Learn Six Sigma

A Lean Six Sigma Black Belt Training Guide

Featuring Examples Using JMP v.13

Lean Sigma Corporation™

Michael Parker

LEARN SIX SIGMA
USING JMP V.13

A LEAN SIX SIGMA BLACK BELT TRAINING GUIDE
FEATURING EXAMPLES FROM JMP V.13

Lean Sigma Corporation

Michael Parker

Lean Sigma Corporation
124 Floyd Smith Office Park Dr.
Suite 325
Charlotte, NC 28262
www.LeanSigmaCorporation.com

Printed in the United States of America
Library of Congress Control Number: 2015906455
U.S. Copyright Registration: TX0007533360
ISBN: 978-0-578-14745-1

Author
Michael Parker

Copy Editor
Cristina Neagu

Other Contributors
Dan Yang
Brian Anderson
Emily Plate
Denise Coleman
Ashley Burnett

Ordering Information:
www.LeanSigmaCorporation.com
980-209-9674

Due to the variable nature of the internet, web addresses, urls, and any other website references such as hyperlinks to data files or support files may have changed since the publication of this book and may no longer be valid.

Table of Contents

0.0 INTRODUCTION

This book has been written to explain the topics of Lean Six Sigma and provide step by step instructions on how perform key statistical analysis techniques using JMP. The content of this book has been updated to include instructions using JMP version 13.

This book will provide the reader with all the necessary knowledge and techniques to become an effective Lean Six Sigma practitioner. For those who have already achieved certification this book is an excellent reference manual as well as companion for all who pass on the virtues of Lean Six Sigma.

 Another valuable component to this publication is the use of numerous step by step analysis instructions for the reader to learn exactly how to perform and interpret statistical analysis techniques using JMP. Anywhere throughout this book where you see the image of a chart on a clipboard followed by the words "Data file:" will be a place where the reader will need the necessary data file to accurately follow the exercise. This data file can be downloaded at https://www.leansigmacorporation.com/support-files/sample-data.xlsx.

1.0 DEFINE PHASE

1.1 SIX SIGMA OVERVIEW

1.1.1 WHAT IS SIX SIGMA?

In statistics, *sigma* (σ) refers to standard deviation, which is a measure of variation. You will come to learn that variation is the enemy of any quality process; it makes it much more difficult to meet a customer's expectation for a product or service. We need to understand, manage, and minimize process variation.

Six Sigma is an aspiration or goal of process performance. A Six Sigma goal is for a process average to operate approximately 6σ away from the customer's high and low specification limits. A process whose average is about 6σ away from the customer's high and low specification limits has abundant room to "float" before approaching the customer's specification limits.

Most people think of Six Sigma as a disciplined, data-driven approach to eliminating defects and solving business problems. If you break down the term, Six Sigma, the two words describe a measure of quality that strives for near perfection.

A Six Sigma process only yields 3.4 defects for every 1 million opportunities! In other words, 99.9997% of the products are defect-free, but some processes require more quality and some require less.

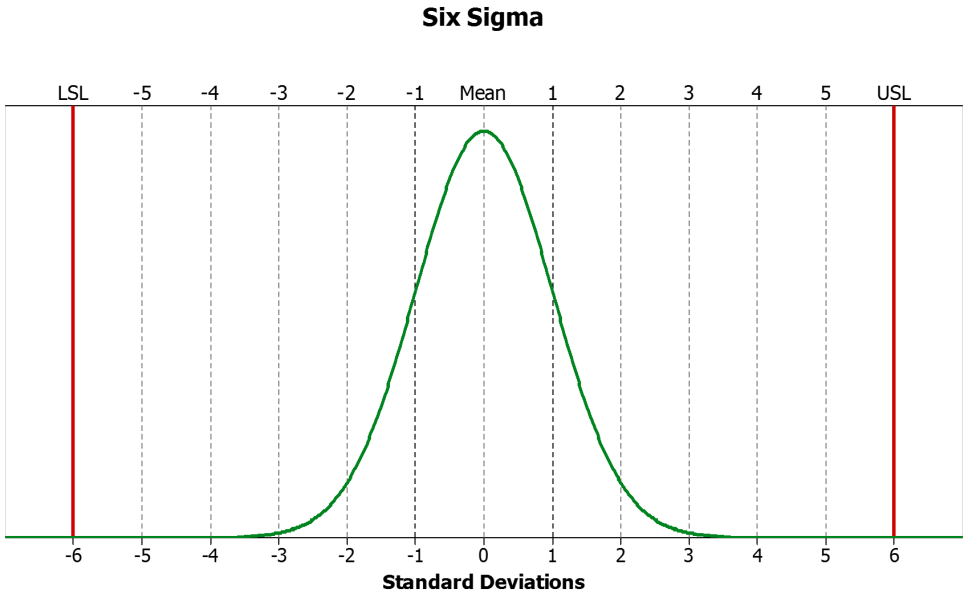


Fig. 1.1 Six Sigma Process with mean 6 standard deviations away from either specification limit.

The more variation that can be reduced in the process (by narrowing the distribution), the more easily the customer's expectations can be met.

It is important to note that a Six Sigma level of quality does not come without cost, so one must consider what level of quality is needed or acceptable and how much can be spent on resources to remove the variation.

What is Six Sigma: Sigma Level

Sigma level measures how many sigma there are between your process average and the nearest customer specification. Let us assume that your customer's upper and lower specifications limits (USL and LSL) were narrower than the width of your process spread. The USL and LSL below stay about one standard deviation away from the process average. Therefore, this process operates at *one sigma*.

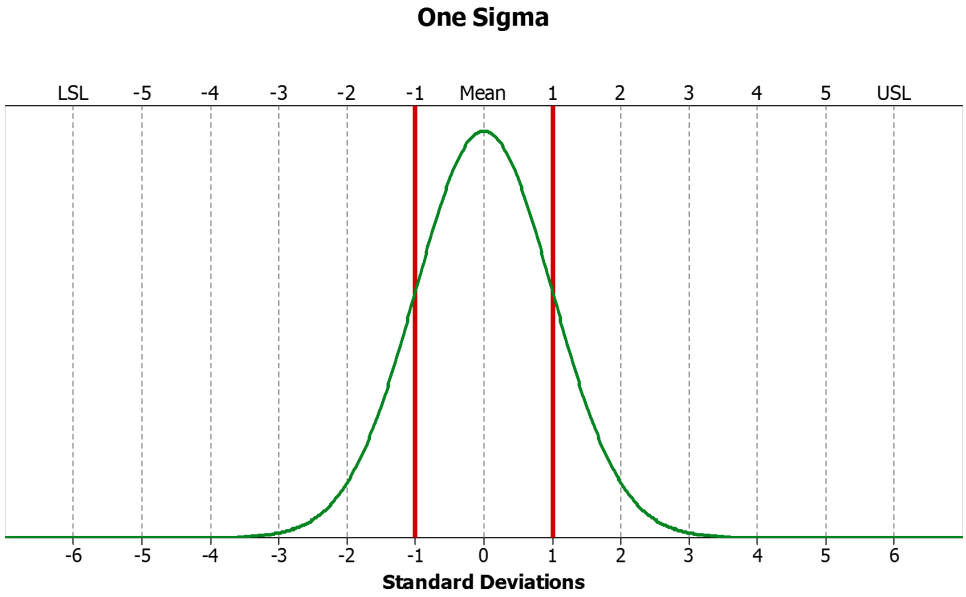


Fig. 1.2 Process operating at 1 sigma with mean one standard deviation away from closest spec limit.

In Fig. 1.1, the LSL and USL were at six standard deviations from the mean. Would that process be more or less forgiving than the one shown in Fig. 1.2?

Answer: *This* process is much less forgiving. It is a one sigma process because the USL and LSL are only one standard deviation from the mean. The area under the blue curve to the left of the LSL and to the right of the USL represents *process defects*.

A process operating at one sigma has a defect rate of approximately 70%. This means that the process will generate defect-free products only 30% of the time, so for every three units that are good, seven are defective. Obviously, a one sigma process is not desirable anywhere because its implications are high customer claims, high process waste, lost reputation, and many others.

What about processes with more than one sigma level? A higher sigma level means a lower defect rate. Let us look at the defect rates of processes at different sigma levels.

Table 1.1 shows each sigma level's corresponding defect rate and DPMO (defects per million opportunities). The higher the sigma level, the lower the defective rate and DPMO. The table shows how dramatically the quality can improve when moving from a one sigma process, to two-sigma, to three, and so on. Next, we will look at how this concept applies to some everyday processes.

Sigma Level	Defect Rate	DPMO
1	69.76%	697612
2	30.87%	308770
3	6.68%	66810
4	0.62%	6209
5	0.023%	232
6	0.00034%	3.4

Table 1.1 Sigma level, Defect Rate and DPMO

Let us look at processes operating at three-sigma, which has a defect rate of approximately 7%. What would happen if processes operated at three-sigma? According to http://www.qualityamerica.com:

- Virtually no modern computer would function.
- 10,800,000 health care claims would be mishandled each year.
- 18,900 US savings bonds would be lost every month.
- 54,000 checks would be lost each night by a single large bank.
- 4,050 invoices would be sent out incorrectly each month by a modest-sized telecommunications company.
- 540,000 erroneous call details would be recorded each day from a regional telecommunications company.
- 270 million erroneous credit card transactions would be recorded each year in the United States.

Just imagine what a three-sigma process would look like in the tech industry, in health care, in banking. Can you apply this concept to a process that you relate to? What if processes operated with 1% defect rate? There would be:

- 20,000 lost articles of mail per hr. (Implementing Six Sigma – Forest W. Breyfogle III).
- Unsafe drinking water almost 15 minutes per day.
- 5,000 incorrect surgical operations per week.
- Short or long landings at most major airports each day.
- 200,000 wrong drug prescriptions each year.
- No electricity for almost seven hours per month.

Even at a 1% defect rate, some processes would be unacceptable to you and many others. Would you drink the water in a community where the local water treatment facility operates at a 1% defect rate? Would you go to a surgeon with a 1% defect rate?

So, what is Six Sigma? Sigma is the measure of quality, and Six is the *goal*.

What is Six Sigma: The Methodology

Six Sigma itself is the *goal*, not the method. To achieve Six Sigma, you need to improve your process performance by:

- Minimizing the process variation so that your process has enough room to fluctuate within customer's spec limits.
- Shifting your process average so that it is centered between your customer's spec limits.

Accomplishing these two process improvements, along with stabilization and control (the ability to maintain the process improvements or level of quality over time), you can achieve Six Sigma. This way, the process becomes more capable of meeting the specification limits that are set by the customer.

The methodology prescribed to achieve Six Sigma is called DMAIC. DMAIC is a systematic and rigorous methodology that can be applied to *any* process in order to achieve Six Sigma. It consists of five phases of a project:

1. **D**efine
2. **M**easure
3. **A**nalyze
4. **I**mprove
5. **C**ontrol

You will be heavily exposed to many concepts, tools, and examples of the DMAIC methodology through this training. The goal of this training is to teach you how to apply DMAIC and leverage the many concepts and tools within it. At the completion of the curriculum, you will become capable of applying the DMAIC methodology to improve the performance of *any* process.

1.1.2 Six Sigma History

The Six Sigma terminology was originally adopted by Bill Smith at Motorola in the late 1980s as a quality management methodology. As the "Father of Six Sigma," Bill forged the path for Six Sigma through Motorola's CEO Bob Galvin who strongly supported Bill's passion and efforts.

Bill Smith originally approached Bob Galvin with what he referred to as the "theory of latent defect." The core principle of the theory is that *variation* in manufacturing processes is the main driver for *defects*, and eliminating variation will help eliminate defects. In turn, it will eliminate the wastes associated with defects, saving money and increasing customer satisfaction. The threshold agreed to by Motorola was 3.4 defects per million opportunities. Does it sound familiar?

Starting from the late 1980s, Motorola extensively applied Six Sigma as a process management discipline throughout the company, leveraging Motorola University. In 1988, Motorola was

recognized with the prestigious Malcolm Baldrige National Quality Award for its achievements in quality improvement.

Six Sigma has been widely adopted by companies as an efficient way of improving the business performance since General Electric implemented the methodology under the leadership of Jack Welch in the 1990s. As GE connected Six Sigma results to its executive compensation and published the financial benefits of Six Sigma implementation in their annual report, Six Sigma became a highly sought-after discipline of quality. Companies across many industries followed suit. In some cases GE taught companies how to deploy the methodology, and in many cases experts from GE and other pioneer companies were heavily recruited to companies that were new to methodology.

Most Six Sigma programs cover the aspects, tools, and topics of Lean or Lean Manufacturing. The two, work hand in hand, benefitting each other. Six Sigma focuses on minimizing process variability, shifting the process average, and delivering within customer's specification limits. Lean focuses on eliminating waste and increasing efficiency.

While the term Lean was coined in the 1990s, the methodology is mainly based on the concepts of the Toyota Production System, which was developed by Taiichi Ohno, Shigeo Shingo, and Eiji Toyoda at Toyota between 1948 and 1975.

Lean and its popularity began to form and gain significant traction in the mid-1960s with the Toyota initiative TPS or Toyota Production System, originally known as the "just-in-time production system." The concepts and methodology of Lean, however, were fundamentally applied much earlier by both Ford and Boeing in the early 1900s.

The DMAIC methodology is essentially an ordered collection of concepts and tools that were not unique to Six Sigma. The concepts and tools are leveraged in a way that leads down a problem-solving path.

Despite the criticism and immaturity of Six Sigma in many aspects, its history continues to be written with every company and organization striving to improve its business performance. The track record for many of these companies after implementing Six Sigma, speaks for itself.

1.1.3 Six Sigma Approach

Six Sigma Approach: Y = f(x)

The Six Sigma approach to problem solving uses a transfer function. A *transfer function* is a mathematical expression of the relationship between the inputs and outputs of a system. The relational transfer function that is used by all Six Sigma practitioners is:

$$Y = f(x)$$

The symbol Y refers to the measure or output of a process. Y is usually your primary metric, the measure of process performance that you are trying to improve. The expression f(x) means

function of x, where x's are factors or inputs that affect the Y. Combined, the Y = f(x) statement reads "Y is a function of x," in simple terms, "My process performance is dependent on certain x's."

The objective in a Six Sigma project is to identify the critical x's that have the most influence on the output (Y) and adjust them so that the Y improves. It is important to keep in mind this concept throughout the DMAIC process. The system (or process) outputs (Y's) are a function of the inputs (x).

Example
Let us look at a simple example of a pizza delivery company that desires to meet customer expectations of on-time delivery.

Measure = on-time pizza deliveries
Y = percent of on-time deliveries

f(x) would be the x's or factors that heavily influence timely deliveries

x1: might be traffic
x2: might be the number of deliveries per driver dispatch
x3: might be the accuracy of directions provided to the driver
x4: might be the reliability of the delivery vehicle

The statement Y = f(x) in this example will refer to the proven x's determined through the steps of a Six Sigma project. Most would agree that an important requirement of pizza delivery is to be fast or on-time. In this example, let us consider factors that might cause variation in Y (percent of on-time deliveries). In the DMAIC process, the goal is to determine which inputs (x's) are the main drivers for the variation in Y.

With this approach, all potential x's are evaluated throughout the DMAIC methodology. The x's should be narrowed down until the vital few x's that significantly influence on-time pizza deliveries are identified.

This approach to problem solving will take you through the process of determining all potential x's that *might* influence on-time deliveries and then determining through measurements and analysis which x's *do* influence on-time deliveries. Those significant x's become the ones used in the Y = f(x) equation.

The Y = f(x) equation is a very powerful concept and requires the ability to measure your output and quantify your inputs. Measuring process inputs and outputs is crucial to effectively determining the significant influences to any process.

1.1.4 Six Sigma Methodology

Six Sigma is a data-driven methodology for solving problems, improving, and optimizing business problems. Six Sigma follows a methodology that is conceptually rooted in the

principles of a five-phase project. Each phase has a specific purpose and specific tools and techniques that aid in achieving the phase objectives.

The five phases of DMAIC are:

1. **D**efine
2. **M**easure
3. **A**nalyze
4. **I**mprove
5. **C**ontrol

The goals of the five phases are:

1. **D**EFINE—To define what the project is setting out to do and scope the effort
2. **M**EASURE—To establish a baseline for the process, ensure the measurement system is reliable, and identify all possible root causes for a problem
3. **A**NALYZE—To narrow down all possible root causes to the critical few that are the primary drivers of the problem
4. **I**MPROVE—To develop the improvements for the process
5. **C**ONTROL—To implement the fix and a control plan to ensure the improvements are sustained over time

In terms of the transfer function, the five phases mean:

1. **D**EFINE—Understand the project Y's and how to measure them
2. **M**EASURE—Prioritize potential x's and measure x's and Y's
3. **A**NALYZE —Test x–Y relationships and verify/quantify important x's
4. **I**MPROVE—Implement solutions to improve Y's and address important x's
5. **C**ONTROL—Monitor important x's and the Y's over time

Six Sigma Methodology: Define Phase

The goal of the *Define* phase is to establish a solid foundation and business case for a Six Sigma project. Define is arguably the most important aspect of any Six Sigma project. It sets the foundation for the project and, if it is not done well and properly thought, it is very easy for a project to go off-track.

All successful projects start with a current state challenge or problem that can be articulated in a quantifiable manner—without a baseline and a goal, it is very difficult to keep boundaries around the project. But it is not enough to just know the problem, you must quantify it and also determine the goal. Without knowing how much improvement is needed or desired, it can be very difficult to control the scope of the project.

Once problems and goals are identified and quantified, the rest of the Define phase will be about valuation, team, scope, project planning, timeline, stakeholders, Voice of the Customer, and Voice of the Business.

Define Phase Tools and Deliverables
1. *Project charter*, which establishes the:

- Business Case
- Problem Statement
- Project Objective
- Project Scope
- Project Timeline
- Project Team

The project charter is essentially a contract formed between a project team, the champion, and the stakeholders. It defines what the project is going to do, why they are going to do it, when it will be done, and by whom. It includes the business case, problem statement, project objective, scope, timeline, and team.

2. *Stakeholder Assessment*

Stakeholder assessment involves the following:

- High-Level Pareto Chart Analysis
- High-Level Process Map
- Voice of the Customer/ Voice of the Business and Critical to Quality Requirements Identified and Defined
- Financial Assessment

A stakeholder assessment is done to understand where there are gaps in stakeholder support and develop strategies to overcome them. High-level Pareto chart and process maps, along with Voice of the Customer, Voice of the Business, and Critical to Quality Requirements also help to develop the scope and put some guardrails around the process.

Six Sigma Methodology: Measure Phase

The goal of the *Measure* phase is to gather baseline information about the process (process performance, inputs, measurements, customer expectations etc.). This phase is necessary to determine if the measurement systems are reliable, if the process is stable, and how capable the process is of meeting the customer's specifications.

Throughout the Measure phase you will seek to achieve a few important objectives:

- Gather All Possible x's
- Assess Measurement System and Data Collection Requirements
- Validate Assumptions
- Validate Improvement Goals
- Determine Cost of Poor Quality
- Refine Process Understanding

- Determine Process Stability
- Determine Process Capability

Measure Phase Tools and Deliverables
The tools and deliverables for this phase are:

- Process Maps, SIPOC, Value Stream Maps—To visualize a process
- Failure Modes and Effects Analysis—To identify possible process failure modes and prioritize them
- Cause-and-Effect Diagram—To brainstorm possible root causes for defects
- XY Matrix—To prioritize possible root causes for defects
- Six Sigma Statistics
- Basic Statistics and Descriptive Statistics—To understand more about your process data
- Measurement Systems Analysis—To establish repeatability, reproducibility, linearity, accuracy, and stability
- Variable and/or Attribute Gage R&R
- Gage Linearity and Accuracy or Stability
- Basic Control Charts—To assess process stability
- Process Capability and Sigma Levels—To assess and quantify a process' ability to meet customer specification limits
- Data Collection Plan—To ensure that when data is collected, it is done properly

Six Sigma Methodology: Analyze Phase

The *Analyze* phase is all about establishing verified drivers. In the DMAIC methodology, the Analyze phase uses statistics and higher-order analytics to discover relationships between process performance and process inputs. In other words, this phase is where the potential root causes are narrowed down to the *critical* root causes.

Statistical tools are used to determine whether there are relationships between process performance and the potential root causes. Where the strong relationships are discovered, these become the foundation for the solution to improve the process.

Ultimately, the Analyze phase establishes a reliable hypothesis for improvement solutions. During the Analyze phase, one needs to:

- Establish the Transfer Function Y = f(x)
- Validate the List of Critical x's and Impacts
- Create a Beta Improvement Plan (e.g., pilot plan)

Analyze Phase Tools and Deliverables
The Analyze phase is about proving and validating critical x's using the appropriate and necessary analysis techniques. The tools that help to formulate a hypothesis about how much improvement can be expected in the Y, given a change in the x include:

- Hypothesis Testing (e.g., t-tests, Chi-Square)
- Parametric and Non-Parametric
- Regression (Simple Linear Regression, Multiple Linear Regression)—To establish quantitative and predictive relationships between the x's and Y's

The Analyze phase is also about establishing a set of solution hypotheses to be tested and further validated in the Improve phase.

Six Sigma Methodology: Improve Phase

The goal of the *Improve* phase is make the improvement. Improve is about designing, testing, and implementing your solution. To this point, you have defined the problem and objective of the project, brainstormed possible x's, analyzed and verified critical x's. Now it's time to make it real! During the Improve phase, the following are necessary:

- Statistically Proven Results from Active Study/Pilot
- Improvement/Implementation Plan
- Updated Stakeholder Assessment
- Revised Business Case with Return on Investment
- Risk Assessment/Updated Failure Modes and Effects Analysis
- New Process Capability and Sigma

Methodologies for the Improve phase include:

- Experiments and planned studies
- Pilots or tests designed to validate relationships and determine how much change in an input is needed to induce the desired result in the output
- Implementation plan to stimulate thoughts and planning for any necessary communications, training, and preparations to implement the process improvement

Improve Phase Tools and Deliverables
The tools and deliverables of this phase are:

- Any appropriate tool from previous phases
    - An updated stakeholder assessment to ensure the right support exists to make the process change
    - An updated business case to justify the change
    - An updated Failure Modes and Effects Analysis to ensure the solution is robust and proper control points are identified
    - A revisited process capability and sigma level to quantify how well the improved process will perform against customer specifications
- Design of Experiment: Full Factorial and Fractional Factorial
- Pilot or Planned Study using Hypothesis Testing and Valid Measurement Systems
- Implementation Plan

Six Sigma Methodology: Control Phase

The last of the five core phases of the DMAIC methodology is the *Control* phase. The purpose of this phase is to establish the mechanisms to ensure the process sustains improved performance and, if it does not, to establish a reaction/mitigation plan to activate by a process owner. The Control phase is a common failure point for projects. Improvements are made, but gains will only be temporary if the proper controls are not in place.

Control Phase Tools and Deliverables
The tools and deliverables of this phase are:

- Statistical Process Control (SPC/Control Charts): IMR, Xbar-S, Xbar-R, P, NP, U, C etc.—For monitoring process inputs and outputs for stability and quickly identifying when the process goes "out of control"
    - Control Plan Documents
    - Control Plan—To ensure control points are identified and accountability for monitoring them is taken
    - Training Plan—To ensure employees are properly trained to perform process changes
    - Communication Plan—To alert any stakeholders of a change
    - Audit Checklist
- Lean Control Methods—To mistake-proof a process
    - Poka-Yoke
    - 5-S—To organize the workplace and make it more visual
    - Kanban

1.1.5 Roles and Responsibilities

The various roles in a Six Sigma program are commonly referred to as "Belts." In addition to Belts, there are also other key roles with specific responsibilities.

It is important to know the many roles in Six Sigma to understand where responsibilities lie to execute a project. Let us explore the different roles and their corresponding responsibilities in a Six Sigma program. Each of the four Six Sigma belts represents a different level of expertise in the field of Six Sigma:

- Six Sigma Master Black Belt
- Six Sigma Black Belt
- Six Sigma Green Belt
- Six Sigma Yellow Belt

Different Belt levels are associated with differing levels of expertise in Six Sigma, with Master Black Belt being the most highly trained expert, to Yellow Belt having very basic knowledge. In addition to Belts, there are other critical and complementary roles, very important to the success of Six Sigma initiatives:

- Champions
- Sponsors
- Stakeholders
- Subject Matter Experts

Roles and Responsibilities: Master Black Belt

The *Master Black Belt* (MBB) is the most experienced, educated, and capable Six Sigma expert, often thought of as a trusted advisor to high-level leaders.

A typical MBB has managed dozens of Black Belt level projects. An MBB can simultaneously lead multiple Six Sigma Belt projects, ensuring the proper and consistent application of Six Sigma across departments, while mentoring and certifying Black Belt and Green Belt candidates.

An MBB typically works with high-level operations directors, senior executives, and business managers to help with assessing and planning business strategies and tactics. This is why the MBB needs to have a strong ability to communicate with and influence leaders at all levels of an organization.

An MBB commonly advises management team on the cost of poor quality of an operation and consults on methods to improve business performance.

While an MBB is typically technically savvy with all of the concepts and tools, it is critical that he or she have the skills to communicate in a practical manner with those that are not well-versed in Six Sigma.

Typical Responsibilities of an MBB
An MBB:

- Identifies and defines the portfolio of projects required to support a business strategy
- Establishes scope, goals, timelines, and milestones
- Assigns and marshals resources
- Trains and mentors Green Belts and Black Belts
- Facilitates tollgates or checkpoints for Belt candidates
- Reports-out/updates stakeholders and executives
- Establishes organization's Six Sigma strategy/roadmap
- Leads the implementation of Six Sigma

Roles and Responsibilities: Black Belt

The *Black Belt* (BB) is the most active and valuable experienced Six Sigma professional among all Six Sigma Belts. The Black Belt is the key to success among all the other Belts.

A typical BB has:

- Led multiple projects
- Trained and mentored various Green Belts candidates
- Understood how to define a problem and drive effective solution

The BB incorporates many skills that are critical to successfully and quickly implementing a process improvement initiative through the DMAIC methodology. The BB is well rounded in terms of project management, statistical analysis, financial analysis, meeting facilitation, prioritization, and a range of other value-added capabilities, which makes a BB highly valuable asset in the business world.

BBs commonly serve as the dedicated resource continuing their line management role while simultaneously achieving a BB certification.

Typical Responsibilities of a BB
A BB has the following responsibilities:

- Project Management
    - Defines projects, scope, teams etc.
    - Marshals resources
    - Establishes goals, timelines, and milestones
    - Provides reports and/or updates to stakeholders and executives
- Task Management
    - Establishes the team's Lean Sigma roadmap
    - Plans and implements the use of Lean Sigma tools
    - Facilitates project meetings
    - Does project management of the team's work
    - Manages progress toward objectives
- Team Management
    - Chooses or recommend team members
    - Defines ground rules for the project team
    - Coaches, mentors, and directs project team
    - Coaches other Six Sigma Belts
    - Manages the team's organizational interfaces

Roles and Responsibilities: Green Belt

The *Green Belt* (GB) is considered as a less intense version of Six Sigma professional compared to the Black Belt (BB). A GB is exposed to all the comprehensive aspects of Six Sigma with less focus on the statistical theories and some other advanced analytical methodologies such as Design of Experiment. When it comes to project management, a GB has almost the same responsibilities as a BB. In general, the GB works on less complicated and challenging business problems than a BB.

While a Black Belt's sole business responsibility could be the implementation of projects, Green Belts often implement projects in addition to other job responsibilities. Therefore, the

Green Belt typically works on less complicated business problems. A Green Belt takes direction and coaching from the Black Belt.

Typical Responsibilities of a Green Belt
A GB has the following responsibilities:

- Project Management
    - Defines the project, scope, team etc.
    - Marshals resources
    - Sets goals, timelines, and milestones
    - Reports-out/updates stakeholders and executives
- Task Management
    - Establishes the team's Lean Sigma Roadmap
    - Plans and implements the use of Lean Sigma tools
    - Facilitates project meetings
    - Does Project Management of the team's work
    - Manages progress toward objectives
- Team Management
    - Chooses or recommends team members
    - Defines ground rules for the project team
    - Coaches, mentors, and directs project team
    - Coaches other Six Sigma Belts
    - Manages the team's organizational interfaces

As you can see, the roles and responsibilities are identical to those of a Black Belt. The key differences are the level of complexity and the dedication that a Black Belt has toward Six Sigma projects. Again, a Green Belt typically does projects in addition to his or her normal line job responsibilities.

Roles and Responsibilities: Yellow Belt

The *Yellow Belt* (YB) understands the basic objectives and methods of a Six Sigma project. YB has an elementary understanding about what other Six Sigma Belts (GB, BB, and MBB) are doing to help them succeed.

In a Six Sigma project, YB usually serves as a subject matter expert regarding some aspects of the process or project. Supervisors, managers, directors, and sometimes executives are usually trained at the YB level. This level of training is important to have in an environment where Six Sigma projects are prevalent, such as when managing Belts.

Typical Responsibilities of a Yellow Belt
A Yellow Belt:

- Helps define process scope and parameters
- Contributes to team selection process

- Assists in information and data collection
- Participates in experiential analysis sessions (Failure Modes and Effects Analysis, Process Mapping, Cause and Effect etc.)
- Assists in assessing and developing solutions
- Delivers solution implementations

Roles and Responsibilities: Champions and Sponsors

Champions and sponsors are those individuals (directors, executives, managers etc.) chartering, funding, or driving the Six Sigma projects that BBs and GBs are conducting. Champions and Sponsors' roles are to set direction for projects and ensure that the path is clear for the Belts to succeed: obtain resources and funding, overcome stakeholder issues, and knock down other barriers to progress.

Champions and sponsors need to have a basic understanding of the concepts, tools, and techniques involved in the DMAIC methodology so that they can provide proper support and direction. Champions and sponsors play critical roles in the successful deployment of Six Sigma. Strong endorsement of Six Sigma from the leadership team is critical for success.

Typical Responsibilities of a Champion or Sponsor
A Champion or Sponsor typically:

- Maintains a strategic oversight
- Establishes strategy and direction for a portfolio of projects
- Clearly defines success
- Provides resolution for issues such as resources or politics
- Establishes routine tollgates or project reviews
- Clears the path for solution implementation
- Assists in project team formation

From a project standpoint, it is important for the Champions and Sponsors to be highly engaged in the work. It is their role to provide oversight and guidance to define success for the project or projects. They resolve issues when necessary, such as resource needs, stakeholder conflicts, or political issues. They create the routines for tollgate reviews to ensure there are regular checkpoints with the project team.

Roles and Responsibilities: Stakeholders

Stakeholders are usually the recipients or beneficiaries of the success of a Six Sigma project. Generally speaking, a stakeholder is anyone who has interest in a project; stakeholders are usually the beneficiaries of the process improvement. A stakeholder can be a person, a group, or an organization that is affected by the project.

Stakeholders are individuals owning the process, function, or production/service line that a Six Sigma Belt focuses on improving the performance of.

BBs and GBs need to keep strong working relationships with stakeholders because without their support, it would be extremely difficult to make the Six Sigma project a success. A lack of stakeholder support can cause a project to fail, either through active or passive resistance, so it is very important to be aware of how stakeholders feel about the work being done so the team can proactively manage their perceptions and level of support.

Roles and Responsibilities: Subject Matter Experts

Subject Matter Experts (SMEs) are commonly known as the experts of the process or subject matter. They play critical roles to the success of a project. Six Sigma Belts should proactively look to key SMEs to round out their working project team. Based on SMEs' extensive knowledge about the process, they have the experience to identify which solutions can work and which cannot work.

SMEs who simply do not speak up can hurt the chances of the process' success. SMEs' input is critical to ensure the team is leveraging accurate information when it comes to the process or evaluating data and information.

SMEs are also the same people who prefer to keep the status quo. Six Sigma Belts may find many of them unwilling to help implement the changes. When selecting SMEs, it is important to choose people who are vocal, but who also bring ideas. SMEs that are closed-minded to change may be vocal, but can be a barrier to making the necessary improvements.

SUMMARY

Throughout this module, we have reviewed the various common roles and corresponding responsibilities in any Six Sigma program:

- Six Sigma Master Black Belt
- Six Sigma Black Belt
- Six Sigma Green Belt
- Six Sigma Yellow Belt
- Champion and Sponsors
- Stakeholders
- Subject Matter Experts

These Six Sigma belts and other roles are designed to deliver value to the business effectively and successfully. For a Six Sigma program to be effective, it is important to assign these roles to individuals who are well equipped to carry out the responsibilities.

1.2 SIX SIGMA FUNDAMENTALS

1.2.1 DEFINING A PROCESS

What is a Process Map?

A *process map* is a graphical representation of a process flow. It is necessary to document a process because process maps:

- Visually show how the business process is accomplished in a step by step manner.
- Describe how information or materials sequentially flow from one entity to the next.
- Illustrate who is responsible for steps and actions.
- Depict the input and output of each individual process step.

In the Measure phase, the project team should map the current state of the process. Teams should be sure not to fall into the trap of mapping an ideal state or what they think the process should look like.

Process Map Basic Symbols

Figure 1.3 below depicts a basic process map that was created with the four most commonly used process mapping symbols. The four symbols are:

- Oval – Used for start and end points
- Square or Rectangle – Used for process steps
- Diamond – Used for decision points
- Arrow – Used to represent flow direction

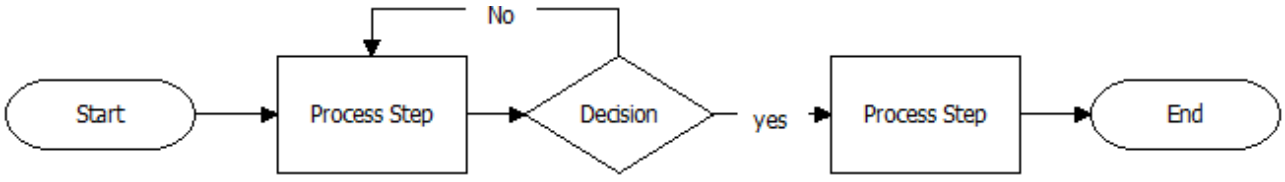


Fig. 1.3 Basic Process Mapping Symbols

Additional Process Mapping Symbols

Below are additional process mapping symbols that may be used to provide further information and detail when developing a process map. Although these symbols are not as common as those shown in figure 1.3, they are universally understood to represent process information as described:


Alternative Process Step: Meaning "there is a different way of performing the step"

 *Predefined Process*: Used to hold the place of a pre-existing process

 *Manual Process Step: A*s opposed to an "automated" step

 *Preparation*: Indicates a preparation operation

 *Delay*: Indicates a waiting period

 *Data (I/O):* Shows the inputs and outputs of a process

 *Document:* Indicates a process step that results in a document

 *Multi-Document:* Indicates a process step that results in multiple documents

 *Stored Data:* Indicates a process step that stores data

 *Magnetic Disk:* Indicates a database

 *Off Page Connector:* Indicates the process flow continues on another page

 *Merge:* Indicates where multiple processes merge into one

 *Extract:* Indicates a process splitting into multiple parallel processes

How to Plot a Process Map

Process Mapping Step 1: Define boundaries of the process you want to map.

- A process map can depict the flow of an entire process or a segment of it.
- You need to identify and define the beginning and ending points of the process before starting to plot.
- Use operational definitions.

The first step in plotting a process map is deciding where the beginning and ending points are for the process you are trying to depict. You might be trying to map a whole process that begins and ends with the customer, or you may be trying to map out a segment of it.

An *operational definition* is an exact description—in other words, it is useful in removing ambiguity and reduces the chances of getting disparate results from a process if performed by different people.

Process Mapping Step 2: Define and sort the process steps with the flow. There are a few ways to do this:

- Consult with process owners and subject matter experts or observe the process in action to understand how the process is performed.
- Record the process steps and sort them according to the order of their occurrence.

Usually, it is a good idea to do both; sometimes hearing what is supposed to happen might be different than what you observe, this often indicating a problem.

Process Mapping Step 3: Fill the step information into the appropriate process symbols and plot the diagram.

- In the team meeting of process mapping, place the sticky notes with different colors on a white board to flexibly adjust the under-construction process map.
- The flow lines are plotted directly on the white board. For the decision step, rotate the sticky note by 45°.
- When the map is completed on the white board, record the map using Excel, PowerPoint, or Visio.

Mapping out the process with sticky notes on a white board is a good practice to do as a team because the map is easy to adjust until finished. Use a dry-erase marker to draw the flow lines between the sticky notes.

Process Mapping Step 4: Identify and record the inputs/outputs and their corresponding specifications for each process step.

- The process map helps in understanding and documenting Y=f(x) of a process where Y represents the outputs and x represents the inputs.
- The inputs of each process step can be controllable or non-controllable, standardized operational procedure, or noise. They are the source of variation in the process and need to be analyzed qualitatively and quantitatively in order to identify the vital few inputs that have significant effect on the outcome of the process.
- The outputs of each process step can be products, information, services, etc. They are the little Y's within the process.

Ideally, the specifications should be recorded for each step so it is understood what the inputs and outputs should look like to ensure process efficiency and quality.

Process Mapping Step 5: Evaluate the process map and adjust it if needed.

- If the process is too complicated to be covered in one single process map, you may create additional detailed sub-process maps for further information.
- Number the process steps in the order of their occurrence for clarity.

Functional Process Map or "Swim Lane"

To illustrate the responsibility of different organizations involved in the process, use a functional process map. This type of map is also commonly referred to as a swim lane map. The value of this type of map is that it adds the dimension of accountability.

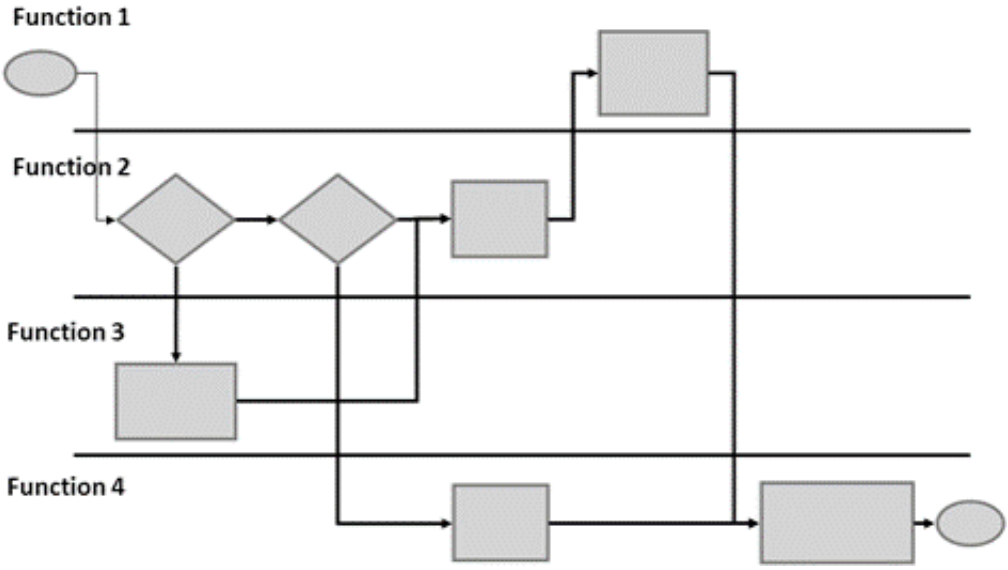


Fig. 1.4 Functional Process Map

Figure 1.4 is a functional process map or "swim lane" diagram which depicts the responsibilities for each step in the process. Swim-lane diagrams can be drawn vertically or horizontally. The swim lanes can show organizations, roles, or functions within a business.

High Level Process Map

Most high-level process maps are also referred to as *flow charts*. The key to a high-level process map is to over-simplify the process being depicted so that it can be understood in its most generic form. As a general rule, high-level process maps should be four–six steps and no more.

Below, figure 1.5a depicts an oversimplified version of a high-level process map for cooking a 10 lb. prime rib for a dozen holiday guests. As mentioned, high level maps are basic and the challenge is summarizing a process well enough to depict it in 4 – 6 steps.

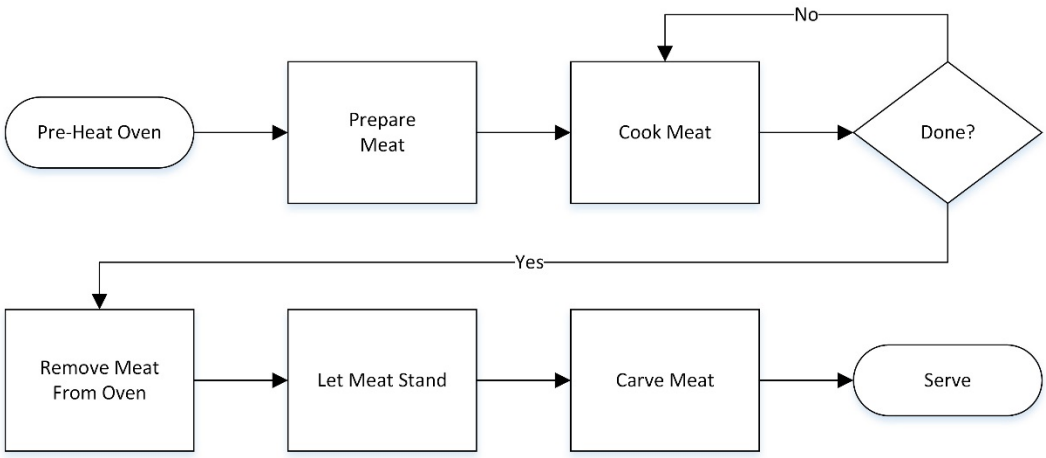


Fig. 1.5a High Level Process Map (cooking prime rib)

Detailed Process Map

Detailed process maps or multi-level maps take the high-level map much further. Detailed maps can be two, three, or more levels deeper than your high-level process map.

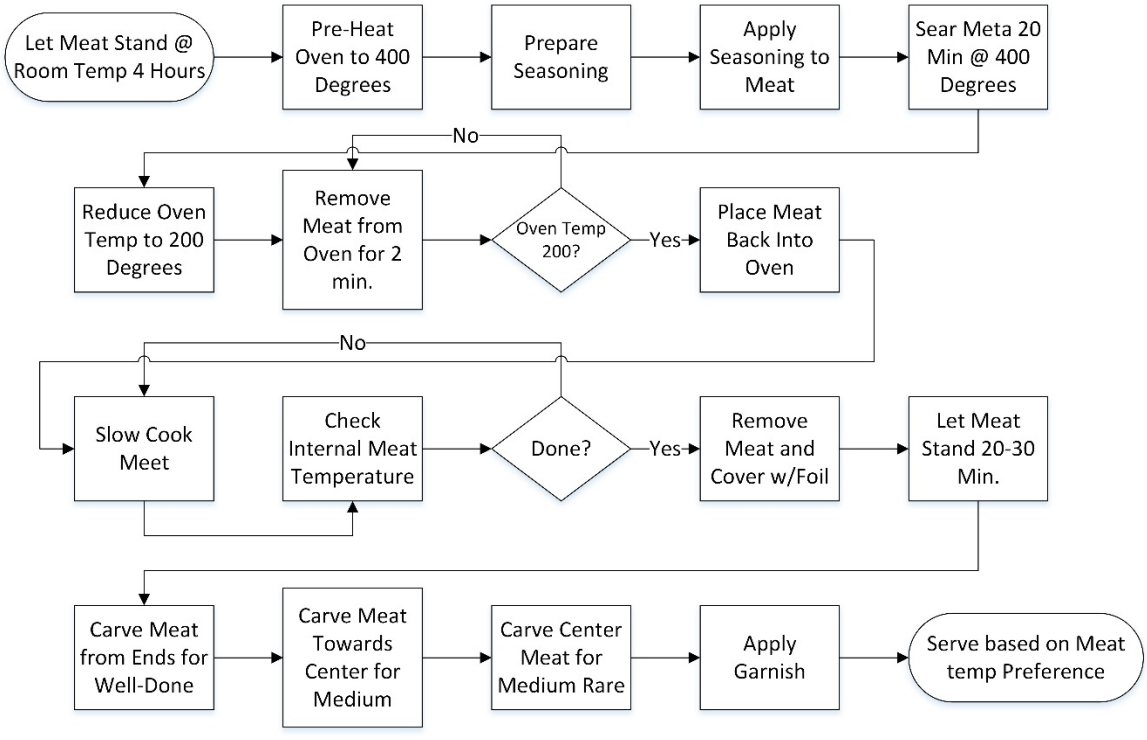


Fig. 1.5b Detailed Process Map (cooking prime rib)

Figure 1.5b is a more detailed representation of our high-level map shown in figure 1.5a. This would be considered a level 2 process map. A good guideline to use to help create the second level is that for each high-level step break it down into another two to four steps each (no more). This is usually a helpful way to index a process, so when you want to understand more about a specific part of a process later, you can find it by starting at the high level and digging

down into the process steps you are interested in. Repeat this process (level 3, level 4 etc.) until reaching the desired level of detail. Some detailed maps are two or three levels deep, others can be five or six levels deep. Obviously, the deeper the levels, the more complex and the more burdensome to create and maintain. However, more detail reveals far more information.

What is SIPOC?

A SIPOC (Suppliers, Inputs, Process, Outputs, and Customers) is a high-level visualization tool to help identify and link the different components in a process. SIPOCs summarize the inputs and outputs of a process.

SIPOCs are usually applied in the Measure phase in order to better understand the current state of the process and define the scope of the project. At the outset of a process improvement effort it is very helpful to provide an overview to people who are unfamiliar or need to become reacquainted with a process.

Key Components of a SIPOC

- **S**uppliers: The vendors who provide the raw material, services, and information. Customers can also be suppliers sometimes.
- **I**nput: The raw materials, information, equipment, services, people, environment involved in the process
- **P**rocess: The high-level sequence of actions and decisions that results in the services or products delivered to the customers
- **O**utput: The services or products delivered to the customers and any other outcomes of the process
- **C**ustomers: The end users or recipients of the services or products. Customers can be external or internal to an organization.

How to Plot a SIPOC Diagram

Creating a SIPOC - The first method:

1. Create a template that can contain the information of the five key components in a clear way.
2. Plot a high-level process map that covers five steps at maximum.
3. Identify the outputs of the process.
4. Identify the receipt of the process.
5. Brainstorm the inputs required to run each process step.
6. Identify the suppliers who provide the inputs.

In the first method of creating a SIPOC, begin with the process steps (five steps at most), determine the outputs, identify who the recipients of the process, then work to the left, identifying the inputs and then the suppliers for those inputs.

Creating a SIPOC - The second method:

1. Create a template that can contain the information of the five key components in a clear way.
2. Identify the receipt of the process.
3. Identify the outputs of the process.
4. Plot a high-level process map that covers five steps at maximum.
5. Brainstorm the inputs required to run each process step.
6. Identify the suppliers who provide the inputs.

In the second method, start with the customer (or the recipient of the process), work your way back by identifying the process outputs, then the high-level process map (five steps maximum), brainstorm the inputs to the process, then identify the suppliers for those inputs.

Creating a SIPOC Diagram

Step 1: Vertically List High-Level Process

If you followed the general rules for a high-level process map, then you should have no more than four to six steps for your process. List those steps in a vertical manner in the middle section labeled "Process" as depicted below.

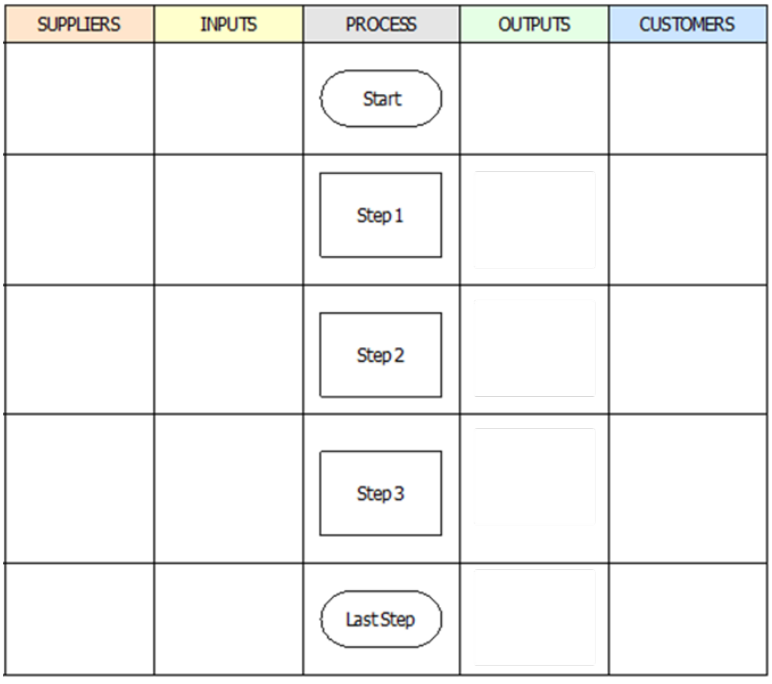


Fig. 1.6 SIPOC High Level Process Map

Step 2: List Process Outputs

In step two, for each process step list any outputs that are generated from the corresponding process step. Outputs can be finished goods, documents, reports, raw materials etc.

SUPPLIERS	INPUTS	PROCESS	OUTPUTS	CUSTOMERS
		Start		
		Step 1	Enter Step 1 Outputs	
		Step 2	Enter Step 2 Outputs	
		Step 3	Enter Step 3 Outputs	
		Last Step	Enter Step 4 Outputs	

Fig. 1.7 SIPOC Outputs

Step 3: List Output Customers

In step 3, list the customers that receive the outputs from each process step. Customers can be internal or external. The important thing to remember is who is the output being delivered to.

SUPPLIERS	INPUTS	PROCESS	OUTPUTS	CUSTOMERS
		Start		
		Step 1	Enter Step 1 Outputs	Enter Step 1 Customers
		Step 2	Enter Step 2 Outputs	Enter Step 2 Customers
		Step 3	Enter Step 3 Outputs	Enter Step 3 Customers
		Last Step	Enter Step 4 Outputs	Enter Step 4 Customers

Fig. 1.8 SIPOC Output Customers

Step 4: List Process Inputs

For the inputs, in step 4 list any inputs that are necessary for any or all process steps. Inputs can be raw materials, information, finished goods from other processes etc.

SUPPLIERS	INPUTS	PROCESS	OUTPUTS	CUSTOMERS
		Start		
	Enter Step 1 Inputs	Step 1	Enter Step 1 Outputs	Enter Step 1 Customers
	Enter Step 2 Inputs	Step 2	Enter Step 2 Outputs	Enter Step 2 Customers
	Enter Step 3 Inputs	Step 3	Enter Step 3 Outputs	Enter Step 3 Customers
	Enter Step 4 Inputs	Last Step	Enter Step 4 Outputs	Enter Step 4 Customers

Fig. 1.9 SIPOC Process Inputs

Step 5: List Suppliers of Inputs

The last step is to identify and list suppliers of the inputs. Who provides the materials or information as inputs to the process? Where do the inputs come from?

SUPPLIERS	INPUTS	PROCESS	OUTPUTS	CUSTOMERS
		Start		
Enter Step 1 Suppliers	Enter Step 1 Inputs	Step 1	Enter Step 1 Outputs	Enter Step 1 Customers
Enter Step 2 Suppliers	Enter Step 2 Inputs	Step 2	Enter Step 2 Outputs	Enter Step 2 Customers
Enter Step 3 Suppliers	Enter Step 3 Inputs	Step 3	Enter Step 3 Outputs	Enter Step 3 Customers
Enter Step 4 Suppliers	Enter Step 4 Inputs	Last Step	Enter Step 4 Outputs	Enter Step 4 Customers

Fig. 1.10 SIPOC Suppliers of Inputs

SIPOC Benefits

A SIPOC has the following benefits, SIPOCs help to:

- Visually communicate project scope
- Identify key inputs and outputs of a process

- Identify key suppliers and customers of a process
- Verify:
    - Inputs match outputs for upstream processes
    - Outputs match inputs for downstream processes.
    - This type of mapping is effective for identifying opportunities for improvement of your process.
- If you have completed your high-level process map, follow the outlined steps to create a process map of **S**uppliers, **I**nputs, **P**rocess, **O**utputs, and **C**ustomer.

SIPOC is more powerful than a simple high-level process map because it conveys the inputs and outputs of a process, *as well as* the suppliers and customers.

What is Value Stream Mapping?

Value stream mapping is a method to visualize and analyze the path of how information and raw materials are transformed into products or services customers receive. It is used to identify, measure, and decrease the non-value-adding steps in the current process.

Non-Value-Added Activities

Non-value-adding activities are activities in a process that, from the customers' perspective, do not add any other value to the products or services customers demand. Examples of non-value-adding activities are:

- Rework—It should have been done right the first time, so a second time certainly does not add value
- Overproduction—Work that is done before it is needed, and could change, leading to rework
- Excess transportation—Moving the product around does not add value
- Excess stock—Similar to overproduction
- Waiting—Wasted time
- Unnecessary motion—Wasted energy

Not all non-value-adding activities are unnecessary. Sometimes, non-value added activities are required. For example, an inspection that is done for regulatory compliance purposes is necessary but does not add value to the product or service the customer receives.

Keys to Plotting a Value Stream Map

Plot the entire high-level process flow from when the customer places the order to when the customer receives the products or services in the end. A value stream map requires more detailed information for each step than the standard process map:

- Cycle time
- Available time
- Demand & Takt Time
- Inventory

- Working time
- Wait time
- Value Added Activities
- Non-Value Added (NVA)
- Preparation type
- Scrap rate
- Rework rate
- Number of operators

Basic Value Stream Map Layout

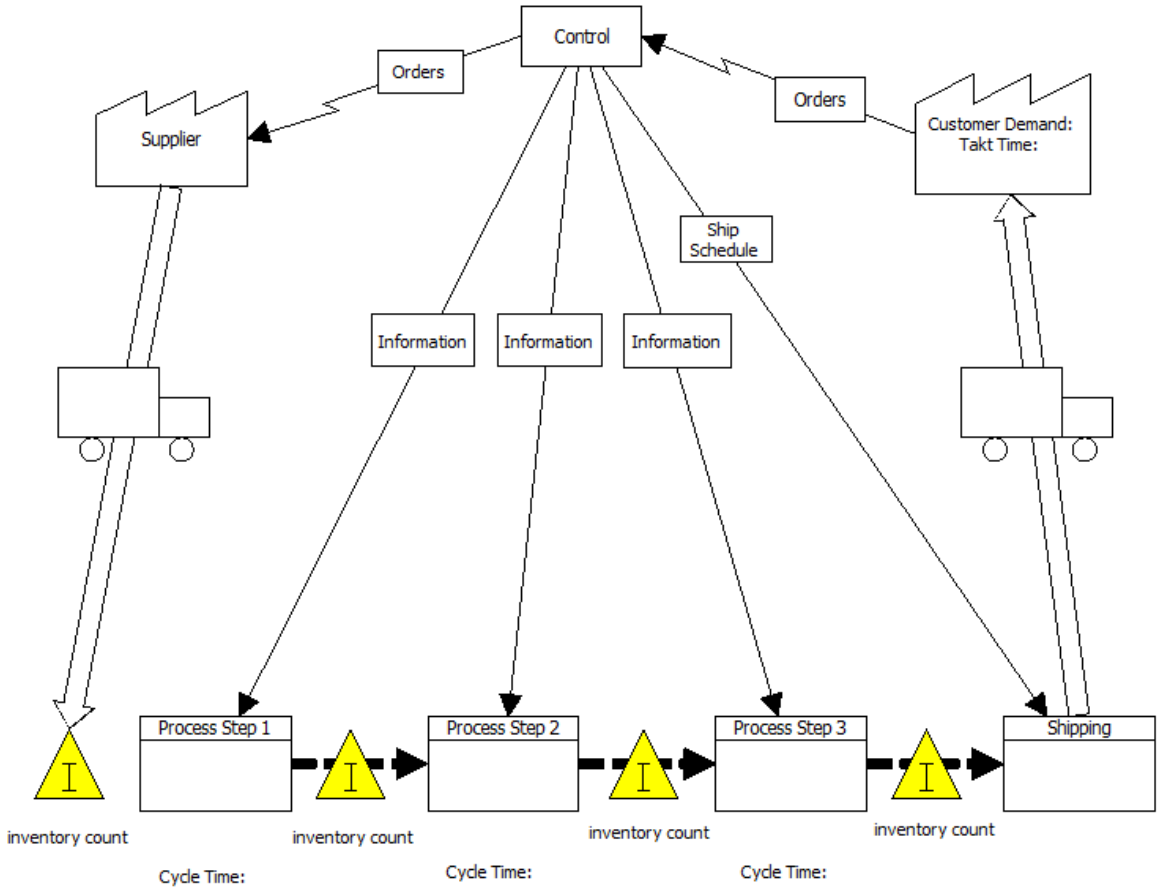


Fig. 1.11 Value Stream Map Layout

Figure 1.11 demonstrates a generic process so that you can see the simple view of the saw-tooth box in the upper right where the customer initiates orders. Orders flow to a control point and which are distributed to supplier orders and throughout the production process of four sub-processes which eventually end with the custom receiving the goods.

Value stream maps are useful to determine:

- What the current process looks like
- Where the process starts and where it ends
- How value flows through the process
- The steps which are value added in the process
- The steps which are non-value added in the process
- What the relationship is between information flow and material flow

- Average cycle time throughout the process
- Inventory levels throughout the process
- Sources of the waste
- Areas in need of the most improvement

The true value of the value stream map is not the map itself but the information gathered and assessed to create the map. This information is vital to ascertaining where inefficiencies and opportunities exist.

Additional Mapping Technique

We will now touch on an additional mapping technique that can come in very handy during a process understanding session; the thought process map.

Thought Process Mapping

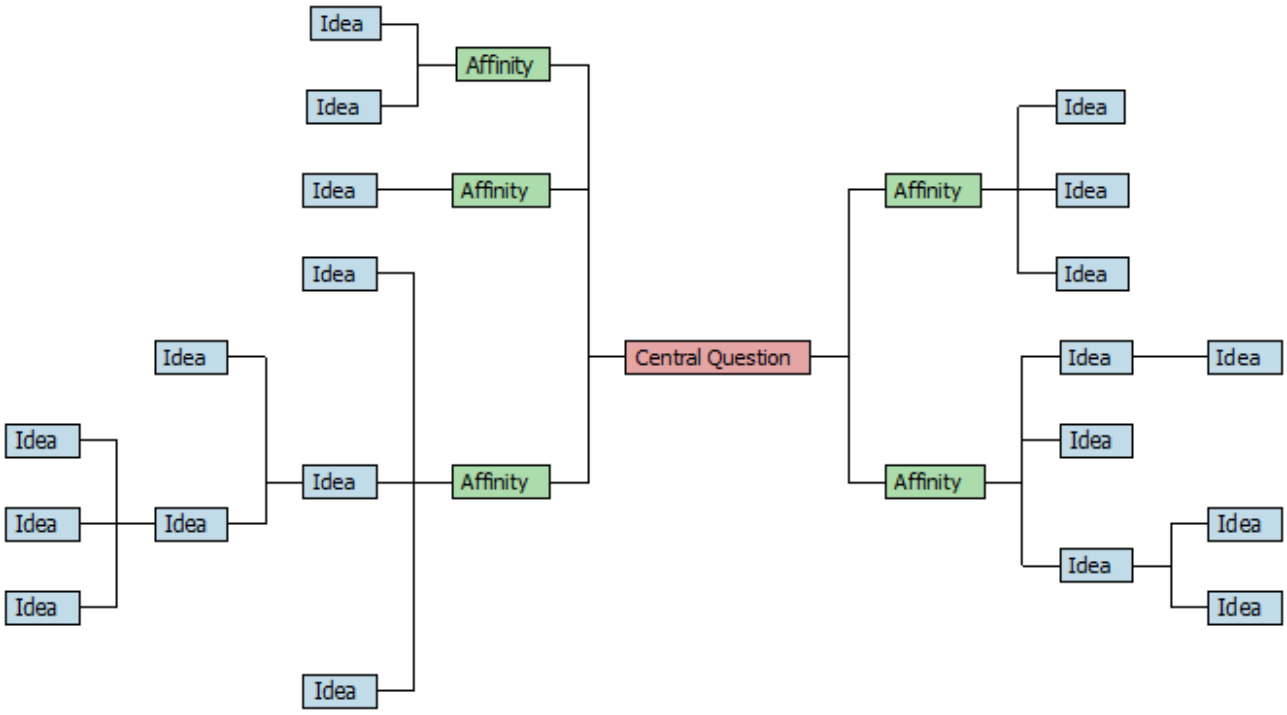


Fig. 1.12 Thought Process Map Example

A *thought process map* is a graphical tool to help brainstorm, organize, and visualize information, ideas, questions, or thoughts regarding reaching a project goal. It helps a project team to understand where they have been, where they still need to go, what questions have been answered, how were they answered, and what is yet to be answered. "Thought Maps" are a popular tool used at any phase of a project to:

- Identify knowns and unknowns
- Communicate assumptions and risks
- Discover potential problems and solutions

- Identify resources, information, and actions required to meet the goal
- Present relationship of thoughts

How to Plot a Thought Process Map

1. Define the project goal.
2. Brainstorm knowns and unknowns about the project.
3. Brainstorm questions and group the unknowns.
4. Sequence the questions below the project goal and link related questions.
5. Identify tools or methods that would be used to answer the questions.
6. Repeat steps 3 to 5 as the project continues until the goal is met.

Thought Process Map Example

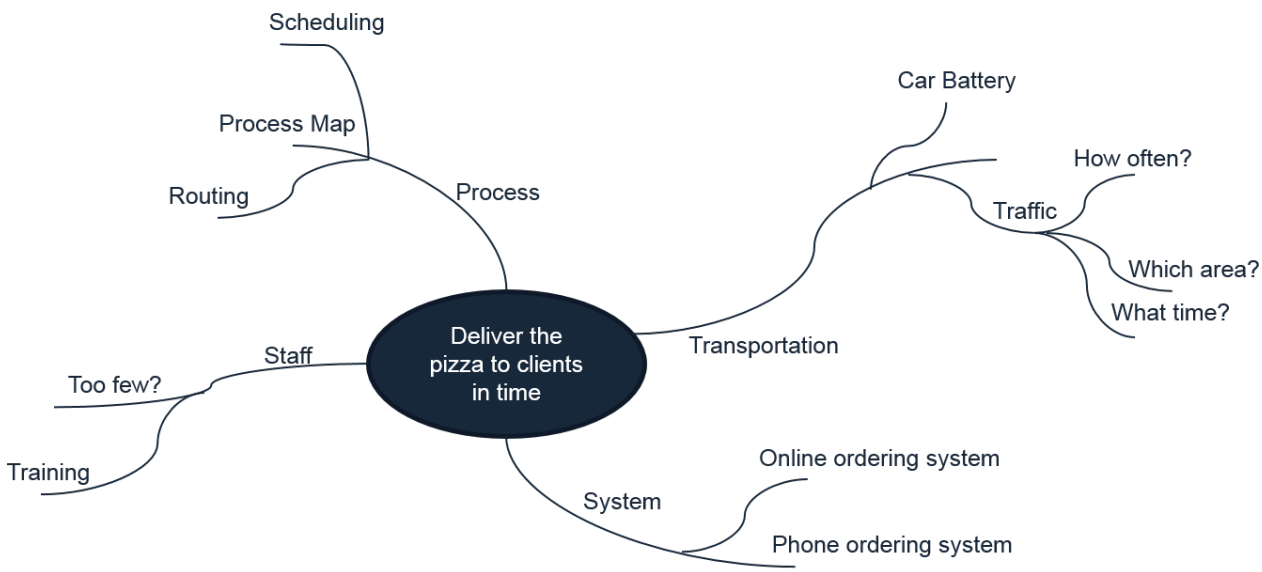


Fig. 1.13 Thought Process Map Example

This is an example of a thought process map. In this case, the main theme is on-time pizza delivery. You can see several themes and concepts explored. Thought process maps serve well to introduce questions and ideas while also being great tools to leave "trails" of where you have been so that later in a project when someone makes a suggestion that has already been evaluated you can refer to the thought map as evidence.

1.2.2 *VOICE OF THE CUSTOMER AND CRITICAL TO QUALITY*

Voice of the Customer

VOC stands for Voice of the Customer. Voice of the customer is a term used for a data-driven plan to discover customer wants and needs. It is *always* important to understand the VOC when designing and improving any Six Sigma process.

There are also other "voices" that need to be heard when conducting projects. The three primary forms are:

- VOC: Voice of the Customer
- VOB: Voice of the Business - What does the business require of a process?
- VOA: Voice of the Associate or Voice of the Employee - What does the employee require of a process (e.g., safety, simplicity)?

Gathering VOC

Gathering VOC should be performed methodically. The two most popular methods of collecting VOC are indirect and direct.

1. Indirect data collection for VOC involves passive information exchange, meaning the VOC was not expressly collected for the purposes of the project, but was gathered through byproducts of the process, such as:

- Warranty claims
- Customer complaints/compliments
- Service calls
- Sales reports

All the above are effective ways to understand how the customer feels about a product or service.

2. Indirect methods are less effective, sometimes dated, require heavy interpretation, and are also more difficult to confirm because the data was not collected to understand VOC specific to a business or process problem. Direct data collection methods for VOC are active and planned customer engagements, such as:

- Conducting interviews
- Conducting customer surveys
- Conducting market research
- Hosting focus groups

Direct methods are typically more effective for several reasons:

- Less need to interpret meaning
- Researchers can go a little deeper when interacting with customers and take a direction that might not have been expected or planned
- Customers are aware of their participation and will respond better upon follow-up
- Researchers can properly plan engagements (questions, sample size, information collection techniques etc.)

There are professional organizations that can perform customer data collection on behalf of a company. They commonly plan carefully by designing the questions, determining the sample size, and considering what technique is most appropriate. If there is a need to get very specific customer feedback, a survey is appropriate. However, if the need is more exploratory, interviews or focus groups are more appropriate.

Gathering VOC

- Gathering VOC requires consideration of many factors such as product or service types, customer segments, manufacturing methods or facilities etc. All this information will influence the sampling strategy because different factors might indicate different customer needs across customer segments, or perhaps the customers' feedback may differ by product type or manufacturing method.
- Consider which factors are important and build a sample size plan around them. Also, consider response rates and adjust the initial sample strategy to ensure adequate input is received.
- Once a sampling plan is in place, collect data via the direct and indirect methods discussed earlier.
- After gathering VOC, it will be necessary to translate it into something meaningful: Critical to Quality.

Critical to Quality

CTQs or *Critical to Quality* are generated from VOC feedback. Because VOC is usually vague, emotional, or simply a generalization about products or services, it is important to translate it into something that is applicable to a process: CTQs, which are the quantifiable, measurable, and meaningful translations of VOC.

To create CTQs, the VOC must be organized first into like groupings. One effective way to organize VOC is to group or bucket it using an *affinity diagram*. Affinity diagrams are ideal for large amounts of soft data resulting from brainstorming sessions or surveys.

Affinity Diagram: Building a CTQ Tree

Steps for conducting an affinity diagram exercise:

1. Clearly define the question or focus of the exercise.
   ("Why are associates late for work?").
2. Record all participant responses on note cards or sticky notes.
   (This is the sloppy part, record everything!).
3. Lay out all note cards or post the sticky notes onto a wall.
4. Look for and identify common themes.
5. Begin moving the notes into the themes until all responses are allocated.
6. Re-evaluate and adjust.

This is an outline of building an affinity diagram (or CTQ tree); it is typically done with a group of people. The group must agree on what question they are trying to answer, or what the focus for the VOC is. It is good to do this exercise in a group because people bring different perspectives and might understand and organize VOC differently than a single person would.

Affinity Diagram: Building a CTQ Tree Step by Step
In summary, to build an affinity diagram:

- Define the question or focus.
- Record responses on note cards or sticky notes.
- Display all note cards or sticky notes on a wall if necessary.

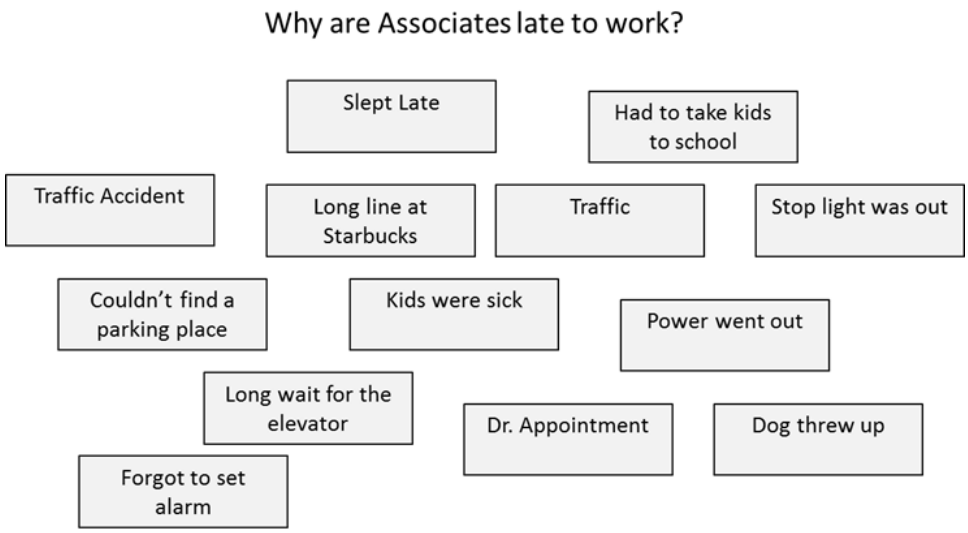


Fig. 1.14 Random reasons prior to an affinity diagram

In figure 1.14 the focus question is "Why are associates late to work?" You can see all of the responses randomly distributed across the page.

- Look for and identify common themes within the responses.

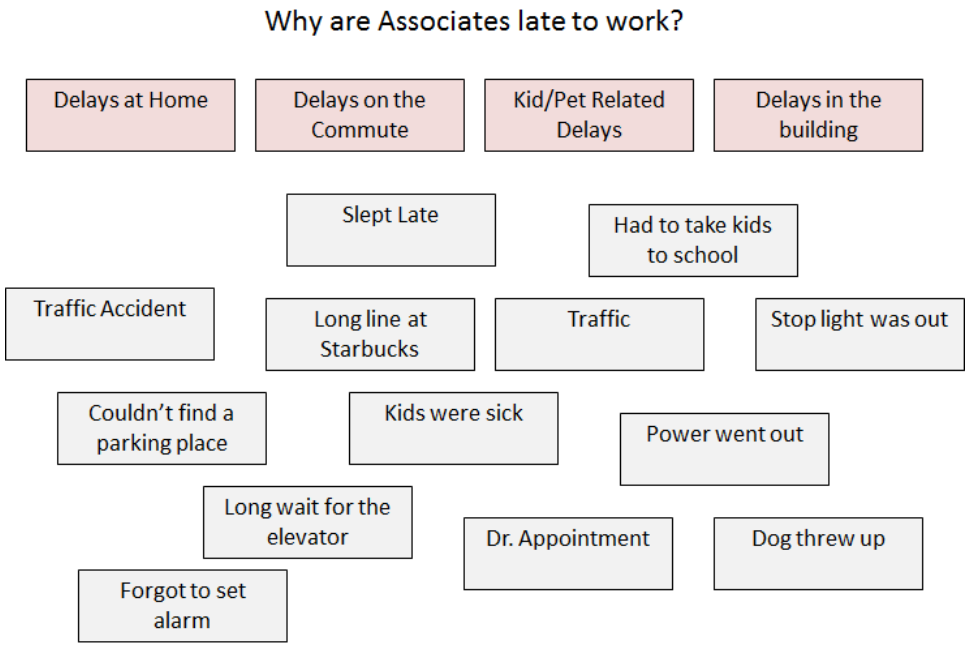


Fig. 1.15 Common themes

In figure 1.15, you can see that some common themes have emerged (delays at home, delays on the commute, kids/pet related delays, or delays in the building).

- Group note cards or sticky notes into themes until all responses are allocated.
- Re-evaluate and make final adjustments.

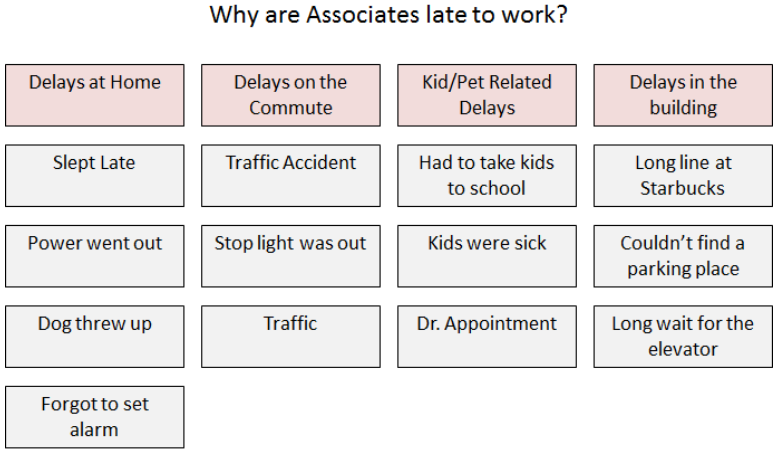


Fig. 1.16 Affinity Diagram

Finally, in figure 1.16 you can now see that the individual responses are lined up under each theme. See any that might be misplaced? How about "dog threw up?" Perhaps "delays at home" and "kid/pet-related delays" do not need to be broken down separately. These are the kinds of things to consider when evaluating and adjusting.

CTQ Tree

The next figure shows an example of a generic CTQ tree transposed from a white board to a software package (*generated with Minitab's Quality Companion 3*)

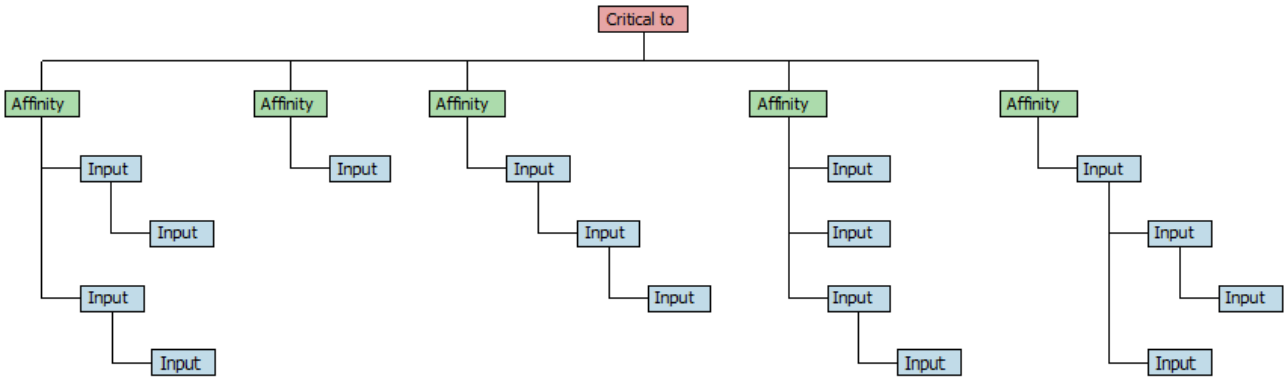


Fig. 1.17 CTQ Tree

Kano

Another VOC categorization technique is the *Kano*. The Kano model was developed by Noriaki Kano in the 1980s and it is a graphic tool that further categorizes VOC and CTQs into three distinct groups:

1. Must Haves
2. Performance Attributes
3. Delighters

The Kano helps to:

- Identify CTQs that add incremental value vs. those that are simply requirements
- Understand that having more is not necessarily better

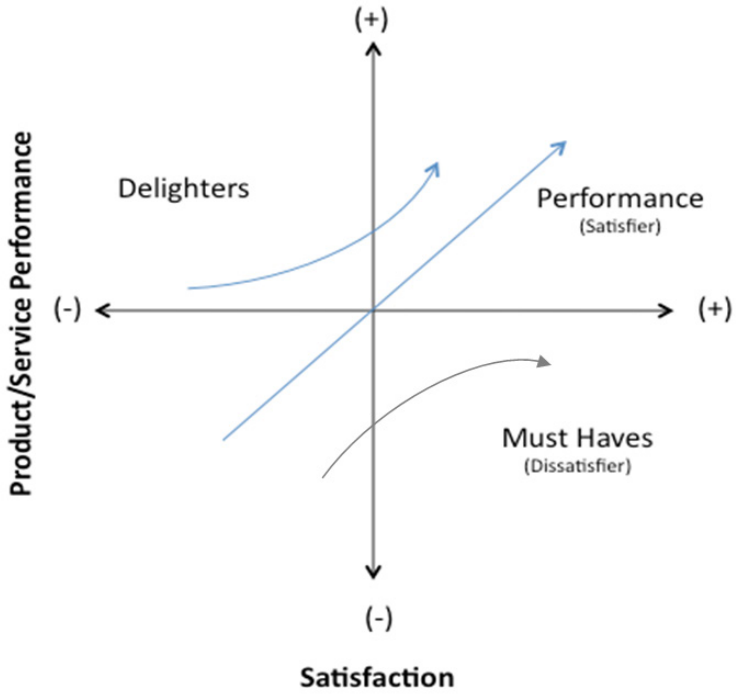


Fig. 1.18 Kano Diagram

It is important to understand what CTQs are non-negotiable (*must haves*) versus those that are nice to have. "Must have" attributes are often taken for granted when they are fulfilled, but result in dissatisfaction when they are not fulfilled. The curve for a "must have" will drop in satisfaction as you move to the left where the need is not fulfilled, but flattens out to a neutral satisfaction level as you move to the right.

A *delighter* is the opposite of the must have—if it is not fulfilled, the customer will not be dissatisfied, but as it is fulfilled, the customer's satisfaction increases. These are things the customer did not realize they wanted or needed, until they have experienced them. Eventually, delighters will become expected, turning them into *performance attributes* or must haves.

Examples:

- How would you classify the steering wheel in a car?
- Low prices at a discount store?
- What about the touch screen on an iPhone (think back to when it came out)?

A performance attribute (or one-dimensional attribute) will increase satisfaction level as the need is fulfilled, and will decrease satisfaction when it is not fulfilled.

Validating VOC and CTQs

Once the process of collecting VOC is complete, and CTQs have been determined, the next step is to *validate* what has been developed. Confirming can be accomplished by conducting surveys through one or more of the following methods:

- Group sessions
- One-on-one meetings
- Phone interviews
- Electronic means (chat, email, social media etc.)
- Physical mail

Consider your confirming audience and try to avoid factors that may influence or bias responses such as inconvenience or overly burdensome time commitments. For example, an overly extensive customer survey might end up getting a rushed response and, therefore, may not represent the true views of an audience.

Translating CTQs to Requirements

To be able to incorporate a CTQ into the design of a process, the next step is to *translate* CTQs to specific process requirements. A *requirements tree* translates CTQs to meaningful and measurable requirements for production processes and products.

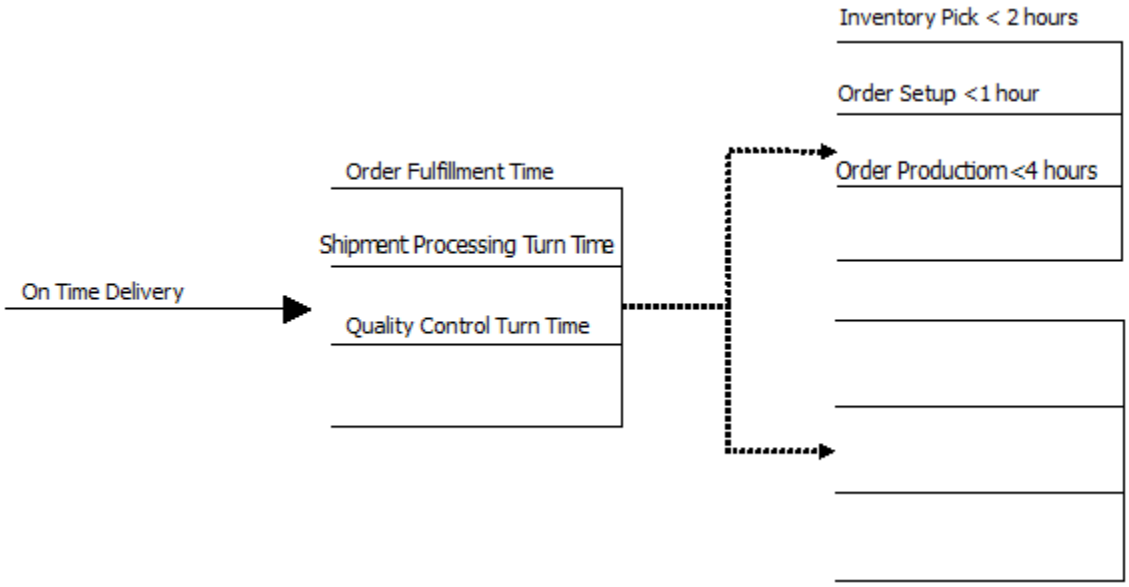


Fig. 1.19 CTQ Requirements Tree

In figure 1.19, the actual feedback from the customer (VOC) might have been something like, "Deliver my product when you say you're going to deliver it". This statement is then translated to something your business can measure such as "on-time delivery."

To achieve on time delivery and ensure that you're meeting your customers' expectations, on time delivery must be translated to the very specific process requirements that your business must achieve to meet on time delivery requirements.

To do so, one must then consider what aspects of the process contribute to the time that elapses between when the customer places an order and when it is delivered. The customer's order must be fulfilled, how does that process occur and what are the necessary performance

requirements? Also, the order must be inspected by quality control and processed for shipping. What are the performance requirements for those process steps?

The requirements tree help to translate a CTQ into process requirements that can be measured and attained to meet customer expectations.

1.2.3 Quality Function Deployment

History of Quality Function Deployment

The QFD (Quality Function Deployment) was developed in 1972 by Shigeru Mizuno (1910–1989) and Yoji Akao (b. 1928) from the TQM practices of a Japanese company, Mitsubishi Heavy Industries Ltd. QFD aims to design products that assure customer satisfaction and value—the first time and every time.

The QFD framework can be used for translating actual customer statements and needs (the voice of the customer) into actions and designs to build and deliver a quality product. QFD is considered a valuable tool fundamental to a business' success, hence it is widely applied.

What is QFD?

- *QFD* is a construction methodology and quantification tool used to identify and measure customer's requirements and transform them into meaningful and measurable parameters. QFD uses planning matrices—it is also called "The House of Quality."
- QFD helps to prioritize actions to advance process or product to meet customer's anticipations.
- QFD is an excellent tool for contact between cross-functional groups.

Purpose of QFD

The quality function deployment has many purposes, some of the most important being:

- Market analysis to establish needs and expectations
- Examination of competitors' abilities
- Identification of key factors for success
- Translation of key factors into product and process characteristics

QFD focuses on customer requirements, prioritizes resources, and uses competitive information effectively.

Phases of QFD

There are four key phases of QFD:

- Phase I: Product Planning Including the "House of Quality" (Requirements Engineering Life Cycle)

- Phase II: Product Design (Design Life Cycles)
- Phase III: Process Planning (Implementation Life Cycle)
- Phase IV: Process Control (Testing Life Cycle)

How to build a House of Quality

The steps to building a House of Quality are:

- Determine Customer Requirements ("What's" from VOC/CTQ)
- Technical Specifications/Design Requirements ("How's")
- Develop Relationship Matrix ("What's" and "How's")
- Prioritize Customer Requirements
- Conduct Competitive Assessments
- Develop Interrelationship ("How's")
- Prioritize Design Requirements

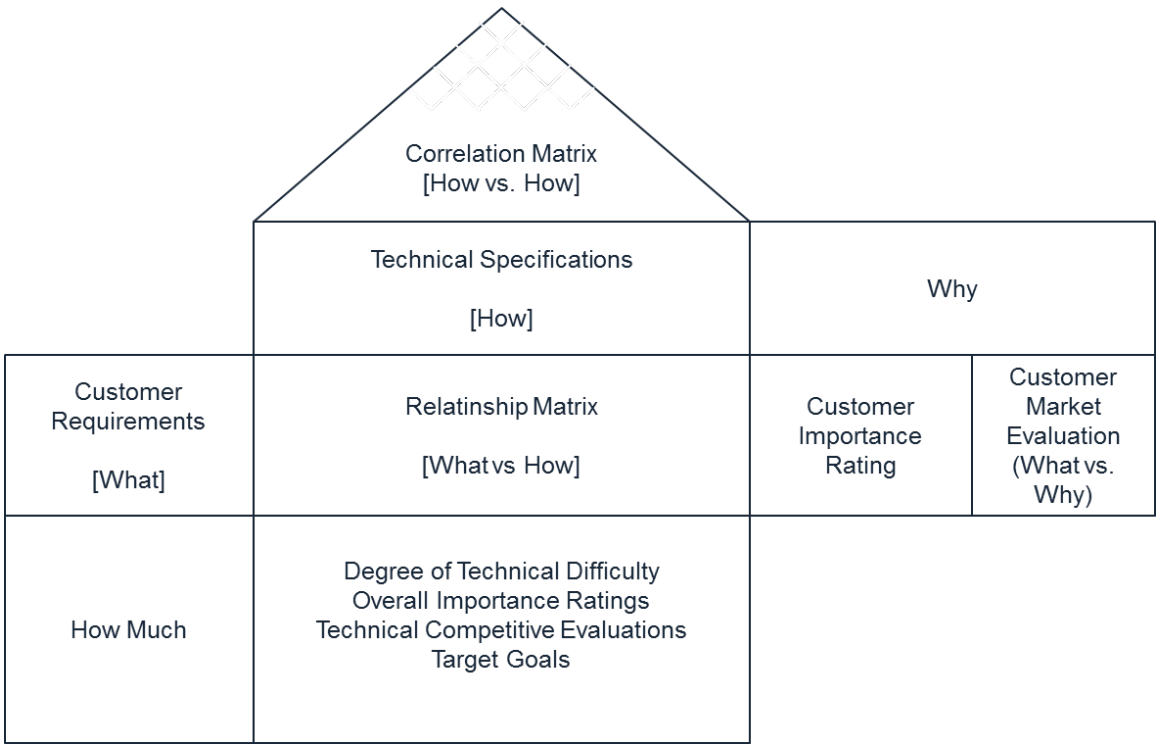


Fig. 1.20 House of Quality

This is the structure of a House of Quality. You can see that it contains the following elements, sometimes referred to as rooms:

- Customer Requirements or the "What's" derived from VOC and CTQs
- Technical specifications or Design Requirements (the "How's")
- Relationship Matrix (the Roof or "What's" and "How's")
- Prioritized Customer Requirements
- Competitive Assessment

- Interrelationships or Correlation Matrix (the "How's")
- Prioritized Design Requirements

How to Build a House of Quality Step by Step
Step 1: Determine Customer Requirements

Identify the important customer requirements. These are the "What's" and are typically determined through the VOC/CTQ process. Use the results from your requirements tree diagram as inputs for the customer requirements in your HOQ.

Fig. 1.21 Customer Requirements

Customers drive the development of the product, not the designers. Hence, it is necessary to determine all customer requirements. Customers will provide the requirements in their own language; you will need to categorize the requirements based on importance. General customer requirements are:

- Human factors
- Physical requirements
- Reliability
- Life-cycle concerns
- Manufacturing requirement

Step 2: Technical Specifications

- Potential choices for product features
- Voice of Designers or Engineers
- Each "What" item must be refined to "How's"

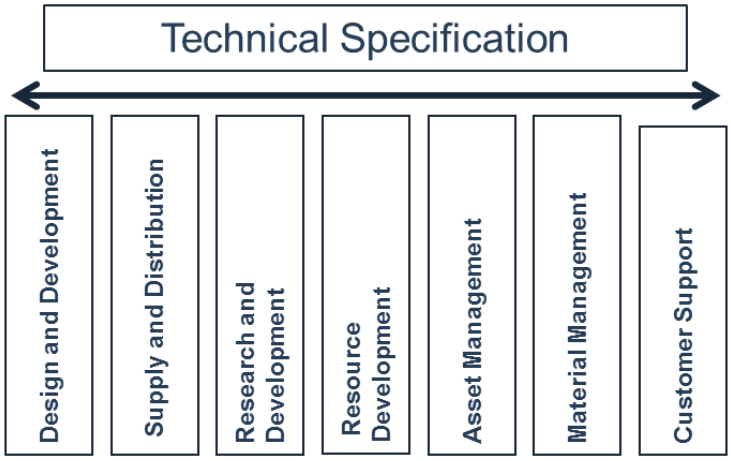


Fig. 1.22 Technical Specifications

The goal is to develop a set of technical specifications from the customer's requirements. Designers and engineers then take the customer requirements and determine technical parameters that need to be designed into the product or process.

Step 3: Develop Relationship Matrix ("What's" and "How's")

This is the center portion of the house. Each cell represents how each technical specification relates to each customer requirement.

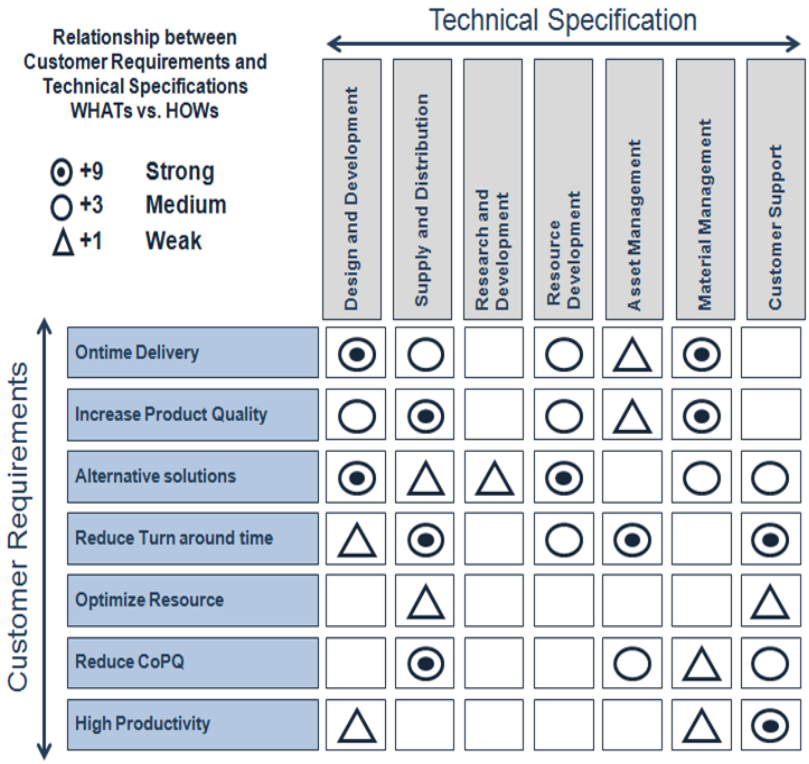


Fig. 1.23 Relationship Matrix

The relationship matrix represents the strength of the relationship between each technical specification and each customer requirement. Use symbolic notations for depicting strong, medium, and weak relationships, such as assigning weights of 1–3–9.

Step 4: Prioritize Customer Requirements

This is the right portion of the house. Each cell represents customer requirements based on relative importance to customers and perceptions of competitive performance.

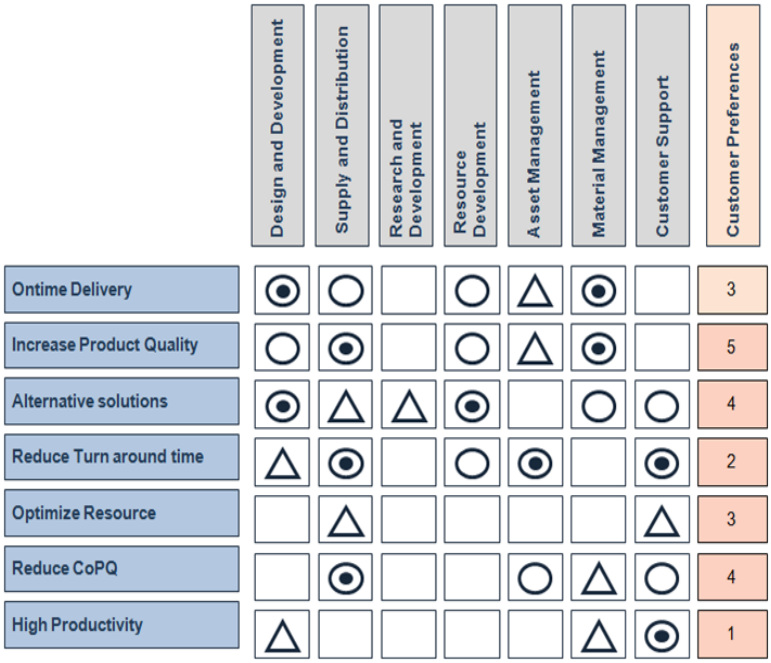


Fig. 1.24 Prioritize Customer Requirements

Prioritize the customer requirements based on importance rating, target value, scale-up factor, sales point.

Step 5: Competitive Assessments

This is the extreme right portion of the house. Comparison of the organization's product to its competitor's products.

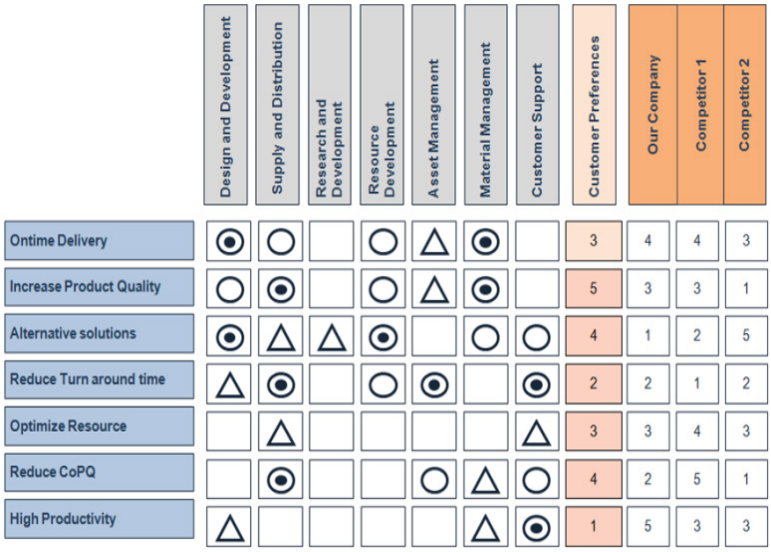


Fig. 1.25 Competitive Assessment

Use the competitive assessment as an opportunity to compare your product/process requirements and specifications with your competitors' products/process. The customer can be used to estimate 1–5 ratings on all products/processes.

Step 6: Correlation Matrix

This is the top portion of the house. It identifies the correlation between "how" items. Some have support (positive) correlations and other may have conflict or negative correlations.

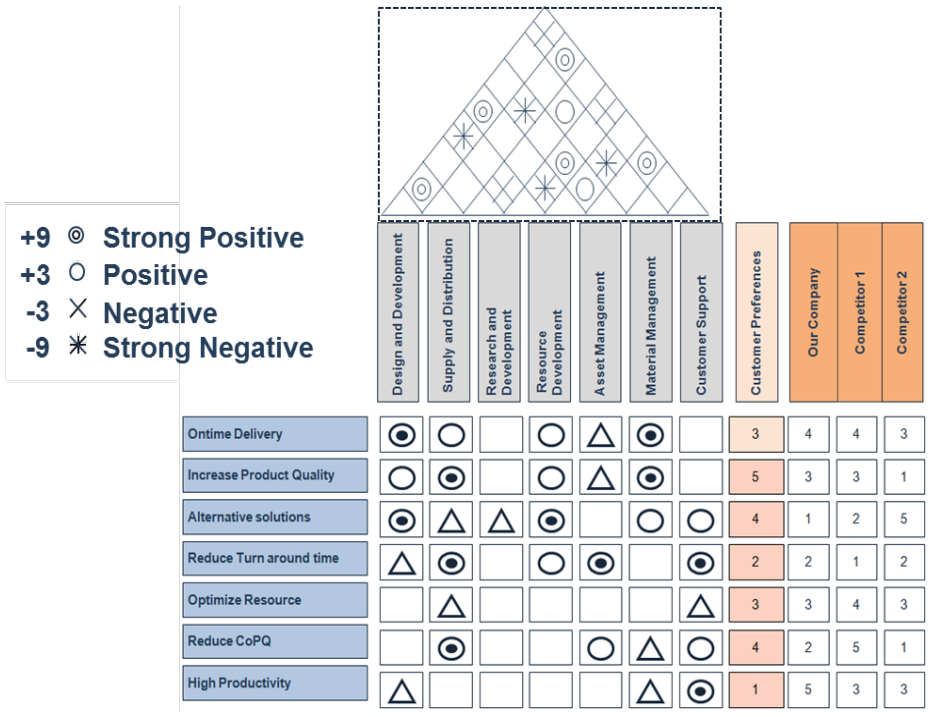


Fig. 1.26 Correlation Matrix

The purpose of the correlation matrix is to identify any inconsistency within the design or technical specifications.

Step 7: Prioritize Design Requirements

- Overall Importance Ratings—Function of relationship ratings and customer prioritization ratings
- Technical Difficulty Assessment—Like customer market competitive evaluations but conducted by the technical team

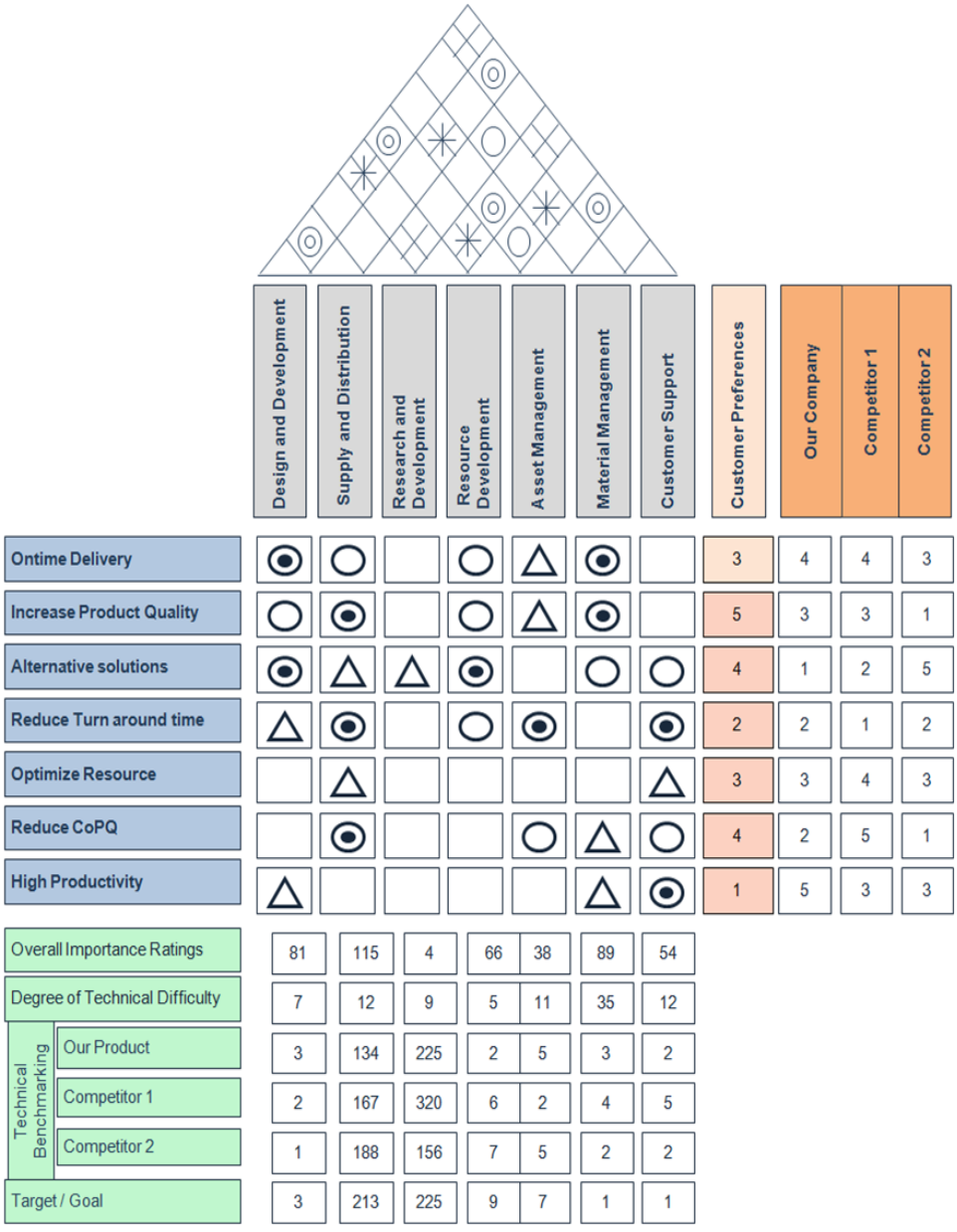


Fig. 1.27 Prioritize Design Requirements

- Overall Importance
- Technical Specification
- Competitive Evaluation
- Target Goals

Each of these prioritization categories helps to establish the feasibility and realization of each "How" item. Target Goals for example can help identify how much is good enough to satisfy the customer. Technical specifications & competitive evaluation should be performed by technical teams. Generally, 1 to 5 ratings should be used. They should be clearly stated in a measurable way.

Figure 1.28 depicts a full House of Quality with each "room" labeled. In summary, the rooms are:

1. Customer Requirements
2. Technical Specifications
3. Relationship Matrix
4. Prioritized Customer Requirements
5. Competitive Assessment
6. Correlation Matrix
7. Prioritized Design Requirements

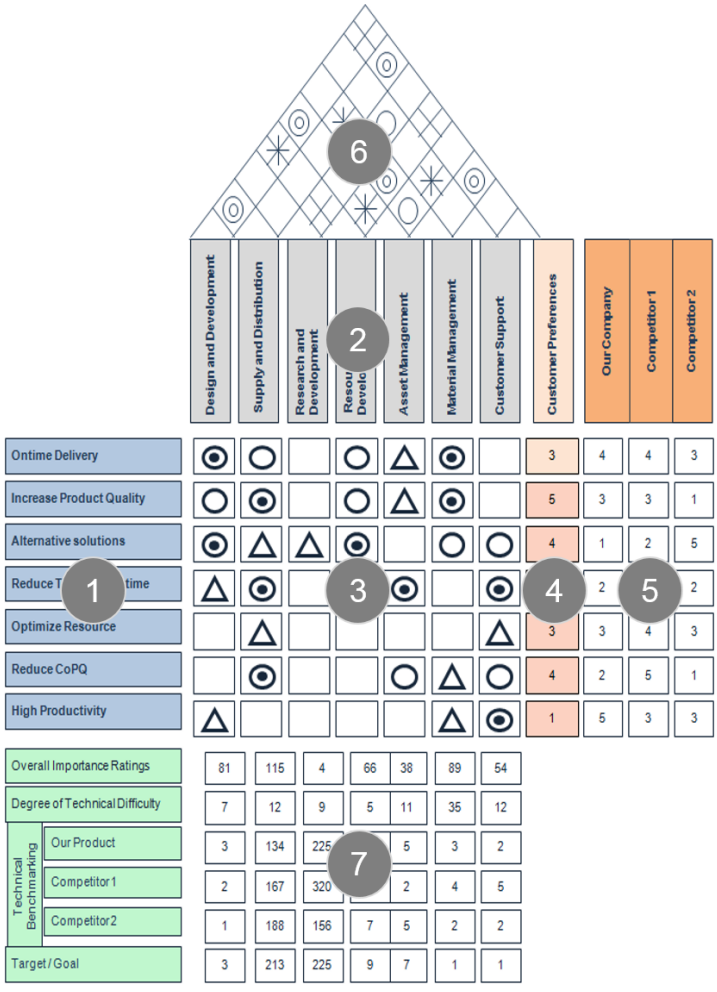


Fig. 1.28 House of Quality

Pros of QFD

- Focuses the design of the product or process on satisfying customer's needs and wants.

- Improves the contact channels between customers, advertising, research and improvement, quality and production departments, which sustains better decision making.
- Reduces the new product development project period and cost.

Cons of QFD

- The relationship matrix can be too obscure with many process inputs and/or many customer constraints.
- It can be very complicated and difficult to implement without experience.
- If throughout the process new ideas, specifications, or requirements are not discovered, you run the risk of losing team members' trust in the process.

QFD Summary

When used properly, the quality function deployment is an extremely valuable approach to product/process design. There are many benefits of QFD that can only be realized when each step of the process is completed thoroughly:

- Logical way of obtaining information and presenting it
- Smallest product development cycle
- Considerably condensed start-up costs
- Fewer engineering alterations
- Reduced chance of supervision during design process
- Collaborating environment
- Preserving everything in characters

1.2.4 Cost of Poor Quality

The *Cost of Poor Quality* (COPQ) is the expense incurred due to waste, inefficiencies, and defects. It is the cost absorbed when a process operates at anything less than optimal: the cost of waste, resource inefficiency, and defects (e.g., repairs, rework, service calls, scrap, warranty claims, and write-offs). COPQ can be staggering when all process inefficiencies are revealed.

Experts have estimated COPQ to be between 5% and 30% of gross sales for most companies. Consider a company that does approximately $100 million in sales per week, and has a COPQ of 20%. That means that one day out of a five-day work week is completely devoted to producing scrap!

Understanding COPQ and where to look for it will help uncover process inefficiencies, defects, and hidden factories within your business. Consider for instance the hidden factories—Do you have procedures for rework, or workarounds? They might seem "hidden" at first, because they have come to be accepted as the normal process for doing things. There are many different types of non-value added activities that can drive waste.

Cost of Poor Quality

There are seven common forms of waste that are often referred to as the *seven deadly muda*. *Muda* is a Japanese word meaning "futility, uselessness, idleness, superfluity, waste, wastage, wastefulness."

Technically, there are more than seven forms of waste but if you can remember these you will capture over 90% of your waste. The seven deadly muda, as defined by Taiichi Ohno (the Toyota Production System), are:

1. Defects
2. Overproduction—Producing more than what your customers are demanding
3. Over-Processing—Unnecessary time spent (e.g., relying on inspections instead of designing a process to eliminate problems)
4. Inventory—Things awaiting further processing or consumption
5. Motion—Extra, unnecessary movement of employees
6. Transportation—Unnecessary movement of goods
7. Waiting—For an upstream process to delivery, for a machine to finish processing, or for an interrupted worker to get back to work

The seven deadly muda are very important to understand. They are the best way to identify the COPQ. The presence of any muda causes many other forms of inefficiencies and hidden factories to manifest themselves. The next section we will discuss how to determine the cost of poor quality related to the seven muda.

COPQ: Costs Related to Production

Costs related to *production* are the direct costs of the presence of muda. These forms of COPQ are usually understood and easily observable. They are in fact the seven deadly muda themselves:

1. Defects—The cost of a defect itself (i.e., the lost labor and materials)
2. Overproduction—Resources spent producing more of something than what is needed, along with the material cost. This is producing "just-in-case" instead of "just-in-time" and creates excessive lead times, results in high storage costs, and makes it difficult to detect defects. Over-producing hides problems in production and takes a lot of courage to produce only what is needed (which will uncover where the issues are).
3. Over-Processing—Resources spent inspecting and fixing instead of getting right the first time
4. Inventory—Work in process on the floor is a symptom of over-processing and waiting
5. Motion—Unnecessary time spent by employees to perform the process; consider motion like bending, stretching, walking, lifting, and reaching

6. Transportation—Adds no value to the product and even increases risk of damage. This can be a hard cost to overcome if facilities are built in a way that requires transportation from one process to another.
7. Waiting—Usually because material flow is poor, production runs are too long, and distances between work centers are too long

COPQ: Costs Related to Prevention

Costs related to the *prevention* of muda are those associated with trying to reduce or eliminate any of the seven deadly muda before they occur:

- Costs for error proofing methods or devices
- Costs for process improvement and quality programs
- Costs for training and certifications, etc.

Any costs directly associated with the prevention of waste and defects should be included in the COPQ calculation.

COPQ: Costs Related to Detection

Costs related to the *detection* of muda are those associated with trying to find or observe any of the seven deadly muda after they occur. Any costs directly associated with the detection of waste and defects should be included in the COPQ calculation. Detection is not as effective as prevention, but sometimes is the only option. It can keep additional costs from being incurred, like continuing to process a defective unit. A few detection examples are:

- Costs for sampling
- Costs for quality control check points
- Costs for inspection costs
- Costs for cycle counts or inventory accuracy inspections, etc.

COPQ: Costs Related to Obligation

Costs related to *obligation* are those associated with addressing the muda that reach a customer. Any costs directly associated with customer obligations should be included in the COPQ calculation. Examples include:

- Repair costs
- Warranty costs
- Replacement costs
- Customer returns and customer service overhead, etc.

COPQ: Types of Cost

There are two types of costs to be considered when determining COPQ:

- Hard Costs—Tangible costs that can be traced to the income statement (e.g., costs of resources or materials)
- Soft Costs—Intangible costs: avoidance, opportunity costs, lost revenue, cost avoidance (the difference between what is actually spent and what *could* have been spent had no action been taken; think of this as slowing the rate of cost increases), opportunity cost (value of lost time that could have been used producing something of value)

Calculating the COPQ

1. Determine the types of waste that are present in your process.
2. Estimate the frequency of waste that occurs.
3. Estimate the cost per event, item, or time frame, whichever is appropriate.
4. Multiply the costs.

1.2.5 Pareto Charts and Analysis

Pareto Principle

The *Pareto principle* is commonly known as the "law of the vital few" or "80:20 rule." It means that the majority (approximately 80%) of effects come from a few (approximately 20%) of the causes. It is a very powerful tool for scoping process improvement efforts because it tells us what the "few" causes are that drive most the effects (or in many cases defects).

This principle was first introduced in early 1900s and has been applied as a rule of thumb in various areas. Business-management consultant Joseph M. Juran suggested the principle and named it after Italian economist Vilfredo Pareto, who observed in 1906 that 80% of the land in Italy was owned by 20% of the population; he developed the principle by observing that 20% of the pea pods in his garden contained 80% of the peas. Examples of applying the Pareto principle:

- 80% of the defects of a process come from 20% of the causes
- 80% of sales come from 20% of customers

The Pareto principle helps us to focus on the vital few items that have the most significant impact. In concept, it also helps us to prioritize potential improvement efforts. Since this 80:20 rule was originally based upon the works of Wilfried Fritz Pareto (or Vilfredo Pareto), all Pareto chart and principle references should be capitalized because Pareto refers to a person (proper noun). Mr. Pareto is also credited for many works associated with the 80:20, some more loosely than others: Pareto's Law, Pareto efficiency, Pareto distribution etc.

Pareto Charts

A *Pareto chart* is a chart of descending bars with an ascending cumulative line on the top. *Sum or Count:* The descending bars on a Pareto chart may be set on a scale that represents the total of all bars or relative to the biggest bucket, depending on the software you are using.

Percent to Total: A Pareto chart shows the percentage to the total for individual bars.

Cumulative Percentage: A Pareto chart also shows the cumulative percentage of each additional bar. The data points of all cumulative percentages are connected into an ascending line on the top of all bars. Not all software packages display Pareto charts the same way. Some show the percent of total for each individual bar, and others show the cumulative percentage.

Pareto Charts Case Study

Next, we will use JMP to run Pareto charts on the same data set. The following table shows the count of defective products by team. Input the tabled data below into your software program and follow the instructions over the next few pages to run Pareto charts.

Count	Category
2	Team 1
12	Team 2
4	Team 3
22	Team 4
2	Team 5
2	Team 6

Table 1.2 Pareto Chart Data

Create a Pareto Chart in JMP

Steps to generate a Pareto chart using JMP:

1. Open the data file "Pareto.jmp"
2. Click Analyze → Quality & Process → Pareto Plot.
3. In the new window, select "Count" as "Frequency"
4. Select "Category" as "Y, cause."
5. Click "OK."

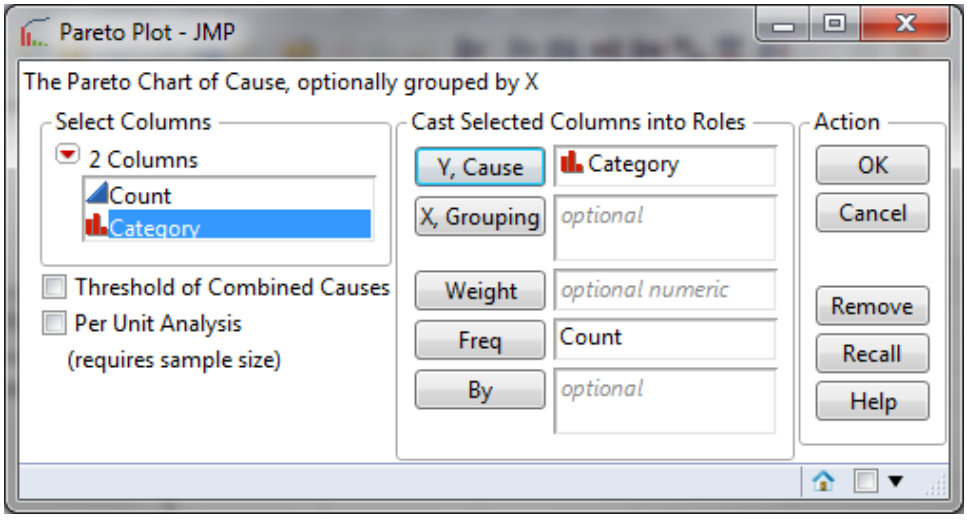


Fig. 1.29 Running a Pareto Chart in JMP

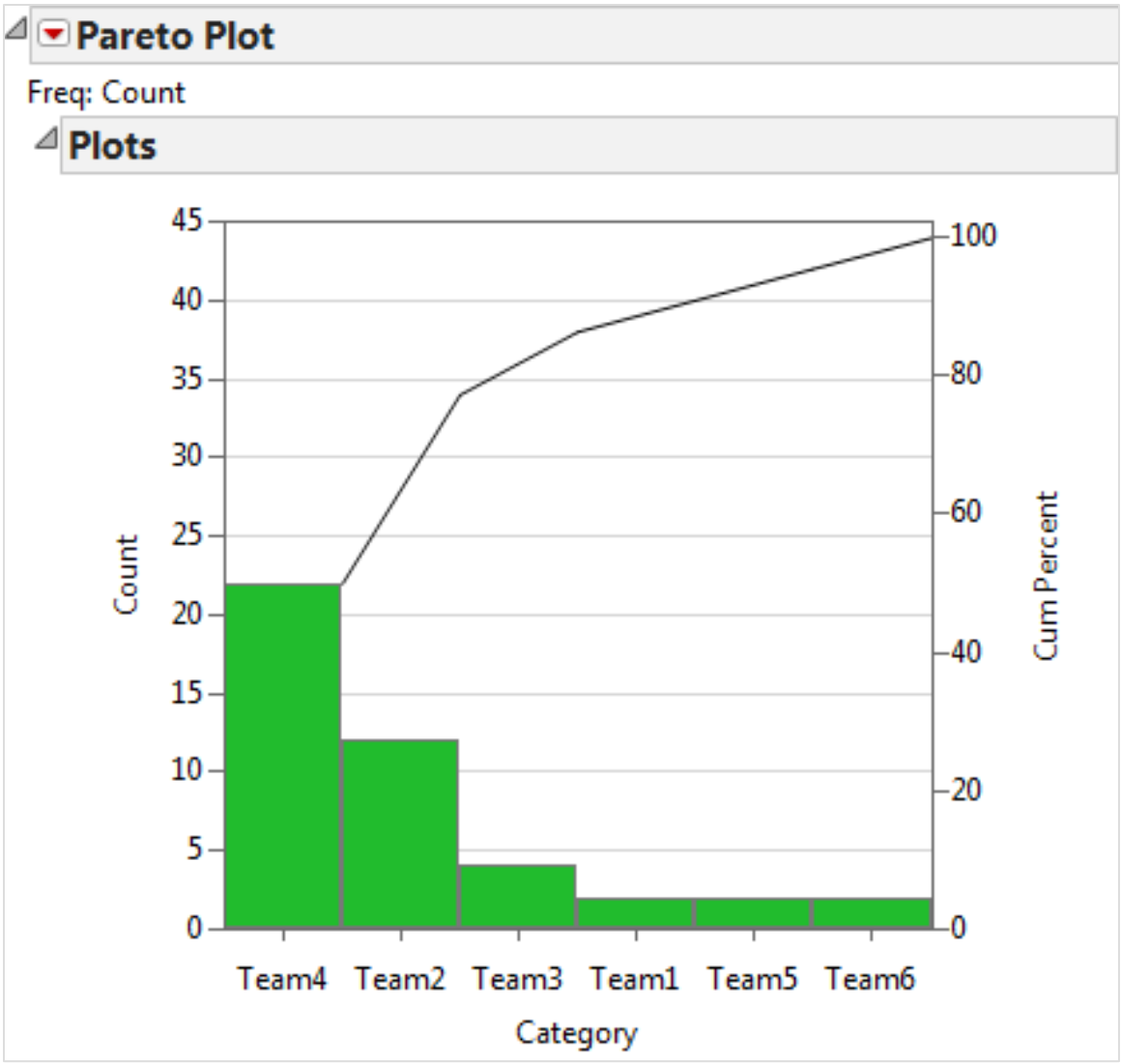


Fig. 1.30 JMP Pareto Chart Output

The Pareto chart above was generated in JMP and represents the count of defective products by team. The bars are descending on a scale with the peak at 45 (approximately the total count of all defective products for all teams). The cumulative percentages make up the black line spanning across the graphic on top of the bars.

Pareto Analysis

The *Pareto analysis* is used to identify root causes by using multiple Pareto charts. In Pareto analysis, we drill down into the bigger buckets of defects and identify the root causes of defects that contribute heavily to total defects. This drill-down approach often effectively solves a significant portion of the problem. Next you will see an example of three-level Pareto analysis.

- The second-level Pareto is a Pareto chart that is a subset of the tallest bar on the first Pareto. Therefore, the chart will represent categorized defects of Team 4.
- The third-level Pareto will then be a subset of the tallest bar of the second-level Pareto.

Pareto Analysis: First Level

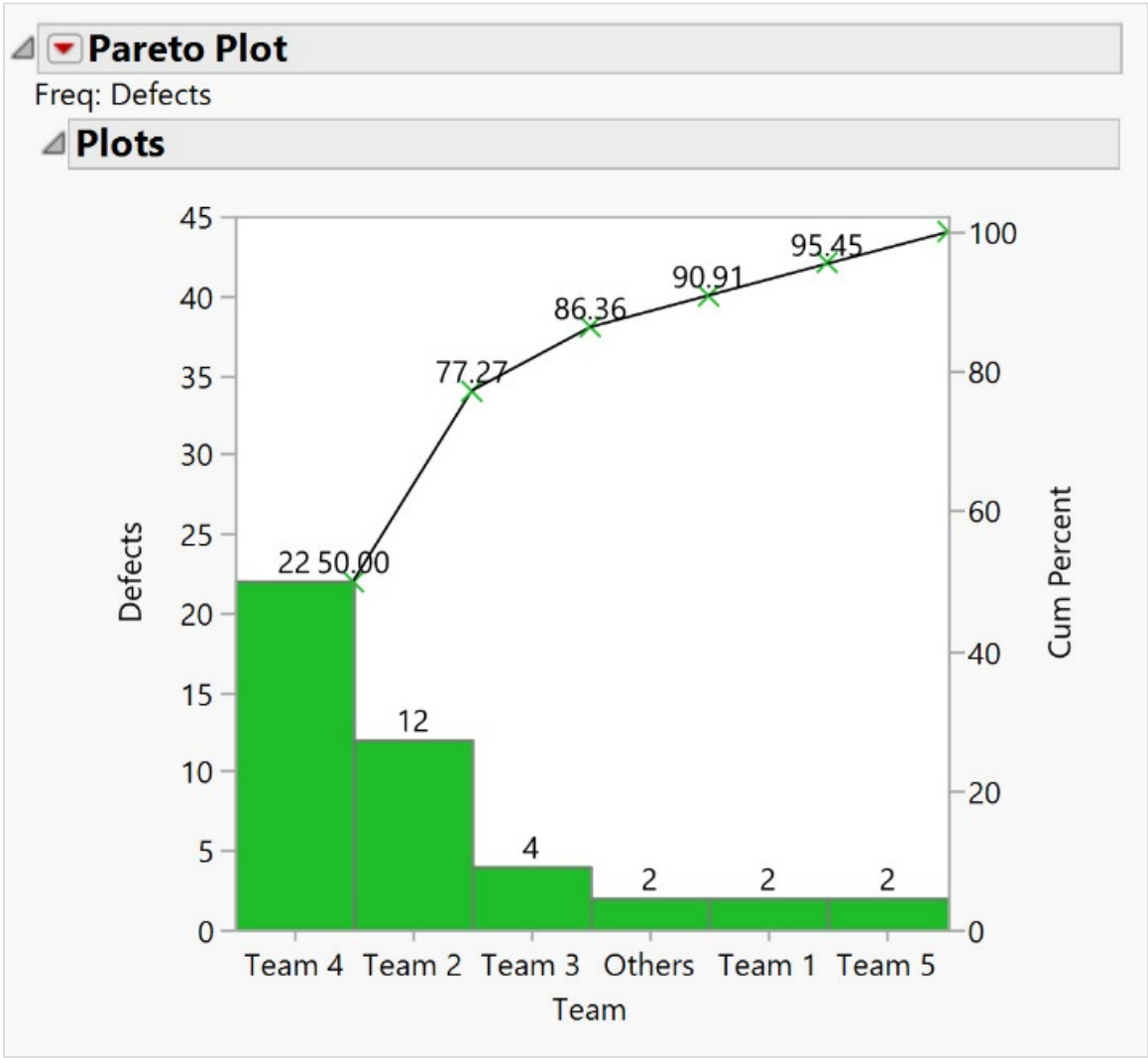


Fig. 1.31 First Level Pareto (defects by team)

The first-level Pareto was already demonstrated in figure 1.31. It showed the count of defective items by team. In that first level Pareto, we concluded that Team 4 accounted for 50% of the total number of defects. The next level Pareto will be a second level and will only show the defective items from Team 4 categorized in a logical manner or in a way that the data allows.

Pareto Analysis: Second Level

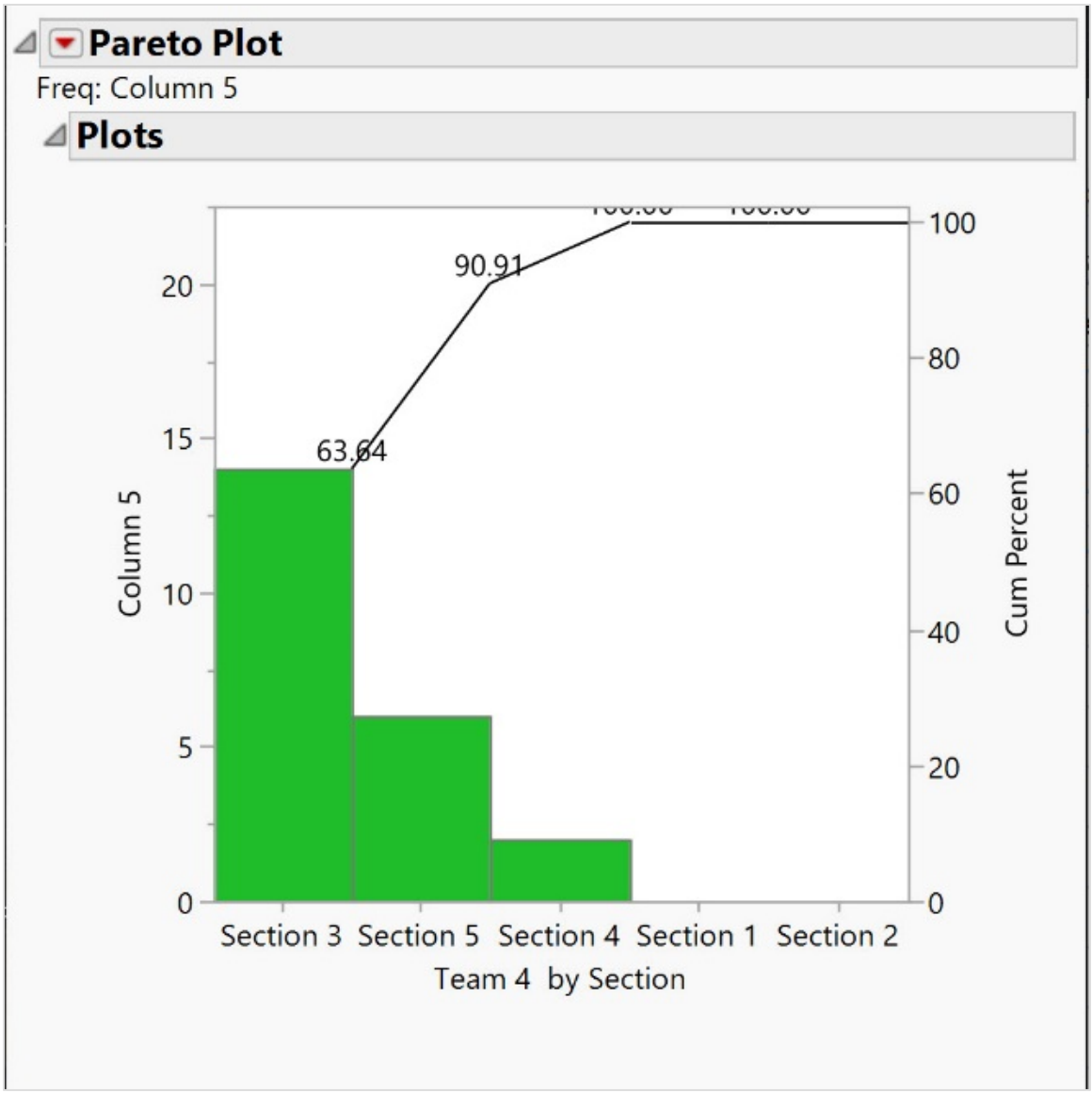


Fig. 1.32 Second Level Pareto (defects by section)

The second-level Pareto shows the count of the defective items by section for only Team 4 (22 defects). You can see that teams are broken down to sections and it appears that section 3 accounts for 63% of the total defects for team 4.

Pareto Analysis: Third Level

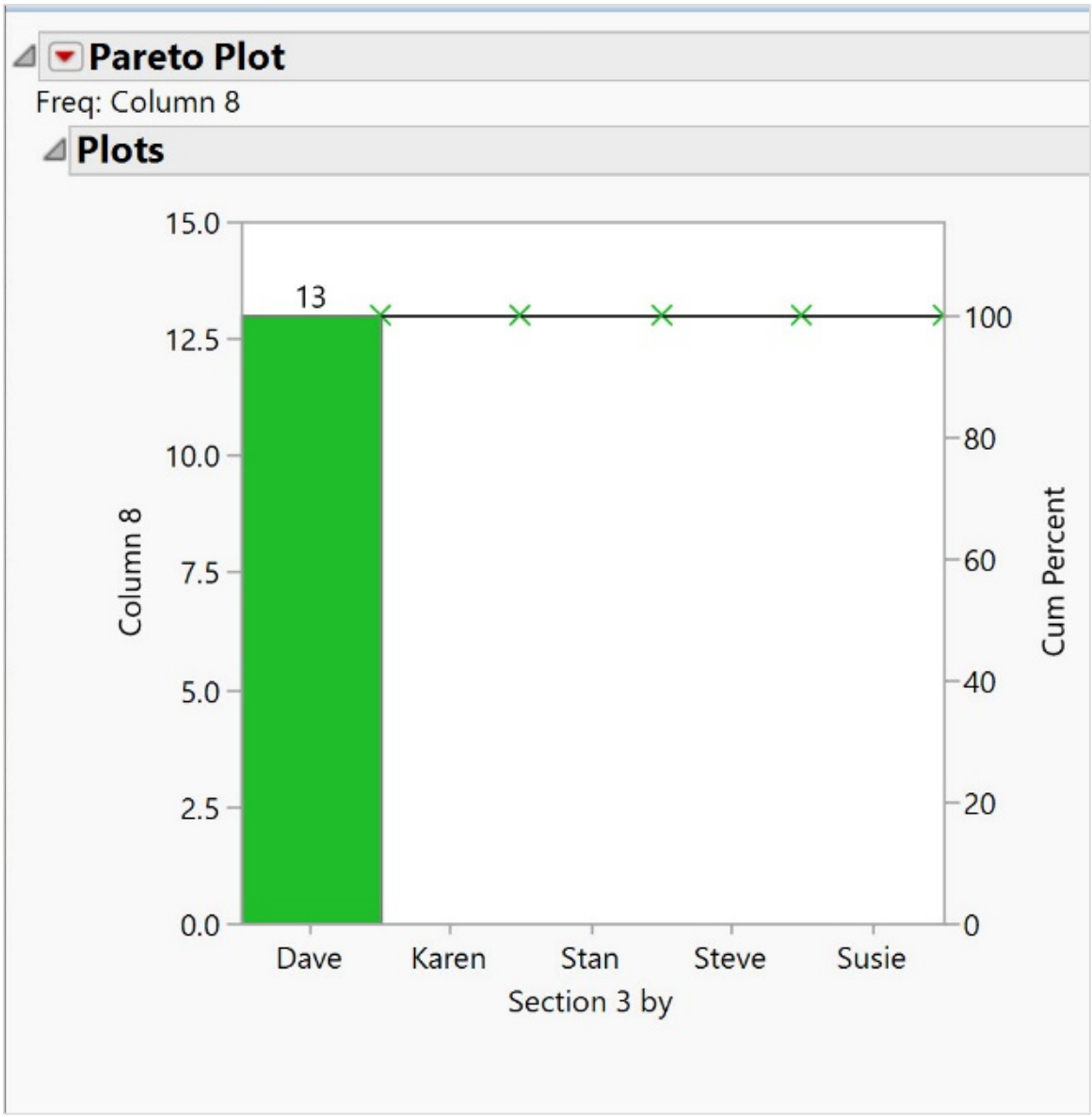


Fig. 1.33 Third Level Pareto (defects by associate)

The third-level Pareto shows the count of defective items by associate for only Section 3 of Team 4. The total defects by associate sum to the total for Section 3 and Dave accounts for 93% of Section 3's defects.

Pareto Analysis: Conclusion

Our analysis clearly shows the root of the problem being associate Dave who has 13 of the 14 defects in Section 3 of Team 4. Through this multi-level Pareto analysis, instead of focusing on all teams to reduce the number of defects, we discovered that we can focus on a single individual who is contributing 13 defects out of a total of 44 which is 30% of the total for all teams. Determining what Dave might be doing differently or wrong and solving that problem could fix about 30% of the entire defective products (13/44). This type of analysis significantly narrowed the focus necessary for reducing defective products.

1.3 SIX SIGMA PROJECTS

1.3.1 SIX SIGMA METRICS

There are many Six Sigma metrics and/or measures of performance used by Six Sigma practitioners. In addition to the ones we will cover here, several others (Sigma level, C_p, C_{pk}, P_p, P_{pk}, takt time, cycle time, utilization etc.) will be covered in other modules throughout this training. The Six Sigma metrics of interest in the *Define* phase are:

- Defects per Unit
- Defects per Million Opportunities
- Yield
- Rolled Throughput Yield

Defects per Unit

DPU, *Defects per Unit*, is the basis for calculating DPMO and RTY, which we will cover in the next few pages. DPU is found by dividing total defects by total units.

$$DPU = \frac{D}{U}$$

For example, if you have a process step that produces an average of 65 defects for every 598 units, then your DPU = 65/598 = 0.109 or 10.9%.

Defects per Million Opportunities

DPMO, *Defects per Million Opportunities*, is one of the few important Six Sigma metrics that you should get comfortable with if you are associated with Six Sigma. Remember 3.4 defects per million opportunities? In order to understand DPMO it is best if you first understand both the nomenclature and the nuances such as the difference between defect and defective.

Nomenclature:

- Defects = D
- Unit = U
- Opportunity to have a defect = O

To properly discuss DPMO, we must first explore the differences between “defects” and “defective.”

Defective
Defective suggests that the value or function of the entire unit or product has been compromised. Defective items will always have at least one defect. Typically, however, it takes multiple defects and/or critical defects to cause an item to be defective.

Defect
A defect is an error, mistake, flaw, fault, or some type of imperfection that reduces the value of a product or unit. A single defect may or may not render the product or unit "defective" depending on the specifications of the customer.

To summarize in simple terms, defect means that part of a unit is bad. Defective means that the whole unit is bad. Now let us turn our attention to defining "opportunities" so that we can fully understand Defects per Million Opportunities (DPMO).

Opportunities
Opportunities are the total number of possible defects. Therefore, if a unit has six possible defects, then each unit produced is equal to six defect opportunities. If we produce 100 units, then there are 600 defect opportunities (100 units × 6 opportunities/unit).

Calculating Defects per Million Opportunities

The equation is:

$$DPMO = \frac{D}{U \times O} \times 1{,}000{,}000$$

For example, let us assume there are six defect opportunities per unit and there is an average of 4 defects every 100 units.

Opportunities = 6 × 100 = 600
Defect rate = 4/600
DPMO = 4/600 × 1,000,000 = 6,667

What is the reason or significance of 1,000,000? Converting defect rates to a "per million" value becomes necessary when the performance of your process approaches Six Sigma. When this happens, the number of defects shrink to virtually nothing. In fact, if you recall from section "1.1 What is Six Sigma", six sigma is equivalent to 3.4 defects per million opportunities. By using 1,000,000 opportunities as the barometer we have the resolution in our measurement to count defects all the way up to Six Sigma.

Rolled Throughput Yield

Rolled Throughput Yield (RTY) is a process performance measure that provides insight into the cumulative effects of an entire process. RTY measures the yield for each of several process steps and provides the probability that a unit will come through that process defect-free.

RTY allows us to expose the "hidden factory" by providing visibility into the yield of each process step. This helps us identify the poorest performing process steps and gives us clues into where to look to find the most impactful process improvement opportunities.

Calculating RTY

RTY is found by multiplying the yields of each process step. Let us take the five-step process below and calculate the RTY (the yield or percent of units that get through without a defect) using the multiplication method mentioned above.

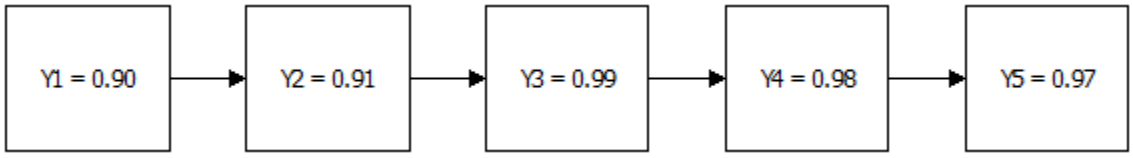


Fig. 1.34 Five Yields for a 5-step process

The RTY calculation for figure 1.34 is: RTY = 0.90 × 0.91 × 0.99 × 0.98 × 0.97 = 0.77

Therefore, RTY = 77%. In this example, 77% of units get through all five process steps without a defect. You may have noticed that to calculate RTY we must determine the yield for each process step. Before we get into calculating yield, there are a few abbreviations that need to be declared.

- Defects = D
- Unit = U
- Defects per Unit = DPU
- Yield = Y
- e = 2.71828 (mathematical constant)

Calculating Yield

The *yield* of a process step is the success rate of that step or the probability that the process step produces no defects. To calculate the yield, we need to know the DPU and then we can apply it to the yield equation below.

$$Y = e^{-dpu}$$

For example, let us assume a process step has a DPU of 0.109 (65/598).

Y = 2.718 ^ −0.109 = 0.8967. Rounded, Y = 90%.

In this example, the DPU (defects per unit) is 0.109 (65 defects/598 units). Plugging in the DPU, and using the mathematical constant for e (2.718), the result is 0.8967, or 90% when rounded. Below in table 1.3, is the above process yield data that we used in the earlier RTY calculation. This table allows us to see the DPU and yield of each step as well as the RTY for the whole process.

Process Step	Defects	Units	DPU	Yield	RTY
1	65	598	0.10870	0.89701	0.90
2	48	533	0.09006	0.91389	0.82
3	5	485	0.01031	0.98974	0.81
4	10	480	0.02083	0.97938	0.79
5	15	471	0.02972	0.97072	0.77

Table 1.3 DPU, Yield and RTY from figure 1.34

The rightmost column of table 1.3 shows the RTY cumulated through all the process steps. Which steps seem to be the greatest opportunities? Notice the yield in steps 1 and 2 are the lowest.

Estimating Yield

Instead of using the equation provided earlier, there is a simpler way to calculate yield using *yield estimation:* It is possible to "estimate" yield by taking the inverse of DPU or simply subtracting DPU from 1.

Yield Estimation = 1 – DPU

Consider the DPU's in table 1.3 and follow the 1-DPU method for estimating yield. You should have something similar to the following equations:

- Yield Estimate for process step 1: 1 – 0.10870 = 0.90
- Yield Estimate for process step 2: 1 – 0.09006 = 0.91
- Yield Estimate for process step 3: 1 – 0.01031 = 0.99
- Yield Estimate for process step 4: 1 – 0.02083 = 0.98
- Yield Estimate for process step 5: 1 – 0.02972 = 0.97

Now, let's calculate RTY using the Yield Estimation Method:

RTY = 0.90 × 0.91 × 0.99 × 0.98 × 0.97 = 0.77 = 77%

The maximum DPU is 1, so to get a yield just subtract DPU from 1. As you can see, by calculating yield this way we get the same result for RTY. The previous method's calculated result was also 77%. This may not be the most exact approach but for business professionals who make decisions based on sound analysis, the estimation method will typically suffice.

1.3.2 Business Case and Charter

Earlier we stated that DMAIC is a structured and rigorous methodology designed to be repeatedly applied to *any* process in order to achieve Six Sigma. We also stated that DMAIC was a methodology that refers to five phases of a project: Define, Measure, Analyze, Improve, and Control.

Given that the premise of the DMAIC methodology is project-based, we must take the necessary steps to define and initiate a project, hence the need for project charters. The project charter is what sets direction for a project.

Project Charter

The purpose of a *project charter* is to provide vital information about a project in a quick and easy-to-comprehend manner. It is a contract of sorts between project champions, sponsors, stakeholders, and the project team. It is a quick reference to tell the story of what, why, who, when, etc. To set the *right* direction for a project, it is critical to define certain guardrails to ensure the project stays on track and is successful. Project charters are used to get approval and buy-in for projects and initiatives as well as declaring:

- Scope of work
- Project teams
- Decision authorities
- Project lead
- Success measure

The key elements of project charters are:

- Title
- Project Lead
- Business Case
- Problem Statement
- Project Objective
- Primary Metric
- Secondary Metrics
- Project Scope
- Project Timeline
- Project Constraints
- Project Team
- Stakeholders
- Approvers
- Constraints
- Dependencies
- Risk

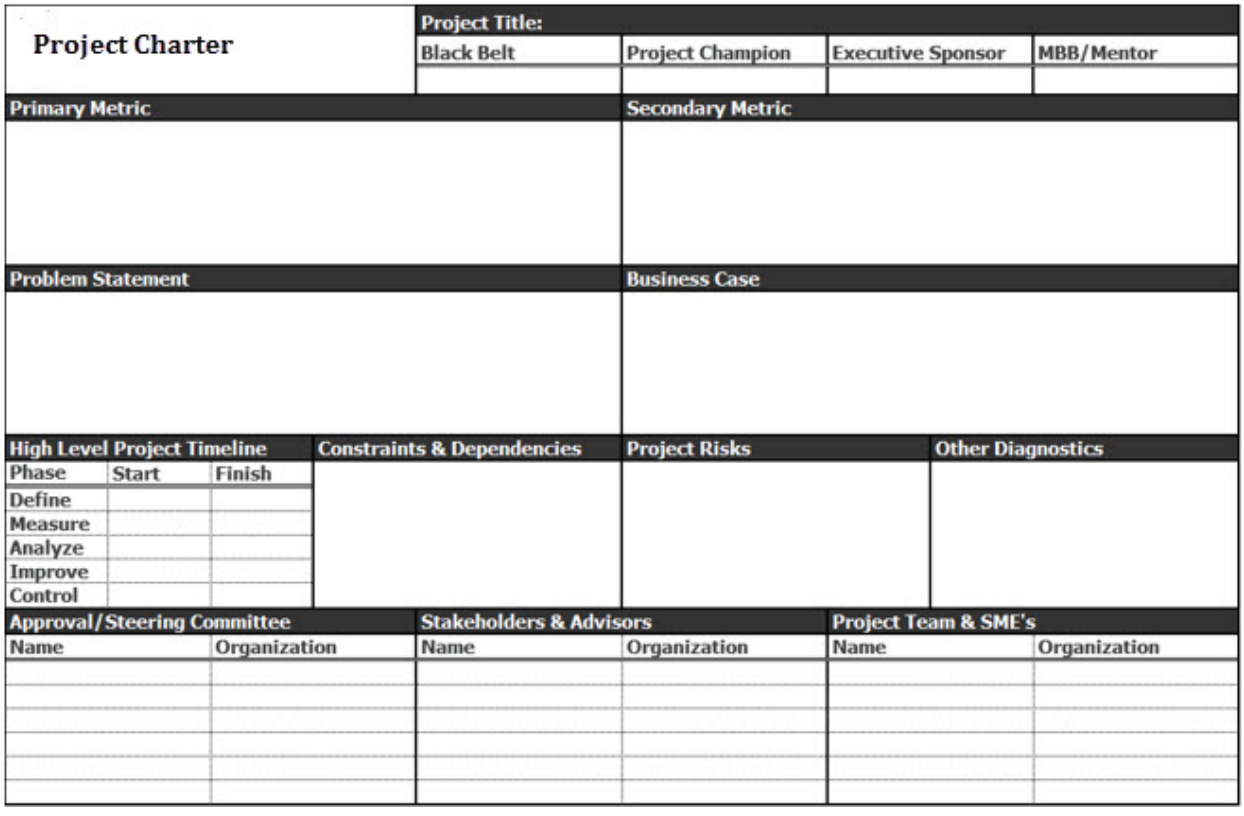


Fig. 1.35 Project Charter Template

Figure 1.35 above shows an example of a project charter with all the key elements. Because of its formatting, it's easy to share, easy to read, and is useful for quickly communicating to people who may not be familiar with the project.

Project Charter: Key Elements
Title: Projects should have a name, title, or some reference that identifies them. Branding can be an important ingredient in the success of a project so be sure your project has a reference name or title. This is how your project will be known to others, and keeping it top of mind is obviously part of ensuring its success.

Leader: All projects need a declared leader or someone who is responsible for project's **RACI** stands for **R**esponsible, **A**ccountable, **C**onsulted, and **I**nformed. RACI identifies the people that play those roles. Every project must have declared leaders indicating who is responsible and who is accountable.

Business Case: A business case is the quantifiable reason why the project is important. It is important for the business case to be quantifiable. Stakeholders have to know what success looks like. Cost of poor quality, as described earlier, is a great approach to quantifying the business case.

Business cases help shed light on problems. They explain why a business should care. A business case clearly articulates why it is necessary to do the project and what is the benefit to the customer, to employees, or to the shareholders.

A problem by itself is not enough to articulate why a project needs to be done. The business case turns the problem into the reason a business should care: What is the cost of the problem, what is the impact to customers, or what is the lost revenue opportunity? Business cases must be quantified and stated succinctly. COPQ is a key method of quantification for any business case.

Problem Statement and Objective: A properly written problem statement has an objective statement woven into it. There should be no question as to the current state or the goal.

A gap should be declared, the gap being the difference between the present state and the goal state. The project objective should be to close the gap or reduce the gap by some reasonable amount. Valuation or COPQ is the monetary value assigned to the gap.

Lastly, a well-written problem statement refers to a timeline expected to be met. A well-written problem statement includes all of the following:

- Declaration of the current state of a problem in terms of a given measure
- A goal for the measure, which establishes a gap from the current state
- An objective that states how much of the gap the project aims to close
- The monetary value of the objective (cost of poor quality)
- A timeline that the project will meet

Project Charter: Problem Statement Examples
Currently, process defect rates are 17% with a goal of 2%. This represents a gap of 15%, costing the business $7.4 million dollars. The goal of this project is to reduce this gap by 50% before November 2010 putting process defect rates at 9.5% and saving $3.7MM.

Process cycle time has averaged 64 minutes since Q1 2009. However, production requirements put the cycle time goals at 48 minutes. This 16-minute gap is estimated to cost the business $296,000. The goal of this project is to reduce cycle time by 16 minutes by Q4 2010 and capture all $296,000 cost savings. Can you identify the key elements of a problem statement?

Metrics: A measure of success is an absolute for any project.

- Metrics give clarity to the purpose of the work.
- Metrics establish how the initiative will be judged.
- Metrics establish a baseline or starting point.

For all Six Sigma projects, metrics are mandatory! How does one know if a project does what it sets out to do? It is critical that *every* project choose the right metrics to determine success. Project metrics are the center point for the work to ensure that the direction and decisions made through the course of the project are based on data. The metrics are what stakeholders will judge the success of the project on.

Primary Metric: The *primary metric* is a generic term for a Six Sigma project's most important measure of success. Choosing the primary metric should not be taken lightly. A Six Sigma expert should play a significant role in determining the primary metric to ensure it meets the characteristics of a good one. The primary metric is defined by the BB, GB, MBB, or Champion.

A primary metric is an absolute must for any project and it should not be taken lightly. Here are a few characteristics of good primary metrics. Primary metrics should be:

- Tied to the problem statement to ensure that what is being measured is a direct indication of whether the problem is being solved or not
- Measurable
- Expressed with an equation, meaning that the primary metric, Y, can be expressed as a function of the x's
- Aligned to business objectives, meaning that the primary metric should be correlated to business results
- Tracked at the proper frequency (hourly, daily, weekly, monthly etc.); in other words, they should be something that can be tracked frequently enough so that action can be taken quickly when the metric change
- Expressed pictorially over time with a run chart, time series, or control chart
- Validated with an MSA

The primary metric is the reason for your work, the success indicator, and your beacon. The primary metric is of utmost importance and should be improved, *but* not at the expense of your secondary metric. Do not trade one problem for another. Solutions need to be balanced. Be sure that you do not create an issue somewhere else when you implement a solution to improve your primary metric. That is why *secondary metrics* are important.

Secondary Metric: The *secondary metric* is the thing you do not want sacrificed on behalf of a primary improvement. A secondary metric is one that makes sure problems are not just "changing forms" or "moving around." The secondary metric keeps us honest and ensures we are not sacrificing too much for our primary metric.

If your primary metric is a cost or speed metric, then your *secondary metric* should probably be some quality measure. For example, if you were accountable for saving energy in an office building and your primary metric was energy consumption then you *could* shut off all the lights and the HVAC system and save tons of energy . . . except that your secondary metrics are probably comfort and functionality of the work environment.

When you think about your primary metric, think about the really easy and obvious ways to improve the primary metric. Chances are that you can easily improve it by sacrificing something else. The elements of a good project charters include:

- Scope Statement—Defined by high-level process map
- Stakeholders Identified—Who is affected by the project
- Approval Authorities Identified—Who makes the final call
- Review Committees Defined—Who is on the review team
- Risks and Dependencies Highlighted—Identify risks and critical path items
- Project Team Declared—Declare team members
- Project Timeline Estimated—Set high-level timeline expectations

1.3.3 PROJECT TEAM SELECTION

Six Sigma project team selection is the cornerstone of a successful Six Sigma project. Team selection is obviously not something that is unique to a Six Sigma project. Team selection is important in any kind of project. You want to choose a group of people that complement each other, share similar goals and objectives, and are not afraid to hold each other accountable for achieving the project's objectives.

Teams and Team Success

A *team* is a group of people who share complementary skills and experience.

- A team will be dedicated to consistent objectives.
- Winning teams share similar and coordinated goals.
- Teams often execute common methods or approaches.

- Team members hold each other accountable for achieving shared goals.

What Makes a Team Successful?

There should not be any surprises about what makes a team successful. These characteristics are all important for a team to be successful:

- Shared goals
- Commitment
- Leadership
- Respect
- Effective communication
- Autonomy
- Diverse knowledge & skills
- Adequate resources

The keys to team success are:

- Agreed focus on the goal or the problem at hand
    - Focus on problems that have meaning to the business
    - Focus on solvable problems within the scope of influence; a successful team does not seek unattainable solutions
- Team Selection
    - Selected teammates have proper skills and knowledge
    - Adequately engaged management
    - Appropriate support and guidance from their direct leader
- Successful teams use reliable methods
    - Follow the prescribed DMAIC methodology
    - Manage data, information, and statistical evidence
- Successful teams always have "exceeds" rated players. Winning teams typically have unusually high standards
    - Have greater expectations of themselves and each other
    - Do not settle for average or even above average results

Teams should be able to agree on what the focus is. The team members all need to understand how the problem translates to something meaningful for the business. They also should be realistic about what they can and should seek to accomplish.

The people that are selected need to have "skin in the game." Not only do they need to be knowledgeable and have the needed skills, but they (and their management) need to be motivated to seek the solution.

The team should use the facts and data to lead the way. Be disciplined to follow the DMAIC approach and be aware of when a gut feeling or emotion is driving their behavior. Of course, teams should be built from the top performers whenever possible.

Principles of Team Selection

- Select team members based on:

    - Skills required to achieve the objective
    - Experience (subject matter expertise)
    - Availability and willingness to participate
    - Team size (usually four to eight members)
    - Don't go at it alone!
    - Don't get too many cooks in the kitchen!
    - Members' ability to navigate
    - The process
    - The company
    - The political landscape
- Be sure to consider the inputs of others
    - Heed advice
    - Seek guidance

A team member should be selected based on the skills and subject matter expertise. For example, if you are focusing on a process improvement, find a top performing employee that performs the process on a regular basis. A team member needs to have capacity to participate on a project. Therefore, it is important to pick people with shared goals. If they have a shared goal, it is likely that they can find the time.

The team size needs to be optimal. Include enough people to generate consensus, but not too many people because they can never reach consensus, and this can paralyze a project. Be open minded, listen to others, be inclusive.

Project Team Development

All teams experience the following four stages of development. It is helpful to understand these phases so that you can anticipate what your team is going to experience. The four stages of team development process are:

1. Forming
2. Storming
3. Norming
4. Performing

Teammates seek something different at each stage:

1. In the forming stage they seek inclusion
2. In the storming stage they seek direction and guidance
3. In the norming stage they seek agreement
4. In the performing stage they seek results

Patterns of a team in the Forming stage:

- Roles and responsibilities are unclear

- Process and procedures are ignored
- Scope and parameter setting is loosely attempted
- Discussions are vague and frustrating
- There is a high dependence on leadership for guidance

In the Forming stage, people are trying to figure out where they fit in; the team is trying to figure out where the social boundaries are with each other; they are all trying to "get to know" each other.

Patterns of a team in the Storming stage:

- Attempts to skip the research and jump to solutions
- Impatience for some team members regarding lack of progress
- Arguments about decisions and actions of the team
- Team members establish their position
- Subgroups or small teams form
- Power struggles exist and resistance is present

After the delay of the forming stage, urgency can set in and they start to seek progress. Team members even seek progress by sacrificing the discipline. Patterns of a team in the Norming stage are:

- Agreement and consensus start to form
- Roles and responsibilities are accepted
- Team members' engagement increases
- Social relationships begin to form
- The leader becomes more enabling and shares authority

In the Norming stage the team is settling into place and into roles with responsibilities. Once roles are clear, engagement increases with comfort level, and the team starts to form relationships with each other. Patterns of a team in the Performing stage are:

- Team is directionally aware and agrees on objectives
- Team is autonomous
- Disagreements are resolved within the team
- Team forms above average expectations of performance

In the Performing stage, the team is aligned and directed. The members are autonomous in that they can resolve issues within and stay on the path to progress. The team is united to exceed expectations of performance.

Well-structured and energized project teams are the essential components of any successful Six Sigma project. To have better chances of executing the project successfully, you will need to understand and effectively manage the team development process. Even with the most

perfectly defined project, a well-structured team can be the difference between success and failure.

1.3.4 PROJECT RISK MANAGEMENT

Risk

Risk is defined as a future event that *can* impact the task/project if it occurs. A broad definition of project risk is an uncertainty that can have a negative *or* positive effect on meeting project objectives. Positive risks are risks that result in good things happening—sometimes called opportunities.

What is Project Risk Management?

The main purpose of *risk management* is to foresee potential risks that may inhibit the project deliverables from being delivered on time, within budget, and at the appropriate level of quality, and then to mitigate these risks by creating, implementing, and monitoring *contingency plans*. Risk management is concerned with identifying, assessing, and monitoring project risks before they develop into issues and impact the project.

Risk analysis helps to identify and manage potential problems that could impact key business initiatives or project goals.

Three Basic Parameters of Risk Analysis

1. Risk Assessment: The process of identifying and evaluating risks, whether in absolute or relative terms
2. Risk Management: Project risk management is the effort of responding to risks throughout the life of a project and in the interest of meeting project goals and objectives
3. Risk Communication: Communication plays a vital role in the risk analysis process because it leads to a good understanding of risk assessment and management decisions

Why is Risk Analysis Necessary?

What can happen if you omit the risk analysis?

- Vulnerabilities cannot be detected
- Mitigation plans are introduced without proper justification
- Customer dissatisfaction
- Not meeting project goals
- Remake the whole system
- Huge cost and time loss

Project Risk Analysis Steps

The project risk analysis process consists of the following steps that evolve through the life cycle of a project.

1. Risk Identification: Identify risks and risk categories, group risks, and define ownership.
2. Risk Assessment: Evaluate and estimate the possible impacts and interactions of risks.
3. Response Planning: Define mitigation and reaction plans
4. Mitigation Actions: Implement action plans and integrate them into the project
5. Tracking and Reporting: Provide visibility to all risks
6. Closing: Close the identified risk

1. Risk Identification
The first action of risk management is the identification of individual events that the project may encounter during its lifecycle. The identification step comprises:

- Identify the risks
- Categorize the risks
- Match the identified risks to categories
- Define ownership for managing the risks

The process of identification, matching, and assigning ownership may involve the full project team in a brainstorming exercise or a workshop. The team should be encouraged to participate in risk identification by actively reporting risks as they arise, and to make suggestions as to the controls needed to rectify the situation.

Source of Risk:
Identification of risk sources provides a basis for systematically examining changing situations over time to uncover circumstances that impact the ability of the project to meet its objectives.

Establishing categories for risks provides a mechanism for collecting and organizing risks as well as ensuring appropriate scrutiny and management attention for those risks that can have more serious consequences on meeting project objectives.

Source of Risk	Description
Human Resources	The risks originated from human resources (e.g. resource availability, skills, training etc.)
Physical Resources	The risks originated from physical resources (e.g., hardware or software, availability of the required number at the right time etc.)
Technology	The risks originated from technology (e.g., development environment, new or complex technologies, performance requirements, tools etc.)
Suppliers	The risks are associated with a supplier (e.g., delays in supplies, capability of suppliers etc.)

Customer	The risks derived from the customer (e.g., unclear requirements, requirement volatility, change in project scope, delays in response etc.)
Security	The risks are associated with information security, security of personnel, security of assets, and security of intellectual property
Legal	The risks are associated with legal issues that may impact the project
Project Management	The risks are associated with project management processes, organizational maturity etc.

Table 1.4 Risk Identification (sources of risk)

The table above shows some examples of source of risks, from those associated with human resources to those related to project management.

Risk Parameters:

Parameters for evaluating, categorizing, and prioritizing risks include the following:

- Risk likelihood (i.e., probability of risk occurrence)
- Risk consequence (i.e., impact and severity of risk occurrence)
- Thresholds to trigger management activities

Risk parameters are used to provide common and consistent criteria for comparing the various risks to be managed and to prioritize the necessary actions required for risk mitigation planning.

2. Risk Assessment

The *risk assessment* consists of evaluating the range of possible impacts should the risk occur.

Follow these steps when assessing risks:

- Define the various impacts of each risk
- Rate each impact based on a logical severity level
- Sort and evaluate risks by severity level
- Determine if any controls already exist
- Define potential mitigation actions

We must prioritize and act on risks in order of priority. In reality, we may be able to fix simple project risks immediately, so we must also use our common sense when prioritizing. Prioritizing is good practice when we need to allocate resources and keep records of the process. The project core team must review *all* risks to ensure the full impact of risk on a project has been identified and estimated.

3. Risk Mitigation Planning

The risk owners are responsible for planning and implementing mitigation actions with support from the project team.

All team members, inclusive of partners and suppliers, may be requested to identify and develop mitigation measures for identified risks. The project core team members are responsible for identifying an appropriate action owner for each identified risk.

After mitigation actions are defined, the project core team will review the actions. The risk owner must track all mitigation actions and expected completion dates. The risk owner and the project core team members must hold all action owners accountable for the risk mitigation planning.

Documenting the process allows us to systematically address the risk that exists in the project. Once the process is documented, action plans can be formulated and responsibilities allocated for controls that need to be implemented to eliminate or reduce the risk of those hazards.

An action plan for each identified risk can include multiple actions involving the team members. In case of risk occurrence, or risk triggers, the risk owner reports to the project manager and implements the response plans.

4. Risk Mitigation Action Implementation
The *action implementation* is the responsibility of the risk owner. The action owners are responsible for the execution of the tasks or activities necessary to complete the mitigation action and eliminate or minimize the risk. The risk owner or the project manager will monitor completion dates of the mitigation action implementation.

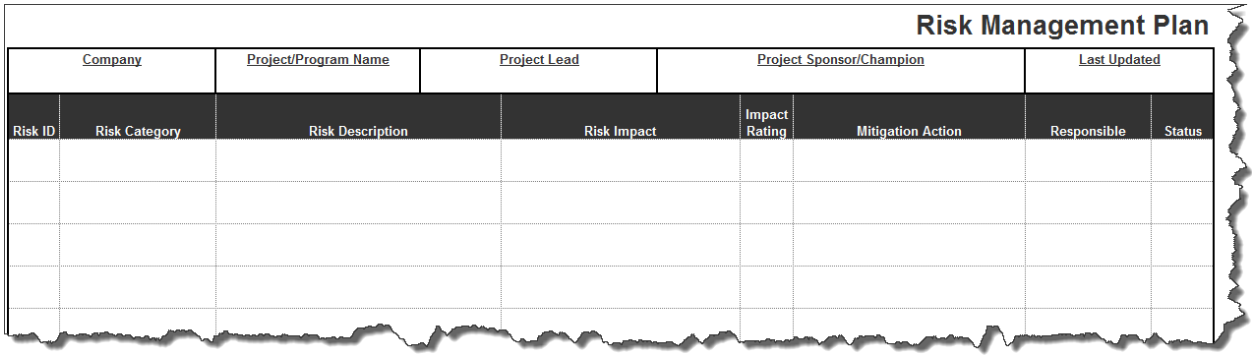
Risk Occurrence and Contingency Plans

Whenever any risk occurs, the project team should implement *contingency plans* to ensure that project deliverables can be met. The details of each occurrence should be recorded in the risk register or other tracking tool.

The *risk register* or *risk management plan* will be maintained by the project manager and reviewed regularly. The risk register is a document that contains the results of various risk management processes. It is often displayed in a table or spreadsheet format.

5. Risk Tracking and Reporting
Risk tracking and reporting provides critical visibility to all risks. Risk owners must report on the status of their mitigation actions.

The most efficient method to track and report risk is to do it as an integral part of project team meetings through all stages in the project lifecycle. Depending on the risk severity, project managers need to report the risk status of each category of risk to senior management in the form of dashboards or through weekly status reports.



Risk Management Plan

Company	Project/Program Name	Project Lead	Project Sponsor/Champion	Last Updated

Risk ID	Risk Category	Risk Description	Risk Impact	Impact Rating	Mitigation Action	Responsible	Status

Fig. 1.36 Risk Management Plan Example

6. Risk Closure

The risk owners are responsible for recommending the risk closure to the project manager.

A risk is *closed* only when the item is not considered a risk to the project anymore. When a risk is closed, the project manager needs to update the risk status in the lessons learned document.

Risk Analysis Features

The risk analysis should be:

- Systematic
- Comprehensive
- Data driven
- Adherent to evidence
- Logically sound
- Practically acceptable
- Open to critique
- Easy to understand

Project Risk Analysis Advantages

The risk analysis:

- Aids strategic & business planning
- Meets customer requirements
- Reduces schedule delays
- Minimizes cost overruns
- Promotes effective use of resources
- Promotes continuous improvement
- Helps to achieve project goals
- Prevents surprises
- Uncovers new opportunities
- Enhances communication

1.3.5 PROJECT PLANNING

What is Project Management?

Project management is the process of defining, planning, organizing, managing, leading, securing, and optimizing the resources to achieve a set of planned goals and objectives. It is the application of knowledge, skills, tools, and techniques to project activities in order to meet project requirements.

Project management involves understanding project needs, planning those needs, and properly allocating resources to achieve the desired results.

The basic steps of project management as we will cover in greater detail later are: initiating the project, planning the project, executing the project, and closing the project.

What is a Project Plan?

A *project plan* is a crucial step in project management for achieving a project's goals. A project plan is a formal approved document used to guide and execute project tasks. It provides an overall framework for managing project tasks, schedules, and costs. Project plans are coordinating tools and communication devices that help teams, contractors, customers, and organizations define the crucial aspects of a project or program. Project planning involves defining clear, distinct tasks and deliverables, and the work needed to complete each task and phase of the project.

Project Planning Stages

The stages of project planning are:

- Determine project scope and objectives: Explore opportunities, identify and prioritize needs, consider project solutions
- Plan the project: Identify input and resources requirements such as human resources, materials, software, hardware, and budgets.
- Prepare the project proposal: Based on stakeholder feedback, plan the necessary resources, timeline, budget etc.
- Implement the project: Implement the project by engaging responsible resources and parties. Ensure execution and compliance of the defined plans.
- Evaluate the project: Regularly review progress and results. Measure the project's effectiveness against *quantifiable* requirements

The five key stages of project planning are:

1. Determining project scope and objectives
2. Project planning
3. Project proposal and approval
4. Project implementation
5. Project evaluation

Planning and Scheduling Objectives

The goals of planning and scheduling are:

- To optimize the use of resources (both human and other resources)
- To increase productivity

- To achieve desired schedules and deliverables
- To establish an approach to minimize long-term maintenance costs
- To minimize the chaos and productivity losses resulting from planned production schedules, priority changes, and non-availability of resources
- To assess current needs and future challenges

Proper project planning is necessary to coordinate all resources and tasks.

Project Planning Activities

Project planning activities involve the creation of:

- Statement of work
- Work breakdown structure
- Resource estimation plan
- Project schedule
- Budget or financial plan
- Communication plan
- Risk management plan

Project Planning Activities: Statement of Work
Statement of Work (SOW) is a formal document often accompanying a contract that outlines specific expectations, limitations, resources and work guidelines intended to define the work or project.

SOWs:

- Define the scope of the project
- Establish expectations and parameters of the project
- Identify technical requirements for the project
- Provide guidance on materials to be used
- Establish timeline expectations

SOWs or statements of work are important "contracts" that outline many critical elements of a project or body of work. SOWs clarify expectations, deliverables, timelines, budgets, and responsibilities.

Project Planning Activities: Work Breakdown Structure
Work Breakdown Structure (WBS) is a decomposition of project components into small and logical bodies of work or tasks.

WBSs:

- Identify all required components of a project
- Cascades components into sub-components and tasks

WBSs are not by themselves project plans or schedules but they are a necessary step to help establish project plans and timelines. WBSs also enable logical reporting and summarizations of project progress.

Level 1	Level 2	Level 3
1.0 Bicycle	1.1 Frame	1.1.1 Set Body Frame
		1.1.2 Assemble Handlebars
		1.1.3 Install Seat & Seat Post
		1.1.4 Assemble Wheel Bearings & Axle
		1.1.5 Install Wheels
		1.1.6 Attach Brake System to Frame Set
	1.2 Wheels	1.2.1 Stage Rims
		1.2.2 Install Spokes
		1.2.3 Insert Tubes in Tires
		1.2.4 Assemble Tires on Rims
		1.2.5 Install Reflectors
	1.3 Brake System	1.3.1 Connect Brake Cables To Hand Levers
		1.3.2 Connect Brake Cables to Brake Harness
		1.3.3 Attach Brake Pads to Brake Harness
		1.3.4 Adjust/Calibrate Brake System

Fig. 1.37 Work Breakdown Structure Example

A WBS is an important parameter in project planning because projects must inevitably be divided into smaller, more manageable tasks and subtasks. These breakdowns must be aligned in a logical manner to meet customer defined target dates. The WBS is required to prepare a project schedule and/or Gantt chart as well as to identify project milestones.

Project Planning Activities: Resource Planning

Resource Estimation Plan

- Estimate Resource Requirements (Use Your WBS)
- Parts, Hardware, Software, Human Resources etc.
- Plan resources
- Establish who is responsible for what and when
- Determine quantity requirements and delivery dates etc.

Level 3	Parts Required	Quantity	Cost	Part Inventory	Responsible
1.1.1 Set Body Frame	Body Frame	1	$9.00	4	John
1.1.2 Assemble Handlebars	Handlebars	1	$4.00	4	John
1.1.3 Install Seat & Seat Post	Seat & Post	1 Each	$3.00	4 Each	John
1.1.4 Assemble Wheel Bearings & Axle	Bearing & Axle	2 Each	$1.00	8 Each	John
1.1.5 Install Wheels	Wheels	2	$0.50	8	John
1.1.6 Attach Brake System to Frame Set					John
1.2.1 Stage Rims	Wheel Rim	2	$0.30	12	Cathy
1.2.2 Install Spokes	Spokes	48	$0.03	72	Cathy
1.2.3 Insert Tubes in Tires	Tubes & Tires	2 Each	$0.33	0 Tubes 8 Tires	Cathy
1.2.4 Assemble Tires on Rims					Cathy
1.2.5 Install Reflectors	Reflectors	4	$0.10	18	Cathy
1.3.1 Connect Brake Cables To Hand Levers	Cables & Levers	2 Each	$0.90	12 Each	Lisa
1.3.2 Connect Brake Cables to Brake Harness	Brake Harness	2	$0.40	8	Lisa
1.3.3 Attach Brake Pads to Brake Harness	Brake Pads	4	$0.15	12	Lisa
1.3.4 Adjust/Calibrate Brake System					Lisa

Fig. 1.38 Resource Estimation Plan Example

The estimation of resource requirements (human, software, hardware, materials etc.) is necessary for each task and is useful for the creation of a project plan. Use your work breakdown structure to boil it down to tasks so that you can determine specific resource requirements, dependencies etc.

Project Planning Activities: Project Scheduling

Project Scheduling

- Assign beginning times to each activity in the WBS (days are used for start and durations times in example below)
- Assign duration times to each activity in the WBS
- Identify People Responsible and set completion dates
- Represent schedules as Gantt charts or network diagrams
- Identify critical dependencies between tasks

Level 2	Level 3	Task Start	Task Duration	Responsible
1.1 Frame	1.1.1 Set Body Frame	0	2	John
	1.1.2 Assemble Handlebars	2	2	John
	1.1.3 Install Seat & Seat Post	4	2	John
	1.1.4 Assemble Wheel Bearings & Axle	6	4	John
	1.1.5 Install Wheels	14	4	John
	1.1.6 Attach Brake System to Frame Set	18	3	John
1.2 Wheels	1.2.1 Stage Rims	0	1	Cathy
	1.2.2 Install Spokes	1	7	Cathy
	1.2.3 Insert Tubes in Tires	10	2	Cathy
	1.2.4 Assemble Tires on Rims	12	2	Cathy
	1.2.5 Install Reflectors	8	2	Cathy
1.3 Brake System	1.3.1 Connect Brake Cables To Hand Levers	0	8	Lisa
	1.3.2 Connect Brake Cables to Brake Harness	8	8	Lisa
	1.3.3 Attach Brake Pads to Brake Harness	16	2	Lisa
	1.3.4 Adjust/Calibrate Brake System	21	2	Lisa

Fig. 1.39 Project Scheduling Example

Scheduling is critical because without it you will miss deadlines and poorly organize dependent tasks and activities. The result of the lack of scheduling, or of poor scheduling, can be catastrophic to a project.

Project Planning Activities: Project Scheduling

Project Schedule – Gantt Chart

The advantage of a Gantt chart is its ability to display the status of each task/activity at a glance. Because it is a graphic representation, it is easy to demonstrate the schedule of tasks and timelines to all the stakeholders.

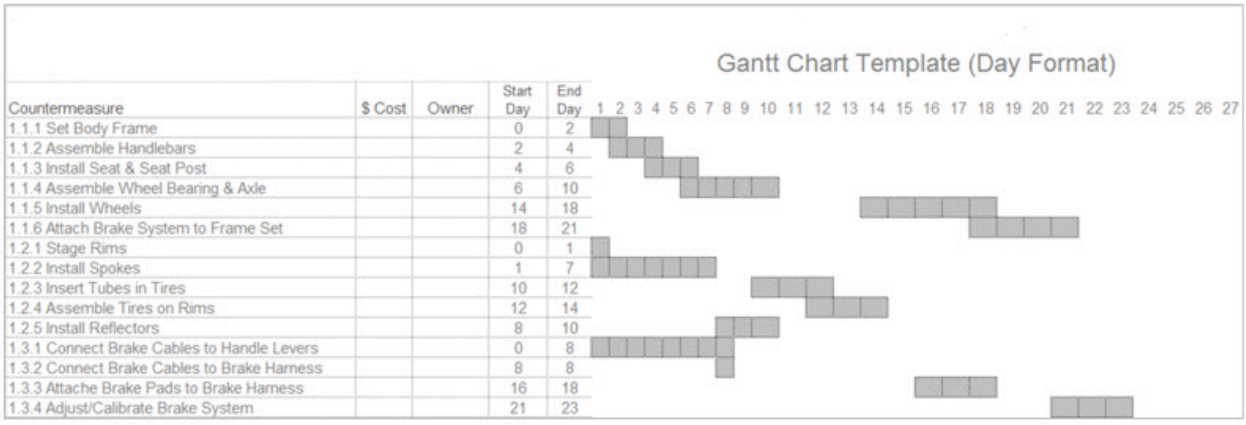


Fig. 1.40 Gantt Chart Example

A Gantt chart offers visibility to a project plan. It displays the tasks of the project plan in a simple and well-organized manner.

There are many tools available to create Gantt charts. Some of the most common and compatible for businesses are Microsoft Project, Visio, and Excel. In Microsoft Excel, there are no native Gantt chart templates so we have established a simple one to use with our bicycle example (shown in fig. 1.40). This template is available for free on our website under the "Tools" link.

Project Scheduling: Critical Path Method

Critical Path Method (CPM)

Critical path method (CPM) is a project modeling technique used to identify the set of activities that are most influential to a project's completion timeline. Critical path tasks are those that others are dependent upon.

Project timelines cannot be shortened without shortening the tasks or activities that are identified as critical path items.

Steps to Using the Critical Path Method

1. List all activities necessary to complete the project.
2. Determine the time or duration of each activity.
3. Identify the dependencies between the activities.

CPM is fundamental to project management. You must understand all required project tasks, their durations and dependencies. The critical path method can help you organize and understand how each task behaves relative to others and how the collection of all tasks fit in the grand scheme of a project.

Critical Path Method Example: Let's continue with our bicycle example referencing fig. 1.40. Note that the tasks in the WBS are already set to start based on their dependencies. For example, installing the wheels in step 1.1.5 requires the whole set of steps in 1.2 to be completed so 1.1.5 starts on day 14 when the wheels are ready.

By using your WBS, you can efficiently evaluate task durations and dependencies to find the critical tasks that will affect the final timeline of your project.

Continuing with this example; the last task, "1.3.4 Adjusting the Brake System" completes on day 23 and cannot begin until day 21 because the brake system must be attached to the bike frame before being calibrated. The reason for such a delay can be found in the earlier example. You should have noticed in the resource plan that tire tubes stock was depleted.

Let us assume the tubes take 21 days to receive from the supplier. Below is the adjustment to our Gantt chart. Notice that the project timeline is pushed out by ten days, finishing on day 33 instead of day 23.

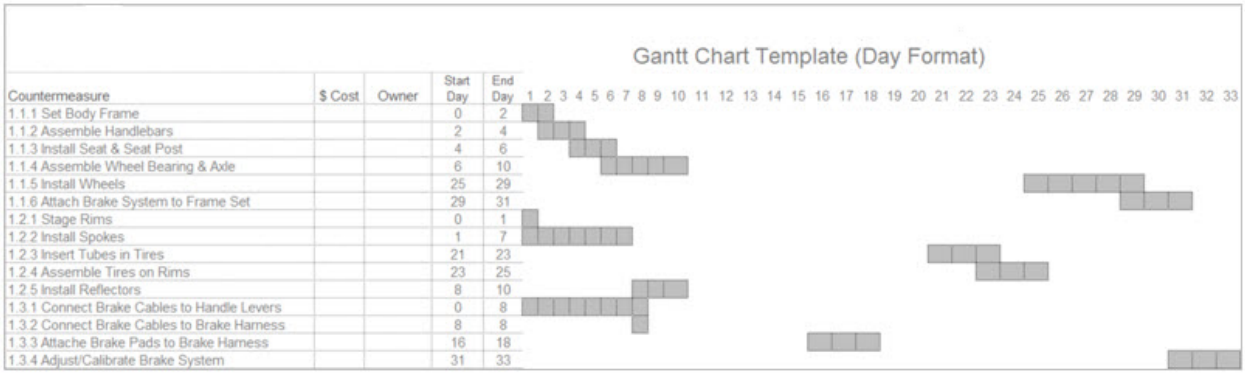


Fig. 1.41 Gantt Chart Revision (21 days for Tubes)

By coordinating the use of all project planning tools and methods covered this far (WBS, Resource Planning, Gantt Chart, CPM) you can not only depict the project tasks and timelines but you can also better manage resources as a result. The Tubes example demonstrates that there is nearly a full FTE resource wasted in waiting time. By knowing this you might be able to re-organize your projects tasks and resources to save on the cost of labor.

Project Scheduling: PERT

Program Evaluation and Review Technique (PERT)

The program evaluation and review technique (PERT) is a method of evaluation that can be applied to time or cost. PERT provides a weighted assessment of time or cost.

PERT uses three parameters for estimation:

1. Optimistic or Best Case Scenario represented by “O”
2. Pessimistic or Worst Case Scenario represented by “P”
3. Most Likely Scenario represented by “ML”

The PERT equation is:

$$E = \frac{O + 4ML + P}{6}$$

Where: E = estimate (of time or cost).

The PERT equation provides for heavier weighting of the most likely scenario but also considers the best and worst cases. In the event that a best or worst case scenario is an extreme situation, the PERT will account for it in a weighted manner.

Using the PERT formula, it is the sum of the estimates of the best case, the worst case, and four times the most likely case, all divided by six.

Let us apply the PERT formula to our estimates of receiving our needed tire tubes from the supplier and then change our schedule based on the result. Your experience with this supplier tells you that they typically overestimate the time required to deliver. Therefore, their 21 day lead time on the tubes should be considered a “worst case scenario.”

Procurement has indicated that tubes have arrived in as few as six days from this supplier. So, your best-case scenario is six days. The most likely scenario you decide will be the median of this supplier’s delivery time, which is 10 days.

Therefore:

$$E = \frac{O + 4ML + P}{6}$$

$$E = \frac{6 + 4\times10 + 21}{6}$$

$$E = \frac{6 + 40 + 21}{6}$$

$$E = \frac{67}{6}$$

$$E = 11$$

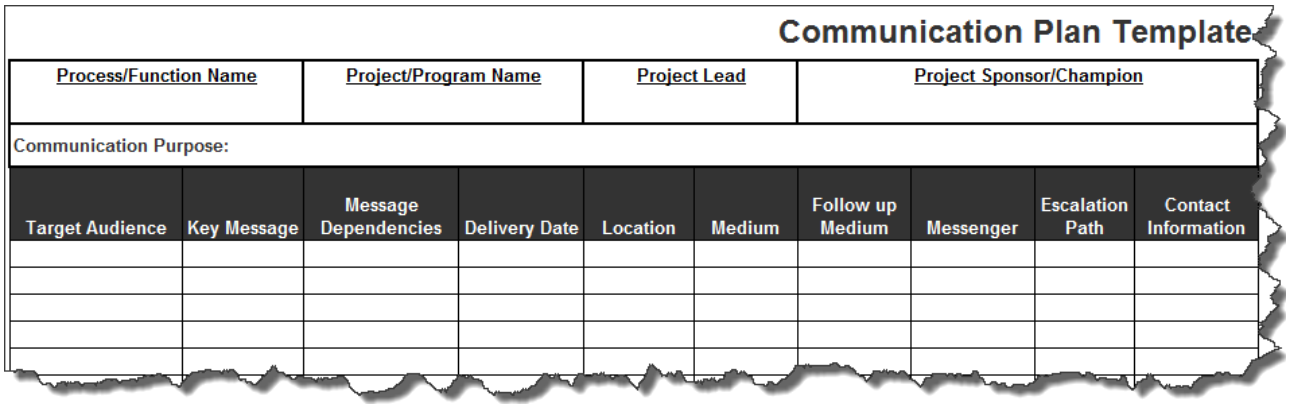
Using the PERT formula in this example: add the best-case scenario (6) to “four times the most likely scenario” (4*10) plus the worst case scenario (21). The result is sixty-seven. We then divide 67 by 6, which yields 11 (rounded). Eleven days is your adjusted estimate of delivery for

the tire tubes. Using 11 days in your project plan instead of 21 gives you a more confident estimate of time and removes tire tubes as the primary detracting critical path item.

Project Planning Activities: Budgeting or Financial Planning
Create a preliminary project budget that outlines the planned expenses and revenues pertaining to the project. Preliminary budgets are commonly used for project justification and future business forecasts.

Most organizations expect to see some financial analysis regarding the payback of a project. IRR (Internal Rate of Return) and NPV (Net Present Value) are two very common and acceptable financial measures that can be used to justify a project.

Project Planning Activities: Communication Plan
Communication plans are used to establish communication procedures among management, team members, and relevant stakeholders. It is appropriate and necessary to determine the communication schedule and define the acceptable modes of communication with your project team, stakeholders and steering committee.



Communication Plan Template

Process/Function Name	Project/Program Name	Project Lead	Project Sponsor/Champion
Communication Purpose:			

Target Audience	Key Message	Message Dependencies	Delivery Date	Location	Medium	Follow up Medium	Messenger	Escalation Path	Contact Information

Fig. 1.42 Communication Plan Example

Effective communication and frequency of communication are vital to any project. Keeping participants, stakeholders, and customers updated with progress and forecasts helps ensure good communication.

Project Planning Activities: Risk Management Plan
Risk management plans help to identify the sources of project risks and estimate the effects of those risks. Risks may arise from new technology, availability of resources, lack of inputs from customers, business risks or other unexpected sources.

It is important to assess the impact of risk to customers and project stakeholders. Calculate the probability of risk occurrence based on previous or similar projects and industry benchmarks. Then put a plan together to manage the risks.

Most projects that fail are due to the lack of, or improper, risk management. Identify sources of risk as early as possible in the project. Determine the severity and frequency of possible risk

occurrences and establish controls or mitigation plans. Prioritize your energy and resources based on risk severity.

Project Planning Tools Advantages

Project planning tools are very useful to organize and communicate project plans, status, and projections. The project planning tools we just reviewed (WBS, Schedule, Gantt, PERT etc.) are some of the most popular and readily available tools and methods used by professionals today.

Project planning tool advantages:

- Help link tasks and sub-tasks or other work elements to get a whole view of what needs to be accomplished
- Allow a more objective comparison of alternative solutions and provide consistent coverage of responsibilities
- Allow for effective scope control and change management
- Facilitate effective communication with all project participants and stakeholders
- Help define management reviews
- Act as an effective monitoring mechanism for the project
- Establish project baselines for progress reviews and control points

1.4 LEAN FUNDAMENTALS

1.4.1 LEAN AND SIX SIGMA

What is Lean?

A Lean enterprise is one which intends to eliminate waste and allow only value to be pulled through its system. A Lean enterprise can be achieved by identifying and effectively eliminating all waste (which will result in a flowing, cost-effective system).

A Lean manufacturing system drives value, flows smoothly, maximizes production, and minimizes waste. Lean manufacturing is characterized by:

- Identifying and driving value
- Establishing flow and pull systems
- Creating production availability and flexibility
- Zero waste
- Waste elimination
- Waste identification and elimination is critical to any successful Lean enterprise
- Elimination of waste enables flow, drives value, cuts cost, and provides flexible and available production

The Five Lean Principles

The following five principles of Lean are taken from the book *Lean Thinking* (1996) by James P. Womack and Daniel T. Jones.

1. Specify value desired by customers.
2. Identify the value stream.
3. Make the product flow continuous.
4. Introduce pull systems where continuous flow is possible.
5. Manage toward perfection so that the number of steps and the amount of time and information needed to serve the customer continually falls.

Principle 1: Specify Value Defined by Customers

Only a small fraction of the total time and effort spent in an organization actually adds value for the end customer. With a clear definition of value (from the customer's perspective), it is much easier to identify where the waste is.

Principle 2: Identify the Value Stream
The value stream is the entire set of activities across all parts of the organization involved in jointly delivering the product or service. It is the end-to-end process that delivers the value to the customer. As stated above, once you understand what determines value to the customer, it is easier to determine where the waste is.

Principle 3: Create Flow
Typically, you will find that only 5% of activities add value, but the percentage can be as high as 45% in a service environment. Eliminating most or all of this waste ensures that your product or service flows better. Thus, bringing timely, desired value to the customer without interruption or waiting.

Principle 4: Create Pull
When possible, understand the customer demand on your product or service. Then, create a process to respond to it. In other words, only produce what the customer wants when they want it.

Principle 5: Pursue perfection
At first, reorganizing process steps can improve flow, and as this is done, layers of waste become exposed and exploited. Continuing the process of eliminating, reorganizing, and exploiting waste will continue to drive down the number of steps and time needed to serve the customer. This continuous cycle should be pursued to perfection.

Lean and Six Sigma

Lean and Six Sigma both have the objectives of producing high value (quality) at lower costs (efficiency). They approach these objectives in somewhat different manners but in the end,

both Lean and Six Sigma drive out waste, reduce defects, improve processes, and stabilize the production environment.

Lean and Six Sigma are a perfect combination of tools for improving quality and efficiency. The two methodologies complement each other to improve quality, efficiency, and ultimately profitability and customer satisfaction.

1.4.2 History of Lean

History of Lean

Lean thinking has been traced as far back as the 1400s but the concepts and principles of Lean were more rigorously and methodically applied starting in the early 1900s. The first person to truly apply the principles in a production process was Henry Ford. He established the first mass production system in 1913 by combining standard parts, conveyors, and work flow. In 1913, Ford brought together the concepts of consistently interchangeable parts, standard work, and moving conveyance to create what he called *flow production*.

Decades later, between 1948 and 1975, Kiichiro Toyoda and Taiichi Ohno at Toyota improved and implemented various new concepts and tools (e.g., value stream, takt time, Kanban etc.) many based on Ford's efforts.

Toyota developed what is known today as the Toyota Production System (TPS) based in Lean principles. TPS advanced the concepts of Ford's system through a series of innovations that provided the continuity of flow *and* a variety of product offerings.

The term Lean, was later coined in the 1990's by John Krafcik who was a graduate student at MIT working on a research project for the book The Machine That Changed the World by Jim Womack. Womack, the founder of the Lean Enterprise Institute helped to refine and advance collectively, the principles known today as Lean.

1.4.3 Seven Deadly Muda

The Seven Deadly Muda

Muda is a Japanese word meaning "futility, uselessness, idleness, superfluity, waste, wastage, wastefulness." The seven commonly recognized forms of waste, also called the "Seven Deadly Muda," as defined by Taiicho Ohno (the Toyota Production System), are:

1. **Defects**
2. **Overproduction**—Producing more than what your customers are demanding
3. **Over-Processing**—Unnecessary time spent (e.g., relying on inspections instead of designing a process to eliminate problems)
4. **Inventory**—Things awaiting further processing or consumption

5. **Motion**—Extra, unnecessary movement of employees
6. **Transportation**—Unnecessary movement of goods
7. **Waiting**—For an upstream process to delivery, for a machine to finish processing, or for an interrupted worker to get back to work

The Seven Deadly Muda: Defects
Defects or defectives are an obvious waste for any working environment or production system. Defects require rework during production or afterwards when the product is returned from an unhappy customer.

Some defects are difficult to solve and often create workarounds and hidden factories. Nobody likes defects. They are an obvious form of waste—wasted time, wasted materials, and wasted money. Defects take time and resources to be fixed, and can be especially costly if they get into the hands of a customer. Eliminating defects is a sure way to improve product quality, customer satisfaction, and production costs.

The Seven Deadly Muda: Overproduction
Overproduction is wasteful because your system expends energy and resources to produce more materials than the customer or next function requires. Overproduction is one of the most detrimental of the seven deadly muda because it leads to other wastes such as overproduction, inventory, transportation and it also exacerbates existing wastes already inherent in the process such as defects, waiting or needless motion.

The Seven Deadly Muda: Over-Processing
Over-processing occurs any time more work is done than is required by the next process step, operation, or consumer. Over-processing also includes being over capacity (scheduling more workers than required or having more machines than necessary).

Another form of over-processing can be buying tools or software that are more precise, complex, or more expensive than required. You might think of this as over-engineering a process or just doing more than is required at a specific step of the process (having more labor available than necessary, having more machines than are needed, or having tools that are not adding incremental value because they do more than the product requires).

The Seven Deadly Muda: Inventory
Inventory is an often-overlooked waste and one that needs to be managed meticulously in order to optimize turns and resources. Take this book for instance. Our distributors require inventory so that they can pack and ship this book when the customer orders it. Unfortunately, inventory is costly up front while revenues lag. If we print too many books and never sell them, then we have excess waste in the form of inventory.

If on the other hand, we create too little inventory and our distributors run out of stock leaving orders unfulfilled then we have an opportunity cost in the form lost or delayed revenues due to poor inventory management.

In either case, inventory has a significant impact on cash flow which is a vital resource for businesses to survive and prosper. Proper inventory management is important to managing waste and opportunity costs.

The 7 Deadly Muda: Motion
Motion is another form of waste often occurring as a result of poor setup, configuration, or operating procedures. Wasted motion can be experienced by machines or humans and is exaggerated by repetition or recurring tasks.

Wasted motion is very common with workers who are unaware of the impact of small unnecessary movements in repetitive tasks. This can result from an improperly planned setup, work area configuration, and operating procedure. Small movements of an employee probably do not seem very wasteful by themselves. But when you consider repetitive motion that occurs over a long period, your perspective may change.

The Seven Deadly Muda: Transportation
Transportation is the unnecessary movement of goods, raw materials or finished products. Transportation is considered wasteful because it does *nothing* to add value or transform the product.

Imagine for a moment driving to and from work twice before getting out of your car to go into work . . . That is waste in the form of transportation. Driving less is better! In a similar way, the less transportation a product must endure, the better. There would be fewer opportunities for delay, destruction, loss, damage etc.

Transportation is wasteful because it takes energy to move something. And, it may take costly tools to move it. Transportation also introduces the risk of something getting damaged, lost, misplaced, delayed etc. Unless something is being transported into the hands of the customer, it probably means it is going somewhere to be stored (which is also wasteful).

The Seven Deadly Muda: Waiting
Waiting is an obvious form of waste and is typically a symptom of an upstream problem. Waiting is usually caused by inefficiency, bottlenecks, or poorly-designed work flows within the value stream, but it can also be caused by inefficient administration.

Reduction in waiting time will require thoughtful applications of Lean and process improvement. Waiting is when a downstream process is being starved because there is something wrong upstream in the process (like a machine having to finish processing or a worker that has stepped away from his or her station). In other words, there is a "bottleneck" that inhibits the flow of the process.

1.4.4 FIVE-S (5S)

What is 5S?

5S is systematic method used to organize, order, clean, and standardize a workplace . . . and, to keep it that way! 5S is a methodology originally developed in Japan and is used for improving the Lean work environment. 5S is summarized in five Japanese words all starting with the letter S:

1. Seiri (sorting)
2. Seiton (straightening)
3. Seiso (shining)
4. Seiketsu (standardizing)
5. Shisuke (sustaining)

To separate the needed from the unnecessary and prioritize what is needed. Sorting in a 5s activity is accomplished by many practitioners using the red tag method whereby red tags are placed on items in a work environment that are no longer useful or need to be removed. A common mantra in the sorting stage is "when in doubt" "throw it out".

Seiton (Set in order or straighten)
To set in order is to organize the work environment. This means to place all tools and equipment (those remaining from the sort stage) into the most useful, logical and accessible locations. Organization should be such that it promotes productivity and avails itself to the concepts of the visual factory.

Another common mantra used in 5S is "A place for everything and everything in its place". This saying keeps the focus on orderliness and not just during the 5S activity but for many years thereafter.

One of the most memorable examples of proper "set in order" results is that of the tool outlines on a peg board. When a tool is removed from the board it becomes very evident that the tool is no longer in its storage location. If it's not in use, then it should be where it can be found by other workers.

Seiso (Shine)
Shine means to clean the work area. This is not a casual sweep and "de-clutter" activity either. Shine in 5S calls for "getting dirty" while cleaning. Examples are machinery to be maintained, equipment cleaned and treated or painted, machinery vacuumed, equipment greased, lubed and brought to optimal performance levels.

This is the stage where a measure of pride becomes evident, when other departments and other work areas begin to express interest in conducting 5S on their own work environments.

Seiketsu (Standardize)
Standardizing is the stage where work practices are staged to be performed consistently. Clear roles and responsibilities are outlined and everyone is expected to know their role.
Standardize also means to make workstations and tool layouts common among like functions

and normalize equipment and parts across all workstations so that workers can be moved around to any workstation at any time and perform the same task.

Shisuke (Sustain)
Sustaining in 5S is also called self-discipline. It takes discipline to maintain 5S. Obviously, it takes time for new habits to form, and there needs to be a culture within the team to follow this methodology. It is critical not to fall back into old ways. Therefore, it is critical that the sustain stage of 5s:

- Creates the culture in the team to follow the first four S's consistently.
- Avoids falling back to the old ways of doing things with cluttered and unorganized work environments.
- Establishes and maintains the momentum of optimizing the workplace.
- Promotes innovations of workplace improvement.
- Sustains the first four stages using:
    - 5S Maps
    - 5S Schedules
    - 5S Job cycle charts
    - Integration of regular work duties
    - 5S Blitz schedules
    - Daily workplace scans to maintain and review standards

Goals and Benefits of 5S

The goals of 5S are to reduced waste, cut unnecessary expense, optimize efficiency and establish a work environment that is:

- Self-explaining
- Self-ordering
- Self-regulating
- Self-improving

Thus, 5S should create a work environment where this is no more wandering and searching for tools or parts. Workers should experience fewer delays or wait time due to equipment down time and the overall workplace will be safer, cleaner and more efficient. Some of the more common benefits reported by 5S participants are:

- Reduced changeovers
- Reduced defects
- Reduced waste
- Reduced delays
- Reduced injuries
- Reduced breakdowns
- Reduced complaints
- Reduced red ink
- Higher quality
- Lower costs
- Safer work environment
- Greater capacity

By following the disciplined 5S method and meeting the goals established by the 5S team, the following quantifiable benefits have been realized:

- Cut in floor space: 60%
- Cut in flow distance: 80%
- Cut in accidents: 70%
- Cut in rack storage: 68%
- Cut in number of forklifts: 45%
- Cut in changeover time: 62%
- Cut in physical inventory: 50%
- Cut in training requirements: 55%
- Cut in nonconformance: 96%
- Increase in test yields: 50%
- Late deliveries: 0%
- Increase in throughput: 15%

5S Disclaimer

As a method, 5S is one of many Lean methods that creates immediate and observable improvements. It is tempting and often attempted by organizations to implement 5S alone without considering the entire value stream. However, it is advisable to consider a well-planned Lean manufacturing approach to the entire production system.

2.0 MEASURE PHASE

2.1 PROCESS DEFINITION

2.1.1 CAUSE AND EFFECT DIAGRAM

What is a Cause and Effect Diagram?

A *cause and effect diagram* is also called a *fishbone diagram* or *Ishikawa diagram*. It was created by Kaoru Ishikawa and is used to identify, organize, and display the potential causes of a specific effect or event in a graphical way similar to a fishbone.

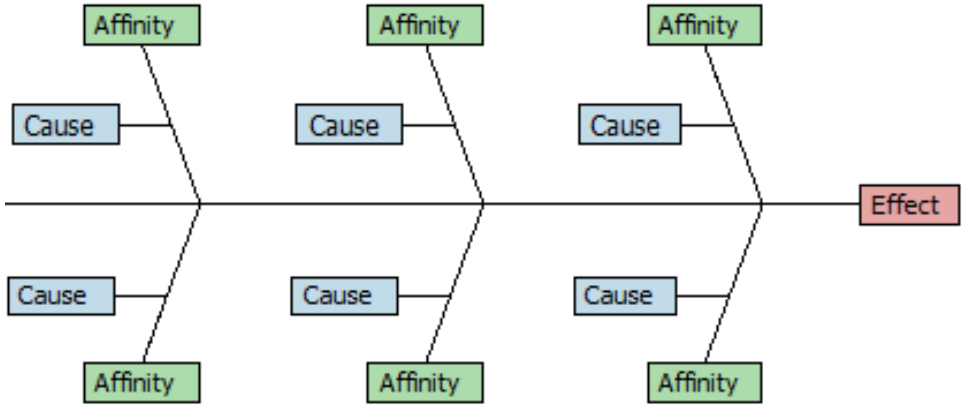


Fig. 2.1 Cause & Effect Diagram

A powerful brainstorming tool, the cause and effect diagram illustrates the relationship between one specified event (output) and its categorized potential causes (inputs) in a visual and systematic way.

Major Categories of Potential Causes

Typically, the possible causes fall into six major categories (or branches in the diagram). The acronym P4ME is a helpful way to remember the following standard categories which can be used when no other categories are known or obvious for your project:

- **P**eople: People who are involved in the process
- **M**ethods: How the process is completed (e.g., procedures, policies, regulations, laws)
- **M**achines: Equipment or tools needed to perform the process
- **M**aterials: Raw materials or information needed to do the job
- **M**easurements: Data collected from the process for inspection or evaluation
- **E**nvironment: Surroundings of the process (e.g., location, time, culture).

How to Plot a Cause and Effect Diagram

Step 1: Identify and define the effect being analyzed.

- Clearly state the operational definition of the effect or event of interest.

- The effect can be the positive outcome desired or negative problem targeted to solve.
- Enter the effect in the end box of the fishbone diagram and draw a spine pointed to it.
- Plotting a cause and effect diagram is a simple four-step process.

Step 2: Brainstorm the potential causes or factors of the effect occurring.

- Identify any factors with a potential impact on the effect and include them in this step.
- Put all the identified potential causes aside for use later.

Step 3: Identify the main categories of causes and group the potential causes accordingly.

- Besides P4ME (i.e., people, methods, machines, materials, measurements, and environment), you can group potential causes into other customized categories.
- Below each major category, you can define sub-categories and then classify them to help you visualize the potential causes.
- Enter each cause category in a box and connect the box to the spine. Link each potential cause to its corresponding cause category.

You can use P4ME, or you can use customized main categories to group the potential causes. Under each main category, you can define sub-categories if additional levels of detail are needed.

Step 4: Analyze the cause and effect diagram.

- A cause and effect diagram includes all the possible factors of the effect being analyzed.
- You can use a Pareto chart to filter the causes the project team needs to focus on.
- Identify causes with high impact that the team can act upon.
- Determine how to measure causes and effects quantitatively.
- Prepare for further statistical analysis.

Benefits to Using Cause and Effect Diagram

The benefits to using a cause and effect diagram to analyze an effect or event are:

- Helps to quickly identify and sort the potential causes of an effect.
- Provides a systematic way to brainstorm potential causes.
- Identifies areas requiring data collection for further quantitative analysis.
- Locates “low-hanging fruit.”

Limitation of Cause and Effect Diagrams

- A cause and effect diagram only provides qualitative analysis of correlation between each cause and the effect. More analysis is needed to quantify relationships between causes and effects.

- One cause and effect diagram can only focus on *one* effect or event at a time. Effects can have different causes.
- Further statistical analysis is required to quantify the relationship between various factors and the effect and identify the root causes.

Cause and Effect Diagram Example

Let us follow an example through the process of using a cause and effect diagram. A real estate company is interested to find the root causes of high energy costs of its properties. The cause and effect diagram is used to identify, organize, and analyze the potential root causes.

Step 1: Identify and define the effect being analyzed: In this example, the effect we are concerned with is high energy costs of buildings.

Step 2: Brainstorm the potential causes or factors of the high-energy costs: A team is assembled to brainstorm possible reasons for high energy costs of buildings. They provided the following list of possible causes:

- Poor Maintenance
- Bad Habits
- Lighting Schedule
- HVAC Schedule
- Temperature Set Point
- Old HVAC Units
- Open Dampers
- Light Bulbs
- Inefficient Building Materials
- Meter Float
- Meter Accuracy
- Increasing Fuel Cost
- Building Air Leakage
- Humidity

Step 3: Identify the main categories of causes and group the potential causes accordingly. The team used the standard P4ME categories for this exercise and bucketed the causes above based on the most appropriate categories.

Step 4: Analyze the cause and effect (C&E) diagram.

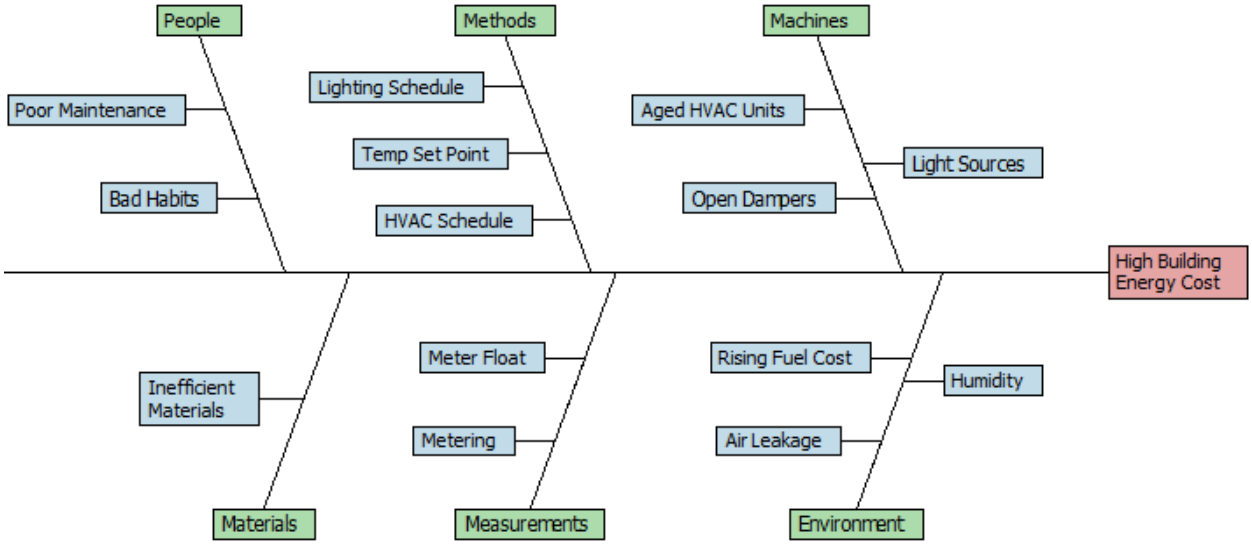


Fig. 2.2 Populated Cause & Effect Diagram with P4ME Categories

After completing the C&E diagram, the real estate company conducts further research on each potential root cause. It is discovered that:

- The utility metering is accurate
- The building materials are acceptable and there is no significant air leakage
- The fuel prices increased recently but were negligible
- Most lights are off during the non-business hours except that some lights have to be on for security purposes
- The temperature set points in the summer and winter are both adequate and reasonable
- The high-energy costs are probably caused by the poor HVAC maintenance on aged units and the wasteful energy consuming habits

Next, the real estate company needs to collect and analyze the data to check whether root causes identified in the C&E diagram have any validity. Data should be used to support this conclusion. Therefore, the team need to consider how to collect data to prove or disprove their assertion.

2.1.2 Cause and Effects Matrix

What is a Cause and Effect Matrix?

The *cause and effect matrix* (XY Matrix) is a tool to help subjectively quantify the relationship of several X's to several Y's. Among the Y's under consideration, two important ones should be the *primary* and *secondary metrics* of your Six Sigma project. The X's should be derived from your cause and effect diagram. Let us take a peek as to what it looks like on the next page.

A cause and effect matrix is a powerful yet easy tool to transition from brainstorming (which you did using the cause and effect diagram) to estimating, ranking, and quantifying the

relationship between those things you found during brainstorming and the Y's that are important to the project.

Cause and Effects Matrix

Cause & Effects Matrix *(XY Matrix)*

Date:
Project:
XY Matrix Owner:

Output Measures (Y's)*	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	
Weighting (1-10):											
Input Variables (X's)#	*For each X, score its impact on each Y listed above (use a 0,3,5,7 scale)*										Score
X_1											0
X_2											0
X_3											0
X_4											0
X_{29}											0
X_{30}											0

XY Matrix Premis: The XY Matrix or "Cause & Effect Matrix functions on the premis of the Y=f(x) equation.
*Rate each "Y" on a scale of 1 to 10 with 1 being the least important output measure
#For each X rate its impact on each Y using a 0,3,5,7 scale (0=No impact, 3=Weak impact, 5=Moderate impact, 7=Strong

Fig. 2.3 Cause & Effects Matrix

Figure 2.3 represents a clean cause and effect matrix where you can translate all of the potential X's brainstormed during the creation of your cause and effect diagram session.

Once you have that information entered into the matrix you can add one or more Y's to the top of the matrix. The Y's should be weighted relative to each other based on their importance to the project.

Once this is complete, you can begin rating each X against each Y on a scale of 0, 3, 5, 7 (0 means no impact; 3 means weak impact; 5 is moderate impact; and 7 is strong impact). Try to use the whole spectrum of your scale during this process because you will need the resolution to help differentiate the X's from one another.

How to Use a Cause and Effect Matrix

1. Across the top enter your output measures. These are the Y's that are important to your project.
2. Next, give each Y a weight. Use a 1–10 scale, 1 being least important and 10 most important.
3. Below, in the leftmost column, enter all the variables you identified with your cause and effect diagram.
4. Within the matrix, rate the strength of the relationship between the X in the row and the corresponding Y in that column. Use a scale of 0, 3, 5, and 7.

5. Lastly, sort the “Score” column to order the most important X’s first.

Cause and Effect Matrix Notes

When complete, sort the X's based on their scores and graph them (use a Pareto chart). What the results will tell you should not be "taken to the bank" as this entire process from brainstorming X's to rating them against Y's is a subjective exercise.

However, if you have chosen your team wisely, it is very common that someone or several SMEs on your project team know the real solutions and/or the best way to narrow down to find them. Therefore, it is also important not to dismiss the results of your cause and effect matrix. Look at the image with comments embedded to summarize what you have just read.

After You Have Completed the C&E Matrix

After you have completed your cause and effects matrix, build a strategy for validating and/or eliminating the x’s as significant variables to the Y=f(x) equation.

- Build a data collection plan
- Prepare and execute planned studies
- Perform analytics
- Review results with SMEs, etc.

The combination of the C&E diagram and the C&E matrix is powerful. This is where some of the best ideas and theories about the Y=f(x) equation begin.

Now, it is time to start eliminating and/or validating these variables. Put a plan together to do so. Use the resources and tools you have already learned as well as those that you are about to learn throughout this course.

2.1.3 Failure Modes and Effects Analysis (FMEA)

What is FMEA?

Failure Modes & Effects Analysis - **FMEA**

Product or Process Step	Potential Failure Mode	Potential Failure Effects	S	Potential Causes	O	Current Controls	D	RPN	Recommended Actions	Responsible	Actions Taken	S	O	D	RPN

Fig. 2.4 Failure Modes & Effects Analysis Template

Figure 2.4 is a template of an FMEA. The *FMEA (Failure Modes and Effects Analysis)* is a tool and an analysis technique used to identify, evaluate, and prioritize potential deficiencies in a process so that the project team can design action plans to reduce the probability of those failures from occurring.

FMEA activity often follows process mapping. A process map is commonly used during the FMEA to determine where potential process breakdowns can occur. At each step, the team must ask, "What can go wrong?" A Pareto analysis may follow to quantitatively validate or disprove the results of the FMEA.

FMEA is best completed in cross-functional brainstorming sessions where attendees have a good understanding of the entire process or of a segment of it.

Important FMEA Terms

- *Process Functions* - Process steps depicted in the process map. FMEA is based on a process map and one step/function is analyzed at a time.
- *Failure Modes* - Potential and actual failure in the process function/step. It usually describes the way in which failure occurs. There might be more than one failure mode for one process function.
- *Failure Effects* - Impact of failure modes on the process or product. One failure mode might trigger multiple failure effects.
- *Failure Causes* - Potential defect of the design that might result in the failure modes occurring. One failure mode might have multiple potential failure causes.
- *Severity Score* - The seriousness of the consequences of a failure mode occurring. Ranges from 3 to 9, with 9 indicating the most severe consequence.
- *Occurrence Score* - The frequency of the failure mode occurring. Ranges from 3 to 9, with 9 indicating the highest frequency.
- *Detection Score* - How easily failure modes can be detected. It is a rating for how effective any controls are in the process. Ranges from 3 to 9, with 9 indicating the most difficult detection.
- *RPN (Risk Prioritization Number)* - The product of the severity, occurrence, and detection scores. It is a calculated value that ultimately is used to prioritize the FMEA. Ranges from 1 to 1000. The higher RPN is, the more focus the particular step/function needs.
- *Recommended Actions* - The action plan recommended to reduce the probability of failure modes occurring or at least enable better controls to detect it when failure occurs.
- *Current Controls* - Procedures currently conducted to prevent failure modes from happening or to detect the failure mode occurring.

How to Conduct an FMEA

Conducting an FMEA is not a difficult process, but can be time consuming when done correctly, particularly the first time it is performed on a process. To demonstrate how an FMEA is used we will be sharing a super-simple example.

Joe is trying to identify, analyze, and eliminate the failure modes he experienced in the past when preparing his work bag before heading to the office every morning. He decides to run an FMEA for his process of work bag preparation. There are only two steps involved in the process: putting the work files in the bag and putting a water bottle in the bag.

Failure Modes & Effects Analysis - **FMEA**

Product or Process Step	Potential Failure Mode	Potential Failure Effects	S	Potential Causes	O	Current Controls	D	RPN	Recommended Actions	Responsible	Actions Taken	S	O	D	RPN

Fig. 2.5 FMEA Header

Figure 2.5 shows a clearer picture of the header of the FMEA template. For the example we are about to share, we will be starting in the left column and for the most part, we will work through the template from left to right.

FMEA Step 1: List the critical functions of the process based on the process map created. Here you see that the process steps are listed under the "Process Function" heading. In this case, the process is extremely simple so that we can focus on how to use FMEA as a tool. The two process steps populated are Joe putting work files in the bag and Joe putting his water bottle in the bag.

Product or Process Step	Potential Failure Mode	Potential Failure Effects
Place files in bag		
Put water bottle in bag		

Fig. 2.6 FMEA List Process Steps

FMEA Step 2: List all the potential failure modes that might occur in each function. In figure 2.7 you can see that we have chosen to list only one failure for each process step. In your case, be sure to be thorough and not overlook any possible failure modes. We could add other failure modes such as placed files in wrong pocket, placed bent/crumpled file in bag etc.

Product or Process Step	Potential Failure Mode	Potential Failure Effects
Place files in bag	Incorrect files put in the bag	
Put water bottle in bag	Water leaks	

Fig. 2.7 FMEA Potential Failure Modes

FMEA Step 3: List all the potential failure effects that might affect the process. In this situation, if Joe places the wrong files in his bag, his work will be delayed when he arrives at his

destination because he will have the wrong files. Or, his files will be destroyed because his water bottle leaked.

Product or Process Step	Potential Failure Mode	Potential Failure Effects
Place files in bag	Incorrect files put in the bag	Work is delyed
Put water bottle in bag	Water leaks	Files in bag damaged

Fig. 2.8 FMEA Failure Effects

FMEA Step 4: List all the possible causes that may lead to the failure mode happening. Joe may have had a loose cap on his water bottle or disorganized files which caused his failures. There can be more than one cause to the same failure. List each one separately.

Potential Failure Effects	S	Potential Causes	O
Work is delyed		Files are not organized well	
Files in bag damaged		Cap on water bottle not tight	

Fig. 2.9 FMEA Failure Causes

FMEA Step 5: List the current control procedures for each failure cause. In many cases, you'll find that there are no controls or weak controls relative to your failure causes. The FMEA will help you and your team identify these weaknesses.

Potential Causes	O	Current Controls	D
Files are not organized well		Check if files are needed	
Cap on water bottle not tight		Check bottle cap before inserting	

Fig. 2.10 FMEA Current Controls

FMEA Step 6: Determine the severity rating for each potential failure mode. At this point it is necessary to attempt to quantify the failure effects in terms of their severity. As you progress, will see that there will be 3 ratings to be assigned in an FMEA (Severity, Occurrence and Detection). These quantifications will result in a total "Risk Priority Number" (RPN) which will help you determine the failure modes that are most detrimental to your process.

For severity rankings. We recommend using a scale range of 3,5,7,9 with 9 representing the most severe and 3 being the least severe. If a failure mode has more than one cause, severity rankings should be the same for each cause because you're ranking the severity of the failure not of the cause.

Potential Failure Effects	S	Potential Causes	O	Current Controls
Work is delyed	9	Files are not organized well		Check if files are needed
Files in bag damaged	7	Cap on water bottle not tight		Check bottle cap before inserting

Fig. 2.11 FMEA Failure Effects Severity Rating

FMEA Step 7: Determine the occurrence rating for each potential failure cause. Using the same 3,5,7,9 scale. Rate each failure causes based on how frequently you believe they occur. A rating of 9 will represent a high occurrence rating and 3 will represent an infrequent or low occurrence rating. If there is more than one cause to a single failure mode, rate each cause based on how frequently the failure occurs due to that particular cause.

Potential Failure Effects	S	Potential Causes	O	Current Controls
Work is delyed	9	Files are not organized well	3	Check if files are needed
Files in bag damaged	7	Cap on water bottle not tight	5	Check bottle cap before inserting

Fig. 2.12 FMEA Failure Occurrence Rating

FMEA Step 8: Determine the detection rating for each current control procedure. Detection ratings are intuitively different and inverse of the others two ratings. High detection ratings mean that control procedure has a poor ability to detect the failure when it occurs. Conversely, a low detection rating of 3 indicates that the control procedure is very capable of detection. This ensures that the product of all 3 ratings will yield an RPN that when ranked, the highest RPN will indicate which failure mode is in need of the most attention.

S	Potential Causes	O	Current Controls	D	RPN
9	Files are not organized well	3	Check if files are needed	5	135
7	Cap on water bottle not tight	5	Check bottle cap before inserting	5	175

Fig. 2.13 FMEA Failure Detection Rating

FMEA Step 9: Calculate the Risk Prioritization Number (RPN). The RPN is calculated by multiplying all 3 ratings together S x O x D = RPN.

S	Potential Causes	O	Current Controls	D	RPN
9	Files are not organized well	3	Check if files are needed	5	135
7	Cap on water bottle not tight	5	Check bottle cap before inserting	5	175

Fig. 2.14 FMEA Risk Priority Number (RPN)

FMEA Step 10: Rank the failures using the RPN and determine the precedence of problems or critical inputs of process. Ranking should be from highest to lowest.

S	Potential Causes	O	Current Controls	D	RPN
7	Cap on water bottle not tight	5	Check bottle cap before inserting	5	175
9	Files are not organized well	3	Check if files are needed	5	135

Fig. 2.15 FMEA Sort by RPN

FMEA Step 11: Brainstorm and create recommended action plans for each failure mode. There can be multiple actions per failure mode, or cause or control procedure. The purpose is to thoroughly address root causes as well as your ability to detect failures. The goal would be to prevent failures.

Current Controls	D	RPN	Recommended Actions
Check bottle cap before inserting	5	175	Obtain new water bottle
Check if files are needed	5	135	Organize & Categorize Files

Fig. 2.16 FMEA Recommended Actions

FMEA Step 12: Determine and assign task owners and projected completion dates for each action item. This is an important accountability step and will be further supported by management routines that review progress and the effectiveness of the action items.

D	RPN	Recommended Actions	Responsible
5	175	Obtain new water bottle	Joe
5	135	Organize & Categorize Files	Joe

Fig. 2.17 FMEA Action Responsibility

FMEA Step 13: Determine new severity ratings after actions have been taken. FMEAs are often considered living and breathing documents because they should continuously address failure modes and require continued updating to actions, rankings and RPNs.

Recommended Actions	Responsible	Actions Taken	S	O	D	RPN
Obtain new water bottle	Joe					0
Organize & Categorize Files	Joe					0

Fig. 2.18 FMEA After Action Failure Severity

FMEA Step 14: Determine the new occurrence ratings after actions are taken.

Recommended Actions	Responsible	Actions Taken	S	O	D	RPN
Obtain new water bottle	Joe					0
Organize & Categorize Files	Joe					0

Fig. 2.19 FMEA After Action Failure Occurrence Rating

FMEA Step 15: Determine new detection ratings after actions are taken.

Recommended Actions	Responsible	Actions Taken	S	O	D	RPN
Obtain new water bottle	Joe					0
Organize & Categorize Files	Joe					0

Fig. 2.20 FMEA After Action Failure Detection Rating

FMEA Step 16: Update the RPN based on new severity, occurrence, and detection ratings.

Recommended Actions	Responsible	Actions Taken	S	O	D	RPN
Obtain new water bottle	Joe					0
Organize & Categorize Files	Joe					0

Fig. 2.21 FMEA List Process Steps

Make sure the corrective action plan is robust. Make sure it is clear what actions are to be taken, who owns the task, and what the completion date is for the task. Re-assess the severity, occurrence, and detection ratings if the corrective action plan is carried out, and update the RPN. Continue this cycle as needed.

2.1.4 Theory of Constraints

What is Theory of Constraints (TOC)?

Processes, systems, and organizations are all vulnerable to their weakest part. Any manageable system is limited by constraints in its ability to produce more (and there is always at least one constraint).

A common analogy for the theory of constraints is, “a chain is no stronger than its weakest link.” This concept was introduced by Eliyahu M. Goldratt in his 1984 book titled *The Goal*.

Theory of Constraints (TOC) is an important tool for improving process flows. For most organizations, the goal is to make money. For some organizations (e.g., non-profits) making money is a necessary condition for pursuing the goal.

Constraints can come in many forms. Some examples are production capacity, material, logistics, the market (demand), employee behavior, or even management policy. The

underlying premise of TOC is that organizations can be measured and controlled through variation of three measures—throughput, operational expense, and inventory.

TOC Performance Measures

Making sound financial decisions based on these three measures is a critical requirement.

- Throughput—Rate at which a system generates money through sales
- Operational Expense—Money spent by the system to turn inventory into throughput
- Inventory—Money the system has invested in purchasing things it intends to sell

Why is throughput defined by sales? Goods are not considered an asset until sold. This contradicts the common accounting practice of listing inventory as an asset even if it may never be sold. Units that are produced but not sold are inventory.

The objective of a firm is to increase throughput and/or decrease inventory and operating expense in such a way as to increase profit, return on investment, and cash flow.

TOC Five Focusing Steps

The objective is to ensure ongoing improvement efforts are focused on the constraints of a system.

1. Identify the system's constraints.
2. Decide how to exploit the constraints.
3. Subordinate everything else to the decision in step 2.
4. Elevate the constraints.
5. If in previous steps a constraint has been broken, return to step 1, but do not allow inertia to cause a system's constraint.

TOC focuses on the output of an entire system versus a discrete unit of the components. The five focusing steps help to identify the constraint that towers above all others within the system. It is an iterative process. Once the largest constraint is addressed (strengthened), then the next weakest link in the chain should be addressed, and so on. It is an ongoing system of process improvement.

Logical Thinking Processes

	Focusing Step	Thinking Process	Tools
1	Identify the system's constraint(s)	• Identify the problems • Find the root causes	Cause and effect diagram
2	Decide how to exploit the constraint(s)	• Develop a solution	Future reality tree
3	Subordinate everything else to the decision in step 2	• Identify the conflict preventing the solution • Remove the conflict	Evaporating cloud
4	Elevate the constraint	• Construct and execute an implementation plan	Prerequisite tree Transition tree
5	If in previous steps a constraint has been broken, return to step 1, but do not allow inertia to cause a system's constraint		

Fig. 2.22 TOC Logical Thinking Processes

Used in conjunction with the five focusing steps. There are many logic tools to help a firm analyze and verbalize cause and effect:

- E-C-E (Effect-Cause-Effect) Diagramming
- Current reality tree, also known as cause and effect diagram -Root cause identification
- Future reality tree—Assess solutions
- Evaporating cloud—Conflict resolution tool
- Prerequisite tree—Implementation tool
- Transition tree—Implementation plan

Simulation Exercise
Resources needed:

- Three "production line" participants
- One timer per each production line participant
- Five small boxes of 15 widgets each (paperclips, pens/pencils, candy, etc.)

Widget Value Chain:

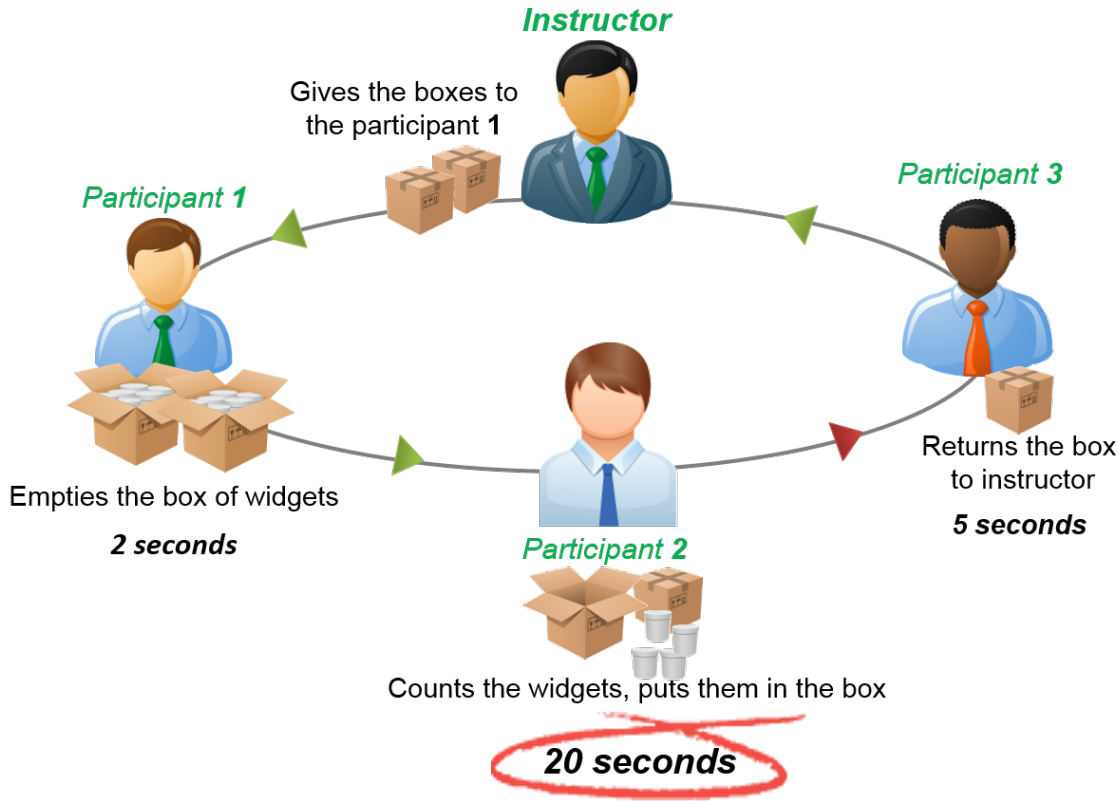


Fig. 2.23 TOC Widget Value Chain Simulation

Perform the process slowly the first time. The instructor gives the boxes, one at a time, to the first participant in the production line who empties the contents and hands the box and widgets to the next participant. The second participant counts the widgets, puts them in the box, and hands it to the third person who returns the box to the instructor. The instructor varies the rates at which the five boxes of widgets are handed to the first person in the "value chain" and times are recorded for each member's part in the process.

Considerations or Observations:

- As the boxes are distributed slowly, it is clear that all participants have plenty of time to do their activities, but they are all being starved of work and the "bottleneck" is external to the process (the instructor).
- As the instructor speeds up the process and hands over the boxes at a much faster rate, all three participants start to work faster, with the person in the middle always busy as the other two participants are waiting for him or her. In essence, the bottleneck has now changed to Person B, and a process constraint or bottleneck has been identified.
- Have the timers provide an average time for each of the three steps. Calculate throughput" for each step in the value chain (follow example below).

Participant 1: Time = 2 seconds (Throughput = 30 boxes/minute)
Participant 2: Time = 20 seconds (Throughput = 3 boxes/minute) ← Bottleneck
Participant 3: Time = 5 seconds (Throughput = 12 boxes/minute)

The efficiency of the process has slowed down to the slowest resource. In the example above, the output of the whole process is three boxes a minute; and Participant 1 and Participant 3 are not contributing to the overall efficiency. If they were to slow down, they would not have an adverse impact on the process, since they will always have idle time. Such resources with extra capacity are non-bottleneck resources.

Key question: What would you do to improve the efficiency of this system? Use the five focusing steps to lead to a conclusion and an action plan.

1. Identify the system's constraints.
2. Decide how to exploit the constraints.
3. Subordinate everything else to the decision in step 2.
4. Elevate the constraints.
5. If in previous steps a constraint has been broken, return to step 1, but do not allow inertia to cause a system's constraint.

Focusing Step 1: Identifying the constraint. In this exercise, "Count widgets and fill box" is the bottleneck; however, it is not always that easy to identify the bottleneck, so brainstorming tools like cause and effect diagrams are helpful.

Focusing Step 2: Determine how to exploit the constraint? This is where you would brainstorm potential solutions.

- More resources to perform step 2 (adds cost to the process).
- Sharing the work with Participant 1 or 3 (no added cost).
- Other ideas?

Focusing Step 3: Determine what conflicts prevent the solution?

- Perhaps there is resistance to spending more on additional resources?
- Maybe resources performing other steps require additional training/certification to perform another function?

Focusing Step 4: Determine the plan to implement?

- Hiring new resources? Training?
- What other considerations for implementing a solution?

Focusing Step 5: Implement the plan and then consider where the new bottleneck is.

2.2 SIX SIGMA STATISTICS

2.2.1 BASIC STATISTICS

What is Statistics?

Statistics is the science of collection, analysis, interpretation, and presentation of data. In Six Sigma, we apply statistical methods and principles to quantitatively measure and analyze process performance to reach statistical conclusions and help solve business problems.

sta·tis·tics (as defined on dictionary.com)

1. The science that deals with the collection, classification, analysis, and interpretation of numerical facts or data, and that, by use of mathematical theories of probability, imposes order and regularity on aggregates of more or less disparate elements.

2. The numerical facts or data themselves.

Types of Statistics

Descriptive statistics simply describe what is happening, while *inferential statistics* help you make comparisons, draw conclusions, or make judgments about data.

Descriptive Statistics
Descriptive statistics is applied to describe the main characteristics of a collection of data. It summarizes the features of the data quantitatively only and it does not make any generalizations beyond the data at hand. The data used for descriptive statistics are for the purpose of representing or reporting.

Inferential Statistics
Inferential statistics is applied to "infer" the characteristics or relationships of the populations from which the data are collected. Inferential statistics draws statistical conclusions about the population by analyzing a sample of data subject to random variation. A complete data analysis includes both descriptive statistics and inferential statistics.

Statistics vs. Parameters

The word *statistic* refers to a numeric measurement calculated using a sample data set, for example, sample mean or sample standard deviation. Its plural is *statistics* refers to the scientific discipline of statistical analysis.

Parameter refers to a numeric metric describing a population, for example, population mean and population standard deviation. Unless you have the full data set of the population, you will not be able to know population parameters. Almost always, you will be working with statistics, but they will allow you to make inferences about the population.

Continuous Variable vs. Discrete Variable

Data can take different forms, which is important to recognize when it is time to analyze the data. Continuous variables are measured and can be divided infinitesimally (they can take any

value within a range) and there is an infinite number of values possible. Examples are temperature, height, weight, money, time etc.

Discrete variables are sometimes referred to as count data which can literally be counted. Discrete data is finite with limited numbers of values available. Examples are count of people, count of countries, count of defects or count of defectives.

Types of Data

Nominal Data: Nominal Data are categorical but have no natural rank order. Examples of nominal data are colors, states, days of the week etc. Nominal comes from the Latin word "nomen" meaning "name" and nominal data are items that are differentiated by a simple naming system.

Be careful of nominal items that have numbers assigned to them and may appear ordinal but are not. These are used to simplify capture and referencing. Nominal items are usually categorical and belong to a definable category.

Ordinal Data: Ordinal Data are rank order data. Examples of ordinal data are the first, second and third place in a race or scores on an exam. Items on an ordinal scale are set into some kind of *order* by their position on the scale. The order of items is often defined by assigning numbers to them to show their relative position. Nominal and Ordinal data are categorical and thus referred to as discrete data.

Interval: Interval data are data measured on a quantifiable scale with equidistant values throughout the scale. Examples of interval data are temperature as measured on the Fahrenheit or Celsius scale.

Ratio Data: Ratio data are the ratio between the magnitude of a continuous value and the unit value of the same category. Examples of ratio data are weight, length, time etc.

In a ratio scale, numbers can be compared as multiples of one another. Thus, one person can be twice as tall as another person. Also important, the number zero has meaning. Thus, the difference between a person of 35 and a person 38 is the same as the difference between people who are 12 and 15. A person can also have an age of zero.

Interval and ratio data measure quantities and hence are continuous. Because they can be measured on a scale, they may also be referred to as scale data.

2.2.2 DESCRIPTIVE STATISTICS

Basics of Descriptive Statistics

Descriptive statistics

Descriptive statistics provide a quantitative summary for the data collected. It summarizes the main features of the collection of data (shape, location and spread). Descriptive statistics is a presentation of data collected and it does *not* provide any inferences about a more general condition. Instead, descriptive statistics describe what the data "looks like."

Shape of the Data

When we describe the "shape" of the data, we typically refer to the *distribution* of the data. Distribution, also called frequency distribution, summarizes the frequency of an individual value or a range of values for a variable (either continuous or discrete). It is sometimes referred to as a "frequency distribution" because it depicts how often a particular value occurs within a data set. Distribution is depicted as a table or graph.

Simple Example of Distribution

We are tossing a fair die. The possible value we obtain from each toss is a value between 1 and 6. Each value between 1 and 6 has a 1/6 chance to be hit for each toss. The distribution of this game describes the relationship between every possible value and the percentage of times the value is hit (or count of times the value is hit).

In this example, there is an equal probability for each value on the die occurring. If you were to roll the die a high number of times, what might the distribution look like? If you rolled it 600 times, you would likely get each value nearly 100 times.

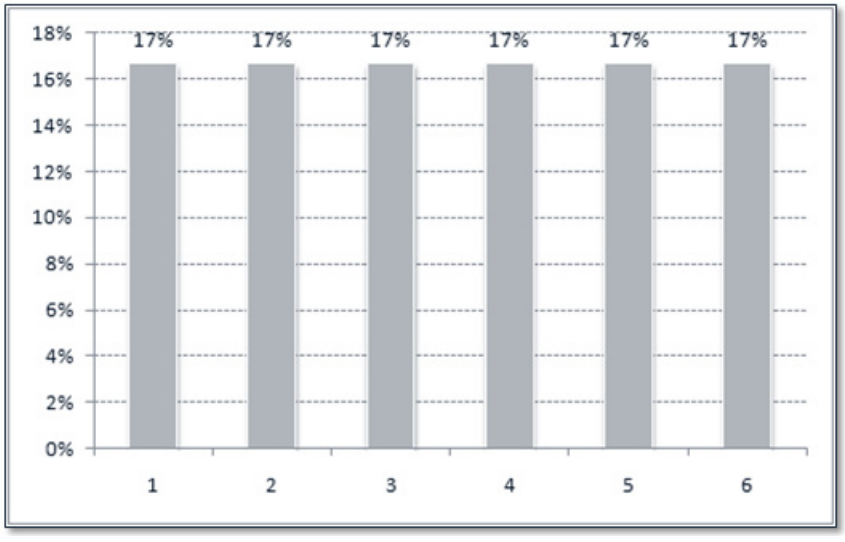


Fig. 2.24 Fair Die Distribution Probability

Common Types of Continuous Distributions

- Normal Distribution
- T distribution
- Chi-square distribution
- F distribution

Common Types of Discrete Distributions

- Binomial distribution
- Poisson distribution

There are different types of distributions that are commonly found in statistics. It is important to recognize that there are different types of data distributions because the statistics can tell you different things about the data and different inferential statistics have to be used for different situations.

Location of the Data
The location (i.e., central tendency) of the data describes the value where the data tend to cluster around. There are multiple measurements to capture the location of the data, the three most common and meaningful are Mean, Median and Mode.

Mean

The mean is the arithmetic average of a data set. It is easily calculated by dividing the sum of the data points by n (n is the number of values in the data set).

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For example, we have a set of data: 2, 3, 5, 8, 5, and 9. The arithmetic mean of the data set is

$$\frac{2+3+5+8+5+9}{6} = 5.33$$

In the example above, there are six data points. The sum of the six data points is 32, and when dividing by the number of data points, 6, the arithmetic mean is 5.33.

Median

The median is the middle value of the data set in numeric order. It separates a finite set of data into two even parts, one with values higher that the median and the other with values lower than the median. For example, we have a set of data: 45, 32, 67, 12, 37, 54, and 28. The median is 37 since it's the middle value of the sorted list of those values (i.e., 12, 28, 32, 37, 45, 54, and 67). There are three values lower and three values higher than the median.

Mode

The mode is the value that occurs most often in the data set. If no number is repeated, there is no mode for the data set. For example, we have a data set: 55, 23, 45, 45, 68, 34, 45, 55. The mode is 45 since it occurs most frequently. The numbers 23, 34, and 68 occur once, 55 occurs twice, and 45 occurs three times.

Spread of the Data

The spread (i.e., variation) of the data describes the degree of data dispersing around the center value. There are multiple measurements to capture the spread of the data, the most common and meaningful are Range, Variance and Standard Deviation.

Range

The range is the numeric difference between the greatest and smallest values in a data set. Only two data values (i.e., the greatest and the smallest values) are accounted for calculating the range. For example, we have a set of data: 34, 45, 23, 12, 32, 78, and 23. The largest value is 78, and the smallest is 12. Therefore, the range is 78 − 12 = 66.

Variance

The variance measures how far on average the data points spread out from the mean. More specifically, it is the average squared deviation of each value from its mean. All data points are accounted for by calculating the variance:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Where: *n* is the number of values in the data set.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

You will notice that the variance and standard deviation are directly related to each other. Why is the term squared? First, because squared terms always give a positive value and we are only trying to measure the distance from the mean (positives and negatives will negate each other). Second, squaring emphasizes the larger terms (or differences) between the mean and each individual data point.

Standard Deviation

Standard deviation describes how far the data points spread away from the mean. It is simply the square root of the variance. All data points are considered by calculating the standard deviation:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_1 - \bar{x})^2}$$

Where: *n* is the number of values in the data set.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2.2.3 NORMAL DISTRIBUTION AND NORMALITY

What is Normal Distribution?

Normal distribution, also called Gaussian distribution, is the probability distribution of a continuous random variable whose values spread symmetrically around the mean. A normal distribution can be completely described by using its mean (μ) and variance (σ^2).

When a variable x is normally distributed, we note $x \sim N(\mu, \sigma^2)$.

Z Distribution

The Z distribution, also known as standard normal distribution, is the simplest normal distribution with the mean equal to zero and the variance equal to one. Any normal distribution can be transferred to a Z distribution by applying the following equation:

$$Z = \frac{x - \mu}{\sigma}$$

Where: $x \sim N(\mu, \sigma^2)$ and $\sigma \neq 0$.

The standard normal distribution is important because the probabilities and quantiles of any normal distribution can be computed from the standard normal distribution if the mean and standard deviation are known.

Shape of Normal Distribution

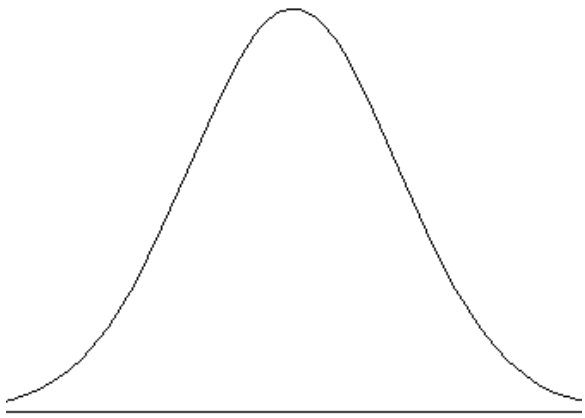


Fig. 2.25 Shape of a Normal Distribution

You might remember teachers from your academic days referring to "the curve" when talking about how they will grade performance. The probability density function curve of normal distribution is bell-shaped (see figure 2.25) and is expressed by the equation:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The equation just means that within the population that is described by this distribution, the highest probability is for values closest to the mean, then the probability drops as you move further from the mean (in either direction).

Location of Normal Distribution

If a variable is normally distributed, the mean, the median, and the mode have approximately the same value. The probability density curve of normal distribution is symmetrical around a center value which is the mean, median, and mode at the same time.

Spread of Normal Distribution

The spread or variation of the normally-distributed data can be described using the variance or the standard deviation (which is the square root of the variance). The smaller the variance or the standard deviation, the less variability in the data set.

The 68–95–99.7 Rule

The *68–95–99.7 rule* or the *empirical rule in statistics* states that for a normal distribution:

- About 68% of the data fall within one standard deviation of the mean
- About 95% of the data fall within two standard deviations of the mean
- About 99.7% of the data fall within three standard deviations of the mean

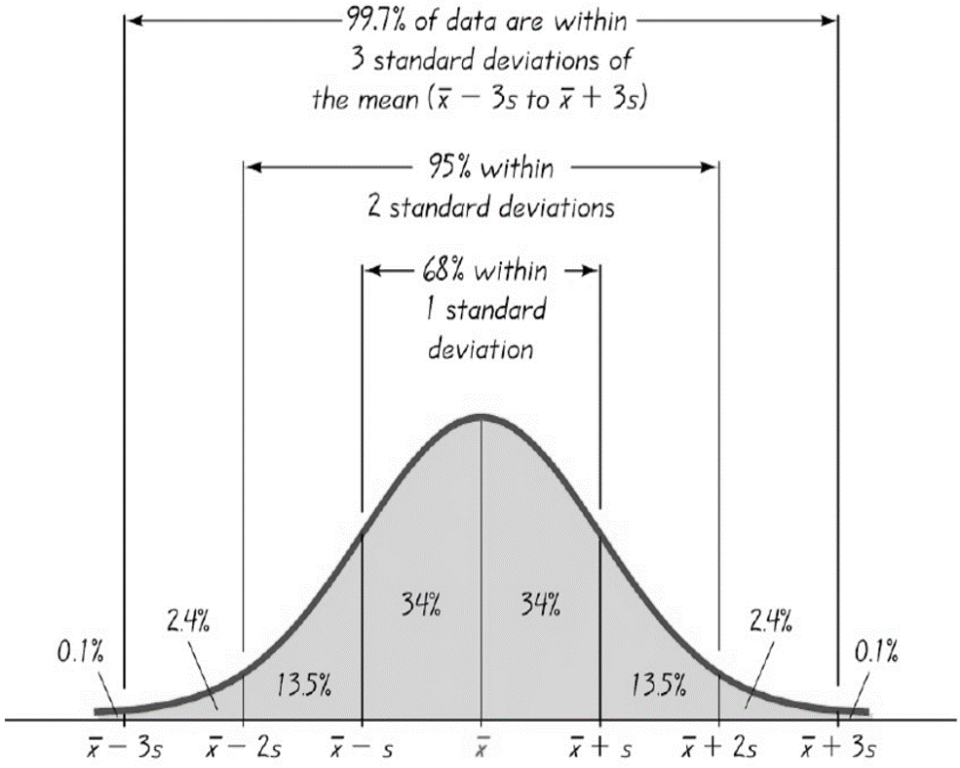


Fig. 2.26 Shape of a Normal Distribution

One way to think about this rule is to consider the area under the normal distribution curve. The 68–95–99.7 rule describes what percent of data fall within one, two, and three standard deviations from the mean, respectively. Most of the data values are grouped around the center.

Normality

Not all the distributions with a bell shape are normal distributions. It is important to understand whether or not data being analyzed fit the normal distribution because there are implications for how the data are analyzed and what inferences are made from the data.

To check whether a group of data points are normally distributed, we need to run a normality test. There are different normality tests available, such as Anderson–Darling test, Sharpiro–Wilk test, or Jarque–Bera test. More details of normality test will be introduced in the Analyze module.

Normality Testing

To check whether the population of our interest is normally distributed, we need to run normality test.

- Null Hypothesis (H_0): The data points *are* normally distributed.
- Alternative Hypothesis (H_a): The data points are *not* normally distributed.

Use JMP to Run a Normality Test

Normality Test Case study: Using a data set provided, we are interested to know whether the height of basketball players is normally distributed. This example demonstrates how to perform this analysis using JMP.

Data File: "OneSampleT-Test.jmp"

- Null Hypothesis (H_0): The height of basketball players is normally distributed.
- Alternative Hypothesis (H_a): The height of basketball players is not normally distributed.

Steps to run a Normality test in JMP:

1. Click Analyze -> Distribution
2. Select "HtBk" as "Y, Columns"
3. Click "OK"

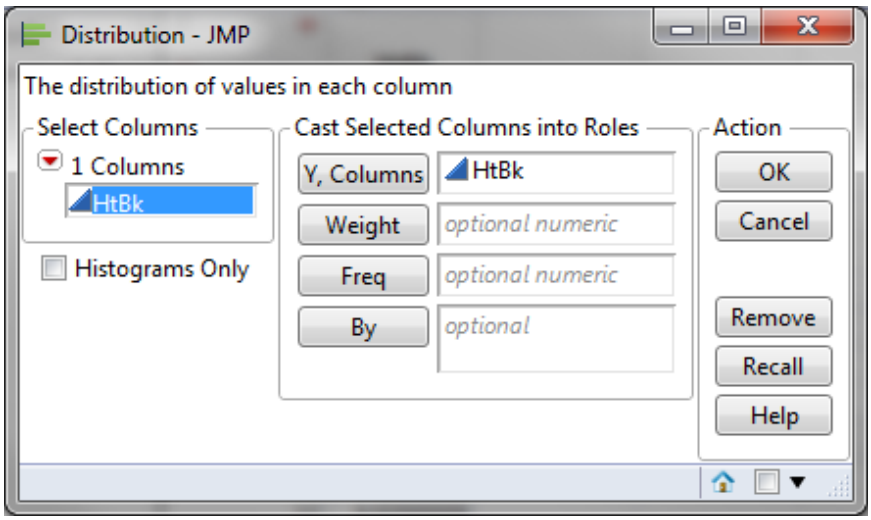


Fig. 2.27 Performing Normality Test in JMP

4. Click on the red triangle button next to "HtBk" in the Distribution page
5. Click Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"

7. Select “Goodness of Fit”

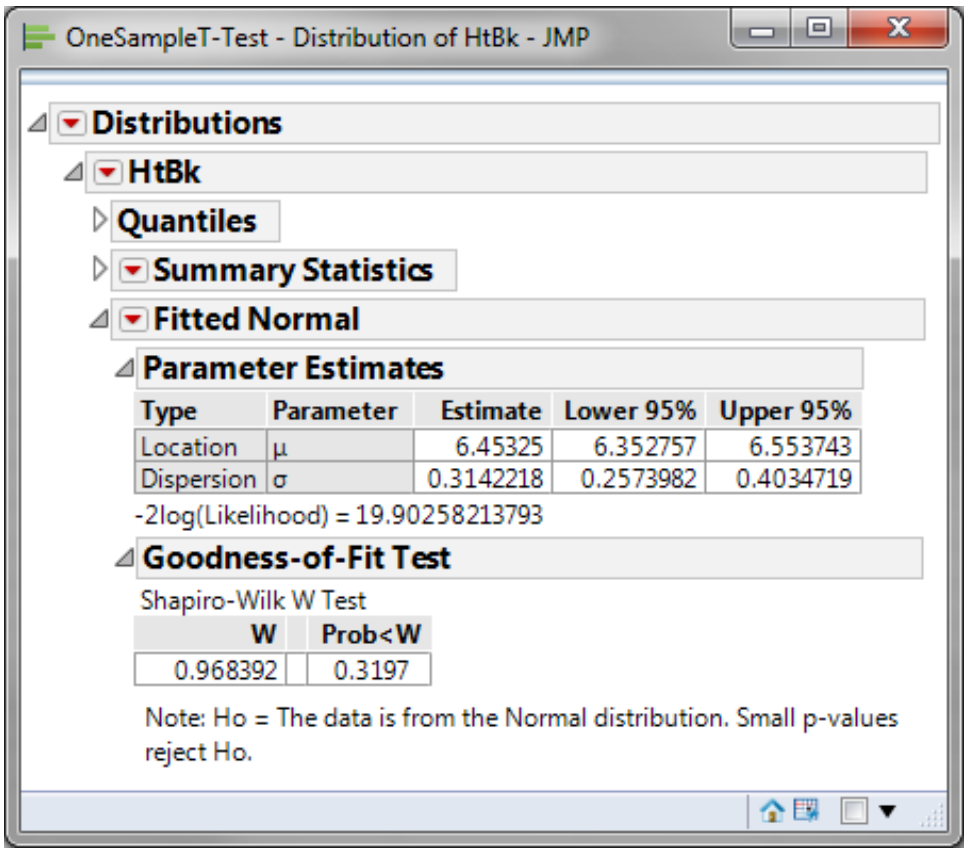


Fig. 2.28 JMP Normality Test Output

Since the p-value of the normality is 0.3197 greater than alpha level (0.05), we fail to reject the null and claim that the data are normally distributed. If the data are not normally distributed, you need to use other hypothesis tests other than one sample t-test.

2.2.4 GRAPHICAL ANALYSIS

What is Graphical Analysis?

In statistics, *graphical analysis* is a method to visualize the quantitative data. Graphical analysis is used to discover structure and patterns in the data. The presence of which may explain or suggest reasons for additional analysis or consideration. A complete statistical analysis includes both quantitative analysis and graphical analysis.

When we think of statistics, we often think of numbers. However, graphical analysis is complementary to purely quantitative statistics because it allows us to visualize structure and patterns in the data. Often, the graphical presentation of data can tell a story on its own, but the total statistical analysis is important to complete the picture. Sometimes our eyes can deceive us. There are various graphical analysis tools available. The most commonly used and hence the ones we shall cover here are Box Plot, Histogram, Scatter Plot and Run Chart.

Box Plot

A *box plot* is a graphical method to summarize a data set by visualizing the minimum value, 25th percentile, median, 75th percentile, the maximum value, and potential outliers. A percentile is the value below which a certain percentage of data fall. For example, if 75% of the observations have values lower than 685 in a data set, then 685 is the 75th percentile of the data. At the 50th percentile, or median, 50% of the values are lower and 50% are higher than that value.

Box Plot Anatomy

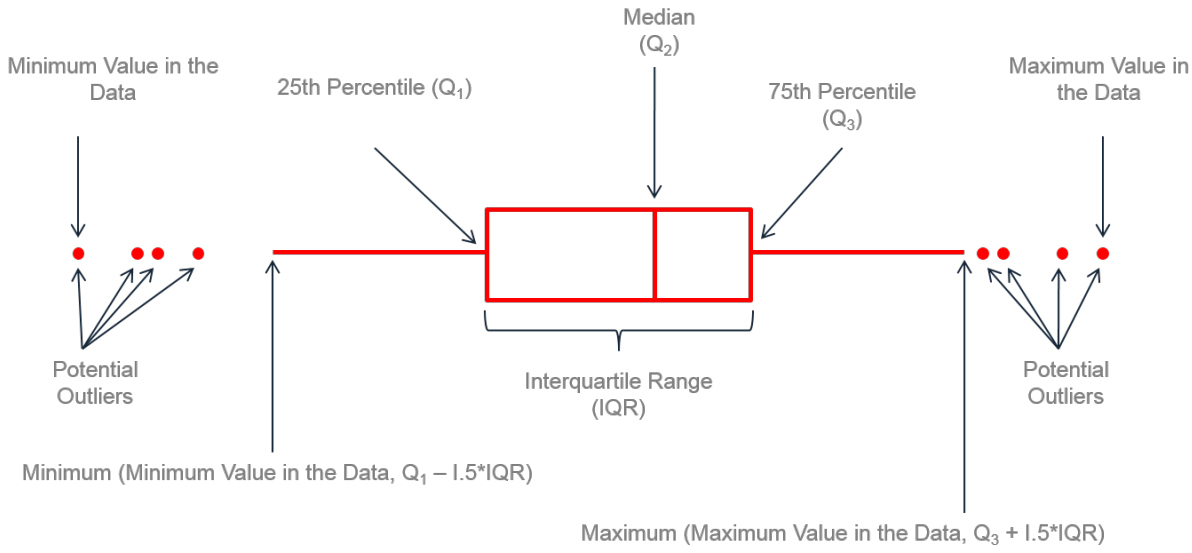


Fig. 2.29 Interpreting a Box Plot

Figure 2.30 describes how to read a box plot. Here are a few explanations that may help. The middle part of the plot, or the "interquartile range," represents the middle quartiles (or the 75th minus the 25th percentile). The line near the middle of the box represents the median (or middle value of the data set). The whiskers on either side of the IQR represent the lowest and highest quartiles of the data. The ends of the whiskers represent the maximum and minimum of the data, and the individual dots beyond the whiskers represent outliers in the data set.

How to Use JMP to Generate a Box Plot

Data File: "BoxPlot.jmp"

Steps to render a Box Plot in JMP:

1. Click Analyze -> Distribution
2. Select "HtBk" as "Y, Columns"

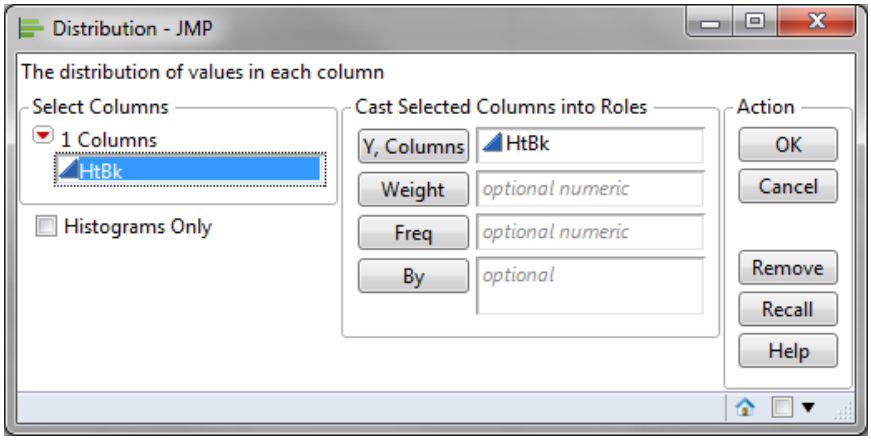


Fig. 2.30 Running a Box Plot in JMP

3. Click "OK"
4. Click on the red triangle button next to "HtBk"
5. Uncheck Histogram Options -> Vertical

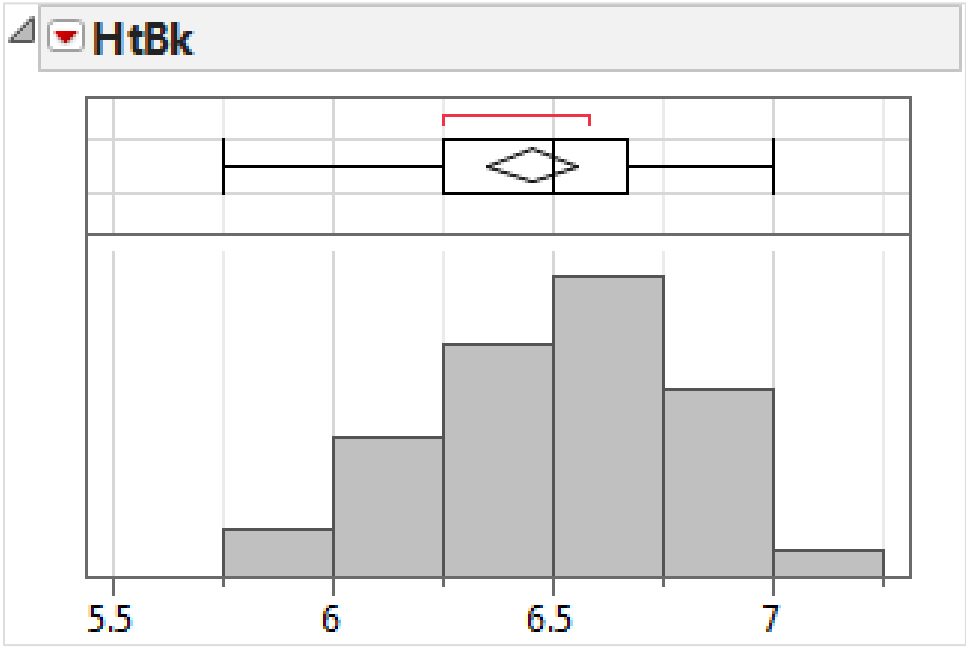


Fig. 2.31 JMP Box Plot Output

Histogram

A *histogram* is a graphical tool to present the distribution of the data. The X axis of a histogram represents the possible values of the variable and the Y axis represents the frequency of the value occurring. A histogram consists of adjacent rectangles erected over intervals with heights equal to the frequency density of the interval. The total area of all the rectangles in a histogram is the number of data values.

A histogram can also be normalized. In the case of normalization, the X axis still represents the possible values of the variable, but the Y axis represents the percentage of observations that

fall into each interval on the X axis. The total area of all the rectangles in a normalized histogram is 1. Using histograms, we have a better understanding of the shape, location, and spread of the data.

How to Use JMP to Generate a Histogram

Data File: "Histogram.jmp"

Steps to render a histogram in JMP:

1. Click Analyze -> Distribution
2. Select "HtBk" as "Y, Columns"

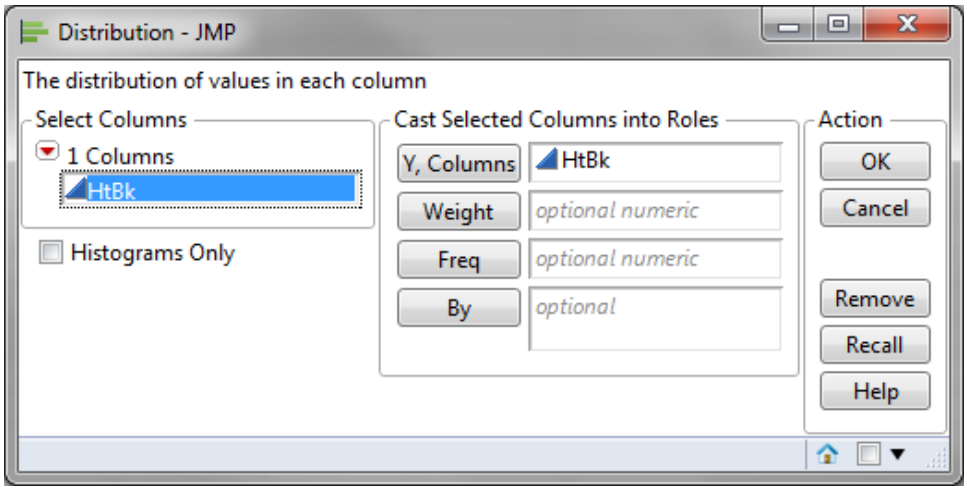


Fig. 2.32 Rendering a Histogram in JMP

3. Click "OK"
4. Click on the red triangle button next to "HtBk"
5. Uncheck Histogram Options -> Vertical

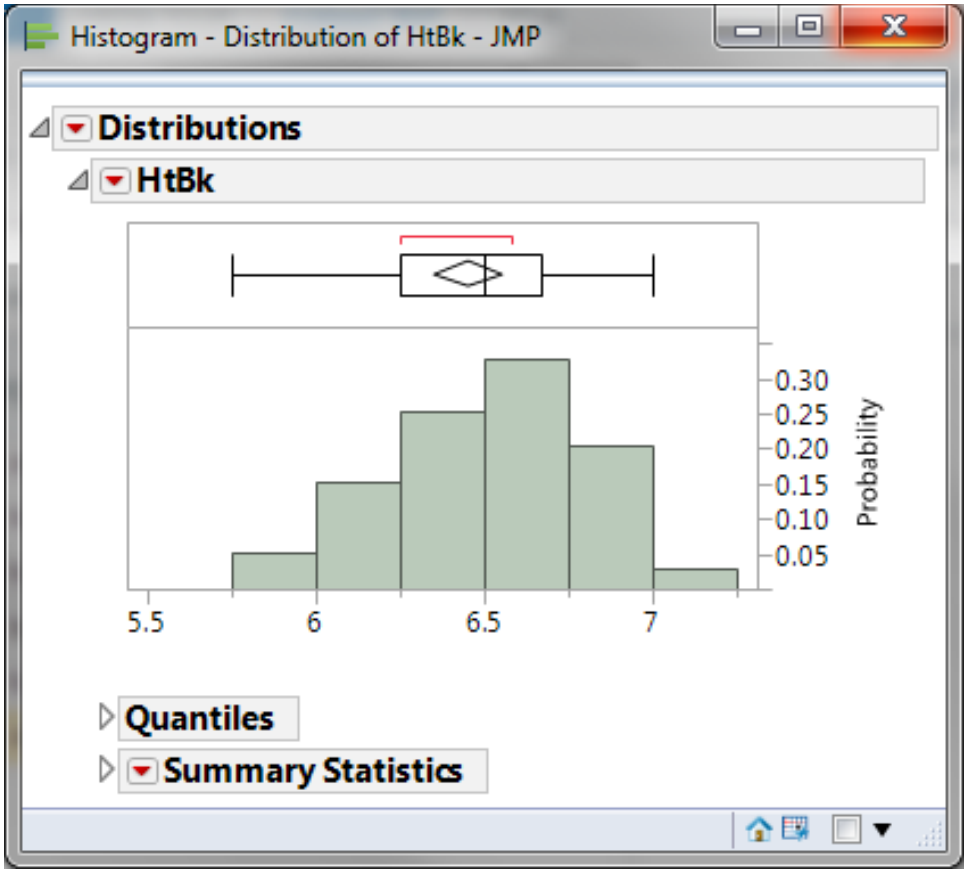


Fig. 2.33 JMP Histogram Output

To display Count as the Y axis:

1. Click on the red triangle button next to "HtBk"
2. Click Histogram Options -> Prob Axis

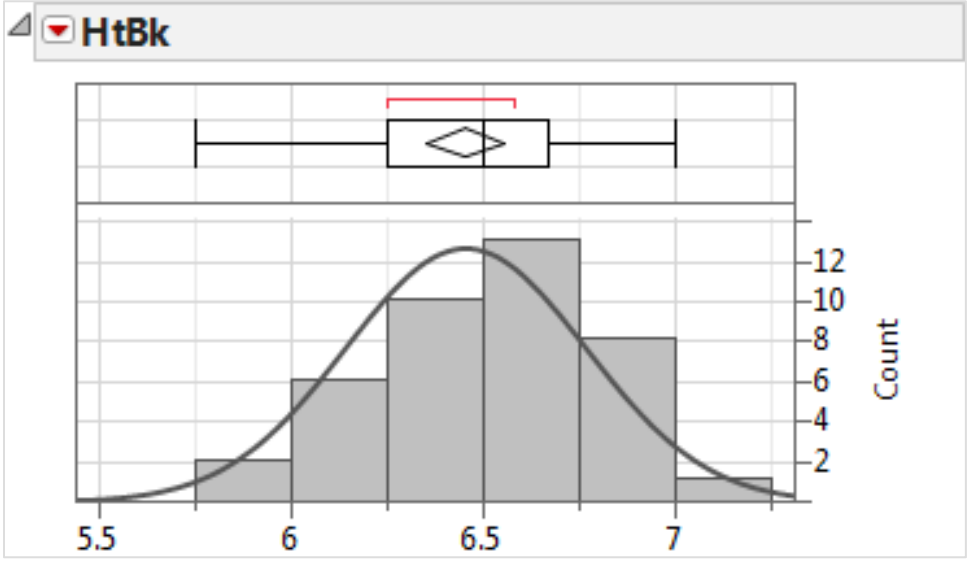


Fig. 2.34 Display Count as the Y axis

To display Probability as the Y axis:

1. Click on the red triangle button next to “HtBk”
2. Click Histogram Options -> Prob Axis

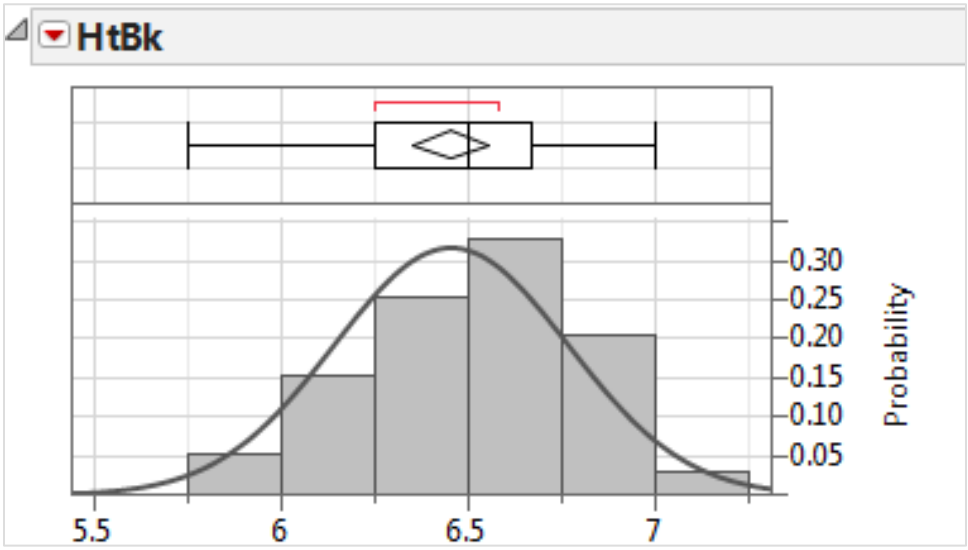


Fig. 2.35 Display Probability as the Y axis

The output from the previous steps has generated a graphical summary report of the data set HtBk. Among the information provided is a histogram. The image shows the frequency of the data for the numerical categories ranging from 5.5 to approximately 7. You can see the shape of the data roughly follows the bell curve.

Scatter Plot

A *scatter plot* is a diagram to present the relationship between two variables of a data set. A scatter plot consists of a set of data points. On the scatter plot, a single observation is presented by a data point with its horizontal position equal to the value of one variable and its vertical position equal to the value of the other variable. A scatter plot helps us to understand:

- Whether the two variables are related to each other or not
- What the strength of their relationship
- The shape of their relationship
- The direction of their relationship
- Whether outliers are present

How to Use JMP to Generate a Scatter Plot

 ***Data File: “ScatterPlot.jmp”***

Steps to render a Scatterplot in JMP:

1. Click Analyze -> Fit Y by X
2. Select MPG as the “Y, Response” and weight as the “X, Factor”

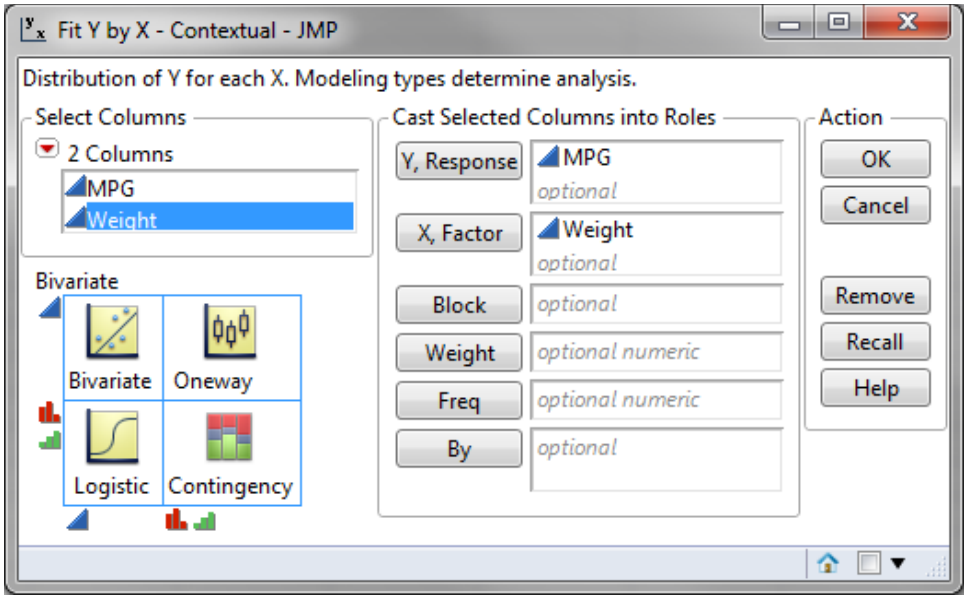


Fig. 2.36 Running a Scatter Plot in JMP

3. Click OK

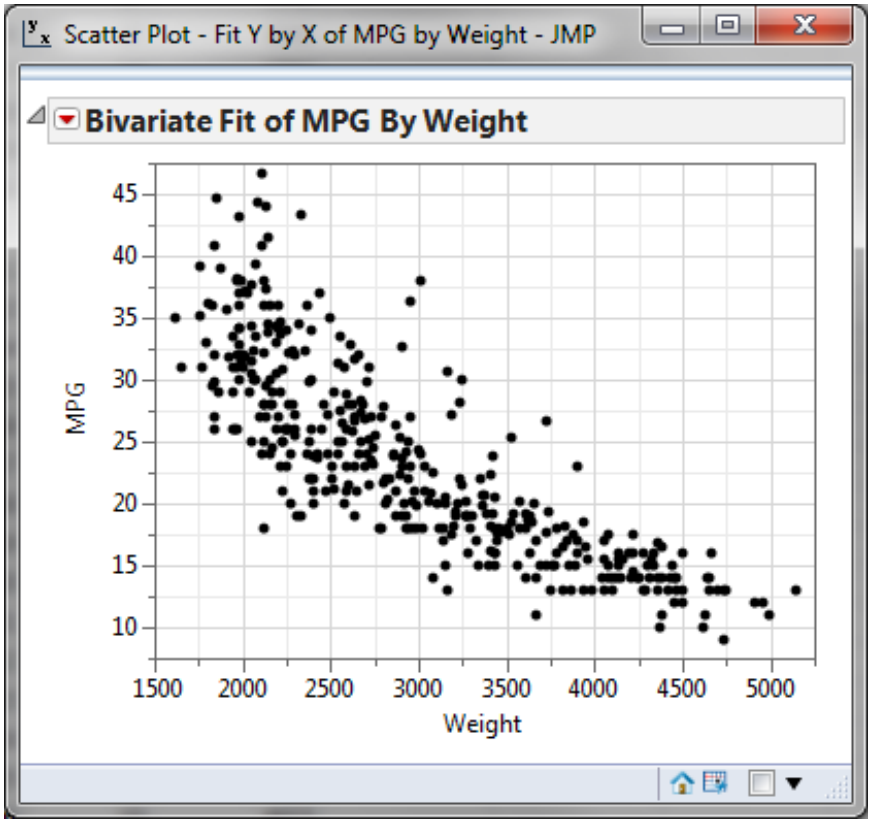


Fig. 2.37 JMP Scatter Plot Output

Figure 2.37 is JMP's output of the scatterplot data. You can immediately see the value of graphical displays of data. This information obtainable by viewing this output shows a relationship between weight and MPG. This scatterplot shows that the heavier the weight the lower the MPG value and vice versa.

Run Chart

A *run chart* is a chart used to present data in time order. Run charts capture process performance over time. The X axis of a run chart indicates time and the Y axis shows the observed values. A run chart is similar to a scatter plot in that it shows the relationship between X and Y. Run charts differ however because they show how the Y variable changes with an X variable of time.

Run charts look similar to control charts except that run charts do not have control limits and they are much easier to produce thank a control chart. A run chart is often used to identify anomalies in the data and discover pattern over time. They help to identify trends, cycles, seasonality and other anomalies.

How to Plot a Run Chart in JMP

Data File: "RunChart.jmp"

Steps to plot a run chart in JMP:

1. Click > Analyze -> Quality & Process -> Control Chart -> Run Chart
2. Select "Measurement" as the "Process"

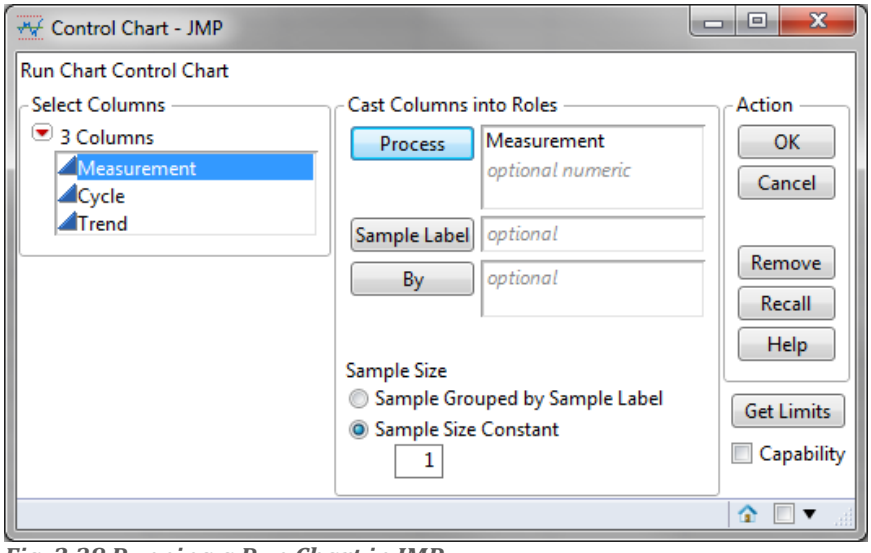


Fig. 2.38 Running a Run Chart in JMP

3. Click "OK"

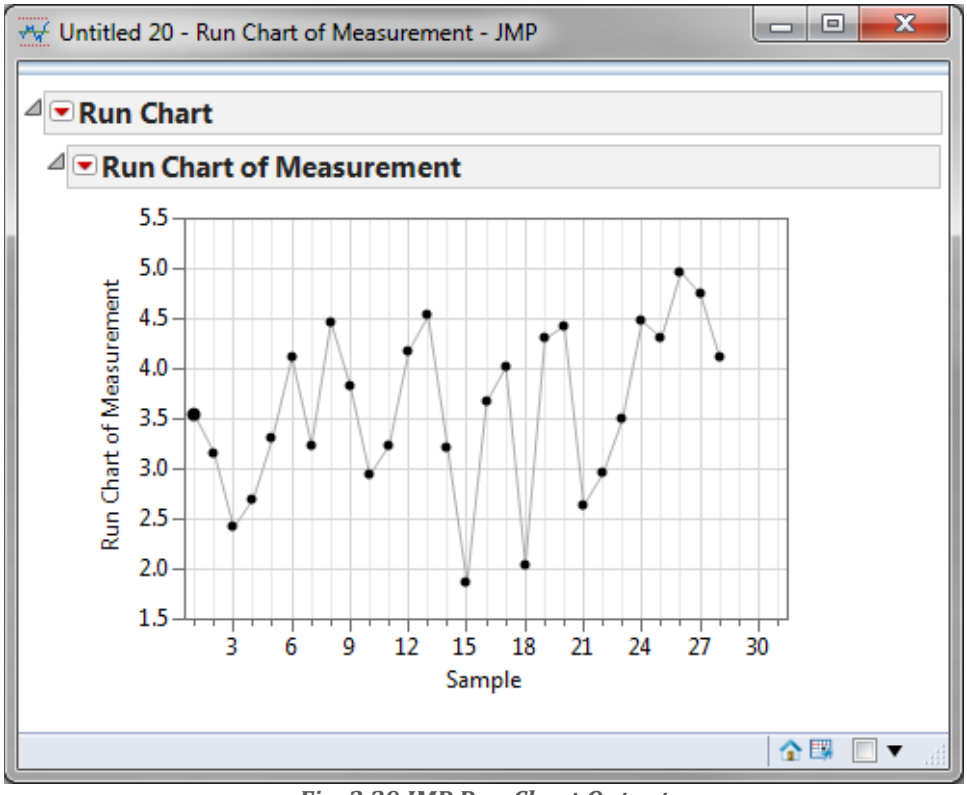


Fig. 2.39 JMP Run Chart Output

Figure 2.39 is a run chart created with JMP. The time series displayed by this chart appears stable. There are no extreme outliers, no visible trending or seasonal patterns. The data points seem to vary randomly over time.

Now, let us take a look at another example which may give us a different perspective. We will create another run chart using the data listed in the column labeled “Cycle”. This column is in the same file used to generate Figure 2.39. Follow the steps used for the first run chart and instead of using “Measurement” use “Cycle” in the Run Chart dialog box.

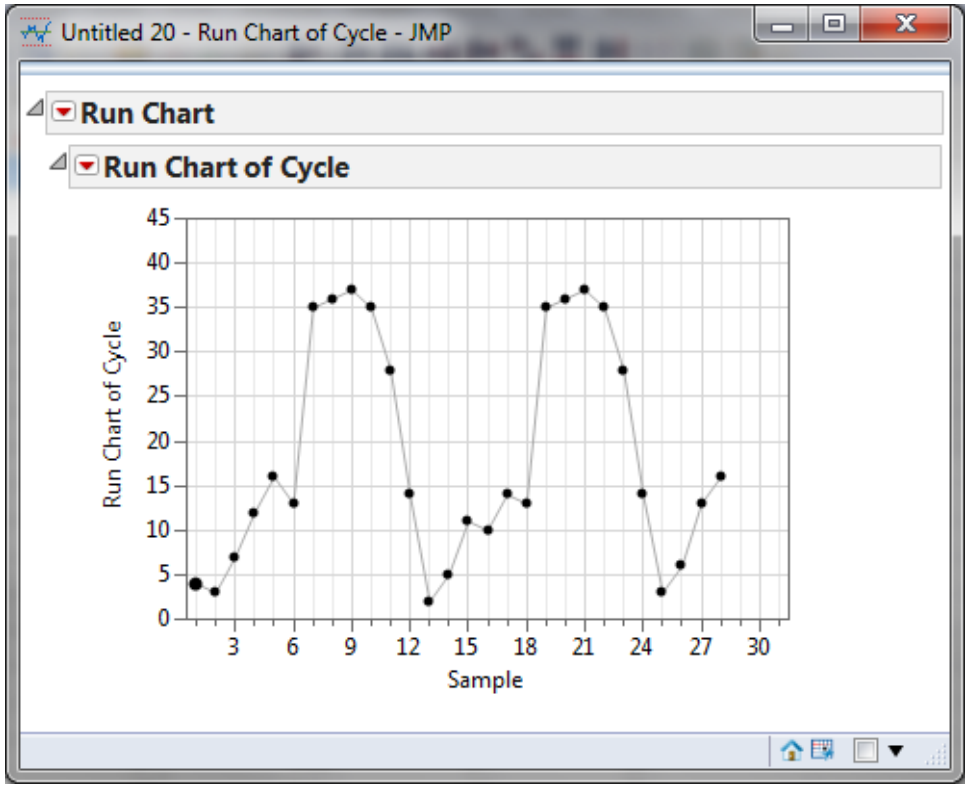


Fig. 2.40 JMP Run Chart for the "Cycle" data

In this example (figure 2.40), the data points are clearly exhibiting a pattern. It could be seasonal or it could be something cyclical. Imagine that the data points are taken monthly and this is a process performing over a period of 2.5 years. Perhaps the data points represent the number of customers buying new homes. The home buying market tends to peak in the summer months and dies down in the winter. Using the same data tab lets create a final run chart. This time use the "Trend" data. Again, follow the steps outlined previously to generate a run chart.

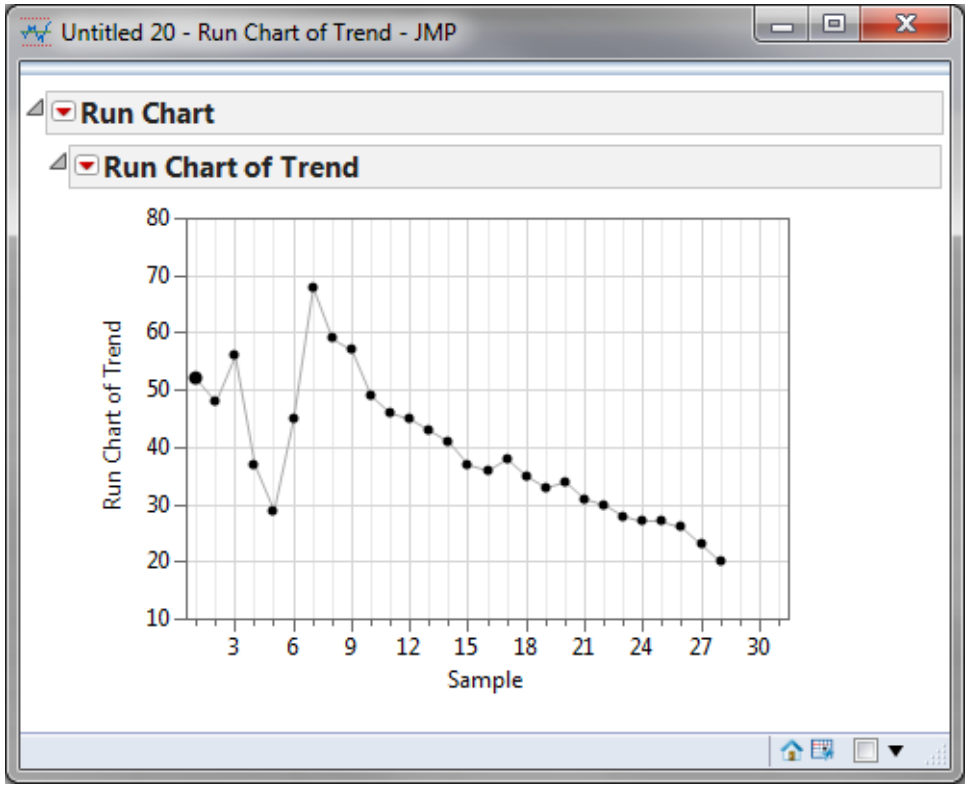


Fig. 2.41 JMP Run Chart for Trending Data

In this example (Fig. 2.41) the process starts out randomly, but after the seventh data point almost every data point has a lower value than the one before it. This clearly illustrates a downward trend. What might this represent? Perhaps a process winding down? Product sales at the end of a product's life cycle? Defects decreasing after introducing a process improvement?

It should be clear through our review of Histograms, Scatterplots and Run Charts, that there is great value in "visualizing" the data. Graphical displays of data can be very telling and offer excellent information.

2.3 Measurement System Analysis

2.3.1 Precision and Accuracy

What is Measurement System Analysis?

Measurement System Analysis (*MSA*) is a systematic method to identify and analyze the variation components of a measurement system. It is a mandatory step in any Six Sigma project to ensure the data are reliable before making any data-based decisions. A MSA is the check point of data quality before we start any further analysis and draw any conclusions from the data. Some good examples of data-based analysis where MSA should be a prerequisite:

- Correlation analysis
- Regression analysis
- Hypothesis testing
- Analysis of variance
- Design of experiments
- Statistical process control

You will see where and how the analysis techniques listed above are used throughout this training program. It is critical at this stage however to know that any variation, anomalies, or trends found in your analysis are due to the data and not due to the inaccuracies or inadequacies of a measurement system. Therefore, the need for a MSA is vital.

Measurement System

A measurement system is a process used to obtain data and quantify a part, product or process. Data obtained with a measurement device or measurement system are the *observed* values. Observed values are comprised of two elements

1. True Value = Actual value of the measured part
2. Measurement Error = Error introduced by the measurement system.

The *true value* is what we are ultimately trying to determine through the measurement system. It reflects the true measurement of the part or performance of the process.

Measurement error is the variation introduced by the measurement system. It is the bias or inaccuracy of the measurement device or measurement process.

The *observed value* is what the measurement system is telling us. It is the measured value obtained by the measurement system. Observed values are represented in various types of measures which can categorized into two primary types discrete and continuous. Continuous measurements are represented by measures of weight, height, money and other types of measures such as ratio measures. Discrete measures on the other hand are categorical such as Red/Yellow/Green, Yes/No or Ratings of 1–10 for example.

A *variable MSA* is designed assess continuous measurement systems while *attribute MSAs* have been designed to assess discrete measurement systems.

Sources of measurement errors
There are numerous ways to introduce error into a measurement system. The following are just a few common ones:

- Human Error - imagine reading a ruler or a scale, or think about systems where a person must make a qualitative judgment
- Environment - Sometimes the temperature of the air or humidity can affect how a measurement system reads or the equipment itself may need to be calibrated from time to time
- Equipment

- Sample - Sampling can introduce variation based on how or where or when we choose to sample a process
- Process
- Materials
- Methods

Fishbone diagrams can help to brainstorm potential factors affecting the accuracy or validity of measurement systems. The more errors a measurement system allows, the less reliable the observed values are. A valid measurement system brings in minimum amount of measurement error. Ultimately, the goal of a measurement system analysis is to understand how much error exists in the system by analyzing the characteristics of the system. Then, decisions can be made as to whether the system needs to be changed, replaced, modified or accepted.

Characteristics of a Measurement System

Any measurement systems can be characterized by two aspects:

- Accuracy (location related)
- Precision (variation related)

A valid measurement system is *both* accurate and precise. Being accurate does not guarantee the measurement system is precise. Being precise does not guarantee the measurement system is accurate.

Accuracy vs. Precision

Accuracy is the level of closeness between the average observed value and the true value. Accuracy tells us how well the observed value reflects the true value.

Precision is the spread of measurement values. It is how consistent the repeated measurements deliver the same values under the same circumstances.

To simplify, *accuracy* describes how close the average observed value is to the true value for what is being measured. *Precision* describes how close or repeatable the measurement is. Let us take a look at a few representations of accuracy and precision.

Accurate and precise (high accuracy and high precision)
Figure 2.42 represents a target with shots landing in the center and all shots close to one another. This is a depiction of both accuracy and precision. For the purposes of an MSA, think of the center of the target as the “true value” and the black dots as “observed values.” All six measurements are located at the center of the target and are grouped tightly together. This means that the measurements are both accurate and precise.

Fig. 2.42 Accurate and Precise

Accurate and not precise (high accuracy and low precision) Figure 2.43 is an example with all the observations located around the center of the target, but they are spread further apart from one another, more so than in figure 2.42. This measurement system would be described as accurate, but not precise.

Fig. 2.43 Accurate but Not Precise

Precise and not accurate (high precision and low accuracy) Figure 2.44 shows all observations grouped closely together, but they are not located near the center of the target. This system is precise, but not accurate.

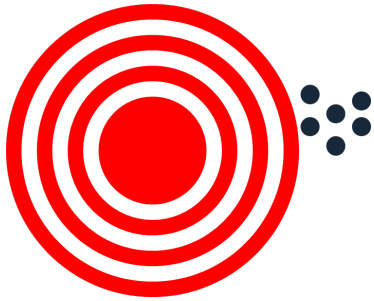


Fig. 2.44 Precise but Not Accurate

Not accurate and not precise (low accuracy and low precision) Lastly, figure 2.45 shows a system that is neither accurate nor precise. The observations are not located around the center and they are spread far apart from each other.

Fig. 2.45 Not Accurate and Not Precise

MSA Conclusions

If the measurement system is considered *both* accurate and precise, we can start the data-based analysis or decision making. If the measurement system is either not accurate or not precise, we need to identify the factors affecting it and calibrate the measurement system until it is both accurate and precise.

Stratifications of Accuracy and Precision

- Accuracy
  - Bias
  - Linearity
  - Stability
- Precision
  - Repeatability
  - Reproducibility

2.3.2 Bias, Linearity, and Stability

Bias

Bias is the difference between the observed value and the true value of a parameter or metric being measured. Think of it as a measurement system's absolute correctness versus a standard (or true value). Bias is calculated by subtracting the reference value from the average value of the measurements.

$$Bias = Grand\ Mean - Reference\ Value$$

where the reference value is a standard agreed upon value.

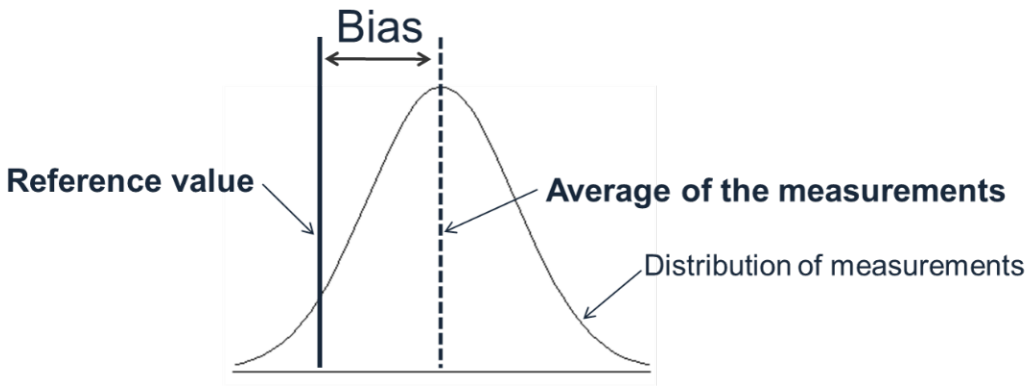


Fig. 2.46 Measurement Bias

In figure 2.49 you can see that the group of observations is consistently to the right of the reference value. The dotted line represents the grand mean described on the last page. The goal is to reduce the distance between the grand mean and the reference value, thereby reducing bias. The closer the average of all measurements is to the reference level, the smaller the bias. The reference level is the average of measurements of the same items using the master or standard instrument.

To determine whether the difference between the average of observed measurement and the reference value is statistically significant (we will explain more details about statistical significance in the Analyze module), we can either conduct a *hypothesis test* or *compare* the reference value against the confidence intervals of the average measurements. If the reference value falls into the confidence intervals, the bias is not statistically significant and can be ignored. Otherwise, the bias is statistically significant and must be fixed.

Potential causes of bias are:

- Errors in measuring the reference value
- Lack of proper training for appraisers
- Damaged equipment or instrument
- Measurement instrument not calibrated precisely
- Appraisers read the data incorrectly

In the first case, for example, there could be something wrong with the standard (or reference value). For example, think about weight standards that are used to calibrate scales. If the weight is damaged in some way such that it no longer represents the standard, then obviously, we will not be able to measure the correct value.

Linearity

Linearity is the degree of the consistency of bias over the entire expected measurement range. It quantifies how the bias changes over the range of measurement. For example, a scale is off by 0.01 pounds when measuring an object of 10 pounds. However, it is off by 10 pounds when measuring an object of 100 pounds. The scale's bias is not linear.

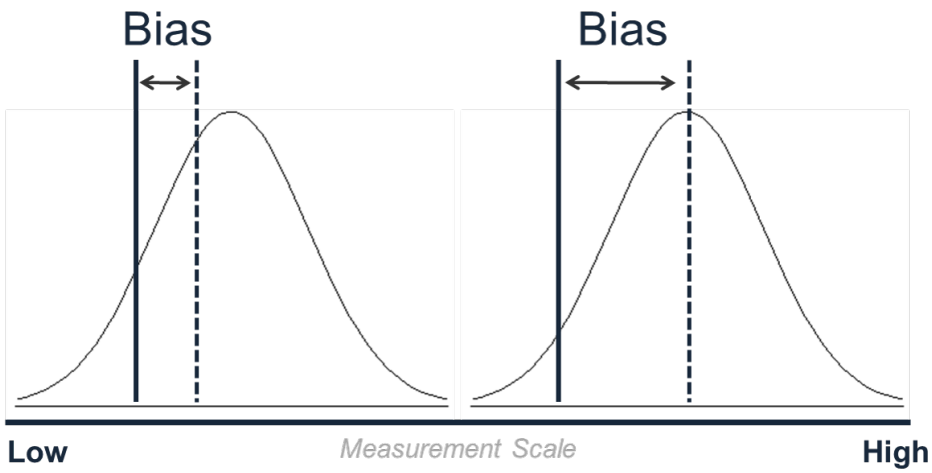


Fig. 2.47 Measurement Linearity

Figure 2.50 represents the bias of a measurement system across its measurement scale from low measures to high measures. The image suggests that as the measurement scale increases so does the bias. The linearity for this measurement system would be positive and undesirable.

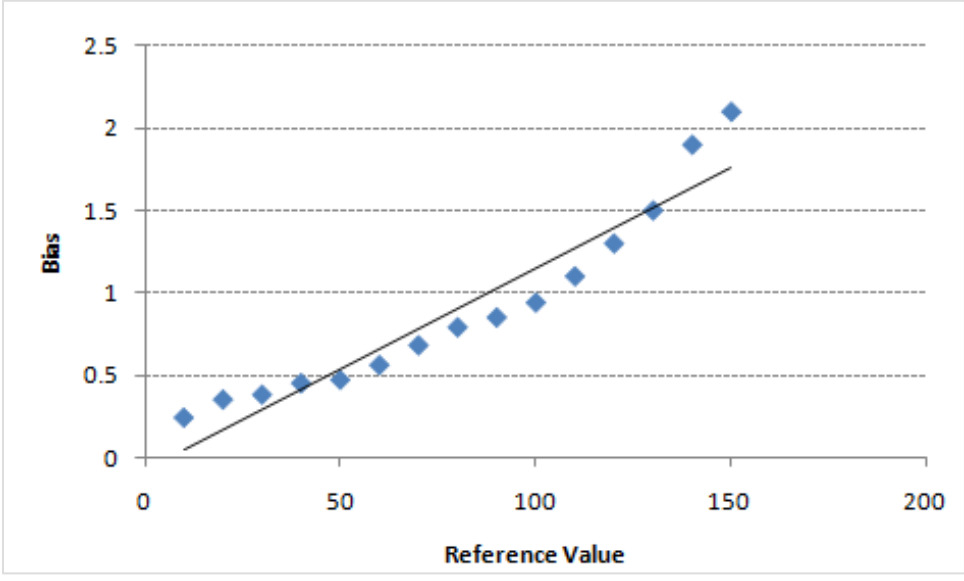


Fig. 2.48 Bias Scatterplot for Linearity

Figure 2.39 uses a scatterplot to demonstrate the linearity of a measurement system. The X axis represents references values for the measurement scale and the Y axis represents bias. The best fit line shows the slope of the bias. The ideal linearity of a measurement system would have a slope of zero which would imply no change in bias over the scale of the measurement system. The slope in figure 2.39 is positive suggesting that as the measurement scale increases so does the bias.

The formula of the linearity of a measurement system is:

$$Linearity = |Slope| \times Process\ Variation$$

Where:

$$Slope = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n}(x_i^2) - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}$$

and where:

- x_i is the reference value
- y_i is the bias at each reference level
- n is the sample size.

As demonstrated in the previous figures, we know that slope is an important computation in determining the linearity of a measurement system. This is based on the fact that the equation of a line defines the relationship between the bias and the reference values of the parts or samples.

Potential causes of linearity
The causes of linearity include:

- Errors in measuring the lower end or higher end of the reference value
- Lack of proper training for appraisers
- Damaged equipment or instrument
- Measurement instrument not calibrated correctly at the lower or higher end of the measurement scale
- Innate nature of the instrument

As you notice, some of the things we discussed as causes for bias are also potential causes for linearity issues: training, damaged or improperly-calibrated equipment. How these issues manifest themselves are the difference, because bias can become an issue at the low and high ends of a measurement range.

Stability

Stability is the consistency level needed to obtain the same values when measuring the same objects over an extended period. A measurement system that has low bias and linearity close to zero but cannot consistently perform well over time would not deliver reliable data.

Much like we discussed linearity being the degree of bias over the range of a measurement scale, stability is the degree of bias over time. Figure 2.41 demonstrates increasing bias as time passes.

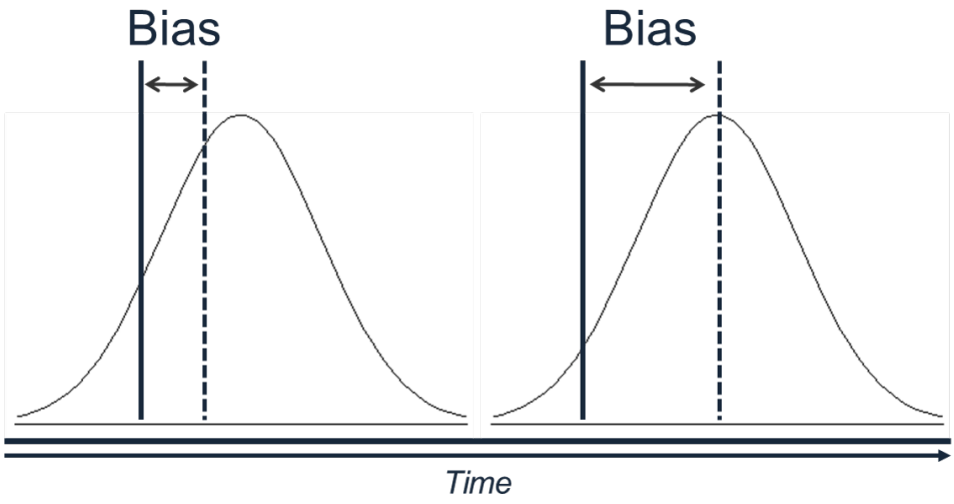


Fig. 2.49 Measurement Stability: Bias over Time

Control charts are the most common tools used to evaluate the stability of a measurement system. By plotting bias calculations of a measurement system over time and using a control chart to visualize scale and variation, it allows us to determine if bias is stable or out of control. As you will learn in the Control phase, if no measures on a control chart are "out of control" then we can consider bias to be stable.

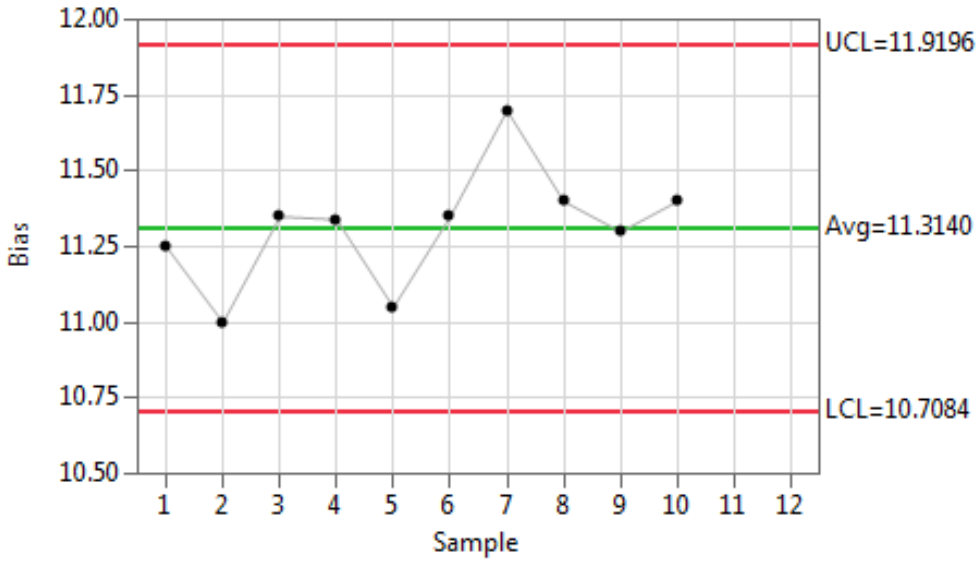


Fig. 2.50 Bias Control Chart: Measurement Stability

Potential causes of instability
The causes of instability include:

- Inconsistent training for appraisers
- Damaged equipment or instrument
- Worn equipment or instrument
- Measurement instrument not calibrated

- Appraisers do not follow the procedure consistently

Consider two appraisers who were trained at two different times, and they would measure differently, or consider two appraisers who do not follow the procedures in the same way. Measurement equipment can wear out over time, lose calibration, or even break, obviously affecting its ability to measure accurately over time.

2.3.3 GAGE REPEATABILITY AND REPRODUCIBILITY

Repeatability

Repeatability evaluates whether the same appraiser can obtain the same value multiple times when measuring the same object using the same equipment under the same environment. It refers to the level of agreement between the repeated measurements of the same appraiser under the same condition. Repeatability measures the inherent variation of the measurement instrument. To simplify repeatability, think of it as your own personal ability to measure something 3 times and get the same answer all three times. If you can't then you're unable to successfully "repeat" a measurement.

Reproducibility

Reproducibility evaluates whether different appraisers can obtain the same value when measuring the same object independently. It refers to the level of agreement between different appraisers. While repeatability is inherent to the measurement system, reproducibility is not caused by the inherent variation of the measurement instrument. It reflects the variability caused by different appraisers, locations, gauges, environments etc. The simple explanation of reproducibility is best described by asking, can you and someone else measure something and get the same answer. If not, the reproducibility of the measurement system is inadequate.

Gage Repeatability & Reproducibility

Gage R&R (*Gage Repeatability & Reproducibility*) is a method used to analyze the variability of a measurement system by partitioning the variation of the measurements using ANOVA (Analysis of Variance). Gage R&R primarily addresses the precision aspect of a measurement system. It is a tool used to understand if a measurement system can repeat and reproduce and if not, help us determine what aspect of the measurement system is broken so that we can fix it.

Gage R&R requires a deliberate study with parts, appraisers and measurements. Measurement data must be collected and analyzed to determine if the measurement system is acceptable. Typically Gage R&Rs are conducted by 3 appraisers measuring 10 samples 3 times each. Then, the results can be compared to determine where the variability is concentrated. The optimal result is for the measurement variability to be due to the parts.

Data collection of a gage R&R study:
Let *k* appraisers' measure *n* random samples independently and repeat the process *p* times. Different appraisers perform the measurement independently. The order of measurement (e.g., sequence of samples and sequence of appraisers) is randomized, as evidenced by the representative letters *k*, *n*, and *p*.

Potential sources of variance in the measurement:
Appraisers: $\sigma^2_{appraisers}$

Parts: σ^2_{parts}

Appraisers × Parts: $\sigma^2_{appraisers \times parts}$

Repeatability: $\sigma^2_{repeatability}$

Variance Components: $\sigma^2_{total} = \sigma^2_{appraisers} + \sigma^2_{parts} + \sigma^2_{appraisers \times parts} + \sigma^2_{repeatability}$

As can be seen in this equation for total measurement variance, there are several components: the appraisers, the parts being measured, the interaction between the appraisers and parts, and the repeatability. A valid measurement system has low variability in both repeatability and reproducibility so that the total variability observed can reflect the true variability in the parts being measured.

$$\sigma^2_{total} = \sigma^2_{reproducibility} + \sigma^2_{repeatability} + \sigma^2_{parts}$$

Where:

$$\sigma^2_{reproducibility} = \sigma^2_{appraisers} + \sigma^2_{appraisers \times parts}$$

Gage R&R variance reflects the precision level of the measurement system.

$$\sigma^2_{R\&R} = \sigma^2_{repeatability} + \sigma^2_{reproducibility}$$

As you can see, we can also show the total variance of the measurement system is a function of the reproducibility, the repeatability, and the variance in parts. It ties back to the equation on the previous page when you substitute variance in the appraisers and the variance in interaction between appraisers and parts with reproducibility. The goal is to minimize reproducibility variance and repeatability variance so you are only left with variance in parts—the variance in the parts is the "true" variance.

The *precision* of the measurement system is represented by the sum of the repeatability variance and reproducibility variance.

Variation Components
The variation of each component can be represented by the standard deviation (or sigma) for each component multiplied by Z_0, which is a sigma multiplier that assumes a specific

confidence level for the spread of the data. This is important because we will use the variation to calculate each component's contribution to the total measurement variation of the system.

$$Variation_{total} = Z_0 \times \sigma_{total}$$

$$Variation_{repeatability} = Z_0 \times \sigma_{repeatability}$$

$$Variation_{reproducibility} = Z_0 \times \sigma_{reproducibility}$$

$$Variation_{parts} = Z_0 \times \sigma_{parts}$$

Where:

$$\sigma^2_{total} = \sigma^2_{reproducibility} + \sigma^2_{repeatability} + \sigma^2_{parts}$$

The percentage of variation R&R contributes to the total variation in the measurement:

$$Contribution\ \%_{R\&R} = \frac{Variation_{R\&R}}{Variation_{total}} \times 100\%$$

Where:

$$Variation_{R\&R} = Z_0 \times \sqrt{\sigma^2_{repeatability} + \sigma^2_{reproducibility}}$$

The goal for the Contribution $\%_{R\&R}$ calculation is to determine how much variation is contributed by repeatability and reproducibility. As you can imagine, the lower the contribution from these components the better, but there are some guidelines for how to assess the quality of the measurement system. Table 2.1 represents the standard interpretations of varying degrees of contribution percentages.

Contribution $\%_{R\&R}$	**Interpretation**
Less than 1%	Satisfactory – Measurement System is Acceptable
Between 1% and 9%	Grey Area – Requires Decisions and Trade-offs
Greater than 10%	Unsatisfactory – Measurement System is Unacceptable

Table 2.1 Gage R&R Contribution% Interpretation Guidelines

2.3.4 Variable and Attribute MSA

Variable Gage R&R

Whenever something is measured repeatedly or by different people or processes, the results of the measurements will vary. Variation comes from two primary sources:

- Differences between the parts being measured
- The measurement system

We can use a gage R&R to conduct a measurement system analysis to determine what portion of the variability comes from the parts and what portion comes from the measurement system. There are key study results that help us determine the components of variation within our measurement system.

Key Measures of a Variable Gage R&R

- %Contribution - The percent of contribution for a source is 100 times the variance component for that source divided by the total variance.
- %Study Var (6*SD) - The percent of study variation for a source is 100 times the study variation for that source divided by the total variation
- %Tolerance (SV/Tolerance) - The percent of spec range taken up by the total width of the distribution of the data based on variation from that source
- Distinct Categories - The number of distinct categories of parts that the measurement system is able to distinguish. If a measurement system is not capable of distinguishing at least five types of parts, it is probably not adequate.

Variable Gage R&R Guidelines (AIAG)

The guidelines for acceptable or unacceptable measurement systems can vary depending on an organizations tolerance or appetite for risk. The common guidelines used for interpretation are published by the Automotive Industry Action Group (AIAG). These guidelines are considered standard for interpreting the results of a measurement system analysis using Variable Gage R&R. Table 2.2 summarizes the AIAG standards.

Measurement System	% Study Var	% Contribution	Distinct Categories
Acceptable	10% or less	1% or Less	5 or Greater
Marginal	10% - 30%	1% - 9%	
Unacceptable	30% or Greater	9% or Greater	Less than 5

Table 2.2 Variable Gage R&R Guidelines

Use JMP to Implement a Variable MSA

Data File: "VariableMSA.jmp"

Let's take a look at an example of a Variable MSA using the data in the Variable MSA tab in your "Sample Data.xlsx" file. In this exercise, we will first walk through how to set up your study using JMP and then we will perform a Variable MSA using 3 operators who all measured 10 parts three times each. The part numbers and operators and measurement trials are all generic so that you can apply the concept to your given industry. First, we need to set up the study.

Steps in JMP to run a Variable MSA:

Step 1: Initiate the MSA study

1. Click: Analyze > Quality & Process > Measurement Systems Analysis
2. Select “Measurement” as “Y, Response”
3. Select “Operator” as “X, Grouping”
4. Select “Part” as “Sample, Part ID”
5. Select “Gauge R&R” as the “MSA Method”
6. Select “Crossed” as “Model Type”

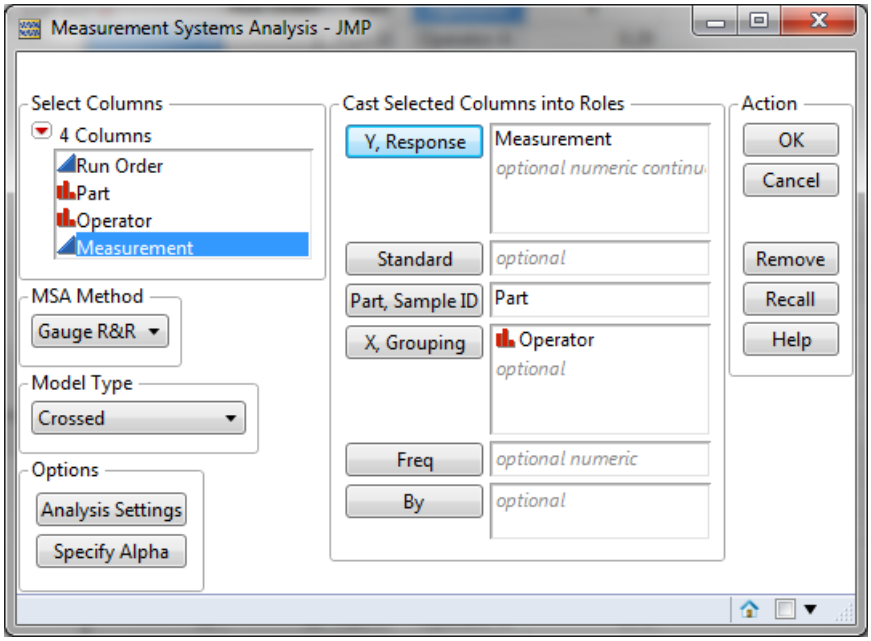


Fig. 2.51 Initiating a Variable MSA in JMP

7. Click “OK”

Step 2: Create the variability chart for measurement

1. Click on the red triangle button next to “Variability Gauge”
2. Click “Connect Cell Means” to link the average measurement for each part together
3. Click “Show Group Means” to display the average for each appraiser (solid line)
4. Click “Show Grand Mean” to display the average for the entire data set (dotted line)

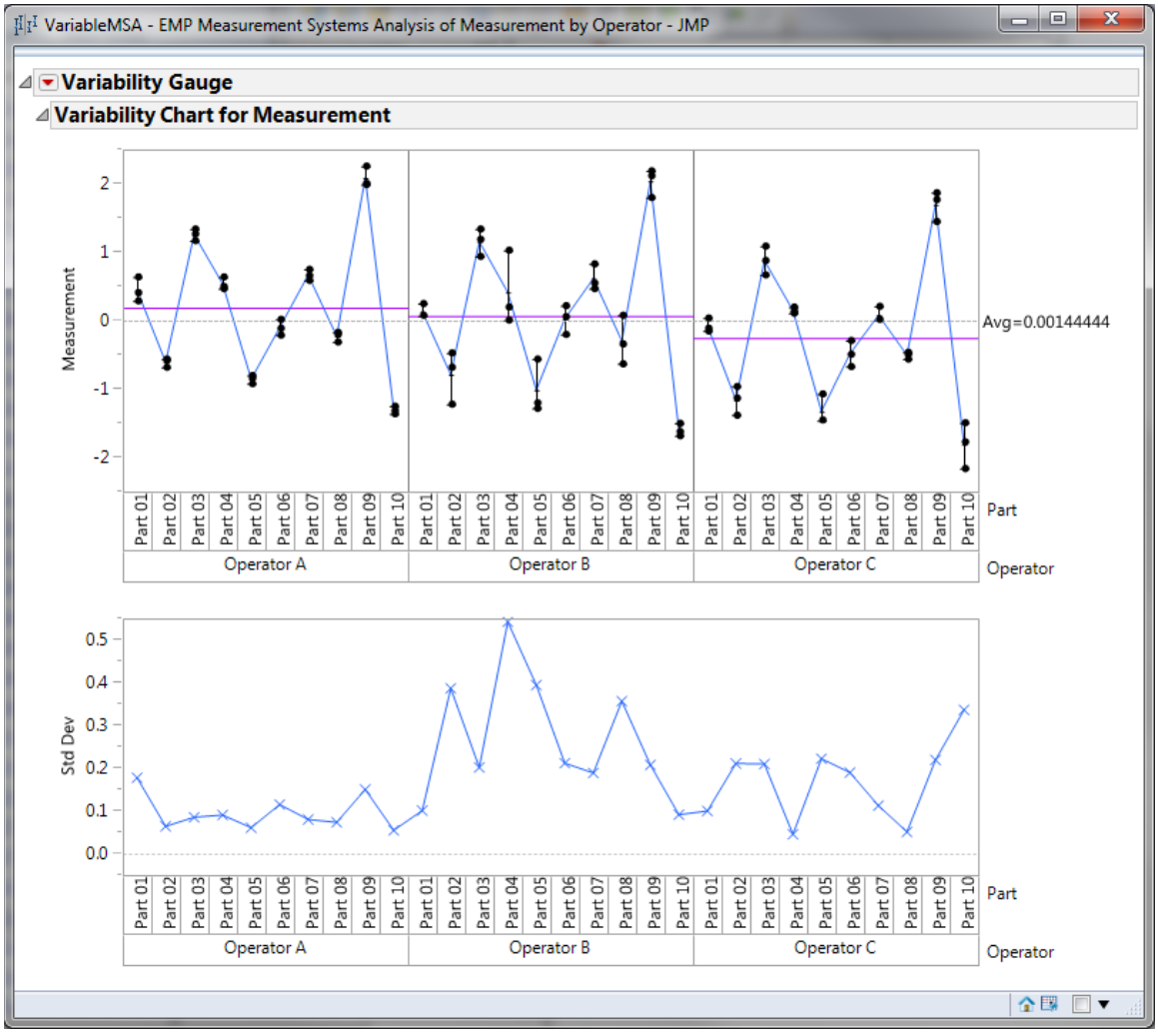


Fig. 2.52 Initiating a Variable MSA in JMP

Step 3: Implement Gauge R&R

1. Click on the red triangle button next to "Variability Gauge"
2. Click "Gauge Studies" -> "Gauge RR"
3. A window named "Enter/Verify Gauge R&R Specifications" opens
4. Enter the specified value into "K, Sigma Multiplier" box. In this example, we use 5.15 to assume a 99% spread of the data.

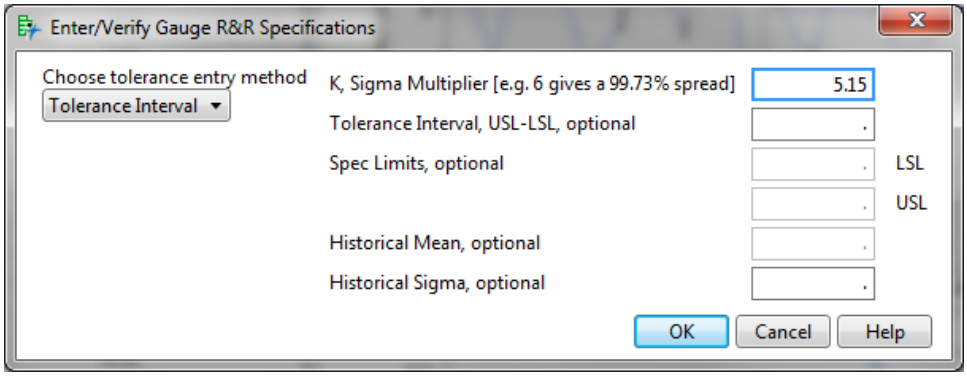


Fig. 2.53 Verify Gauge R&R Specifications

5. Click “OK”

VariableMSA - EMP Measurement Systems Analysis of Measurement by Operator - JMP

Variability Gauge

Variability Chart for Measurement

Gauge R&R

Measurement Source		Variation (5.15*StdDev)		which is 5.15*sqrt of
Repeatability	(EV)	1.0075306	Equipment Variation	V(Within)
Reproducibility	(AV)	0.8408528	Appraiser Variation	V(Operator) + V(Operator*Part)
Operator		0.8085026		V(Operator)
Operator*Part		0.2309915		V(Operator*Part)
Gauge R&R	(RR)	1.3123076	Measurement Variation	V(Within) + V(Operator) + V(Operator*Part)
Part Variation	(PV)	4.8442316	Part Variation	V(Part)
Total Variation	(TV)	5.0188376	Total Variation	V(Within) + V(Operator) + V(Operator*Part) + V(Part)

5.15	k
26.1476	% Gauge R&R = 100*(RR/TV)
0.2709	Precision to Part Variation = RR/PV
5	Number of Distinct Categories = 1.41(PV/RR)

Using last column 'Part' for Part.

Variance Components for Gauge R&R

Component	Var Component	% of Total	20 40 60 80
Gauge R&R	0.06493171	6.84	
Repeatability	0.03827384	4.03	
Reproducibility	0.02665787	2.81	
Part-to-Part	0.88478009	93.16	

Fig. 2.54 JMP Variable Gauge R&R Output

Step 4: Create Mean Plots for further analysis

1. Click on the red triangle button next to “Variability Gauge”
2. Click “Gauge Studies” -> “Gauge R&R Plots” -> “Mean Plots”
3. Three plots appear

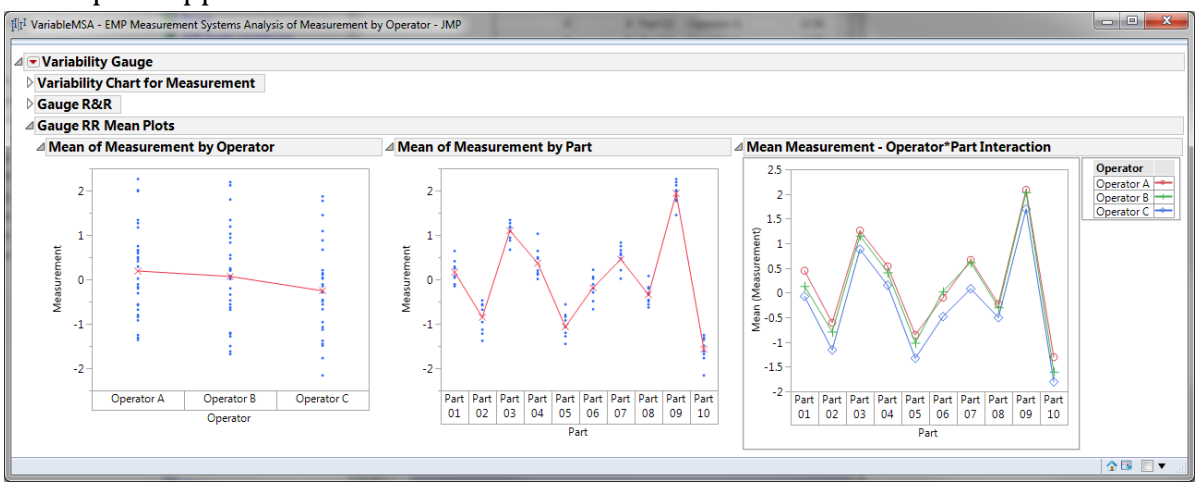


Fig. 2.55 Mean Plots by Operator, Part and Operator by Part Interaction

The result of this Gage R&R study leaves room for consideration on one key measure. As noted in previous pages, the targeted percent contribution R&R should be less than 9% and study

variation less than 30%. With % contribution at 7% it is below our 9% unacceptable threshold and similarly, Study variation at 26.1476% is also below the threshold of 30% but this result is at best marginal and should be heavily scrutinized by the business before concluding that the measurement system does not warrant further improvement.

Visual evaluation of this measurement system is another effective method of evaluation but can at times be misleading without the statistics to support it. Diagnosing the mean plots above should help in the consideration of measurement system acceptability, you may benefit from taking a closer look at operator C.

Use JMP to Implement an Attribute MSA

 ***Data File: "AttributeMSA.jmp"***

Steps in JMP to run an attribute MSA:

1. Click Analyze -> Quality & Process ->Variability/Attribute Gauge Chart
2. Select "Appraiser A", "Appraiser B" and "Appraiser C" as "Y, Response"
3. Select "Part" as "X,Grouping"
4. Select "Reference" as "Standard"
5. Select "Attribute" as the "Chart Type"
6. Click "OK"

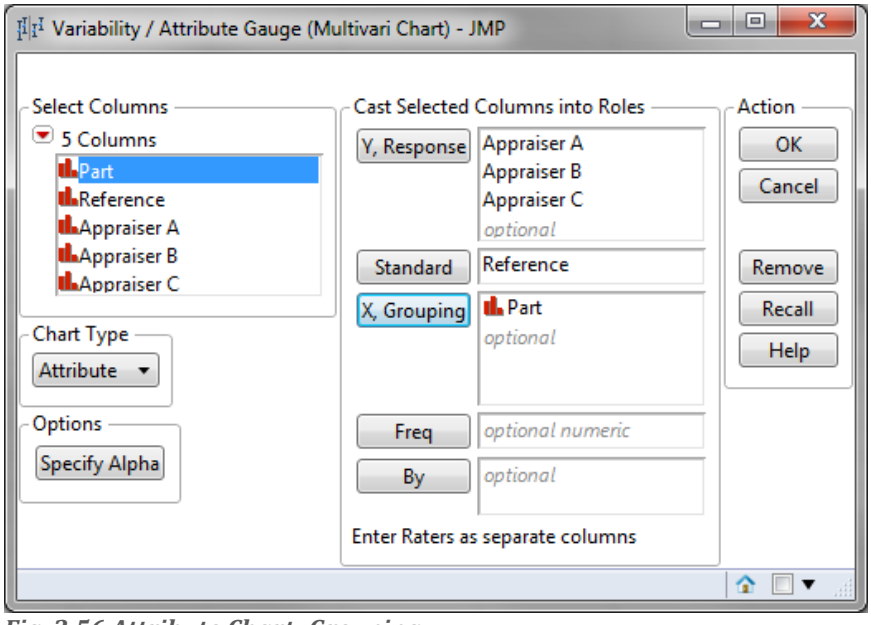


Fig. 2.56 Attribute Chart- Grouping

7. Click on the red triangle button next to "Attribute Gauge"
8. Click "Show Effectiveness Points"
9. Click "Connect Effectiveness Points"

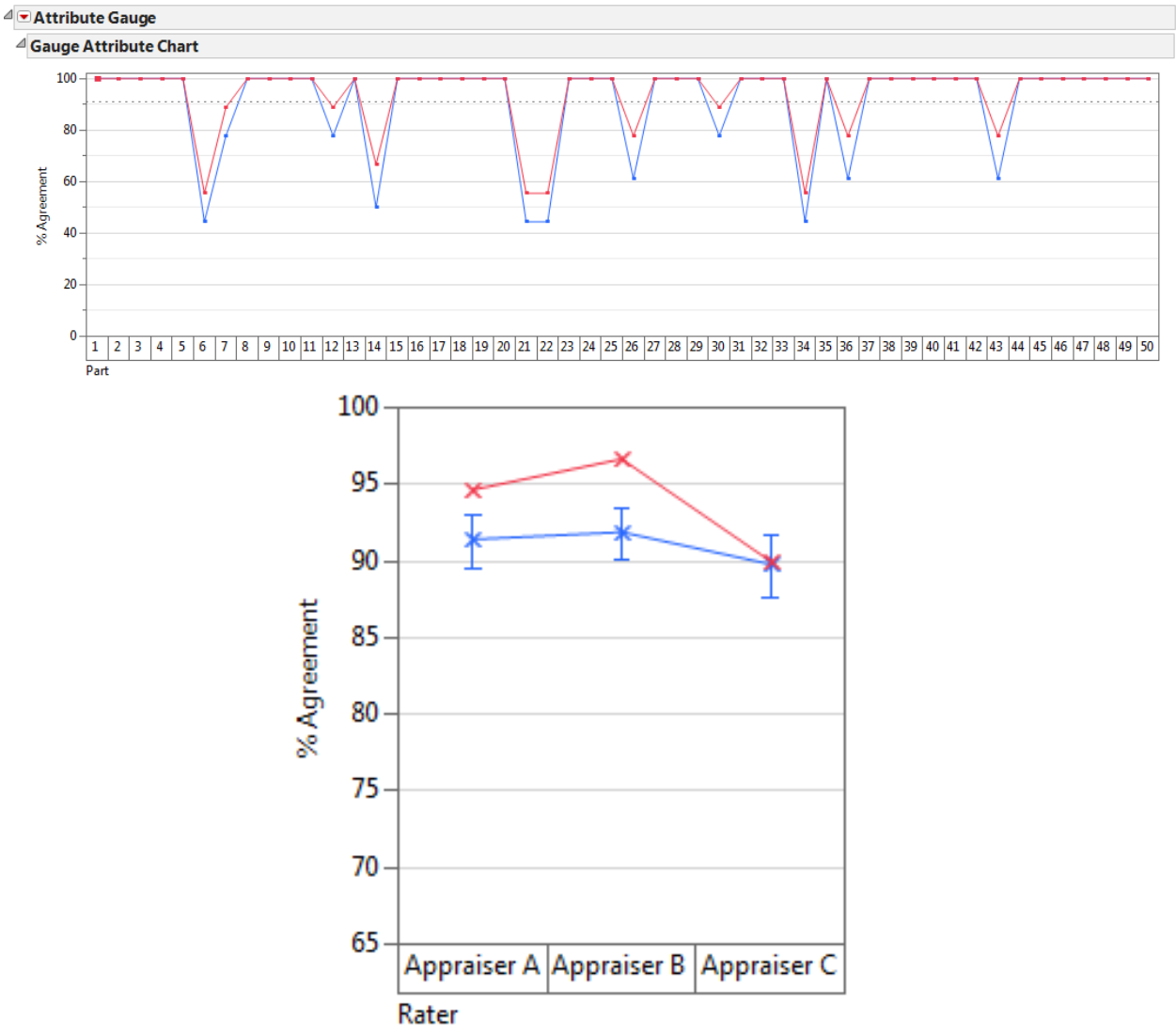


Fig. 2.57 Gauge Attribute Chart

Percentage of agreement by appraiser

- Red line: the percentage of agreement with the reference level
- Blue line: the percentage of agreement between and within the appraisers
- When both lines are at 100% level across parts and appraisers, the measurement system is perfect.

Agreement Report

Rater	% Agreement	95% Lower CI	95% Upper CI
Appraiser A	91.4286	89.5082	93.0248
Appraiser B	91.9048	90.0502	93.4388
Appraiser C	89.8095	87.6057	91.6588

Number Inspected	Number Matched	% Agreement	95% Lower CI	95% Upper CI
50	39	78.000	64.758	87.246

Agreement within Raters

Rater	Number Inspected	Number Matched	Rater Score	95% Lower CI	95% Upper CI
Appraiser A	50	42	84.0000	71.4858	91.6626
Appraiser B	50	45	90.0000	78.6398	95.6524
Appraiser C	50	40	80.0000	66.9629	88.7562

Fig. 2.58 Attribute MSA Results

- % Agreement: Overall agreement percentage of both within and between appraisers. It reflects how precise the measurement system performs.
- In this example, 78% of items inspected have the same measurement across different appraisers and also within each individual appraiser.
- Rater Score: the agreement percentage within each individual appraiser.

Kappa statistic is a coefficient indicating the agreement percentage above the expected agreement by chance. Kappa ranges from −1 (perfect disagreement) to 1 (perfect agreement). When the observed agreement is less than the chance agreement, Kappa is negative. When the observed agreement is greater than the chance agreement, Kappa is positive. Rule of thumb: If Kappa is greater than 0.7, the measurement system is acceptable. If Kappa is greater than 0.9, the measurement system is excellent.

Agreement Comparisons

Rater	Compared with Rater	Kappa	.2 .4 .6 .8	Standard Error
Appraiser A	Appraiser B	0.8629		0.0442
Appraiser A	Appraiser C	0.7761		0.0547
Appraiser B	Appraiser C	0.7880		0.0537

Rater	Compared with Standard	Kappa	.2 .4 .6 .8	Standard Error
Appraiser A	Reference	0.8788		0.0416
Appraiser B	Reference	0.9230		0.0338
Appraiser C	Reference	0.7740		0.0551

Agreement across Categories

Category	Kappa	.2 .4 .6 .8	Standard Error
0	0.7936		0.0236
1	0.7936		0.0236
Overall	0.7936		0.0236

Fig. 2.59 Implementing an Attribute MSA using JMP

The first table shows the Kappa statistic of the agreement between appraisers. The second table shows the Kappa statistic of the agreement between individual appraiser and the standard. The bottom table shows the categorical kappa statistic to indicate which category in the measurement has worse results.

Agreement Counts

Rater	Correct(0)	Correct(1)	Total Correct	Incorrect(0)	Incorrect(1)	Grand Total
Appraiser A	45	97	142	3	5	150
Appraiser B	45	100	145	3	2	150
Appraiser C	42	93	135	6	9	150

Effectiveness

Rater	Effectiveness	95% Lower CI	95% Upper CI	Error rate
Appraiser A	94.6667	89.8296	97.2730	0.0533
Appraiser B	96.6667	92.4348	98.5680	0.0333
Appraiser C	90.0000	84.1565	93.8459	0.1000
Overall	93.7778	91.1542	95.6603	0.0622

Misclassifications

Standard Level	0	1
0	.	16
1	12	.
Other	0	0

Conformance Report

Rater	P(False Alarms)	P(Misses)
Appraiser A	0.0490	0.0625
Appraiser B	0.0196	0.0625
Appraiser C	0.0882	0.1250

Assumptions	
NonConform =	0
Conform =	1

Fig. 2.60 Attribute MSA results

Count of true positives, true negatives, false positives and false negatives. The effectiveness shows the percentage of the agreement between each appraiser and the standard. It reflects the accuracy of the measurement system.

2.4 PROCESS CAPABILITY

2.4.1 CAPABILITY ANALYSIS

What is Process Capability?

Process capability measures how well the process performs to meet given specified outcome. It indicates the conformance of a process to meet given requirements or specifications. *Capability analysis* helps to better understand the performance of the process with respect to meeting customer's specifications and identify the process improvement opportunities.

Process Capability Analysis Steps

Step 1: Determine the metric or parameter to measure and analyze.
Step 2: Collect the historical data for the parameter of interest.
Step 3: Prove the process is statistically stable (i.e., in control).
Step 4: Calculate the process capability indices.
Step 5: Monitor the process and ensure it remains in control over time. Update the process capability indices if needed.

Process Capability Indices

Process capability can be presented using various indices depending on the nature of the process and the goal of the analysis. Popular process capability indices are:

- C_p
- P_p
- C_{pk}
- P_{pk}
- C_{pm}

Capability of the Process (Cp)

C_p can be calculated with the formula:

$$C_p = \frac{USL - LSL}{6 \times \sigma_{within}}$$

Where:

$$\sigma_{within} = \frac{s_p}{c_4(d+1)}$$

$$s_p = \sqrt{\frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)}{\sum_i (n_i - 1)}}$$

$$d = \sum_i (n_i - 1)$$

$$c_4 = \frac{4(n-1)}{(4n-3)}$$

The C_p index is process capability. It assumes the process mean is centered between the specification limits and essentially is the ratio of the distance between the specification limits to six process standard deviations. Obviously, the higher this value the better, because it means you can fit the process variation between the spec limits more easily. C_p measures the process' potential capability to meet the two-sided specifications. It does not take the process average into consideration.

High C_p indicates the small spread of the process with respect to the spread of the customer specifications. C_p is recommended when the process is centered between the specification limits. C_p works when there are both upper and lower specification limits. The higher C_p the better, meaning the spread of the process is smaller relative to the spread of the specifications.

Performance of the Process (P_p)

P_p can be calculated with the formula:

$$P_p = \frac{USL - LSL}{6 \times \sigma_{overall}}$$

Where:

$$\sigma_{overall} = \frac{s}{c_4(n)}$$

$$s = \sqrt{\sum_i \sum_j \frac{\left(x_{ij} - \bar{x}\right)^2}{n-1}}$$

$$c_4 = \frac{4(n-1)}{(4n-3)}$$

Like C_p, P_p assumes the process mean is centered between the spec limits and essentially is the ratio of the distance between the spec limits to six process standard deviations. Obviously, the higher this value the better, because it means you can fit the process variation between the spec limits more easily. Similar to C_p, P_p measures the capability of the process to meet the two-sided specifications. It only focuses on the spread and does not take the process centralization into consideration. It is recommended when the process is centered between the specification limits.

C_p considers the within-subgroup standard deviation and P_p considers the total standard deviation from the sample data. P_p works when there are both upper and lower specification limits. P_p is a more conservative method for determining capability because it considers the total variation of a process instead of a subgroup of data.

Capability of the Process with a k Factor Adjustment (C_{pk})

C_{pk} can be calculated with the formula:

$$C_{pk} = (1 - k) \times C_p$$

Where:

$$k = \frac{|m - \mu|}{\frac{USL - LSL}{2}}$$

$$m = \frac{USL + LSL}{2}$$

The *m* is essentially the center point between the spec limits (it is calculated like an average). When calculating *k*, the numerator is the absolute value of the difference between the center point between the spec limits (*m*) and the process mean (μ); the denominator is half the distance between the specs. As you can guess from the calculation of *k*, it takes into consideration the location of the mean between the spec limits. The formulas to calculate C_{pk} can also be expressed as follows:

$$C_{pk} = \min(\frac{USL - \mu}{3 \times \sigma_{within}}, \frac{\mu - LSL}{3 \times \sigma_{within}}$$

Where:

$$\sigma_{within} = \frac{s_p}{c_4(d + 1)}$$

$$s_p = \sqrt{\frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)}{\sum_i (n_i - 1)}}$$

$$d = \sum_i (n_i - 1)$$

$$c_4 = \frac{4(n - 1)}{(4n - 3)}$$

The expression for C_{pk} is indicating that the limiting factor is the specification limit that the mean is closest to. C_{pk} measures the process' actual capability by taking both the variation and average of the process into consideration. The process does not need to be centered between the specification limits to make the index meaningful. C_{pk} is recommended when the process is not in the center between the specification limits. When there is only a one-sided limit, C_{pk} is calculated using C_{pu} or C_{pl}. C_{pk} for upper specification limit:

$$C_{pu} = \frac{USL - \mu}{3 \times \sigma_{within}}$$

C_{pk} for lower specification limit:

$$C_{pl} = \frac{\mu - LSL}{3 \times \sigma_{within}}$$

Where:

- *USL* and *LSL* are the upper and lower specification limits.
- μ is the process mean.

Performance of the Process with a k Factor Adjustment (P_{pk})

Ppk can be calculated with the formula:

$$P_{pk} = (1 - k) \times P_p$$

Where:

$$k = \frac{|m - \mu|}{\frac{USL - LSL}{2}}$$

$$m = \frac{USL + LSL}{2}$$

Where:

- *USL* and *LSL* are the upper and lower specification limits
- μ is the process mean.

P_{pk} is to P_p as C_{pk} is to C_p. The *k* factor is the same used earlier to calculate C_{pk}.

The formulas to calculate P_{pk} can also be expressed as follows:

$$P_{pk} = \min(\frac{USL - \mu}{3 \times \sigma_{overall}}, \frac{\mu - LSL}{3 \times \sigma_{overall}}$$

$$\sigma_{overall} = \frac{s}{c_4(n)}$$

$$s = \sqrt{\sum_i \sum_j \frac{\left(x_{ij} - \bar{x}\right)^2}{n - 1}}$$

$$c_4 = \frac{4(n - 1)}{(4n - 3)}$$

Similar to C_{pk}, P_{pk} measures the process capability by taking both the variation and the average of the process into consideration. P_{pk} solves the decentralization problem P_p cannot overcome. C_{pk} considers the within-subgroup standard deviation, while P_{pk} considers the total standard deviation from the sample data. When there is only a one-sided specification limit, P_{pk} is calculated using P_{pu} or P_{pl}. P_{pk} for upper specification limit:

$$P_{pu} = \frac{USL - \mu}{3 \times \sigma_{overall}}$$

P_{pk} for lower specification limit:

$$P_{pl} = \frac{\mu - LSL}{3 \times \sigma_{overall}}$$

Where:

- *USL* and *LSL* are the upper and lower specification limits
- μ is the process mean.

Taguchi's Capability Index (C_{pm})

C_p, P_p, C_{pk}, and P_{pk} all consider the variation of the process. C_{pk} and P_{pk} take both the variation and the average of the process into consideration when measuring the process capability. It is possible that the process average fails to meet the target customers require while the process still remains between the specification limits. C_{pm} (Taguchi's capability index) helps to capture the variation from the specified target. Rather than only telling us if a measurement is good (between the spec limits), C_{pm} helps us understand *how good* the measurement is. C_{pm} can be calculated with the formula:

$$C_{pm} = \frac{\min(T - LSL, USL - T)}{3 \times \sqrt{s^2 + (\mu - T)^2}}$$

Where:

- *USL* and *LSL* are the upper and lower specification limits
- *T* is the specified target
- μ is the process mean.

Note: C_{pm} can work only if there is a target value specified.

Capability Interpretation

Interpreting the results of a capability analysis is dependent upon the nature of the business, product and customer. The Automotive Industry Action Group (AIAG) suggests that P_p and P_{pk} should be > 1.67 while C_p and C_{pk} should be > 1.33. The reality is that anything greater than 1.0 is a fairly-capable process but your business needs to assess the costs vs. benefits of achieving capability greater than 1.67 or even higher. If you're dealing with customer safety or life and death influences then obviously the product necessitates a capability greater than 2.0. Below is a simple table for quick interpretation with sigma level references.

Index	Value	Interpretation	Sigma Level
C_{pk}	<1.0	Not Very Capable	<3
C_{pk}	1.0-1.99	Capable	3-6
C_{pk}	>2.0	Very Capable	>6

Table 2.3 Capability Index Interpretation

Use JMP to Run a Process Capability Analysis

Data File: "CapabilityAnalysis.jmp"

Steps in JMP to run a process capability analysis:

1. Click Analyze -> Distribution
2. Select "HtBk" as "Y, Columns"
3. Click "OK"

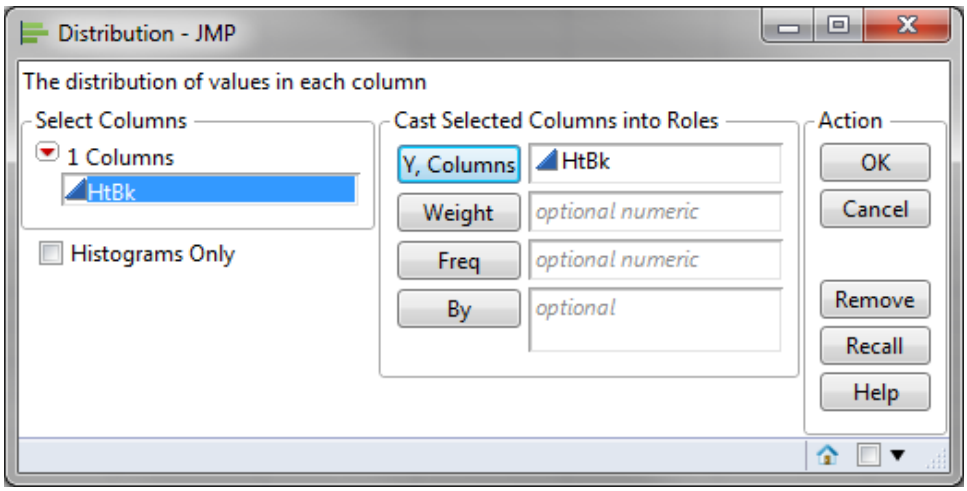


Fig. 2.61 Process Capability Columns Window

4. Click on the red button next to "HtBk"
5. Click "Capability Analysis"
6. A window named "Capability Analysis, Setting Specification Limits" pops up
7. Enter the 6.0 as the "Lower Spec Limit", 6.5 as the "Target" and 7.0 as the "Upper Spec Limit"
8. Click OK

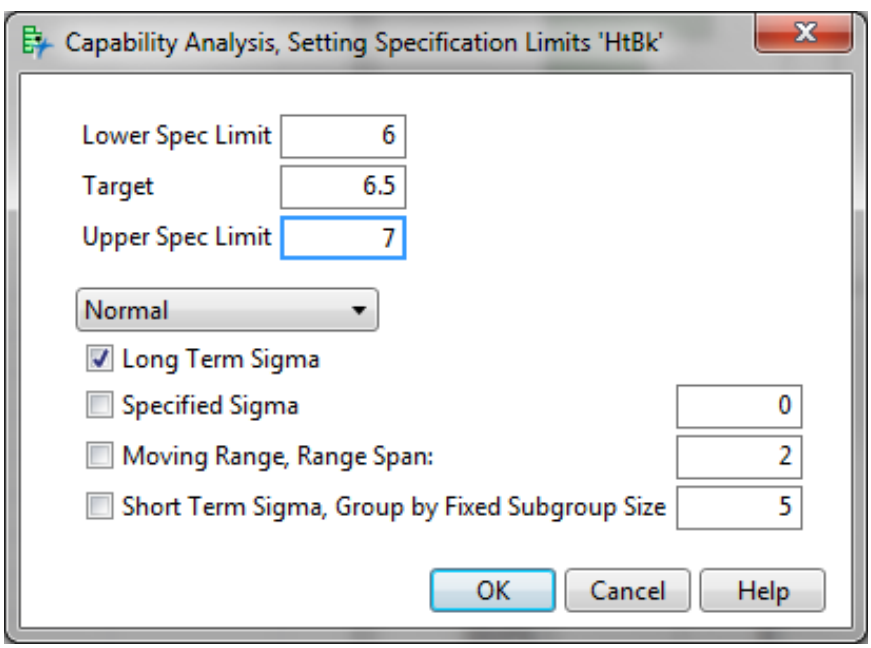
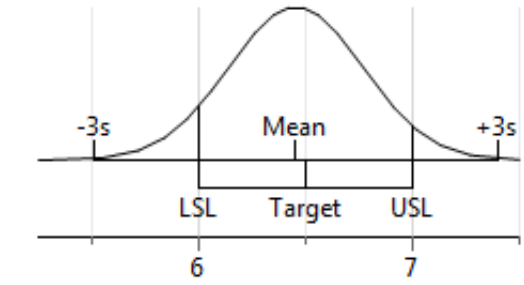


Fig. 2.62 Selection Window for Specification Limits

Capability Analysis

Specification	Value	Portion	% Actual
Lower Spec Limit	6	Below LSL	5.0000
Spec Target	6.5	Above USL	0.0000
Upper Spec Limit	7	Total Outside	5.0000

Long Term Sigma



Sigma = 0.31422

Capability	Index	Lower CI	Upper CI
CP	0.530	0.413	0.648
CPK	0.481	0.332	0.629
CPM	0.525	0.415	0.647
CPL	0.481	0.331	0.628
CPU	0.580	0.414	0.743

Portion	Percent	PPM	Sigma Quality
Below LSL	7.4587	74587.383	2.942
Above USL	4.0928	40928.354	3.240
Total Outside	11.5516	115515.74	2.698

Fig. 2.63 Process Capability Analysis

With a CPK of less than 1.0 we can conclude that the capability of this process is not very good. Anything less than 1.0 should be considered not capable and we should strive for a CPK to reach levels of greater than 1 and preferably over 1.33.

2.4.2 Concept of Stability

What is Process Stability?

A process is said to be stable when:

- The process is in control
- The future behavior of the process is predictable at least between some limits
- There is only random variation involved in the process.
- The causes of variation in the process are only due to chance or common causes
- There are not any trends, patterns, or outliers in the control chart of the process

Stability means that the process is in control; there are no special causes (only random variation) influencing the process outcome. Future process outcomes are predictable within a range of values, and there are no trends, patterns, or outliers in a control chart of the process.

Root Causes of Variation in the Process

To discuss stability, it is important that we define and differentiate common cause and special cause variation. Common causes are unable to be eliminated from the process. Typical common causes are:

- Chance

- Random and anticipated
- Natural noise
- Inherent in the process

Special causes are:

- Assignable cause
- Unanticipated
- Unnatural pattern
- The signal of changes in the process

Special causes are able to be eliminated from the process. Common cause variation is left to chance. It is random, it is expected, and inherent to the process; therefore, it is unable to be eliminated. Special cause, on the other hand, can be attributed to an "assignable cause." It is not expected (or anticipated) and is external to the process. These causes signal changes in the process and can be eliminated.

Control Charts

Control charts are the graphical tools to analyze the stability of a process. A control chart is used to identify the presence of potential special causes in the process and to determine whether the process is statistically in control. If the samples or calculations of samples are all in control, the process is stable and the data from the process can be used to predict the future performance of the process.

Popular Control Charts

- I-MR Chart—Individuals and moving range
- Xbar-R Chart—Xbar is the average and R is range
- Xbar-S Chart—Xbar is the average and S is standard deviation
- C Chart—Uses count data to chart non-conformances or defects)
- U Chart—U stands for units. It is the same as a C chart, but a C chart assumes subgroup side is constant, while a U chart assumes uneven subgroup sizes.
- P Chart—Proportions chart, which shows the proportion defective
- NP Chart—Shows number defective (with equal subgroup size)
- EWMA Chart—Exponentially weighted moving average
- CUSUM Chart—Cumulative sum

Process Stability vs. Process Capability

Process stability indicates how stable a process performed in the past. When the process is stable, we can use the data from the process to predict its future behavior. Process capability indicates how well a process performs with respect to meeting the customer's specifications. The process capability analysis is valid only if the process is statistically stable (i.e., in control,

predictable). Being stable does *not* guarantee that the process is also capable. However, being stable is the prerequisite to determine whether a process is capable.

2.4.3 Attribute and Discrete Capability

Process Capability Analysis for Binomial Data

If we are measuring the count of defectives in each sample set to assess the process performance of meeting the customer specifications, we use "%Defective" (percentage of items in the samples that are defective) as the process capability index. %Defective is given by a very simple calculation: determine the number of defective units divided by the number of units overall.

$$\%\ Defective = \frac{N_{defectives}}{N_{overall}}$$

Where:

- $N_{defectives}$ is the total count of defectives in the samples
- $N_{overall}$ is the sum of all the sample sizes.

Process Capability Analysis for Poisson Data

If we are measuring the count of defects in each sample set to assess the process performance of meeting the customer specifications, we use Mean DPU (defects per unit of measurement) as the process capability index:

$$DPU = \frac{N_{defects}}{N_{overall}}$$

Where:

- $N_{defects}$ is the total count of defects in the samples
- $N_{overall}$ is the sum of all the units in the samples.

2.4.4 Monitoring Techniques

Capability and Monitoring

In the Measure phase of the project, process stability analysis and process capability analysis are used to baseline the performance of current process. In the Control phase of the project, process stability analysis and process capability analysis are combined to monitor whether the improved process is maintained consistently as expected.

3.0 ANALYZE PHASE

3.1 PATTERNS OF VARIATION

3.1.1 MULTI-VARI ANALYSIS

What is Multi-Vari Analysis?

Multi-vari analysis is a graphic-driven method to analyze the effects of categorical inputs on a continuous output. It studies how the variation in the output changes across different inputs and helps us quantitatively determine the major source of variability in the output. Multi-vari charts are used to visualize the source of variation. They work for both crossed and nested hierarchies.

- Y: continuous variable
- X's: discrete categorical variables. One X may have multiple levels

Hierarchy

Hierarchy is a structure of objects in which the relationship of objects can be expressed similar to an organization tree. Each object in the hierarchy is described as above, below, or at the same level as another one. If object A is above object B and they are directly connected to each other in the hierarchy tree, A is B's parent and B is A's child. In Multi-vari analysis, we use the hierarchy to present the relationship between categorical factors (inputs). Each object in the hierarchy tree indicates a specific level of a factor (input). There are generally two types of hierarchies: crossed and nested.

Crossed Hierarchy
In the hierarchy tree, if one child item has more than one parent item at the higher level, it is a crossed relationship.

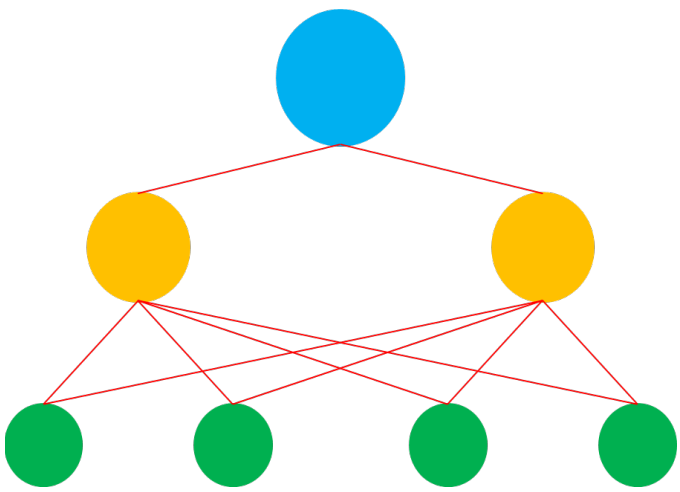


Fig. 3.1 Crossed Hierarchy

Consider the top dot is a school district, the yellow dots are schools, and the middles dots are neighborhoods. In many cities, there are choice systems that allow children from the same

neighborhood to attend different schools, so the neighborhoods cannot be perfectly nested under a school.

Nested Hierarchy
In the hierarchy tree, if one child item only has one parent item at the higher level, it is a nested relationship.

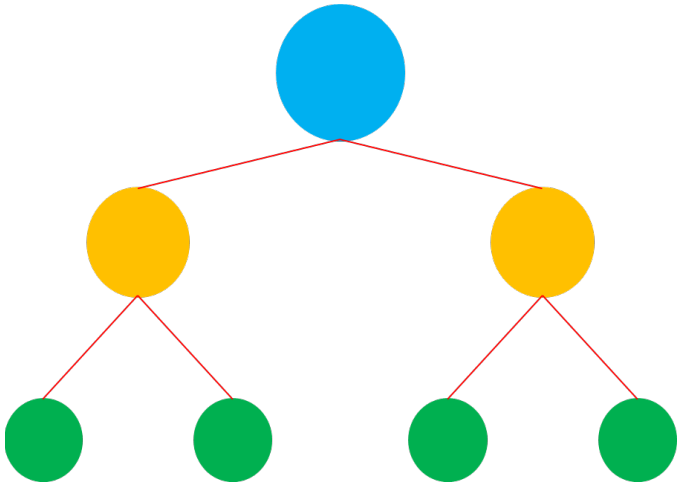


Fig. 3.2 Nested Hierarchy

Consider the example of a call center (the top dot), the middle dots represent units in the call center, and the bottom dots represent associates in the unit. Each associate can only be a part of one unit, and each unit can only be a part of the one site.

Use JMP to Perform Multi-Vari Analysis

Data File: "Multi-Vari.jmp"

Case study: ABC Company produces 10 kinds of units with different weights. Operators measure the weights of the units before sending them out to customers. Multiple factors could have an impact on the weight measurements. The ABC Company wants to have a better understanding of the main source of variability existing in the weight measurement. The ABC Company randomly selects three operators (Joe, John, and Jack) each of whom measures the weights of 10 different units. For each unit, there are three items sampled.

Steps in JMP to perform a Multi-Vari Analysis:

1. Click Analyze -> Variability/Gauge Chart
2. Select "Measurement" as "Y, Response"
3. Select both "Operator" and "Unit" as "X, Grouping"

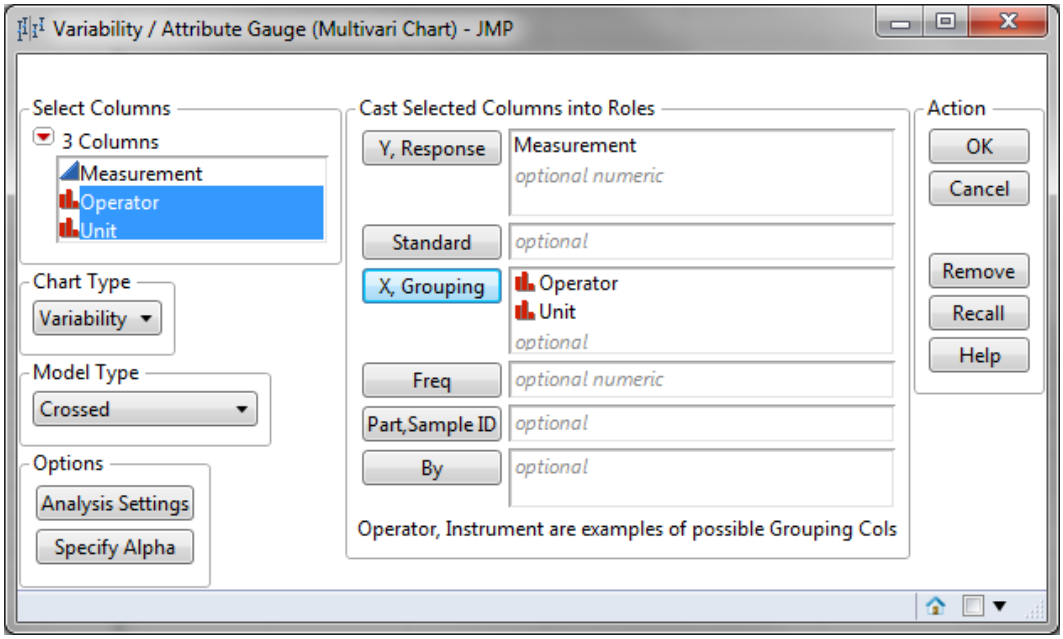


Fig. 3.3 Multi-Vari Charts variable selection box

4. Select "Crossed" as "Model Type" since it's a crossed hierarchy
5. Click "OK"
6. In the red triangle button next to "Variability Gauge"
7. Click "Connect Cell Means" to create lines connecting individual observation
8. Click "Show Group Means" to display the average within individual group
9. Click "Show Grand Mean" to display the average across all the groups
10. Click "Variance Components" to display the proportion table of variance components

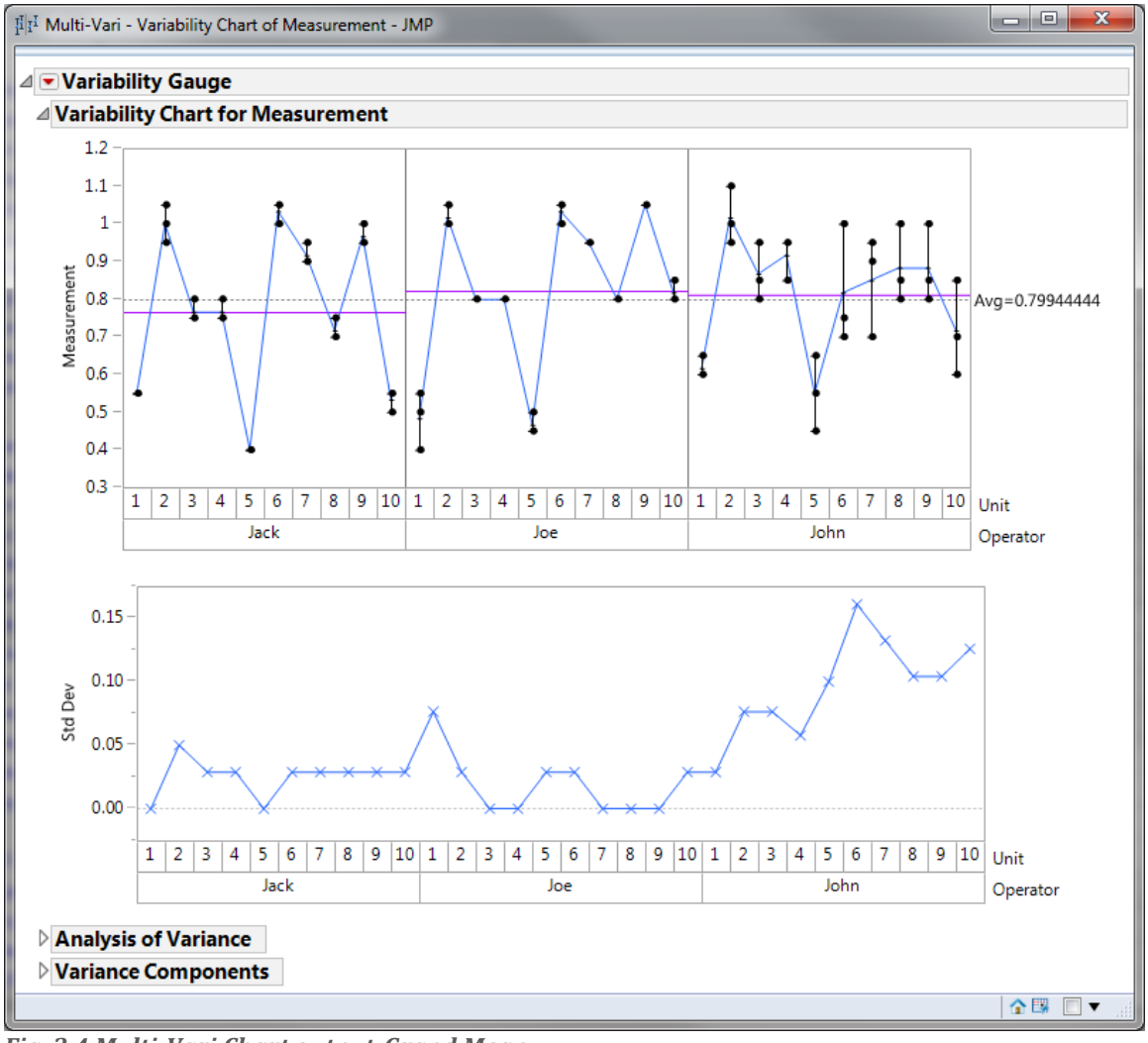


Fig. 3.4 Multi-Vari Chart output-Grand Mean

Connecting the cell means enables you to see the variation within and between units, and within operators. When adding the group means you can see the variation between operators as well. The grand mean on the right of the chart provides a broad reference point.

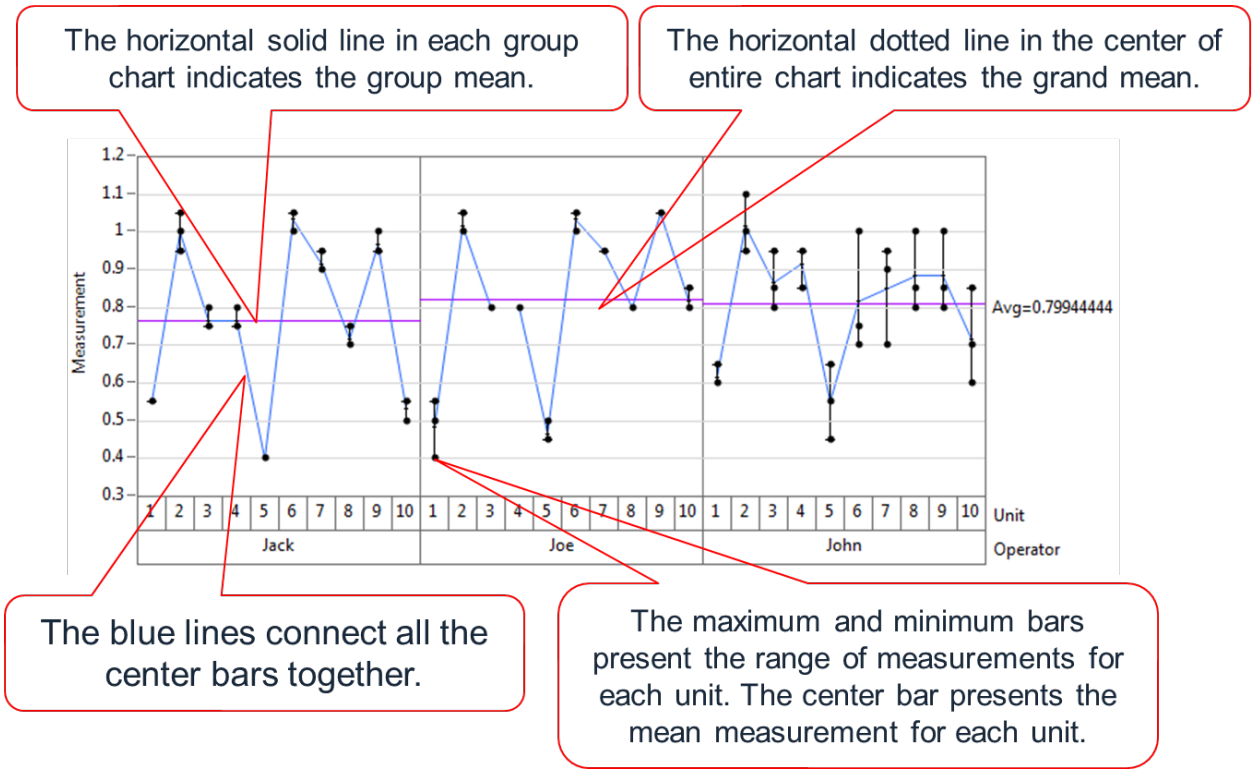
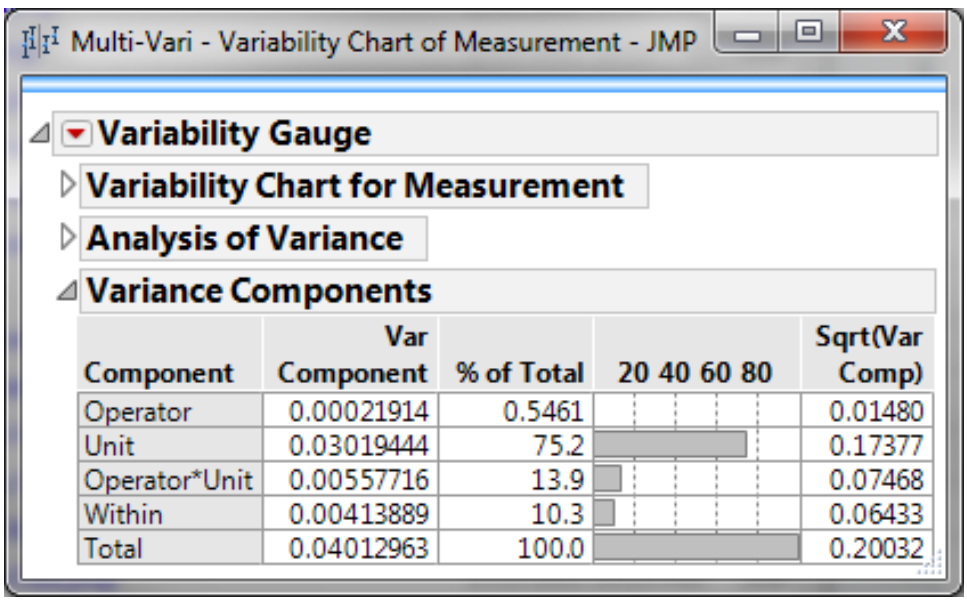


Fig. 3.5 Multi-Vari Chart output interpretation

Based on the Multi-Vari chart, the measurement of units range from 0.3 to 1.2. Joe's and John's mean measurements stay between 0.8 and 0.9. Jack's mean is slightly lower than both Joe's and John's. John has the worst variation when measuring the same kind of unit because John has the highest difference between the maximum and minimum bars for any kind of unit. By observing the black lines of three operators, it seems like all three operators' measurements follow the same pattern. The operator-to-operator variability is not large. The unit-to-unit variability is large and it could be the main source of variation in measurements.

Quantifying Variance Components



Component	Var Component	% of Total	20 40 60 80	Sqrt(Var Comp)
Operator	0.00021914	0.5461		0.01480
Unit	0.03019444	75.2		0.17377
Operator*Unit	0.00557716	13.9		0.07468
Within	0.00413889	10.3		0.06433
Total	0.04012963	100.0		0.20032

Fig. 3.6 Variance Components Data

JMP can then quantify the amount of variance that each component contributes to the system. The unit, as you can see from the table above, contributes 75% of the variance.

3.1.2 Classes of Distribution

Probability is the likelihood of an event occurring. The probability of event A occurring is written as P(A). Probability is a percentage between 0% and 100%. The higher the probability, the more likely the event will happen.

- When probability is equal to 0%, it indicates that the event will take place with no chance
- When probability is equal to 100%, it indicates the event will definitely take place without any uncertainty

Example of probability: When tossing a coin, there is 50% chance that the head lands face up and 50% chance that the head lands face down.

Probability Property

The probability of event A not occurring is:

$$P(\bar{A}) = 1 - P(A)$$

The probability of event A OR B OR both events occurring:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

The probability of A OR B, OR "A and B" occurring is written as: the probability of A plus the probability of B minus the probability of A and B. If the events are mutually exclusive, meaning only one can occur, the probability of A and B occurring is zero.

$$P(A \cap B) = 0$$

and

$$P(A \cup B) = P(A) + P(B)$$

Consider rolling a die. What is the probability of rolling a 2 or a 4? Because you cannot role more than one value at a time, they are mutually exclusive; therefore, the probability of rolling a 2 or a 4 is written as P(2) + P(4).

The probability of event A AND B occurring together:

$$P(A \cap B) = P(A|B)P(B)$$

If events A and B are independent of each other:

$$P(A|B) = P(A)$$

and

$$P(A \cap B) = P(A)P(B)$$

P(A|B) refers to the conditional probability that event A occurs, given that event B has occurred.

The probability of event A occurring given event B taking place:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Where:

$$P(B) \neq 0$$

What is Distribution?

Distribution (or *probability distribution*) describes the range of possible events and the possibility of each event occurring. In statistical terms, distribution is the probability of each possible value of a random variable when the variable is discrete, or the probability of a value falling in a specific interval when the variable is continuous.

Probability Mass Function

PMF (*Probability Mass Function*) is a function describing the probability of a discrete random variable equal to a particular value. Below is the probability mass function chart for a fair die. There are six possible outcomes of the die, and with a fair die, each outcome has equal probability.

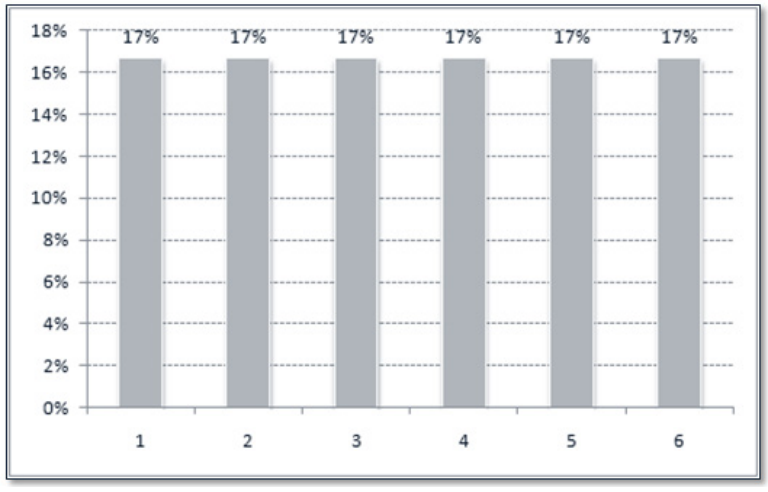


Fig. 3.7 Probability Mass Function Graph

Probability Density Function

PDF (*Probability Density Function*) is a function used to retrieve the probability of a continuous random variable falling within a particular interval. The probability of a variable falling within

a particular region is the integral of the probability density function over the region. Here is the probability density function chart of a standard normal distribution.

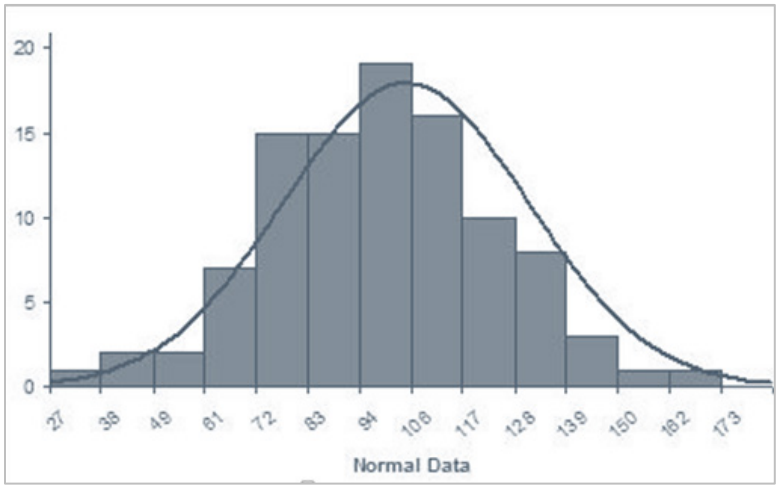


Fig. 3.8 Probability Density Function Graph

A *probability mass function* is to discrete data as a probability density function is to continuous data. The bars represent the probability of a value occurring within a particular interval. In this example of a standard normal distribution, there is nearly a 20% chance that a value will fall between 94 and 108.

Advantages of Distribution

Distributions capture the most basic features of data:

- Shape
    - Which distribution family may the data belong to?
    - Is the data symmetric?
    - Are the tails of the data flat?
- Center
    - Where is the midpoint of the data?
- Scale
    - What is the range of data?

The *shape* of the distribution tells us what type of distribution it is and how we should analyze the data. The *symmetry* of the distribution tells us the probabilities are the same or different at the low and high end of the range of data values. The *center*, or midpoint, tells us what is most likely to occur—average, or median, for example. The *scale* refers to how much spread there is in the data—think of range or standard deviation.

Measures of Skewness and Kurtosis

The shape, center (i.e., location), and scale (i.e., variability) are three basic data characteristics that probability distributions capture. Skewness and kurtosis are the two more advanced characteristics of the probability distribution.

Skewness

Skewness is a measure of the asymmetry degree of the probability distribution.

The mathematical definition of the skewness is:

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{(E[(X-\mu)^2])^{3/2}} = \frac{k_3}{k_2^{3/2}}$$

Where: μ_3 is the third moment about the mean and σ is the standard deviation.

The skewness of a normal distribution is zero. When the distribution looks symmetric to the left side and the right side of the center point, the skewness is close to zero. Skewness can be caused by extreme values or bounds on one side of the distribution, such as when no negative values are possible.

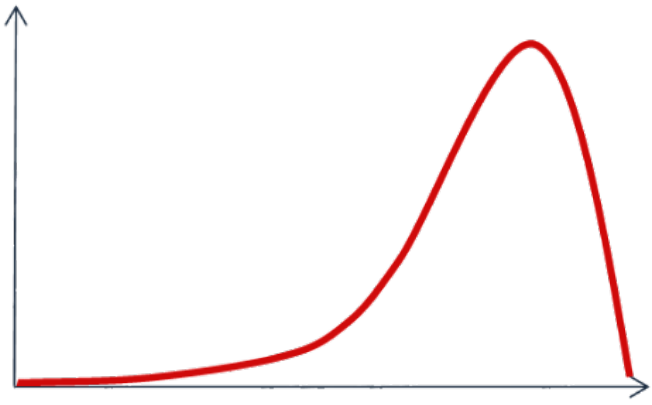


Fig. 3.9 Negatively Skewed Graph

A negative value of skewness indicates that the distribution is left skewed, meaning the left tail is longer than the right. It could be caused by having an upper bound on the distribution. A positive value of skewness indicates that the distribution is right skewed, meaning the right tail is longer than the left. It could be caused by a lower bound of zero.

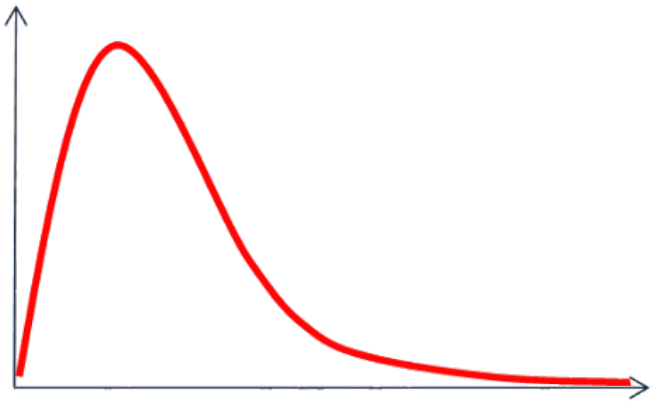


Fig. 3.10 Positively Skewed Graph

Kurtosis

Kurtosis is a measure of the peakedness of the probability distribution.

The mathematical definition of the excess kurtosis is:

$$\gamma_2 = \frac{k_4}{k_2^2} = \frac{\mu_4}{\sigma^4} - 3$$

The kurtosis of a normal distribution is zero. Distributions with a zero-excess kurtosis are called *mesokurtic*. The distribution with a high kurtosis has a more distinct peak and fatter, longer tails. The distribution with a low kurtosis has a more rounded peak and thinner, shorter tails.

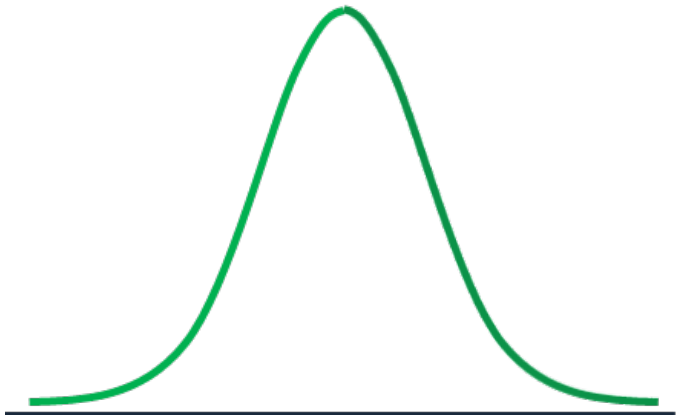


Fig. 3.11 Leptokurtic Kurtosis

As you can see, the distribution with excess kurtosis greater than zero looks more narrow, less rounded at its peak vs. the normal distribution. This type of kurtosis is called *leptokurtic*. The distribution with excess kurtosis less than zero looks more squat than the normal distribution. The type is called *platykurtic*.

Multimodal Distribution

In statistics, a continuous probability distribution with two or more different unimodal distributions is called a *multimodal distribution*. A *unimodal distribution* has only one local maximum which is equal to its global maximum. A multimodal probability distribution has more than one local maximum.

Example: a bivariate, multimodal distribution

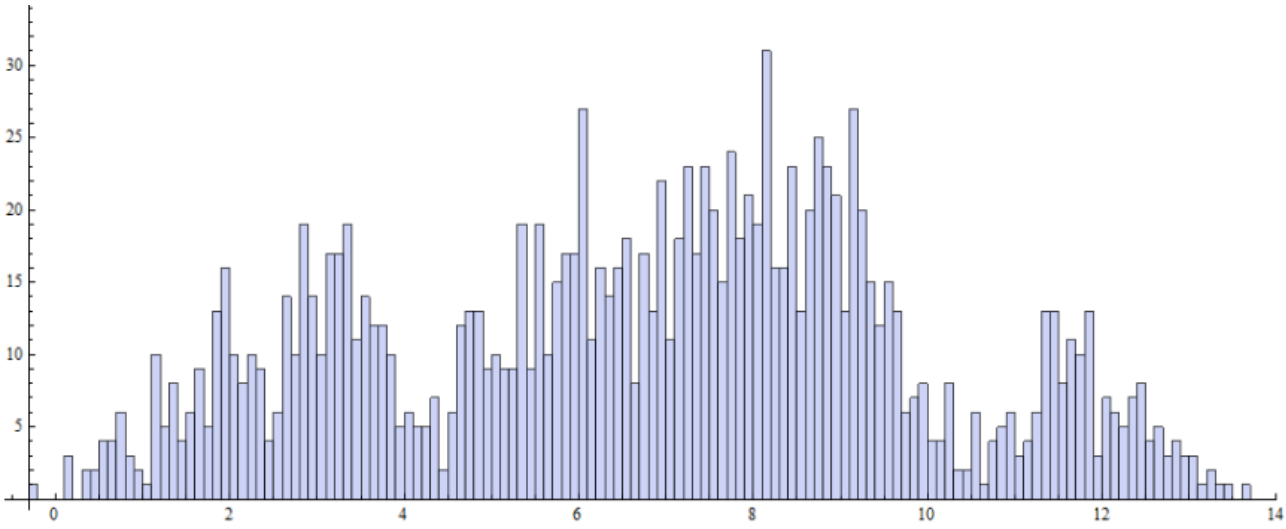


Fig. 3.12 Multimodal Distribution
(source: S&P 500 Composite 20-year total returns)

Bimodal Distribution

In statistics, a continuous probability distribution with two different unimodal distributions is called a *bimodal distribution*. A bimodal probability distribution has two local maxima.

Example:

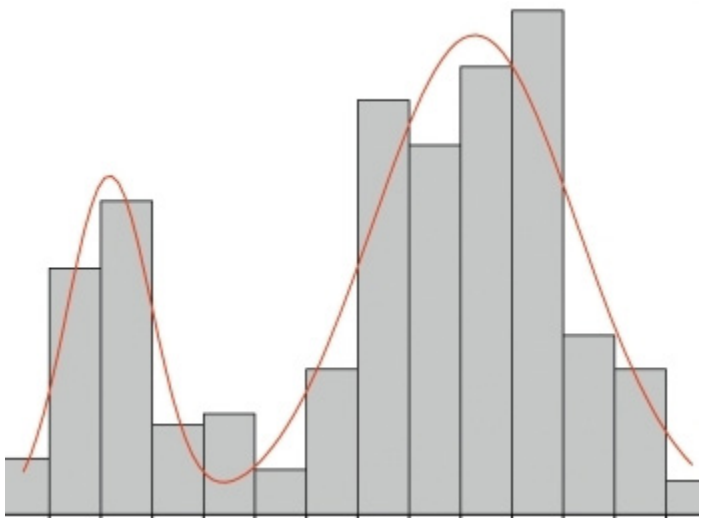


Fig. 3.13 Bimodal Distribution
(source: https://openi.nlm.nih.gov)

By mixing two unimodal distributions with different means, the mixture distribution is not necessarily a bimodal distribution. The mixture of two normal distributions with the same standard deviation is a bimodal distribution only if the difference between their means is at least twice their common standard deviation.

Examples of discrete distribution are:

- Binomial Distribution
- Poisson Distribution

Binomial Distribution

Assume there is an experiment with *n* independent trials, each of which only has two potential outcomes (i.e., yes/no, fail/pass). *p* is the probability of one outcome and (1 – *p*) is the probability of the other.

Binomial distribution is a discrete probability distribution describing the probability of any outcome of the experiment.

Excel formula for binomial distribution:

BINOMDIST(number_s, trials, probability_s, cumulative)

In probability theory and statistics, the binomial distribution is the discrete probability distribution of the number of successes in a sequence of *n* independent yes/no experiments, each of which yields success with probability *p*. Such a success/failure experiment is also called a *Bernoulli experiment* or *Bernoulli trial*; when *n* = 1, the binomial distribution is a Bernoulli distribution.

Consider the coin flip example. There are only two possible outcomes of a trial. The Excel formula will return the probability of a desired result, given the known probability of success. For example, if a batter's average is 0.300, and we want to know the probability that the batter will get seven hits out of 10 at-bats. The entry of the formula would look like this: BINOMDIST(7,10,0.3,0), where the 0 will return the probability mass function. The result is 0.009, meaning less than a 1% chance of getting 10 at-bats.

PMF of Binomial Distribution
The probability of getting *k* successes in *n* trials:

$$f(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Where:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Mean of Binomial Distribution

$$n \times p$$

Variance of Binomial Distribution

$$n \times p \times (1-p)$$

In the example we used, *n* = 10, *k* = 7, and *p* = 0.3. Carrying out the calculations, you will get 0.009. On this page, you will also calculate the mean of a binomial distribution. For a given

number of trials, it is just the number of trials times the probability of success. For our example, the mean is 10 × 0.3 = 3. For the variance, it is the mean times the difference between 1 minus the probability of success.

For our example, the variance is 10 × 0.3 × 0.7 = 2.1

Poisson Distribution

Poisson distribution is a discrete probability distribution describing the probability of a number of events occurring at a known average rate and in a fixed period of time. The expected number of occurrences in this fixed period of time is λ. The Greek letter lambda is used to denote the number of occurrences in a fixed interval of time.

Excel formula for Poisson distribution

POISSON(x, mean, cumulative)

The Excel formula is shown here, where x is the number of events for which we are trying to determine the probability, mean is the expected number of events, and the last argument is telling Excel whether to return the PMF or the cumulative Poisson probability.

An example to consider: Let us assume an office worker typically gets 25 emails a day, and we want to determine the probability of getting 40 emails, Excel returns the result of 0.001.

PMF of Poisson Distribution
The probability of getting k occurrences in *n* trials is:

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Here is the formula for calculating the probability mass function, where lambda (λ) is the mean of Poisson distribution (or expected number of occurrences, 25 in our email example) and also the variance of the Poisson distribution.

Examples of Continuous Distribution

- Normal Distribution
- Exponential Distribution
- Weibull Distribution
- Student's t distribution
- Chi-square distribution
- F distribution

Normal Distribution

Normal distribution is a continuous probability distribution describing random variables which cluster around the mean. In probability theory, the *normal (or Gaussian) distribution* is a continuous probability distribution that has a bell-shaped probability density function, known

as the Gaussian function or more commonly the "bell curve." Its probability density function is a bell-shaped curve. But not all the bell-shaped PDFs are from normal distributions. The normal distribution is the most extensively used distribution.

Excel formula for normal distribution:

NORMDIST(x, mean, standard_dev, cumulative)

The 68–95–99.7 rule for normal distribution:

- About 68% of the data stay within σ from the mean
- About 95% of the data stay within 2σ from the mean
- About 99.7% of the data stay within 3σ from the mean

The Excel formula here can be used to determine the probability of a value X occurring, when the mean and standard deviation are known. As an example, consider the height of professional basketball players. The average height is 77 inches, and the standard deviation is 4 inches. Let us determine the probability that a player is 8 feet tall (or 84 inches). The probability is 0.02.

Remember the 68–95–99.7 rule? These represent the percentage of data that fall within one, two, and three standard deviations from the mean.

PDF of Normal Distribution

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Mean of Normal Distribution:

$$\mu$$

Variance of Normal Distribution

$$\sigma^2$$

Here is the formula for calculating the PDF for a normal distribution. It a little easier to do with Excel, where the Greek letter mu (μ) is used for the mean, and sigma squared (σ^2) is the variance.

Exponential Distribution

Exponential distribution is a continuous probability distribution describing the probability of events occurring at a known constant average rate between points of time. The exponential distribution is very similar to the Poisson distribution, except that the former is built on continuous variables and the latter is built on discrete variables.

Excel formula for exponential distribution

EXPONDIST(x, lambda, cumulative)

In probability theory and statistics, the *exponential distribution* (or negative exponential distribution) is a family of continuous probability distributions. It describes the time between events in a Poisson process, i.e., a process in which events occur continuously and independently at a constant average rate.

Example: Busses arrive at a bus stop at an average of three busses per hour. (a) What is the expected time between busses? (b) What is the probability that the next bus will arrive within five minutes?

Solution: Bus arrivals represent rare events, thus the time T between them is exponential, with rate three busses/hour. So the expected time between arrivals is 1/3 hours or 20 minutes. To determine the probability that the next bus will arrive in five minutes or 1/12 of an hour, let us plug the value of 1/12 in for X, 1/3 in for lambda. The result is 0.32.

In real-world scenarios, the constant rate assumption is rarely satisfied. For example, the rate of incoming phone calls differs according to the time of day. But if we focus on a time interval during which the rate is roughly constant, such as from 2 p.m. to 4 p.m. during work days, the exponential distribution can be used as a good approximate model for the time until the next phone call arrives. In the real world, there are always caveats, but the key is that we can estimate or approximate using distribution models:

- The time until a radioactive particle decays, or the time between clicks of a Geiger counter
- The time it takes before your next telephone call
- The time until default (on payment to company debt holders) in reduced form credit risk modeling

Exponential variables can also be used to model situations where certain events occur with a constant probability per unit length, such as the distance between road kills on a given road. In queuing theory, the service times of agents in a system (e.g., how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially-distributed variables.

PDF of Exponential Distribution

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Mean of Exponential Distribution

$$\frac{1}{\lambda}$$

Variance of Exponential Distribution

$$\frac{1}{\lambda^2}$$

Here is the formula for the PDF of the exponential distribution, where lambda (λ) is the known rate, the mean is 1/lambda, and variance is 1/lambda-squared.

Weibull Distribution

Weibull distribution is a continuous probability distribution which is widely used to model a great variety of data due to its flexibility. This distribution is named after Waloddi Weibull, who described it in detail in 1951, although it was first identified by Fréchet (1927) and first applied by Rosin and Rammler (1933) to describe the size distribution of particles. It is used in survival analysis, reliability/failure analysis, weather analysis, and particle size distributions to name a few.

Excel formula for Weibull distribution

WEIBULL(x, alpha, beta, cumulative)

In Excel, the function is as described on the page. *X* is the value at which we look to evaluate the probability, and alpha and beta are parameters of the distribution. You can see the PDF for the Weibull here, where k > 0 is the shape parameter and λ > 0 is the scale parameter of the distribution.

PDF of Weibull Distribution

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Student's T Distribution

Student's t distribution is a continuous probability distribution that resembles a normal distribution. When *n* random samples were drawn from a normally-distributed population with the mean μ and the standard deviation σ, then

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

follows a t distribution with (n – 1) degrees of freedom.

Its probability density function is a bell-shaped curve but with heavier tails than those of the normal distribution. In probability and statistics, Student's t-distribution (or t-distribution) is a family of continuous probability distributions that arise when estimating the mean of a normally-distributed population in situations where the sample size is small and population standard deviation is unknown.

It plays a role in several widely used statistical analyses, including the Student's t-test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervals for the difference between two population means, and in linear regression analysis.

The t-distribution is symmetric and bell-shaped, like the normal distribution, but has heavier tails, meaning that it is more prone to producing values that fall far from its mean. This makes

it useful for understanding the statistical behavior of certain types of ratios of random quantities, in which variation in the denominator is amplified and may produce outlying values when the denominator of the ratio falls close to zero.

PDF of Student's T Distribution

$$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}$$

The Greek letter nu (ν) is the number of degrees of freedom and gamma is the "gamma function."

Mean of Student's T Distribution: 0 for degrees of freedom greater than 1, otherwise undefined

Variance of Student's T Distribution

$$\begin{cases} \frac{\nu}{\nu-2} & \nu > 2 \\ \infty & 1 < \nu \leq 2 \\ undefined & otherwise \end{cases}$$

Chi-Square Distribution

Chi-square distribution (also *chi-square* or χ^2*-distribution*) is a continuous probability distribution of the sum of squares of multiple independent standard normal random variables.

If $Y_1, Y_2 \ldots Y_k$ are a group of independent normal distributed variables, each with mean 0 and standard deviation 1, then

$$\chi^2 = \sum_{i=1}^{k} Y_i^2$$

follows a chi-square distribution with *k* degrees of freedom.

Chi-square distribution is one of the most widely used probability distributions in inferential statistics, e.g., in hypothesis testing or in construction of confidence intervals. The chi-squared distribution is used in the common chi-squared tests for goodness of fit of an observed distribution to a theoretical one, the independence of two criteria of classification of qualitative data, and in confidence interval estimation for a population standard deviation of a normal distribution from a sample standard deviation. Many other statistical tests also use this distribution.

PDF of Chi-Square Distribution

$$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}$$

Mean of Chi-Square Distribution

$$k$$

Variance of Chi-Square Distribution

$$2k$$

The Greek letter gamma (Γ) is the gamma function.

F Distribution

F distribution is a continuous probability distribution which arises in the analysis of variance or test of equality between two variances.

If $X^2_{\nu_1}$ and $X^2_{\nu_2}$ are independent variables with chi-square distributions with ν_1 and ν_2 degrees of freedom,

$$F = \frac{\chi^2_{\nu_1} / \nu_1}{\chi^2_{\nu_2} / \nu_2}$$

follows a F distribution.

PDF of F Distribution

$$\frac{\sqrt{\frac{(\nu_1 x)^{\nu_1} \nu_2^{\nu_2}}{(\nu_1 x + \nu_2)^{\nu_1 + \nu_2}}}}{xB\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}$$

Mean of F Distribution

$$\frac{\nu_2}{\nu_2 - 2}$$

Variance of F Distribution

$$\frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$$

The PDF function for an F distribution, where ν_1 and ν_2 represent the degrees of freedom for two chi-squared distributions, and beta represents the beta function.

Use JMP to Fit a Distribution

Case study: We are interested to fit the height data of basketball players into a distribution.

Data File: "OneSampleT-Test.jmp"

Steps to fit a distribution in JMP:

1. Click Analyze -> Distribution

2. Select “HtBk” as “Y, Columns”

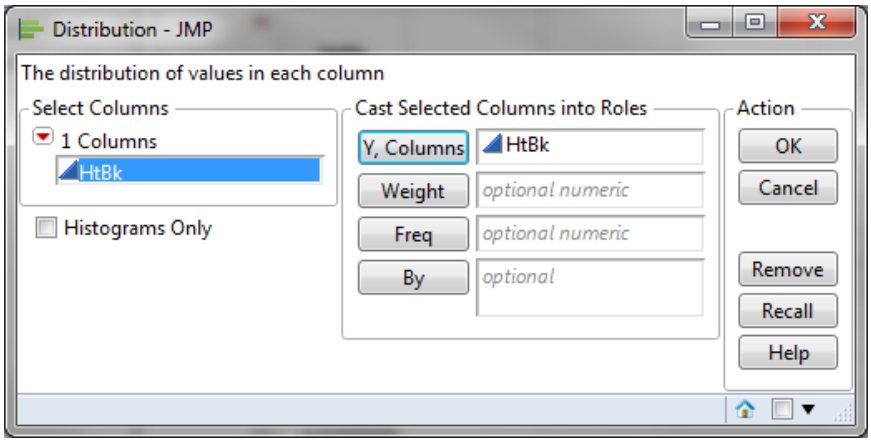
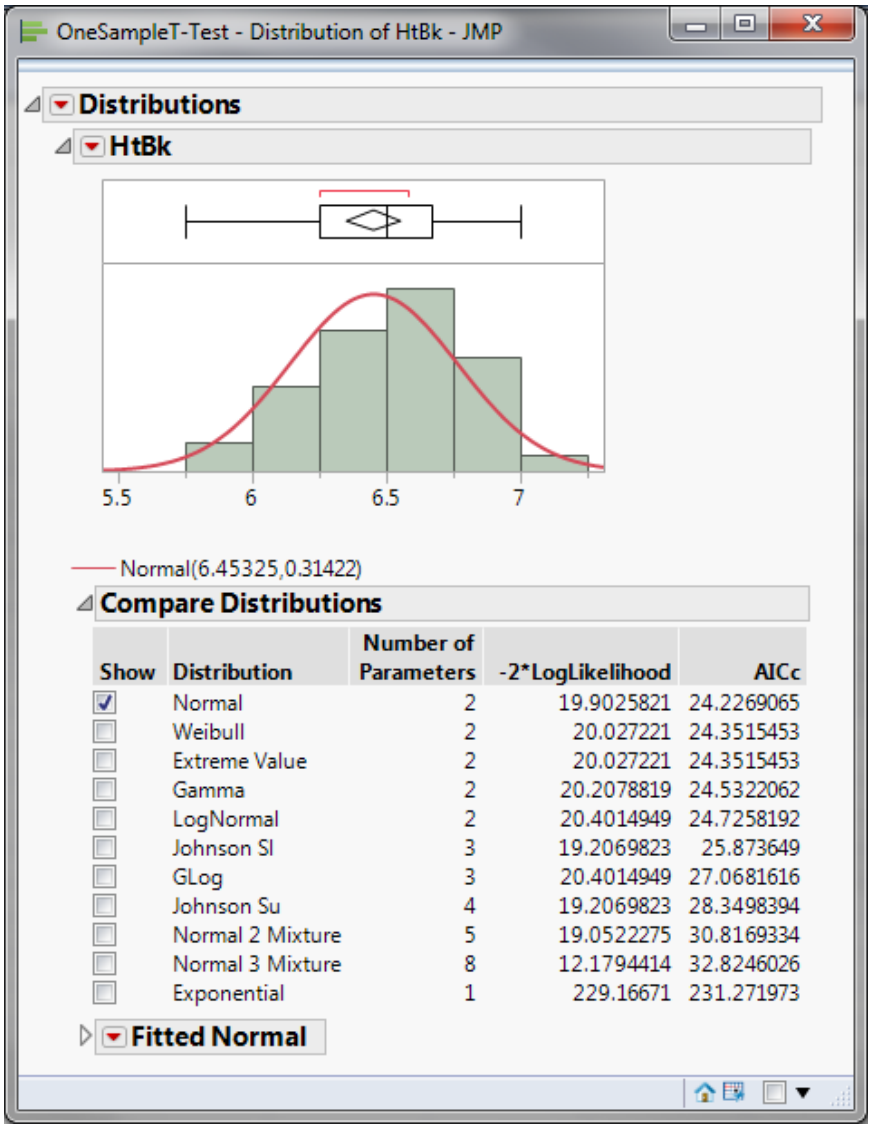


Fig. 3.14 Distribution data selection window

3. Click “OK”
4. Click on the red triangle button next to “HtBk”
5. Click Continuous Fit -> All



Show	Distribution	Number of Parameters	-2*LogLikelihood	AICc
☑	Normal	2	19.9025821	24.2269065
☐	Weibull	2	20.027221	24.3515453
☐	Extreme Value	2	20.027221	24.3515453
☐	Gamma	2	20.2078819	24.5322062
☐	LogNormal	2	20.4014949	24.7258192
☐	Johnson SI	3	19.2069823	25.873649
☐	GLog	3	20.4014949	27.0681616
☐	Johnson Su	4	19.2069823	28.3498394
☐	Normal 2 Mixture	5	19.0522275	30.8169334
☐	Normal 3 Mixture	8	12.1794414	32.8246026
☐	Exponential	1	229.16671	231.271973

Fig. 3.15 Distribution Fit Normal output

The distribution the data fit the best is listed in the first row of the table "Compare Distribution" with a small check-mark and the smallest AIC_C value. In this example, the data fit the normal distribution best.

3.2 INFERENTIAL STATISTICS

3.2.1 UNDERSTANDING INFERENCE

What is Statistical Inference?

Statistical inference is the process of making inferences regarding the characteristics of an unobservable population based on the characteristics of an observed sample. Statistical inference is widely used since it is difficult or sometimes impossible to collect the entire population data. It is rare that we ever know the characteristics of a population, so we need to take limited data and infer what the population looks like.

Outcome of Statistical Inference

The outcome or conclusion of statistical inference is a *statistical proposition* about the population, such as an estimate of the population mean or standard deviation.

Examples of statistical propositions:

- Estimating a population parameter
- Identifying an interval or a region where the true population parameter would fall with some certainty
- Deciding whether to reject a hypothesis made on characteristics of the population of interest
- Making predictions
- Clustering or partitioning data into different groups

Population and Sample

A *statistical population* is an entire set of objects or observations about which statistical inferences are to be drawn based on its sample. It is usually impractical or impossible to obtain the data for the entire population. For example, if we are interested in analyzing the population of all the trees, it is extremely difficult to collect the data for all the trees that existed in the past, exist now, and will exist in the future.

A *sample* is a subset of the population (like a piece of the pie above). It is necessary for samples to be *representative* of the population. The process of selecting a subset of observations within a population is referred to as *sampling*.

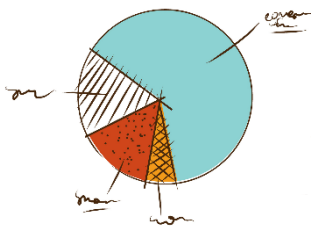


Fig. 3.16 Sampling a statistical population

When we talk about statistical inference, it is because we are trying to make inferences about a population based upon a sample size. Consider the presidential elections. The “population” includes every single person that casts a vote. However, on Election Day you will see the news outlets making projections (inferences) about how people have voted, based upon sampling being done throughout the day. Unless the population is very small, it is typically impractical or impossible to obtain measurements on an entire population.

Population and Sample

Population Parameters (Greek letters)
Mean: μ
Standard deviation: σ

Variance: σ^2
Median: η
Sample Statistics (Roman letters)
Sample Mean: $\bar{X}$
Standard deviation: s
Variance: S^2
Median: $\tilde{X}$
The *population parameter* is the numerical summary of a population. The *sample statistic* is the numerical measurement calculated based on a sample of that population. It is used to estimate the true population parameter. There is a different nomenclature and there are different symbols that we use when referring to a population or a sample.

Descriptive Statistics vs. Statistical Inference

Descriptive Statistics
Descriptive statistics summarize the characteristics of a collection of data. Descriptive statistics are descriptive only and they do not make any generalizations beyond the data at hand. Data used for descriptive statistics are for the purpose of representing or reporting.

Statistical Inference
Statistical inference makes generalizations from a sample at hand of a population. Data used for statistical inference are for the purpose of making inferences on the entire population of interest. A complete statistical analysis includes *both* descriptive statistics and statistical inference.

Error Sources of Statistical Inference

Statistical inference uses sample data to best approximate the true features of the population. A valid sample must be unbiased and representative of the population, meaning that the data should not be selected in a way that only represents a certain segment of the population. For example, what would happen if we wanted to understand income for the population of the US? In our sampling, if we only pulled data for professional athletes, or if we only pulled our sample in Beverly Hills, CA, what would happen to our data?

The two sources of error in statistical inference are:

1. Random Sampling Error
    - Is variation due to observations being selected randomly.
    - Is inherent to the sampling process and beyond one's control
2. Selection Bias
    - Non-random variation due to inadequate design of sampling
    - It can be improved by adjusting the sampling size and sampling strategy

Random sampling error is expected when observations are selected randomly. This error is expected when we are not able to look at an entire population.

Selection bias is when we do not consider the profile of the population in the design of our sampling strategy, meaning the design is inadequate to represent the population.

3.2.2 SAMPLING TECHNIQUES

What is Sampling?

Sampling is the process of selecting objects or observations from a population of interest about which we wish to make a statistical inference. It is extensively used to collect information about a population, which can represent physical or intangible objects.

Fig. 3.17 Sampling process

We know that we usually cannot get data for an entire population. When we want to understand the population, we need to make sure we are getting a good sample that will allow us to make a statistical inference.

Advantages of Sampling

It is usually impractical or impossible to collect the data of an entire population because it is costly and time consuming, but also because of the unavailability of historical records and the dynamic nature of the population.

Advantages of sampling a representative subset of the population:

- Lower cost
- Faster data collection
- Easier to manipulate

Uses of Samples

With valid samples collected, we can draw statistical conclusions about the population of our interest. There are two major uses of samples in making statistical inference:

- *Estimation*: Estimating the population parameters using the sample statistics
- *Hypothesis testing*: Testing a statement about the population characteristics using sample data

This module covers sample size calculation for estimation purposes only. Sample size calculations for hypothesis testing purposes will be covered in the Hypothesis Testing module.

Basic Sampling Steps

1. Determine the population of interest
2. Determine the sampling frame
3. Determine the sampling strategy
4. Calculate the sample size
5. Conduct sampling

Over the next few pages, we will go into some detail about these basic steps to follow when doing sampling.

- What population are we interested in knowing about?
- What are the "guardrails" we put around the sample?
- What is the sampling strategy?
- What is the size of the sample needed?
- Do the sampling.

Population

Population in statistics is the entire set of objects or observations about which we are interested to draw conclusions or generalize based on some representative sample data. The population covers all the items with characteristics we are interested to analyze.

A population can be either physical or intangible.

- Physical: trees, customers, monitors etc.
- Intangible: credit score, pass/fail decisions etc.

A population can be static or dynamic.

- Characteristics of individuals are relatively static over time.
- Items making up the population continue to change or be generated over time. For instance, the world population is constantly changing. On the other hand, the average stock price in 2011 is no longer changing.

Sampling Frame

In the ideal situation, the scope of a population could be defined. For example, Mary is interested to know how the soup she is cooking tastes. The population is simply the pot of soup. However, in some other situations, the population cannot be identified or defined precisely. For example, to collect the information for an opinion poll, we do not have a list of all the people in the world at hand.

A *sampling frame* is a list of items of the population (preferably the entire population; if not, approximate population). For example, the telephone directory would be a sampling frame for opinion poll data collection.

Basic Sampling Strategies

- Simple Random Sampling
- Matched Random Sampling
- Stratified Sampling
- Systematic Sampling
- Proportional to Size Sampling
- Cluster Sampling

Simple Random Sampling

Simple random samples are selected in such a way that each item in the population has an equal chance of being selected. There is no bias involved in the sample selection. Such selection minimizes the variation between the characteristics of samples and the population. It is the basic sampling strategy.

Matched Random Sampling

Matched random samples are samples randomly selected in pairs, each of which has the same attribute. For example, researchers are interested in understanding the weight of twins; researchers are interested in understanding the patients' blood pressure before and after taking some medicine.

Stratified Sampling

A population can be grouped or "stratified" into distinct and independent categories. An individual category can be considered as a sub-population. *Stratified samples* are randomly selected in each category of the population. The categories can be gender, region, income level etc.

Stratified sampling requires advanced knowledge of the population characteristics, for example: A fruit store wants to measure the quality of all their oranges. They decide to use stratified sampling by region to collect sample data. Since about 40% of their oranges are from California, 40% of the sample is selected from the California oranges sub-population.

Systematic Sampling

Systematic samples are selected at regular intervals based on an ordered list where items in the population are arranged according to a certain criterion. Systematic sampling has a random start and then every i^{th} item is selected going forward, for example, we are sampling the every 5^{th} unit produced on the production line.

Cluster Sampling

Cluster sampling is a sampling method in which samples are only selected from certain clusters or groups of the population. It reduces the cost and time spent on the sampling but bears the risk that the selected clusters are biased, for example, selecting samples from the region where researchers are located so that the cost and time spent on travelling is reduced.

Sampling Strategy Decision Factors

When determining the sampling strategy, we need to consider the following factors:

- Cost and time constraints
- Nature of the population of interest

- Availability of advanced knowledge of the population
- Accuracy requirement

So how does one decide what sampling strategy to use? Consider how much time is available to collect the data and what cost restrictions are in place. Consider the nature of the population, and what you know about the population. Are there specific segments you are interested in? Are the samples streaming like in a manufacturing operation? Determine how accurate you need to be—this can compete with how much you must spend on sampling, but usually this will determine how big of a sample you should collect.

Sample Size

The *sample size* is a critical element that can influence the results of statistical inference. The smaller the sample size, the higher the risk that the sample statistic will not reflect the true population parameter. The greater the sample size, the more time and money we will spend on collecting the samples.

Sample Size Factors

There are a few factors to consider when determining the right sample size:

- Is the variable of interest continuous or discrete?
- How large is the population size?
- How much risk do you want to take regarding missing the true population parameters?
- What is the acceptable margin of error you want to detect?
- How much is the variation in the population? The larger the variation in the population, the larger the risk in sample variation.

Sample Size Calculation for Continuous Data

Sample size equation for continuous data

$$n_0 = \left(\frac{Z_{\alpha/2} \times s}{d} \right)^2$$

Where:

- n is the sample size we are trying to determine
- Z is a value that we look up based upon the confidence level. If we want 95% confidence, then alpha is 1 – 0.95 = 0.05 (or 5%). Therefore, we will look up the Z value for when alpha is 0.05, which is 1.96.
- s is the estimated standard deviation for the population
- d is the acceptable margin of error.

Sample Size Calculation for Continuous Data

When the sample size calculated using the formula

$$n_0 = \left(\frac{Z_{\alpha/2} \times s}{d} \right)^2$$

exceeds 5% of the population size, we use this formula to calculate the sample size:

$$n = \frac{n_0}{\left(1 + \frac{n_0}{N} \right)}$$

Where:

n_0 represents the sample size calculated using the formula on the previous page

N is the size of the population.

Sample Size Calculation for Discrete Data
Sample size equation for continuous data

$$n_0 = \left(\frac{Z_{\alpha/2}}{d} \right)^2 \times p \times (1 - p)$$

Where:

- n_0 is the number of samples
- $Z_{\alpha/2}$ is the *Z* score when risk level is *α/2*. When α is 0.05, it is 1.96. When *α* is 0.10, it is 1.65.
- *s* is the estimation of standard deviation in the population
- *d* is the acceptable margin of error
- *p* is the proportion of one type of event occurring (e.g., proportion of passes)
- $p \times (1 - p)$ is the estimate of variance.

Sample Size Calculation for Discrete Data

Sample size calculated using the formula:

$$n_0 = \left(\frac{Z_{\alpha/2}}{d} \right)^2 \times p \times (1 - p)$$

When the sample size exceeds 5% of the population size, we use a correction formula to calculate the final sample size.

$$n = \frac{n_0}{\left(1 + \frac{n_0}{N}\right)}$$

Where:

- n_0 is the sample size calculated using the previous equation
- N is the population size.

Sampling Errors

Random Sampling Error

- Is random variation due to observations being selected randomly
- Is inherent in the sampling process and beyond one's control

Selection Bias

- Is non-random variation due to inadequate design of sampling
- Can be improved by adjusting the sampling size and sampling strategy

3.2.3 SAMPLE SIZE

Sample Size

In the last model, we talked about sample size considerations for when we are looking to estimate parameters for a population. Now we want to use sample size considerations for hypothesis tests. The sample size is a critical element that can influence the results of hypothesis testing. The smaller the sample size, the higher the risk that the statistical conclusions will not reflect the population relationship. The greater the sample size, the more time and money we will spend on collecting the samples. So, we must figure out where the right balance is.

Sample Size Calculation

General sample size formula for *continuous data*:

$$n = \left(\frac{\left(Z_{\alpha/2} + Z_{\beta}\right) \times s}{d} \right)^2$$

General sample size formula for *discrete data*:

$$n = \left(\frac{Z_{\alpha/2} + Z_{\beta}}{d} \right)^2 \times p \times (1 - p)$$

Where:

- n is the number of observations in the sample
- α is the risk of committing a false positive error (or Type I error). A false positive error, commonly called a false alarm, is a result that indicates a given condition has been fulfilled, when it has not been fulfilled. In the case of "crying wolf," the condition tested for was "Is there a wolf near the herd?" The actual result was that there had not been a wolf near the herd. The shepherd wrongly indicated there was one, by calling "Wolf, wolf!"
- β is the risk of committing a false negative error (or Type II error). A false negative error is where a test result indicates that a condition failed, while it was successful. An example is a guilty prisoner freed from jail. The condition, "Is the prisoner guilty?" had a positive result (yes, he is guilty). But the test failed to realize this, and wrongly decided the prisoner was not guilty.
- s is the estimation of standard deviation in the population
- d is the size of effect you want to be able to detect
- p is the proportion of one type of event occurring (e.g., proportion of passes)

Use JMP to Calculate the Sample Size

Case study: We are interested in comparing the average retail price of a product between two states. We will run a hypothesis test on the two sample means to determine whether there is a statistically significant difference between the retail prices in the two states. The average retail price of the product is $23 based on our estimation and the standard deviation is 3. We want to detect at least a $2 difference with a 90% chance when it is true and we can tolerate the alpha risk at 5%. What should the sample size be? To use the software to calculate the sample size, consider this example where we are trying to determine if there is a statistical difference between two sample means.

Steps to calculate the sample size in JMP:

1. Click DOE -> Design Diagnostics->Sample Size and Power
2. Click "Two Sample Means"

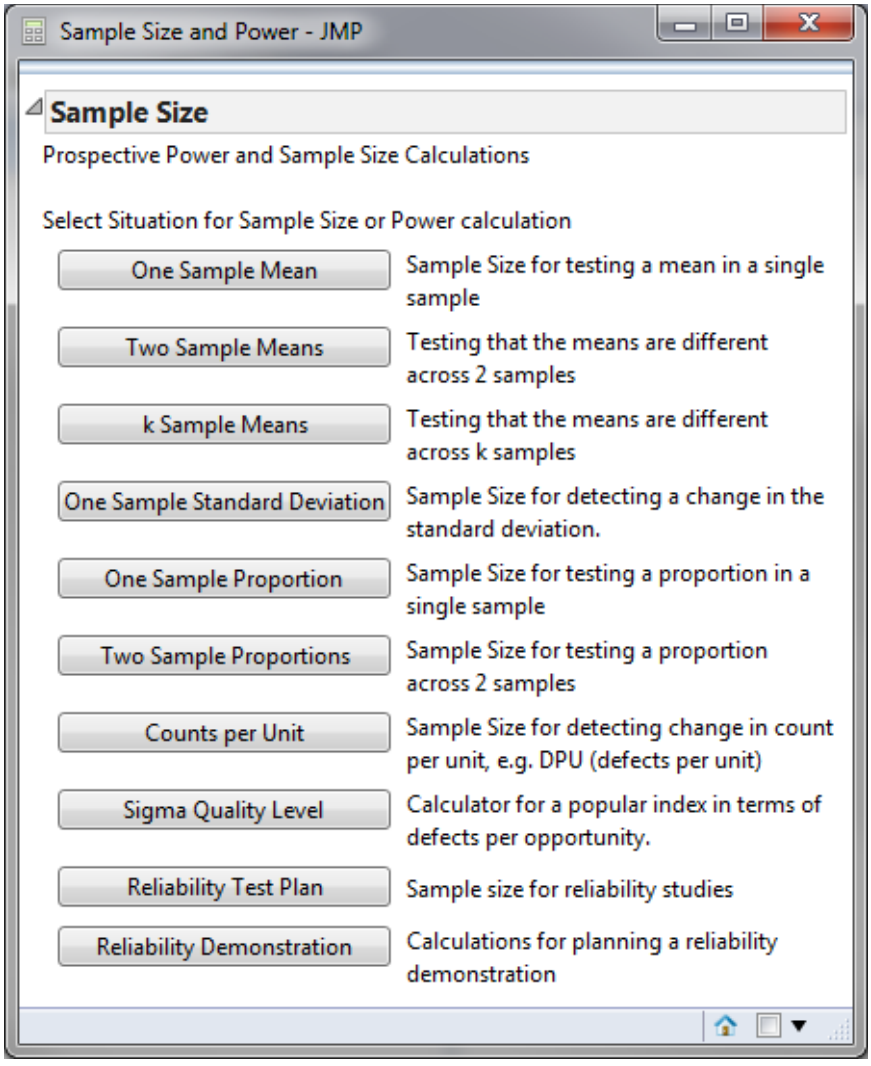


Fig. 3.18 Sample Size selection box

3. Enter the alpha level: 0.05
4. Enter the standard deviation value: 3
5. Enter the difference to detect: 2
6. Enter the power: 0.90

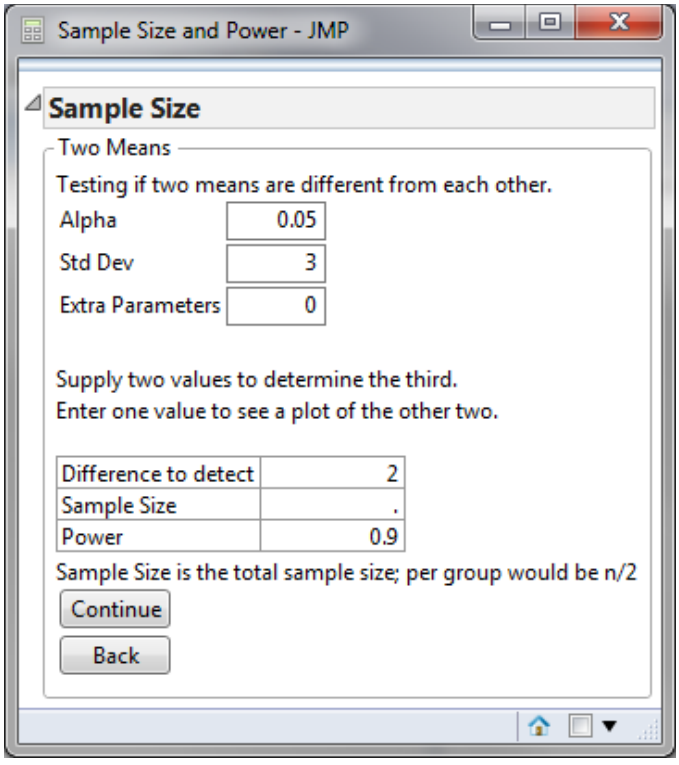


Fig. 3.19 Power and Sample Size for 2 Sample T with variables

7. Click "Continue"
8. The sample size result appears in the box next to "Sample Size"

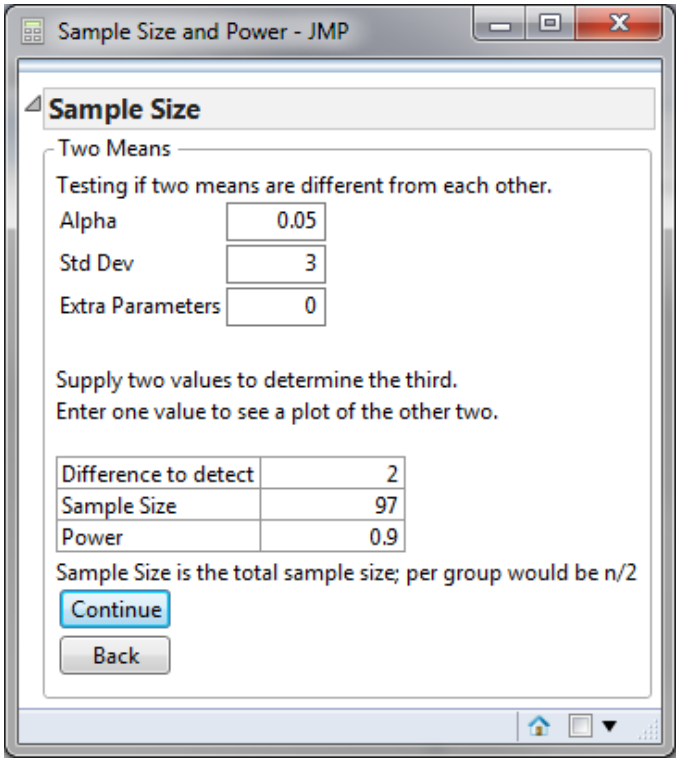


Fig. 3.20 Sample Size Results

If the value of the difference to detect is not entered, JMP will return a chart with X axis being the difference to detect and Y axis being the sample size. When the difference to detect decreases, the required sample size increases.

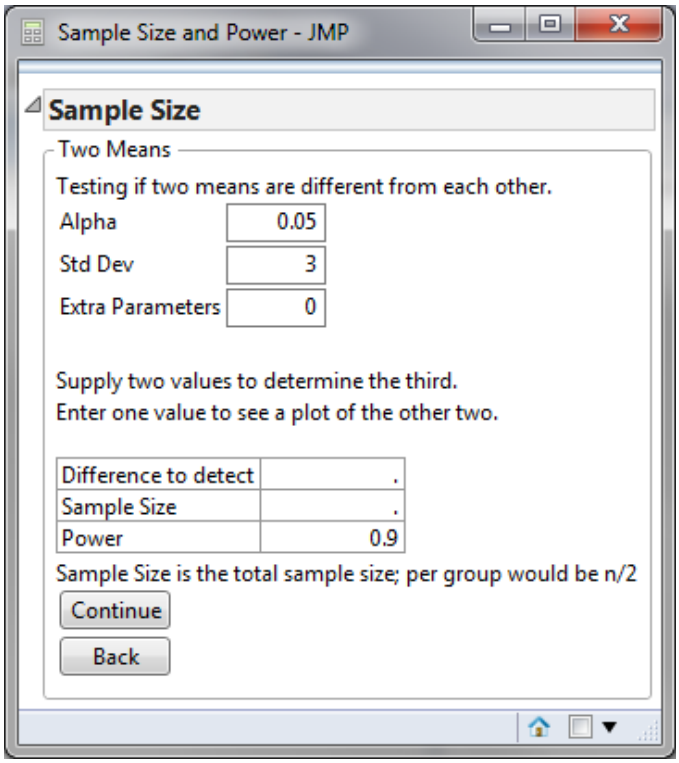


Fig. 3.21 Power and Sample Size for 2-Sample t with variables

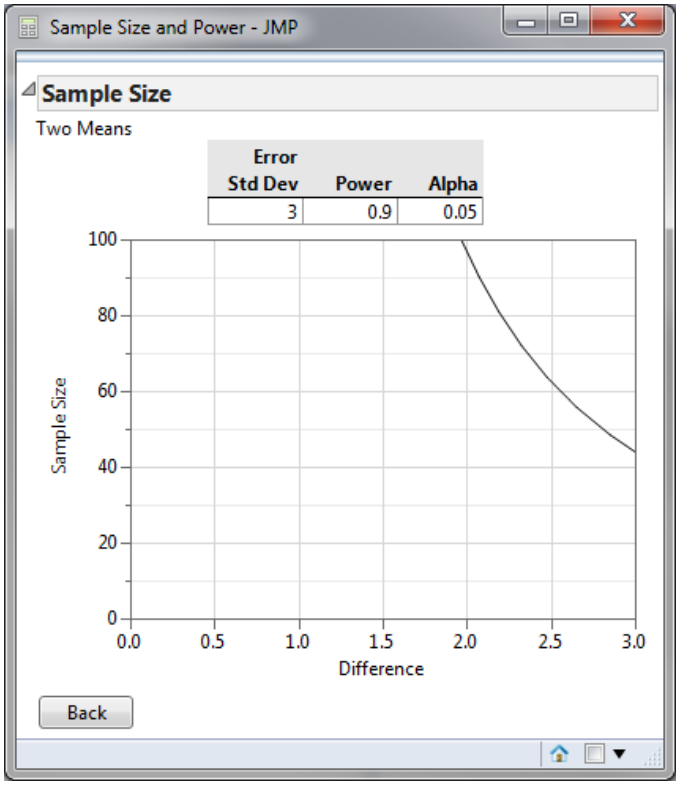


Fig. 3.22 Relationship between the difference to detect and the sample size needed

In this example…when the "difference to detect'" is not entered, the chart above is given as output. It shows graphically the relationship between the difference to detect and the sample size needed.

3.2.4 CENTRAL LIMIT THEOREM

What is Central Limit Theorem?

The *Central Limit Theorem* is one of the fundamental theorems of probability theory. It states a condition under which the mean of a large number of independent and identically-distributed random variables, each of which has a finite mean and variance, would be approximately normally distributed. Let us assume $Y_1, Y_2 \ldots Y_n$ is a sequence of n i.i.d. random variables, each of which has finite mean μ and variance σ^2, where $\sigma^2 > 0$. When n increases, the sample average of the n random variables is approximately normally distributed, with the mean equal to μ and variance equal to σ^2/n, regardless of the common distribution Y_i follows where $i = 1, 2 \ldots n$.

Independent and Identically Distributed

A sequence of random variables is *independent and identically distributed* (i.i.d.) if each random variable is independent of others and has the same probability distribution as others. It is one of the basic assumptions in Central Limit Theorem. Consider the law of large numbers (LLN)—It is a theorem that describes the result of performing the same experiment a large number of times. According to the LLN, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed. The following example will explain this further.

Central Limit Theorem Example

Let us assume we have 10 fair die at hand. Each time we roll all 10 die together we record the average of the 10 die. We repeat rolling the die 50 times until we will have 50 data points. Upon doing so, we will discover that the probability distribution of the sample average approximates the normal distribution even though a single roll of a fair die follows a discrete uniform distribution. Knowing that each die has six possible values (1, 2, 3, 4, 5, 6), when we record the average of the 10 dice over time, we would expect the number to start approximating 3.5 (the average of all possible values). The more rolls we perform, the closer the distribution would be to a normal distribution centered on a mean of 3.5.

Central Limit Theorem Explained in Formulas

Let us assume $Y_1, Y_2 \ldots Y_n$ are i.i.d. random variables with

$$E(Y_i) = \mu_Y \quad where -\infty < \mu_Y < \infty$$

$$var(Y_i) = \sigma_Y^2 \quad where\ 0 < \sigma_Y^2 < \infty$$

As $n \to \infty$, the distribution of $\overline{Y}$ becomes approximately normally distributed with

$$E\left(\overline{Y}\right) = \mu_Y$$

$$\text{var}(\overline{Y}) = \sigma_{\overline{Y}}^2 = \frac{\sigma_Y^2}{n}$$

Where:

- Y_1, Y_2, Y_n are independent and identically distributed random variables
- $E(Y_i)$ = the expected value for Y is the mean for Y
- $Var(Y_i)$ = the expected variance for Y.

As the *n* gets infinitely large, the distribution of Y means becomes approximately a normal distribution.

Central Limit Theorem Application

Use the sample mean to estimate the population mean if the assumptions of Central Limit Theorem are met:

$$E(Y_i) = E(\bar{y})$$

Where: $i = 1, 2 \ldots n$.

Use standard error of the mean to measure the standard deviation of the sample mean estimate of a population mean.

$$SE_{\overline{Y}} = \frac{s}{\sqrt{n}}$$

Where: *s* is the standard deviation of the sample and *n* is the sample size.

Standard deviation of the population mean:

$$SD_{\overline{Y}} = \frac{\sigma}{\sqrt{n}}$$

Where:

- σ is the standard deviation of the population
- *n* is the sample size.

Use a larger sample size, if economically feasible, to decrease the variance of the sampling distribution. The larger the sample size, the more precise the estimation of the population parameter. Use a confidence interval to describe the region which the population parameter

would fall in. The sample distribution approximates the normal distribution in which 95% of the data stays within two standard deviations from the center. Population mean would fall in the interval of two standard errors of the mean away from the sample mean, 95% of the time.

Confidence Interval

The *confidence interval* is an interval where the true population parameter would fall within a certain confidence level. A 95% confidence interval, the most commonly used confidence level, indicates that the population parameter would fall in that region 95% of the time or we are 95% confident that the population parameter would fall in that region. The confidence interval is used to describe the reliability of a statistical estimate of a population parameter.

The width of a confidence interval depends on the:

- Confidence level—The higher the confidence level, the wider the confidence interval
- Sample size—The smaller the sample size, the wider the confidence interval
- Variability in the data—The more variability, the wider the confidence interval

Confidence Interval of the Mean

Confidence interval of the population mean μ_Y of a continuous variable Y is:

$$\left[\overline{Y} - Z_{\alpha/2} \times \left(\frac{\sigma}{\sqrt{n}}\right), \overline{Y} + Z_{\alpha/2} \times \left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

Where:

- $\overline{Y}$ is the sample mean
- σ is the standard deviation of the population
- *n* is the sample size
- $Z_{\alpha/2}$ is the Z score when risk level is *α/2*.
  When *α* is 0.05, it is 1.96.
  When *α* is 0.10, it is 1.65.
- *α* is (1 – confidence level). When confidence level is 95%, *α* is 5%.
  When the confidence level is 90%, *α* is 10%.

JMP: Calculate the Confidence Interval of the Mean

Data File: "CentralLimitTheorem.jmp"

Steps to calculate the confidence interval of the Mean:

1. Click Analyze -> Distribution
2. Select "Cycle Time (Minutes)" as the "Y, Columns"

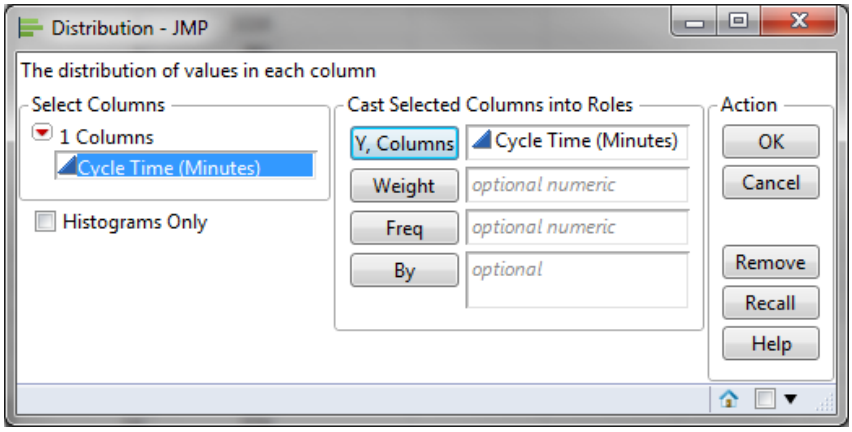


Fig. 3.23 Summary dialog box with variable

3. Click "OK"
4. "Upper 95% Mean" and "Lower 95% Mean" at the bottom of the newly generated window are the upper and lower boundaries of 95% confidence interval.

Summary Statistics

Mean	703.16667
Std Dev	1180.9511
Std Err Mean	215.61119
Upper 95% Mean	1144.1411
Lower 95% Mean	262.19226
N	30

Fig. 3.24 Summary results

In JMP, the confidence level is 95% by default. In order to see the confidence interval of "Cycle Time (Minutes)" at other confidence level, we need to:

1. Click on the red triangle button next to "Cycle Time (Minutes)"
2. Select Confidence Interval -> the confidence level of interest (e.g. 90%, 95%, 99% etc.)
3. The confidence interval at the selected confidence level appears at the bottom of the distribution analysis page

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	703.1667	336.8159	1069.517	0.900
Std Dev	1180.951	974.8675	1511.269	0.900

Fig. 3.25 Summary output with new confidence interval

3.3 HYPOTHESIS TESTING

3.3.1 GOALS OF HYPOTHESIS TESTING

What is Hypothesis Testing?

A *hypothesis test* is a statistical method in which a specific hypothesis is formulated about a population, and the decision of whether to reject the hypothesis is made based on sample data. Hypothesis tests help to determine whether a hypothesis about a population or multiple populations is true with certain confidence level based on sample data. Hypothesis testing is a critical tool in the Six Sigma tool belt. It helps us separate fact from fiction, and special cause from noise, when we are looking to make decisions based on data.

Hypothesis Testing Examples

Hypothesis testing tries to answer whether there is a difference between different groups or there is some change occurring.

- Are the average SAT scores of graduates from high school A and B the same?
- Is the error rate of one group of operators higher than that of another group?
- Are there any non-random causes influencing the height of kids in one state?

What is Statistical Hypothesis?

A *statistical hypothesis* is an assumption about one or multiple population. It is a statement about whether there is any difference between different groups. It can be a conjecture about the population parameters or the nature of the population distributions.

A statistical hypothesis is formulated in pairs:

- Null Hypothesis
- Alternative Hypothesis

Statements like the following reflect a null hypothesis:

- The data are normally distributed.
- The population mean is the same as X.
- Group A and Group B are the same.

And the alternative hypotheses are:

- The data are not normally distributed
- The population mean is not the same as X.
- Group A and Group B are not the same.

Null and Alternative Hypotheses

Null Hypothesis (H_0) states that:

- There is no difference in the measurement of different groups
- No changes occurred

- Sample observations result from random chance

Alternative Hypothesis (H_1 or H_a) states that:

- There is a difference in the measurement of different groups
- Some changes occurred
- Sample observations are affected by non-random causes

A statistical hypothesis can be expressed in mathematical language by using population parameters (Greek letters) and mathematical symbols.

- Population Parameters (Greek letters)
- Mean: μ
- Standard deviation: σ
- Variance: σ^2
- Median: η
- Mathematical Symbols
- Equal: =
- Not equal: ≠
- Greater than: >
- Smaller than: <

The following are examples of null and alternative hypotheses written in mathematical language.

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$$\begin{cases} H_0: \sigma_1 = 0 \\ H_1: \sigma_1 \neq 0 \end{cases}$$

$$\begin{cases} H_0: \eta_1 = 10 \\ H_1: \eta_1 > 10 \end{cases}$$

$$\begin{cases} H_0: \mu_1 = 10 \\ H_1: \mu_1 < 10 \end{cases}$$

The first pair reads: Null hypothesis where population mean 1 is equal to population mean 2. Alternative hypothesis is population mean 1 is NOT equal to population mean 2. Or in the third pair: Null hypothesis where population mean 1 = 10. Alternative hypothesis is population mean greater than 10.

Hypothesis Testing Conclusion

There are two possible conclusions of hypothesis testing:

1. Reject the null

2. Fail to reject the null

When there is enough evidence based on the sample information to prove the alternative hypothesis, we reject the null. When there is *not* enough evidence or the sample information is *not* sufficiently persuasive, we fail to reject the null.

Decision Rules in Hypothesis Testing

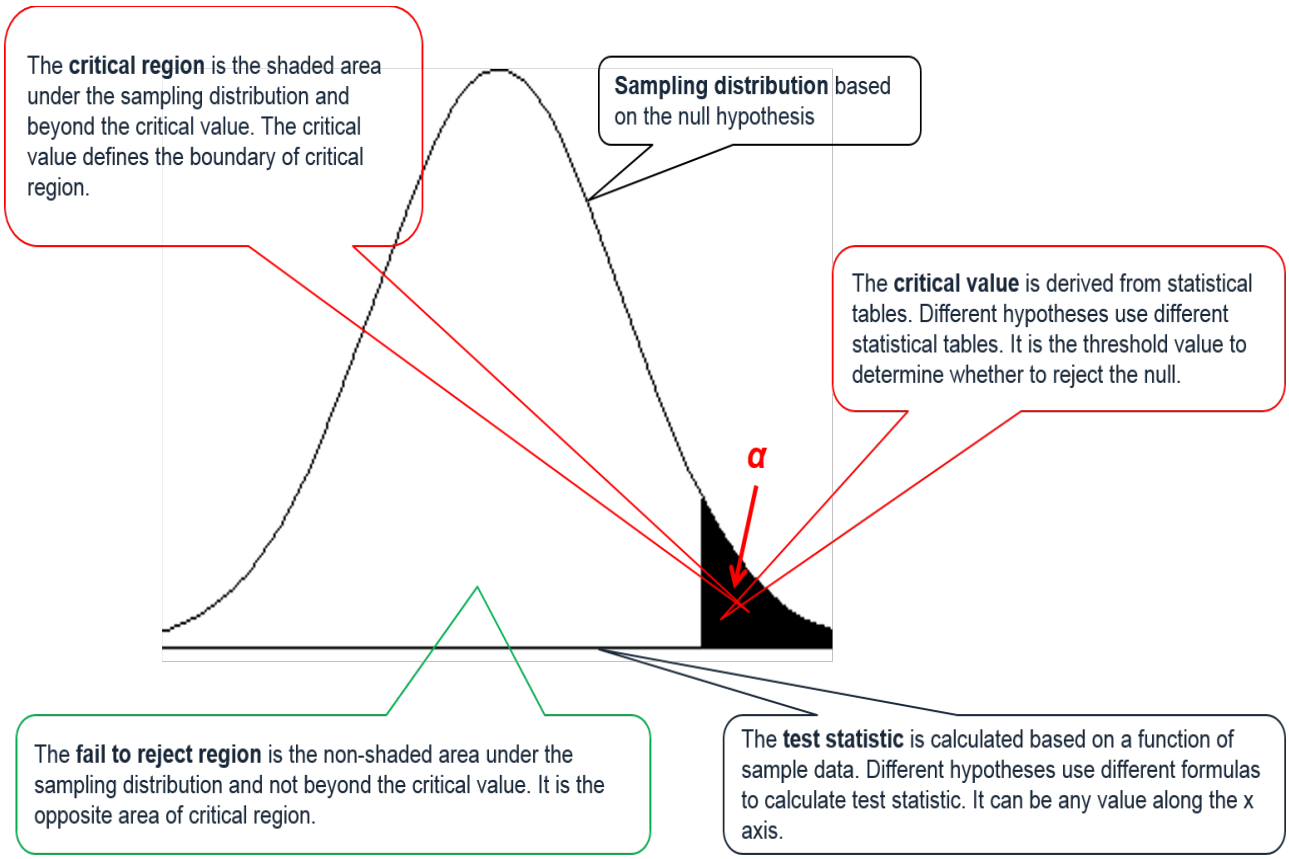


Fig. 3.26 Rules in Hypothesis Testing

As we get into the technicalities of hypothesis testing, it is important to understand some terms as they pertain to a distribution. Consider in this picture a normal distribution, and what we know about the distribution. Remember 68–95–99.7? The amount of data that falls within +/−1, 2, and 3 standard deviations of the mean.

The *test statistic* in hypothesis testing is a value calculated using a function of the sample. Test statistics are considered the sample data's numerical summary that can be used in hypothesis testing. Different hypothesis tests have different formulas to calculate the test statistic.

The *critical value* in hypothesis testing is a threshold value to which the test statistic is compared in order to determine whether the null hypothesis is rejected. The critical value is obtained from statistical tables. Different hypothesis tests need different statistical tables for critical values.

Hypothesis testing is made easy with our fancy software packages, but it is important that you understand how the theory and rules work behind the software.

- When the test statistic falls into the fail to reject region, we fail to reject the null and claim that there is no statistically significant difference between the groups.
- When the test statistic falls into the critical region, we reject the null and claim that there is a statistically significant difference between the groups.

The proportion of the area under the sampling distribution and beyond the critical value indicates α risk (also called *α level*). The most commonly selected α level is 5%. To illustrate, let us assume the test statistic is equal the critical value of 5%. In this case, there is a 5% risk that we will reject the null hypothesis when it is true. So, as the test statistic falls in the critical region, the risk of rejecting the null when it is actually true is less than 5%. Another way to state it is that we are 95% confident that we should reject the null hypothesis.

The proportion of the area under the sampling distribution and beyond the test statistic is the *p-value*. It is the probability of getting a test statistic at least as extreme as the observed one, given the null is true.

- When the p-value is smaller than the α level, we reject the null and claim that there is a statistically significant difference between different groups.
- When the p-value is higher than the α level, we fail to reject the null and claim that there is no statistically significant difference between different groups.

Steps in Hypothesis Testing

Step 1: State the null and alternative hypothesis.
Step 2: Determine α level.
Step 3: Collect sample data.
Step 4: Select a proper hypothesis test.
Step 5: Run the hypothesis test.
Step 6: Determine whether to reject the null.

3.3.2 STATISTICAL SIGNIFICANCE

Statistical Significance

In statistics, an observed difference is *statistically significant* if it is unlikely that the difference occurred by pure chance, given a predetermined probability threshold. Statistical significance indicates that there are some non-random factors causing the result to take place. The statistical significance level in hypothesis testing indicates the amount of evidence which is sufficiently persuasive to prove that a difference between groups exists not due to random chance alone.

When we talk about determining if something is statistically significant, we will use the p-value. The p-value is the probability that the difference between two groups is driven by chance. The lower the p-value, the more likely that there is some non-random factor that is causing the difference.

Practical Significance

An observed difference is *practically significant* when it is large enough to make a practical difference. A difference between groups that is *statistically significant* might not be large enough to be practically significant. Sometimes we determine with statistics that there are differences. However, while those differences might be statistically significant, they may not be large enough to make a practical difference from a business standpoint. When analysis determines that a difference is statistically significant, ask yourself the question, "Why does it matter?" or "Why should we care?"

Example: You started to use the premium gas recently, which was supposed to make your car run better. After running a controlled experiment to measure the performance of the car before and after using the premium gas, you performed a statistical hypothesis test and found that the difference before and after was statistically significant. Using premium gas did improve the performance. However, due to the high cost of the premium gas, you decided that the difference was not large enough to make you pay extra money for it. In other words, the difference is not practically significant.

In this example, the premium gas did make a difference in performance. However, you must ask what is important. If cost is important, does the increased cost of premium gas have long term savings benefit—better gas mileage, lower maintenance, and repair cost, longer engine life?

3.3.3 Risk; Alpha and Beta

Errors in Hypothesis Testing

In statistical hypothesis testing, there are two types of errors:

- *Type I Error*—A null hypothesis is rejected when it is true in fact
- *Type II Error*—A null hypothesis is not rejected when it is not true in fact

	Null hypothesis is true	Alternative hypothesis is true
Fail to reject null hypothesis	Correct	Incorrect (Type II Error)
Reject null hypothesis	Incorrect (Type I Error)	Correct

Fig. 3.27 Types of Errors in Hypothesis Testing

Type I error is when we react as though there is a change or a difference, when in fact it is only chance that is driving the difference.

Type II error is when we do not react when we should, meaning there is a difference but we act as though there is not.

Type I Error

Type I error is also called false positive, false alarm, or alpha (*α)* error. Type I error is associated with the risk of accepting false positives. It occurs when we think there is a difference between groups but in fact there is none, for example, telling a patient he is sick and in fact he is not. Type I error leads to "tampering," meaning we make changes or adjustments when we should not. We create a moving target and make it even more difficult to control the process.

Alpha (α)

Alpha (α) indicates the probability of making a type I error. It ranges from 0 to 1. *α* risk is the risk of making a type I error.

The most commonly used *α is 5%.* It corresponds to the fact that most commonly we use a 95% confidence level. The confidence level used to calculate the confidence intervals is (*100% – α*).

When making a decision on whether to reject the null, we compare the p-value against *α:*

- If p-value is smaller than *α*, we reject the null
- If p-value is greater than *α*, we fail to reject the null

To reduce the *α* risk, we decrease the *α* value to which the p-value is compared. When we make a decision about statistical significance, we compare the p-value against the alpha value. So, if our confidence level is 95%, our alpha is 5%; therefore, our p-value is 0.05. If p-value is less than alpha, then we reject the null hypothesis. To reduce risk, or increase confidence, we lower the alpha value that the p-value is compared to. To be 99% confident, our alpha would be (1 – 0.99) = 0.01.

Type II Error

Type II error is also called false negative, oversight, or beta (β) error. Type II error is associated with the risk of accepting false negatives. It occurs when we think there is not any difference between groups but in fact there is, for example, telling a patient he is not sick and in fact he is. Type II error describes when we are unresponsive to situations that we should respond to.

Beta (β)

β indicates the probability of making a type II error. It ranges from 0 to 1. β risk is the risk of making a type II error. The most commonly used β is *10%.* It corresponds to the fact that most commonly we use a 90% power. The expression (*100%* – β) is called *power,* which denotes the probability of detecting a difference between groups when in fact the difference truly exists. To

reduce the β risk, we increase the sample size. When holding other factors constant, β is inversely related to α.

3.3.4 TYPES OF HYPOTHESIS TESTS

Two Types of Hypothesis Tests

When we talk about these two type of hypothesis tests, two-tailed and one-tailed, they are a direct reference to the tails of a data distribution.

Two-tailed hypothesis test

- Is also called two-sided hypothesis test
- Is a statistical hypothesis test in which the critical region is split into two equal areas, each of which stays on one side of the sampling distribution

One-tailed hypothesis test

- It is also called one-sided hypothesis test
- It is a statistical hypothesis test in which the critical region is only on one side of the sampling distribution

Two-Tailed Hypothesis Test

A two-tailed hypothesis test is used when we care about whether there is a difference between groups and we do not care about the direction of the difference.

Examples of two-tailed hypothesis tests:

$$\begin{cases} H_0: \mu_1 = 10 \\ H_a: \mu_1 \neq 10 \end{cases}$$

The null hypothesis (H_0) is rejected when:

- The test statistic falls into either half of the critical region
- The test statistic is either sufficiently small or sufficiently large
- The absolute value of the test statistic is greater than the absolute value of the critical value

A two-tailed hypothesis test is the most commonly used hypothesis test. In the next modules, we will cover more details about it. When we are comparing group A and group B, and we only care to know if they are different, vs. one specific group being greater than another, we use a two-tailed test.

Two-Tailed Hypothesis Test

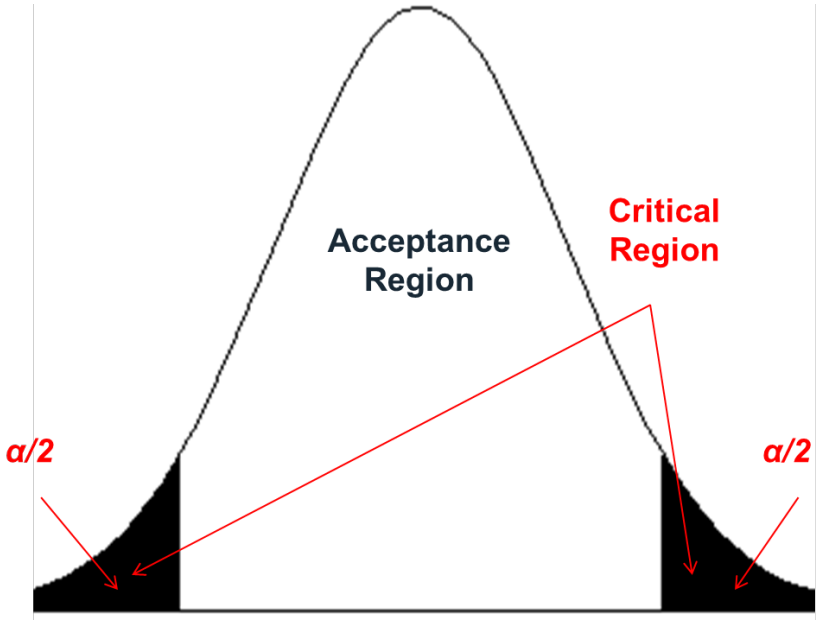


Fig. 3.28 Two-Tailed Hypothesis Test

As described on the previous page, the critical regions are on both (two) tails of the sample distribution.

One-Tailed Hypothesis Test

A one-tailed hypothesis test is used when we care about one direction of the difference between groups. Examples of one-tailed hypothesis tests:

$$\begin{cases} H_0: \mu_1 = 10 \\ H_a: \mu_1 > 10 \end{cases}$$

The null hypothesis is rejected when the test statistic:

- Falls into the critical region which only exits on the right side of the distribution
- Is sufficiently large
- Is greater than the critical value

We use a one-tailed hypothesis test when we only care about one direction when comparing two groups, for instance, a null hypothesis where a population mean is equal to 10. An alternative hypothesis is where the population mean is not just unequal to 10, but greater than 10.

Of course, we look for the test statistic to fall into the critical region to reject the hypothesis. But in this situation, we are only looking for the test statistic to be sufficiently large—greater than the critical value.

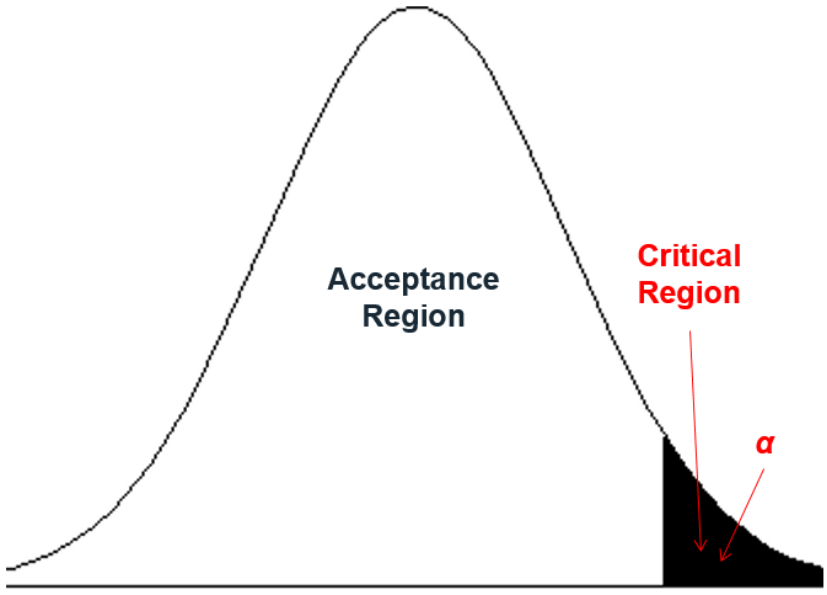


Fig. 3.29 One-Tailed Hypothesis Test

The critical region is only on one side of the distribution. The following are examples of one-tailed hypothesis test.

$$\begin{cases} H_0: \mu_1 = 10 \\ H_a: \mu_1 < 10 \end{cases}$$

The null hypothesis is rejected when the test statistic:

- Falls into the critical region which only exits on the left side of the distribution
- Is sufficiently small
- Is smaller than the critical value

This is the other side of the one-tailed hypothesis test, where the alternative hypothesis is that the population mean is less than 10.

We look for the test statistic to fall into the critical region to reject the hypothesis. But in this situation, we are only looking for the test statistic to be sufficiently small—less than the critical value.

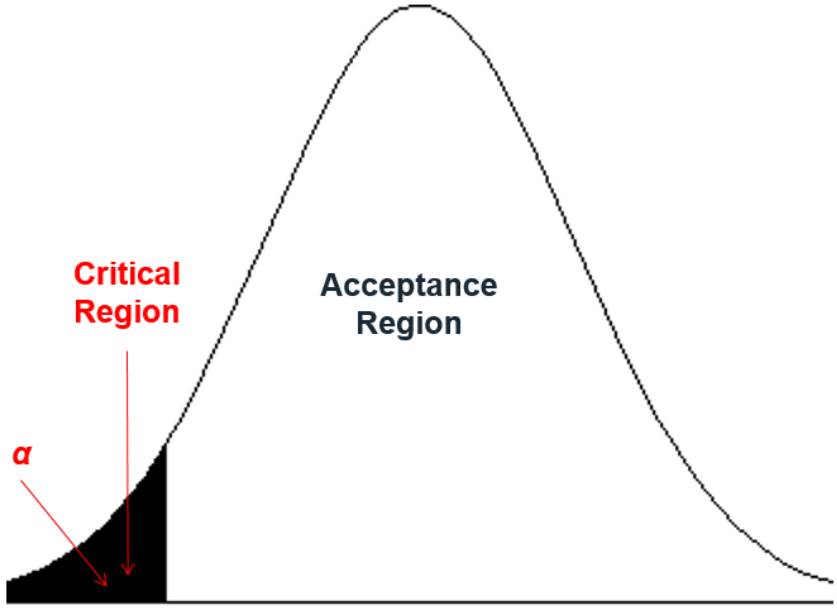


Fig. 3.30 One-Tailed Hypothesis Test

The critical region is on the other side of the distribution.

3.4 HYPOTHESIS TESTS: NORMAL DATA

3.4.1 ONE AND TWO SAMPLE T-TESTS

What is a T-Test?

In statistics, a *t-test* is a hypothesis test in which the test statistic follows a *Student's t* distribution if the null hypothesis is true. We apply a t-test when the population variance (σ) is unknown and we use the sample standard deviation (*s*) instead.

The *t-statistic* was introduced in 1908 by William Sealy Gosset, a chemist working for the Guinness brewery in Dublin, Ireland ("Student" was his pen name). Gosset had been hired due to Claude Guinness's policy of recruiting the best graduates from Oxford and Cambridge to apply biochemistry and statistics to Guinness's industrial processes. Gosset devised the t-test as a cheap way to monitor the quality of stout. He published the test in Biometrika in 1908, but was forced to use a pen name by his employer, who regarded the fact that they were using statistics as a trade secret. In fact, Gosset's identity was known to fellow statisticians (http://en.wikipedia.org/wiki/Student's_t-test).

What is One Sample T-Test?

One sample t-test is a hypothesis test to study whether there is a statistically significant difference between a population mean and a specified value.

- Null Hypothesis (H_0): $\mu = \mu_0$

- Alternative Hypothesis (H_a): $\mu \neq \mu_0$

Where:

- μ is the mean of a population of our interest
- μ_0 is the specific value we want to compare against.

Assumptions of One Sample T-Test

- The sample data of the population of interest are unbiased and representative.
- The data of the population are continuous.
- The data of the population are normally distributed.
- The variance of the population of our interest is unknown.
- One sample t-test is more robust than the z-test when the sample size is small (< 30).

Normality Test

To check whether the population of our interest is normally distributed, we need to run normality test. While there are many normality tests available, such as Anderson–Darling, Sharpiro–Wilk, and Jarque–Bera, our examples will default to using the Anderson-Darling test for normality.

- Null Hypothesis (H_0): The data are normally distributed.
- Alternative Hypothesis (H_a): The data are not normally distributed.

Test Statistic and Critical Value of One Sample T-Test

To understand what is happening when you run a t-test with your software, the formulas here will walk you through the key calculations and how to determine if the null hypothesis should be accepted or rejected. To determine significance, you must calculate the t-statistic and compare it to the critical value, which is a reference value based on the alpha value and degrees of freedom (n – 1). The t-statistic is calculated based on the sample mean, the sample standard deviation, and the sample size.

Test statistic is calculated with the formula:

$$t_{calc} = \frac{\overline{Y}}{s/\sqrt{n}}$$

Where:

$\overline{Y}$ is the sample mean, n is the sample size, and s is the sample standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n-1}}$$

Critical value

- t_{crit} is the t-value in a Student's t distribution with the predetermined significance level α and degrees of freedom (n –1).
- t_{crit} values for a two-sided and a one-sided hypothesis test with the same significance level α and degrees of freedom (n – 1) are different.

Decision Rules of One Sample T-Test

Based on the sample data, we calculated the test statistic t_{calc}, which is compared against t_{crit} to make a decision of whether to reject the null.

- Null Hypothesis (H_0): $\mu = \mu_0$
- Alternative Hypothesis (H_a): $\mu \neq \mu_0$
- If $|t_{calc}| > t_{crit}$, we reject the null and claim there is a statistically significant difference between the population mean μ and the specified value μ_0.
- If $|t_{calc}| < t_{crit}$, we fail to reject the null and claim there is not any statistically significant difference between the population mean μ and the specified value μ_0.

Use JMP to Run a One-Sample T-Test

Case study: We want to compare the average height of basketball players against 7 feet.

Data File: "OneSampleT-Test.jmp"

- Null Hypothesis (H_0): $\mu = 7$
- Alternative Hypothesis (H_a): $\mu \neq 7$

Step 1: Test whether the data are normally distributed

1. Click Analyze -> Distribution
2. Select "HtBk" as "Y, Columns"

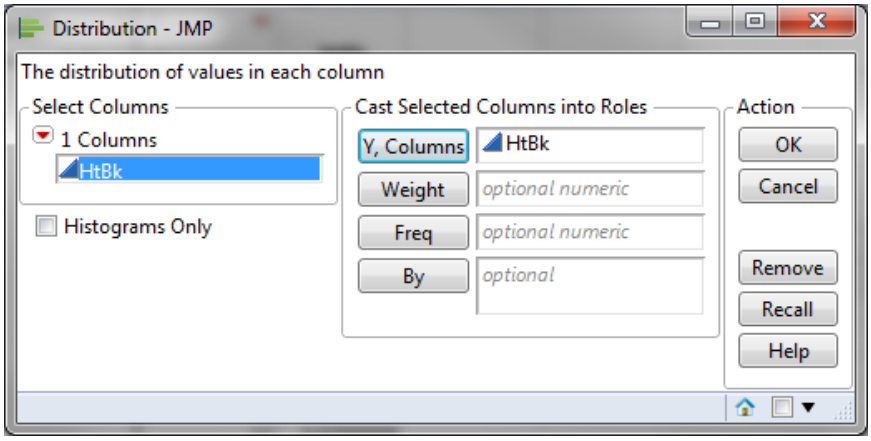


Fig. 3.31 Distribution selection box

3. Click "OK"
4. Click on the red triangle button next to "HtBk" in the Distribution page
5. Click Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

- Null Hypothesis (H_0): The data are normally distributed.
- Alternative Hypothesis (H_a): The data are not normally distributed.

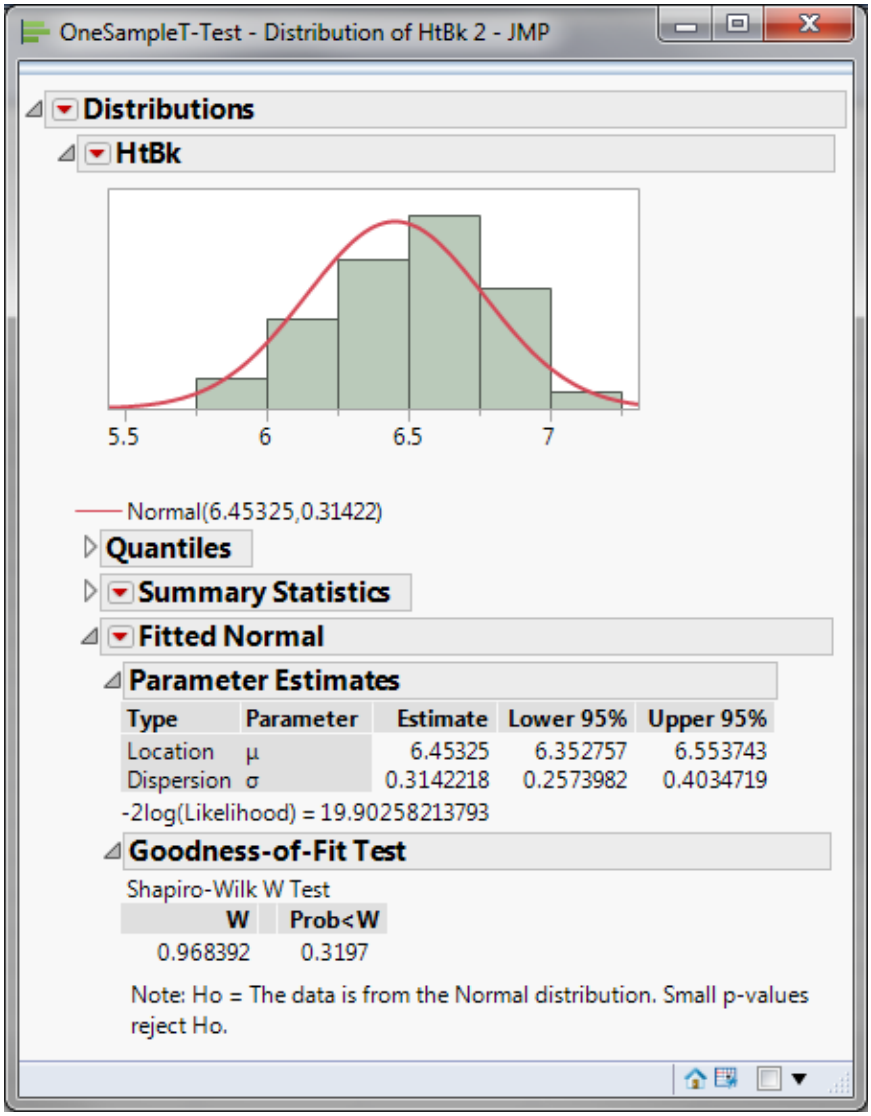


Fig. 3.32 Normality Test output

Since the p-value of the normality is 0.3197 and greater than the alpha level (0.05), we fail to reject the null and claim that the data are normally distributed. If the data are not normally distributed, you need to use hypothesis tests other than the one sample t-test.

Now we can run the one-sample t-test, knowing the data are normally distributed.

Step 2: Run the one-sample t-test

1. Click on the red triangle button next to "HtBk"
2. Select "Test Mean"
3. A new window named "Test Mean" pops up
4. Enter the specific value we want to compare against in the box next to "Specify Hypothesized Mean"

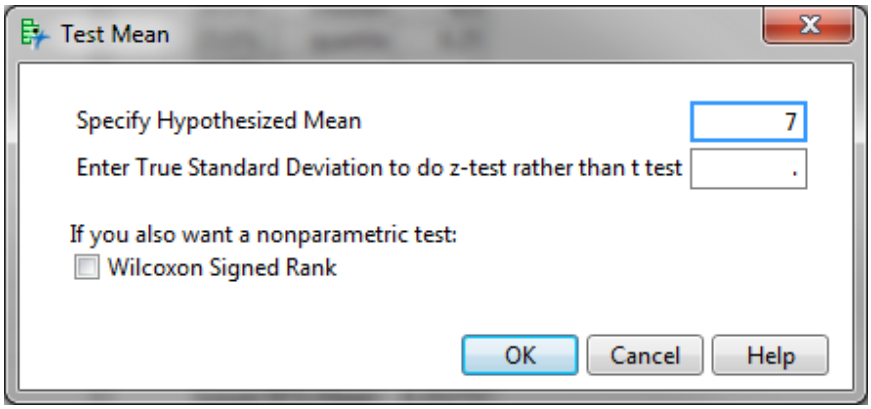


Fig. 3.33 Hypothesized value for One-Sample t

5. Click "OK"

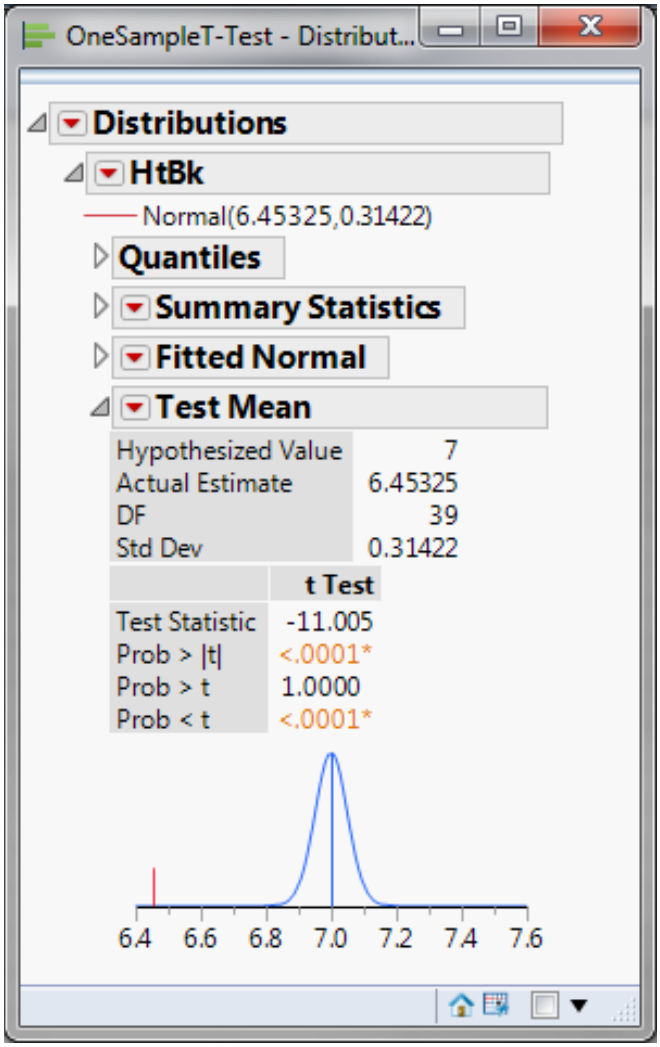


Fig. 3.34 One-Sample t for the Mean output

- Null Hypothesis (H_0): $\mu = 7$
- Alternative Hypothesis (H_a): $\mu \neq 7$

Since the p-value is smaller than alpha level (0.05), we reject the null hypothesis and claim that average of basketball players is statistically different from 7 feet.

What is Two Sample T-Test?

Two sample t-test is a hypothesis test to study whether there is a statistically significant difference between the means of two populations.

- Null Hypothesis (H_0): $\mu_1 = \mu_2$
- Alternative Hypothesis (H_a): $\mu_1 \neq \mu_2$

Where: μ_1 is the mean of one population and μ_2 is the mean of the other population of our interest.

Assumptions of Two Sample T-Tests

- The sample data drawn from both populations are unbiased and representative.
- The data of both populations are continuous.
- The data of both populations are normally distributed.
- The variances of both populations are unknown.
- Two sample t-test is more robust than a z-test when the sample size is small (< 30).

Three Types of Two Sample T-Tests

1. Two sample t-test when the variances of two populations are unknown but equal

Two sample t-test, equal variances; (when $\sigma^2_1 = \sigma^2_2$)

2. Two sample t-test when the variances of the two population are unknown and unequal

Two sample t-test, unequal variances; (when $\sigma^2_1 \neq \sigma^2_2$)

3. Paired t-test when the two populations are dependent of each other, so each data point from one distribution corresponds to a data point in the other distribution.

Test of Equal Variance

To check whether the variances of two populations of interest are significantly different on a statistical basis, we use the test of equal variance. An F-test is used to test this equality between two normally distributed populations.

- Null Hypothesis (H_0): $\sigma_1^2 = \sigma_2^2$
- Alternative Hypothesis (H_1): $\sigma_1^2 \neq \sigma_2^2$

An *F-test* is a statistic hypothesis test in which the test statistic follows an F-distribution when the null hypothesis is true. The most known F-test is the test of equal variance for two normally distributed populations. The F-test is very sensitive to non-normality. When any one

of the two populations is not normal, we use the Brown–Forsythe test for checking the equality of variances.

Test Statistic
To calculate the F statistic, you divide the squared standard deviation of one population by the other.

$$F_{calc} = \frac{s_1^2}{s_2^2}$$

Where: s_1 and s_2 are the sample standard deviations.

Critical Value
F_{crit} is the F value in a F distribution with the predetermined significance level α and degrees of freedom ($n_1 - 1$) and ($n_2 - 1$).

F_{crit} values for a two-sided and a one-sided F-test with the same significance level α and degrees of freedom ($n_1 - 1$) and ($n_2 - 1$) are different.

The critical value is a reference value that you will find in a table, using the prescribed parameters—the significance level alpha, and degrees of freedom for both samples. Based on the sample data, we calculated the test statistic F_{calc}, which is compared against F_{crit} to make a decision of whether to reject the null.

- Null Hypothesis (H_0): $\sigma_1^2 = \sigma_2^2$
- Alternative Hypothesis (H_a): $\sigma_1^2 \neq \sigma_2^2$
- If $F_{calc} > F_{crit}$, we reject the null and claim there is a statistically significant difference between the variances of the two populations.
- If $F_{calc} < F_{crit}$, we fail to reject the null and claim there is not any statistically significant difference between the variances of the two populations.

Test Statistic and Critical Value of a Two Sample T-Test when $\sigma_1^2 = \sigma_2^2$

Test Statistic
For a two-sample t-test, the *test statistic* is calculated using the sample means for the two populations, the pooled standard deviation, and the sample sizes.

$$t_{calc} = \frac{\overline{Y}_1 - \overline{Y}_2}{s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{(n_1 + n_2 - 2)}}$$

Where:

- $\overline{Y}_1$ and $\overline{Y}_2$ are the sample means of the two populations of our interest.
- n_1 and n_2 are the sample sizes. n_1 is not necessarily equal to n_2.

s_{pooled} is a pooled estimate of variance. s_1 and s_2 are the sample standard deviations

Critical Value

- t_{crit} is the t value in a Student's t distribution with the predetermined significance level α and degrees of freedom ($n_1 + n_2 - 2$).
- t_{crit} values for a two-sided and a one-sided t-test with the same significance level α and degrees of freedom ($n_1 + n_2 - 2$) are different

Pooled variance is a method for estimating variance given several different samples taken in different circumstances where the mean may vary between samples but the true variance is assumed to remain the same. The critical t value is pulled from the Student's t distribution using the desired significance level and the degrees of freedom calculated by ($n_1 + n_2 - 2$).

Test Statistic and Critical Value of a Two Sample T-Test when $\sigma_1^2 \neq \sigma_2^2$

Test Statistic

For a two-sample t-test with unknown variances or with known unequal variances, the test statistic is not calculated using the pooled standard deviation.

$$t_{calc} = \frac{\overline{Y}_1 - \overline{Y}_2}{\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}}$$

$$df = \frac{\left(\frac{s_1^{\,2}}{n_1} + \frac{s_2^{\,2}}{n_2}\right)^2}{\frac{\left(s_1 / n_1\right)^2}{n_1 - 1} + \frac{\left(s_2 / n_2\right)^2}{n_2 - 1}}$$

Where:

- $\overline{Y}_1$ and $\overline{Y}_2$ are the sample means of the two populations of our interest
- n_1 and n_2 are the sample sizes. n_1 is not necessarily equal to n_2.

s_{pooled} is a pooled estimate of variance. s_1 and s_2 are the sample standard deviations

Critical Value

- t_{crit} is the t value in a Student's t distribution with the predetermined significance level α and degrees of freedom df calculated using the formula above.

- t_{crit} values for a two-sided and a one-sided t-test with the same significance level α and degrees of freedom *df are different*.

Pooled variance is a method for estimating variance given several different samples taken in different circumstances where the mean may vary between samples but the true variance is assumed to remain the same. When we cannot assume equal variances or if we know they are unequal we cannot make the same assertions by using pooled variances. Therefore, we use the standard deviation of each sample separately in our t_{calc} equation.

Test Statistic and Critical Value of a Paired T-Test

Test Statistic
When using a paired t-test, the test statistic is calculated using the average difference between the data pairs, the standard deviation of the differences, and the sample size of either population.

$$t_{calc} = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Where:

- d is the difference between each pair of data
- $\bar{d}$ is the average of d
- n is the sample size of either population of interest.

Critical Value
- t_{crit} is the t value in a Student's t distribution with the predetermined significance level α and degrees of freedom (n – 1).
- t_{crit} values for a two-sided and a one-sided t-test with the same significance level α and degrees of freedom (n – 1) are different.

The critical t is looked up using the significance level alpha and degrees of freedom (n – 1).

Decision Rules of a Two-Sample T-Test

Based on the sample data, we calculated the test statistic t_{calc}, which is compared against t_{crit} to make a decision of whether to reject the null.

- Null Hypothesis (H_0): $\mu_1 = \mu_2$
- Alternative Hypothesis (H_a): $\mu_1 \neq \mu_2$

If $|t_{calc}| > t_{crit}$, we reject the null and claim there is a statistically significant difference between the means of the two populations.

If $|t_{calc}| < t_{crit}$, we fail to reject the null and claim there is not any statistically significant difference between the means of the two populations.

Use JMP to Run a Two-Sample T-Test

Case study: Compare the average retail price of a product in state A and state B.

Data File: "TwoSampleT-Test.jmp"

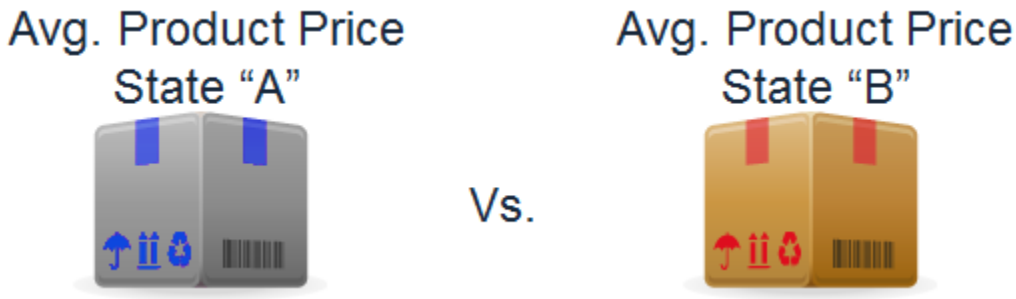


Fig. 3.35 Two-Sample T-Test comparison

- Null Hypothesis (H_0): $\mu_1 = \mu_2$
- Alternative Hypothesis (H_a): $\mu_1 \neq \mu_2$

In this example, we will be comparing the average price of a product in two different states. The null hypothesis is that the price in state A is equal to the price in state B.

Step 1: Test the normality of the retail price for both state A and B.

1. Click Analyze -> Distribution
2. Select "Retail Price" as "Y, Columns"
3. Select "State" as "By"

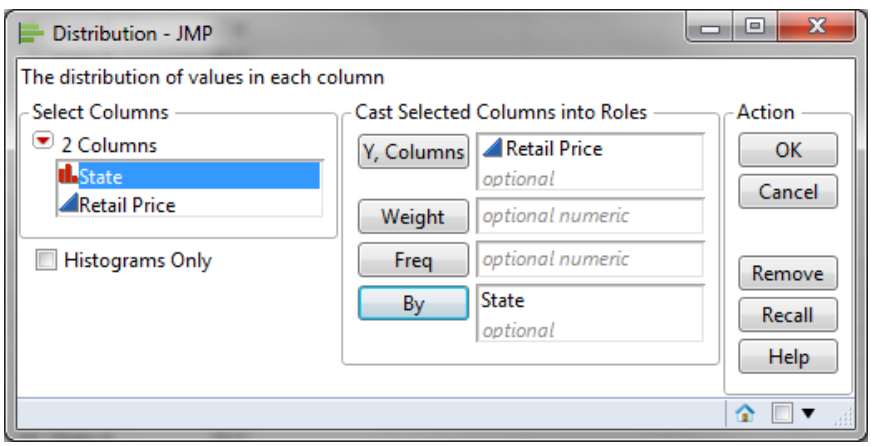


Fig. 3.36 Summary dialog box with variables

4. Click "OK"

5. Click on the red triangle button next to “Retail Price” in the Distribution page for “State = State A”
6. Click Continuous Fit -> Normal
7. Click on the red triangle button next to “Fitted Normal”
8. Select “Goodness of Fit”
9. Repeat the same to test the normality for “State = State B”

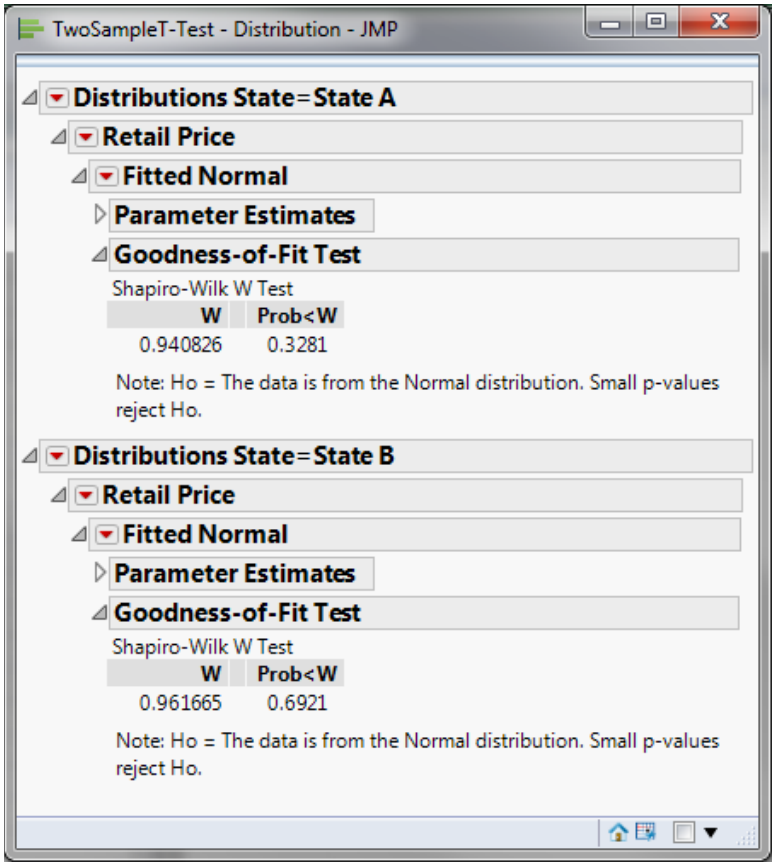


Fig. 3.37 Data results for State B

- Null Hypothesis (H_0): The data are normally distributed.
- Alternative Hypothesis (H_a): The data are not normally distributed.

Both retail price data of state A and B are normally distributed since the p-values are both greater than alpha level (0.05). If any of the data series is not normally distributed, we need to use other hypothesis testing methods other than the two sample t-test.

By following the instructions on the previous page, we first determine if the data follow a normal distribution. In this case, you can see that the p-value for both is higher than 0.05, so we fail to reject the null hypothesis that the data are normally distributed. If the data are not normally distributed, we must use a different test.

Step 2: Test whether the variances of the two data sets are equal.

- Null Hypothesis (H_0): $\sigma_1^2 = \sigma_2^2$

- Alternative Hypothesis (H_a): $\sigma_1^2 \neq \sigma_2^2$

1. Click Analyze -> Fit Y by X
2. Select "Retail Price" as "Y, Response"
3. Select "State" as "X, Factor"

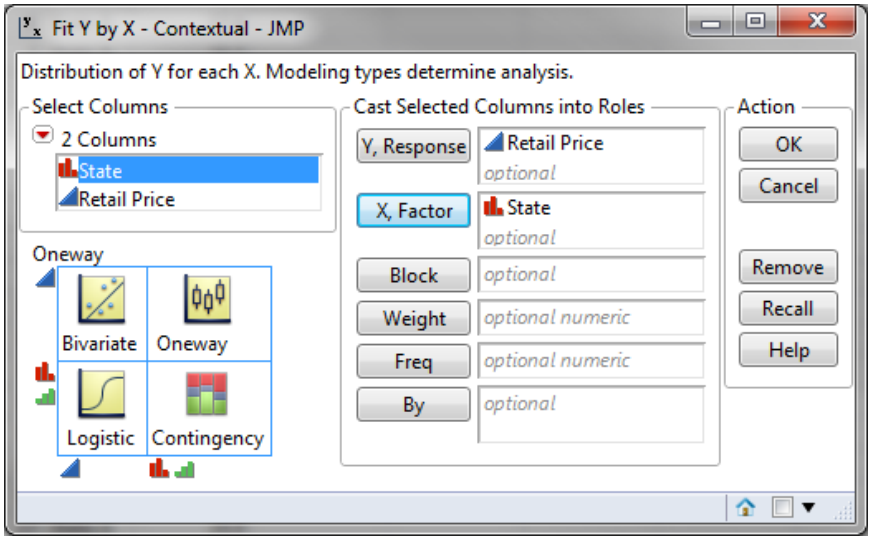


Fig. 3.38 Two-Sample variance with sample variance

4. Click "OK"
5. Click on the red triangle button next to "One-Way Analysis of Retail Price by State"
6. Select "Unequal Variances"

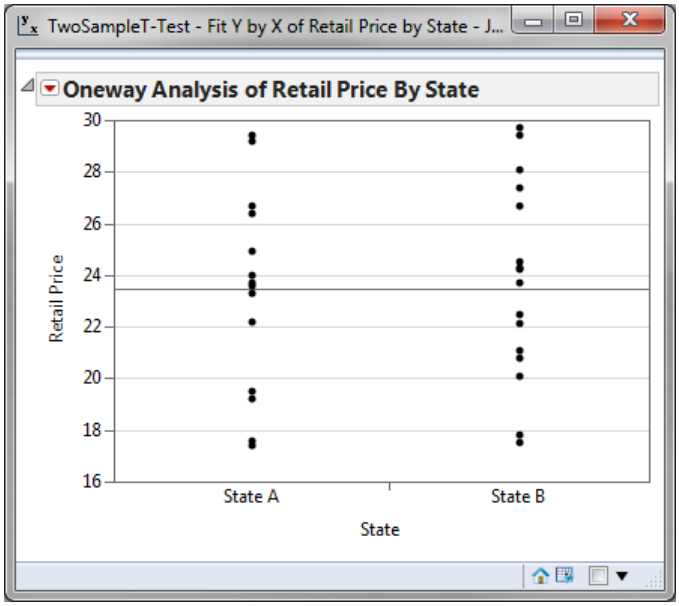


Fig. 3.39 Two-Sample variance output

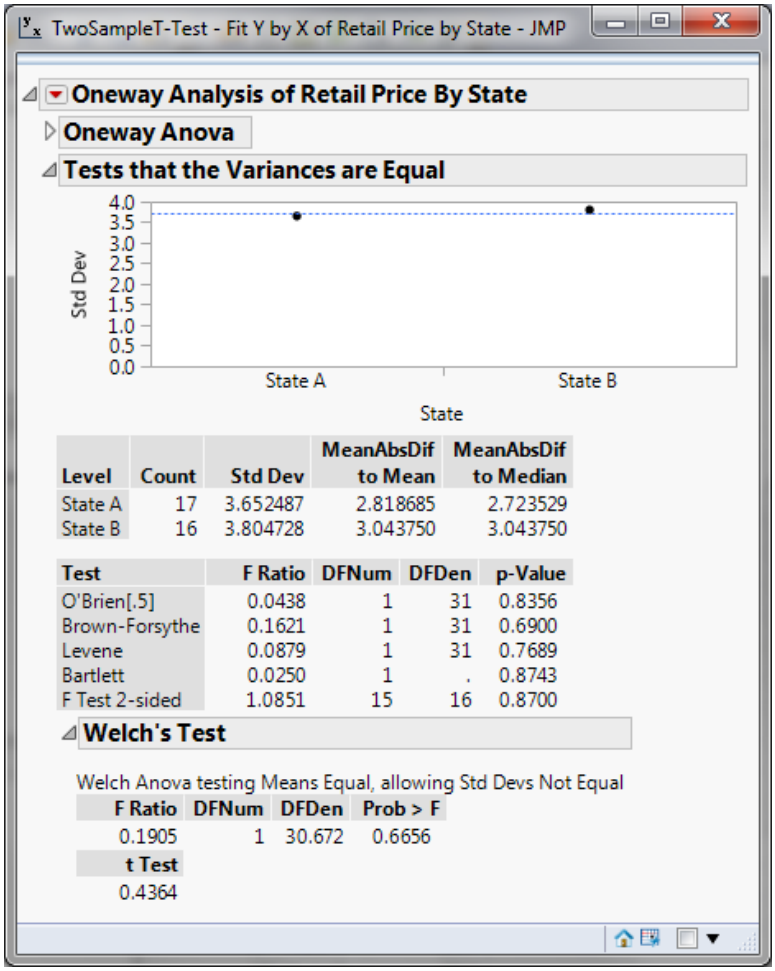


Fig. 3.40 Two-Sample t variance output

Because the retail prices at state A and state B are both normally distributed, an F test is used to test their variance equality. The p-value of F test is 0.870, greater than the alpha level (0.05), so we fail to reject the null hypothesis and we claim that the variances of the two data sets are equal. We will use the two sample t-test (when $\sigma^2_1 = \sigma^2_2$) to compare the means of the two groups. If $\sigma^2_1 \neq \sigma^2_2$, we will use the two sample t-test (when $\sigma^2_1 \neq \sigma^2_2$) to compare the means of the two groups.

Step 3: Run two-sample t-test to compare the means of two groups.

1. Click on the red triangle button next to "One-Way Analysis of Retail Price By State"
2. Select "Means/Anova/Pooled t"

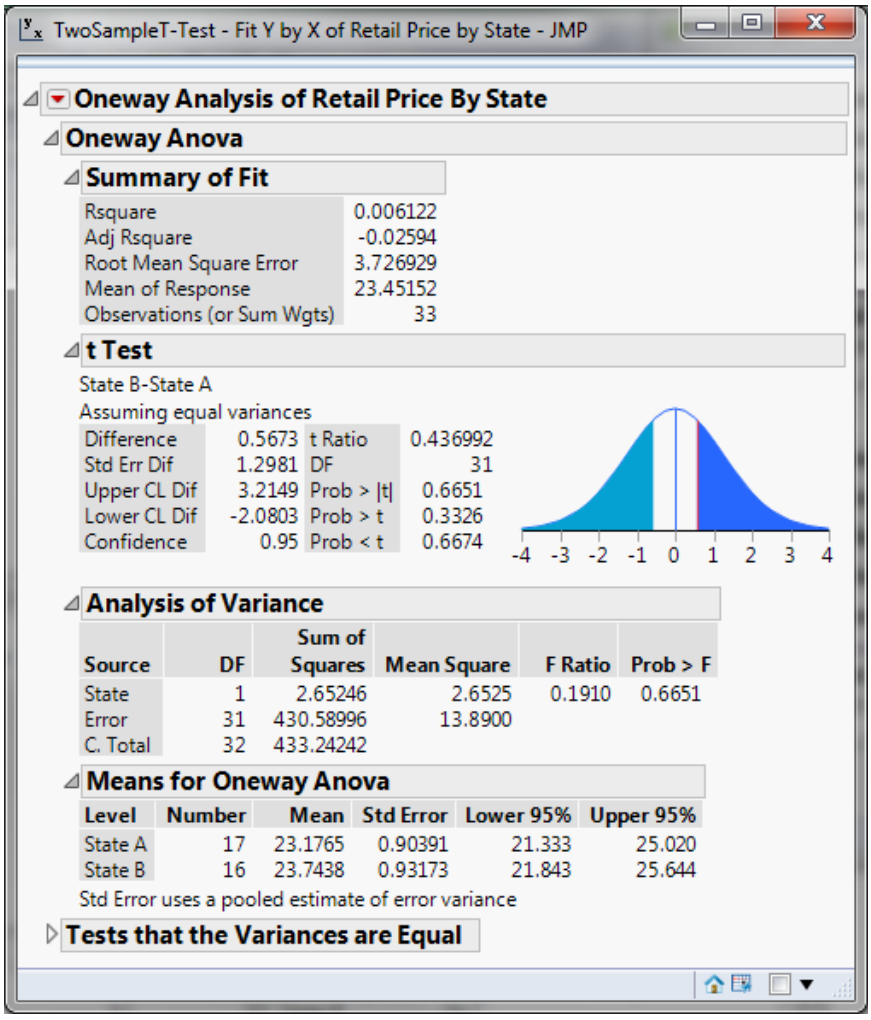


Fig. 3.41 ANOVA results assuming equal variances

Since the p-value of the t-test (assuming equal variance) is 0.665, greater than the alpha level (0.05), we fail to reject the null hypothesis and we claim that the means of the two data sets are equal. If the variances of the two groups do not equal, we will need to use the two-sample t-test (when $\sigma_1 \neq \sigma_2$) to compare the means of the two groups.

- Click on the red triangle button next to "One-Way Analysis of Retail Price By State"
- Select "t test"

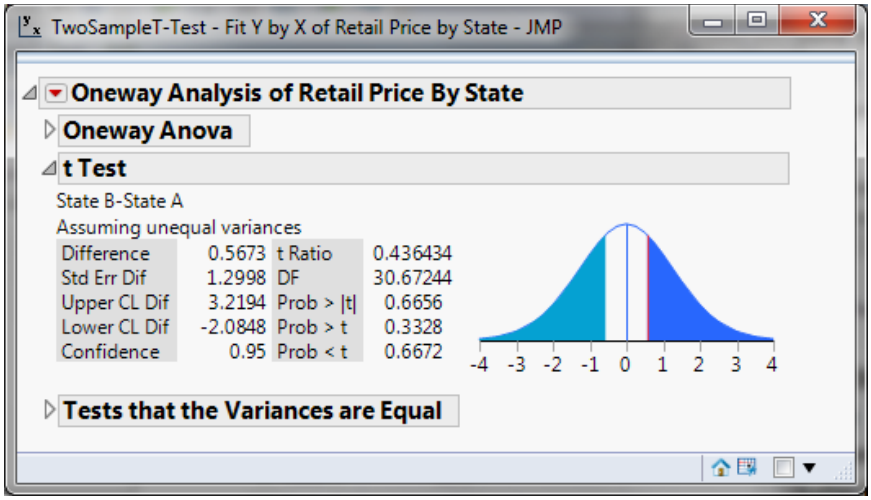


Fig. 3.42 JMP Output for Two-Sample T-Test assuming unequal variances

Since the p-value of the t-test (assuming unequal variance) is 0.666, greater than the alpha level (0.05), we fail to reject the null hypothesis and we claim that the means of two groups are equal.

Case study: We are interested to know whether the average salaries ($1000/yr.) of male and female professors at the same university are the same.

Data File: "PairedT-Test.jmp"

The data were randomly collected from 22 universities. For each university, the salaries of male and female professors were randomly selected.

- Null Hypothesis (H_0): $\mu_{male} - \mu_{female} = 0$
- Alternative Hypothesis (H_a): $\mu_{male} - \mu_{female} \neq 0$

Use JMP to Run a Paired T-Test

Step 1: Create a new column for the difference between the two data sets of interest. (MALES – FEMALES = Difference)

Step 2: Test whether the difference is normally distributed.

- Null Hypothesis (H_0): The difference between two data sets is normally distributed.
- Alternative Hypothesis (H_a): The difference between two data sets is not normally distributed.

1. Click Analyze -> Distribution
2. Select "Difference" as "Y, Columns"

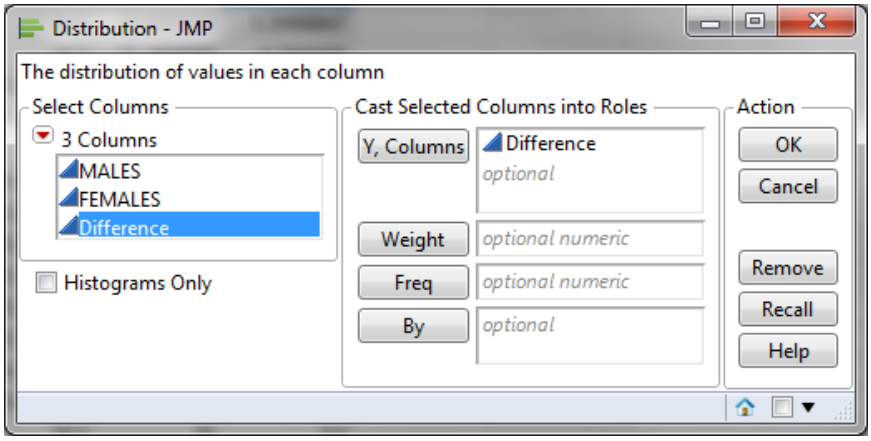


Fig. 3.43 Difference selection

3. Click "OK"
4. Click on the red triangle button next to "Difference" in the Distribution page
5. Click Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

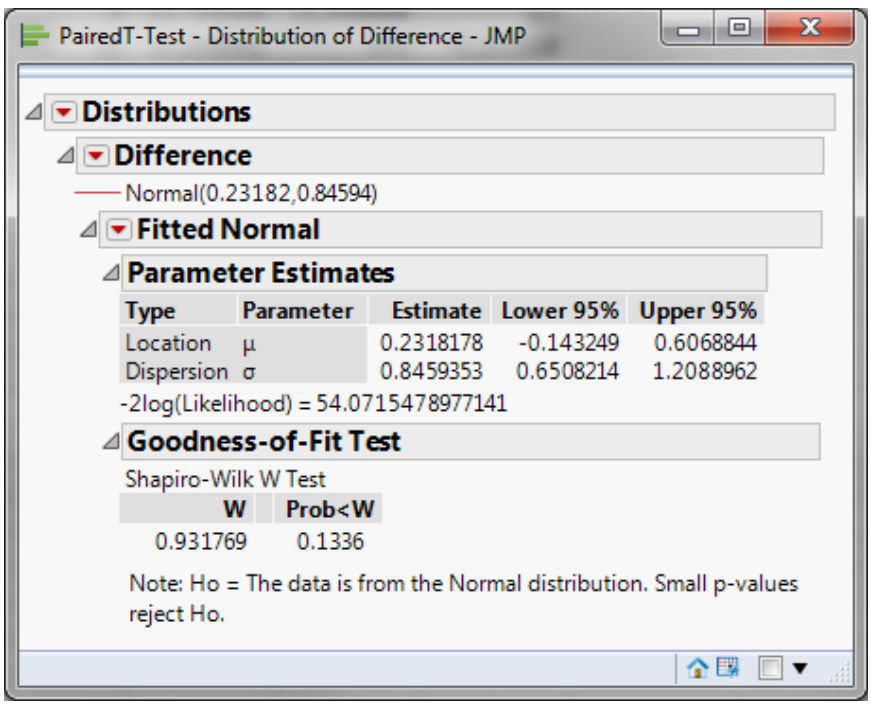


Fig. 3.44 Paired T-Test data output

The p-value of the normality test is 0.1336, greater than the alpha level (0.05), so we fail to reject the null hypothesis and we claim that the difference is normally distributed. When the difference is not normally distributed, we need other hypothesis testing methods.

Step 3: Run the paired t-test to compare the means of two dependent data sets.

1. Click Analyze -> Specialized Modeling->Matched Pairs

2. Select both “MALES” and “FEMALES” as “Y, Paired Response”

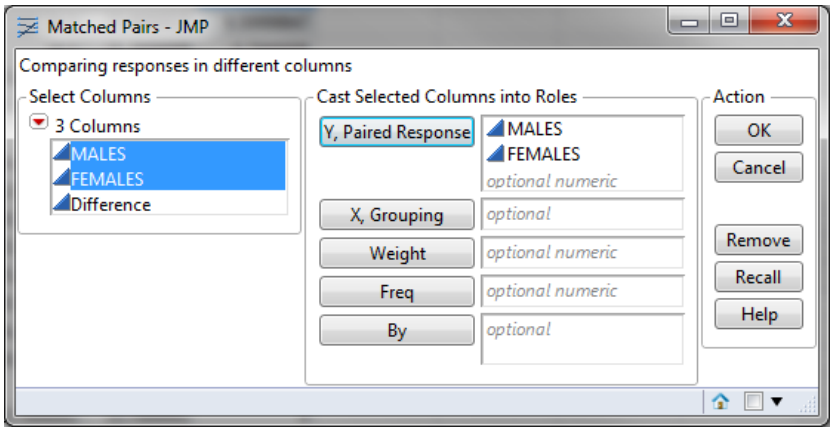


Fig. 3.45 Comparison Selections

3. Click “OK”

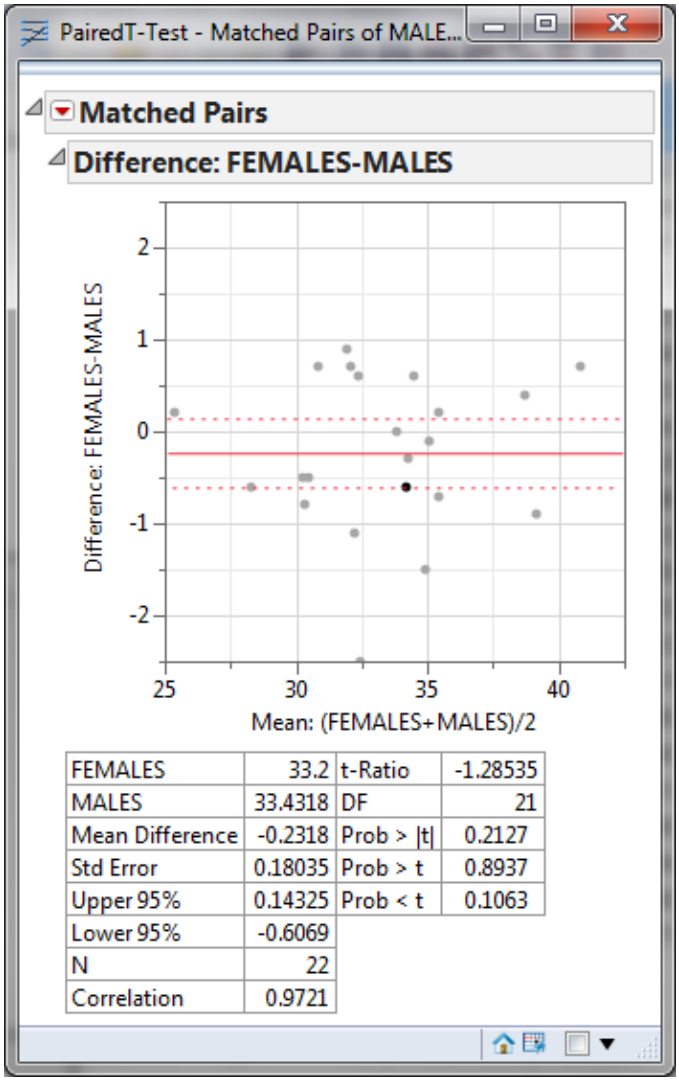


Fig. 3.46 Paired T-Test Output

The p-value of the paired t-test is 0.2127, greater than the alpha level (0.05), so we fail to reject the null hypothesis and we claim that there is no statistically significant difference between the

salaries of male and those of female professors. Because the p-value is higher than 0.05, we again fail to reject the null hypothesis, meaning that there is no statistically significant difference between the salaries of male and female professors.

3.4.2 ONE SAMPLE VARIANCE

What is One Sample Variance Test?

One sample variance test is a hypothesis testing method to study whether there is a statistically significant difference between a population variance and a specified value.

- Null Hypothesis (H_0): $\sigma^2 = \sigma_0^2$
- Alternative Hypothesis (H_a): $\sigma^2 \neq \sigma_0^2$

Where: σ^2 is the variance of a population of our interest and σ_0^2 is the specific value we want to compare against. The null hypothesis is that the population variance is equal to a specified value.

Chi-Square Test

A *chi-square test* is a statistical hypothesis test in which the test statistic follows a chi-square distribution when the null hypothesis is true. A chi-square test can be used to test the equality between the variance of a normally distributed population and a specified value. In a chi-square test, the test statistic follows a chi-square distribution when the null hypothesis is true. The test can be used to test the equality of a variance to a specified value. For the test to be valid, the distribution must be normally distributed.

Test Statistic
The test statistic is calculated using the observed variance of the sample from the population, as well as the sample size and the specified value for variance we are comparing against.

$$\chi_{calc}^{\ 2} = \frac{(n-1)s^2}{\sigma_0^{\ 2}}$$

Where:

- s^2 is the observed variance and n is the sample size
- σ_0^2 is the specified value we compare against.

Critical Value

- χ_{crit}^2 is the χ^2 value in a χ^2 distribution with the predetermined significance level α and degrees of freedom (n – 1).
- χ_{crit}^2 values for a two-sided and a one-sided χ^2-test with the same significance level α and degrees of freedom (*n –1) are different.*

Like in other tests we discussed, we compare the calculated chi-square statistic against a critical value, which is a chi-square table value that can be located based on the determined significance level and (n – 1) degrees of freedom. Based on the sample data, we calculated the test statistic χ_{calc}^2, which is compared against χ_{crit}^2 to make a decision of whether to reject the null.

- Null Hypothesis (H_0): $\sigma^2 = \sigma_0^2$
- Alternative Hypothesis (H_a): $\sigma^2 \neq \sigma_0^2$
- If $\left|\chi_{calc}^{\ 2}\right| > \chi_{crit}^{\ 2}$, we reject the null and claim there is a statistically significant difference between the population variance and the specified value.
- If $\left|\chi_{calc}^{\ 2}\right| < \chi_{crit}^{\ 2}$, we fail to reject the null and claim there is not any statistically significant difference between the population variance and specified value.

In other words, if the absolute value of the statistic is greater that the critical value, then we reject the null and claim that the population variance is different than the specified value. Otherwise, we fail to reject the null.

Use JMP to Run a One-Sample Variance Test

Case study: We are interested in comparing the variance of the height of basketball players with zero.

 ***Data File: "OneSampleT-Test.jmp"***

- Null Hypothesis (H_0): $\sigma^2 = 0$
- Alternative Hypothesis (H_a): $\sigma^2 \neq 0$

In this case study, we will use the software to determine if the variance in the height of basketball players is zero or different from zero.

Step 1: Test whether the data are normally distributed

1. Click Analyze -> Distribution
2. Select "HtBk" as "Y, Columns"

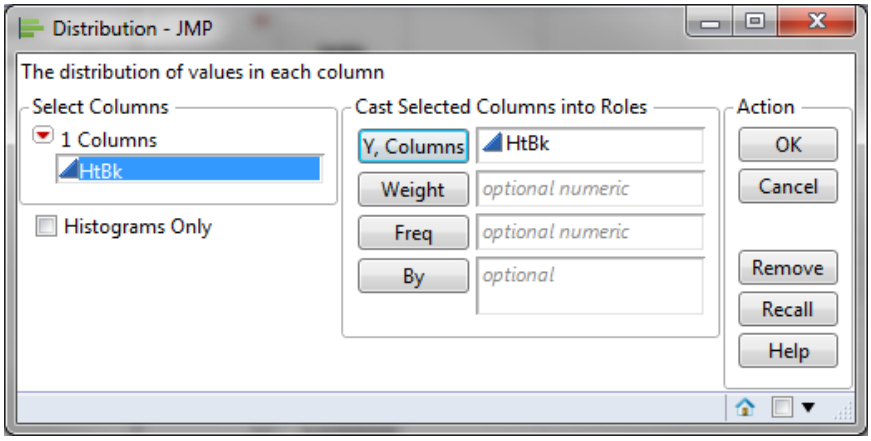


Fig. 3.47 Distribution selections

3. Click "OK"
4. Click on the red triangle button next to "HtBk" in the Distribution page
5. Click Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

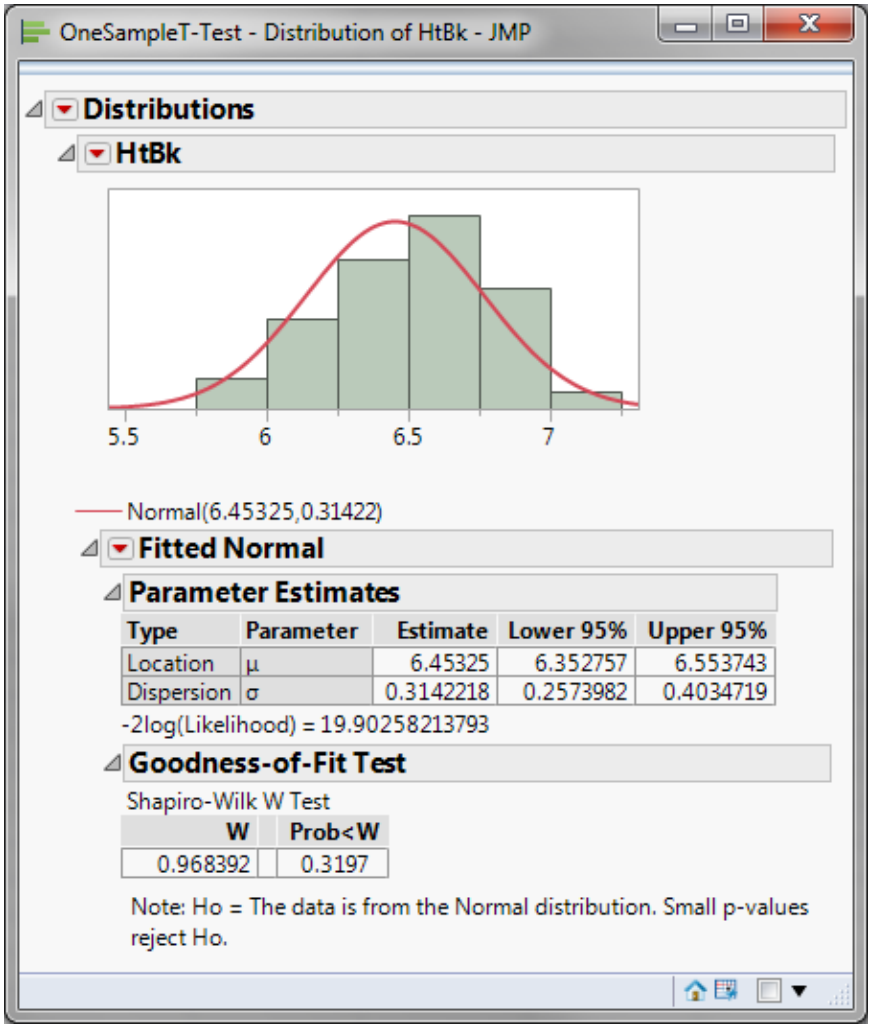


Fig. 3.48 Normality Test Results

- Null Hypothesis (H_0): The data are normally distributed.
- Alternative Hypothesis (H_a): The data are not normally distributed.

Since the p-value of the normality is 0.3197, greater than alpha level (0.05), we fail to reject the null and claim that the data are normally distributed. If the data are not normally distributed, you need to use other hypothesis tests. Recall that the data must be normally distributed in order for a one-sample variance test to be valid. As you can see from the results, the p-value is greater than 0.05, so we fail to reject the null hypothesis that the data are normally distributed.

Step 2: Now check whether the specified value (zero) is within the 95% Confidence interval.

- If the square root of the specified value is between the upper and lower confidence interval boundaries of the standard deviation in "Parameter Estimates" section, we fail to reject the null hypothesis and claim that the variance of the population is not statistically different from the specified value.
- In this case, the square root of zero is out of the confidence interval of the population standard deviation and we claim that the population variance is statistically different from zero.

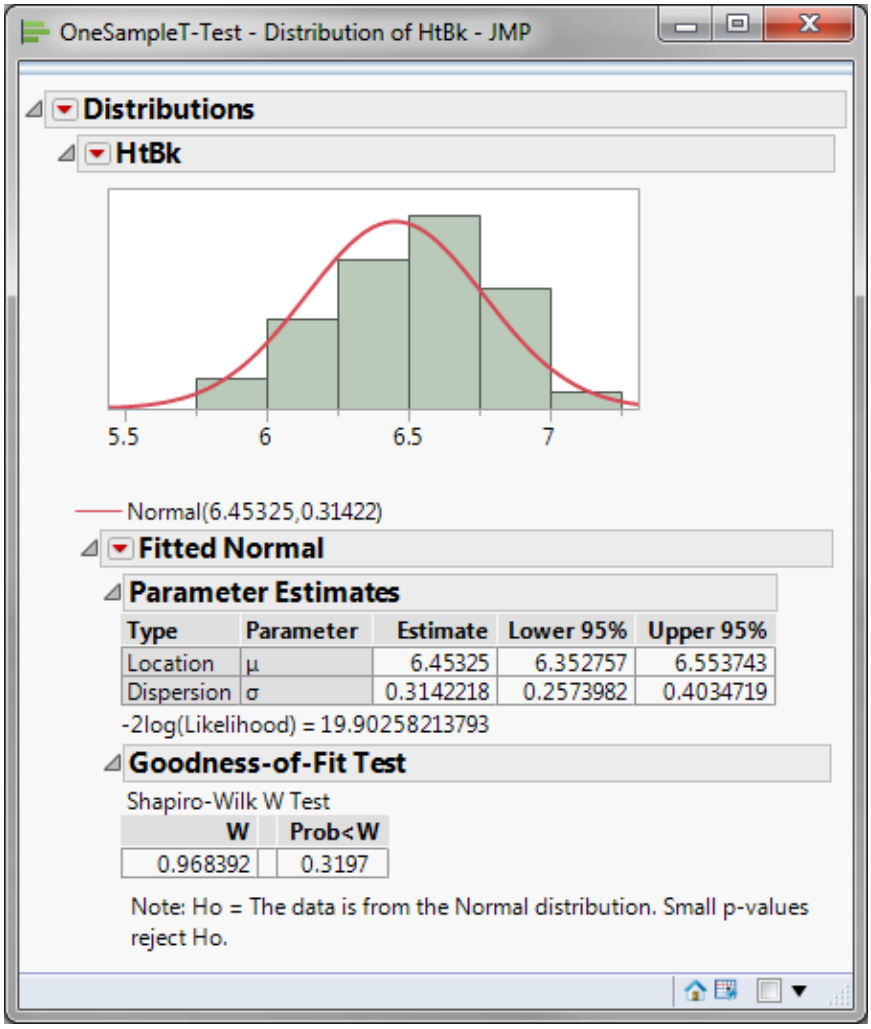


Fig. 3.49 One Sample Variance results

Another way to see whether the population variance is statistically different from zero is to check whether zero stays between the upper and lower confidence interval boundaries of the variance. If yes, we fail to reject the null hypothesis and claim that the population variance is not statistically different from zero. Since the p-value is less than the alpha level, we must reject the null hypothesis that the population variance is equal to zero.

Another way to interpret the results: we can check to see if the specified variance falls within the confidence interval (CI) for the population variance. If it does, then the population variance is statistically the same as the specified value.

3.4.3 One-Way ANOVA

What is One-Way ANOVA?

One-way ANOVA is a statistical method to compare means of two or more populations.

- Null Hypothesis (H_0): $\mu_1 = \mu_2 = \ldots = \mu_k$

- Alternative Hypothesis (H_a): At least one μ_i is different, where i is any value from 1 to k.

It is a generalized form of the two-sample t-test since a two-sample t-test compares two population means and one-way ANOVA compares k population means where $k \geq 2$.

Assumptions of One-Way ANOVA

- The sample data drawn from k populations are unbiased and representative.
- The data of k populations are continuous.
- The data of k populations are normally distributed.
- The variances of k populations are equal.

How ANOVA Works

ANOVA compares the means of different groups by analyzing the variances between and within groups. Let us say we are interested in comparing the means of three normally distributed populations. We randomly collected one sample for each population of our interest.

- Null Hypothesis (H_0): $\mu_1 = \mu_2 = \mu_3$
- Alternative Hypothesis (H_a): One of the μ is different from the others.

Based on the sample data, the means of the three populations might look different because of two variation sources.

1. Variation between groups there are non-random factors leading to the variation between groups.
2. Variation within groups there are random errors resulting in the variation within each individual group.

What we care about the most is the variation between groups since we are interested in whether the groups are statistically different from each other. Variation between groups is the *signal* we want to detect and variation within groups is the *noise* which corrupts the signal.

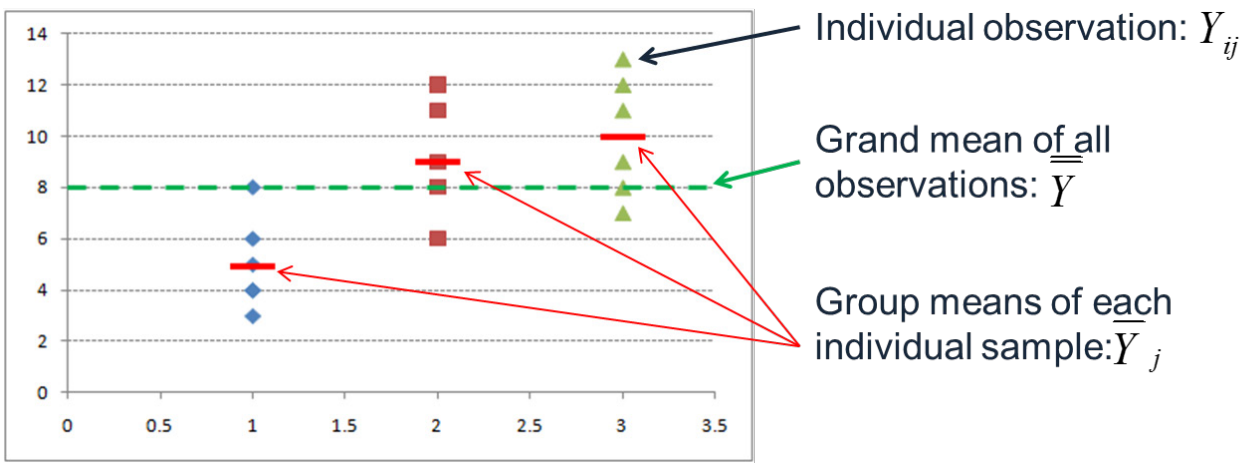


Fig. 3.50 How ANOVA works

$$Total\ Variation = SS(Total) = \sum_{j=1}^{k}\sum_{i=1}^{n_i}\left(Y_{ij} - \bar{\bar{Y}}\right)^2$$

$$Between\ Variation = SS(Between) = \sum_{j=1}^{k} n_j\left(Y_j - \bar{\bar{Y}}\right)^2$$

$$Within\ Variation = SS(Within) = \sum_{j=1}^{k}\sum_{i=1}^{n_i}\left(Y_{ij} - \bar{Y}_j\right)^2$$

Using this visual, the purpose of using the ANOVA is to determine if there is a difference between the data points (the group means of the individual samples). The variation within the groups can make it difficult to detect the difference between the group means.

Variation Components

$$Total\ Variation = Variation\ Between\ Groups + Variation\ Within\ Groups$$

$$\sum_{j=1}^{k}\sum_{i=1}^{n_i}\left(Y_{ij} - \bar{\bar{Y}}\right)^2 = \sum_{j=1}^{k} n_j\left(Y_j - \bar{\bar{Y}}\right)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_i}\left(Y_{ij} - \bar{Y}_j\right)^2$$

Total Variation = Sums of squares of the vertical difference between the individual observation and the grand mean Variation Between Groups = Sums of squares of the vertical difference between the group mean and the grand mean. Variation Within Groups = Sum of squares of the vertical difference between the individual observation and the group mean

Degrees of Freedom (DF)

In statistics, the degrees of freedom are the number of unrestricted values in the calculation of a statistic.

Degrees of Freedom Components

$$DF_{total} = DF_{between} + DF_{within}$$

$$DF_{total} = n - 1$$

$$DF_{between} = k - 1$$

$$DF_{within} = n - k$$

Where:

- n is the total number of observations
- k is the number of groups.

Signal-to-Noise Ratio (SNR)

Earlier, the terms signal and noise were introduced. The *signal* is the variation between groups that we are trying to detect. The *noise* is the variation within a group that makes it difficult to see the signal.

SNR denotes the ratio of a signal to the noise corrupting the signal. It measures how much a signal has been corrupted by the noise. When it is higher than 1, there is more signal than noise. The higher the SNR, the less the signal has been corrupted by the noise.

F-ratio is the SNR in ANOVA

$$F = \frac{MS_{between}}{MS_{within}} = \frac{SS_{between} \Big/ DF_{between}}{SS_{within} \Big/ DF_{within}} = \frac{\sum_{j=1}^{k} (\overline{Y}_j - \overline{\overline{Y}})^2 \Big/ (k-1)}{\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(Y_{ij} - \overline{Y}_j\right)^2 \Big/ (n-k)}$$

In ANOVA, we use the F-test to compare the means of different groups. The F-ratio calculated as above is the test statistic F_{calc}.

The critical value (F_{cri}) in an F-test can be derived from the F table with predetermined significance level (α) and with (k –1) degrees of freedom in the numerator and (n – k) degrees of freedom in the denominator.

How ANOVA Works

- Null Hypothesis (H_0): $\mu_1 = \mu_2 = \ldots = \mu_k$
- Alternative Hypothesis (H_a): At least one μ_i is different, where *i* is any value from 1 to *k*.
- If $|F_{calc}| < F_{crit}$, we fail to reject the null and claim that the means of all the populations of our interest are the same.
- If $|F_{calc}| > F_{crit}$, we reject the null and claim that there is at least one mean different from the others.

Using the F-ratio, we can compare the calculated value against the critical value. If the absolute value of the calculated value is less than the critical value, then we fail to reject the null hypothesis. If the absolute value of the calculated value is greater than the critical value, then we reject the null hypothesis.

Model Validation

ANOVA is a modeling procedure, which means we are using a model to try to predict results. To make sure the conclusions made in ANOVA are reliable, we need to perform residuals analysis.

Good residuals:

- Have a mean of zero
- Are normally distributed

- Are independent of each other
- Have equal variance

The difference between the actual and predicted result is called a *residual* or unexplained variation.

Use JMP to Run an ANOVA

Case study: We are interested in comparing the average startup costs of five kinds of business.

Data File: "OneWayANOVA.jmp"

- Null Hypothesis (H_0): $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
- Alternative Hypothesis (H_a): At least one of the five means is different from others.

Step 1: Test whether the data for each level are normally distributed.

1. Click Analyze -> Distribution
2. Select "Cost" as "Y, Columns"
3. Select "Business" as "By"

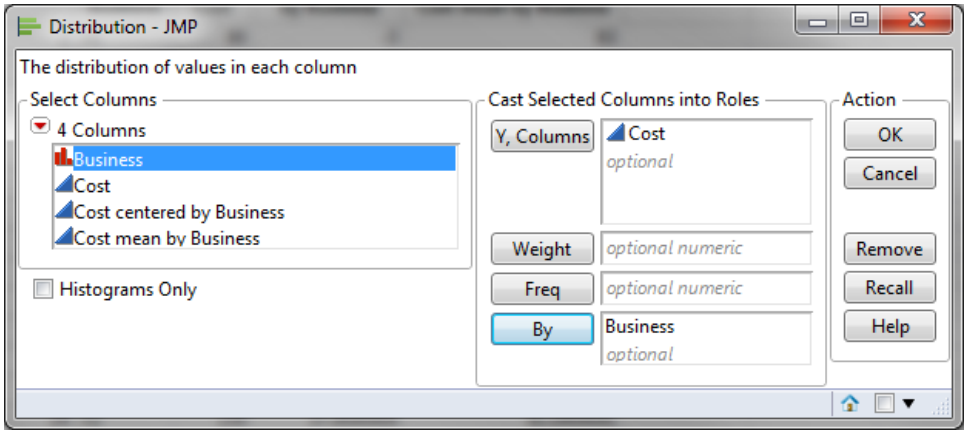


Fig. 3.51 Distribution selections

4. Click "OK"
5. Click on the red triangle button next to "Cost" in the Distribution page for "Business = X_i"
6. Click Continuous Fit -> Normal
7. Click on the red triangle button next to "Fitted Normal"
8. Select "Goodness of Fit"

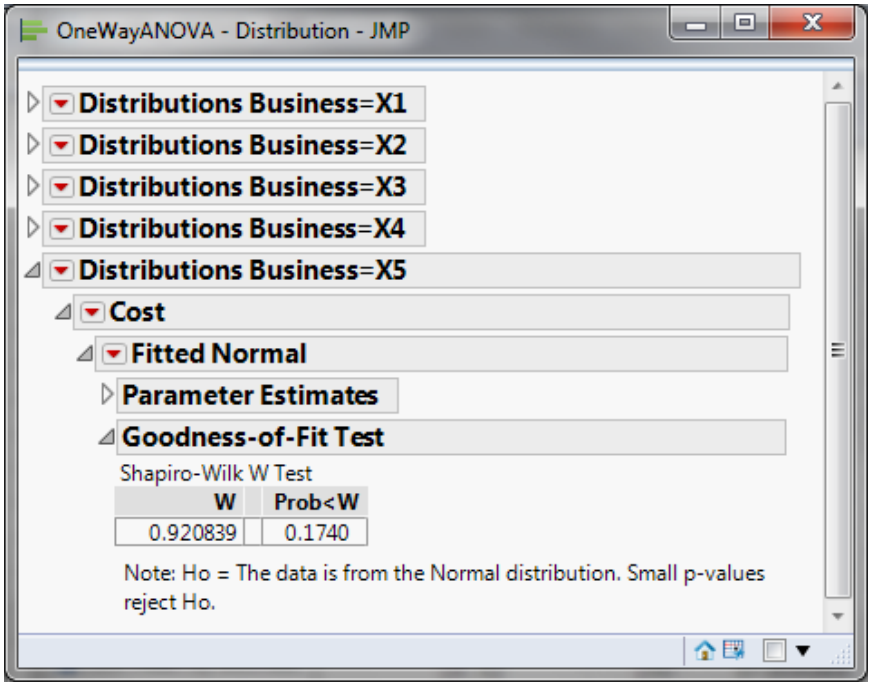


Fig. 3.52 ANOVA Output for all 5 Businesses

Notice all of the p-values are greater than 0.05; therefore, we fail to reject the null hypothesis that the data are normally distributed.

- Null Hypothesis (H_0): The data are normally distributed.
- Alternative Hypothesis (H_a): The data are not normally distributed.

Since the p-values of normality tests for the five data sets are higher than alpha level (0.05), we fail to reject the null hypothesis and claim that the startup costs for any of the five businesses are normally distributed. If any of the five data sets are not normally distributed, we need to use other hypothesis testing methods other than one-way ANOVA. In this example, all five data sets are normally distributed; however, if *any* of them were not normally distributed, we would need to use another hypothesis test.

Step 2: Test whether the variance of the data for each level is equal to the variance of other levels.

- Null Hypothesis (H_0): $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$
- Alternative Hypothesis (H_a): at least one of the variances is different from others.

1. Click Analyze -> Fit Y by X
2. Select “Cost” as “Y, Response”
3. Select “Business” as “X, Factor”

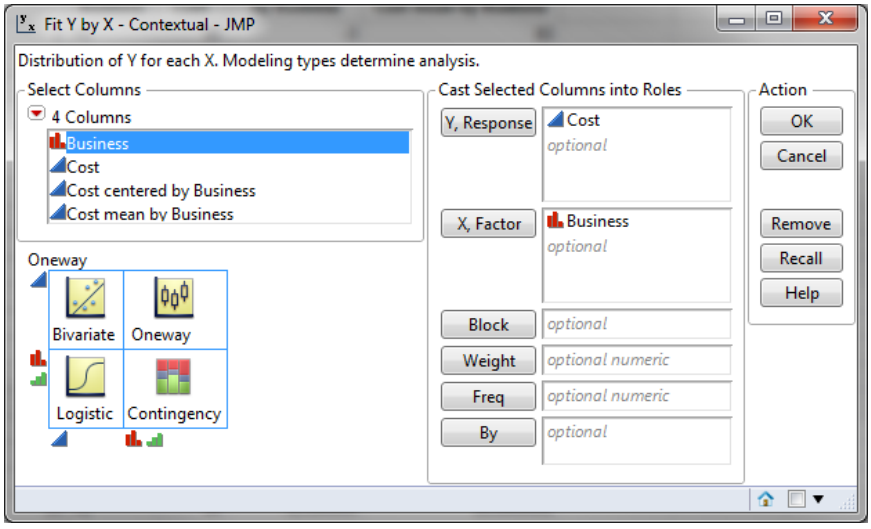
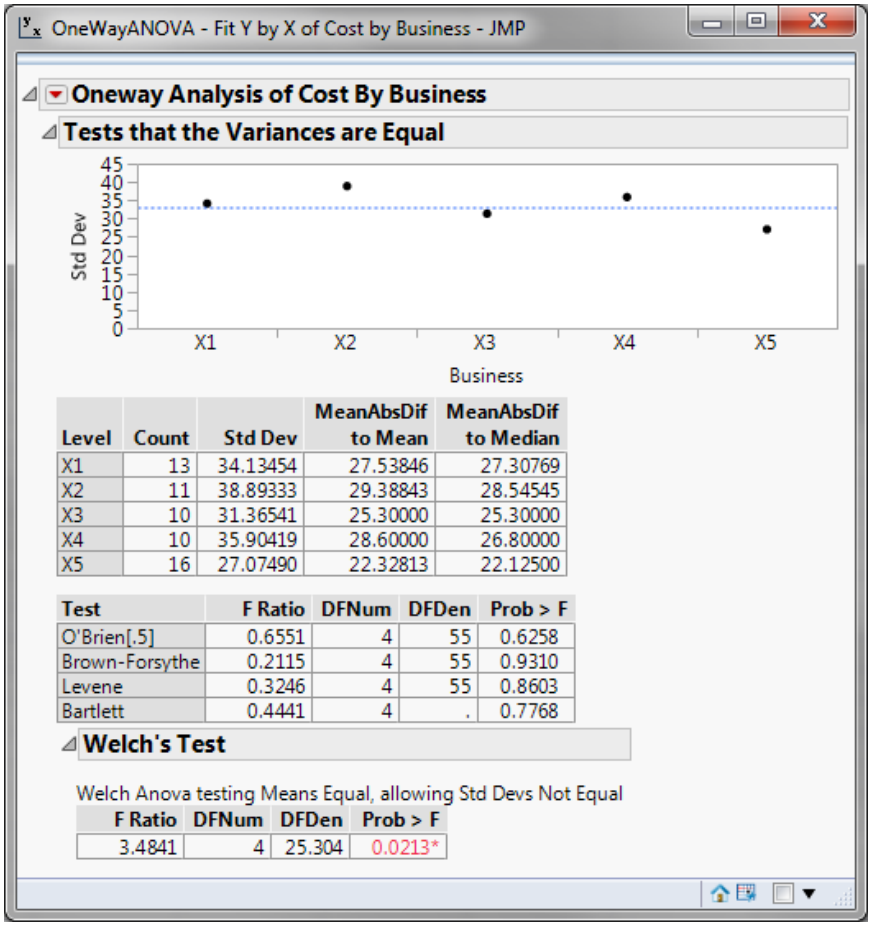


Fig. 3.53 Test for Equal Variances dialog box with response variable

4. Click "OK"
5. Click on the red triangle button next to "One-Way Analysis of Cost by Business"
6. Click "Unequal Variances"



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
X1	13	34.13454	27.53846	27.30769
X2	11	38.89333	29.38843	28.54545
X3	10	31.36541	25.30000	25.30000
X4	10	35.90419	28.60000	26.80000
X5	16	27.07490	22.32813	22.12500

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.6551	4	55	0.6258
Brown-Forsythe	0.2115	4	55	0.9310
Levene	0.3246	4	55	0.8603
Bartlett	0.4441	4	.	0.7768

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
3.4841	4	25.304	0.0213*

Fig. 3.54 JMP ANOVA Results

Use the Bartlett's test for testing the equal variances between five levels in this case since there are more than two levels in the data and the data of each level are normally distributed. The p-value of Bartlett's test is 0.777, greater than the alpha level (0.05), so we fail to reject the null hypothesis and we claim that the variances of five groups are equal. If the variances are not all equal, we need to use other hypothesis testing methods other than one-way ANOVA. If this test suggested that at least one variance was different, then we would need to use a different hypothesis test to evaluate the group means.

Step 3: Test whether the mean of the data for each level is equal to the means of other levels.

- Null Hypothesis (H_0): $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
- Alternative Hypothesis (H_a): at least one of the means is different from others

1. Click on the red triangle button next to "One-Way Analysis of Cost by Business"
2. Select "Mean/Anova"

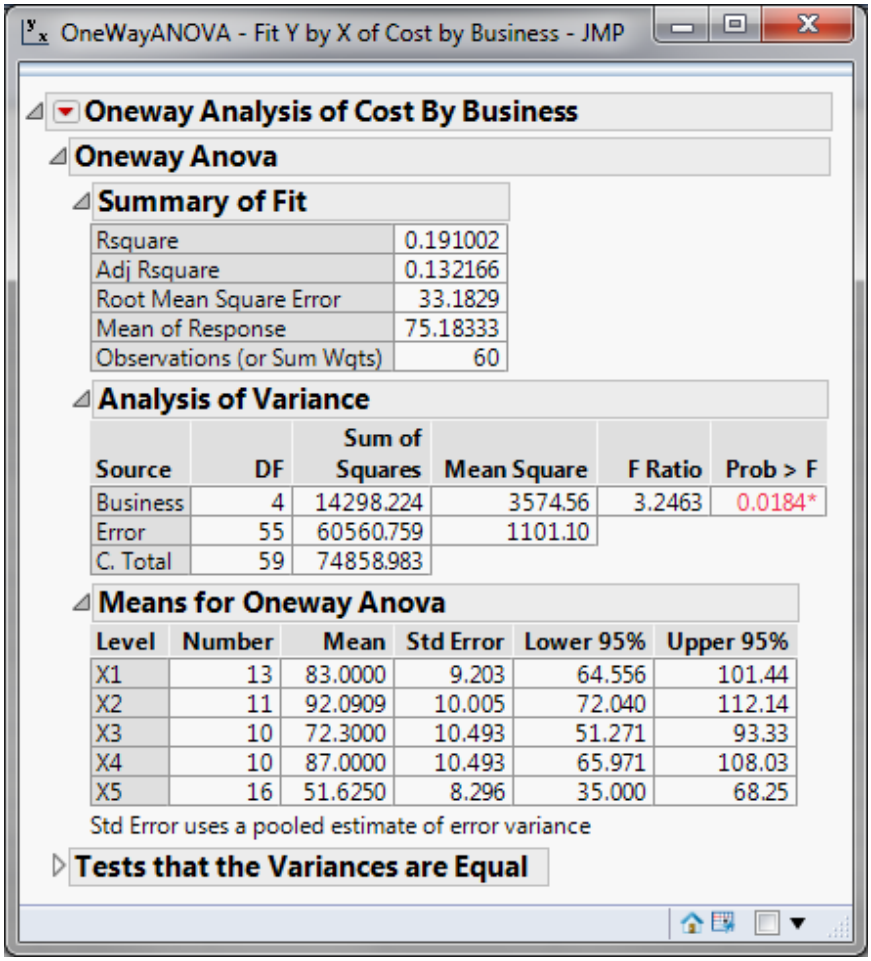


Fig. 3.55 One-Way ANOVA results

Since the p-value of the F test is 0.018, lower than the alpha level (0.05), the null hypothesis is rejected and we conclude that the at least one of the means of the five groups is different from others.

Step 4: Save the residuals and predicted values after ANOVA. The predicted value for each level is the group mean.

1. Click on the red triangle button next to “One-way Analysis of Cost by Business”
2. Select Save -> Save Residuals
3. Select Save -> Save Predicted

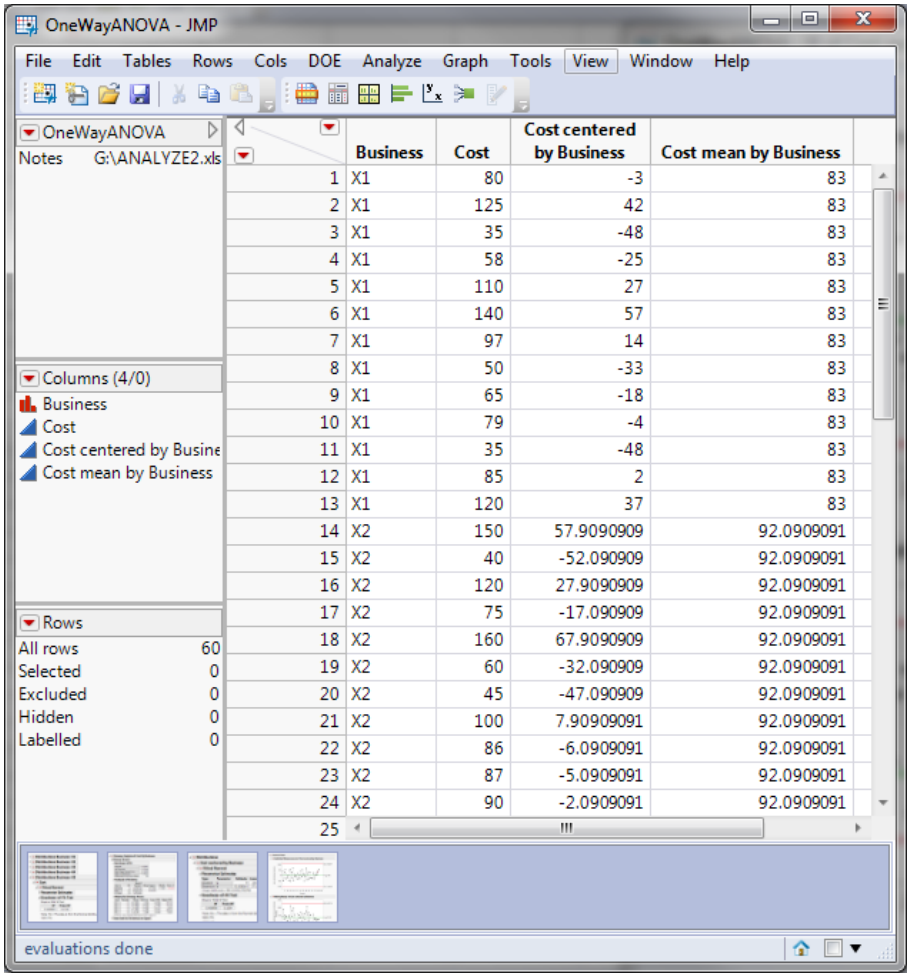


Fig. 3.56 ANOVA Residuals and Predictions Values

Step 5: Test whether the residuals are normally distributed with mean equal zero

1. Click Analyze -> Distribution
2. Select “Cost Centered by Business” as “Y, Columns”

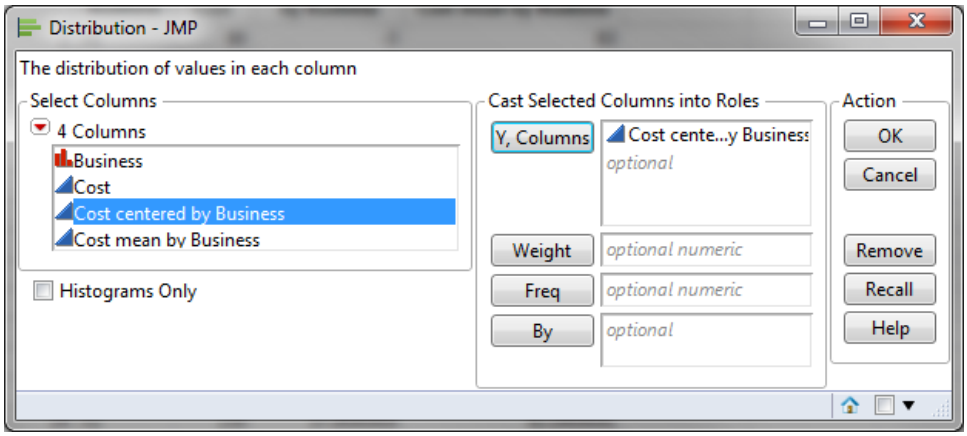


Fig. 3.57 Distribution selections

3. Click "OK"
4. Click on the red triangle button next to "Cost Centered by Business" in the Distribution page
5. Click Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

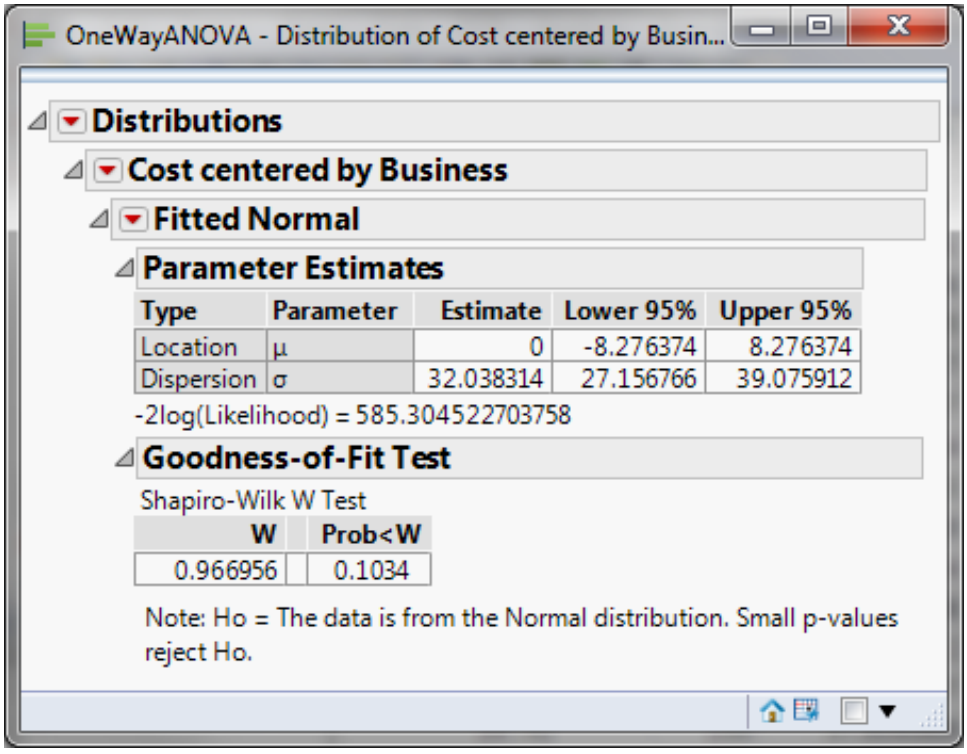


Fig. 3.58 ANOVA Residuals Results

The p-value of the normality test is 0.1034, greater than the alpha level (0.05), and we conclude that the residuals are normally distributed. The mean of the residuals is 0.0000.

Step 6: Check whether the residuals are independent of each other.

- If the data are in time order, then we can plot it in an IR chart to check the independence. When no data points on the IR chart fail any tests, the residuals are independent of each other.
- If the data are not in time order, the IR chart cannot deliver reliable conclusion on the independence.

1. Click Analyze -> Quality & Process - > Control Chart -> IR
2. Select "Cost centered by Business" as "Process"

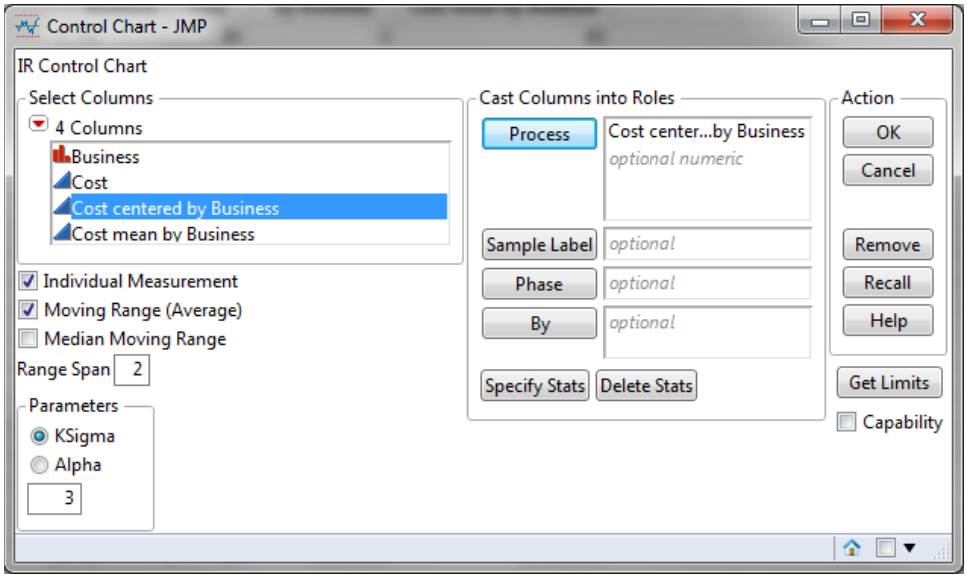


Fig. 3.59 Control Chart selections

3. Click "OK"
4. Click on the red triangle button next to "Individual Measurements of Cost centered by Business"
5. Select "Tests" -> "All Tests"

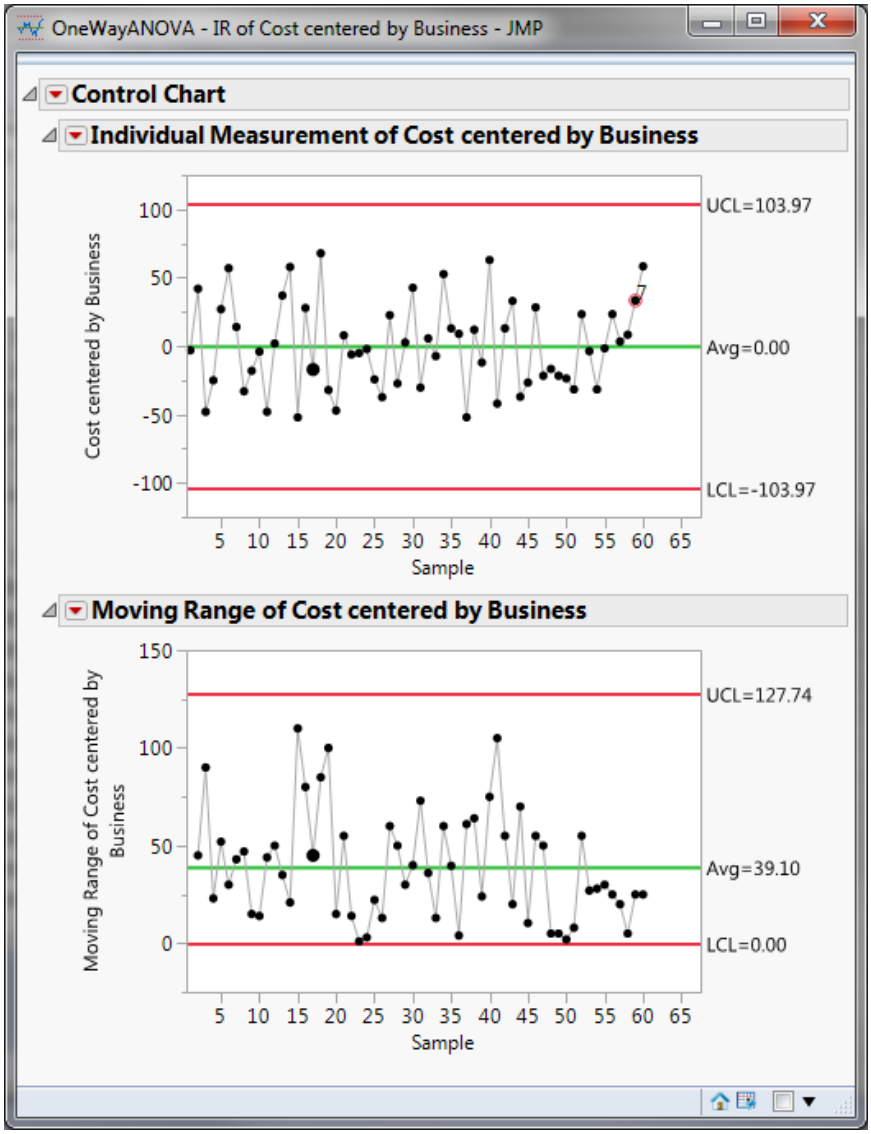


Fig. 3.60 Individuals-Moving Range Chart output

If the residuals are in time order, we can plot IR charts to check the independence. When no data points on the IR charts fail any tests, the residuals are independent of each other. If the residuals are not in time order, the IR charts cannot deliver reliable conclusion on the independence.

Step 7: Plot residuals versus fitted values and check whether there is any systematic pattern.

1. Click Analyze-> Fit Y by X
2. Select “Cost centered by Business” as “Y, Columns”
3. Select “Cost mean by Business” as “X”

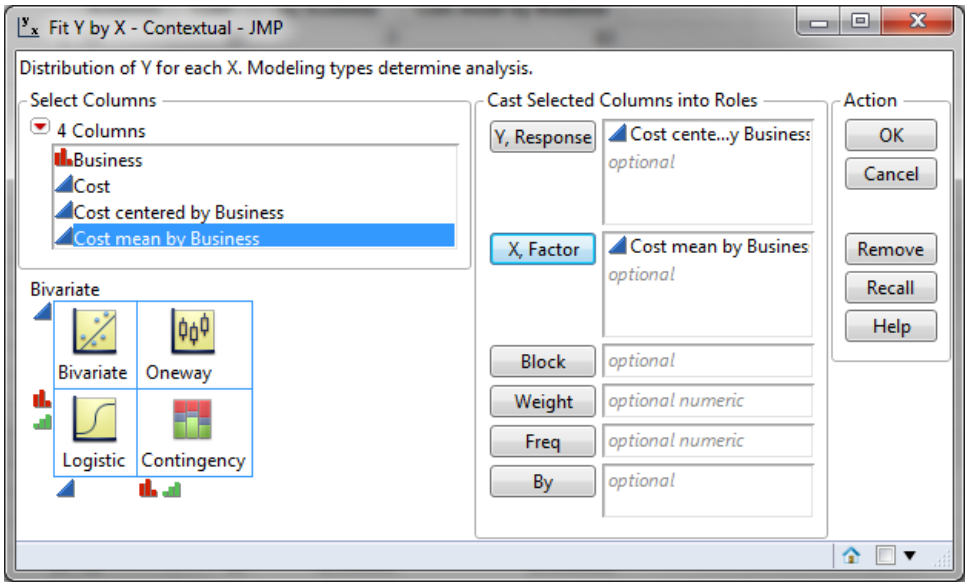


Fig. 3.61 Plot Variables

4. Click "OK"

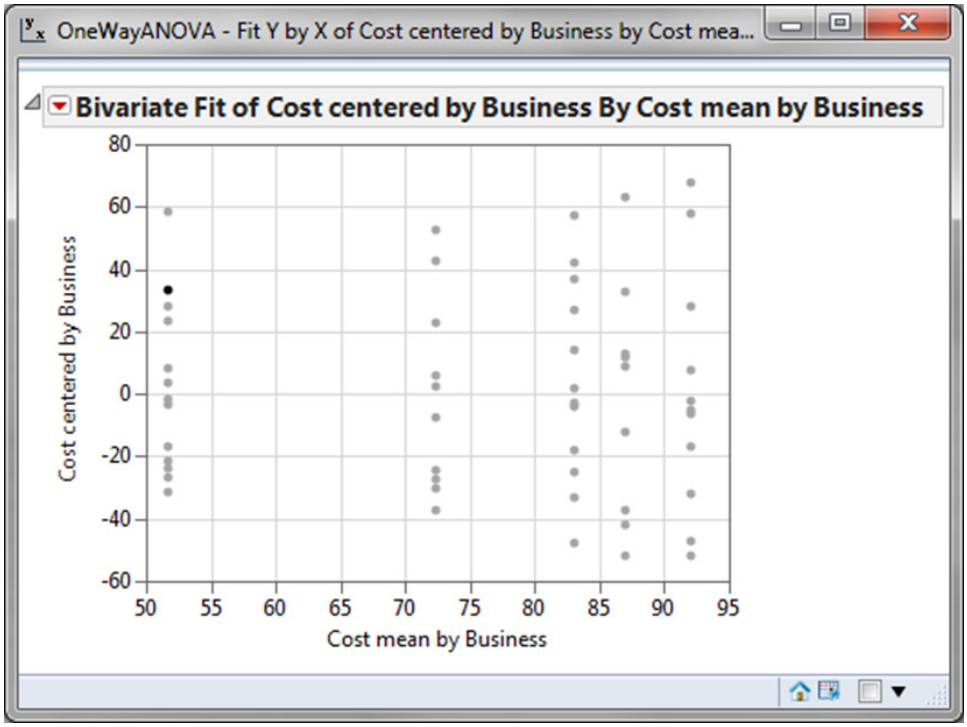


Fig. 3.62 ANOVA Output in JMP

The data points are spread out evenly at any of the five levels. Therefore, we can claim that the residuals have equal variances across all five levels.

3.5 HYPOTHESIS TESTING NON-NORMAL DATA

The tests we have discussed up to this point all apply when we are working with data that follow normal distributions. So, what do we do if that is not the case? The following pages describe tests available to us when the data do not follow the normal distribution.

3.5.1 *Mann–Whitney*

What is the Mann–Whitney Test?

The *Mann–Whitney test* (also called Mann–Whitney U test or Wilcoxon rank-sum test) is a statistical hypothesis test to compare the medians of two populations that are not normally distributed. In a non-normal distribution, the median is the better representation of the center of the distribution.

- Null Hypothesis (H_0): $\eta_1 = \eta_2$
- Alternative Hypothesis (H_a): $\eta_1 \neq \eta_2$

Where:

- η_1 is the median of one population
- η_2 is the median of the other population.

The null hypothesis is that the medians are equal, and the alternative is that they are not.

Mann–Whitney Test Assumptions

- The sample data drawn from the populations of interest are unbiased and representative.
- The data of both populations are continuous or ordinal when the spacing between adjacent values is not constant. (Reminder: Ordinal data—A set of data is said to be ordinal if the values can be ranked or have a rating scale attached. You can count and order, but not measure, ordinal data.)
- The two populations are independent to each other.
- The Mann–Whitney test is robust for the non-normally distributed population.
- The Mann–Whitney test can be used when shapes of the two populations' distributions are different.

How Mann–Whitney Test Works

Step 1: Group the two samples from two populations (sample 1 is from population 1 and sample 2 is from population 2) into a single data set and then sort the data in ascending order ranked from 1 to n, where n is the total number of observations.

Step 2: Add up the ranks for all the observations from sample 1 and call it R_1. Add up the ranks for all the observations from sample 2 and call it R_2.

Step 3: Calculate the test statistics

$$U = \min(U_1, U_2)$$

Where:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

and where:

- n_1 and n_2 are the sample sizes
- R_1 and R_2 are the sum of ranks for observations from sample 1 and 2 respectively.

Step 4: Make a decision on whether to reject the null hypothesis.

- Null Hypothesis (H_0): $\eta_1 = \eta_2$
- Alternative Hypothesis (H_a): $\eta_1 \neq \eta_2$

If both of the sample sizes are smaller than 10, the distribution of U under the null hypothesis is tabulated.

- The test statistic is U and, by using the Mann–Whitney table, we would find the p-value.
- If the p-value is smaller than alpha level (0.05), we reject the null hypothesis.
- If the p-value is greater than alpha level (0.05), we fail to reject the null hypothesis.

If both sample sizes are greater than 10, the distribution of U can be approximated by a normal distribution. In other words, $\frac{U-\mu}{\sigma}$ follows a standard normal distribution.

$$Z_{calc} = \frac{U - \mu}{\sigma}$$

Where:

$$\mu = \frac{n_1 n_2}{2}$$

$$\sigma = \sqrt{\frac{\sqrt{n_1 n_2 (n_1 + n_2 + 1)}}{12}}$$

If the sample sizes are greater than 10, then the distribution of U can be approximated by a normal distribution. The U value is then plugged into the formula seen here to calculate a Z statistic.

When $|Z_{calc}|$ is greater than Z value at α/2 level (e.g., when α = 5%, the z value we compare $|Z_{calc}|$ to is 1.96), we reject the null hypothesis.

Use JMP to Run a Mann–Whitney Test

Case study: We are interested in comparing customer satisfaction between two types of customers using a nonparametric (i.e., distribution-free) hypothesis test: Mann–Whitney test.

Data File: "Mann–Whitney.jmp"

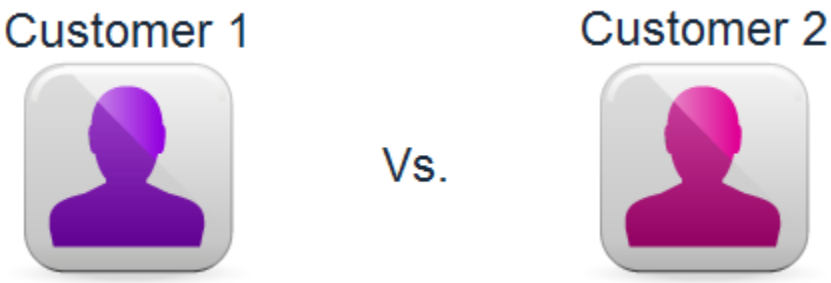


Fig. 3.63 Mann-Whitney Test

- Null Hypothesis (H_0): $\eta_1 = \eta_2$
- Alternative Hypothesis (H_a): $\eta_1 \neq \eta_2$

Steps to run a Mann–Whitney Test in JMP:

1. Click Analyze -> Fit Y by X
2. Select "Overall Satisfaction" as "Y, Response"
3. Select "Customer Type" as "X, Factor"

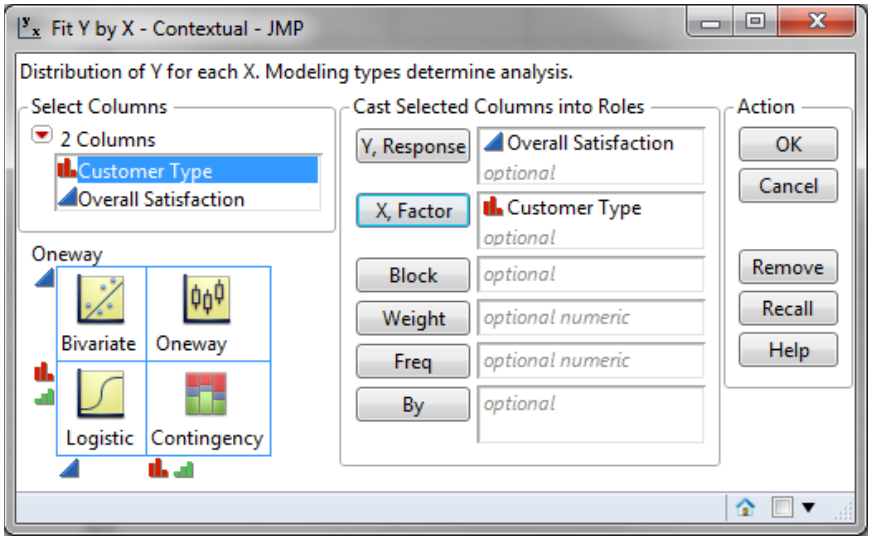


Fig. 3.64 Selections to run a Mann-Whitney Test

4. Click "OK"

5. Click on the red triangle button next to “One-Way Analysis of Overall Satisfaction by Customer Type”
6. Click Nonparametric -> Wilcoxon Test

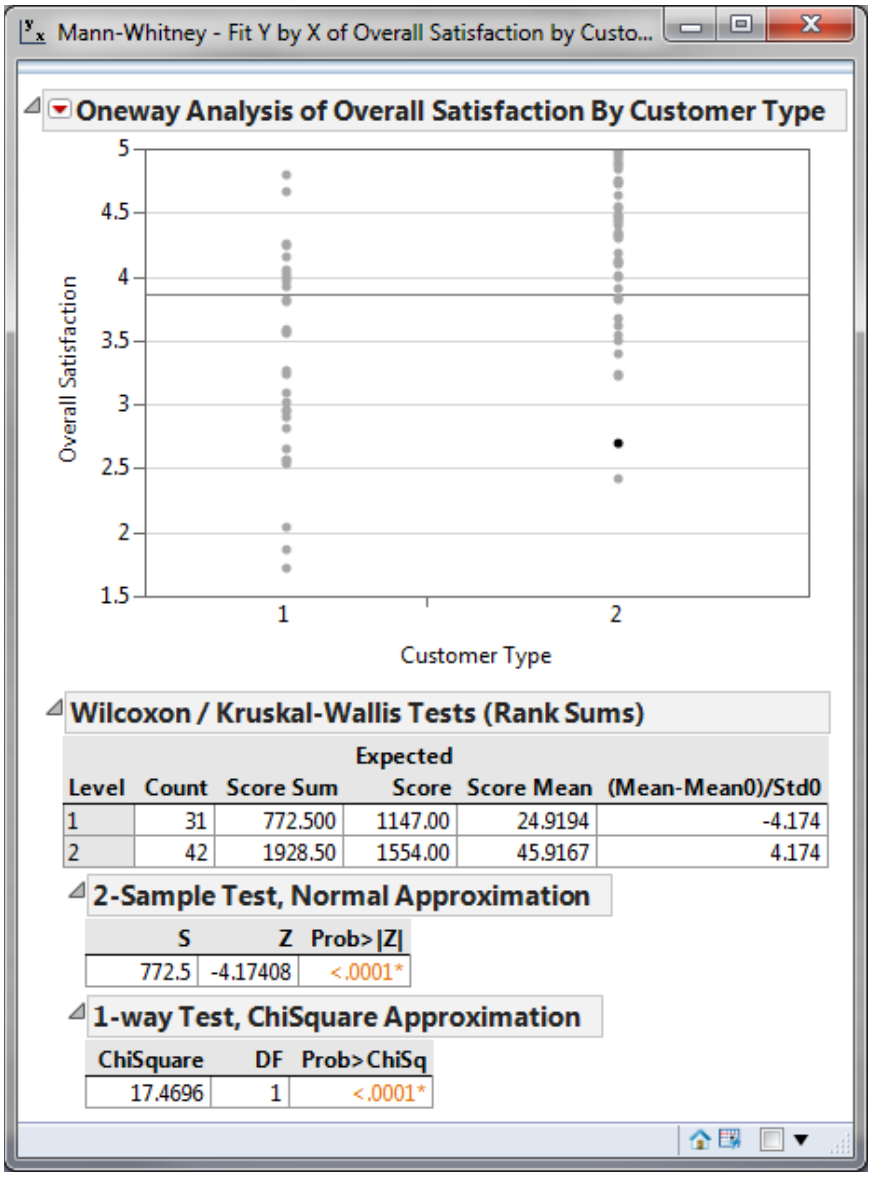


Fig. 3.65 Mann-Whitney Test Results

The p-value of the test is lower than alpha level (0.05); so we reject the null hypothesis and conclude that there is a statistically significant difference between the overall satisfaction medians of the two customer types.

The result of the test is boxed in: The p-value is lower than the alpha value of 0.05; therefore, we must reject the null hypothesis and claim that there is a significant difference between the median customer satisfaction levels of the two groups.

3.5.2 KRUSKAL–WALLIS

Kruskal–Wallis One-Way Analysis of Variance

The *Kruskal–Wallis one-way analysis of variance* is a statistical hypothesis test to compare the medians among more than two groups.

- Null Hypothesis (H_0): $\eta_1 = \eta_2 = \ldots = \eta_k$
- Alternative Hypothesis (H_a): at least one of the medians is different from others.

Where:

- η_i is the median of population i
- k is the number of groups of our interest.

It is an extension of the Mann–Whitney test. While the Mann–Whitney test allows us to compare the samples of two populations, the Kruskal–Wallis test allows us to compare the samples of more than two populations.

One key difference between this test and the Mann–Whitney test is the robustness of the test when the populations are not identically shaped. If this is the case, there is a different test, called Mood's median, which is more appropriate.

Kruskal–Wallis One-Way Analysis of Variance: Assumptions

- The sample data drawn from the populations of interest are unbiased and representative.
- The data of k populations are continuous or ordinal when the spacing between adjacent values is not constant.
- The k populations are independent to each other.
- The Kruskal–Wallis test is robust for the non-normally distributed population.

How Kruskal–Wallis One-Way ANOVA Works

The Kruskal–Wallis test works very similarly to the Mann–Whitney Test.

Step 1: Group the k samples from k populations (sample i is from population i) into one single data set and then sort the data in ascending order ranked from 1 to N, where N is the total number of observations across k groups.

Step 2: Add up the ranks for all the observations from sample i and call it r_i, where i can be any integer between 1 and k.

Step 3: Calculate the test statistic

$$T = (N-1)\frac{\sum_{i=1}^{k} n_i(\bar{r}_i - \bar{r})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(r_{ij} - \bar{r})^2}$$

Where:

- k is the number of groups
- n_i is the sample size of sample i
- N is the total number of all the observations across k groups
- r_{ij} is the rank (among all the observations) of observation j from group i.

Step 4: Make a decision of whether to reject the null hypothesis.

- Null Hypothesis (H_0): $\eta_1 = \eta_2 = \ldots = \eta_k$
- Alternative Hypothesis (H_a): at least one of the medians is different from others.

The test statistic follows chi-square distribution when the null hypothesis is true. If T is greater than χ^2_{k-1} (the critical chi-square statistic), we reject the null and claim there is at least one median statistically different from other medians. If T is smaller than χ^2_{k-1}, we fail to reject the null and claim the medians of k groups are equal.

Use JMP to Run a Kruskal–Wallis One-Way ANOVA

Case study: We are interested in comparing customer satisfaction among three types of customers using a nonparametric (i.e., distribution-free) hypothesis test: Kruskal–Wallis one-way ANOVA.

Data File: "Kruskal–Wallis.jmp"

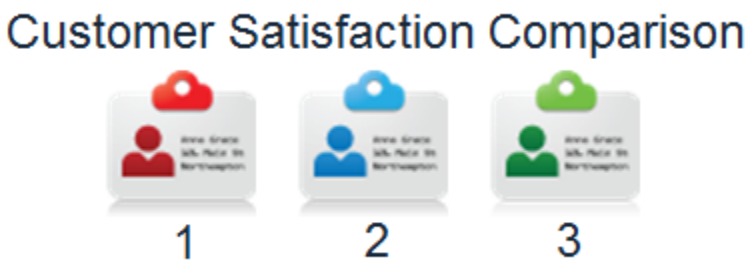


Fig. 3.66 Kruskal-Wallis Test

- Null Hypothesis (H_0): $\eta_1 = \eta_2 = \eta_3$
- Alternative Hypothesis (H_a): at least one of the customer types has different overall satisfaction levels from the others.

Steps to run a Kruskal–Wallis One-Way ANOVA in JMP:

1. Click Analyze -> Fit Y by X
2. Select "Overall Satisfaction" as "Y, Response"
3. Select "Customer Type" as "X, Factor"

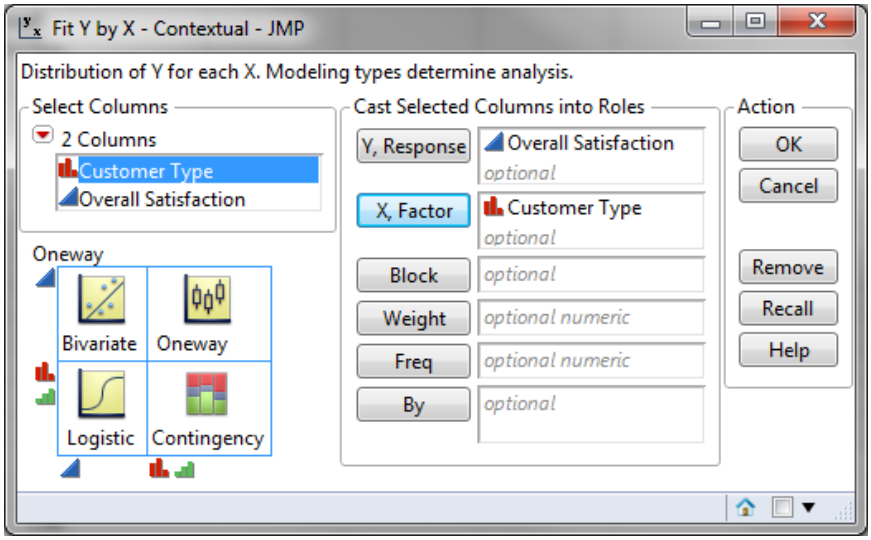


Fig. 3.67 Distributions for Y and X

4. Click "OK"
5. Click on the red triangle button next to "One-Way Analysis of Overall Satisfaction by Customer Type"
6. Click Nonparametric -> Wilcoxon Test

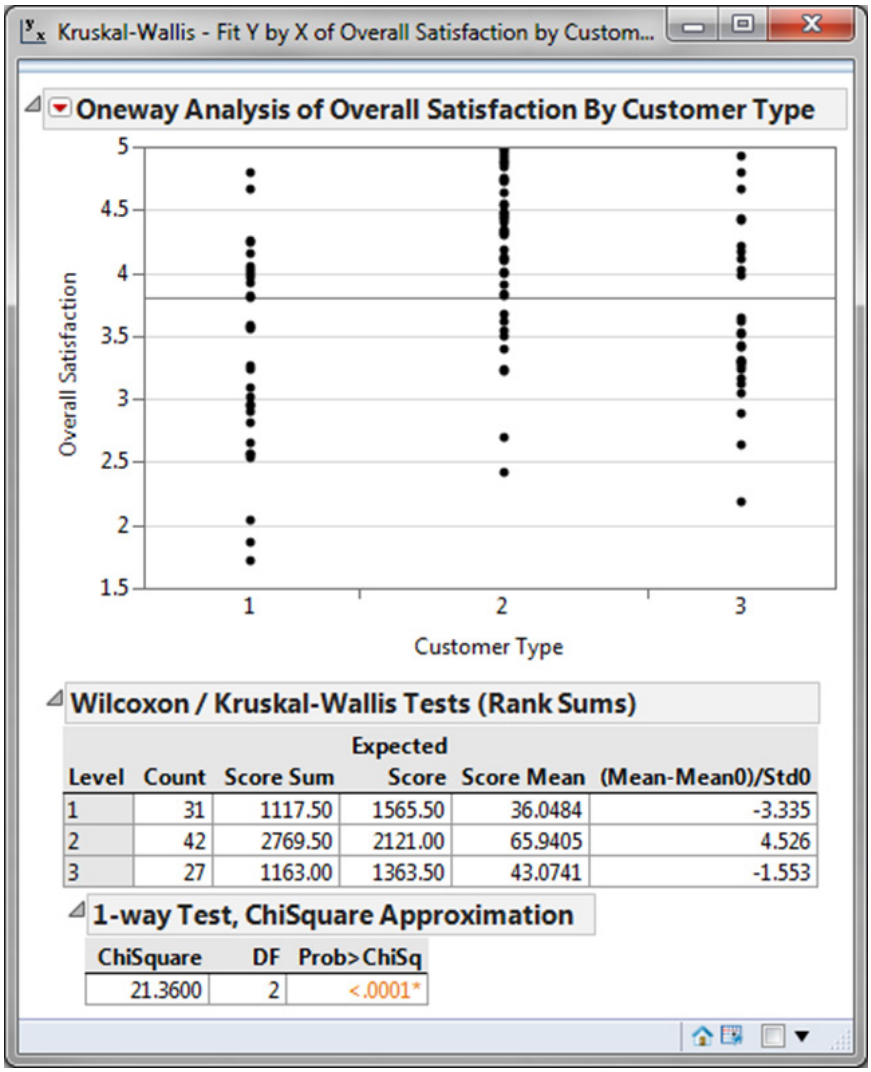


Fig. 3.68 Kruskal-Wallis Results in JMP

The p-value of the test is lower than alpha level (0.05), and we reject the null hypothesis and conclude that at least the overall satisfaction median of one customer type is statistically different from the others. The result of the test is boxed in: The p-value is lower than the alpha value of 0.05; therefore, we must reject the null hypothesis and claim that at least one of the customer types has a different level of satisfaction than the others.

3.5.3 Mood's Median

What is Mood's Median Test?

Mood's median test is a statistical test to compare the medians of two or more populations.

- Null Hypothesis (H_0): $\eta_1 = \ldots = \eta_k$
- Alternative Hypothesis (H_a): At least one of the medians is different from the others.

The symbol *k* is the number of groups of our interest and is equal to or greater than two. Mood's median is an alternative to Kruskal–Wallis. For the data with outliers, Mood's median test is more robust than the Kruskal–Wallis test. It is the extension of one sample sign test. In a one-sample sign test, the null hypothesis is that the probability of a random value from the population being above the specified value is equal to the probability of a random value being below the specified value.

Mood's Median Test Assumptions

- The sample data drawn from the populations of interest are unbiased and representative.
- The data of k populations are continuous or ordinal when the spacing between adjacent values is not constant.
- The k populations are independent to each other.
- The distributions of k populations have the same shape.
- Mood's median test is robust for non-normally distributed populations.
- Mood's median test is robust for data with outliers.

How Mood's Median Test Works

Step 1: Group the k samples from k populations (sample i is from population i) into one single data set and get the median of this combined data set. Step 2: Separate the data in each sample into two groups. One consists of all the observations with values higher than the grand median and the other consists of all the observations with values lower than the grand median.

Step 2: Run a Pearson's chi-square test to determine whether to reject the null hypothesis.

- Null Hypothesis (H_0): $\eta_1 = \ldots = \eta_k$
- Alternative Hypothesis (H_a): At least one of the medians is different from the others.
- If χ^2_{calc} is greater than χ^2_{crit} , we reject the null hypothesis and claim that at least one median is different from the others.
- If χ^2_{calc} is smaller than χ^2_{crit} , we fail to reject the null hypothesis and claim that the medians of k populations are not statistically different.

Use JMP to Run a Mood's Median Test

Case study: We are interested in comparing customer satisfaction among three types of customers using a nonparametric (i.e., distribution-free) hypothesis test: Mood's median test.

Data File: "MedianTest.jmp"

- Null Hypothesis (H_0): $\eta_1 = \eta_2 = \eta_3$
- Alternative Hypothesis (H_a): At least one customer type has different overall satisfaction from the others.

Steps to run a Mood's median test in JMP:

1. Click Analyze -> Fit Y by X
2. Select "Overall Satisfaction" as "Y, Response"
3. Select "Customer Type" as "X, Factor"

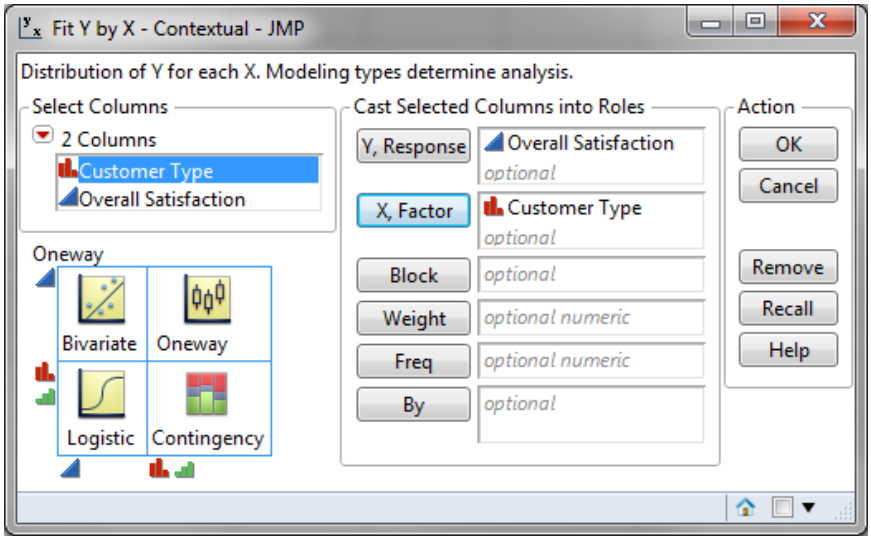


Fig. 3.69 Distributions for Y and X

4. Click "OK"
5. Click on the red triangle button next to "One-Way Analysis of Overall Satisfaction by Customer Type"
6. Click Nonparametric -> Median Test

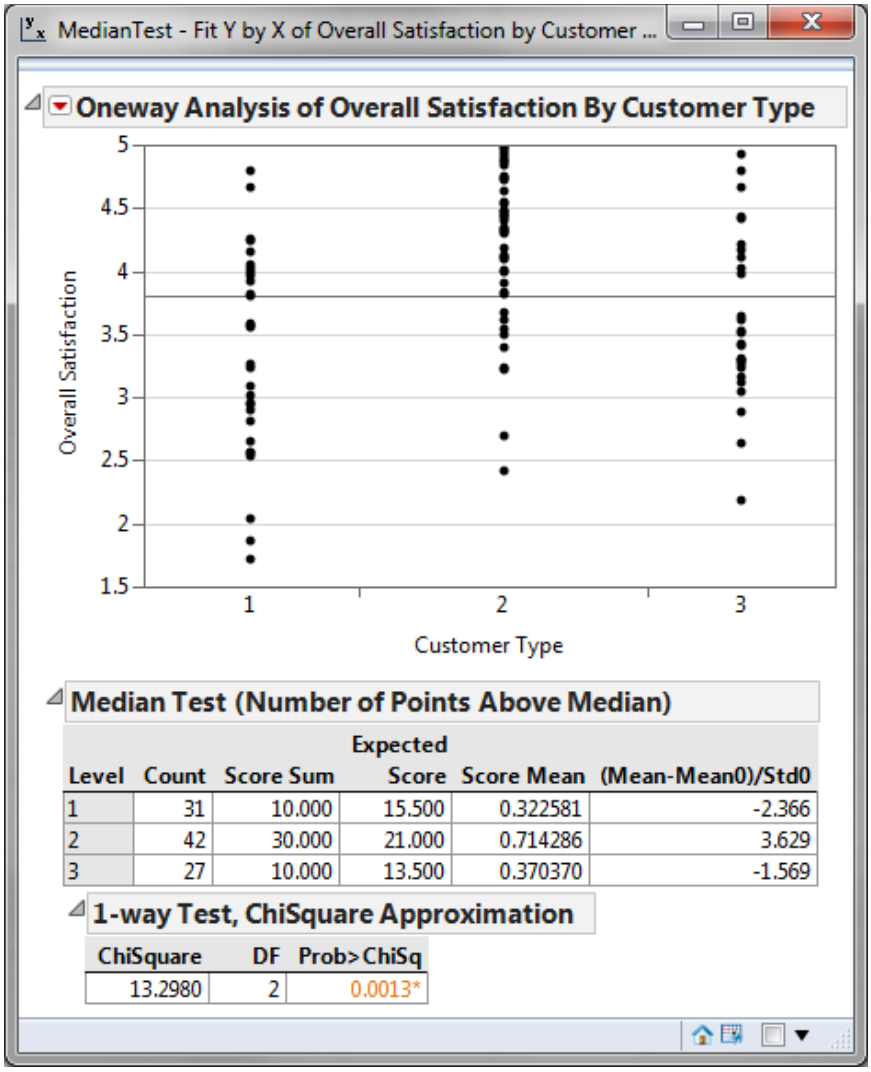


Fig. 3.70 Output from Mood's Median Test

The p-value of the test is lower than alpha level (0.05), and we reject the null hypothesis and conclude that at least the overall satisfaction median of one customer type is statistically different from the others. Since the p-value is less than 0.05, we reject the null hypothesis and claim that at least one customer group has a different level of satisfaction than the others.

3.5.4 FRIEDMAN

What is the Friedman Test?

The *Friedman test* is a hypothesis test used to detect the differences in the medians of various groups across multiple attempts. It is a non-parametric test developed by economist Milton Friedman.

- Null Hypothesis (H_0): The treatments have identical effects (i.e., $\eta_1 = \ldots = \eta_k$).
- Alternative Hypothesis (H_a): At least one of the treatments has different effects from the others (i.e., at least one of the medians is statistically different from others).

It is used as an alternative of the parametric repeated measures ANOVA when the assumption of normality or variance equality is not met.

Friedman Test Assumptions

- The sample data drawn from the populations of interest are unbiased and representative.
- The data are continuous or ordinal when the spacing between adjacent values is not constant.
- Results in one block are independent of the results in another.
- The Friedman test is robust for the non-normally distributed population.
- The Friedman test is robust for populations with unequal variances.

How Friedman Test Works

Step 1: Organize the data into a tabular view with n rows indicating the blocks and k columns indicating the treatments. Each observation x_{ij} is filled into the intersection of a specific block i and a specific treatment j.

Step 2: Calculate the ranks of each observation within each block.

Step 3: Replace the values in the table created in step 1 with the order r_{ij} (within a block) of the corresponding observation x_{ij}.

Step 4: Calculate the test statistic

$$Q = \frac{n\sum_{j=1}^{k}(\bar{r}_j - \bar{r})^2}{\frac{1}{n(k-1)}\sum_{i=1}^{n}\sum_{j=1}^{k}(r_{ij} - \bar{r})^2}$$

Where:

$$\bar{r}_j = \frac{1}{n}\sum_{i=1}^{n} r_{ij}$$

$$\bar{r} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} r_{ij}$$

Step 5: Make a decision of whether to reject the null hypothesis.

When n > 15 or k > 4, the test statistic Q follows a chi-square distribution if the null hypothesis is true and the p-value is $P\left(\chi^2_{k-1} \geq Q\right)$.

When n < 15 and k < 4, the test statistic Q does not approximate a chi-square distribution and the p-value can be obtained from tables of Q for the Friedman test. If p-value > alpha level (0.05), we fail to reject the null.

Friedman Test Examples

A number of n water testers judge the quality of k different water samples, each of which is from a distinct water source. We will apply a Friedman test to determine whether the water qualities of the k sources are the same. There are n blocks and k treatments in this experiment. One tester's decision would not have any influence on other testers. When running the experiment, each tester judges the water in a random sequence.

3.5.5 One Sample Sign

What is the One Sample Sign Test?

The *one sample sign test* is a hypothesis test to compare the median of a population with a specified value. However, with this test we are comparing the median of the population instead of the mean.

- Null Hypothesis (H_0): $\eta = \eta_0$
- Alternative Hypothesis (H_a): $\eta \neq \eta_0$

It is an alternative test of one sample t-test when the distribution of the data is non-normal. It is robust for the data with non-symmetric distribution.

One Sample Sign Test Assumptions

- The sample data drawn from the population of interest are unbiased and representative.
- The data are continuous or ordinal when the spacing between adjacent values is not constant.
- The one sample sign test is robust for the non-normally distributed population.
- The one sample sign test does not have any assumptions on the distribution. It is a distribution-free test.

Assumptions for this test are very similar to those of other non-parametrics; however, a key difference is that there are no assumptions about the shape of the distribution.

How the One Sample Sign Test Works

Step 1: Separate the sample set of data into two groups: one with values greater than and the other with values less than the hypothesized median η_0. Count the number of observations in each group.

Step 2: Calculate the test statistic.

If the null hypothesis is true, the number of observations in each group should not be significantly different from half of the total sample size. The test statistic follows a binomial distribution when the null is true. When n is large, we use the normal distribution to approximate the binomial distribution.

$$Z_{calc} = \frac{n_+ - np}{\sqrt{np(1-p)}}$$

Where:

- n_+ is the number of observations with values greater than the hypothesized median
- n is the total number of observations
- p is 0.5.

Step 3: Make a decision on whether to reject the null hypothesis. If the $|Z_{calc}|$ is smaller than Z_{crit}, we fail to reject the null hypothesis and claim that there is no significant difference between the population median and the hypothesized median.

3.5.6 One Sample Wilcoxon

What is the One Sample Wilcoxon Test?

The *one sample Wilcoxon test* is a hypothesis test to compare the median of one population with a specified value.

- Null Hypothesis (H_0): $\eta = \eta_0$
- Alternative Hypothesis (H_a): $\eta \neq \eta_0$

It is an alternative test of one sample t-test when the distribution of the data is non-normal. It is more powerful than one sample sign test but it assumes the distribution of the data is symmetric.

One Sample Wilcoxon Test Assumptions

- The sample data drawn from the population of interest are unbiased and representative.
- The data are continuous or ordinal when the spacing between adjacent values is not constant.
- The distribution of the data is symmetric about a median.
- The one sample Wilcoxon test is robust for the non-normally distributed population.

The assumptions are similar to other non-parametric tests, again the difference about the symmetry of the distribution.

How the One Sample Wilcoxon Test Works

Step 1: Create the following columns one by one:

- Column 1: All the raw observations (X)
- Column 2: The differences between each observation value and the hypothesized median ($X - \eta_0$)
- Column 3: The signs (+ or –) of column 2
- Column 4: The absolute value of column 2
- Column 5: The ranks of each item in column 4 in ascending order
- Column 6: The product of column 3 and column 5

Step 2: Calculate the test statistic W_{calc}, which is the sum of all the non-negative values in column 6.

Step 3: Make a decision on whether to reject the null hypothesis. Use the table of critical values for the Wilcoxon test to get the W_{crit} with predetermined alpha level and number of observations.

If the W_{calc} is smaller than the W_{crit}, we fail to reject the null hypothesis and claim that there is no significant difference between the population median and the hypothesized median. To calculate the W statistic, sum all of the non-negative values in column 6. To assess the hypothesis, the W statistic is compared against the critical W value.

Use JMP to Run a One Sample Wilcoxon Test

Case study: We are interested in comparing the overall satisfaction of customer type 1 against a specified benchmark satisfaction (3.5) using a nonparametric (i.e., distribution-free) hypothesis test: one sample Wilcoxon test.

Data File: "OneSampleWilcoxon.jmp"

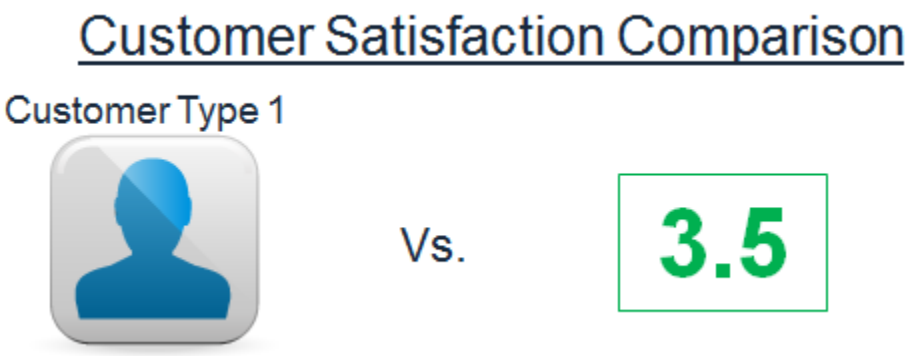


Fig. 3.71 One Sample Wilcoxon

- Null Hypothesis (H_0): $\eta_1 = 3.5$
- Alternative Hypothesis (H_a): $\eta_1 \neq 3.5$

Steps to run a one sample Wilcoxon test in JMP:

1. Click Analyze -> Distribution
2. Select "Overall Satisfaction" as "Y, Columns"

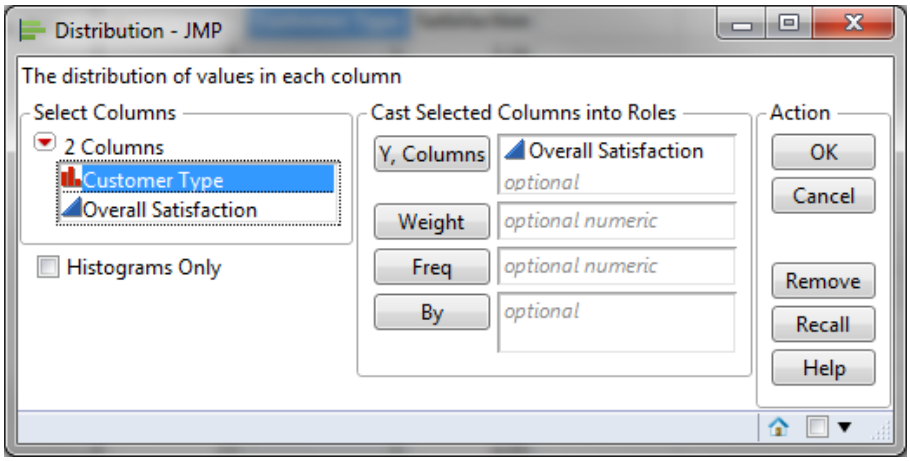


Fig. 3.72 Distribution selections

3. Click OK
4. Click on the red triangle button next to "Overall Satisfaction"
5. Click "Test Mean"
6. In the new pop-up window, enter the specified median (3.5) and check the box "Wilcoxon Signed Rank"

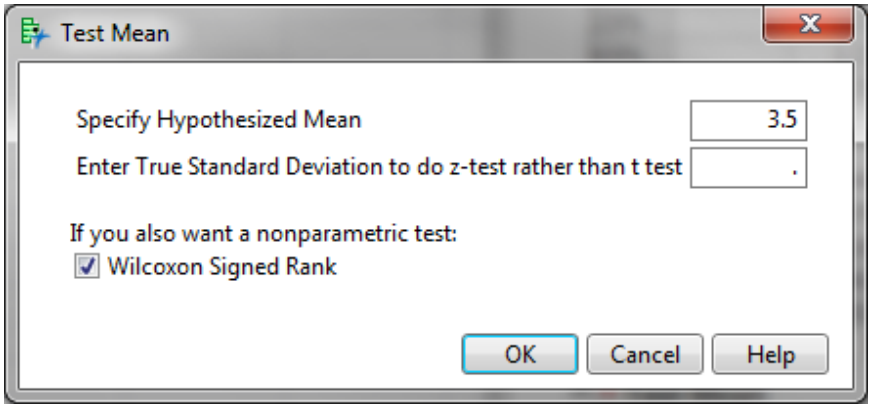


Fig. 3.73 Mean Specification window and Wilcoxon Signed Rank check box

7. Click OK

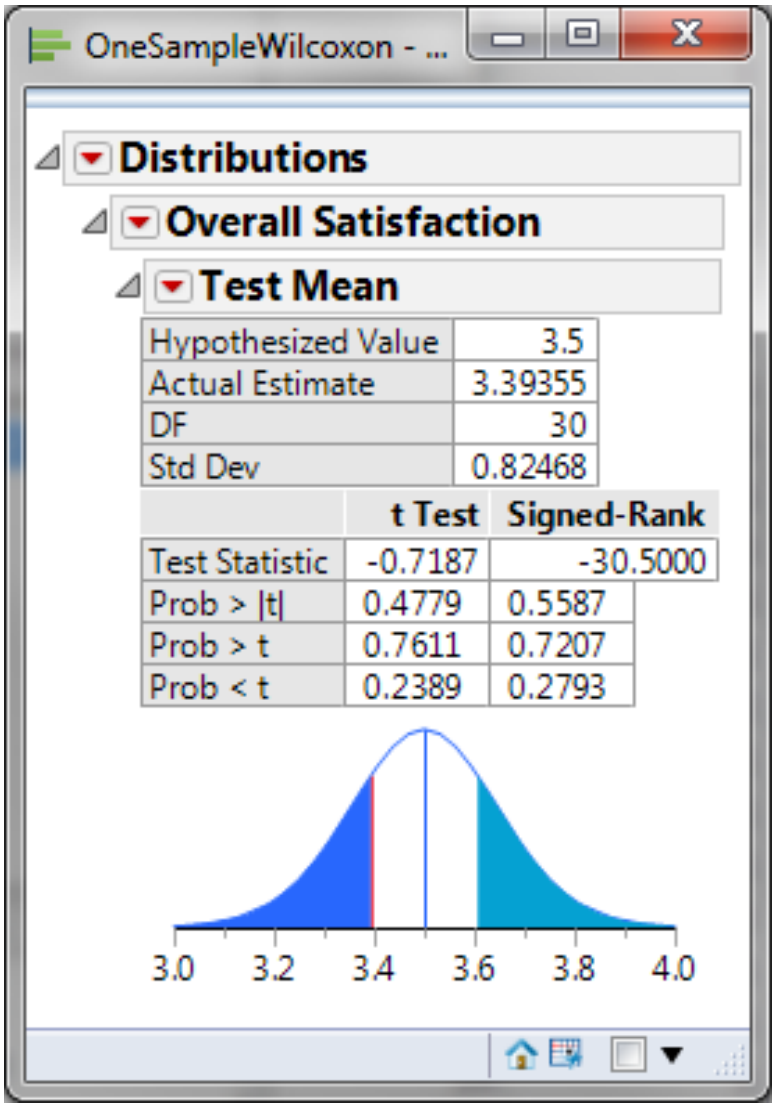


Fig. 3.74 One Sample Wilcoxon Test Results

The p-value of the one sample Wilcoxon test is 0.5587, higher than the alpha level (0.05), and we fail to reject the null hypothesis. There is not any statistically significant difference between the overall satisfaction of customer type 1 and the benchmark satisfaction level. In this example, the software returns a p-value of 0.5587, much higher than the alpha value of 0.05. We fail to reject the null and claim that the population median is not statistically different than the specified value of 3.5.

3.5.7 One and Two Sample Proportion

What is the One Sample Proportion Test?

One sample proportion test is a hypothesis test to compare the proportion of one certain outcome (e.g., the number of successes per the number of trials, or the number of defects per the total number of opportunities) occurring in a population following the binomial distribution with a specified proportion.

- Null Hypothesis (H_0): $p = p_0$
- Alternative Hypothesis (H_a): $p \neq p_0$

One Sample Proportion Test Assumptions

- The sample data drawn from the population of interest are unbiased and representative.
- There are only two possible outcomes in each trial: success/failure, yes/no, and defective/non-defective etc.
- The underlying distribution of the population is binomial distribution.
- When np ≥ 5 and np(1 – p) ≥ 5, the binomial distribution can be approximated by the normal distribution.

How the One Sample Proportion Test Works

When np ≥ 5 and np(1 – p) ≥ 5, we use normal distribution to approximate the underlying binomial distribution of the population.

Test Statistic

$$Z_{calc} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Where:

- $\hat{p}$ is the observed probability of one certain outcome occurring?
- p_0 is the hypothesized probability
- n is the number of trials.

When $|Z_{calc}|$ (absolute value of the calculated test statistic) is smaller than Z_{crit} (critical value), we fail to reject the null hypothesis and claim that there is no statistically significant difference between the population proportion and the hypothesized proportion.

Use JMP to Run a One Sample Proportion Test

Case study: We are interested in comparing the exam pass rate of a high school this month against a specified rate (70%) using a nonparametric (i.e., distribution-free) hypothesis test: one sample proportion test.

Data File: "OneSampleProportion.jmp"

- Null Hypothesis (H_0): p = 70%
- Alternative Hypothesis (H_a): p ≠ 70%

Steps to run a one sample proportion test in JMP:

1. Click Analyze -> Distribution
2. Select “Result” as “Y, Columns”
3. Select “Count” as “Freq”

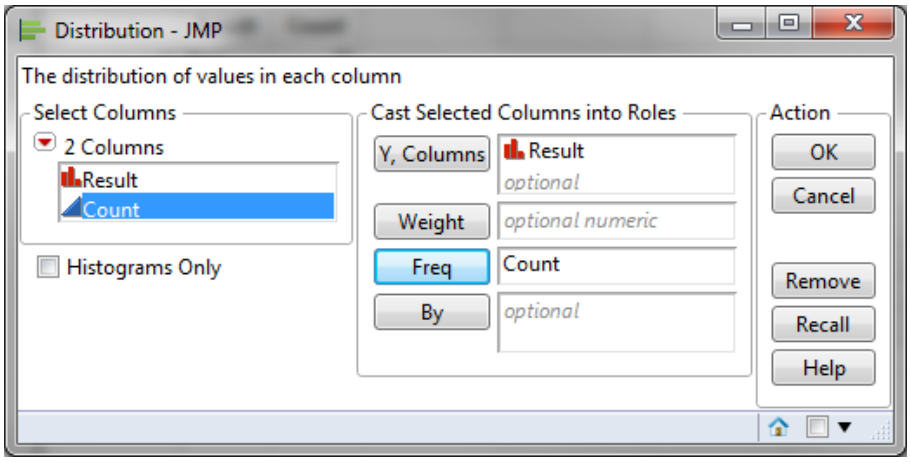


Fig. 3.75 Distribution selections

4. Click “OK”
5. Click on the red triangle button next to “Result”
6. Click “Test Probabilities”
7. In the new pop-up window, enter the hypothesized passing rate and failing rate.
8. Select the alternative hypothesis. In this case, we select “probabilities not equal to hyothesized value (two-sided chi-square test)

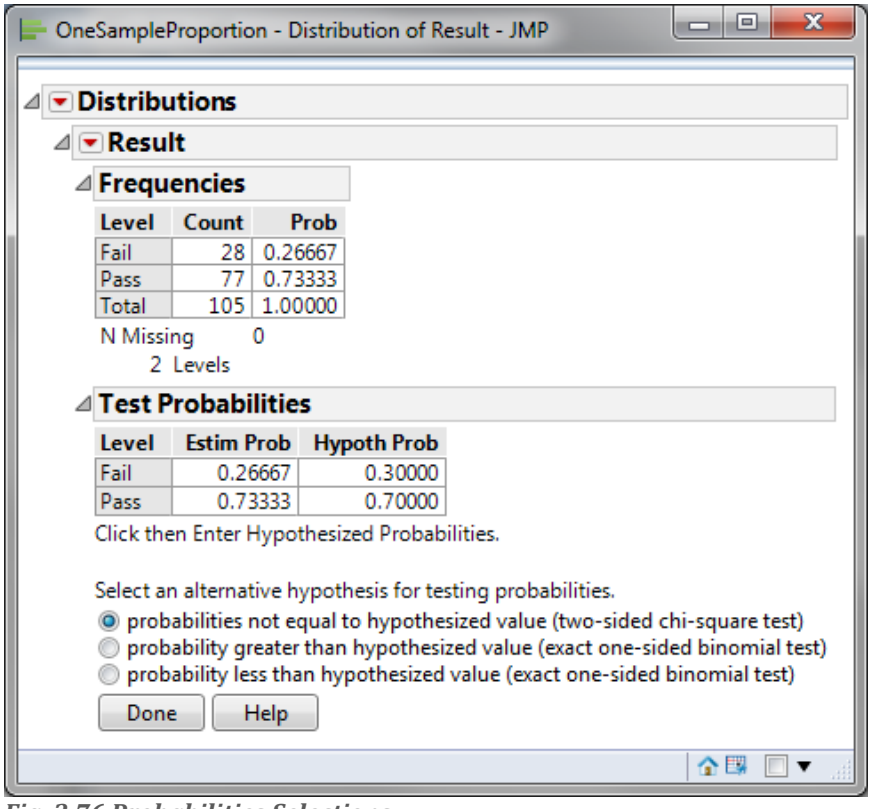


Fig. 3.76 Probabilities Selections

9. Click "Done"

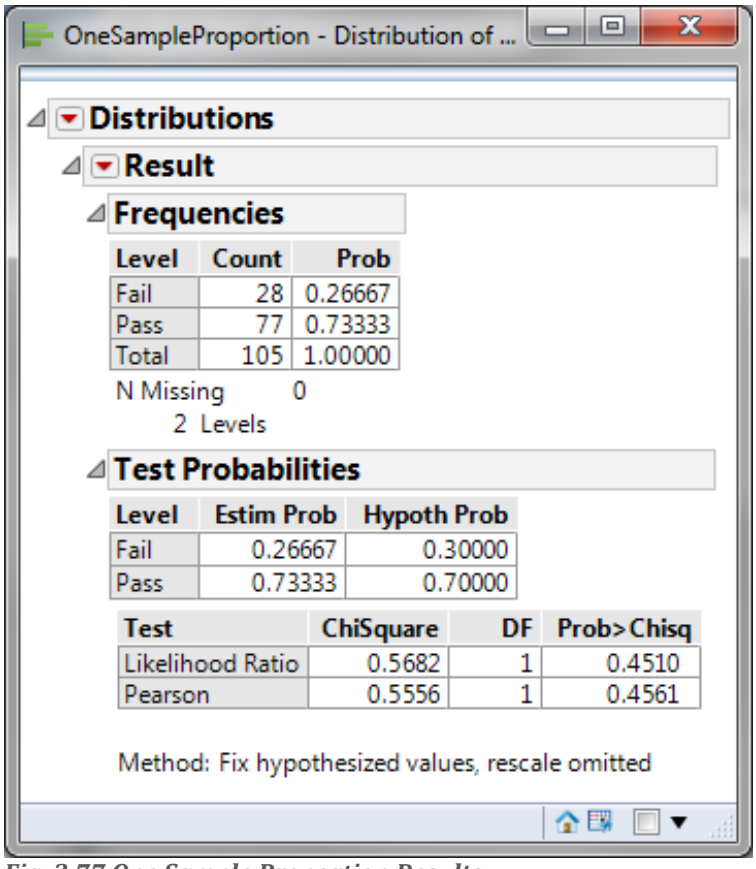


Fig. 3.77 One Sample Proportion Results

The p-value of the one sample proportion test is 0.460, greater than the alpha level (0.05), and we fail to reject the null hypothesis. We conclude that the exam pass rate of the high school this month is not statistically different from 70%. In the output from the test we see the p-value is higher than the alpha value of 0.05; therefore, we fail to reject the null hypothesis and claim there is not a difference between the schools exam pass rate and 70%.

What is the Two Sample Proportion Test?

The *two-sample proportion test* is a hypothesis test to compare the proportions of one certain event occurring in two populations following the binomial distribution.

- Null Hypothesis (H_0): $p_1 = p_2$
- Alternative Hypothesis (H_a): $p_1 \neq p_2$

Two-Sample Proportion Test Assumptions

- The sample data drawn from the populations of interest are unbiased and representative.
- There are only two possible outcomes in each trial for both populations: success/failure, yes/no, and defective/non-defective etc.
- The underlying distributions of both populations are binomial distribution.

- When np ≥ 5 and np(1 – p) ≥ 5, the binomial distribution can be approximated by the normal distribution.

How the Two-Sample Proportion Test Works

When np ≥ 5 and np(1 – p) ≥ 5, we use normal distribution to approximate the underlying binomial distributions of the populations.

Test Statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Where:

$$\hat{p}_0 = \frac{x_1 + x_2}{n_1 + n_2}$$

and where:

- $\hat{p}_1$ and $\hat{p}_2$ are the observed proportions of events in the two samples
- n_1 and n_2 is the number of trials in the two samples respectively
- x_1 and x_2 is the number of events in the two samples respectively.

When $|Z_{calc}|$ is smaller than Z_{crit}, we fail to reject the null hypothesis.

Use JMP to Run a Two-Sample Proportion Test

Case study: We are interested in comparing the exam pass rates of a high school in March and April using a nonparametric (i.e., distribution-free) hypothesis test: two sample proportion test.

Data File: "TwoSampleProportion.jmp"

- Null Hypothesis (H_0): $p_{March} = p_{April}$
- Alternative Hypothesis (H_a): $p_{March} \neq p_{April}$

Steps to run a two-sample proportion test in JMP:

1. Click Analyze -> Fit Y by X
2. Select "Results" as "Y, Responses"
3. Select "Month" as "X, Factor"

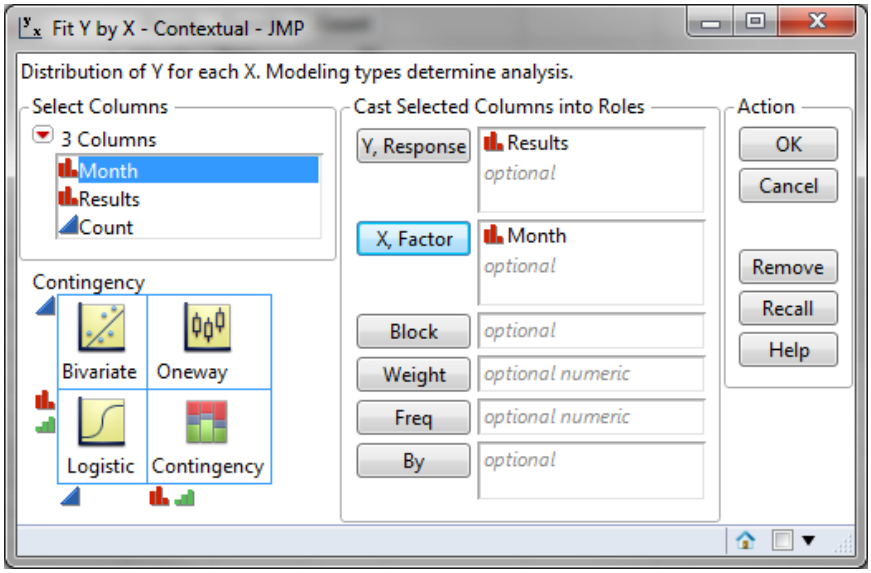


Fig. 3.78 Distribution for X and Y

4. Click "OK"

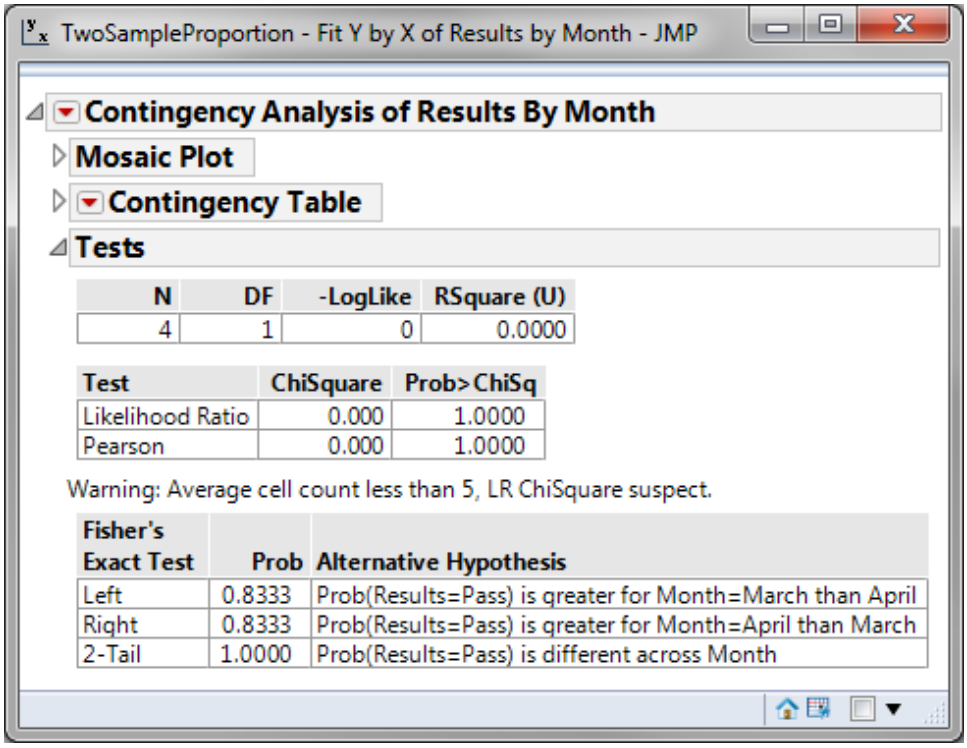


Fig. 3.79 Two Sample Proportion Test Results

The p-value of the two sample proportion test is 0.849, greater than the alpha level (0.05), and we fail to reject the null hypothesis. We conclude that the exam pass rates of the high school in March and April are not statistically different. In the output from the test we see the p-value is higher than the alpha value of 0.05; therefore, we fail to reject the null hypothesis and claim there is not a difference between the school's exam pass rate in March and April.

3.5.8 Chi-Squared (Contingency Tables)

What is the Chi-Square Test?

We have looked at hypothesis tests to analyze the proportion of one population vs. a specified value, and the proportions of two populations, but what do we do if we want to analyze more than two populations? A *chi-square test* is a hypothesis test in which the sampling distribution of the test statistic follows a chi-square distribution when the null hypothesis is true. There are multiple chi-square tests available and in this module, we will cover the Pearson's chi-square test used in contingency analysis.

- Null Hypothesis (H_0): $p_1 = p_2 = \ldots = p_k$
- Alternative Hypothesis (H_a): At least one of the proportions is different from others.

The symbol k is the number of populations of our interest; $k \geq 2$.

What is the Chi-Square Test?

The chi-square test can also be used to test whether two factors are independent of each other. In other words, it can be used to test whether there is any statistically significant relationship between two discrete factors.

- Null Hypothesis (H_0): Factor 1 is independent of factor 2.
- Alternative Hypothesis (H_a): Factor 1 is not independent of factor 2.

Chi-Square Test Assumptions

- The sample data drawn from the populations of interest are unbiased and representative.
- There are only two possible outcomes in each trial for an individual population: success/failure, yes/no, and defective/non-defective etc.
- The underlying distribution of each population is binomial distribution.
- When $np \geq 5$ and $np(1 - p) \geq 5$, the binomial distribution can be approximated by the normal distribution.

How Chi-Square Test Works

Test Statistic

$$\chi^2_{calc} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where:

- O_i is an observed frequency
- E_i is an expected frequency
- N is the number of cells in the contingency table.

If χ^2_{calc} (calculated chi-square statistic) is smaller than χ^2_{crit} (critical value), we fail to reject the null hypothesis. The test statistic is calculated with the observed and expected frequency.

Use JMP to Run a Chi-Square Test

Case study 1: We are interested in comparing the product quality exam pass rates of three suppliers A, B, and C using a nonparametric (i.e., distribution-free) hypothesis test: chi-square test.

Data File: "ChiSquareTest1.jmp"

- Null Hypothesis (H_0): $p_A = p_B = p_C$
- Alternative Hypothesis (H_a): At least one of the suppliers has different pass rates from the others.

Steps to run a chi-square test in JMP:

1. Click Analyze -> Fit Y by X
2. Select "Results" as "Y, Columns"
3. Select "Supplier" as "X, Factor"
4. Select "Count" as "Freq"

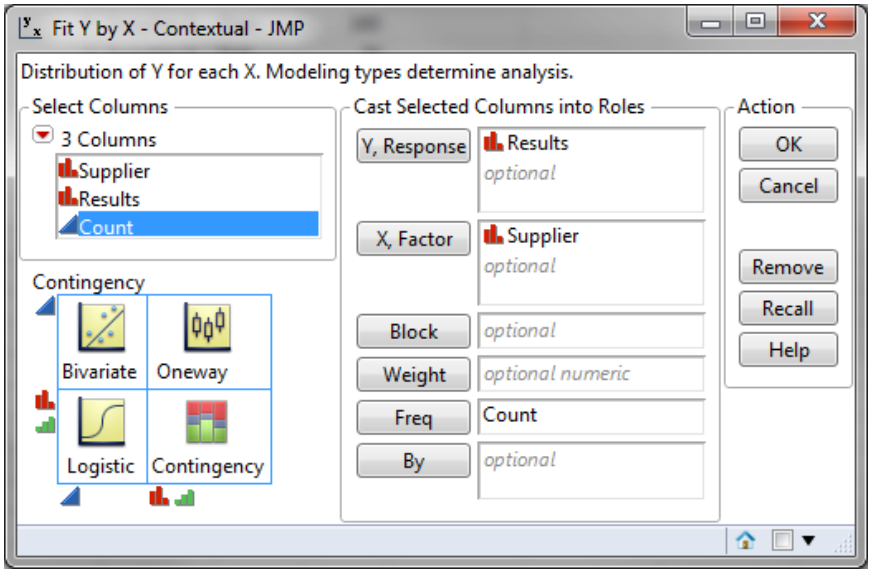


Fig. 3.80 Distribution for Y and X

5. Click "OK"

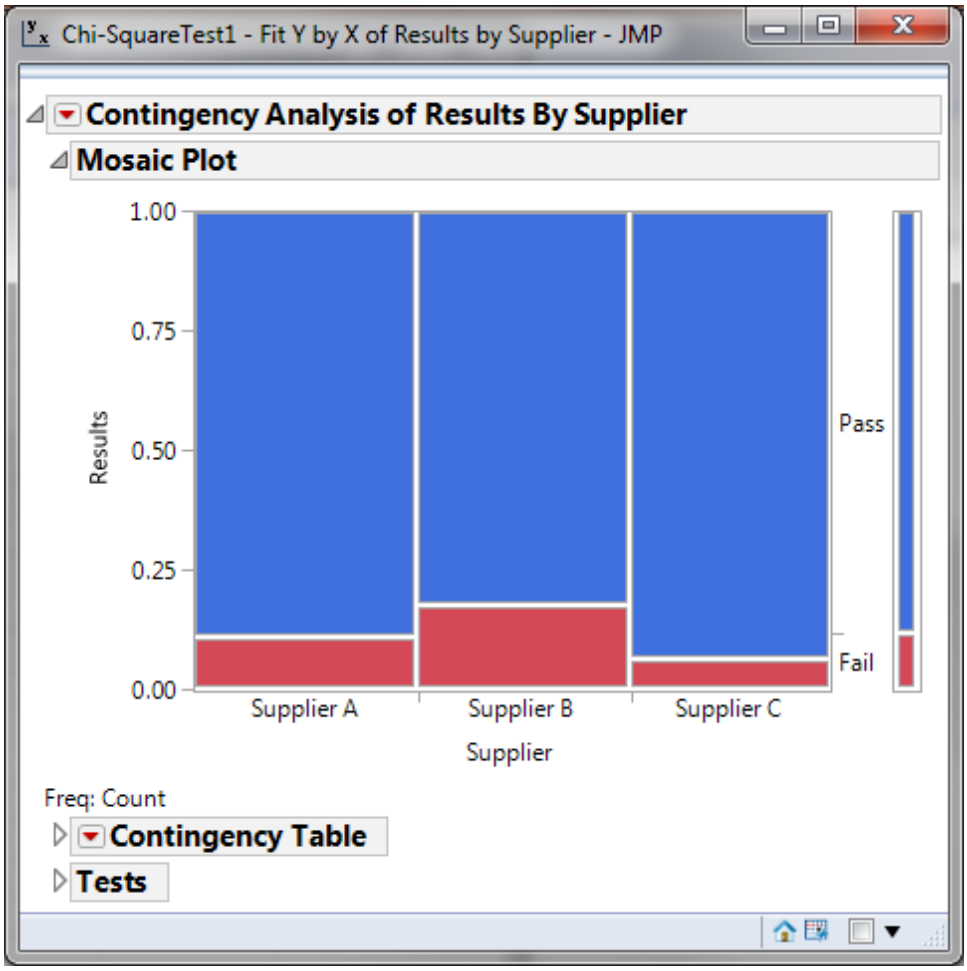


Fig. 3.81 JMP Mosaic Plot results

Mosaic plot is a graphical tool to divide the frequency data into smaller segments each of which is represented by a rectangle with the area proportional to the frequency of a specific outcome.

- The Y axis displays the classification of response. In our example, Y has two possible values: pass or fail.
- The X axis displays the different levels of a factor “Supplier”.

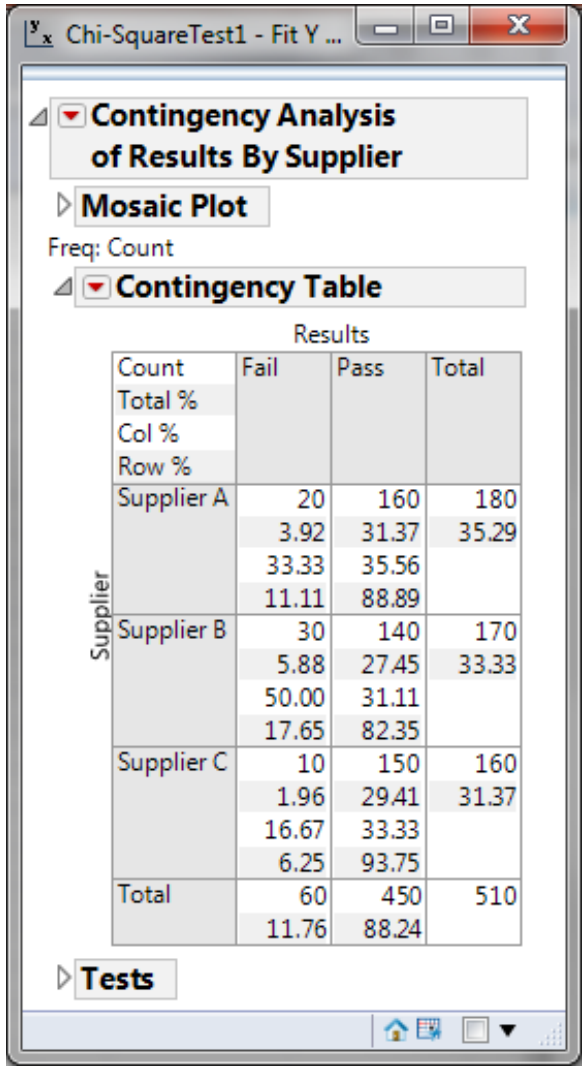


Fig. 3.82 Contingency Table result per supplier

Optional Step: To see more statistics in the contingency table, click on the red triangle button next to "Contingency Table" and select the statistics of interest.

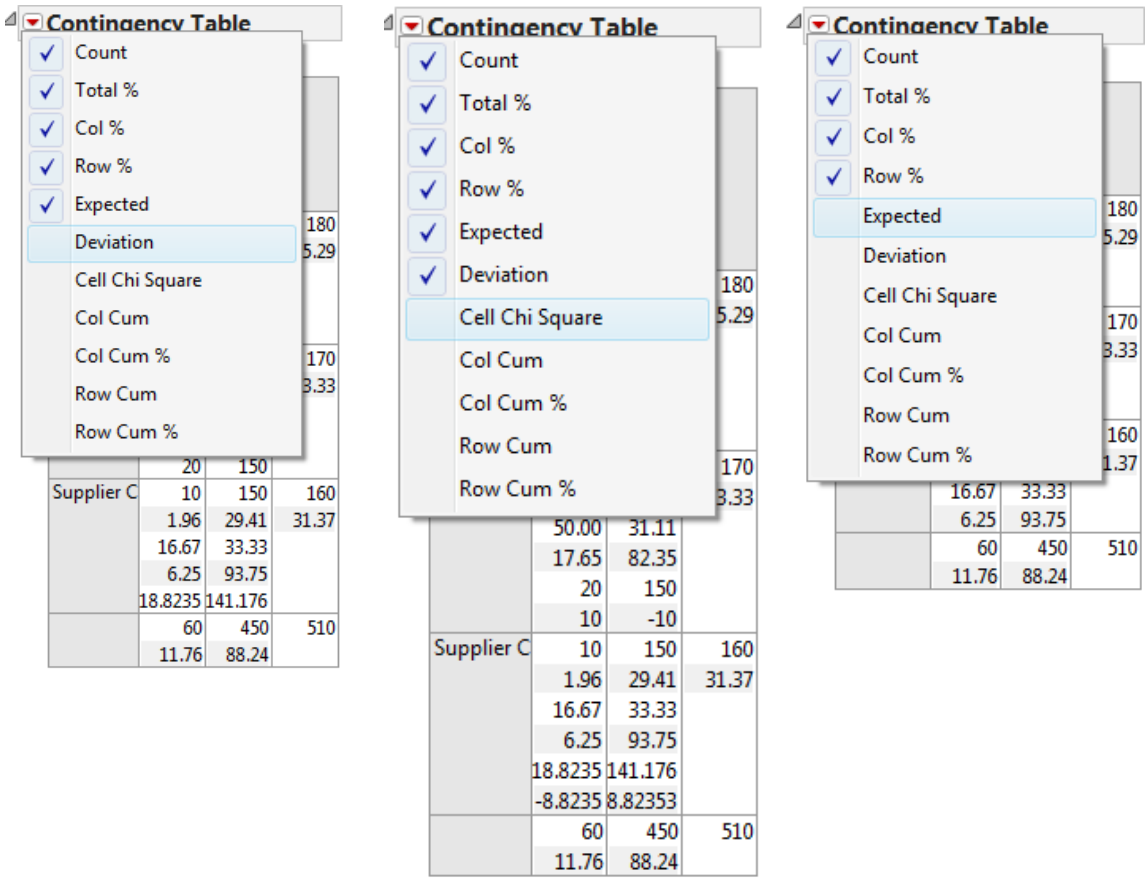


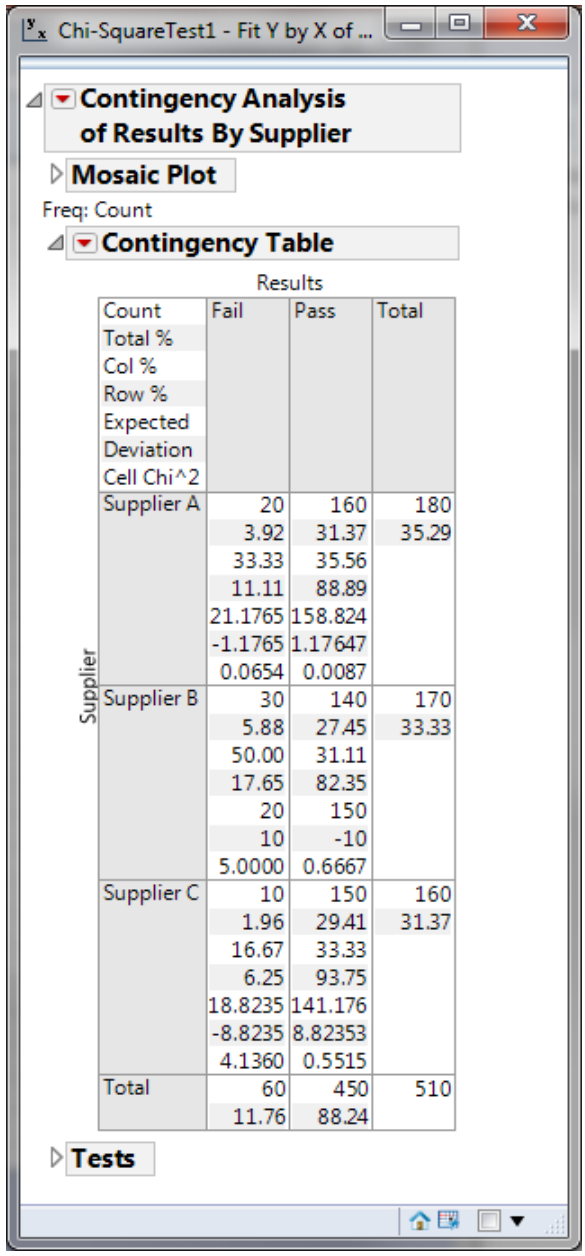
Fig. 3.83, 3.84, 3.85 Additional Statistics in Contingency Tables

After performing the optional step, three more measurements appear in the cells of contingency table:

Expected (E): expected cell frequency under the assumption that the null hypothesis is true. It is the product of the corresponding row and column total, divided by the grand total.

Deviation: the difference between the observed cell frequency(O) and the expected cell frequency (E).

Cell Chi Square: chi-square value for one cell. It is computed as $(O-E)^2/E$.



Supplier	Count / Total % / Col % / Row % / Expected / Deviation / Cell Chi^2	Fail	Pass	Total
Supplier A		20	160	180
		3.92	31.37	35.29
		33.33	35.56	
		11.11	88.89	
		21.1765	158.824	
		-1.1765	1.17647	
		0.0654	0.0087	
Supplier B		30	140	170
		5.88	27.45	33.33
		50.00	31.11	
		17.65	82.35	
		20	150	
		10	-10	
		5.0000	0.6667	
Supplier C		10	150	160
		1.96	29.41	31.37
		16.67	33.33	
		6.25	93.75	
		18.8235	141.176	
		-8.8235	8.82353	
		4.1360	0.5515	
Total		60	450	510
		11.76	88.24	

Fig. 3.86 Contingency Table data results

There are multiple tests available to test the significance of the difference among the proportions displayed in both the Mosaic plot and contingency table. For example, Pearson's chi-square test, Fisher's exact test, Barnard's test etc. We recommend using the p-value of the Likelihood Ratio when the sample size is small. Since both p-values are smaller than alpha level (0.05), we reject the null and claim that at least one supplier has different pass rate from others.

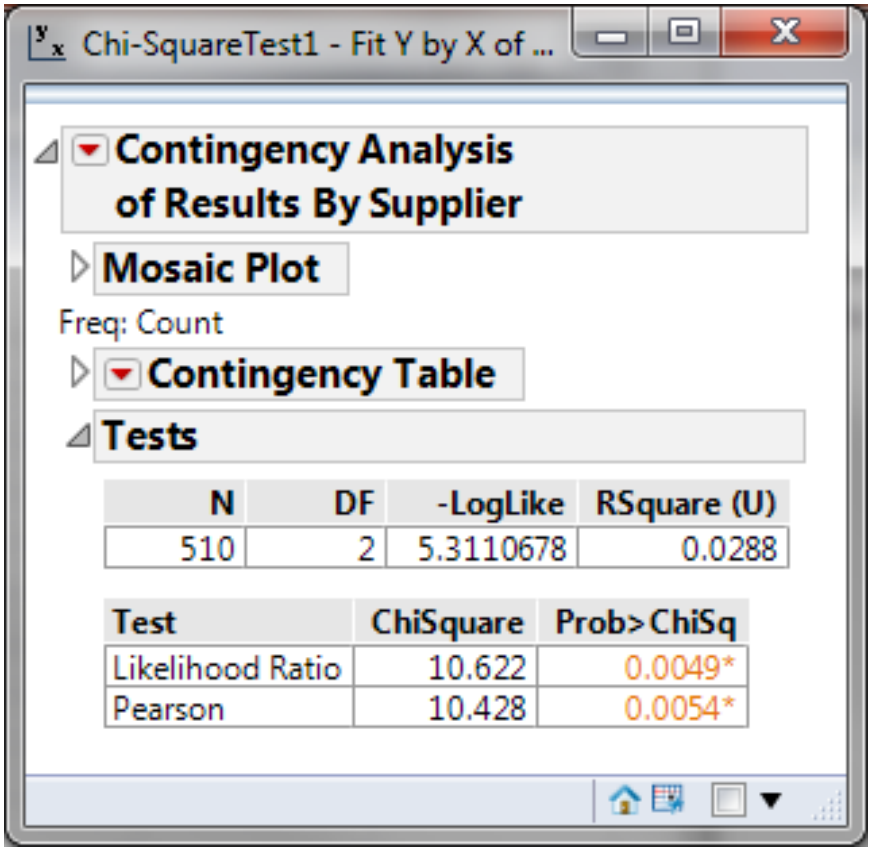


Fig. 3.87 Chi-Square Test Output

Case study 2: We are trying to check whether there is a relationship between the suppliers and the results of the product quality exam using nonparametric (i.e., distribution-free) hypothesis test: chi-square test.

Data File: "ChiSquareTest2.jmp"

- Null Hypothesis (H_0): Product quality exam results are independent of the suppliers.
- Alternative Hypothesis (H_a): Product quality exam results depend on the suppliers.

Steps to run a chi-square test in JMP:

1. Click Analyze -> Fit Y by X
2. Select "Results" as "Y, Columns"
3. Select "Supplier" as "X, Factor"
4. Select "Count" as "Freq"

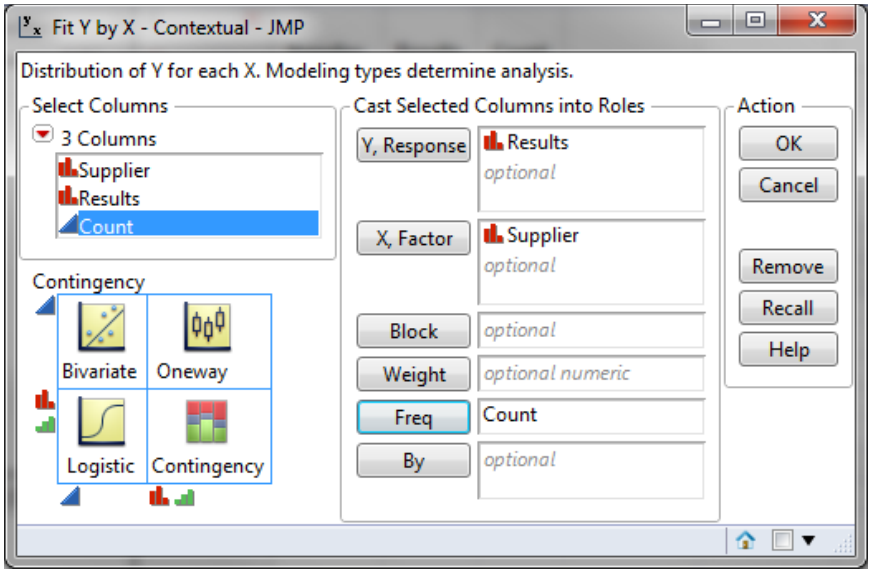


Fig. 3.88 Distribution for Y and X

5. Click "OK"

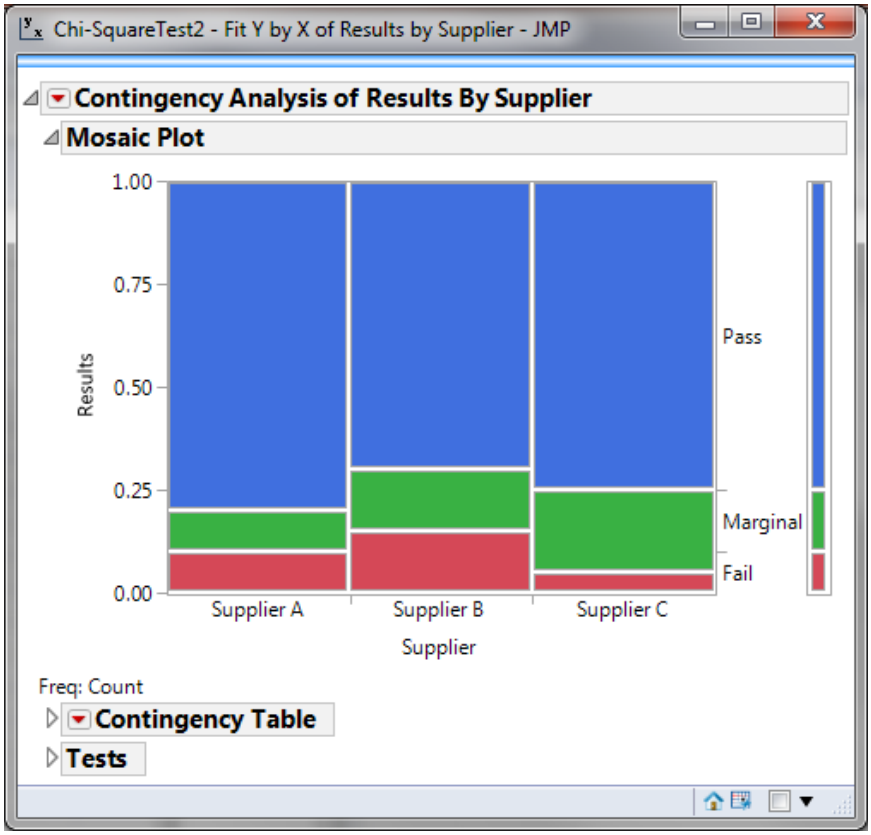


Fig. 3.89 Mosaic Plot output

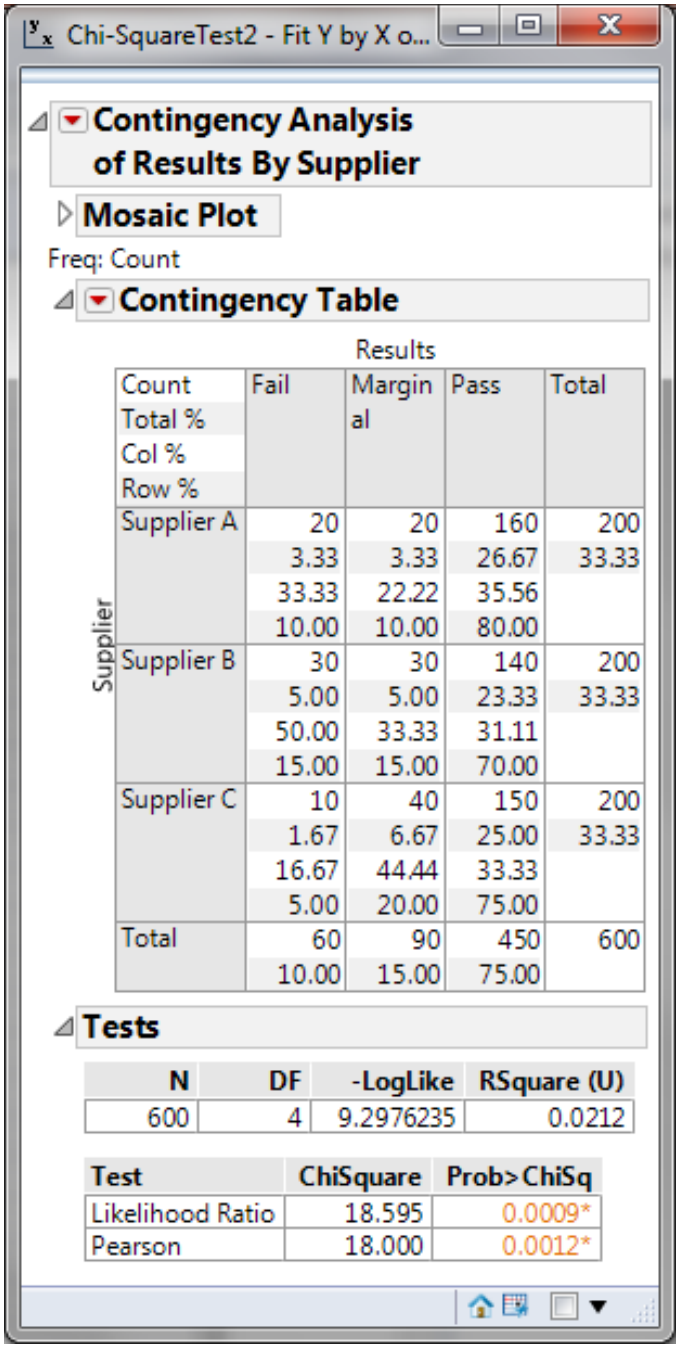


Fig. 3.90 Contingency Table Output

The p-value is smaller than the alpha level (0.05) and we reject the null hypothesis. The product quality exam results are not independent of the suppliers. These results indicate the danger that we can get into when using discrete data. Not everything is as simple as yes/no or pass/fail. Even though supplier C has a lower fail rate of 10, you can see that the number of marginal results is higher. However, the p-value tells us that we must reject the null hypothesis and claim that the quality exam results *are* dependent on the suppliers.

3.5.9 Tests of Equal Variance

What are Tests of Equal Variance?

Tests of equal variance are a family of hypothesis tests used to check whether there is a statistically significant difference between the variances of two or more populations.

- Null Hypothesis (H_0): $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma k^2$
- Alternative Hypothesis (H_a): At least the variance of one population is different from others.

k is the number of populations of interest; $k \geq 2$.

A test of equal variance can be used alone but most of the time it is used with other statistical methods to verify or support the assumption about the variance equality.

F-Test

The *F-test* is used to compare the variances between two normally distributed populations. It is extremely sensitive to non-normality and serves as a preliminary step for two sample t-test. Therefore, it is crucial to validate that the data is normal before using this test.

Test Statistic

$$F_{calc} = \frac{s_1^2}{s_2^2}$$

Where:

- s_1 and s_2 are the sample standard deviations
- The sampling distribution of the test statistic follows F distribution when the null is true.

Bartlett's Test

Bartlett's test is used to compare the variances among two or more normally distributed populations. It is sensitive to non-normality and it serves as a preliminary step for ANOVA.

Test Statistic

$$\chi^2 = \frac{(N-k)\ln(S_p^2) - \sum_{i=1}^{k}(n_i - 1)\ln(S_i^2)}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\left(\frac{1}{n_i - 1}\right) - \frac{1}{N-k}\right)}$$

Where:

$$N = \sum_{i=1}^{k} n_i \text{ and } S_p^2 = \frac{1}{N-k}\sum_i (n_i - 1)S_i^2$$

The sampling distribution of test statistic follows chi square distribution when the null is true.

Brown–Forsythe Test

The *Brown–Forsythe test* is used to compare the variances between two or more populations with any distributions. It is not as sensitive to non-normality as Bartlett's test. The test statistic is the model F statistic from the ANOVA on the transformed response:

$$z_{ij} = \left| y_{ij} - \tilde{y}_i \right|$$

Where: $\tilde{y}_i$ is the median response at ith level.

Levene's Test

Levene's test is used to compare the variances between two or more populations with any distributions. It is not as sensitive to non-normality as Bartlett's test. The test statistic is the model F statistic from the ANOVA on the transformed response:

$$z_{ij} = \left| y_{ij} - \bar{y}_i \right|$$

Where: $\bar{y}_i$ is the mean response at ith level.

Brown–Forsythe Test vs. Levene's Test

$$F = \frac{(N-k)}{(k-1)} \frac{\sum_{i=1}^{k} N_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^{k}\sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}$$

Where:

- N is the total number of observations.
- k is the number of groups.
- N_i is the number of observations in the ith group
- $Z_{i.}$ is the group mean of the ith group
- $Z_{..}$ is the grand mean of all the observations.

The F statistic for both the Brown–Forsythe and Levene's test are calculated the same way; however, you can see at the bottom of the slide how Z represents a different value for each.

In Brown–Forsythe test, $Z_{ij} = \left| Y_{ij} - \widetilde{Y}_{ij} \right|$, where $\widetilde{Y}_{ij}$ is the group median of the i[th] group. In Levene's test, $Z_{ij} = \left| Y_{ij} - \overline{Y}_{ij} \right|$, where $\overline{Y}_{ij}$ is the group mean of the i[th] group.

Use JMP to Run Tests of Equal Variance

Case study: We are interested in comparing the variances of the retail price of a product in state A and state B.

Data File: "TwoSampleT-Test.jmp"

- Null Hypothesis (H_0): $\sigma_A^2 = \sigma_B^2$
- Alternative Hypothesis (H_a): $\sigma_A^2 = \sigma_B^2$

Steps to run tests of equal variance in JMP:

1. Click Analyze -> Fit Y by X
2. Select "Retail Price" as "Y, Response"
3. Select "State" as "X, Factor"

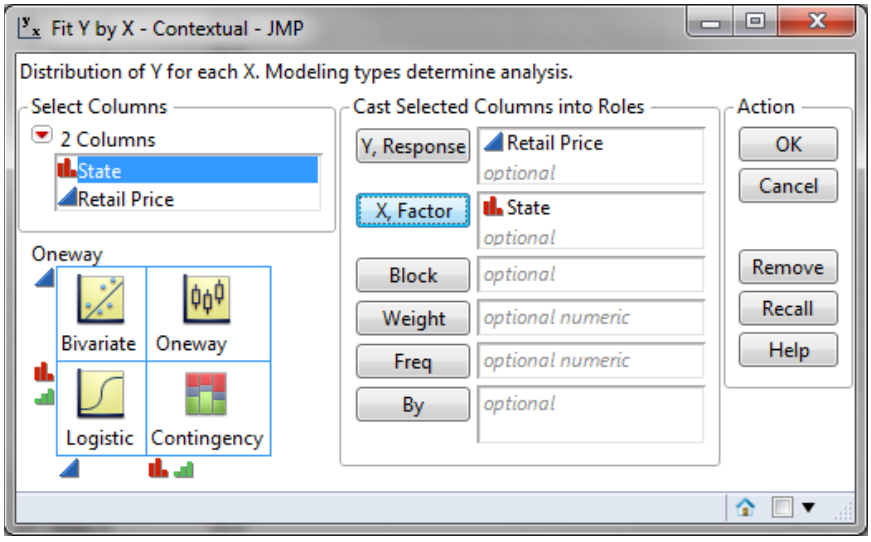
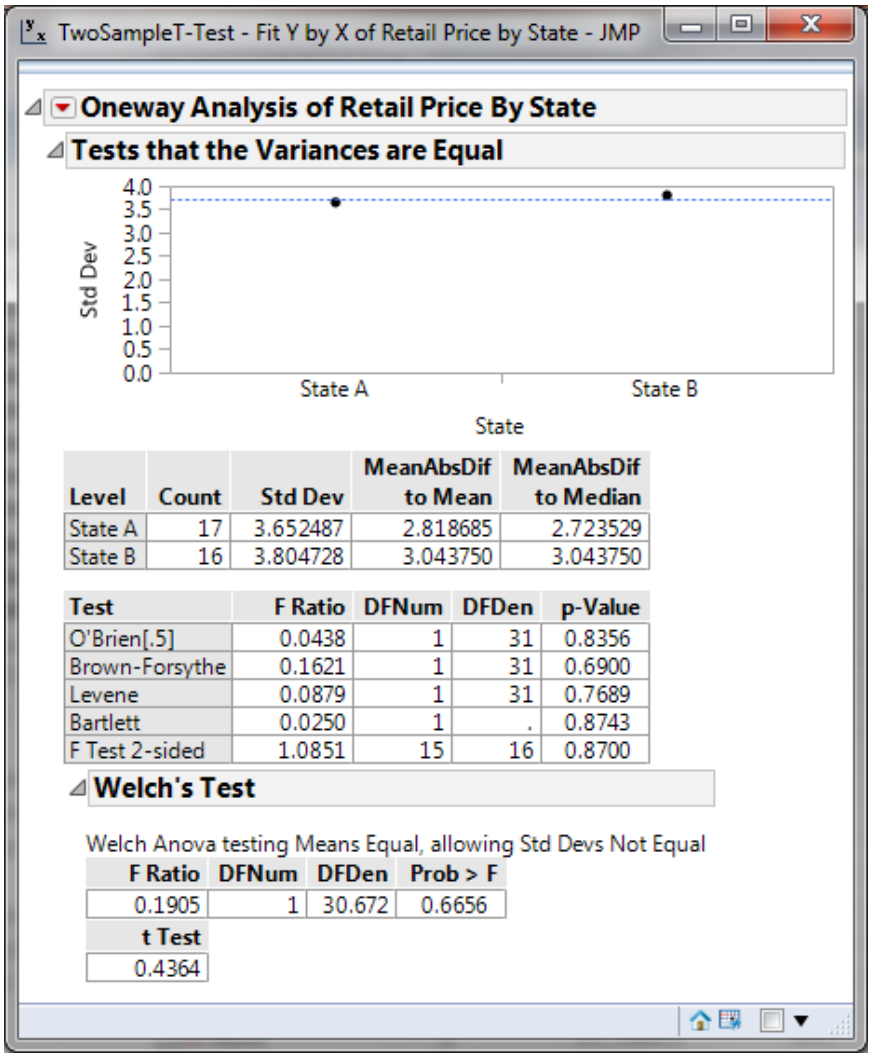


Fig. 3.91 Distribution for Y and X

4. Click "OK"
5. Click on the red triangle button next to "One-Way Analysis of Retail Price By State"
6. Select "Unequal Variances"



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
State A	17	3.652487	2.818685	2.723529
State B	16	3.804728	3.043750	3.043750

Test	F Ratio	DFNum	DFDen	p-Value
O'Brien[.5]	0.0438	1	31	0.8356
Brown-Forsythe	0.1621	1	31	0.6900
Levene	0.0879	1	31	0.7689
Bartlett	0.0250	1	.	0.8743
F Test 2-sided	1.0851	15	16	0.8700

F Ratio	DFNum	DFDen	Prob > F
0.1905	1	30.672	0.6656

t Test
0.4364

Fig. 3.92 JMP Results of the 5 Tests for Equal Variance

Both retail price data of state A and B are normally distributed since the p-values are both greater than alpha level (0.05). When the p-value is greater than the alpha value, we fail to reject the null hypothesis and claim that the distributions are normal. Both samples from state A and state B are normally distributed.

If all the groups of data are normally distributed, use the F-test or Bartlett's test in JMP to test the equality of the variances. If at least one of the groups is not normally distributed, use Levene's test in JMP to test the equality of the variances. If the p-value of the variances equality test is greater than the alpha level (0.05), we fail to reject the null hypothesis and conclude that the variances of different groups are identical.

Hypothesis Testing Roadmap: Putting it all together

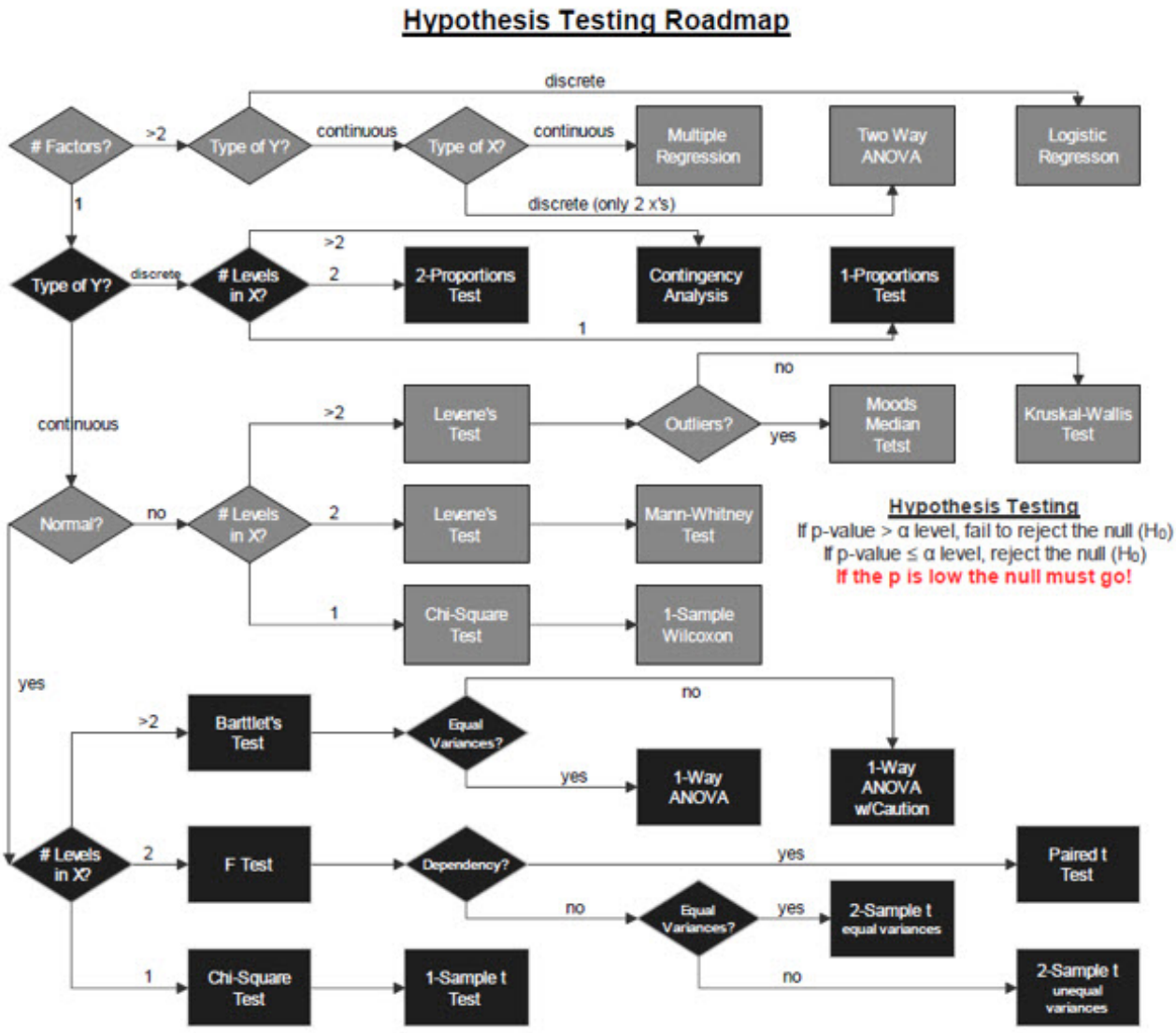


Fig. 3.93 Hypothesis Testing Roadmap

The roadmap is a great way to quickly decide what type of hypothesis test fits a given situation by evaluating the type of data (discrete vs. continuous), the normality of the data, the number of levels, and other factors.

4.0 IMPROVE PHASE

4.1 SIMPLE LINEAR REGRESSION

After identifying and validating the critical Xs during the Analyze phase, we will use some statistical tools during the Improve phase to define the settings needed to achieve the desired output. One such tool is simple linear regression.

4.1.1 CORRELATION

We will start with correlation before regression, because it is a tool that helps us understand whether a relationship exists between two variables without having to know causality (which variable causes a response in the other). There are many similarities between correlation and simple linear regression.

What is Correlation?

Correlation is a statistical technique that describes whether and how strongly two or more variables are related.

Correlation analysis helps to understand the direction and degree of association between variables, and it suggests whether one variable can be used to predict another. Of the different metrics to measure correlation, *Pearson's correlation coefficient* is the most popular. It measures the linear relationship between two variables.

Pearson's Correlation Coefficient

Pearson's correlation coefficient is also called Pearson's r or coefficient of correlation and Pearson's product moment correlation coefficient (r), where r is a statistic measuring the linear relationship between two variables.

Correlation coefficients range from −1 to 1.

- If r = 0, there is no linear relationship between the variables.

The *sign* of r indicates the *direction* of the relationship:

- If r < 0, there is a negative linear correlation.
- If r > 0, there is a positive linear correlation.

The *absolute value* of r describes the *strength* of the relationship:

- If |r| ≤ 0.5, there is a weak linear correlation.
- If |r| > 0.5, there is a strong linear correlation.
- If |r| = 1, there is a perfect linear correlation.

When the correlation is *strong*, the data points on a scatter plot will be close together (tight). The closer r is to −1 or 1, the stronger the relationship.

- −1 Strong inverse relationship
- +1 Strong direct relationship

When the correlation is *weak*, the data points are spread apart more (loose). The closer the correlation is to 0, the weaker the relationship.

Fig. 4.1 Examples of Types of Correlation

This Figure demonstrates the relationships between variables as the Pearson r value ranges from 1 to 0 and to −1. Notice that at −1 and 1 the points form a perfectly straight line.

- At 0 the data points are completely random.
- At 0.8 and −0.8, notice how you can see a directional relationship, but there is some noise around where a line would be.
- At 0.4 and −0.4, it looks like the scattering of data points is leaning to one direction or the other, but it is more difficult to see a relationship because of all the noise.

Pearson's correlation coefficient is only sensitive to the *linear* dependence between two variables. It is possible that two variables have a perfect non-linear relationship when the correlation coefficient is low. Notice the scatter plots below with correlation equal to 0. There are clearly *relationships* but they are not linear and therefore cannot be determined with Pearson's correlation coefficient.

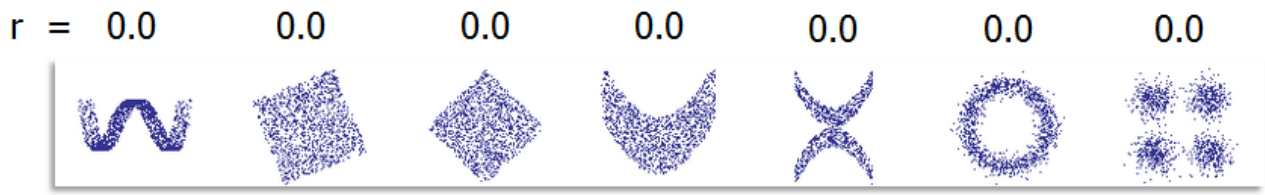


Fig. 4.2 Examples of Types of Relationships

Correlation and Causation

Correlation *does not* imply causation.

If variable A is highly correlated with variable B, it does not necessarily mean A causes B or vice versa. It is possible that an unknown third variable C is causing both A and B to change. For example, if ice cream sales at the beach are highly correlated with the number of shark

attacks, it does not imply that increased ice cream sales causes increased shark attacks. They are triggered by a third factor: summer.

This example demonstrates a common mistake that people make: assuming causation when they see correlation. In this example, it is hot weather that is a common factor. As the weather is hotter, more people consume ice cream *and* more people swim in the ocean, making them susceptible to shark attacks.

Correlation and Dependence

If two variables are independent, the correlation coefficient is zero.

WARNING! If the correlation coefficient of two variables is zero, it does not imply they are independent. The correlation coefficient only indicates the linear dependence between two variables. When variables are non-linearly related, they are not independent of each other but their correlation coefficient could be zero.

Correlation Coefficient and X-Y Diagram

The correlation coefficient indicates the direction and strength of the linear dependence between two variables but it does not cover all the existing relationship patterns. With the same correlation coefficient, two variables might have completely different dependence patterns. A scatter plot or X-Y diagram can help to discover and understand additional characteristics of the relationship between variables. The correlation coefficient is not a replacement for examining the scatter plot to study the variables' relationship.

The correlation coefficient by itself does not tell us everything about the relationship between two variables. Two relationships could have the same correlation coefficient, but completely different patterns.

Statistical Significance of the Correlation Coefficient

The correlation coefficient could be high or low by chance (randomness). It may have been calculated based on two small samples that do not provide good inference on the correlation between two populations.

To test whether there is a statistically significant relationship between two variables, we need to run a hypothesis test to determine whether the correlation coefficient is statistically different from zero.

Hypothesis Test Statements

- H_0: $r = 0$: Null Hypothesis: There is *no* correlation.
- H_1: $r \neq 0$: Alternate Hypothesis: There is a correlation.

Hypothesis tests will produce p-values as a result of the statistical significance test on r. When the p-value for a test is low (less than 0.05), we can reject the null hypothesis and conclude that r is significant; there is a correlation. When the p-value for a test is > 0.05, then we fail to reject the null hypothesis; there is no correlation.

We can also use the t statistic to draw the same conclusions regarding our test for significance of the correlation coefficient. To use the t-test to determine the statistical significance of the Pearson correlation, calculate the t statistic using the Pearson r value and the sample size, n.

Test Statistic

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

Critical Statistic

Is the t-value in t-table with (n – 2) degrees of freedom.

If the absolute value of the calculated t value is less than or equal to the critical t value, then we fail to reject the null and claim no statistically significant linear relationship between X and Y.

- If $|t| \leq t_{critical}$, we fail to reject the null. There is no statistically significant linear relationship between X and Y.
- If $|t| > t_{critical}$, we reject the null. There is a statistically significant linear relationship between X and Y.

Using Software to Calculate the Correlation Coefficient

We are interested in understanding whether there is linear dependence between a car's MPG and its weight and if so, how they are related. The MPG and weight data are stored in the "Correlation Coefficient" tab in "Sample Data.xlsx." We will discuss three ways to get the results.

Use Excel to Calculate the Correlation Coefficient

The formula CORREL in Excel calculates the sample correlation coefficient of two data series. The correlation coefficient between the two data series is −0.83, which indicates a strong negative linear relationship between MPG and weight. In other words, as weight gets larger, gas mileage gets smaller.

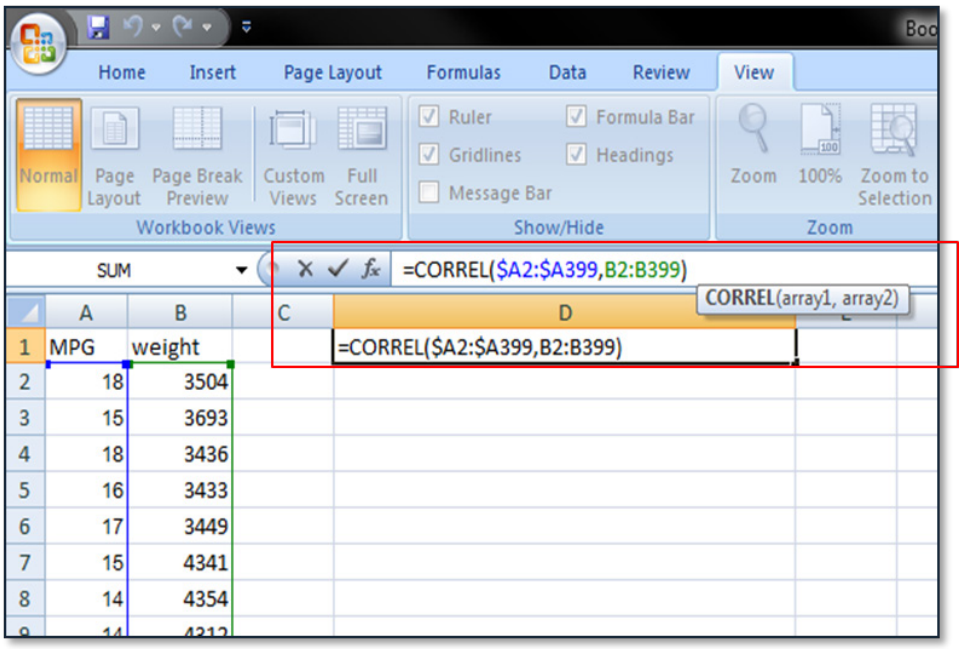


Fig. 4.3 Correlation coefficient in Excel

Use JMP to Calculate the Correlation Coefficient

Steps to run tests of equal variance in JMP:

1. Analyze → Multivariate Methods → Multivariate.
2. Select the two variables of interest and click "OK."

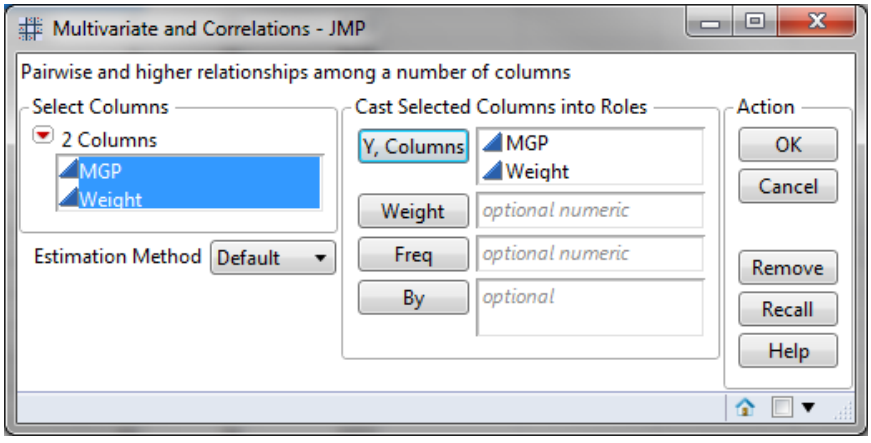


Fig. 4.4 Variables selection

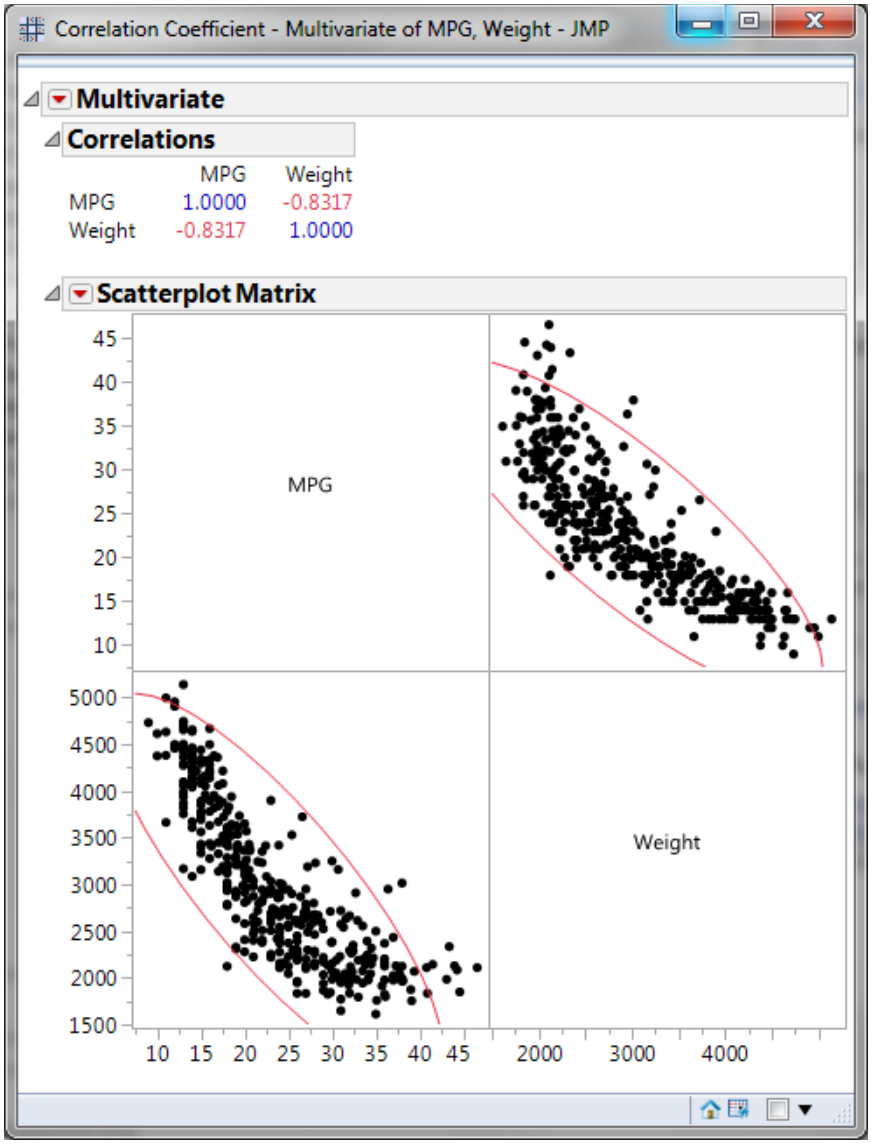


Fig. 4.5 Correlation Analysis

When the correlation analysis is done through JMP, the program presents the results in two ways.

- It presents the coefficient in the correlation matrix just above the scatter plots.
- It also plots a scatter plot so we can have a visual of the relationship between the two variables

The correlation coefficient result (−0.832) appears in the session window. The p-value (0.000) is lower than the alpha level (0.05), indicating the linear correlation is statistically significant. We can claim that there is a linear relationship between mileage and weight.

Interpreting Results

How do we interpret results and make decisions based Pearson's correlation coefficient (r) and p-values?

Let us look at a few examples:

- r = −0.832, p = 0.000 (previous example). The two variables are inversely related and the linear relationship is strong. Also, this conclusion is significant as supported by p-value of 0.00.
- r = −0.832, p = 0.71. Based on r, you should conclude the linear relationship between the two variables is strong and inversely related. However, with a p-value of 0.71, you should then conclude that r is not significant and that your sample size may be too small to accurately characterize the relationship.
- r = 0.5, p = 0.00. Moderately positive linear relationship, r is statistically significant.
- r = 0.92, p = 0.61. Strong positive linear relationship but r is not statistically significant. Get more data.
- r = 1.0, p = 0.00. The two variables have a perfect linear relationship and r is significant.

Correlation Coefficient Calculation

Population Correlation Coefficient (ρ)

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Sample Correlation Coefficient (r)

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

It is only defined when the standard deviations of both X and Y are non-zero and finite. When covariance of X and Y is zero, the correlation coefficient is zero.

4.1.2 X-Y DIAGRAM

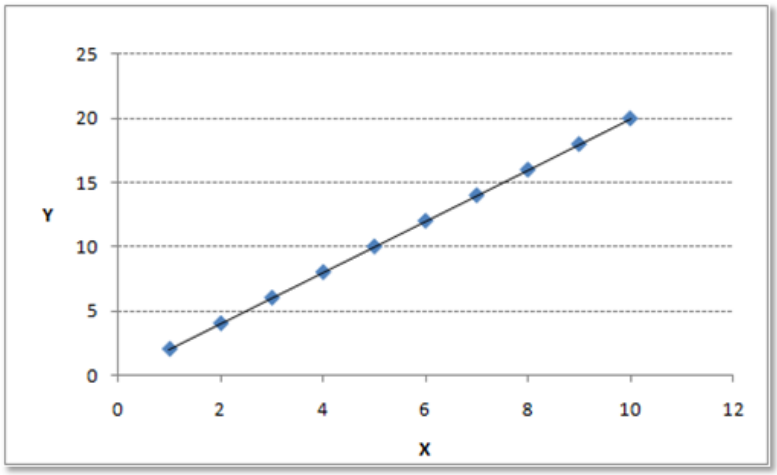
What is an X-Y Diagram?

An *X-Y diagram* is a scatter plot depicting the relationship between two variables (i.e., X and Y). Each point on the X-Y diagram represents a pair of X and Y values, with X plotted on the horizontal axis and Y plotted on the vertical axis.

With an X-Y diagram, you can qualitatively assess both the strength and direction of the relationship between X and Y. To quantitatively measure the relationship between X and Y, you may need to calculate the correlation coefficient.

Example 1: Perfect Linear Correlation
In the chart below, there are 10 data points depicted (10 pairs of X and Y values), and they were created using the equation Y = 2X. As a result, all the data points fall onto the straight line of Y = 2X.

The chart demonstrates a perfect positive linear correlation between X and Y since the relationship between X and Y can be perfectly described by a linear equation in a format of Y = a × X + b where a ≠ 0. If they were negatively correlated, the line would slope down to the right, meaning as X increases, Y decreases.



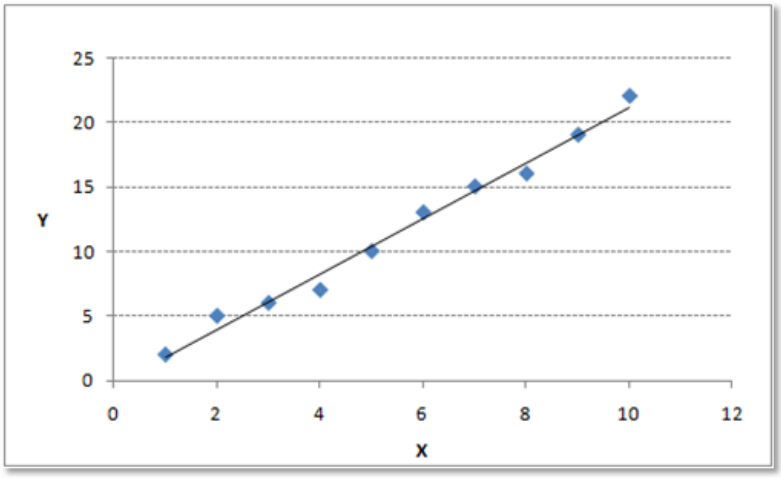
X	Y
1	2
2	4
3	6
4	8
5	10
6	12
7	14
8	16
9	18
10	20

Fig. 4.6 Perfect Positive Linear Correlation

Example 2: Strong Linear Correlation

In this chart, the data points scatter closely around a straight line. When X increases, Y increases accordingly.

This chart demonstrates a strong positive linear correlation between X and Y. The straight line is the best fitting trend that shows approximately how Y changes with X.



X	Y
1	2
2	5
3	6
4	7
5	10
6	13
7	15
8	16
9	19
10	22

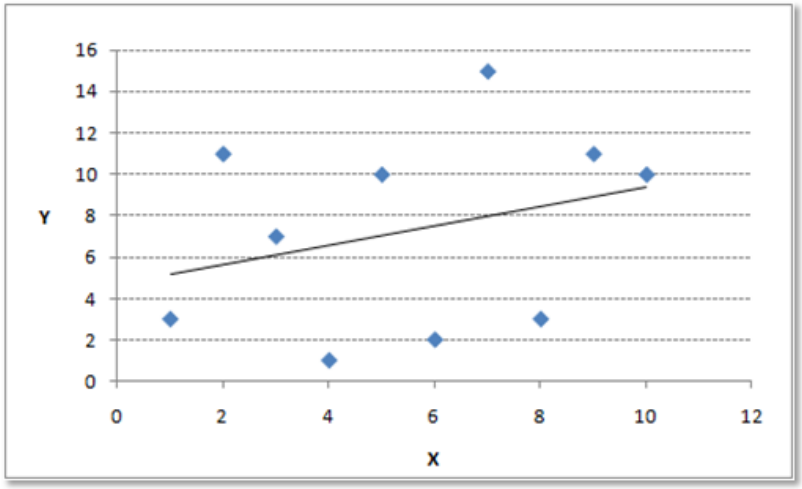
Fig. 4.7 Strong Positive Linear Correlation

In this example, there is a strong correlation, but not perfect. Notice how the points do not all line up perfectly.

Example 3: Weak Linear Correlation

In this chart, the data points scatter remotely around a straight line. When X increases, Y increases accordingly.

This chart demonstrates a weak positive linear correlation between X and Y since the distance between the data points and the trend line is relatively far on average.

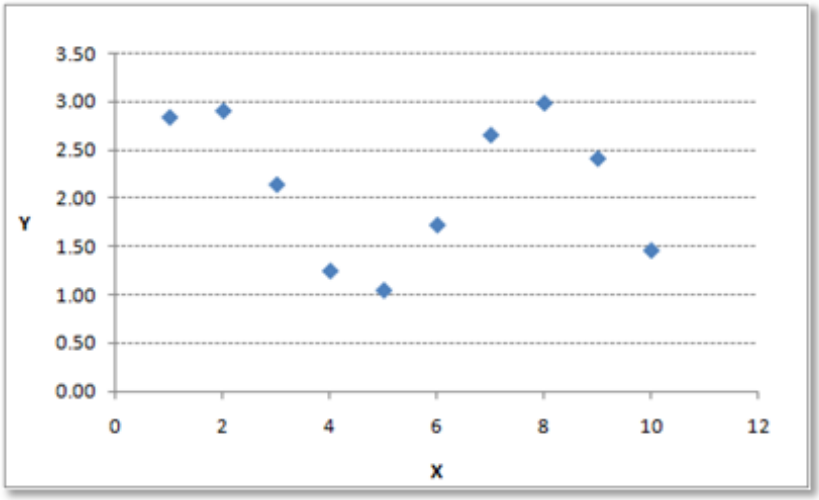


X	Y
1	3
2	11
3	7
4	1
5	10
6	2
7	15
8	3
9	11
10	10

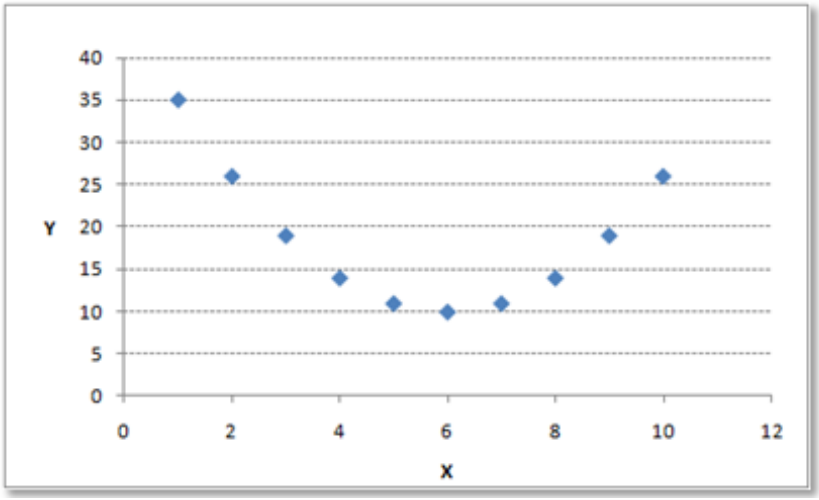
Fig. 4.8 Weak Positive Linear Correlation

The distance between the data points and the best fitting line is much higher than on the previous chart, which is means that the line does not describe the relationship between the X and Y as tightly.

Example 4: Non-Linear Correlation
The X-Y diagram also helps to identify any nonlinear relationship between X and Y.



X	Y
1	2.84
2	2.91
3	2.14
4	1.24
5	1.04
6	1.72
7	2.66
8	2.99
9	2.41
10	1.46



X	Y
1	35
2	26
3	19
4	14
5	11
6	10
7	11
8	14
9	19
10	26

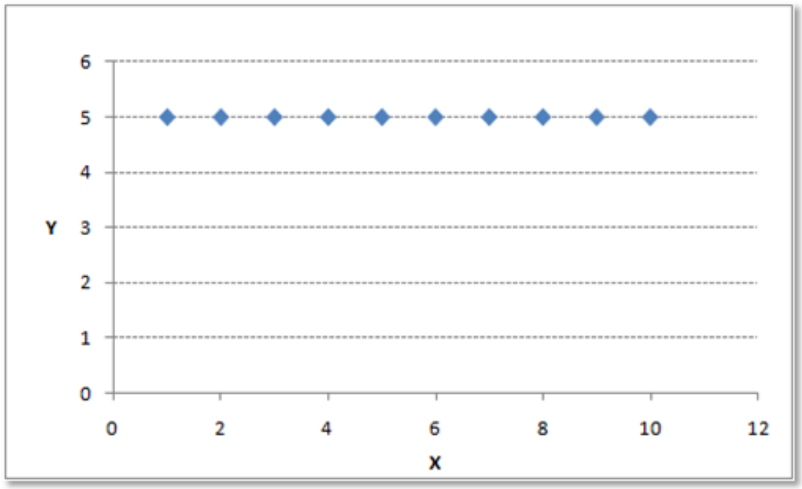
Fig. 4.9 Non-Linear Correlation

The examples show some non-linear correlations. Notice there is a pattern for each relationship, but it is not a line that describes it.

Example 5: Uncorrelated

In this chart, the Y value of each data point is a constant regardless of what the X value is.

Changes in X do not show any relative impact on Y. As a result, there is no correlation between X and Y.



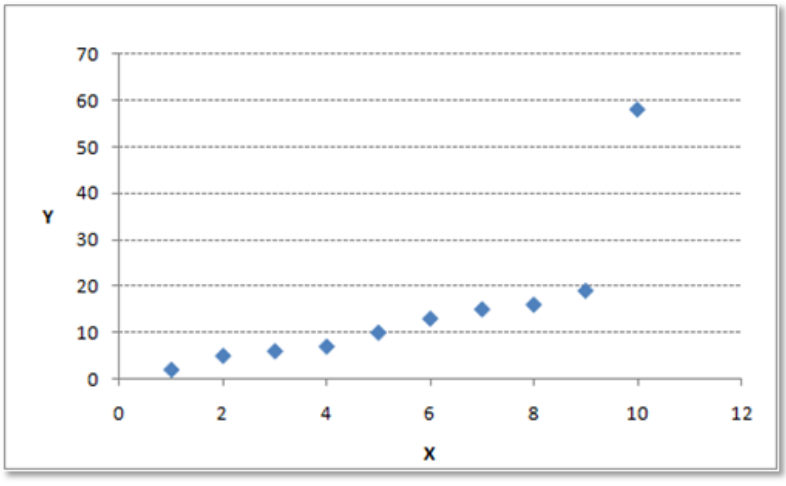
X	Y
1	5
2	5
3	5
4	5
5	5
6	5
7	5
8	5
9	5
10	5

Fig. 4.10 Uncorrelated Correlation

Example 6: Outlier Identification

Another feature of the X-Y diagram is that it allows the user to identify an outlier (or a data point that does not follow the same trend as the rest of the other data points).

In this chart, the last data point does not seem to follow the trend of other data points. When an outlier is spotted, it is good practice to investigate and understand what is causing it.



X	Y
1	2
2	5
3	6
4	7
5	10
6	13
7	15
8	16
9	19
10	58

Fig. 4.11 Outlier Identification

Benefits to Using an X-Y Diagram

- An X-Y diagram graphically demonstrates the relationship between two variables (no relationship, linear relationship, nonlinear relationship).
- It suggests whether two variables are associated and helps to identify the linear or nonlinear correlation between X and Y.
- It captures the strength and direction of the relationship between X and Y.
- It helps identify any outliers in the data.

Limitations of the X-Y Diagram

Although the X-Y diagram helps to "spot" interesting features in the data, it does not provide any quantitative conclusions about the data and further statistical analysis is needed to:

- Assess whether the association between variables is statistically significant.
- Measure the strength of the relationship between variables.
- Determine whether outliers exist in the data.
- Quantitatively describe the pattern of the data.

4.1.3 Regression Equations

Correlation and Regression Analysis

The correlation coefficient answers the following questions:

1. Are two variables correlated?
2. How strong is the relationship between two variables?
3. When one variable increases, does the other variable increase or decrease?
4. The correlation coefficient *cannot* address the following questions:
5. How much does one variable changes when the other variable changes by one unit?
6. How can we set the value of one variable to obtain a targeted value of the other variable?
7. How can we use the relationship between two variables to make predictions?
8. The simple linear regression analysis helps to answer these questions.

What is Simple Linear Regression?

Simple linear regression is a statistical technique to fit a straight line through the data points. It models the quantitative relationship between two variables. It is simple because only one predictor variable is involved. It describes how one variable changes according to the change of another variable. Both variables need to be continuous; there are other types of regression to model discrete data.

Simple Linear Regression Equation

The simple linear regression analysis fits the data to a regression equation in the form

$$Y = \alpha \times X + \beta + e$$

Where:

- Y is the dependent variable (the response) and X is the single independent variable (the predictor)
- α is the slope describing the steepness of the fitting line. β is the intercept indicating the Y value when X is equal to 0

- *e* stands for error (residual). It is the difference between the actual *Y* and the fitted *Y* (i.e., the vertical difference between the data point and the fitting line).

Ordinary Least Squares

The *ordinary least squares* is a statistical method used in linear regression analysis to find the best fitting line for the data points. It estimates the unknown parameters of the regression equation by minimizing the sum of squared residuals (i.e., the vertical difference between the data point and the fitting line).

In mathematical language, we look for α and β that satisfy the following criteria:

$$\min_{\alpha,\beta} Q(\alpha,\beta) where\ Q(\alpha,\beta) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \alpha - \beta_{x_i}\right)^2$$

The actual value of the dependent variable:

$$Y_i = \alpha \times X_i + \beta + e_i$$

Where: *i = 1, 2 . . . n*.

The fitted value of the dependent variable:

$$\hat{y}_i = \alpha \times X_i + \beta$$

Where: *i = 1, 2 . . . n*.

By using calculus, it can be shown the sum of squared error is minimal when

$$\beta = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

and

$$\alpha = \bar{Y} - \beta\,\bar{X}$$

ANOVA in Simple Linear Regression

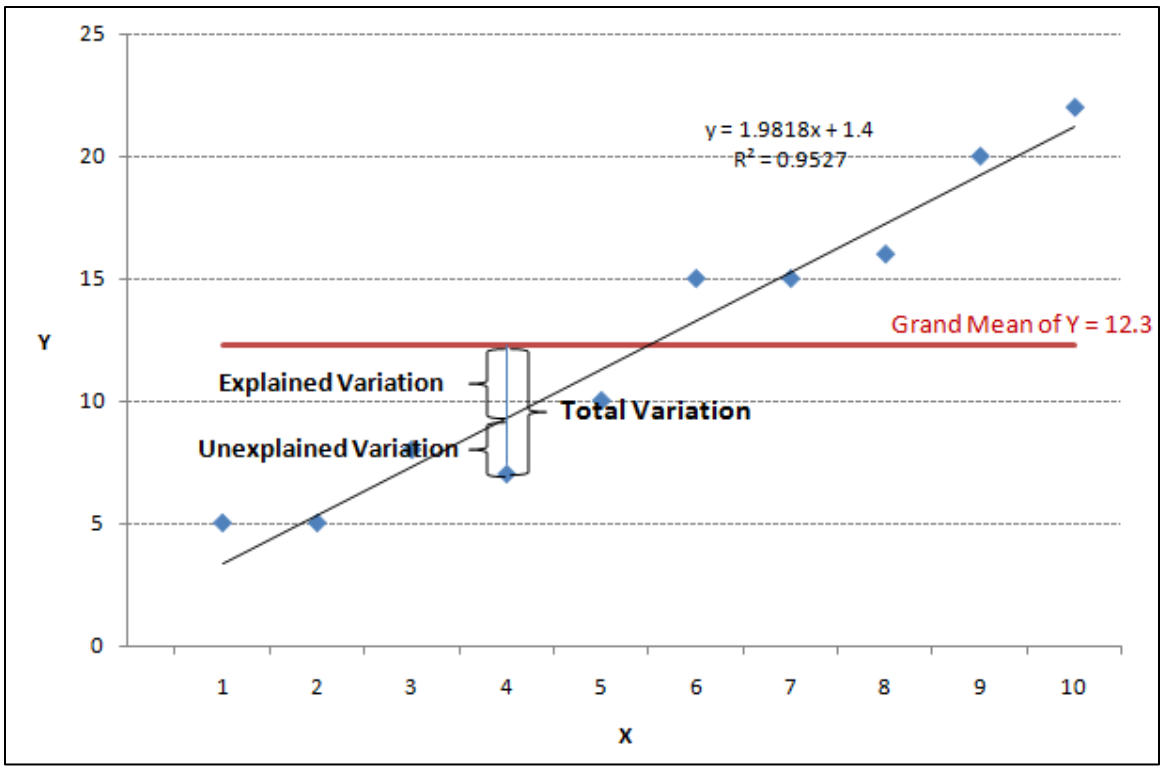


Fig. 4.12 ANOVA in Simple Linear Regression

X: the independent variable that we use to predict;

Y: the dependent variable that we want to predict.

The variance in simple linear regression can be expressed as a relationship between the actual value, the fitted value, and the grand mean—all in terms of Y.

Total Variation = Total Sums of Squares = $\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

Explained Variation = Regression Sums of Squares = $\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$

Unexplained Variation = Error Sums of Squares = $\sum_{i=1}^{n}(Y_i - \hat{Y})^2$

Regression follows the same methodology as ANOVA and the hypothesis tests behind it use the same assumptions.

Variation Components

$$Total\ Variation = Explained\ Variation + Unexplained\ Variation$$

i.e., Total Sums of Squares = Regression Sums of Squares + Error Sums of Squares

Degrees of Freedom Components

$$Total\ Degrees\ of\ Freedom = Regression\ Degrees\ of\ Freedom + Residual\ Degrees\ of\ Freedom$$

i.e., n – 1 = (k – 1) + (n – k), where n is the number of data points, k is the number of predictors

Whether the overall model is statistically significant can be tested by using F-test of ANOVA.

- H_0: The model is not statistically significant.
- H_a: The model is statistically significant.

Test Statistic

$$F = \frac{MSR}{MSE} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \Big/ (k-1)}{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \Big/ (n-k)}$$

Critical Statistic
Is represented by F value in F table with (k – 1) degrees of freedom in the numerator and (n – k) degrees of freedom in the denominator.

- If $F \leq F_{critical}$ (calculated F is less than or equal to the critical F), we fail to reject the null. There is no statistically significant relationship between X and Y.
- If $F > F_{critical}$, we reject the null. There is a statistically significant relationship between X and Y.

Coefficient of Determination

R-squared or R^2 (also called coefficient of determination) measures the proportion of variability in the data that can be explained by the model.

R^2 ranges from 0 to 1. The higher R^2 is, the better the model can fit the actual data.

R^2 can be calculated with the formula:

$$R^2 = \frac{SS_{regression}}{SS_{total}} = 1 - \frac{SS_{error}}{SS_{total}} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Use JMP to Run a Simple Linear Regression

Case study: We want to see whether the score on exam one has any statistically significant relationship with the score on the final exam. If yes, how much impact does exam one have on the final exam?

Data File: "Simple Linear Regression.jmp"

Step 1: Determine the dependent and independent variables. Both should be continuous variables.

- Y (dependent variable) is the score of final exam.
- X (independent variable) is the score of exam one.

Step 2: Create a scatter plot to visualize whether there seems to be a linear relationship between X and Y.

1. Click Analyze -> Fit Y by X
2. Select FINAL as "Y, Response" and EXAM1 as "X, Factor"

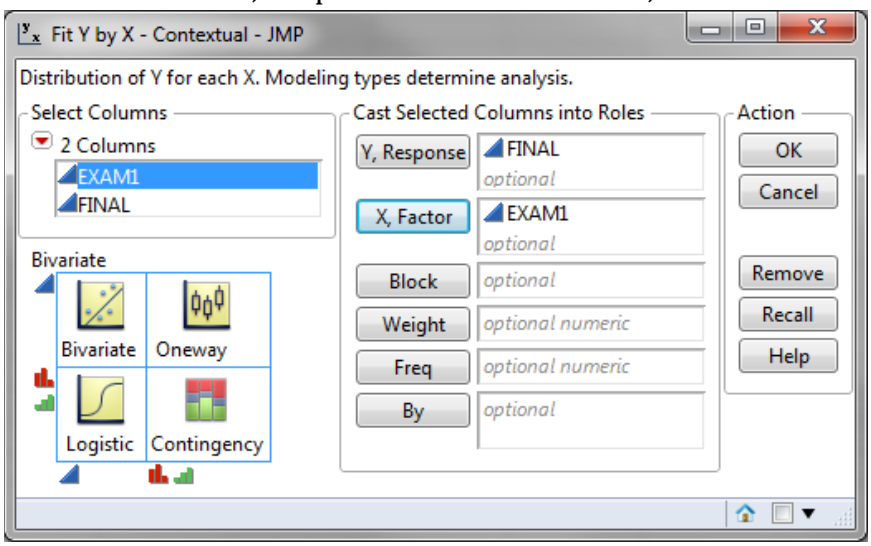


Fig. 4.13 Distribution for Y and X

3. Click "OK"

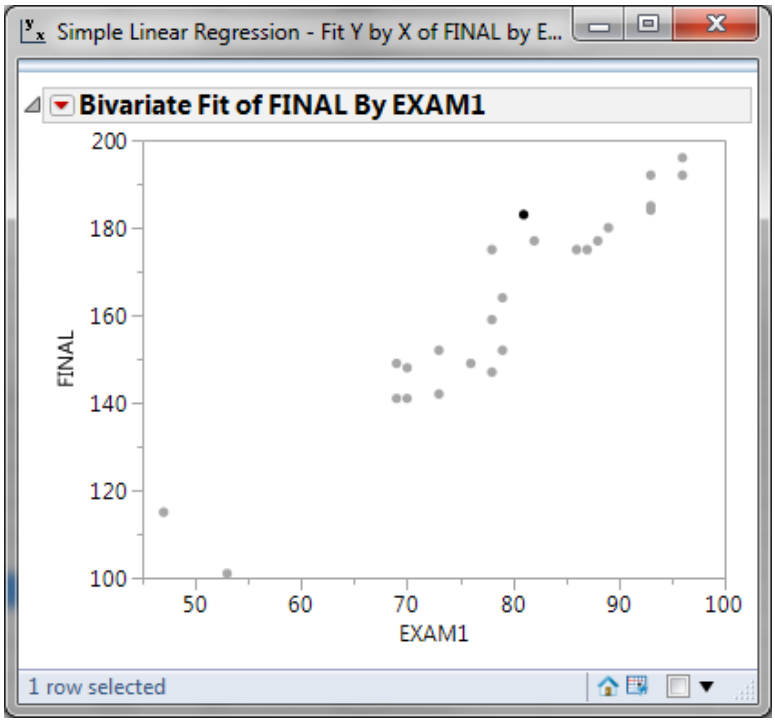


Fig. 4.14 JMP Scatter Plot Results

Based on the scatter plot, the relationship between exam one and final seems linear. The higher the score on exam one, the higher the score on the final. It appears you could "fit" a line through these data points.

Step 3: Run the simple linear regression analysis.

1. Click Analyze -> Fit Model
2. Select FINAL as Y and add EXAM1 as the factor

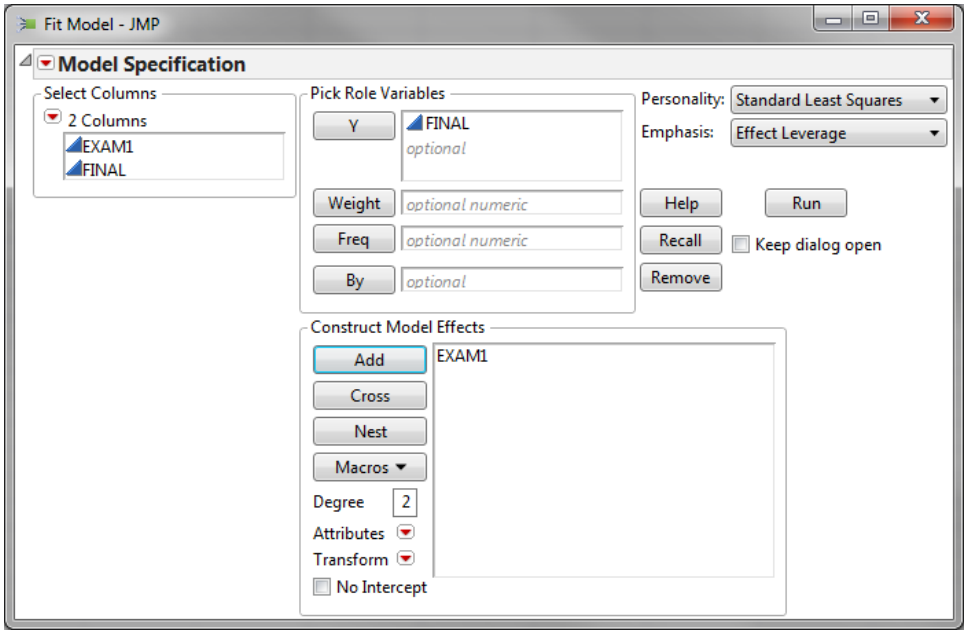


Fig. 4.15 Model Specifications

3. Click “Run”

Step 4: Check whether the model is statistically significant. If not significant, we will need to re-examine the predictor or look for new predictors before continuing.

- The Summary of Fit section provides R2 which measures the percentage of variation in the data set which can be explained by the model.
- 89.5% of the variability in the data can be accounted for by this linear regression model.
- Analysis of Variance section provides a ANOVA table covering degrees of freedom, sum of squares and mean square information for total, regression and error.
- The p-value of the F-test is lower than α level (0.05) indicating that the model is statistically significant.

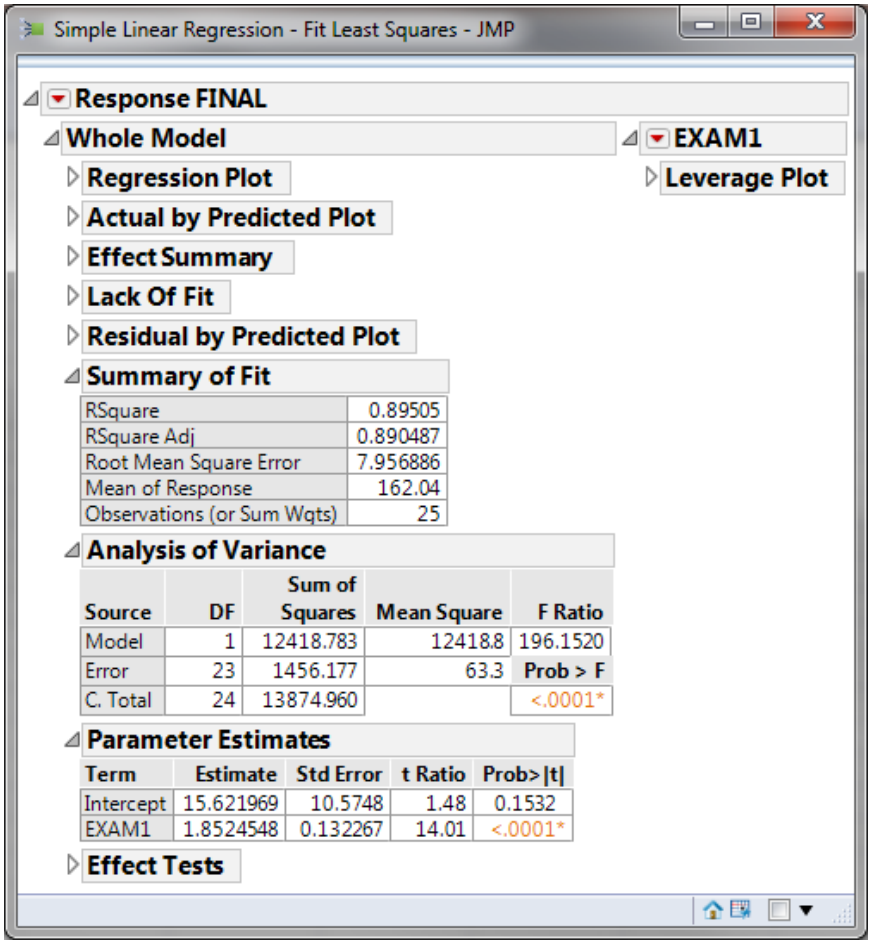


Fig. 4.16 Simple Linear Regression output

The p-value is 0.0001; therefore, we reject the null and claim the model is statistically significant. The R square value says that 89.5% of the variability can be explained by this model.

Step 5: Display regression equation

1. Click on the red triangle button next to “Response FINAL”
2. Select “Estimates” and then “Show Prediction Expression”

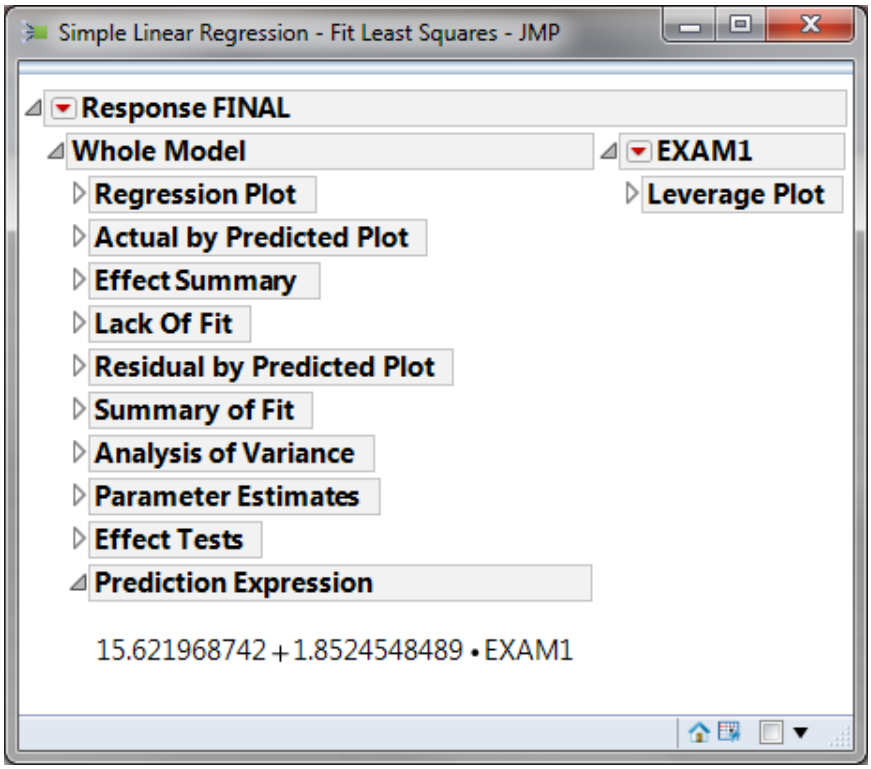


Fig. 4.17 Prediction Results for Exam 1

The estimates of slope and intercept are shown in “Parameter Estimate” section. In this example, Y = 15.6 + 1.85 × X, where X is the score on Exam 1 and Y is the final exam score. One unit increase in the score of Exam1 would increase the final score by 1.85.

Interpreting the Results
Let us say you are the professor and you want to use this prediction equation to estimate what two of your students might get on their final exam.

Rsquare Adj = 89.0%

- 89% of the variation in FINAL can be explained by EXAM1

P-value of the F-test = 0.000

- We have a statistically significant model

Prediction Equation: 15.6 + 1.85 × EXAM1

- 15.6 is the Y intercept, all equations will start with 15.6
- 1.85 is the EXAM1 Coefficient: multiply it by EXAM1 score

Because the model is significant, and it explains 89% of the variability, we can use the model to predict final exam scores based on the results of Exam1.

Let us assume the following:

- Student "A" exam 1 results were: 79
- Student "B" exam 1 results were: 94

Remember our prediction equation 15.6 + 1.85 × Exam1?

Now apply the equation to each student:

Student "A" Estimate: 15.6 + (1.85 × 79) = 161.8

Student "B" Estimate: 15.6 + (1.85 × 94) = 189.5

By simply replacing exam 1 scores into the equation we can predict their final exam scores. But the key thing about the model is whether or not it is useful. In this case, the professor can use the results to Figure out where to spend his time helping students.

4.1.4 Residuals Analysis

What are Residuals?

Residuals are the vertical differences between actual values and the predicted values or the "fitted line" created by the regression model.

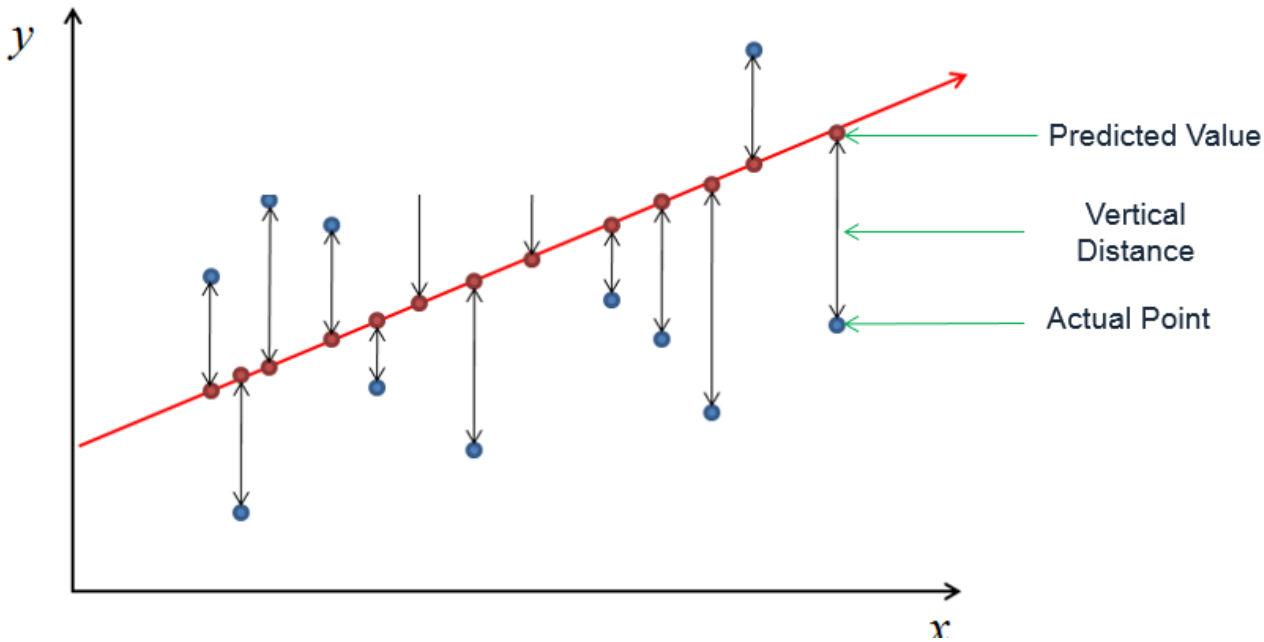


Fig. 4.18 Understanding Residuals

A residual is the vertical (y-axis) distance between an actual value and a value predicted by a regression model.

Why Perform Residuals Analysis?

Regression equations are generated on the basis of certain statistical assumptions. *Residuals analysis* helps to determine the validity of these assumptions.

The assumptions are:

- The residuals are normally distributed, mean equal to zero.
- The residuals are independent.
- The residuals have a constant variance.
- The underlying population relationship is linear.
- If residuals performance does not meet the requirements, we will need to rebuild the model by replacing the predictor with a new one, adding new predictors, building non-linear models, and so on.

A linear regression model is not always appropriate for the data; therefore, assess the appropriateness of the model by analyzing the residuals.

Use JMP to Perform Residuals Analysis

The residuals of the model are saved in the last column of the following data table.

Step 1: Generate the column of residuals in the data set.

1. Click on the red triangle button next to "Response FINAL"
2. Select "Save Columns" -> "Residuals"
3. Select "Save Columns -> "Predicted Values"
4. A new column will called "Residual FINAL" appears in the JMP data sheet along with a column labeled "PredictedFINAL."

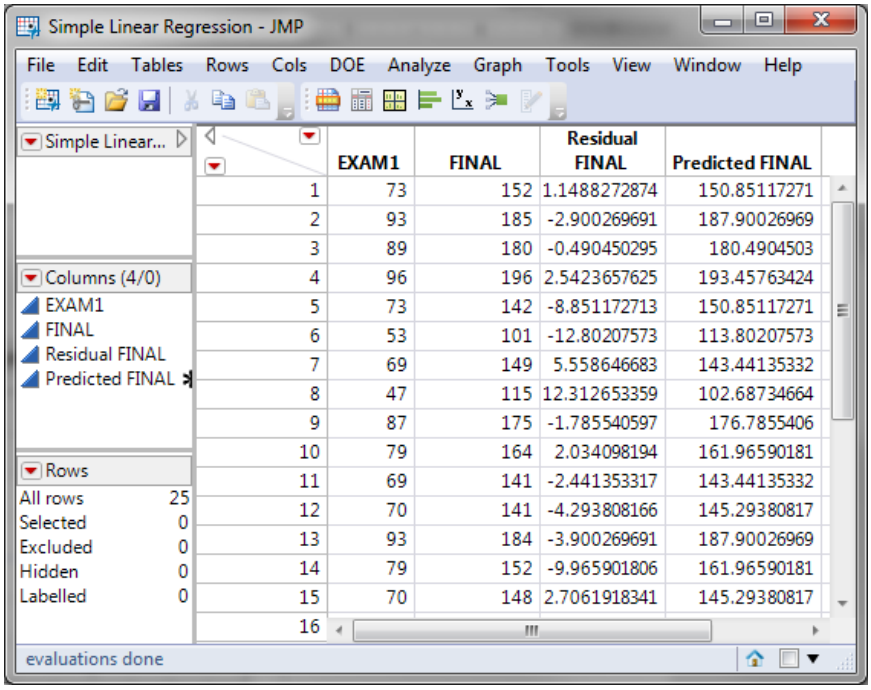


Fig. 4.19 Residual Analysis Output

Step 2: Check whether residuals are normally distributed around the mean of zero.

1. Click Analyze -> Distribution
2. Select Residual FINAL as "Y, Columns"

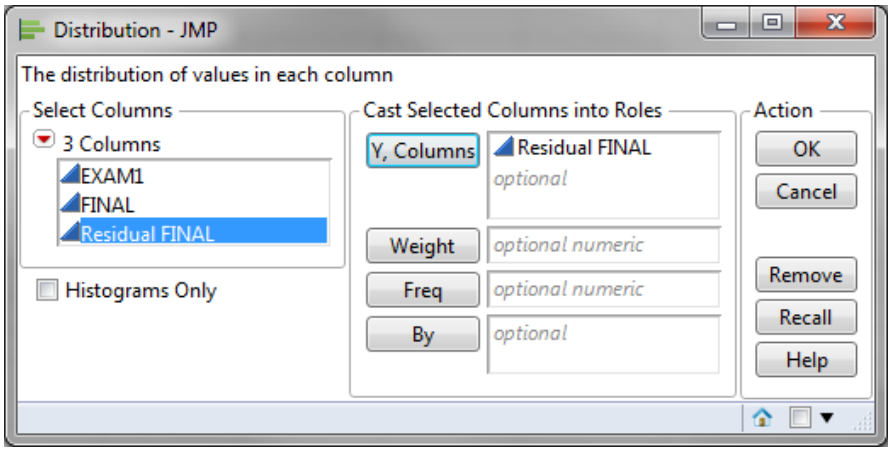


Fig. 4.20 Distribution selections

3. Click "OK"
4. Click on the red triangle button next to "Residual FINAL"
5. Select Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

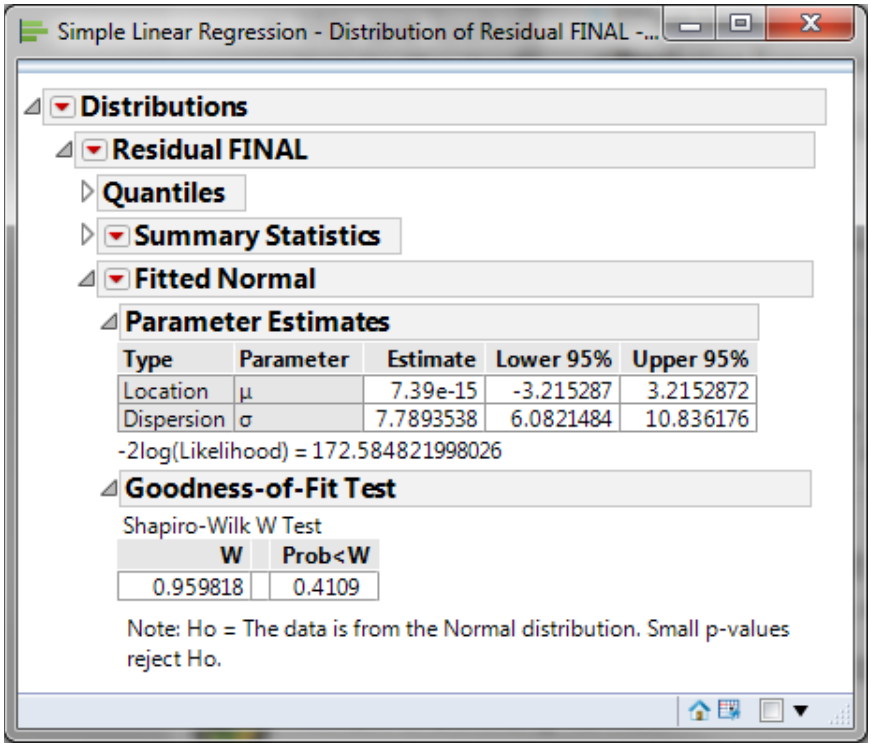


Fig. 4.21 Results of Residual Analysis

The Parameter Estimates section provides the mean of residuals. In this case, it is small enough to be considered zero. The Goodness-of-Fit Test section provides the Normality Test results.

Since the p-value (0.41) is greater than the alpha level (0.05), we fail to reject the null and the residuals are normally distributed.

- H0: The residuals are normally distributed.
- H1: The residuals are not normally distributed.

Step 3: If the data are in time order, run the IR chart to check whether residuals are independent.

1. Click Analyze -> Quality & Process -> Control Chart -> IR
2. Select Residual FINAL as "Process"

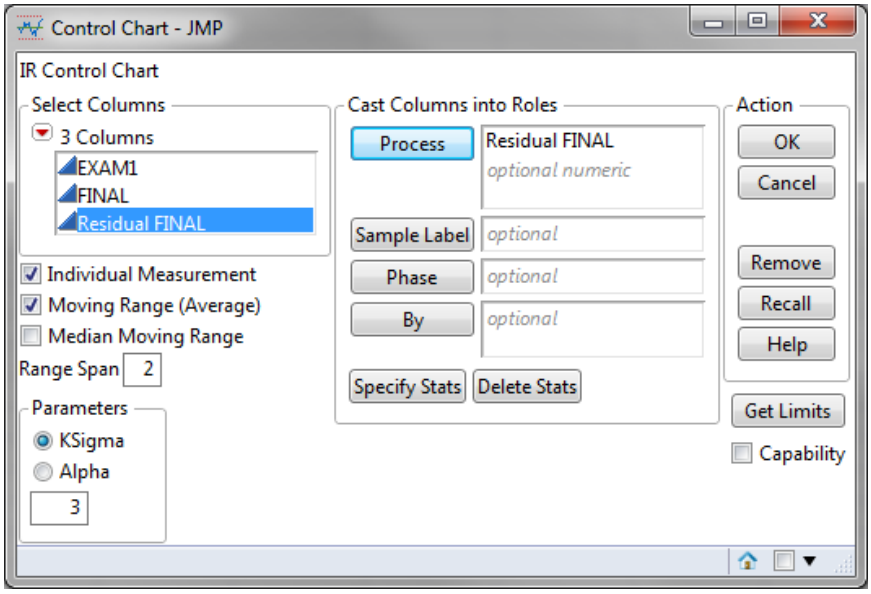


Fig. 4.22 Control Chart Residuals

3. Click "OK"
4. Click on the red triangle button next to "Individual Measurement of Residual FINAL"
5. Select Tests -> All Tests

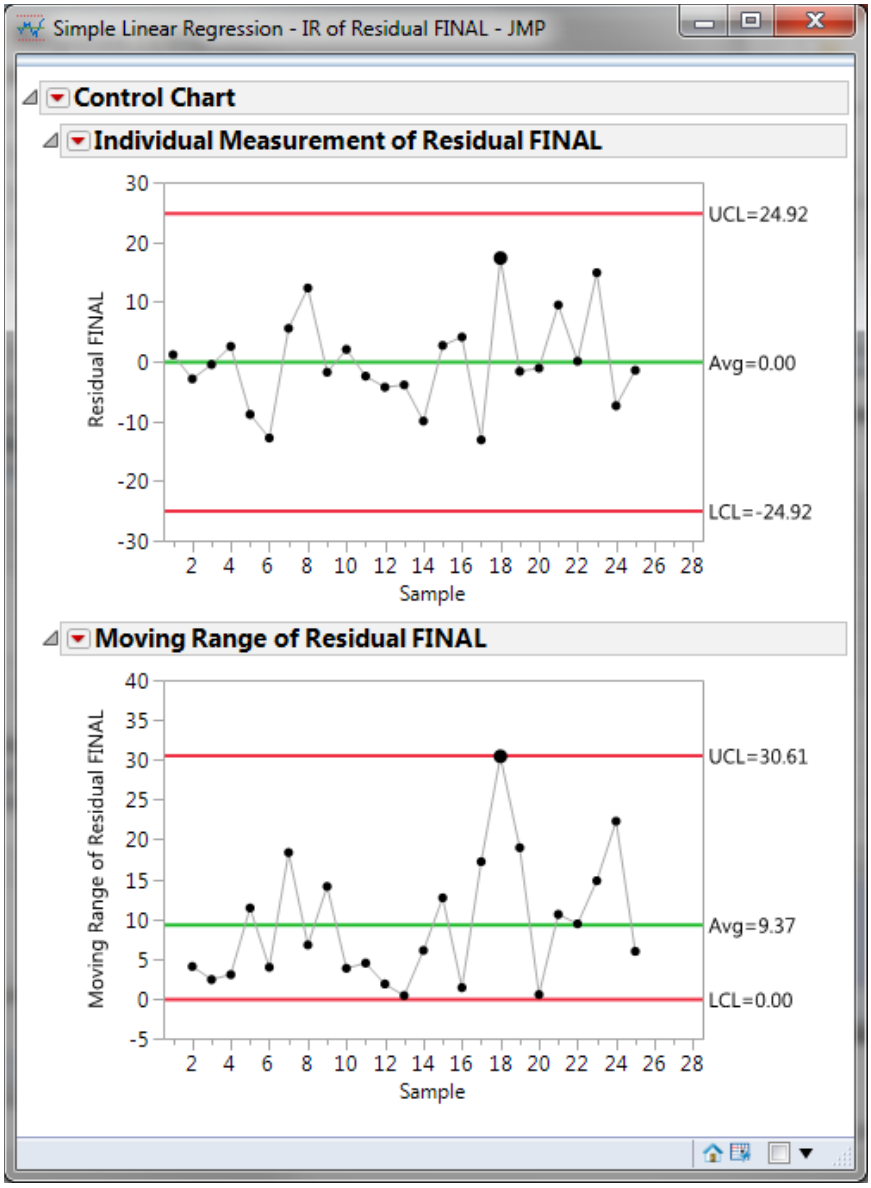


Fig. 4.23 Individual and MR Charts Output

If no data points are out of control in both the I-Chart and MR charts, the residuals are independent of each other. If the residuals are not independent, it is possible that some important predicators are not included in the model. In this example, since the IR chart is in control, residuals are independent.

Step 4: Check whether residuals have equal variance across the predicted responses.

1. Click “Graph” -> “Scatterplot Matrix”
2. A window named “Scatterplot Matrix” opens
3. Select “Residuals FINAL” as “Y”
4. Select “Predicted FINAL” as “X”

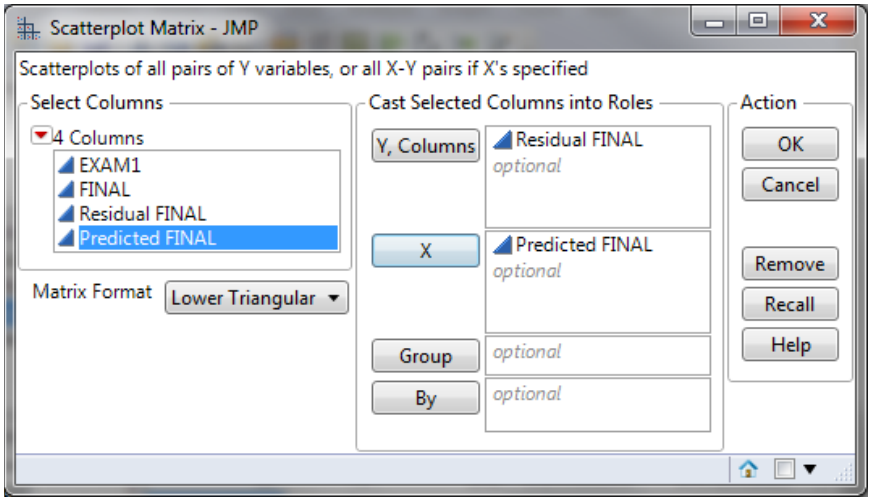


Fig. 4.24 Scatterplot Matrix Dialog Box

5. Click "OK"

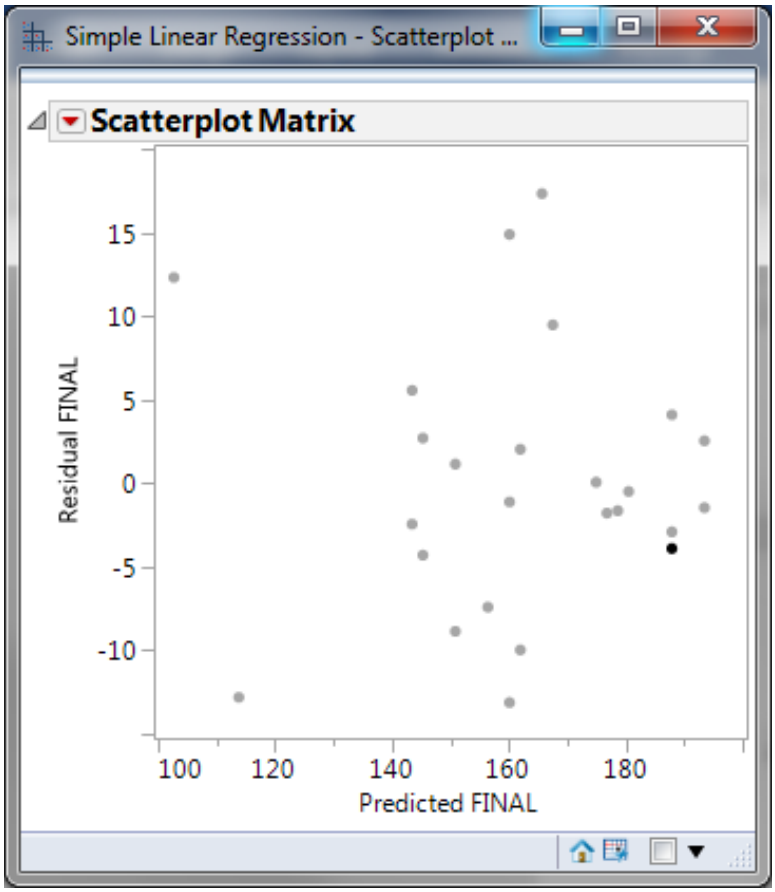


Fig. 4.25 Multiple Regression Residuals

We are looking for the pattern in which residuals spread out evenly around zero from the top to the bottom. Finally, we verify that the residuals have equal variance across the predicted responses. In this example, you can see how the residual values disperse evenly around zero.

Heteroscedasticity is the condition where the assumption of equal variance is violated, and can lead us to believe a variable is a predictor when it is not.

4.2 MULTIPLE REGRESSION ANALYSIS

4.2.1 NON-LINEAR REGRESSION

Linear and Non-Linear

The word *linear* originally comes from Latin word *linearis* meaning "created by lines."

A linear function in mathematics follows the following pattern (i.e., the output is proportional to its input):

$$f(x) = \alpha * x + \beta$$

$$f(x_1, x_2, \ldots, x_n) = \alpha_1 * x_1 + \alpha_2 * x_2 + \ldots + \alpha_n * x_n + \beta$$

A non-linear function does not follow the above pattern. There are usually exponents, logarithms, power, polynomial components, and other non-linear functions of the independent variables and parameters.

Non-Linear Relationships Using Linear Models

Many non-linear relationships can be transformed into linear relationships, and from there we can use linear regression methods to model the relationship. Some non-linear relationships cannot be transformed to linear ones and we need to apply other methods to build the non-linear models. In this section, we will focus on building non-linear regression models using linear transformation (i.e., transforming the independent or dependent variables or parameters to generate a linear function).

Assumptions in Using Non-Linear Regression

The population relationship is non-linear based on a reliable underlying theory. Across the range of all the possible values of the independent variables, the non-linear relationship applies. It is possible that at some extreme values the relationship between variables changes dramatically.

Non-Linear Functions: Transforming to Linear

Examples of non-linear functions that can be transformed to linear functions:

- Exponential Function
- Inverse Function
- Polynomial Function

- Power Function

The following section illustrates how to mathematically transform several non-linear functions to linear functions.

Exponential Function
Exponential Function:

$$Y = a \times b^X$$

Transformation:

$$\log Y = \log a + X \times \log b$$

Inverse Function
Inverse Function:

$$Y = a + b \times \frac{1}{X}$$

Transformation:

$$Y = a + b \times Z \text{ , where } Z = \frac{1}{X}$$

Polynomial Function
Polynomial Function:

$$Y = a + b \times X + c \times X^2$$

Transformation:

$$Y = a + b \times X + c \times Z \text{ , where } Z = X^2$$

Power Function
Power Function:

$$Y = a \times X^b$$

Transformation:

$$\log Y = \log\ a + b \times \log X$$

4.2.2 MULTIPLE LINEAR REGRESSION

What is Multiple Linear Regression?

Multiple linear regression is a statistical technique to model the relationship between one dependent variable and two or more independent variables by fitting the data set into a linear equation.

The difference between simple linear regression and multiple linear regression:

- Simple linear regression only has one predictor.
- Multiple linear regression has two or more predictors.

Multiple Linear Regression Equation

$$Y = \alpha_1 * X_1 + \alpha_2 * X_2 + \ldots + \alpha_p * X_p + \beta + e$$

Where:

- Y is the dependent variable (response)
- $X_1, X_2 \ldots X_p$ are the independent variables (predictors). There are p predictors in total.

Both dependent and independent variables are continuous.

- β is the intercept indicating the Y value when all the predictors are zeros
- $\alpha_1, \alpha_2 \ldots \alpha_p$ are the coefficients of predictors. They reflect the contribution of each independent variable in predicting the dependent variable.
- e is the residual term indicating the difference between the actual and the fitted response value.

Use JMP to Run a Multiple Linear Regression

Case study: We want to see whether the scores in exam one, two, and three have any statistically significant relationship with the score in final exam. If so, how are they related to final exam score? Can we use the scores in exam one, two, and three to predict the score in final exam?

Data File: "Multiple Regression Analysis.jmp"

Step 1: Determine the dependent and independent variables, all should be continuous. Y (dependent variable) is the score of final exam. X_1, X_2, and X_3 (independent variables) are the scores of exam one, two, and three respectively. All x variables are continuous.

Step 2: Start building the multiple linear regression model

1. Click Analyze -> Fit Model
2. Select FINAL as Y and EXAM1, EXAM2 and EXAM3 as predictors

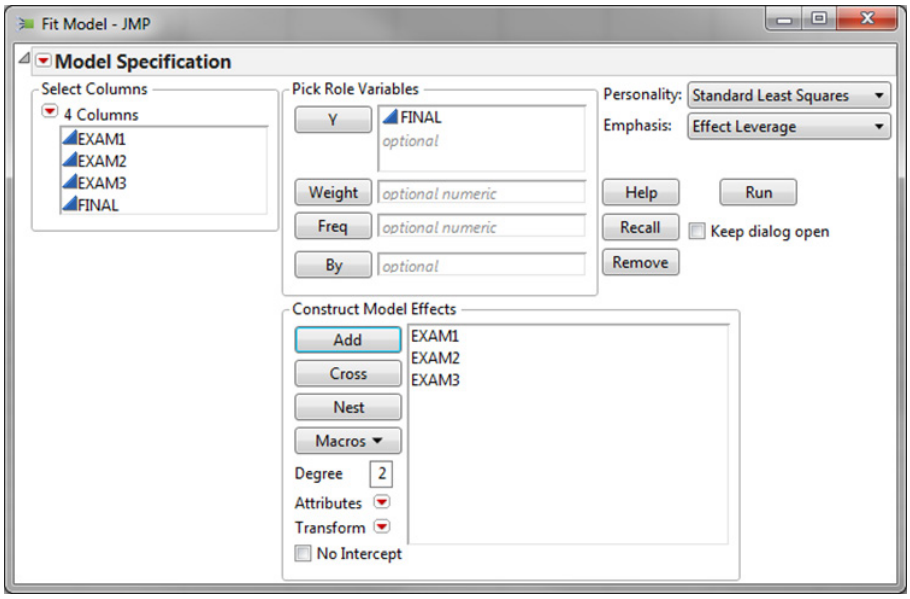


Fig. 4.26 Model Specifications

3. Click "Run"

Step 3: Check whether the whole model is statistically significant. If not, we need to re-examine the predictors or look for new predictors before continuing.

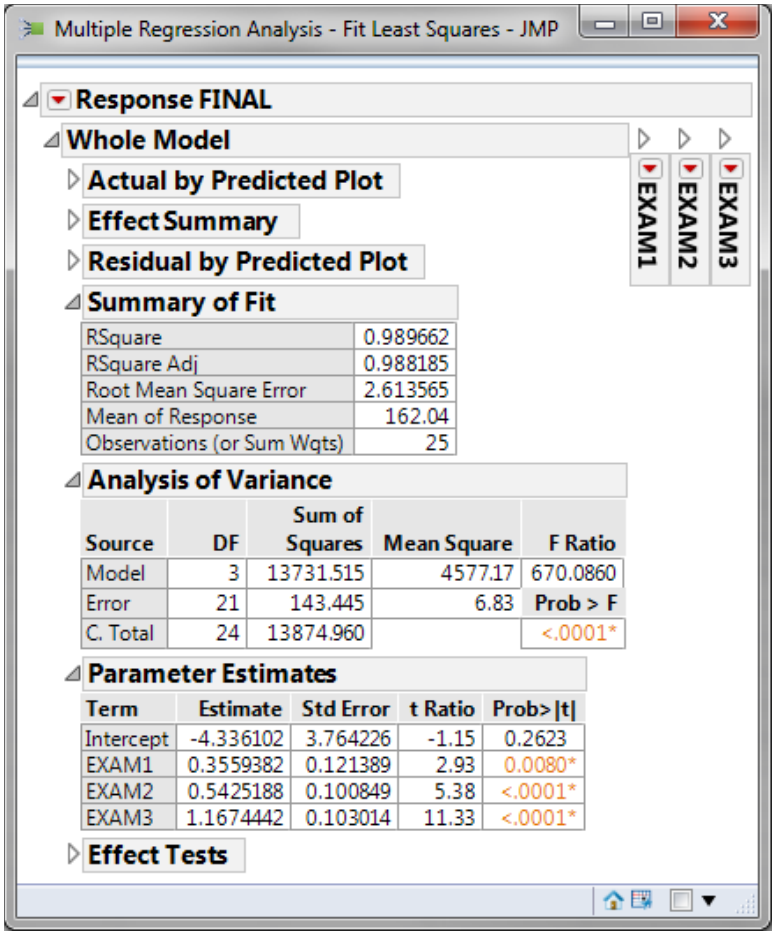


Fig. 4.27 Analysis of Variance for Model

- H_0: The model is not statistically significant (i.e., all the parameters of predictors are not significantly different from zeros).
- H_1: The model is statistically significant (i.e., at least one predictor parameter is significantly different from zero).

In this example, p-value is much smaller than alpha level (0.05), hence we reject the null hypothesis; the model is statistically significant.

Step 4: Check whether multicollinearity exists in the model.

1. Right click on the Parameter Estimates section.
2. Select Columns -> VIF
3. A new column with the heading "VIF" will appear in the table of Parameter Estimates.

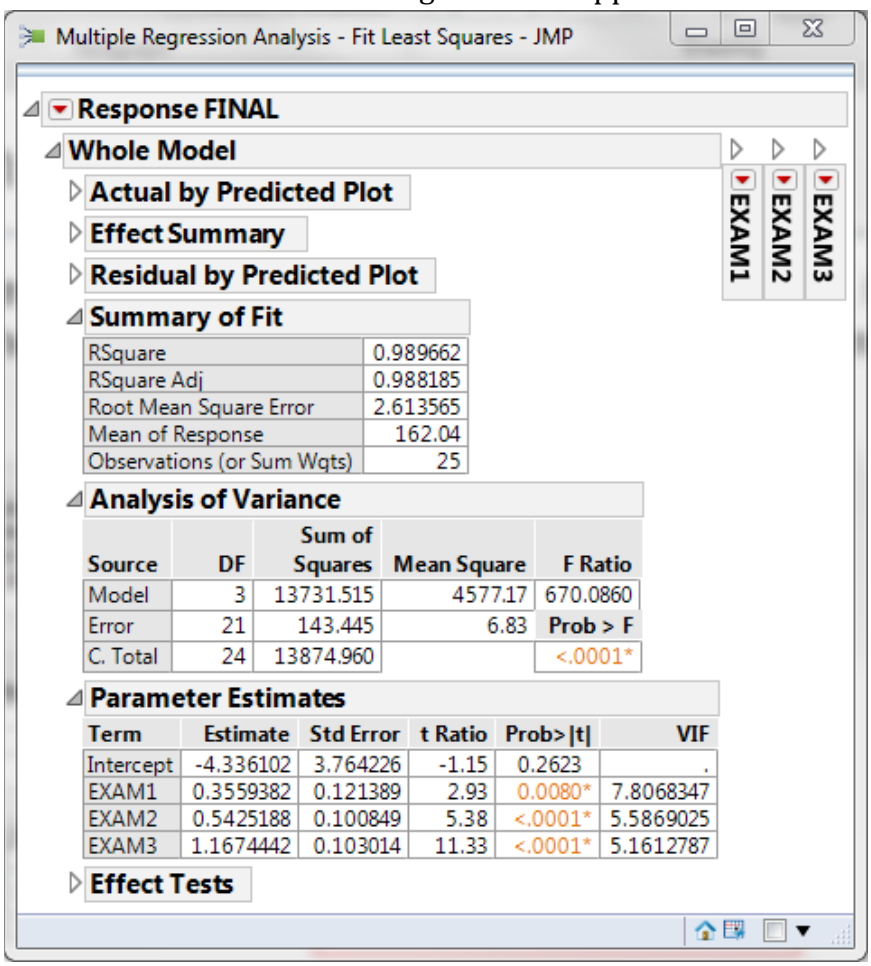


Fig. 4.28 Parameter Estimates

We use the VIF (Variance Inflation Factor) to determine if multicollinearity exists.

Multicollinearity

Multicollinearity is the situation when two or more independent variables in a multiple regression model are correlated with each other. Although multicollinearity does not necessarily reduce the predictability for the model as a whole, it may mislead the calculation

for individual independent variables. To detect multicollinearity, we use VIF (Variance Inflation Factor) to quantify its severity in the model.

Variance Inflation Factor (1)
VIF quantifies the degree of multicollinearity for each individual independent variable in the model.

VIF calculation:

Assume we are building a multiple linear regression model using p predictors.

$$Y = \alpha_1 \times X_1 + \alpha_2 \times X_2 + \ldots + \alpha_p \times X_p + \beta$$

Two steps are needed to calculate VIF for X_1.

Step 1: Build a multiple linear regression model for X_1 by using $X_2, X_3 \ldots X_p$ as predictors.

$$X_1 = a_2 \times X_2 + a_3 \times X_3 + \ldots + a_p \times X_p + b$$

Step 2: Use the R^2 generated by the linear model in step 1 to calculate the VIF for X_1.

$$VIF = \frac{1}{1 - R^2}$$

Apply the same methods to obtain the VIFs for other Xs. The VIF value ranges from one to positive infinity.

Variance Inflation Factor (2)
Rules of thumb to analyze variance inflation factor (VIF):

- If VIF = 1, there is no multicollinearity.
- If 1 < VIF < 5, there is small multicollinearity.
- If VIF ≥ 5, there is medium multicollinearity.
- If VIF ≥ 10, there is large multicollinearity.

How to Deal with Multicollinearity

1. Increase the sample size.
2. Collect samples with a broader range for some predictors.
3. Remove the variable with high multicollinearity and high p-value.
4. Remove variables that are included more than once.
5. Combine correlated variables to create a new one.

In this section, we will focus on removing variables with high VIF and high p-value.

Step 5: Deal with multicollinearity:

1. Identify a list of independent variables with VIF higher than 5. If no variable has VIF higher than 5, go to Step 6 directly.
2. Among variables identified in Step 5.1, remove the one with the highest p-value.
3. Run the model again, check the VIFs and repeat Step 5.1.

Note: we only remove one independent variable at a time.

In this example, all three predictors have VIF higher than 5. Among them, EXAM1 has the highest p-value. We will remove EXAM1 from the equation and run the model again.

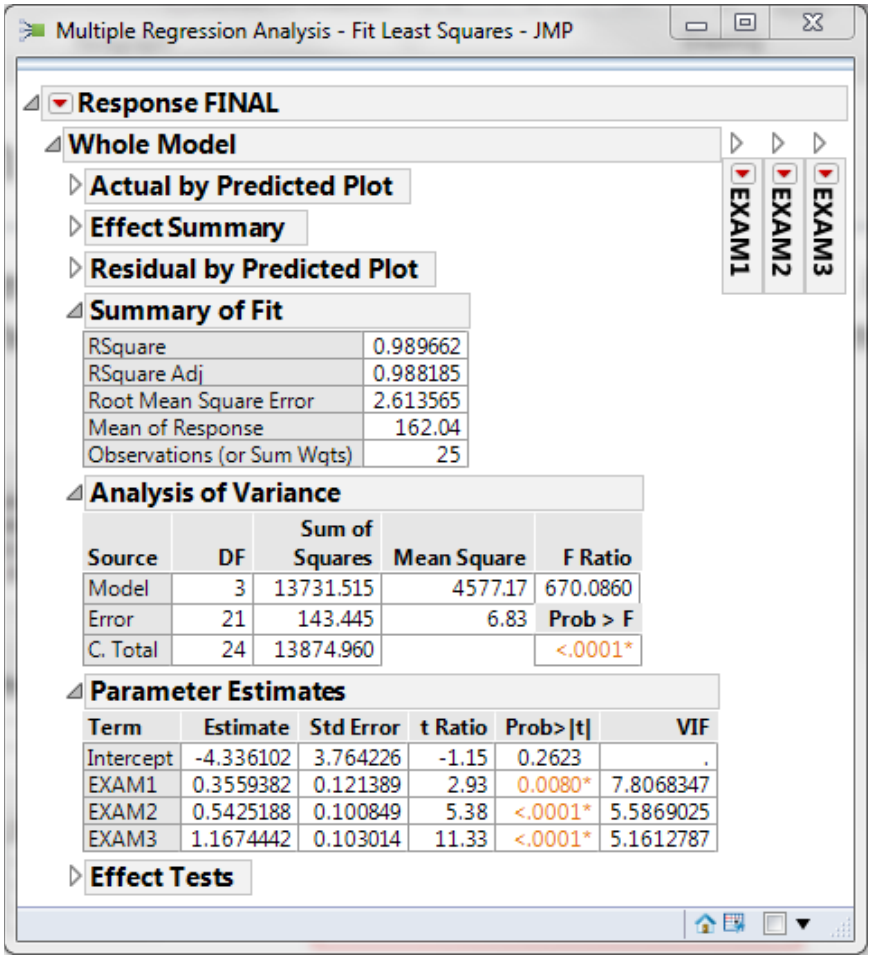


Fig. 4.29 Regression Analysis Final vs. Exam 1, 2, and 3

Run the new multiple linear regression with only two predictors (i.e., EXAM2 and EXAM3).

Check the VIFs of EXAM2 AND EXAM3. They are both smaller than 5; hence, there is little multicollinearity existing in the model.

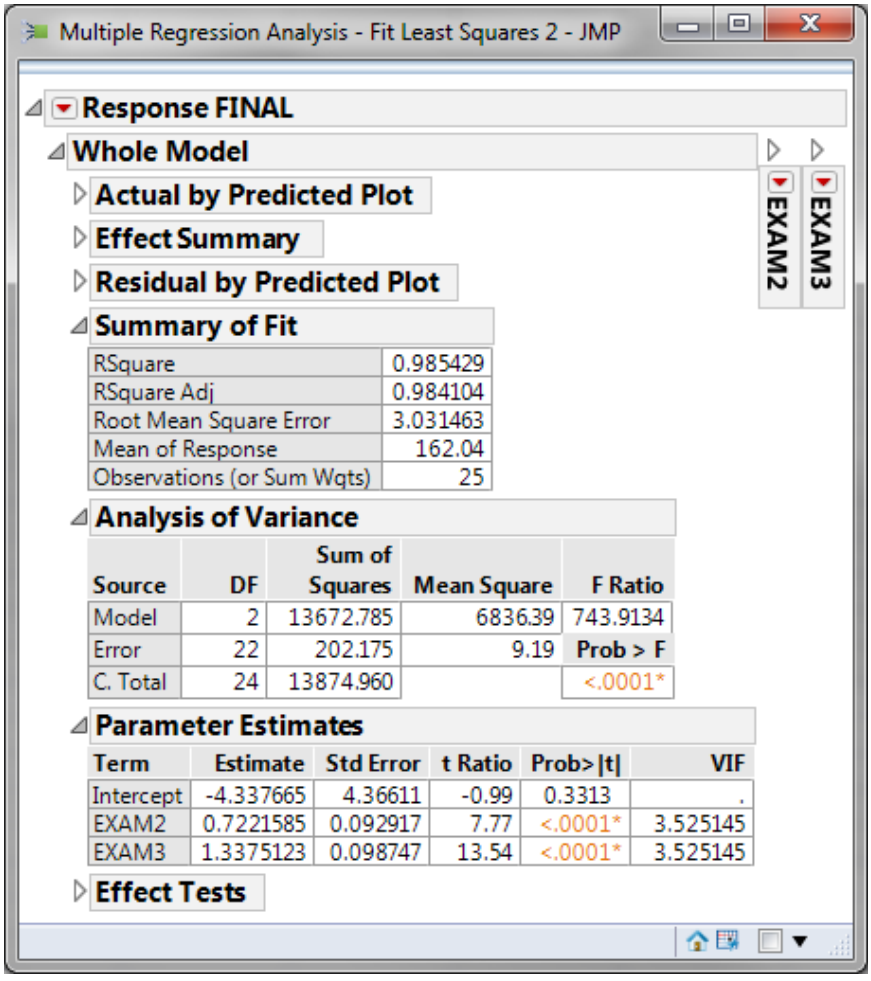


Fig. 4.30 Regression Analysis Final vs. Exam 2, and 3

Step 6: Identify the statistically insignificant predictors. Remove one insignificant predictor at a time and run the model again. Repeat this step until all the predictors in the model are statistically significant.

Insignificant predictors are the ones with p-value higher than alpha level (0.05). When p > alpha level, we fail to reject the null hypothesis; the predictor is not significant.

- H_0: The predictor is not statistically significant.
- H_1: The predictor is statistically significant.

As long as the p-value is greater than 0.05, remove the insignificant variables one at a time in the order of the highest p-value. Once one insignificant variable is eliminated from the model, we need to run the model again to obtain new p-values for other predictors left in the new model. In this example, both predictors' p-values are smaller than alpha level (0.05). As a result, we do not need to eliminate any variables from the model.

Step 7: Interpret the regression equation

1. Click on the red triangle button next to "Response FINAL"

2. Select “Estimates” and then “Show Prediction Expression”

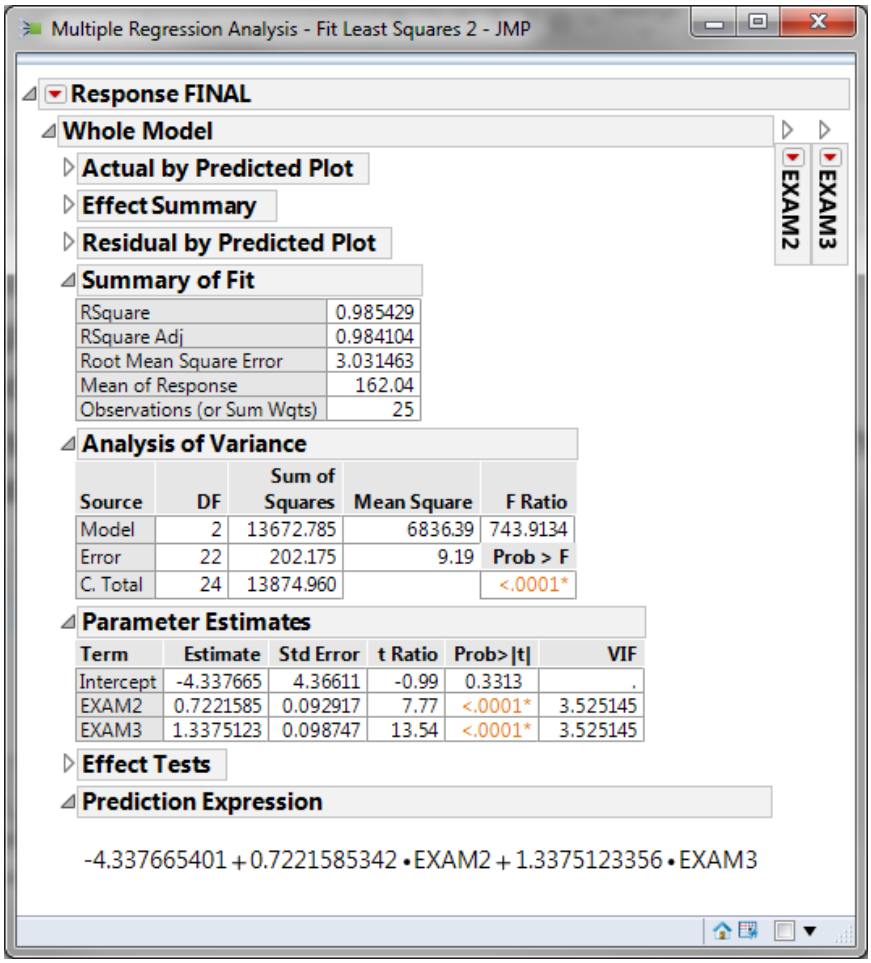


Fig. 4.31 Parameter Estimates for Exam 2 & 3 and Prediction Expressions for Exam 2 & 3

The multiple linear regression equation appears automatically at the top of the session window. “Parameter Estimates” section provides the estimates of parameters in the linear regression equation.

Now that we have removed multicollinearity and all of the insignificant predictors, we have the parameters for the regression equation.

Interpreting the Results
Rsquare Adj = 98.4%

- 98% of the variation in FINAL can be explained by the predictor variables EXAM2 & EXAM3.

P-value of the F-test = 0.000

- We have a statistically significant model.

Variables p-value:

- Both are significant (less than 0.05).

VIF

- EXAM2 and EXAM3 are both below 5; we're in good shape!

Equation: −4.34 + 0.722*EXAM2 + 1.34*EXAM3

- −4.34 is the Y intercept, all equations will start with −4.34.
- 0.722 is the EXAM2 coefficient; multiply it by EXAM2 score.
- 1.34 is the EXAM3 coefficient; multiply it by EXAM3 score.

Let us say you are the professor again, and this time you want to use your prediction equation to estimate what one of your students might get on their final exam.

Assume the following:

- Exam 2 results were: 84
- Exam 3 results were: 102

Use your equation: −4.34 + 0.722*EXAM2 + 1.34*EXAM3

Predict your student's final exam score:

−4.34 + (0.722*84) + (1.34*102) =−4.34 + 60.648 + 136.68 = 192.988

Nice work again! Now you can use your "magic" as the smart and efficient professor and allocate your time to other students because this one projects to perform much better than the average score of 162. Now that we know that exams two and three are statistically significant predictors, we can plug them into the regression equation to predict the results of the final exam for any student.

Confidence and Prediction Intervals

Prediction

The purpose of building a regression model is not only to understand what happened in the past but more importantly to *predict* the future based on the past. By plugging the values of independent variables into the regression equation, we obtain the estimation/prediction of the dependent variable.

Uncertainty of Prediction

We build the regression model using the sample data to describe as close as possible the true population relationship between dependent and independent variables. Due to noise in the data, the prediction will probably differ from the true response value. However, the true

response value might fall in a range around the prediction with some certainty. To measure the uncertainty of the prediction, we need confidence interval and prediction interval.

Confidence Interval

The *confidence interval* of the prediction is a range in which the population mean of the dependent variable would fall with some certainty, given specified values of the independent variables.

The width of confidence interval is related to:

- Sample size
- Confidence level
- Variation in the data

We build the model based on a sample set $\{y_1, y_2 \ldots y_n\}$. The confidence interval is used to estimate the value of the population mean μ of the underlying population. The focus of the confidence interval is represented by the unobservable population parameters. The confidence interval accounts for the uncertainty in the estimates of regression parameters.

Prediction Interval

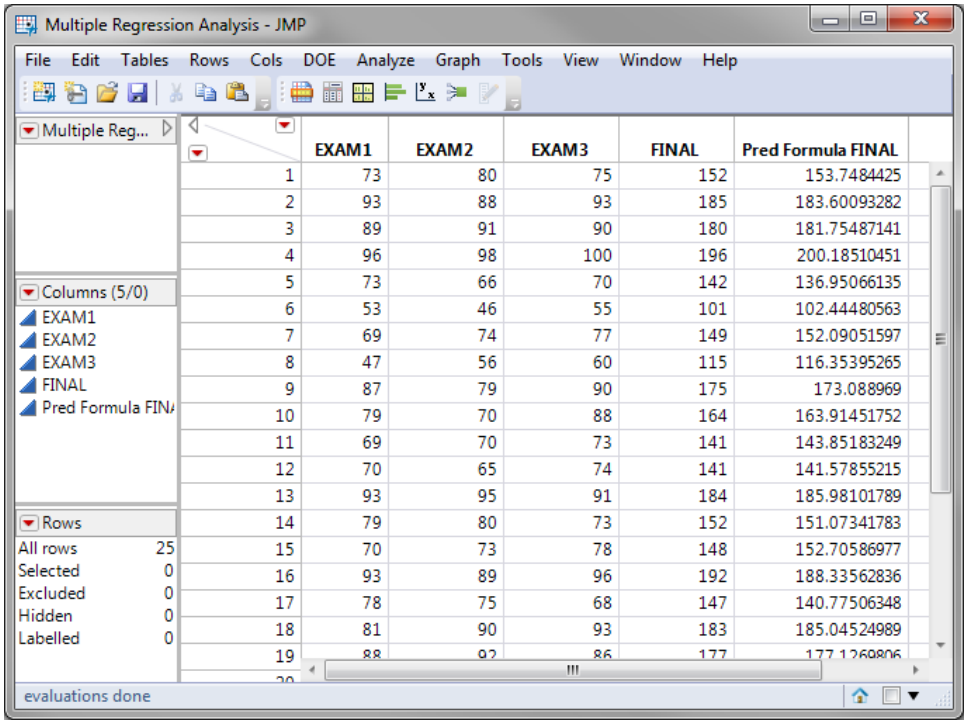
The *prediction interval* is a range in which future values of the dependent variable would fall with some certainty, given specified values of the independent variables. We build the model based on a sample set {y1, y2,..., yn}. The prediction interval is used to estimate the value of future observation y_{n+1}. The focus of the prediction interval is represented by the future observations. Prediction interval is wider than confidence interval because it accounts for the uncertainty in the estimates of regression parameters and the uncertainty of the new measurement.

Use JMP to Obtain Prediction and Confidence Interval

Steps in JMP to obtain prediction, confidence interval and prediction interval after running Regression tests:

1. Click on the red triangle button next to “Response FINAL”
2. Select Save Columns -> Prediction Formula

3. A new column named “Pred Formula FINAL” appears in the data table.



	EXAM1	EXAM2	EXAM3	FINAL	Pred Formula FINAL
1	73	80	75	152	153.7484425
2	93	88	93	185	183.60093282
3	89	91	90	180	181.75487141
4	96	98	100	196	200.18510451
5	73	66	70	142	136.95066135
6	53	46	55	101	102.44480563
7	69	74	77	149	152.09051597
8	47	56	60	115	116.35395265
9	87	79	90	175	173.088969
10	79	70	88	164	163.91451752
11	69	70	73	141	143.85183249
12	70	65	74	141	141.57855215
13	93	95	91	184	185.98101789
14	79	80	73	152	151.07341783
15	70	73	78	148	152.70586977
16	93	89	96	192	188.33562836
17	78	75	68	147	140.77506348
18	81	90	93	183	185.04524989
19	88	92	86	177	177.1269806

Fig. 4.32 Prediction output

4.2.3 Residuals Analysis

Remember what Residuals Are?

Residuals are the vertical difference between actual values and the predicted values or the “fitted line” created by the regression model.

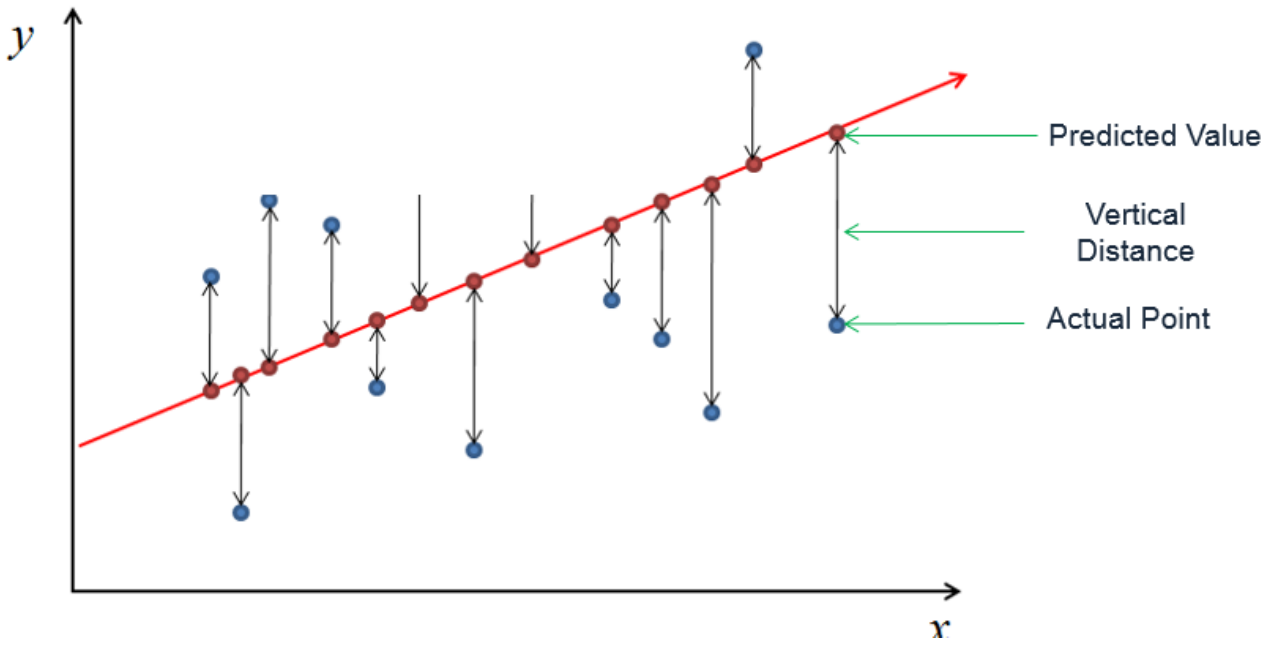


Fig. 4.33 Residuals

Residuals are an important component of evaluating the quality of a regression model. A residual is the vertical (y-axis) distance between an actual value and a value predicted by a regression model.

Why Perform Residuals Analysis?

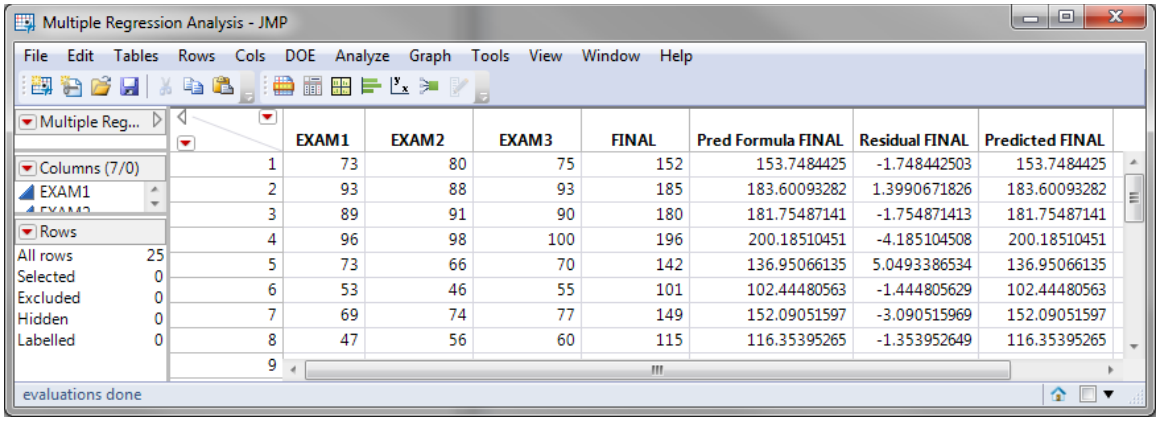
The *regression equation* generated based on the sample data can make accurate statistical inference only if certain assumptions are met. Residuals analysis can help to validate these assumptions. The following assumptions must be met to ensure the reliability of the linear regression model:

- The errors are normally distributed with mean equal to zero.
- The errors are independent.
- The errors have a constant variance.
- The underlying population relationship is linear.
- If the residuals performance does not meet the requirement, we will need to rebuild the model by replacing the predictors with new ones, adding new predictors, building non-linear models, and so on.

Use JMP to Perform Residuals Analysis

Step 1: Create the column of residuals in the data set.

1. Click on the red triangle button next to “Response FINAL”
2. Select “Save Columns” -> “Residuals”
3. Select “Save Columns” -> “Predicted Values”
4. New columns “Residual FINAL” and “Predicted FINAL” add to your data.



	EXAM1	EXAM2	EXAM3	FINAL	Pred Formula FINAL	Residual FINAL	Predicted FINAL
1	73	80	75	152	153.7484425	-1.748442503	153.7484425
2	93	88	93	185	183.60093282	1.3990671826	183.60093282
3	89	91	90	180	181.75487141	-1.754871413	181.75487141
4	96	98	100	196	200.18510451	-4.185104508	200.18510451
5	73	66	70	142	136.95066135	5.0493386534	136.95066135
6	53	46	55	101	102.44480563	-1.444805629	102.44480563
7	69	74	77	149	152.09051597	-3.090515969	152.09051597
8	47	56	60	115	116.35395265	-1.353952649	116.35395265
9							

Fig. 4.34 Residual output

Step 2: Check whether residuals are normally distributed around the mean of zero.

1. Click Analyze -> Distribution
2. Select Residual FINAL as "Y, Columns"

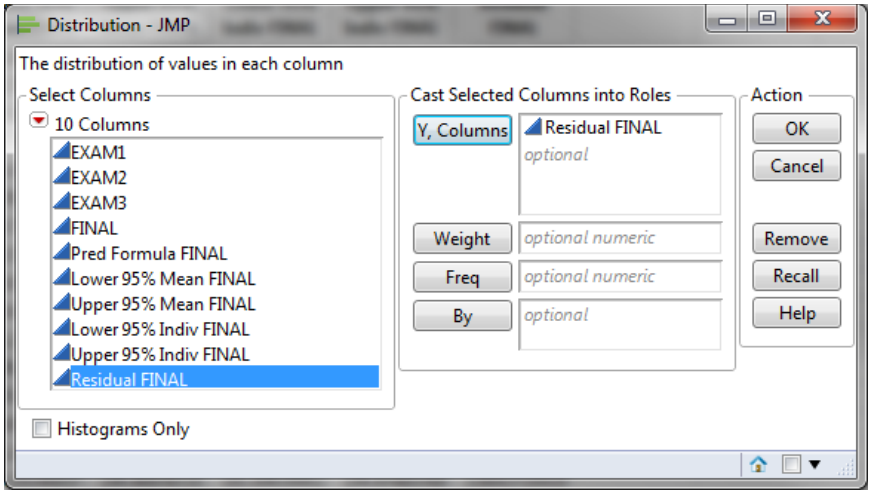


Fig. 4.35 Distribution selections

3. Click "OK"
4. Click on the red triangle button next to "Residual FINAL"
5. Select Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

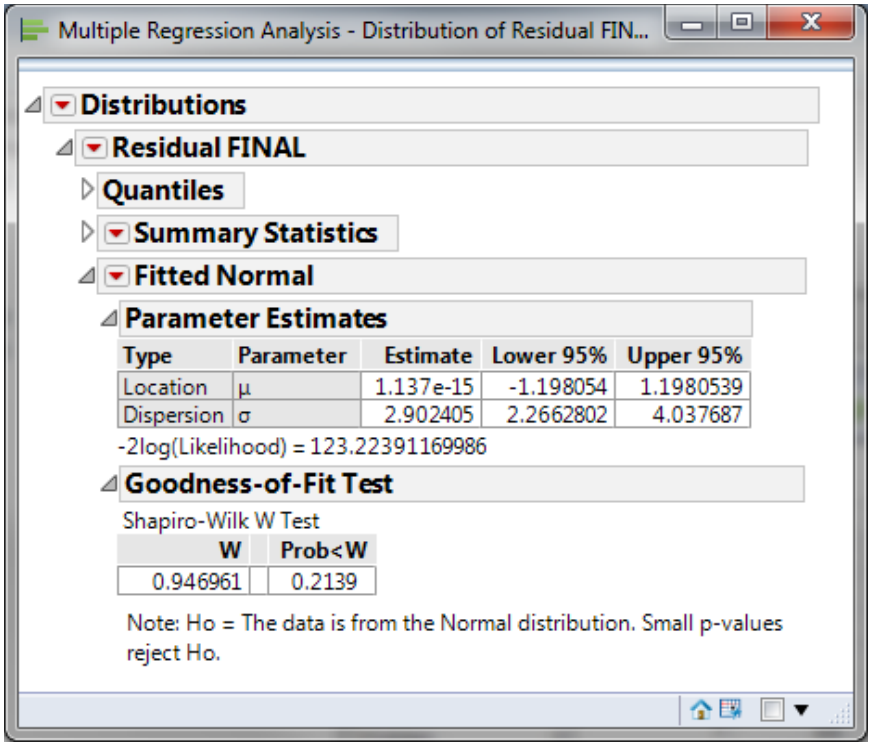


Fig. 4.36 Residual Analysis Results

The Parameter Estimates section provides the mean of residuals. In this case, it is small enough to be considered as zero.

The Goodness-of-Fit Test section provides the Normality Test results. Since p-value (0.21) is greater than the alpha level (0.05), we fail to reject the null and the residuals are normally distributed

- H_0: The residuals are normally distributed.
- H_1: The residuals are not normally distributed.

The residuals are normally distributed and the mean is near zero.

Step 3: If the data are in time order, run the IR chart to check whether residuals are independent.

1. Click Analyze -> Quality & Process -> Control Chart -> IR
2. Select Residual FINAL as “Process”

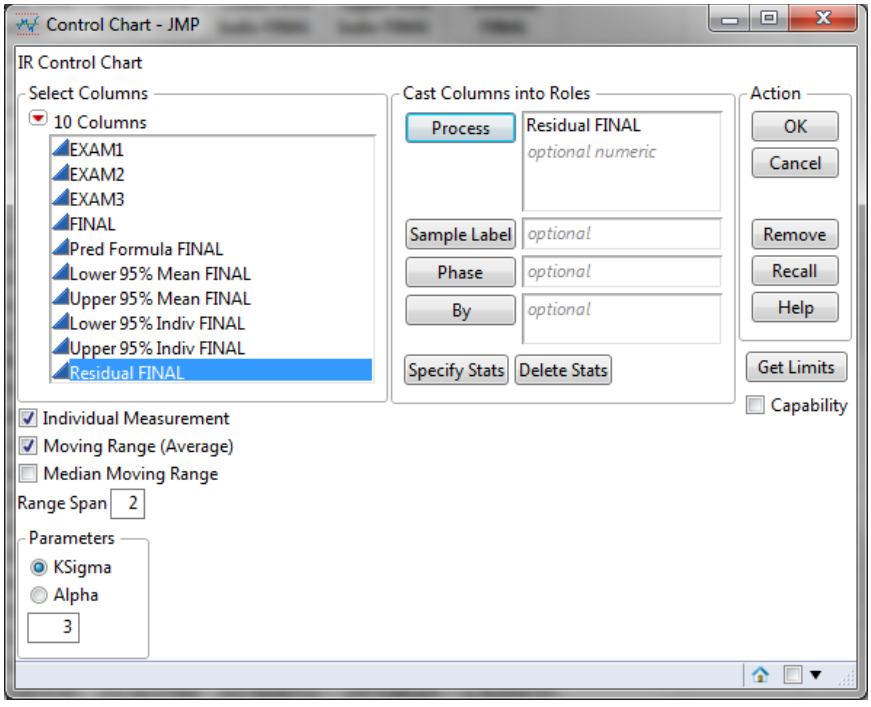


Fig. 4.37 Residuals selection

3. Click "OK"
4. Click on the red triangle button next to "Individual Measurement of Residual FINAL"
5. Select Tests -> All Tests

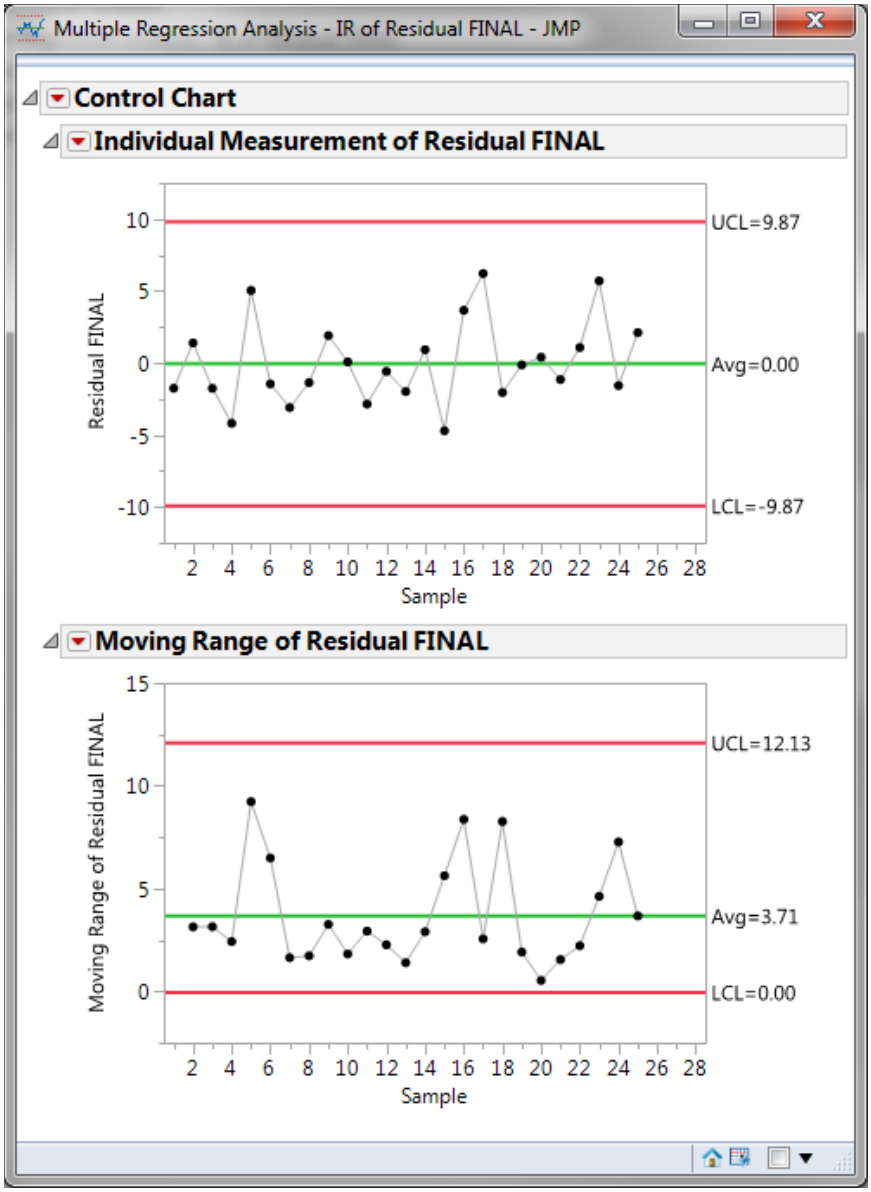


Fig. 4.38 Individual & MR Charts Output

If no data points are out of control in both the I-chart and MR chart, the residuals are independent of each other. If the residuals are not independent, it is possible that some important predictors are not included in the model. In this example, since the IR chart is in control, residuals are independent. The residuals are independent.

Step 4: Check whether residuals have equal variance across the predicted responses.

1. Click “Graph” -> “Scatterplot Matrix”
2. A window named “Scatterplot Matrix” opens
3. Select “Residuals FINAL” as “Y”
4. Select “Predicted FINAL” as “X”

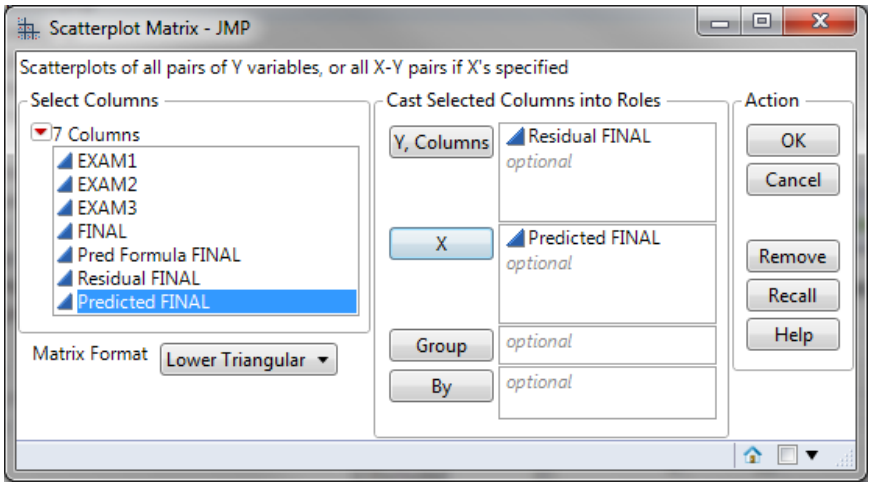


Fig. 4.39 Scatterplot Matrix Dialog Box

5. Click OK

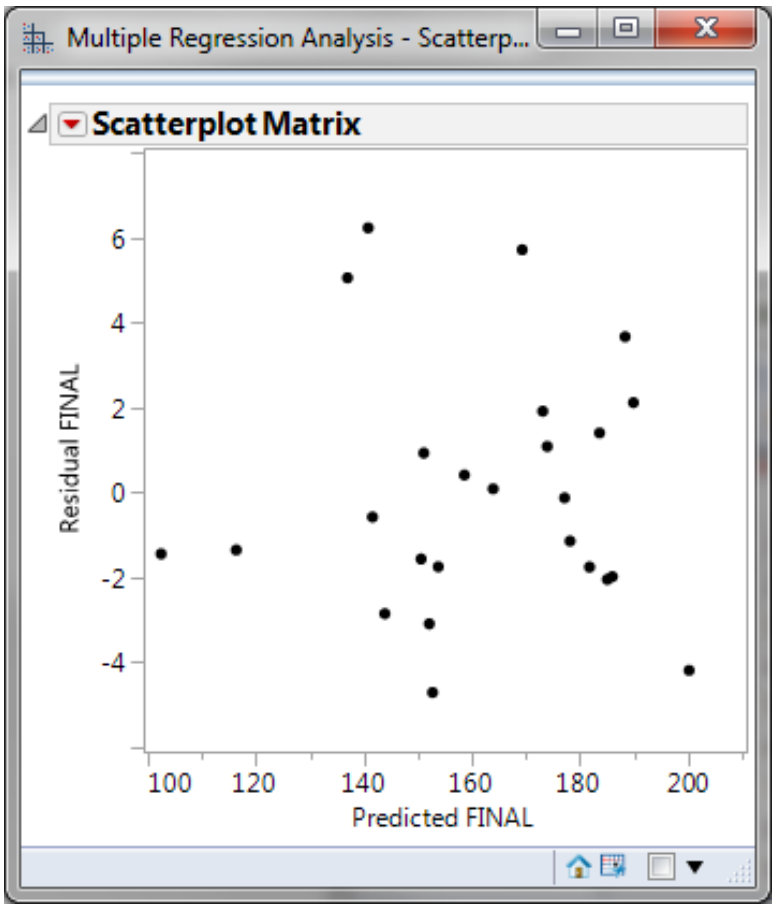


Fig. 4.40 Regular Residuals vs. Predicted Values for: FINAL

We are looking for the pattern in which residuals spread out evenly around zero from the top to the bottom. The residuals are evenly distributed around zero; so, residuals pass the tests.

4.2.4 Data Transformation

What is the Box-Cox Transformation?

Data transforms are usually applied so that the data appear to more closely meet assumptions of a statistical inference model to be applied or to improve the interpretability or appearance of graphs.

Power transformation is a class of transformation functions that raise the response to some power. For example, a square root transformation converts X to $X^{1/2}$.

Box-Cox transformation is a popular power transformation method developed by George E. P. Box and David Cox.

Box-Cox Transformation Formula

The formula of the Box-Cox transformation is:

$$\begin{cases} y = \dfrac{x^{\lambda} - 1}{\lambda} \text{ where } \lambda \neq 0 \\ y = \ln x \text{ where } \lambda = 0 \end{cases}$$

Where:

- y is the transformation result
- x is the variable under transformation
- λ is the transformation parameter.

Use JMP to Perform a Box-Cox Transformation

JMP provides the best Box-Cox transformation with an optimal λ that minimizes the model SSE (sum of squared error). Here is an example of how we transform the non-normally distributed response to normal data using Box-Cox method.

 ***Data File: "Box-Cox.jmp"***

Step 1: Test the normality of the original data set.

1. Click Analyze -> Distribution
2. Select Y as "Y, Column"

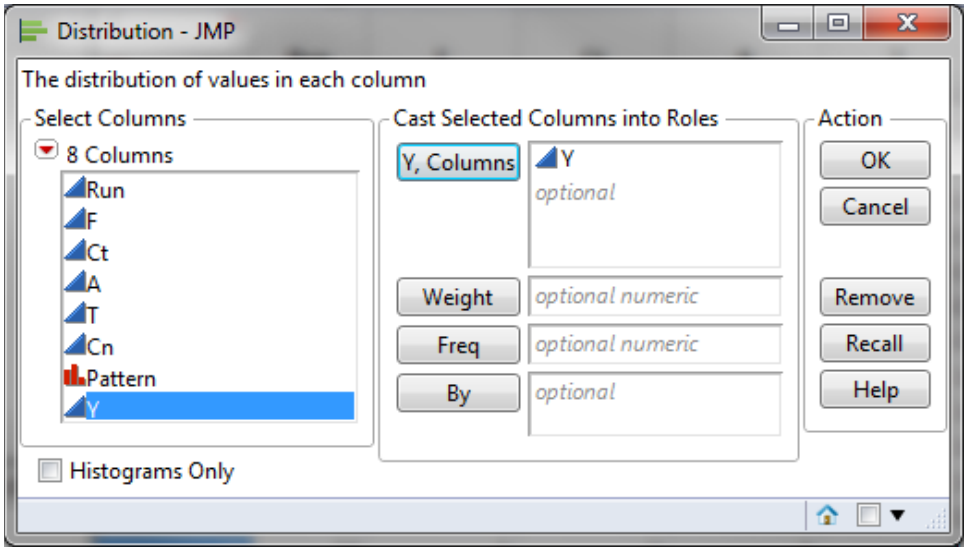


Fig. 4.41 Distribution Selection

3. Click "OK"
4. Click on the red triangle button next to "Y"
5. Select Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

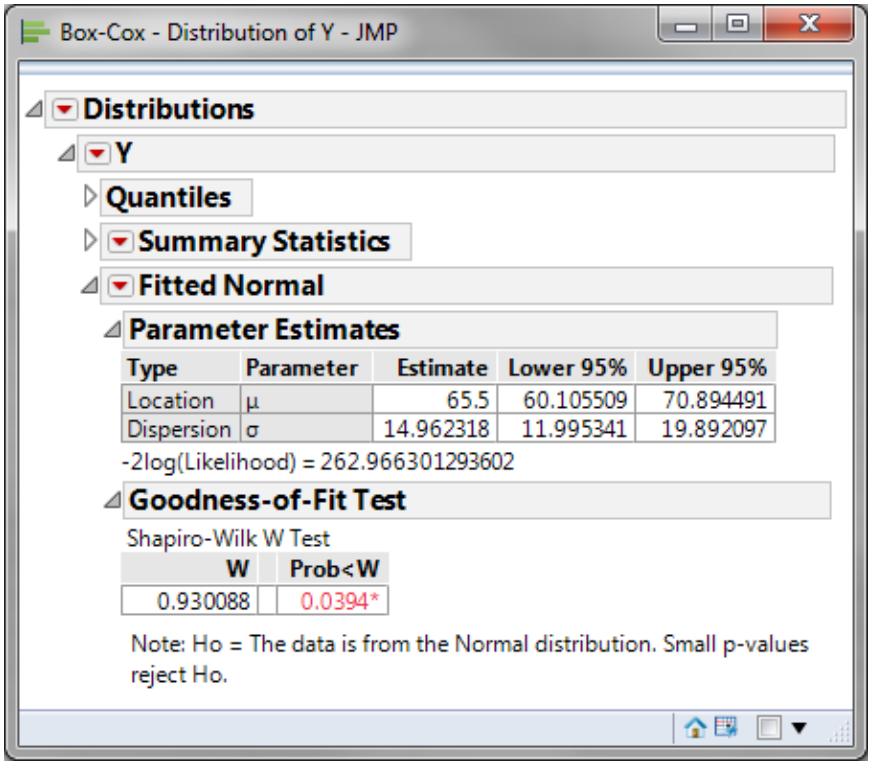


Fig. 4.42 Data output for Box-Cox Transformation

Normality Test:

- H0: The data are normally distributed.

- H1: The data are not normally distributed.

If p-value > alpha level (0.05), we fail to reject the null hypothesis. Otherwise, we reject the null. In this example, p-value = 0.039 < alpha level (0.05). The data are not normally distributed.

Step 2: Run the Box-Cox Transformation:

1. Click Analyze -> Fit Model
2. Select Y as "Y"
3. Click "OK"

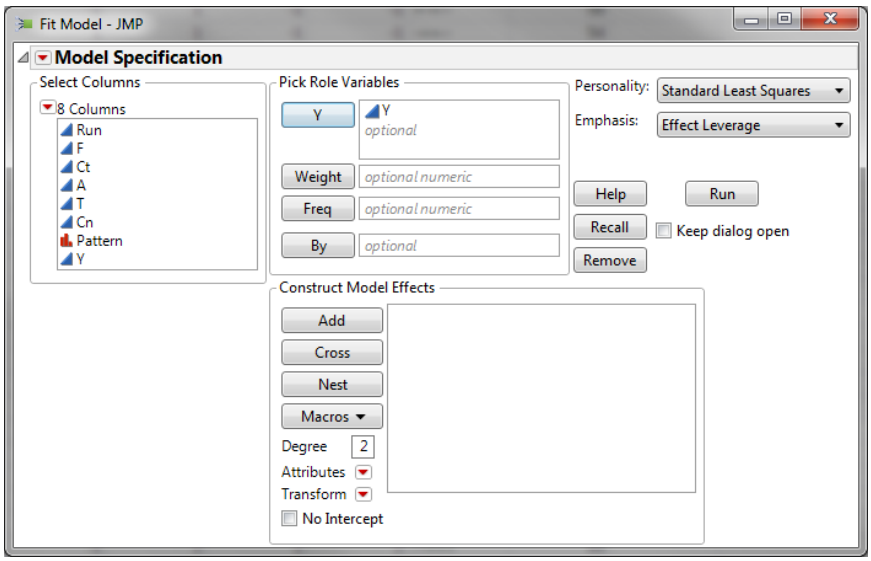


Fig. 4.43 Model Specifications

4. Click on the red triangle button next to "Response Y"
5. Select Factor Profiling -> Box Cox Y Transformation
6. A chart of sum of squared error (SSE) and λ is plotted at the bottom of the analysis output.
7. Click on the red triangle button next to "Box-Cox Transformations"
8. Select "Save Best Transformation"
9. A new column called "Y X" will be added to the data table.

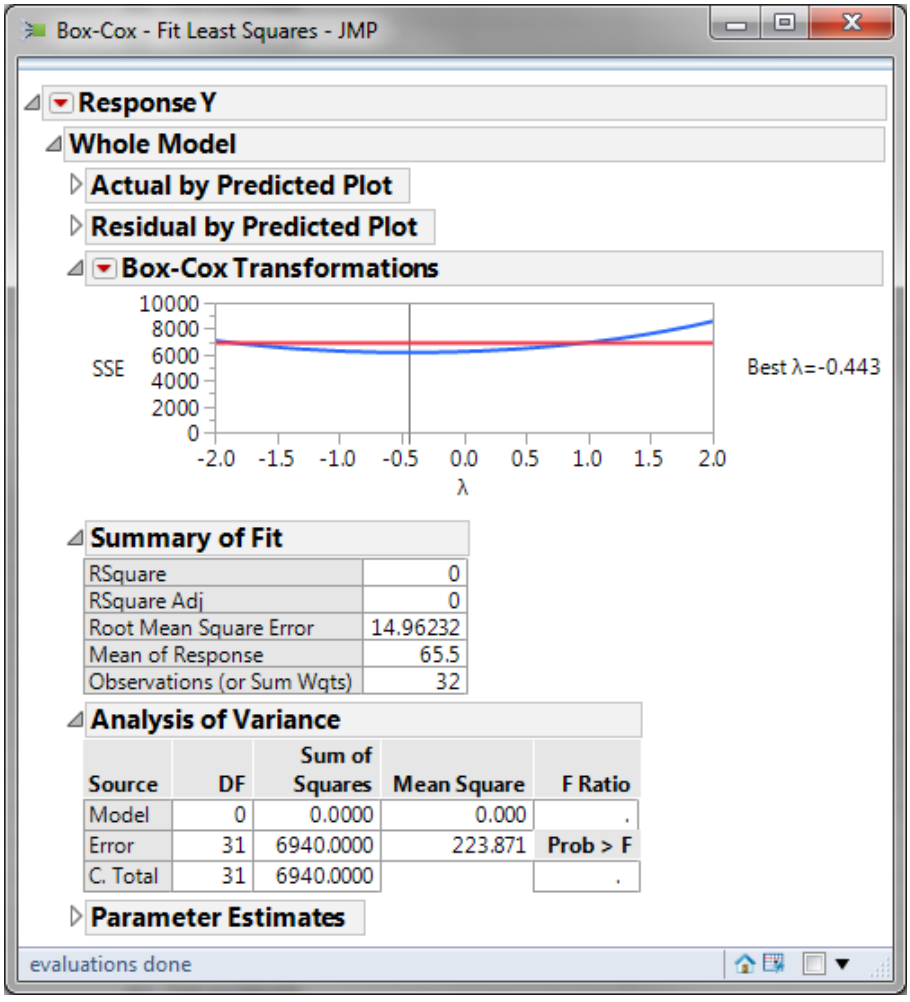
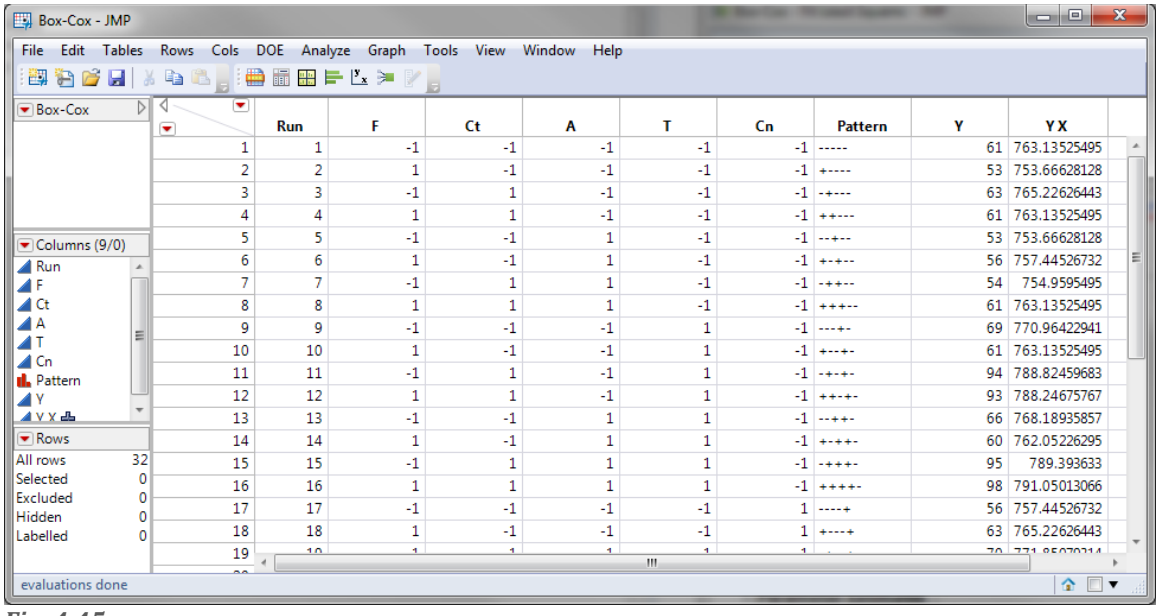


Fig. 4.44 Box-Cox Results

The software looks for the optimal value of lambda that minimizes the SSE (Sum of Squares of Error). In this case the minimum value is 0.12. The transformed Y can also be saved in another column.



	Run	F	Ct	A	T	Cn	Pattern	Y	Y X
1	1	-1	-1	-1	-1	-1	-----	61	763.13525495
2	2	1	-1	-1	-1	-1	+----	53	753.66628128
3	3	-1	1	-1	-1	-1	-+---	63	765.22626443
4	4	1	1	-1	-1	-1	++---	61	763.13525495
5	5	-1	-1	1	-1	-1	--+--	53	753.66628128
6	6	1	-1	1	-1	-1	+-+--	56	757.44526732
7	7	-1	1	1	-1	-1	-++--	54	754.9595495
8	8	1	1	1	-1	-1	+++--	61	763.13525495
9	9	-1	-1	-1	1	-1	---+-	69	770.96422941
10	10	1	-1	-1	1	-1	+--+-	61	763.13525495
11	11	-1	1	-1	1	-1	-+-+-	94	788.82459683
12	12	1	1	-1	1	-1	++-+-	93	788.24675767
13	13	-1	-1	1	1	-1	--++-	66	768.18935857
14	14	1	-1	1	1	-1	+-++-	60	762.05226295
15	15	-1	1	1	1	-1	-+++-	95	789.393633
16	16	1	1	1	1	-1	++++-	98	791.05013066
17	17	-1	-1	-1	-1	1	----+	56	757.44526732
18	18	1	-1	-1	-1	1	+---+	63	765.22626443
19									

Fig. 4.45

The Box-Cox Transformations chart presents how the sum of squared errors change across λ ranging from -2 to 2. Based on the chart, when λ is between -0.5 and 0.0, the transformation is the best with minimum SSE.

Step 3: Test the normality of the newly transformed data set.

1. Click Analyze -> Distribution
2. Select "Y X" as "Y, Column"

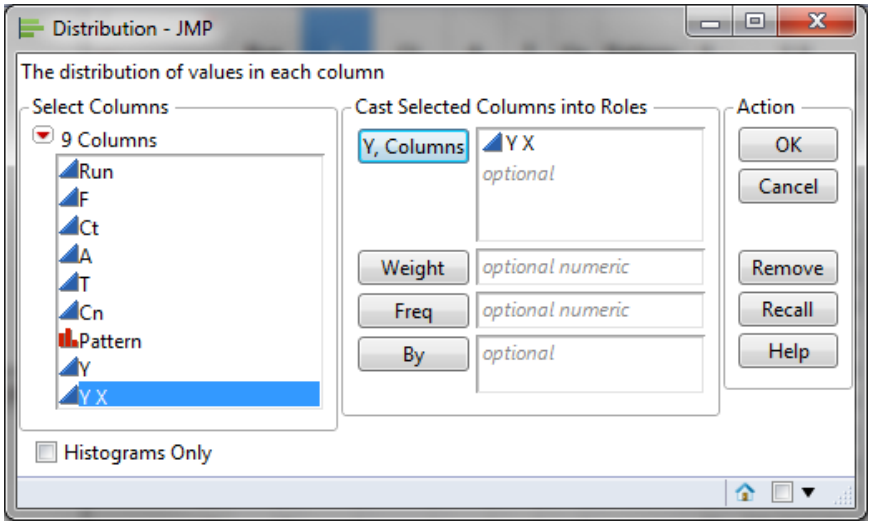


Fig. 4.46 Distribution Selection

3. Click "OK"
4. Click on the red triangle button next to "Y X"
5. Select Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Select "Goodness of Fit"

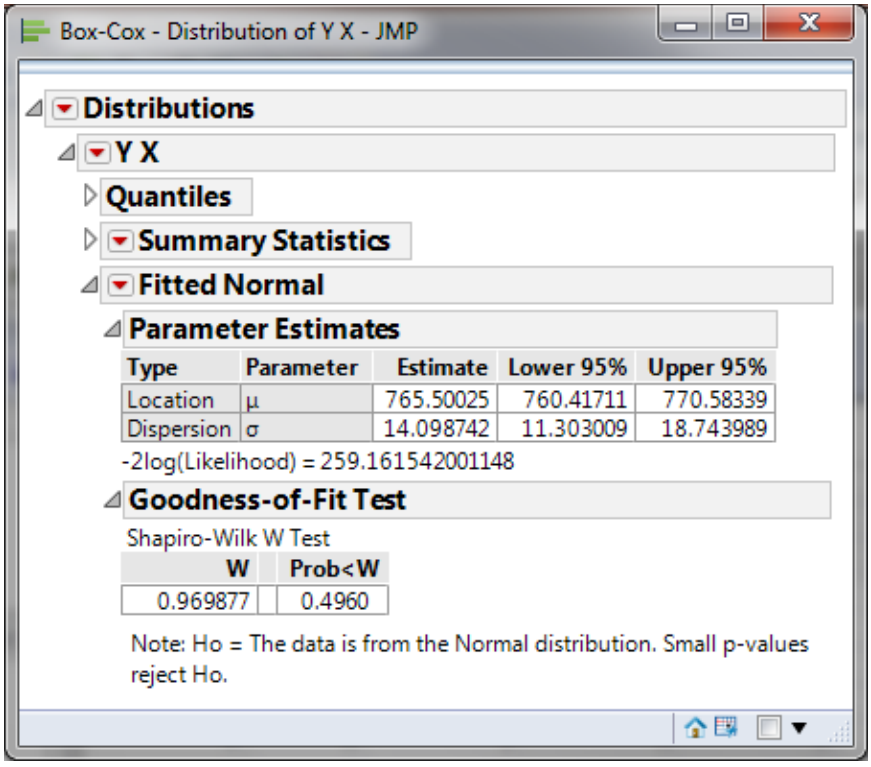


Fig. 4.47 Transformation Data

- H_0: The data are normally distributed.
- H_1: The data are not normally distributed.

If p-value > alpha level (0.05), we fail to reject the null. Otherwise, we reject the null. In this example, p-value = 0.4897 > alpha level (0.05). The data are normally distributed.

4.2.5 Stepwise Regression

What is Stepwise Regression?

Stepwise regression is a statistical method to automatically select regression models with the best sets of predictive variables from a large set of potential variables. There are different statistical methods used in stepwise regression to evaluate the potential variables in the model:

- F-test
- T-test
- R-square
- AIC

Three Approaches to Stepwise Regression

Forward Selection

Bring in potential predictors one by one and keep them if they have significant impact on improving the model.

Backward Selection
Try out potential predictors one by one and eliminate them if they are insignificant to improve the fit.

Mixed Selection
Is a combination of both forward selection and backward selection. Add and remove variables based on pre-defined significance threshold levels.

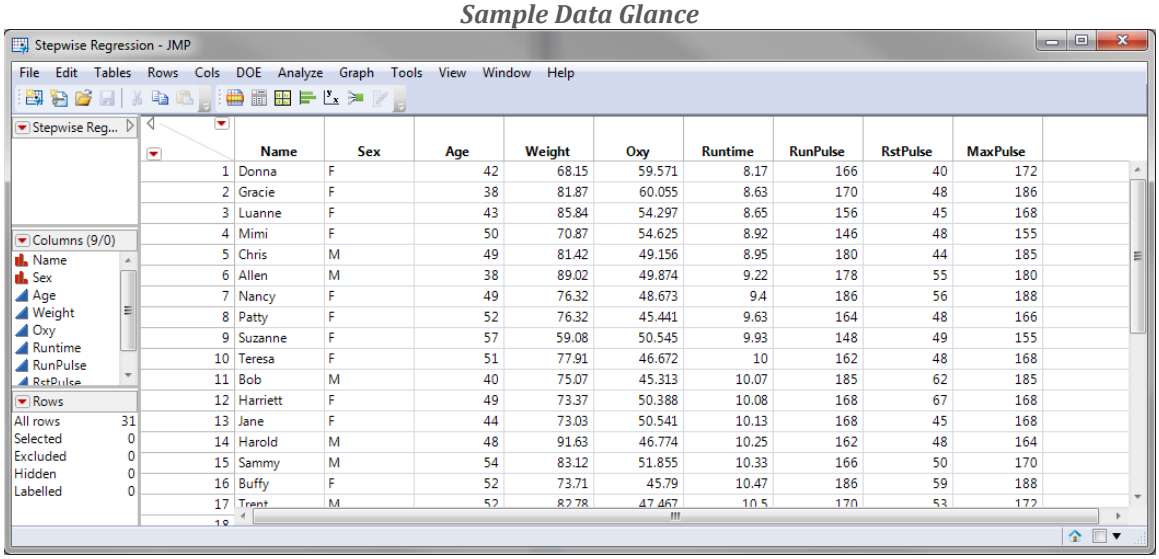
How to Use JMP to Run a Stepwise Regression

Case study: We want to build a regression model to predict the oxygen uptake of a person who runs 1.5 miles. The potential predictors are: Age

- Age
- Weight
- Runtime
- Runpulse
- RstPulse
- MaxPuls

Data File: "StepwiseRegression.jmp"

Sample Data Glance



	Name	Sex	Age	Weight	Oxy	Runtime	RunPulse	RstPulse	MaxPulse
1	Donna	F	42	68.15	59.571	8.17	166	40	172
2	Gracie	F	38	81.87	60.055	8.63	170	48	186
3	Luanne	F	43	85.84	54.297	8.65	156	45	168
4	Mimi	F	50	70.87	54.625	8.92	146	48	155
5	Chris	M	49	81.42	49.156	8.95	180	44	185
6	Allen	M	38	89.02	49.874	9.22	178	55	180
7	Nancy	F	49	76.32	48.673	9.4	186	56	188
8	Patty	F	52	76.32	45.441	9.63	164	48	166
9	Suzanne	F	57	59.08	50.545	9.93	148	49	155
10	Teresa	F	51	77.91	46.672	10	162	48	168
11	Bob	M	40	75.07	45.313	10.07	185	62	185
12	Harriett	F	49	73.37	50.388	10.08	168	67	168
13	Jane	F	44	73.03	50.541	10.13	168	45	168
14	Harold	M	48	91.63	46.774	10.25	162	48	164
15	Sammy	M	54	83.12	51.855	10.33	166	50	170
16	Buffy	F	52	73.71	45.79	10.47	186	59	188
17	Trent	M	52	82.78	47.467	10.5	170	53	172

Fig. 4.48 Stepwise Data at a glance

Steps to run stepwise regression in JMP:

Step 1:

1. Click Analyze -> Fit Model
2. Model Specification window appears.
3. Select "Oxy" as the Y and add the potential factors to the model effects box
4. Select "Stepwise" in the "Personality" dropdown box

5. Click “Run”

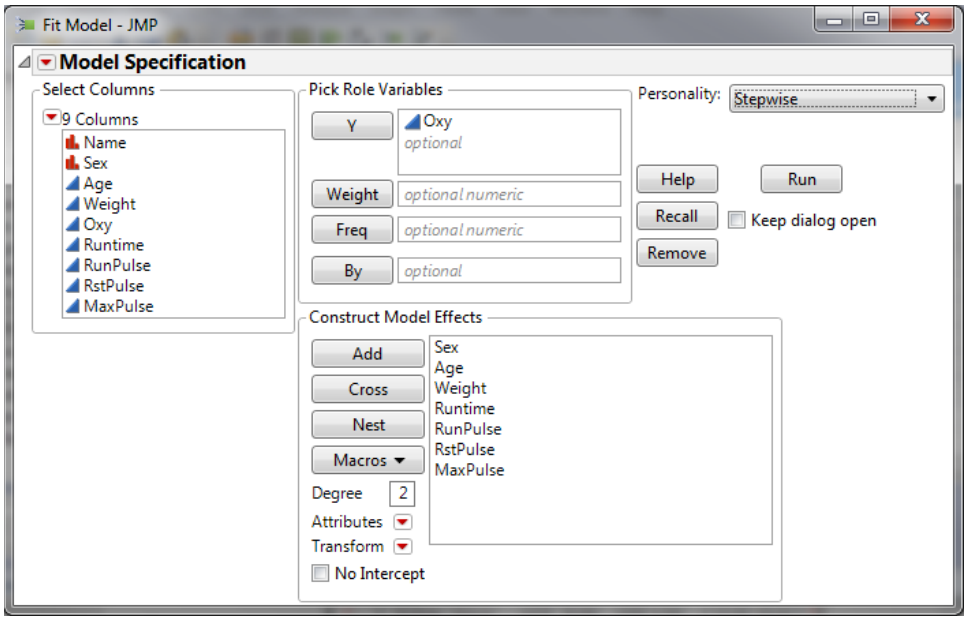


Fig. 4.49 Run Model

Step 2:

1. The “Stepwise Fit” page shows up
2. Select P-value Threshold for Stopping Rule
3. Enter the “Prob to Enter” and “Prob to leave” thresholds into the corresponding text boxes, 0.25 to enter, 0.1 to leave.
4. Select the stepwise regression direction, “Forward”
5. Click “Go” button to let JMP automatically find the set of predictors satisfying the pre-defined significance probability thresholds

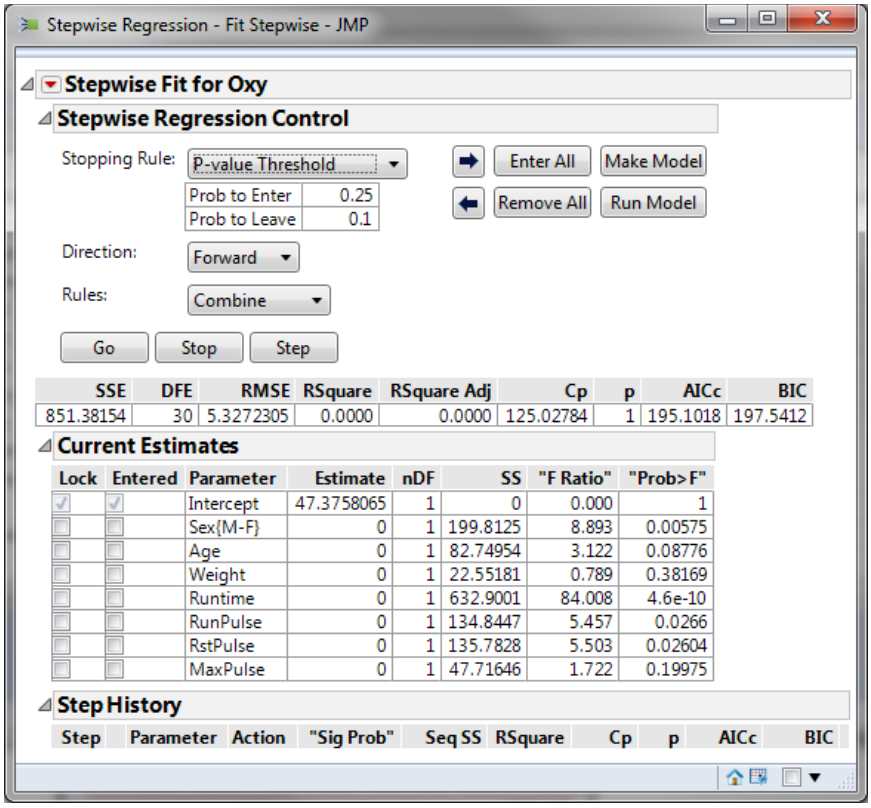


Fig. 4.50 Stepwise Selections

Stepwise Regression Results: Two of seven potential factors are not statistically significant.

Step History: Step-by-step records on how to come up with the final model.

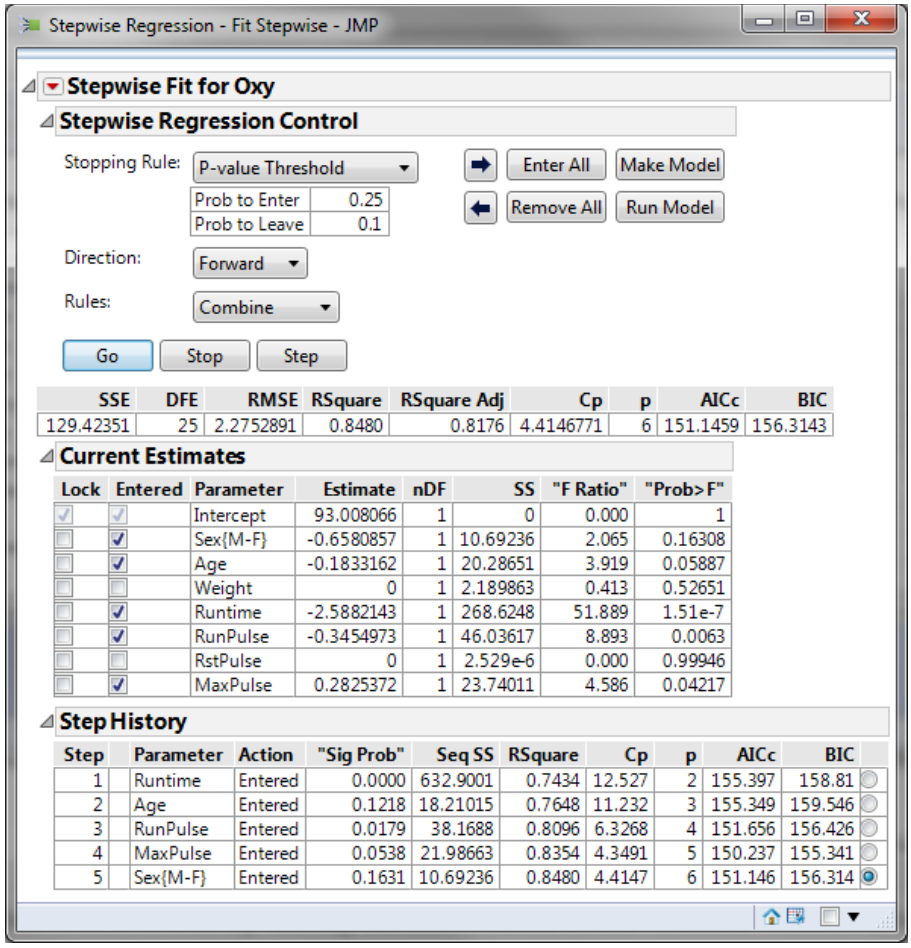


Fig. 4.51 Stepwise Regression Results

4.2.6 LOGISTIC REGRESSION

What is Logistic Regression?

Logistic regression is a statistical method to predict the probability of an event occurring by fitting the data to a logistic curve using logistic function. The regression analysis used for predicting the outcome of a categorical dependent variable, based on one or more predictor variables. The logistic function used to model the probabilities describes the possible outcome of a single trial as a function of explanatory variables. The dependent variable in a logistic regression can be binary (e.g., 1/0, yes/no, pass/fail), nominal (blue/yellow/green), or ordinal (satisfied/neutral/dissatisfied). The independent variables can be either continuous or discrete.

Logistic Function

$$f(z) = \frac{1}{1 + e^{-z}}$$

Where: z can be any value ranging from negative infinity to positive infinity.

The value of *f(z)* ranges from 0 to 1, which matches exactly the nature of probability (i.e., 0 ≤ P ≤ 1).

Logistic Regression Equation

Based on the logistic function,

$$f(z) = \frac{1}{1 + e^{-z}}$$

We define f(z) as the *probability* of an event occurring and *z* is the weighted sum of the significant predictive variables.

$$z = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \cdots + \beta_k \times x_k$$

Where: Z represents the weighted sum of all of the predictive variables.

Logistic Regression

Another of way of representing f(z) is by replacing the z with the sum of the predictive variables.

$$Y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \cdots + \beta_k \times x_k)}}$$

Where: *Y* is the probability of an event occurring and *x's* are the significant predictors.

Notes:

- When building the regression model, we use the actual *Y,* which is discrete (e.g., binary, nominal, ordinal).
- After completing building the model, the fitted *Y* calculated using the logistic regression equation is the probability ranging from 0 to 1. To transfer the probability back to the discrete value, we need SMEs' inputs to select the probability cut point.

Logistic Curve

The logistic curve for binary logistic regression with one continuous predictor is illustrated by the following Figure.

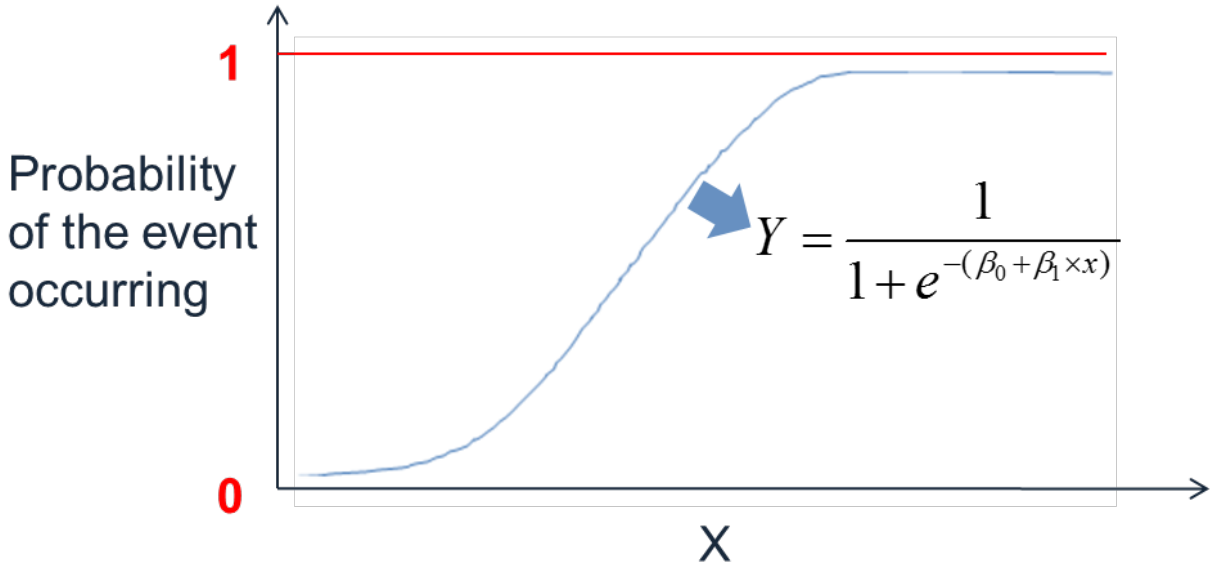


Fig. 4.52 Logistic Curve

Odds

Odds is the probability of an event occurring divided by the probability of the event not occurring.

$$Odds = \frac{P}{1-P}$$

Odds range from 0 to positive infinity.

Probability can be calculated using odds.

$$P = \frac{odds}{1+odds} = \frac{e^{\ln(odds)}}{1+e^{\ln(odds)}} = \frac{1}{1+e^{-\ln(odds)}}$$

Because probability can be expressed by the odds, and we can express probability through the logistic function, we can equate probability, odds, and ultimately the sum of the independent variables.

Since in logistic regression model

$$P = \frac{1}{1+e^{-(\beta_0+\beta_1 \times x_1+\beta_2 \times x_2+\cdots+\beta_k \times x_k)}},$$

therefore

$$\ln(odds) = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \cdots + \beta_k \times x_k$$

Three Types of Logistic Regression

- Binary Logistic Regression
  - Binary response variable
  - Example: yes/no, pass/fail, female/male
- Nominal Logistic Regression
  - Nominal response variable
  - Example: set of colors, set of countries
- Ordinal Logistic Regression
  - Ordinal response variable
  - Example: satisfied/neutral/dissatisfied

All three logistic regression models can use multiple continuous or discrete independent variables and can be developed in JMP using the same steps.

How to Run a Logistic Regression in JMP

We want to build a logistic regression model using the potential factors to predict the probability that the person measured is female or male.

 ***Data File: "LogisticRegression.jmp"***

Response and potential factors

- Response (Y): Female/Male
- Potential Factors (Xs):
  - Age
  - Weight
  - Oxy
  - Runtime
  - RunPulse
  - RstPulse
  - MaxPulse

Step 1:

1. Click Analyze -> Fit Model
2. Select "Sex" as the Y and all the potential factors into the model effects box
3. Click "Run" button

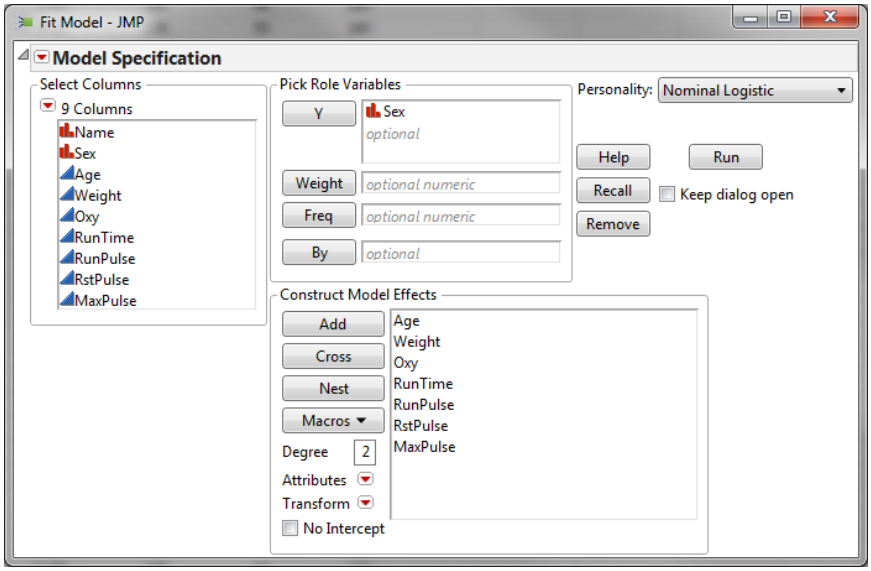


Fig. 4.53 Model Specifications

Step 2:

1. The results of the regression model appear automatically
2. Check the p-value of all the independent variables in the model
3. Remove the insignificant independent variable one at a time from the model and rerun the model
4. Repeat step 2.1 until all the independent variables in the model are statistically significant.

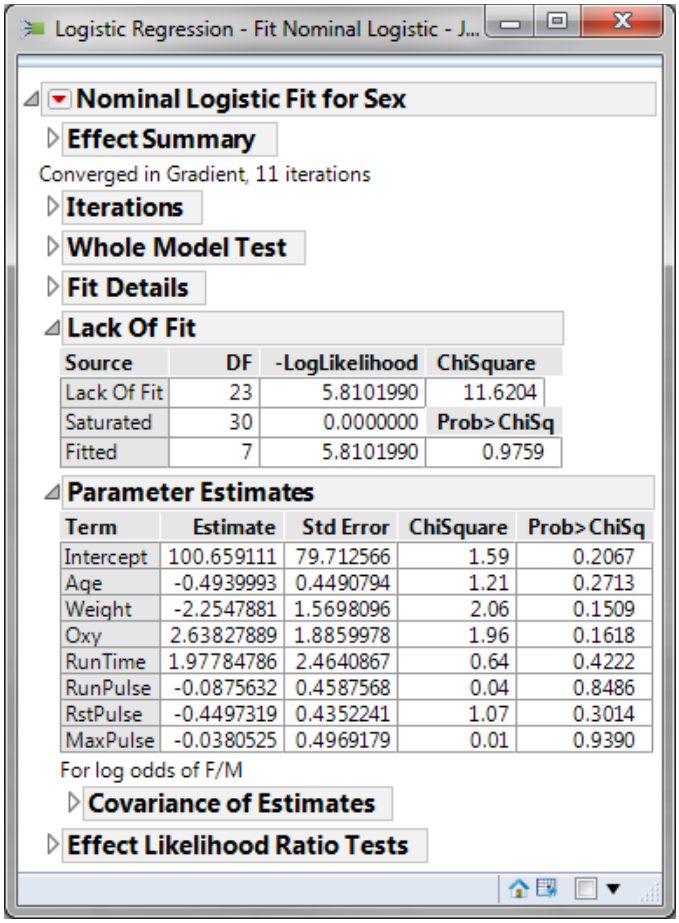


Fig. 4.54 Binary Logistic Regression output

Since the p-values of all the independent variables are higher than the alpha level (0.05), we need to remove the insignificant independent variables one at a time from the model, starting from the one with the highest p-value. MaxPulse has the highest p-value (0.9390), so it would be removed from the model first.

Step 3: Re-Open Your Last Dialog Box

1. Click the red triangle next to "Nominal Logistic Fit for Sex"
2. Select "Model Dialog"
3. The dialog box used to create your last model opens.
4. Remove the variable MaxPulse
5. Follow these same steps with each successive Logistic Regression output screen until you no longer have to remove variables.

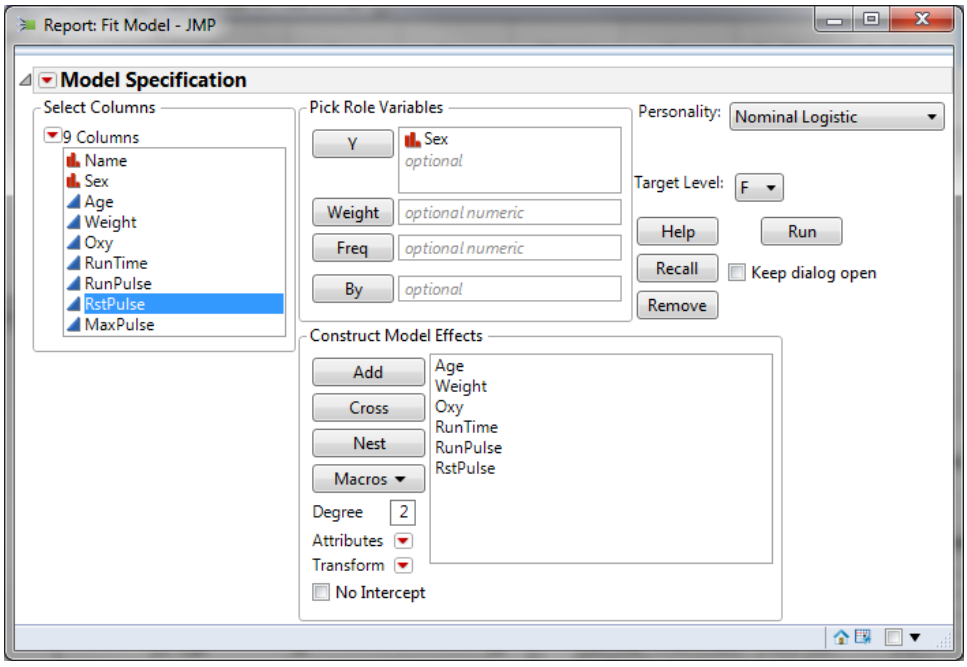


Fig. 4.55 Remove Variable-MaxPulse

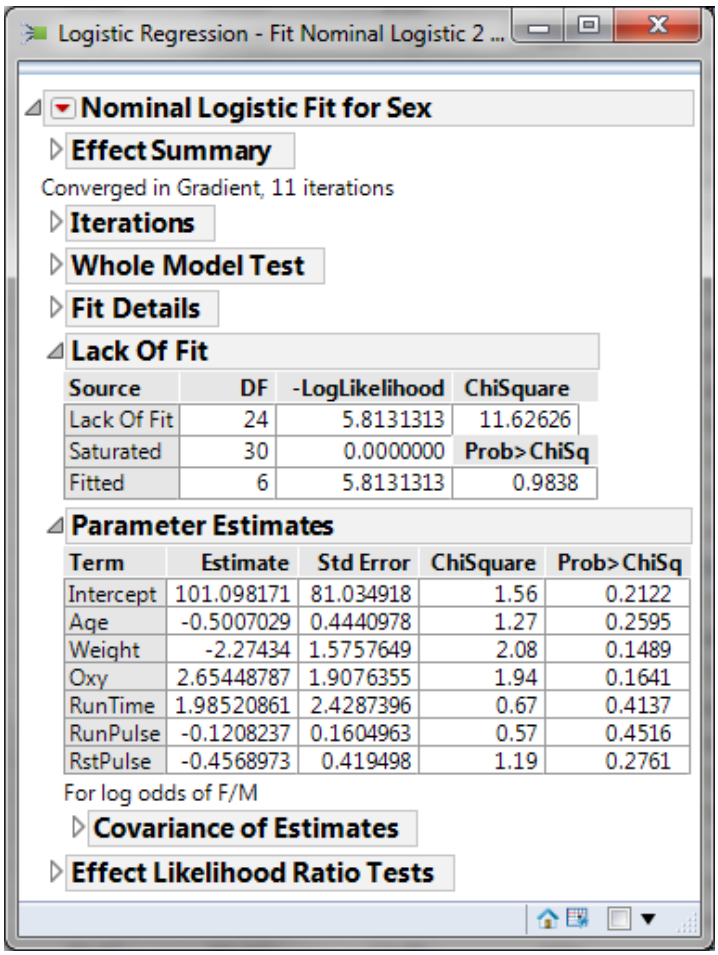


Fig. 4.56 Binary Logistic Regression output without MaxPulse

After removing MaxPulse from the model, the p-values of all the independent variables are still higher than the alpha level (0.05). We need to continue removing the insignificant independent variables one at a time from the model, starting from the one with the highest p-value. RunPulse has the highest p-value (0.4516), so it would be removed from the model next.

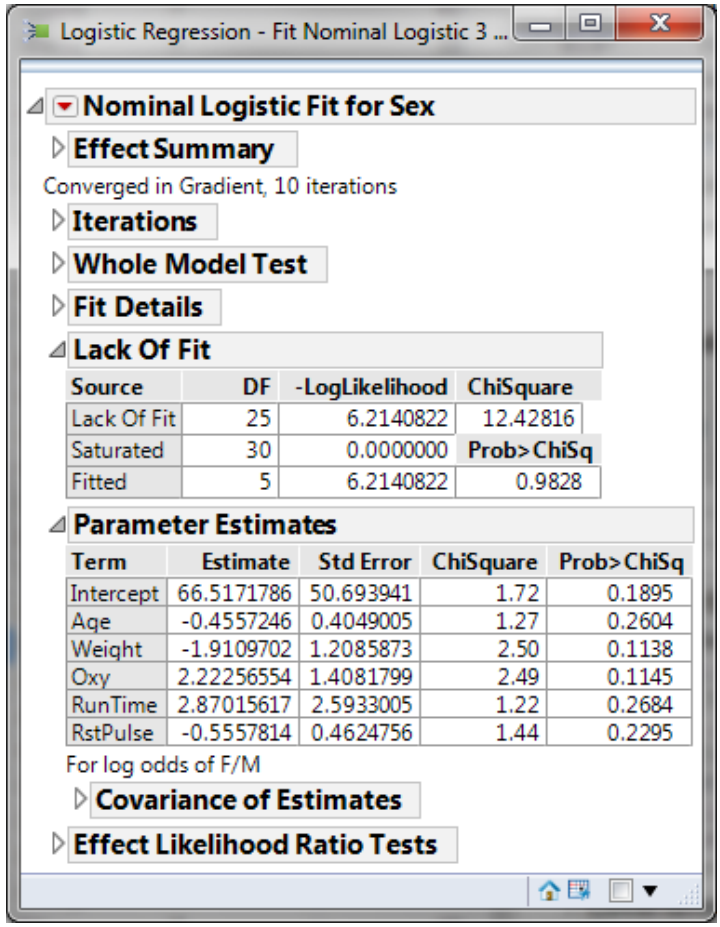


Fig. 4.57 Binary Logistic Regression output without RunPulse

After removing RunPulse from the model, the p-values of all the independent variables are still higher than the alpha level (0.05). We need to continue removing the insignificant independent variables one at a time from the model, starting from the one with the highest p-value. RunTime has the highest p-value (0.2684) so it would be removed from the model next.

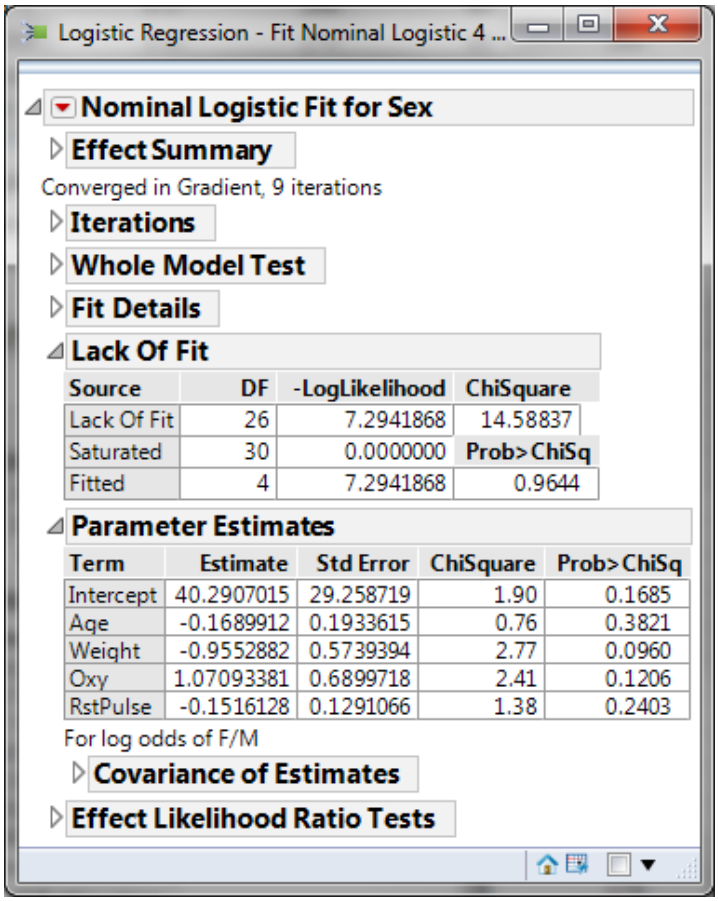


Fig. 4.58 Remove Variable- RunTime

After removing RunTime from the model, the p-values of all the independent variables are still higher than the alpha level (0.05). We need to continue removing the insignificant independent variables one at a time from the model, starting from the one with the highest p-value. Age has the highest p-value (0.3821), so it would be removed from the model next.

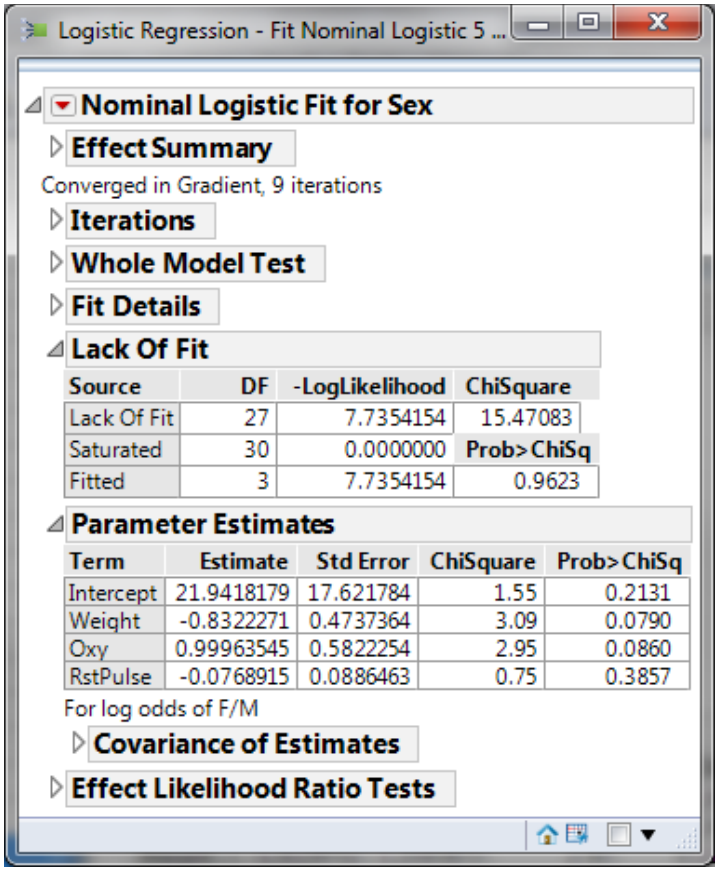


Fig. 4.59 Binary Logistic Regression output without Age

After removing Age from the model, the p-values of all the independent variables are still higher than the alpha level (0.05). We need to continue removing the insignificant independent variables one at a time from the model, starting from the one with the highest p-value. RstPulse has the highest p-value (0.3857), so it would be removed from the model next.

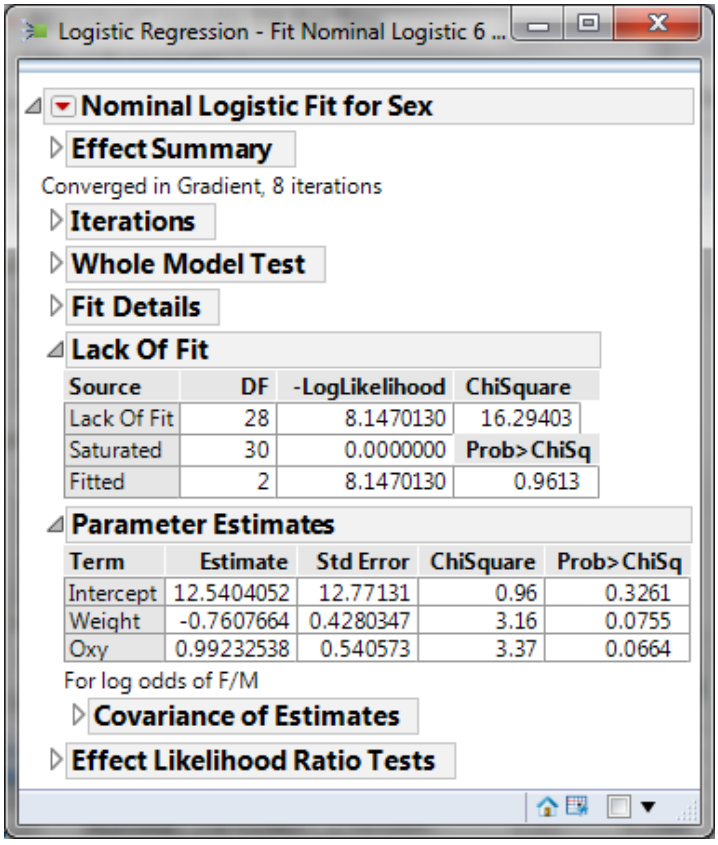


Fig. 4.60 Binary Logistic Regression output without RstPulse

After removing RstPulse from the model, the p-values of all the independent variables are still higher than the alpha level (0.05). We need to continue removing the insignificant independent variables one at a time from the model, starting from the one with the highest p-value. Weight has the highest p-value (0.0755), so it would be removed from the model next.

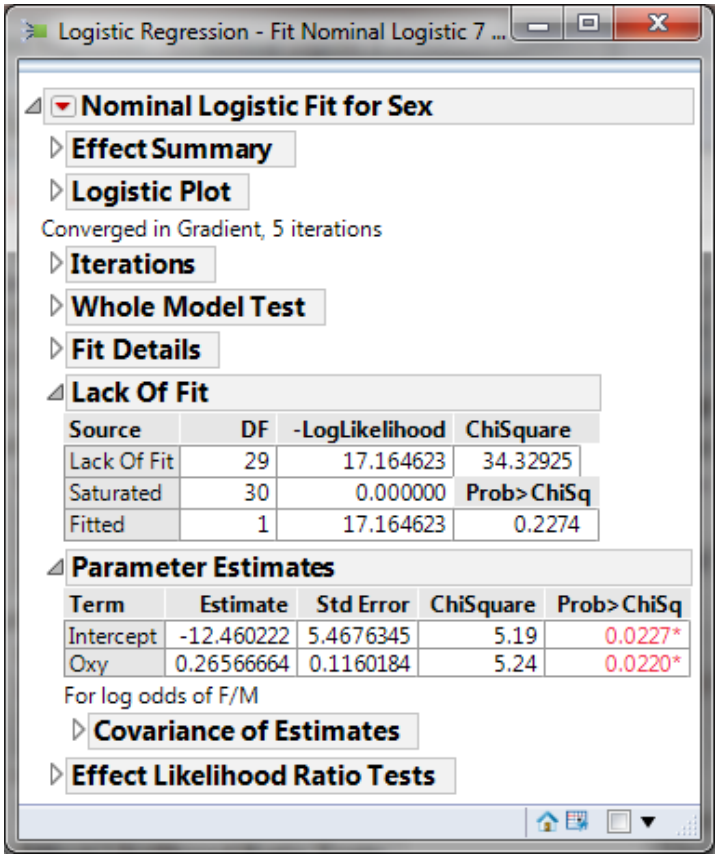


Fig. 4.61 Binary Logistic Regression output without Weight

After removing Weight from the model, the p-value of the only independent variable "Oxy" is lower than the alpha level (0.05). There is no need to remove "Oxy" from the model.

Step 4:

- Analyze the model results and see how the probability change with the change of the independent variable.
- The logistic curve with Y axis indicates the probability of being male and the X axis indicating the value of oxygen uptake.
- When the amount of oxygen uptake increases, the probability of the person measured being male increases accordingly.
- The R-squared is 20.06%. The R-squared of logistic regression is in general lower than the R-squared of the traditional multiple linear regression model.

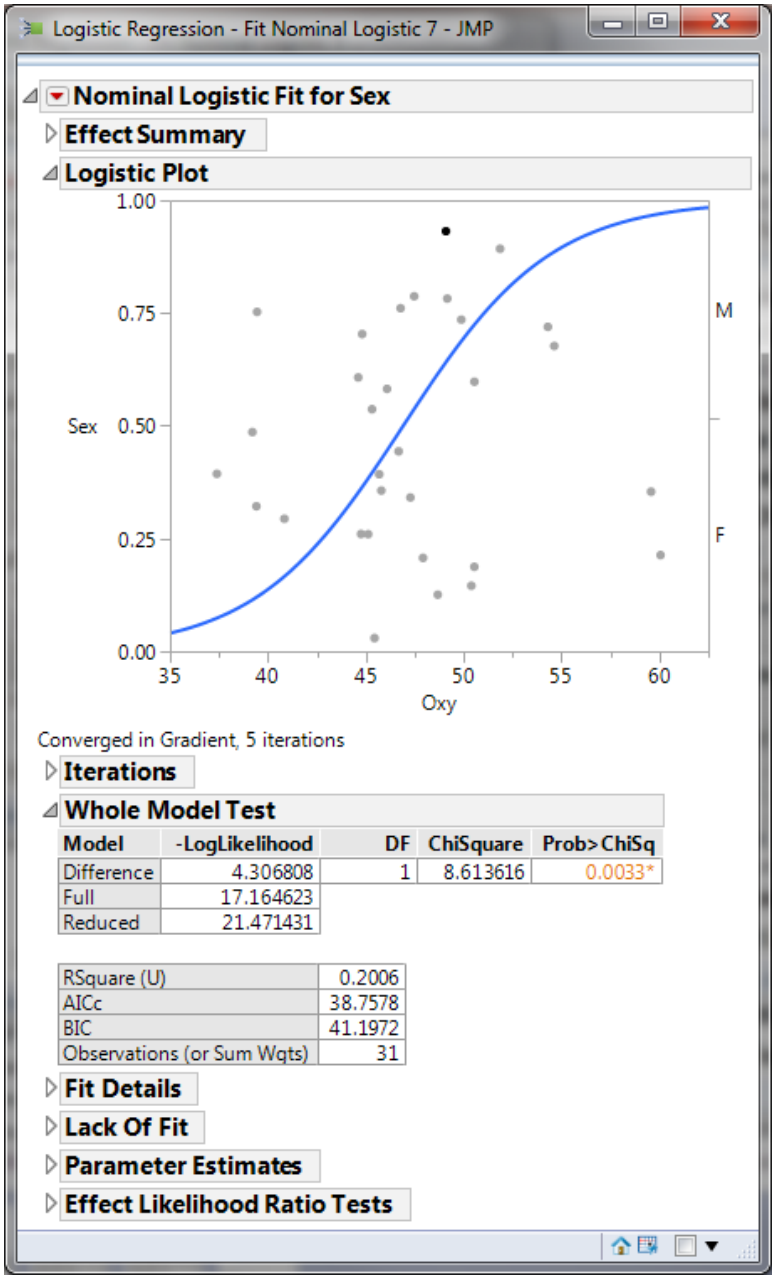
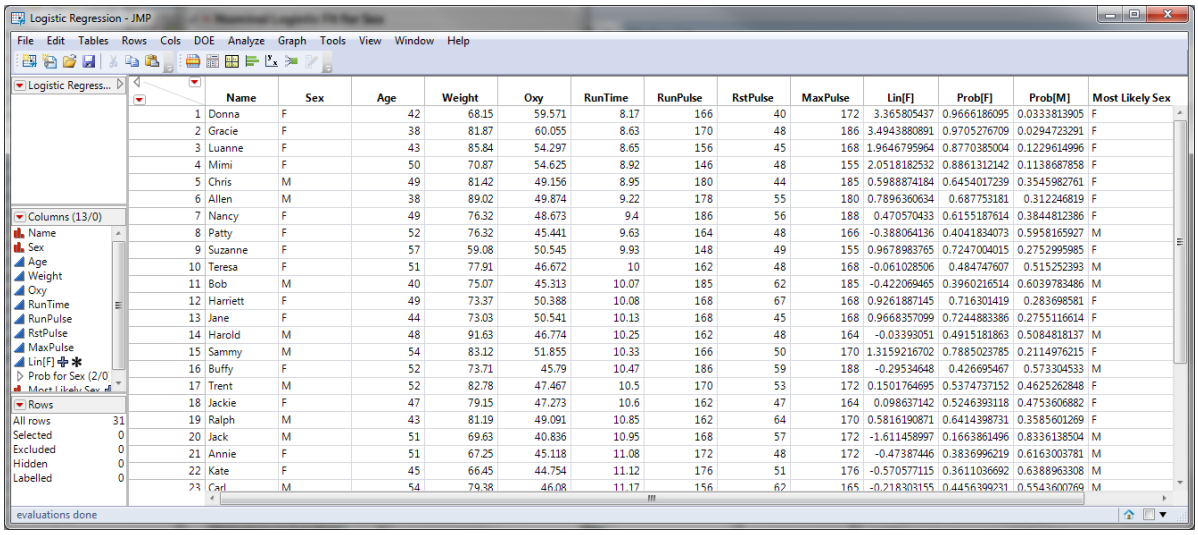


Fig. 4.62 Logistic Regression Analysis in JMP

Step 5:

1. Click on the red triangle button next to “Nominal Logistic Fit for Sex”
2. Click on “Save Probability Formula”
3. The probabilities of the person measured being male and female are added to the right end of the data table.
4. If the probability of a person measured being male is higher than 0.5, it is predicted that the person is male and the prediction is in the last column in the data table.



	Name	Sex	Age	Weight	Oxy	RunTime	RunPulse	RstPulse	MaxPulse	Lin[F]	Prob[F]	Prob[M]	Most Likely Sex
1	Donna	F	42	68.15	59.571	8.17	166	40	172	3.365805437	0.9666186095	0.0333813905	F
2	Gracie	F	38	81.87	60.055	8.63	170	48	186	3.4943880891	0.9705276709	0.0294723291	F
3	Luanne	F	43	85.84	54.297	8.65	156	45	168	1.9646795964	0.8770385004	0.1229614996	F
4	Mimi	F	50	70.87	54.625	8.92	146	48	155	2.0518182532	0.8861312142	0.1138687858	F
5	Chris	M	49	81.42	49.156	8.95	180	44	185	0.5988874184	0.6454017239	0.3545982761	F
6	Allen	M	38	89.02	49.874	9.22	178	55	180	0.7896360634	0.687753181	0.312246819	F
7	Nancy	F	49	76.32	48.673	9.4	186	56	188	0.470570433	0.6155187614	0.3844812386	F
8	Patty	F	52	76.32	45.441	9.63	164	48	166	-0.388064136	0.4041834073	0.5958165927	M
9	Suzanne	F	57	59.08	50.545	9.93	148	49	155	0.9678983765	0.7247004015	0.2752995985	F
10	Teresa	F	51	77.91	46.672	10	162	48	168	-0.061028506	0.484747607	0.515252393	M
11	Bob	M	40	75.07	45.313	10.07	185	62	185	-0.422069465	0.3960216514	0.6039783486	M
12	Harriett	F	49	73.37	50.388	10.08	168	67	168	0.9261887145	0.716301419	0.283698581	F
13	Jane	F	44	73.03	50.541	10.13	168	45	168	0.9668357099	0.7244883386	0.2755116614	F
14	Harold	M	48	91.63	46.774	10.25	162	48	164	-0.03393051	0.4915181863	0.5084818137	M
15	Sammy	M	54	83.12	51.855	10.33	166	50	170	1.3159216702	0.7885023785	0.2114976215	F
16	Buffy	F	52	73.71	45.79	10.47	186	59	188	-0.29534648	0.426695467	0.573304533	M
17	Trent	M	52	82.78	47.467	10.5	170	53	172	0.1501764695	0.5374737152	0.4625262848	F
18	Jackie	F	47	79.15	47.273	10.6	162	47	164	0.098637142	0.5246393118	0.4753606882	F
19	Ralph	M	43	81.19	49.091	10.85	162	64	170	0.5816190871	0.6414398731	0.3585601269	F
20	Jack	M	51	69.63	40.836	10.95	168	57	172	-1.611458997	0.1663861496	0.8336138504	M
21	Annie	F	51	67.25	45.118	11.08	172	48	172	-0.47387446	0.3836996219	0.6163003781	M
22	Kate	F	45	66.45	44.754	11.12	176	51	176	-0.570577115	0.3611036692	0.6388963308	M
23	Carl	M	54	79.38	46.08	11.17	156	62	165	-0.218303155	0.4456399231	0.5543600769	M

Fig. 4.63 Logistic Regression Predictions

Step 6:

1. Click on the red triangle button next to “Nominal Logistic Fit for Sex”
2. Select “Profiler”
3. The Prediction Profiler appears

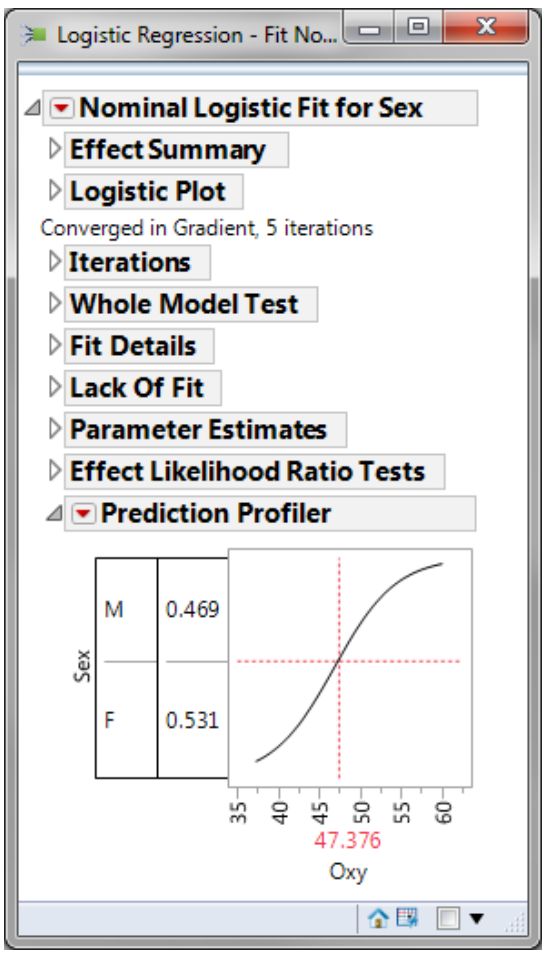


Fig. 4.64 Prediction Profiler output

By changing the value of oxygen uptake, the probability of the person being male or female changes accordingly.

4.3 DESIGNED EXPERIMENTS

Designed experiments are used as information gathering exercises in situations where variation is present. Sometimes the experimenter is controlling the situation, other times not. The goal is understanding how certain input factors affect an output factor.

4.3.1 EXPERIMENT OBJECTIVES

What is an Experiment?

An *experiment* is a scientific exercise to gather data to test a hypothesis, theory, or previous results. Experiments are planned studies in which data are collected actively and purposefully. A typical experiment follows this sequence:

1. A problem arises.
2. A hypothesis is stated.
3. An experiment is designed and implemented to collect data.
4. The analysis is performed to test the hypothesis.
5. Conclusions are drawn based on the analysis results.

An output variable is measured at different levels of input variables, while holding other factors constant. The data are collected with the purpose of understanding how an input variable will impact the output variable.

Why Use Experiments?

- Resolve Problems
    - Eliminate defects or defectives.
    - Shift performance means to meet customer expectations.
    - Squeeze variation, making process more predictable.
- Optimize Performance
    - Use statistical methods to get desired process response.
    - Minimize undesirable conditions, waste, or costs.
    - Reduce variation to eliminate out-of-spec conditions.
- Intelligent Design
    - Design processes around variables that have significant impact on process outputs.
    - Design processes around variables that are easier or more cost effective to manage.

Traditional Methods of Learning

Passive Learning

- Study or observe events while they occur.
- Study or analyze events after they occur.

OFAT Experiment (One Factor at a Time)

- An experimental style limited to controlling one factor independently while others remain either static or are allowed to vary.

Design of Experiments

- Purposefully and proactively manipulate variables so their effect on the dependent variable can be studied.
- Encourages an informative event to occur within specific planned parameters.

What is the Design of Experiment (DOE)?

Design of experiment (DOE) is a systematic and cost-effective approach to plan the exercises necessary to collect data and quantify the cause-and-effect relationship between the response and different factors.

In DOE, we observe the changes occurring in the output (i.e., response, dependent variable) of a process by changing one or multiple inputs (i.e., factors, independent variables). Using DOE we are able to manipulate several factors with every run and determine their effect on the "Y" or output.

Common DOE Terminology

Response: Y, dependent variable, process output measurement

Effect: The change in the average response across two levels of a factor or between experimental conditions.

Factor: X's, inputs, independent variables

Fixed Factor: Factor that can be controlled during the study

Random Factor: Factor that cannot be controlled during the study

Level: factor settings, usually high and low, + and −, 1 and −1

Treatment Combination: setting of all factors to obtain one response measurement (also referred to as a "run")

Replication: Running the same treatment combination more than once (sequence of runs is typically randomized)

Repeat: Non-randomized replicate of all treatment combinations

Inference Space: Operating range of factors under study

Y = f (Xs)

In DOE we are looking to understand the impact of the Xs (continuous or discrete input variables) on the Y (continuous output variable).

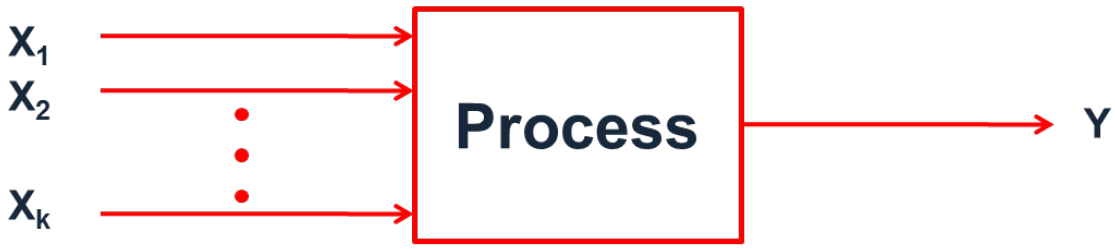


Fig. 4.65 Impact of X's on Y

In Design of Experiments (DOEs);

- $X_1, X_2 \ldots X_k$ are the inputs of the process. They can be either continuous or discrete variables.
- Y is the output of the process. Y is a continuous variable.
- A single X or a group of X's potentially have statistically significant impact on the Y with different levels of influence.

Objectives of Experiments

1. Active Learning: Learn as much as possible with as little resources as possible. Efficiency is the name of the game.
2. Identify Critical X's: Identify the critical factors that drive the output or dependent variable ("Y").
3. Quantify Relationships: Generate an equation that characterizes the relationship between the X's and the Y.
4. Optimize: Determine the required settings of all the X's to achieve the optimal output or response.
5. Validate: Prove results through confirmation or pilot implementation before fully implementing solutions.

Principle of Parsimony

The *principle of parsimony* (also known as Occam's Razor or Ockham's Razor) means that we should not look for complex solutions when a simple one will do. The principle applies to experimentation as well.

The ultimate goal of DOE is to use the minimum amount of data to discover the maximum amount of information. The experiment is designed to obtain data as parsimoniously as

possible to draw as much information about the relationship between the response and factors as possible. DOE is often used when the resources to collect the data and build the model are limited.

Tradeoffs

- The objective of a specific DOE study should be determined by the team.
- The more complex the objective is, the more data are required. As a result, more time and money are needed to collect the data.
- By prioritizing the objectives and filtering factors which obviously do not have any effect on the response, a considerable amount of cost of running the DOE would be reduced.

4.3.2 Experimental Methods

Planning a DOE

1. ***Problem Statement***: Quantifiably define the problem
2. ***Objective***: State the objective of the experiment
3. ***Primary Metric***: Determine the response variable or "Y"
4. ***Key Process Input Variables (KPIV)***: Select the input variables or X's
5. ***Factor Levels***: Determine level settings for all factors
6. ***Design***: Select the experimental design for your DOE
7. ***Project Plan***: Prepare a project plan for your experiment – human and capital resource allocation, timing, duration etc.
8. ***Gain Buy-in***: Your experiment will need support and buy-in from several places, get it!
9. Run the experiment
10. Analyze, Interpret, and Share Results

DOE Problem Statement

Much like the overall problem statement discussed in the Define phase, a DOE problem statement should specific and have quantifiable elements. Make sure your problem statement ties to a business goal or business performance indicator. Problem statements should *not* include conclusions, solutions, and causes—they are all possible outputs of a DOE.

DOE Objectives

State the objective of the experiment

The following are DOE objective examples:

- "To determine the effects of tire pressure on gas mileage"
- "To determine which factors have the most impact on gas mileage"
- "To determine the effects of a pick-up truck's tailgate (up vs. down) on gas mileage"
- "To determine which HVAC and Lighting factors are contributing most to Kwh consumption"

DOE objective statements should:

- Clearly define what you want to learn from the experiment
- Declare study parameters or scope.

DOE Primary Metric

Recall our discussion of primary metrics in the Define phase? We said something to this effect: "the primary metric is the most important measure of success." This statement still holds true when discussing the primary metric for a DOE.

The primary metric for a DOE is your experiment's output or response under study. It is the "Y" in your transfer function *Y = f(x).* It is critical to understand the following things about your metric:

- Measurement accuracy, R&R, and/or bias
- Measurement frequency
- Measurement resolution

DOE Primary Metric

Before embarking on a design of experiment you should have validated your measurement system. If you have not, now would be a wise time to back track and do so. DOEs can be costly on a firm's time and resources. DOEs not planned and run properly will create waste and defective products because treatment combinations should "test" boundaries.

Do not allow your entire study to be wasted because of a poor measurement system or one that cannot respond frequently enough for your test. Measurement resolution is another key factor to consider for your DOE. Make sure your primary metric can decipher the size of the main effects and interactions you are looking for.

There is still even more to understand about your primary metric before jumping into a DOE. Be sure to answer these questions before proceeding:

- Is it discrete or continuous?
- Do you want to shift its mean?
- Do you want to squeeze variation?
- What is the baseline?
- Is it in control?
- Is there seasonality, trending, or any cycle?
- How much change do you want to detect?
- Is your metric normally distributed?

DOE Input Variables—KPIVs (key process input variables)

Factor selection is an important element in planning for a design of experiment. Factors are the "things" or metrics that your study responds to and can be discrete or continuous. Factors should have largely been determined through the use of tools and analytics used throughout the DMAIC roadmap:

- Failure modes and effects analysis
- Cause and effect diagram (fishbone)
- X-Y matrix
- Process mapping
- Planned studies
- Passive analytics etc.

DOE Factor Settings

Factor settings are the range of values selected for each factor and need to be predetermined before every experiment. Factor settings are the values at which the factors will be set during the experiment. The settings are typically a high, low, and sometimes a medium setting. The range of settings should be wide enough to represent a typical operating range that will achieve a measurable response, but narrow enough to be practical. Let us use this paper airplane as an example. Our factors under study for max distance as a response might be:

- Number of folds
- Paper weight
- Paperclip (yes or no)
- Thrust
- Launch height
- Launch angle etc.

Factor settings are important because each setting selection should provide value to the study. Understanding the "margin" where that value can be found is where an experienced Black Belt will earn their keep. For example, what would we learn if we chose the test two levels for "number of folds" and we tested zero folds and 80 folds as our two settings for that factor? Right, nothing!

DOE—Design Considerations

Design selection is an aspect of properly planned design of experiments. Next is a high-level overview of design types (you will learn more about these in later modules).

- Screening Designs
    - Fractional factorials
    - Intended to narrow down the many factors to the vital few
    - Screening designs enable many factors to be evaluated at a lower cost because of the nature of the design (main effects)

- Characterization Designs
    - Higher resolution fractional factorials
    - Full factorials with fewer variables
    - Main effect and interactions
- Optimization Designs
    - Full factorials
    - Response surface designs

DOE Project Plan

DOEs can require a fair amount of coordination between human and capital resources. Because of the amount of work and coordination necessary to conduct a DOE, a project plan can be useful. Make sure you identify critical dependencies like interfering with production or operating hours etc. Use project management software or, at a minimum, a Gantt chart to help you identify *critical path items*.

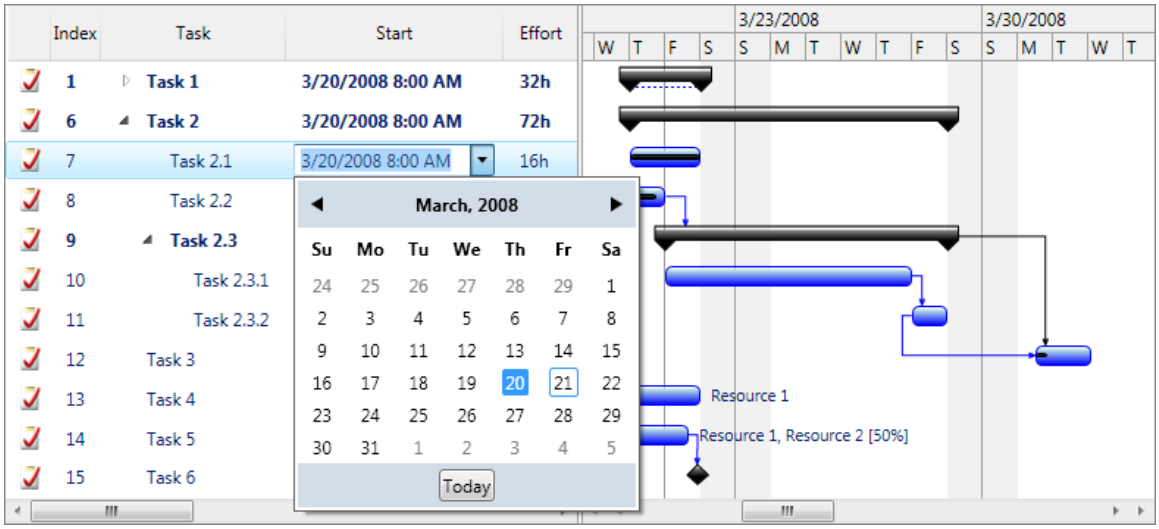


Fig. 4.66 Critical Path Items in Project Plan

DOE—Getting Support

By running a DOE you are about to "discover" something new and better or validate an existing belief or assumption. *(If you do not believe this you should not be running your DOE.)*

You are also about to cost your firm money or disrupt someone's routine . . . Get everyone on the same page before moving forward! Hopefully, you have already established the framework for the following items. If not, follow these steps:

- Business Case and Justification
- Stakeholder Analysis—Figure out who is with you and your project and who is not
- Communication Plan—Build and execute your communication plan based on the first two items.

Next, run your experiment and analyze, interpret, and share your DOE results.

DOE Iterations

One single DOE might not be enough to completely answer the questions due to the following reasons:

1. None of the potential factors identified are statistically significant to the response.
2. More factors need to be included in the model.
3. Residuals of the model are not performing well.
4. The objectives of DOE changes due to business reasons.

4.3.3 Experiment Design Considerations

Considerations of Experiments

There are many factors to consider when deciding what type of experimental approach to take:

- Objectives of the experiments
- Resource availability to run the experiments
- Potential costs to implement the experiments
- Stakeholders' support for successful experiments
- Accuracy and precision of the measurement system
- Statistical stability of the process
- Unexpected plans or changes
- Sequence of running the experiments
- Simplicity of the model
- Inclusion of potential significant factors
- Settings of factors
- Well-behaved residuals

4.4 Full Factorial Experiments

4.4.1 2k Full Factorial Designs

With a 2k Full Factorial Design:

- Each factor has two levels (often labeled + and −)
- It is possible to get a useful design for preliminary analysis and weed out the unimportant factors
- A 2k Full Factorial Design also allows initial study of interactions.

DOE Key Terms

- ***Response***: Y, dependent variable, process output measurement
- ***Effect***: The change in the average response across two levels of a factor or between experimental conditions
- ***Factor***: X's, inputs, independent variables
- ***Fixed Factor***: Factor that can be controlled during the study
- ***Random Factor***: Factor that cannot be controlled during the study
- ***Factor Levels*:** Factor settings, usually high and low, + and −, 1 and −1
- ***Treatment Combination***: setting of all factors to obtain one response measurement (also referred to as a "run")
- ***Replication***: Running the same treatment combination more than once (sequence of runs is typically randomized)
- ***Repeat***: Non-randomized replicate of all treatment combinations
- ***Inference Space***: Operating range of factors under study
- ***Main Effect***: The average change from one level setting to another for a single factor
- ***Interaction***: The combined effect of two factors independent of the main effect of each factor

More About Factors

Factors in DOE are the potential causes that might have significant impact on the outcome of a process or a product. Factors are the independent variables in an experiment, the controllable inputs that are changed in order to understand the impact on the response variable. Only include the factors that might have a significant impact on the response variable: they can be continuous or discrete variables.

Most experiments include at least two or more factors to maximize learning from the experiment; however, they must be balanced with cost. Ask subject matter experts' opinions for main factors selection. The goal of a DOE is to measure the effects of factors and interactions of factors on the outcome.

In DOE:

1. First, we create a set of different combinations of factors each at different levels.
2. Second, we run the experiment and collect the response values in different scenarios.
3. Third, we analyze the results and test how the response changes accordingly when factors change.

Factor Levels

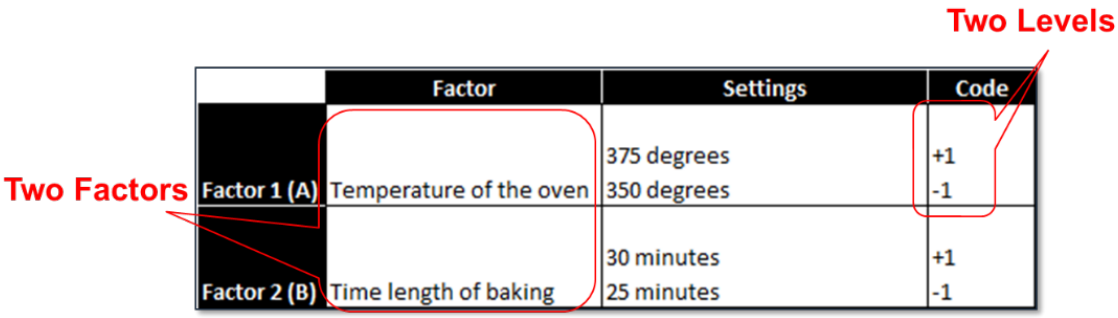
Factor levels are the selected settings of a factor we are testing in the experiment. The more levels a factor has, the more scenarios we have to create and test. As a result, more resources are required to collect samples and the model is more complicated.

The most popular DOE is two-level design, meaning only two levels for each factor.

- Code of factor levels

- High vs. Low
- (+1) vs. (−1)
- (+) vs. (-)

- Example of factor levels: a study on how two factors affect the taste of cakes



	Factor	Settings	Code
Factor 1 (A)	Temperature of the oven	375 degrees	+1
		350 degrees	-1
Factor 2 (B)	Time length of baking	30 minutes	+1
		25 minutes	-1

Fig. 4.67 Example of Factor levels

Treatment Combination

A *treatment* is a combination of different factors at different levels. It is a unique scenario setting in an experiment.

Coding treatments:

- If there are two factors in an experiment, we name one factor A and the other factor B.
- A single treatment in an experiment is named for each factor:
  - −−, +−, −+ and ++
  - l, a, b and ab

Example of treatment: a study on how two factors affect the taste of cakes.

Treatment	Factors	
	Temperature of the oven	Time length of baking
l	350	25
a	375	25
b	350	30
ab	375	30

Treatment	Factors	
	Factor A	Factor B
l	-1	-1
a	+1	-1
b	-1	+1
ab	+1	+1

Four treatments

Fig. 4.68 Example of Treatment

Response

A *response* is the output of a process, a system, or a product. It is the outcome we are interested to improve by optimizing the settings of factors. In an experiment, we observe, record, and

analyze the corresponding changes occurring in the response after changing the settings of different factors.

Example of treatment: a study on how two factors affect the taste of cakes. We use a predefined metric to measure the cake's tastiness. After running each treatment, we obtain and record the resulting response value.

Treatment	Factors		Response
	Temperature of the oven	Time length of baking	
I	350	25	
a	375	25	18
b	350	30	22
ab	375	30	12

Treatment	Factors		Response
	Factor A	Factor B	
I	-1	-1	20
a	+1	-1	18
b	-1	+1	22
ab	+1	+1	12

One response

Fig. 4.69 Example of Response

Main Effect

Main effect is the average change in the response resulting from the change in the levels of a factor. It captures the effect of a factor alone on the response without taking into consideration other factors.

In the baking cake example, the main effects of temperature and time length are

$$MainEffect_A = \frac{18+12}{2} - \frac{20+22}{2} = -6$$

$$MainEffect_B = \frac{22+12}{2} - \frac{20+18}{2} = -2$$

Interpretation of main effect in the example:

By changing the temperature level of the oven from high to low, the response (i.e., tastiness of the cake) increases by 6 measurement units. By changing the time length of baking from high to low, the response (i.e., tastiness of the cake) increases by 2 measurement units.

Interaction Effect

Interaction effect is the average change in the response resulting from the change in the interaction of multiple factors. It occurs when the change in the response triggered by one factor depends on the change in other factors.

The *interaction effect* is the combined or simultaneous effect of two or more factors acting together on a response variable. Consider two factors related to changing a response variable—the flavor of a cup of coffee. The factors are stirring the coffee and addition of sugar

to the coffee. Adding sugar alone, or stirring alone, does not change the flavor of the coffee; however, the combination of the two has a significant impact on the taste of the coffee.

Example of interaction effect

Treatment	Factors		Interaction (A*B)	Response
	Factor A	Factor B		
I	-1	-1	+1	20
a	+1	-1	-1	18
b	-1	+1	-1	22
ab	+1	+1	+1	12

One interaction

Fig. 4.70 Example of Interaction

$$InteractionEffect = \frac{20+12}{2} - \frac{22+18}{2} = -4$$

By changing the interaction from high to low, the response (i.e., tastiness of the cake) increases by 4 measurement units.

Full Factorial DOE

In a *full factorial experiment*, all of the possible combinations of factors and levels are created and tested. For example, for two-level design (i.e., each factor has two levels) with k factors, there are 2^k possible scenarios or treatments.

1. Two factors, each with two levels, we have $2^2 = 4$ treatments
2. Three factors, each with two levels, we have $2^3 = 8$ treatments
3. k factors, each with two levels, we have 2^k treatments

2^k Full Factorial DOE

Full factorial DOE is used to discover the cause-and-effect relationship between the response and both individual factors and the interaction of factors. Generate an equation to describe the relationship between Y and the important Xs:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_p X_1 X_2 \cdots X_k + \varepsilon$$

Where:

- Y is the response and $X_1, X_2 \ldots X_k$ are the factors
- α_0 is the intercept and $\alpha_1, \alpha_2 \ldots \alpha_p$ are the coefficients of the factors and interactions
- ε is the error of the model.

Two-Level Two-Factor Full Factorial
Below is a design pattern of a two-level *two-factor* full factorial experiment.

Run	Treatment	Factors	
		A	B
1	I	-1	-1
2	a	+1	-1
3	b	-1	+1
4	ab	+1	+1

Fig. 4.71 Two-Level Two-Factor Full Factorial

2 (level) raised to 2 (factors) = 4 treatment combinations.

Two-Level Three-Factor Full Factorial
Below is a design pattern of a two-level *three-factor* full factorial experiment.

Run	Treatment	Factors		
		A	B	C
1	I	-1	-1	-1
2	a	+1	-1	-1
3	b	-1	+1	-1
4	ab	+1	+1	-1
5	c	-1	-1	+1
6	ac	+1	-1	+1
7	bc	-1	+1	+1
8	abc	+1	+1	+1

Fig. 4.72 Two-Level Three-Factor Full Factorial

2 (levels) raised to 3 (factors) = 8 treatment combinations.

Two-Level Four-Factor Full Factorial
Below is a design pattern of a two-level *four-factor* full factorial experiment

Run	Treatment	Factors			
		A	B	C	D
1	I	-1	-1	-1	-1
2	a	+1	-1	-1	-1
3	b	-1	+1	-1	-1
4	ab	+1	+1	-1	-1
5	c	-1	-1	+1	-1
6	ac	+1	-1	+1	-1
7	bc	-1	+1	+1	-1
8	abc	+1	+1	+1	-1
9	d	-1	-1	-1	+1
10	ad	+1	-1	-1	+1
11	bd	-1	+1	-1	+1
12	abd	+1	+1	-1	+1
13	cd	-1	-1	+1	+1
14	acd	+1	-1	+1	+1
15	bcd	-1	+1	+1	+1
16	abcd	+1	+1	+1	+1

Fig. 4.73 Two-Level Four-Factor Full Factorial

2 (levels) raised to 4 (factors) = 16 treatment combinations

Two-Level Five-Factor Full Factorial

Below is a design pattern of a two-level *five-factor* full factorial experiment

Run	Treatment	Factors				
		A	B	C	D	E
1	I	-1	-1	-1	-1	-1
2	a	+1	-1	-1	-1	-1
3	b	-1	+1	-1	-1	-1
4	ab	+1	+1	-1	-1	-1
5	c	-1	-1	+1	-1	-1
6	ac	+1	-1	+1	-1	-1
7	bc	-1	+1	+1	-1	-1
8	abc	+1	+1	+1	-1	-1
9	d	-1	-1	-1	+1	-1
10	ad	+1	-1	-1	+1	-1
11	bd	-1	+1	-1	+1	-1
12	abd	+1	+1	-1	+1	-1
13	cd	-1	-1	+1	+1	-1
14	acd	+1	-1	+1	+1	-1
15	bcd	-1	+1	+1	+1	-1
16	abcd	+1	+1	+1	+1	-1
17	e	-1	-1	-1	-1	+1
18	ae	+1	-1	-1	-1	+1
19	be	-1	+1	-1	-1	+1
20	abe	+1	+1	-1	-1	+1
21	ce	-1	-1	+1	-1	+1
22	ace	+1	-1	+1	-1	+1
23	bce	-1	+1	+1	-1	+1
24	abce	+1	+1	+1	-1	+1
25	de	-1	-1	-1	+1	+1
26	ade	+1	-1	-1	+1	+1
27	bde	-1	+1	-1	+1	+1
28	abde	+1	+1	-1	+1	+1
29	cde	-1	-1	+1	+1	+1
30	acde	+1	-1	+1	+1	+1
31	bcde	-1	+1	+1	+1	+1
32	abcde	+1	+1	+1	+1	+1

Fig. 4.74 Two-Level Five-Factor Full Factorial

2 (levels) raised to 5 (factors) = 32 treatment combinations

Order to Run Experiments

The four design patterns shown earlier are listed in the standard order. *Standard order* is used to design the combinations/treatments before experiments start. When actually running the experiments, *randomizing* the standard order is recommended to minimize the noise.

Replication in Experiments

Each treatment can be tested multiple times in an experiment to increase the degrees of freedom and improve the capability of analysis. We call this method *replication*.

Replicates are the number of repetitions of running an individual treatment, which increase the power of the experimental responses. The order to run the treatments in an experiment should be randomized to minimize the noise.

Advantages of replication include: helps to better identify the true sources of variation, helps estimate the true impacts of the factors on the response, and overall improves the reliability and validity of the experimental results.

2^2 Full Factorial DOE

Case study: We are running a 2^2 full factorial DOE to discover the cause-and-effect relationship between the cake tastiness and two factors: temperature of the oven and time length of baking. Each factor has two levels and there are four treatments in total.

We decide to run each treatment twice so that we have enough degrees of freedom to measure the impact of two factors and the interaction between two factors. Therefore, there are eight observations in response eventually.

	Factor	Settings	Code
Factor 1 (A)	Temperature of the oven	375 degrees 350 degrees	+1 -1
Factor 2 (B)	Time length of baking	30 minutes 25 minutes	+1 -1

Fig. 4.75 Full Factorial DOE Observations

The objective is to understand the main effects and the interactions of these factors on the response variable. After running the four treatments twice in a random order, we obtain the following results

Treatment	Factors		Interaction (A*B)	Response
	Factor A	Factor B		
I	-1	-1	+1	20
a	+1	-1	-1	18
b	-1	+1	-1	22
ab	+1	+1	+1	12
I	-1	-1	+1	21
a	+1	-1	-1	17
b	-1	+1	-1	22
ab	+1	+1	+1	13

Fig. 4.76 Full Factorial DOE Results

There are two factors and two levels, so there would be 2^2 = 4 treatment combinations. With replicates, each treatment combination is repeated once; therefore, there are in total 8 runs in this experiment. The experiment results are consolidated into the following table

Treatment	Factors		Interaction (A*B)	Response		
	Factor A	Factor B		Run 1	Run 2	Total
I	-1	-1	+1	20	21	41
a	+1	-1	-1	18	17	35
b	-1	+1	-1	22	22	44
ab	+1	+1	+1	12	13	25

Fig. 4.77 Experiment Results

The main effect of factor A is computed by averaging the difference between combinations where A was at its high settings and where A was at its low settings.

Main effect of factor A (temperature of the oven):

$$MainEffect_A = \frac{(a+ab)-(b+l)}{2^{k-1}r}$$

$$= \frac{(35+25)-(44+41)}{2^{2-1}\times 2}$$

$$= -6.25$$

Where:

- k is the number of factors
- r is the number of times individual treatments are being run.

Using the formula provided, the main effect of increasing the temperature of the oven is to decrease tastiness of the cake by −6.25.

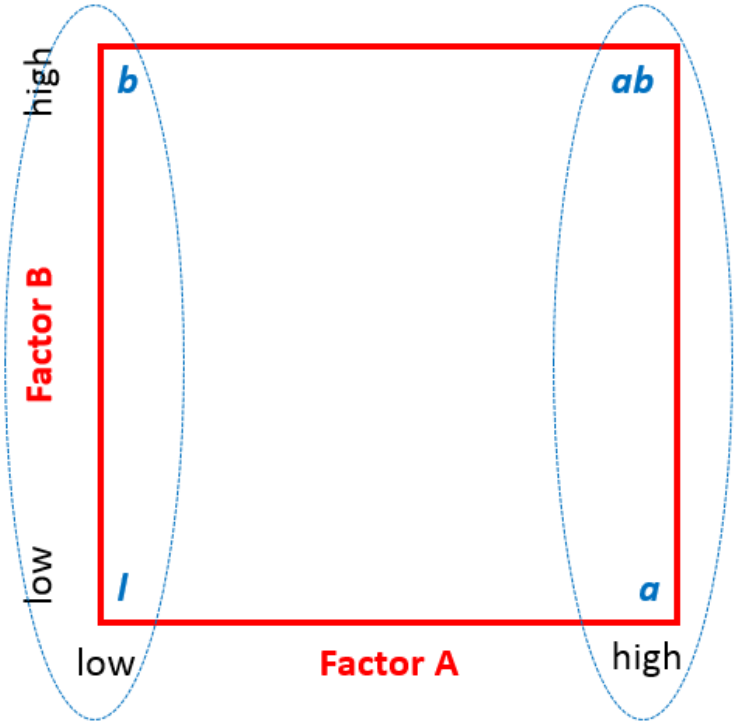


Fig. 4.78 Main Effect of Factor A

The main effect of factor B, similar to A, is computed by averaging the difference between combinations where B was at its high settings and where B was at its low settings.

Main effect of factor B (time length of baking):

$$MainEffect_B = \frac{(b+ab)-(a+l)}{2^{k-1}r}$$

$$= \frac{(44+25)-(35+41)}{2^{2-1}\times 2}$$

$$= -1.75$$

Where:

- k is the number of factors
- r is the number of times individual treatments are being run.

Using the formula provided, the main effect of increasing the baking time is to decrease the tastiness of the cake by −1.75

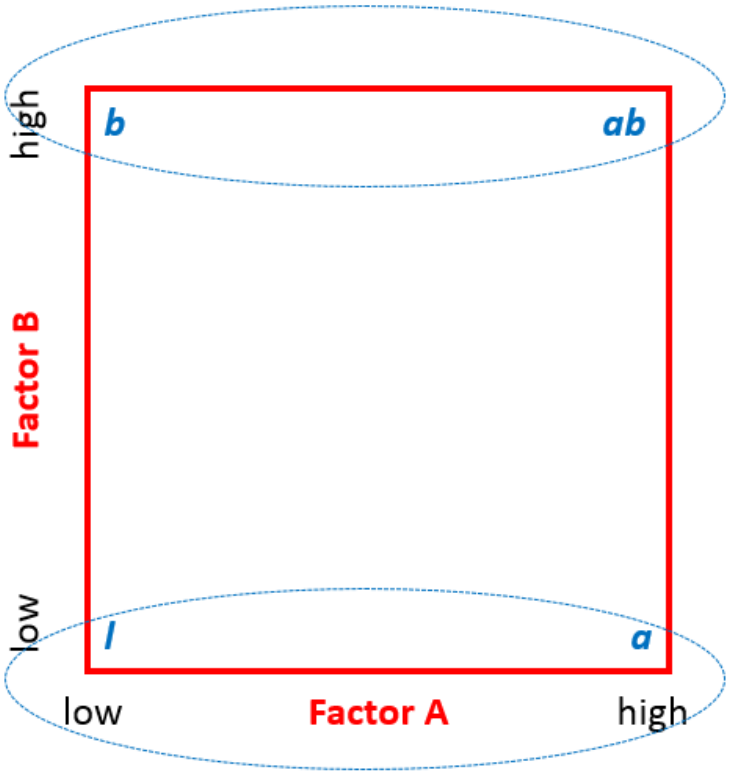


Fig. 4.79 Main Effect of Factor B

The interaction effect is computed by averaging the difference between combinations where A and B were at opposite settings (low and high).

Interaction (i.e., A*B) effect:

$$InteractionEffect = \frac{(l + ab) - (a + b)}{2^{k-1} r}$$

$$= \frac{(41 + 25) - (35 + 44)}{2^{2-1} \times 2}$$

$$= -3.25$$

Where:

- k is the number of factors
- r is the number of times individual treatments are being run.

Using the formula provided, the interaction effect of the temperature and time variables on tastiness was −3.25.

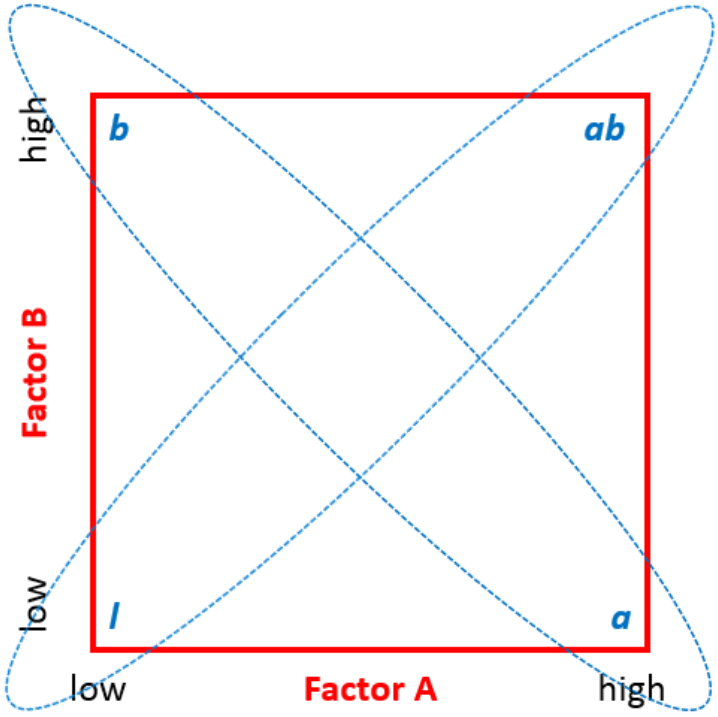


Fig. 4.80 Interaction Effect of A and B

Sum of squares of factors and interaction

$$SS_A = \frac{(a + ab - b - l)^2}{2^k r} = \frac{(35 + 25 - 44 - 41)^2}{2^2 \times 2} = 78.125$$

$$SS_B = \frac{(b + ab - a - l)^2}{2^k r} = \frac{(44 + 25 - 35 - 41)^2}{2^2 \times 2} = 6.125$$

$$SS_{Interaction} = \frac{(l + ab - a - b)^2}{2^k r} = \frac{(41 + 25 - 35 - 44)^2}{2^2 \times 2} = 21.125$$

Where:

- k is the number of factors
- r is the number of times individual treatments are being run.

The sum of squares tells us the relative strength of each main effect and interaction. A has the strongest effect as indicated by the high SS value. The *degrees of freedom* are necessary to determine the mean squares value.

Degrees of freedom of factors and interaction:

$$df_A = 1$$

$$df_B = 1$$

$$df_{Interaction} = 1$$

$$df_{error} = 1$$

Four degrees of freedom are necessary because there are three effects we are looking to understand: factor A, factor B, and the interaction between them.

$$Y = \alpha_0 + \alpha_1 * A + \alpha_2 * B + \alpha_3 * A * B + Error$$

Mean squares of factors and interaction:

$$MS_A = \frac{SS_A}{df_A} = \frac{78.125}{1} = 78.125$$

$$MS_B = \frac{SS_B}{df_B} = \frac{6.125}{1} = 6.125$$

$$MS_{Interaction} = \frac{SS_{Interaction}}{df_{Interaction}} = \frac{21.125}{1} = 21.125$$

Use JMP to Run a 2^k Full Factorial DOE

Step 1: Initiate the experiment design

1. Click DOE -> Custom Design
2. A window named "DOE – Custom Design" pops up
3. Double-click the name box and enter "Tastiness" into the "Response Name" box.
4. Because the tastiness metric indicates how delicious the cake is, we want to maximize it.
5. Click on the button "Add Factor"
6. Click Categorical-> 2 level
7. Enter the factor name "Temp" in the name box
8. Enter the two levels' codes "Low" and "High" in the values boxes
9. Repeat the same steps for the second factor "Time" with two levels "Short" and "Long"
10. Click Continue

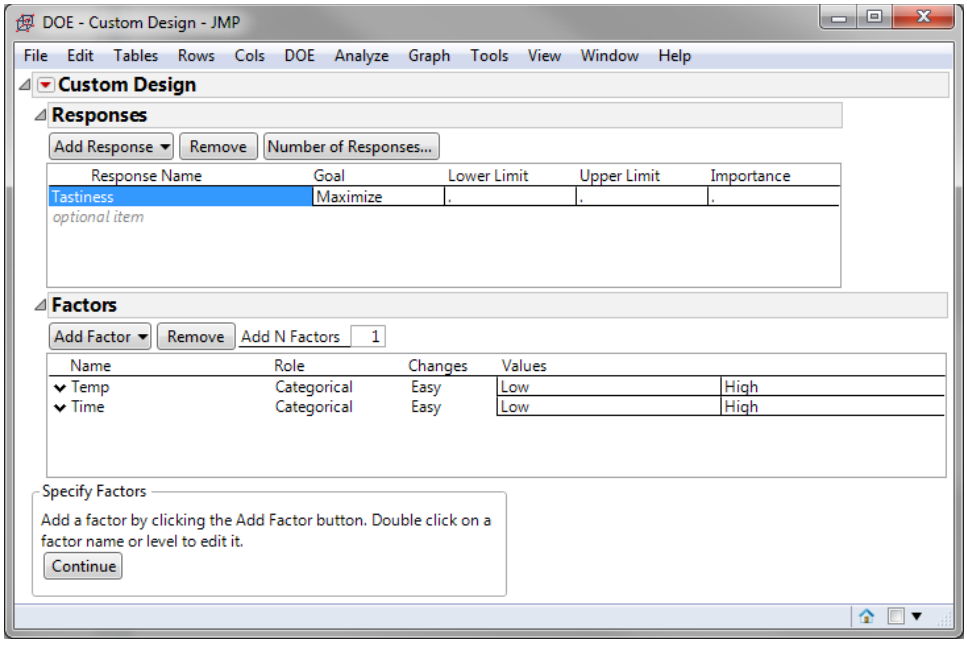


Fig. 4.81 Second Factors with Variables

Step 2: Select interaction between the two factors.

1. Click on the drop down box of "Interactions"
2. Select "2nd"
3. The interaction variable "Temp* Time" appears in the list of factors.

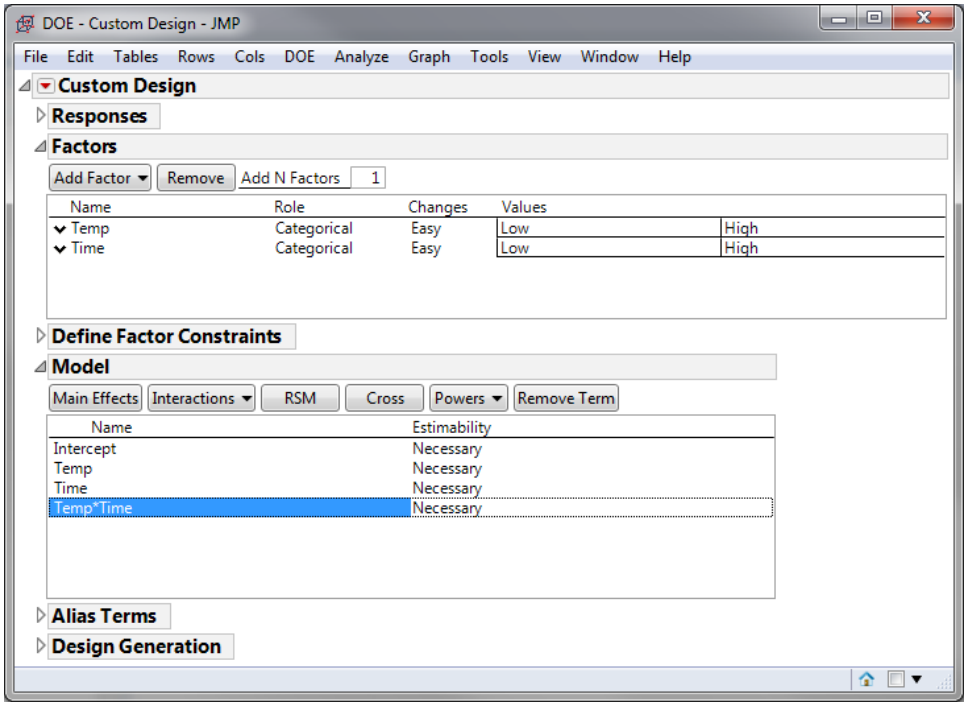


Fig. 4.82 Interaction Variables

Step 3: Make the design and make the table

1. Click on the "Make Design" button

2. The design pattern of the experiment appears
3. Keep “Randomize” in the “Run Order” box
4. Click on “Make Table” button
5. A table with pre-populated treatments in random run order is generated.

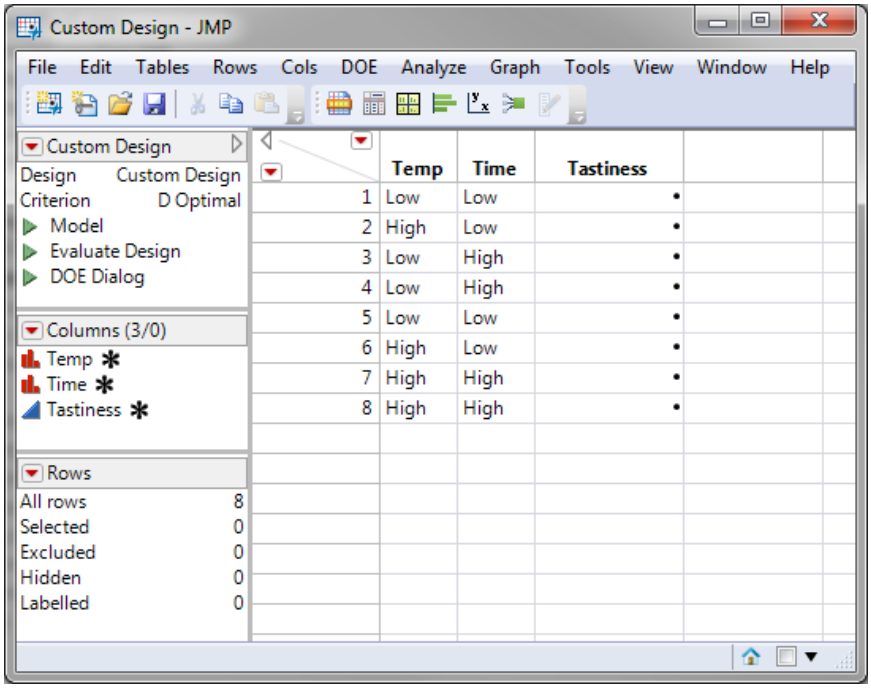


Fig. 4.83 JMP Design and Table

Step 4: Run the experiment and record the responses in the table created by JMP.

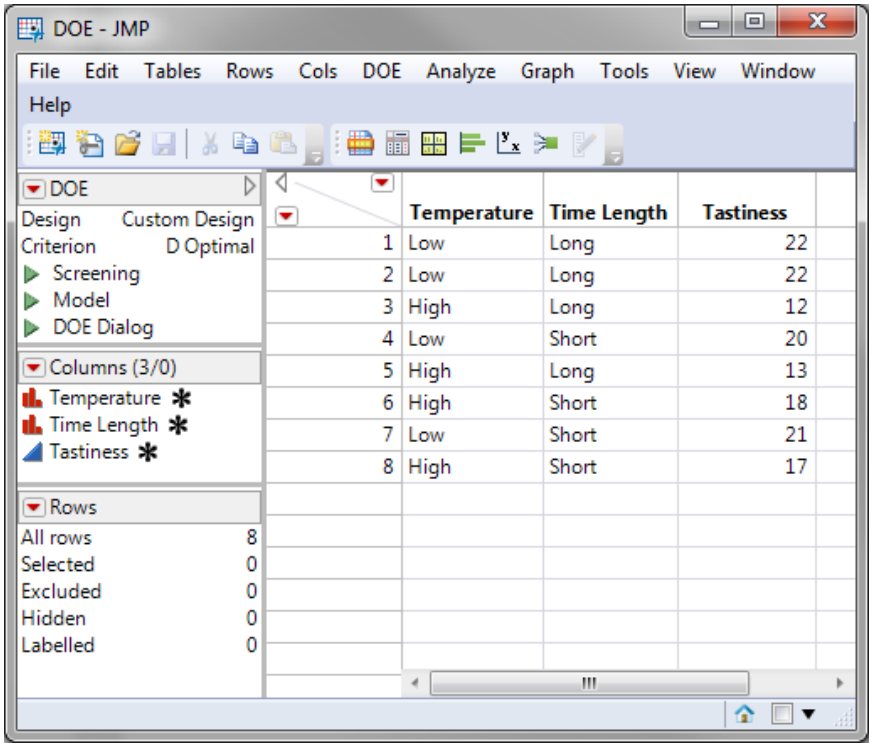


Fig. 4.84 Steps in JMP to Run the Experiment

Step 5: Analyze the experiment results

1. Click Analyze -> Fit Model
2. A box named “Fit Model” pops up in which the response, the two factors and their interaction term are all pre-populated into the “Y” and “Construct Model Effects” boxes respectively
3. Click on “Run” button
4. The statistics of the model appear
5. Analyze the model results

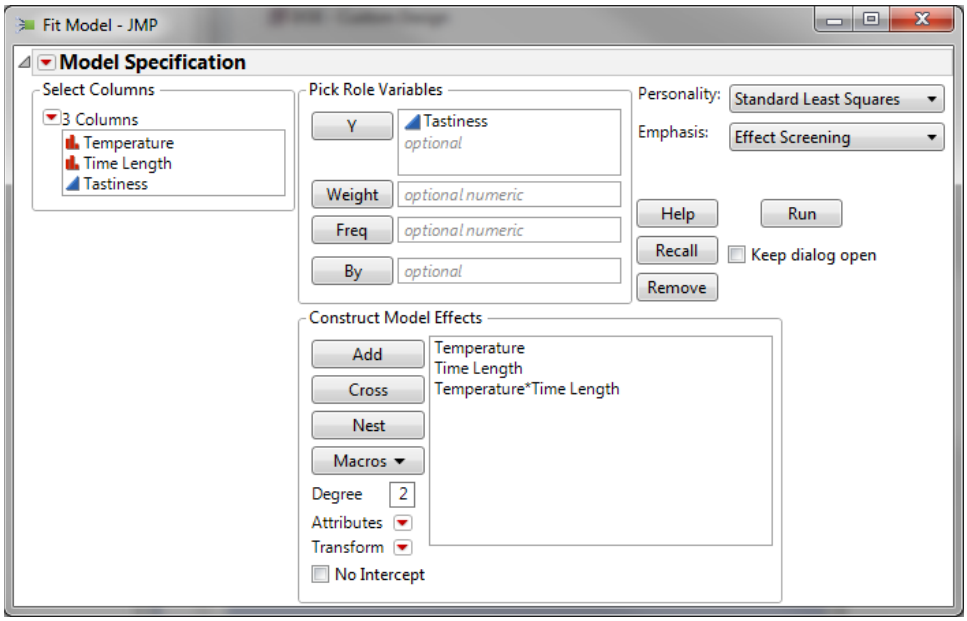


Fig. 4.85 Run the Model

Step 6: Analyze the experiment results

1. Click the red triangle next to “Response Tastiness”
2. Select “Regression Reports”
3. Select “Summary of Fit”
4. Select “Analysis of Variance”
5. Now your “Summary of Fit” and “Analysis of Variance” tables appear.

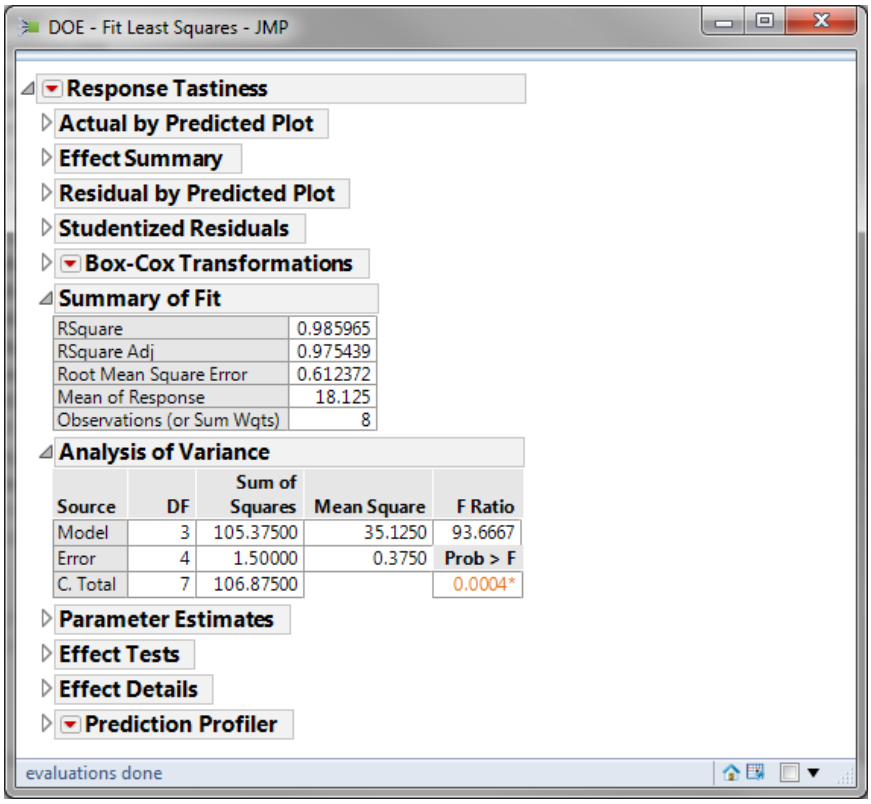


Fig. 4.86 Analysis Variance

High R^2 value shows around 98% of the variation in the response can be explained by the model. A p-value smaller than the alpha level (0.05) indicates that the model is statistically significant (e.g. at least one of the independent variables in the model has a coefficient statistically different from zero).

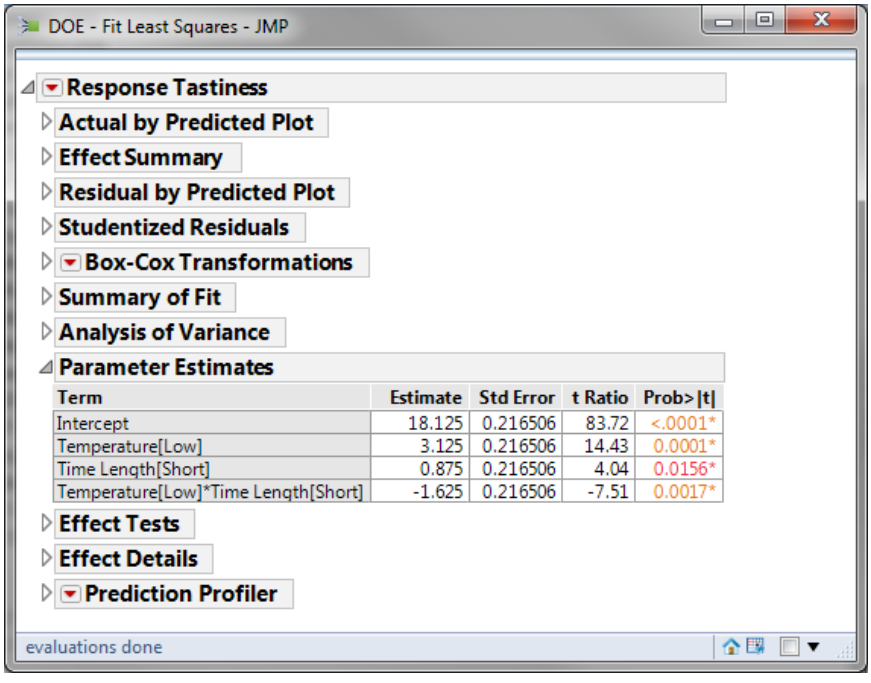


Fig. 4.87 DOE Analysis results

The Parameter Estimates table shows the intercept and coefficients of independent variables in the model. Since p values of all the independent variables in the mode are smaller than alpha level (0.05), both factors and their interaction have statistically significant impact on the response. The output provides intercept values and coefficients for a regression model equation.

4.4.2 Linear and Quadratic Models

Linear DOE

If we assume the relationship between the response and the factors is linear, we use a two-level design experiment in which there are two settings for individual factors. The linear model of a two-level design with three factors is expressed by the formula:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_{12} X_1 X_2 + \alpha_{13} X_1 X_3 + \alpha_{23} X_2 X_3 + \alpha_{123} X_1 X_2 X_3 + \varepsilon$$

Where:

- Y is the response
- X_1, X_2, X_3 are the three two factors
- α_0 is the intercept
- $\alpha_1, \alpha_2 \ldots \alpha_{123}$ are the coefficients of the factors and interactions
- ε is the error of the model.

Quadratic DOE

It is possible that the relationship between the response and the factors is non-linear. To test the non-linear relationship between the response and the factors, we need at least a three-level design, i.e., quadratic design of experiment.

The quadratic DOE adds the $X_1^2, X_2^2 \ldots X_k^2$ and corresponding interactions into the model as potential significant factor of the response.

4.4.3 Balanced and Orthogonal Designs

Balanced Design

In DOE, an experimental design is *balanced* if individual treatments have the same number of observations. In other words, if an experiment is balanced, each level of each factor is run the same number of times. The following design is balanced since the two factors are both run twice at each level.

Run	Treatment	Factors	
		A	B
1	I	-1	-1
2	a	+1	-1
3	b	-1	+1
4	ab	+1	+1

Fig. 4.88 Balanced Design

Orthogonal Design

In DOE, an experiment design is *orthogonal* if the main effect of one factor can be evaluated independently of other factors. In other words, for each level (high setting or low setting) of a factor, the number of high settings and low settings in any other factors must be the same.

In the following design, for the higher level (+1) of factor A, the number of "+1" in factor B is 2 but the number of "−1" in factor B is 0. Therefore, this is *not* an orthogonal design.

Run	Treatment	Factors	
		A	B
1	I	-1	-1
2	a	-1	-1
3	b	+1	+1
4	ab	+1	+1

Fig. 4.89 Non-orthogonal Design

In the following design, for the higher level (+1) of factor A, the number of "+1" in factor B is 1 and the number of "−1" in factor B is 1 as well. Therefore, this is an orthogonal design.

Run	Treatment	Factors	
		A	B
1	I	-1	-1
2	a	+1	-1
3	b	-1	+1
4	ab	+1	+1

Fig. 4.90 Orthogonal Design

Checking for an Orthogonal Design

We can check to see if our design is orthogonal with some very simple math. In an orthogonal design, the sum of products for each run should equal zero. Let us check the earlier design that we said was orthogonal:

Run	Treatment	Factors	
		A	B
1	I	-1	-1
2	a	-1	-1
3	b	+1	+1
4	ab	+1	+1

Fig. 4.91 Checking for an Orthogonal Design

Run 1: −1 * −1 = 1
Run 2: +1 * −1 = −1
Run 3: −1 * +1 = −1
Run 4: +1 * +1 = 1

Sum = **0**

Therefore, the design is orthogonal.

Now let us check the earlier design that we said was not orthogonal:

Run	Treatment	Factors	
		A	B
1	I	-1	-1
2	a	-1	-1
3	b	+1	+1
4	ab	+1	+1

Fig. 4.92 Checking for Non-Orthogonal design

Run 1: −1 * −1 = 1
Run 2: −1 * −1 = 1
Run 3: +1 * +1 = 1
Run 4: +1 * +1 = 1

Sum = **4**

Therefore, the design is not orthogonal.

4.4.4 Fit, Diagnosis, and Center Points

Use JMP to Fit the Model of DOE

In the example introduced in last module, we used “Analyze -> Fit Model” to generate a linear model as follows for the response, i.e., the tastiness of the cake.

If one of the factors or interactions has a p-value larger than alpha level (0.05), it indicates that the particular factor or interaction does *not* have statistically significant impact on the response. To fit the model better, we need to remove the insignificant independent variables one at a time and run the model again until all the independent variables in the model are statistically significant (i.e., p-value smaller than alpha level). Therefore, remove insignificant factors and interactions, one at a time, starting with highest p-value > 0.05, until all factors and interactions are statistically significant.

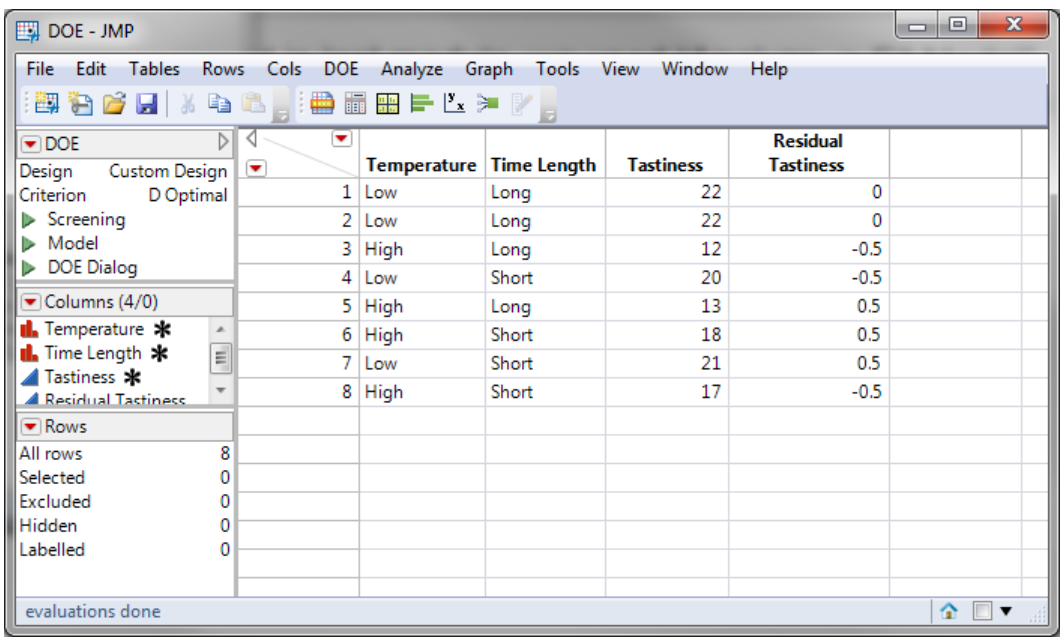
Use JMP to Diagnose the Model

To ensure that the quality of the model is acceptable, we need to conduct the residuals analysis. Residual is the difference between the actual response value and the fitted response value. Well-performing residuals:

- Are normally distributed
- Have a mean equal to zero
- Are independent
- Have equal variance across the fitted values

Step 1: Save the residuals in the data table

1. Click on the red triangle button next to "Response Tastiness"
2. Click "Save Columns" -> "Residuals"
3. A new column of residuals would pop into the data table.



	Temperature	Time Length	Tastiness	Residual Tastiness
1	Low	Long	22	0
2	Low	Long	22	0
3	High	Long	12	-0.5
4	Low	Short	20	-0.5
5	High	Long	13	0.5
6	High	Short	18	0.5
7	Low	Short	21	0.5
8	High	Short	17	-0.5

Fig. 4.93 Residuals data table

Step 2: Check whether the mean of residuals is zero and residuals are normally distributed.

1. Click Analyze -> Distribution

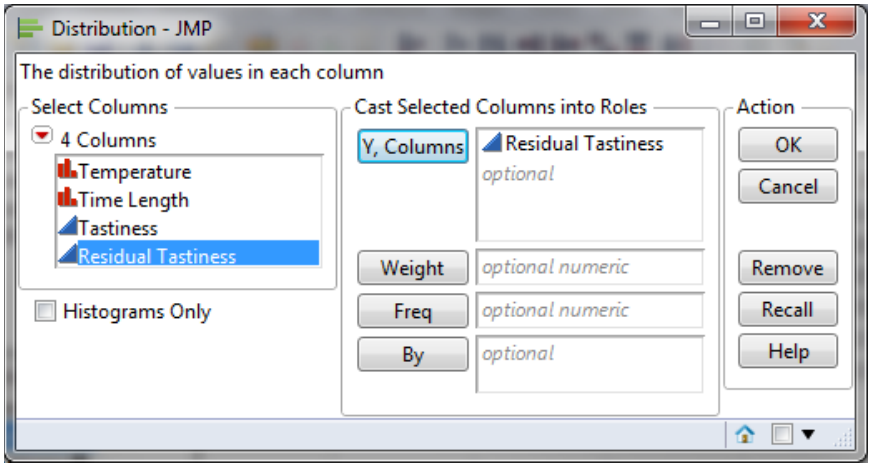


Fig. 4.94 Select "Residual Tastiness" as "Y, Columns"

2. Click "OK"
3. Click on the red triangle button next to "Residual Tastiness"
4. Click Continuous Fit -> Normal
5. Click on the red triangle button next to "Fitted Normal"
6. Select "Goodness of Fit"

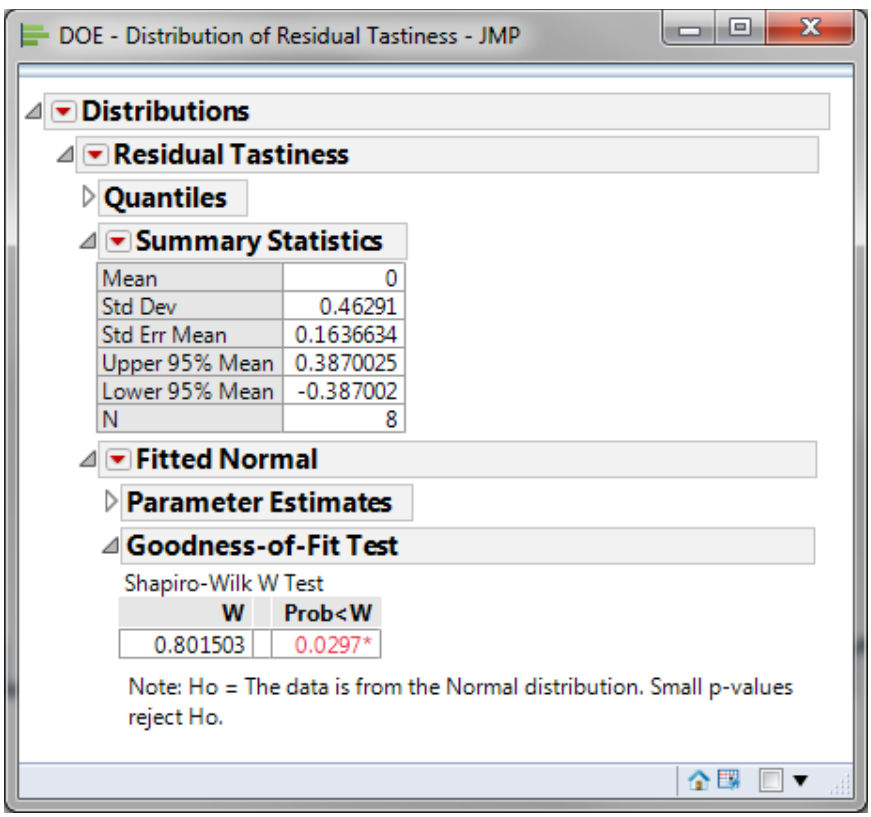


Fig. 4.95 Normality Test with Variable

When the p-value of the normality test is greater than alpha level, the residuals are normally distributed.

Step 3: Check whether the residuals are independent.

1. Click Analyze -> Quality & Process -> Control Chart -> IR
2. Select “Residual Tastiness” as “Process”

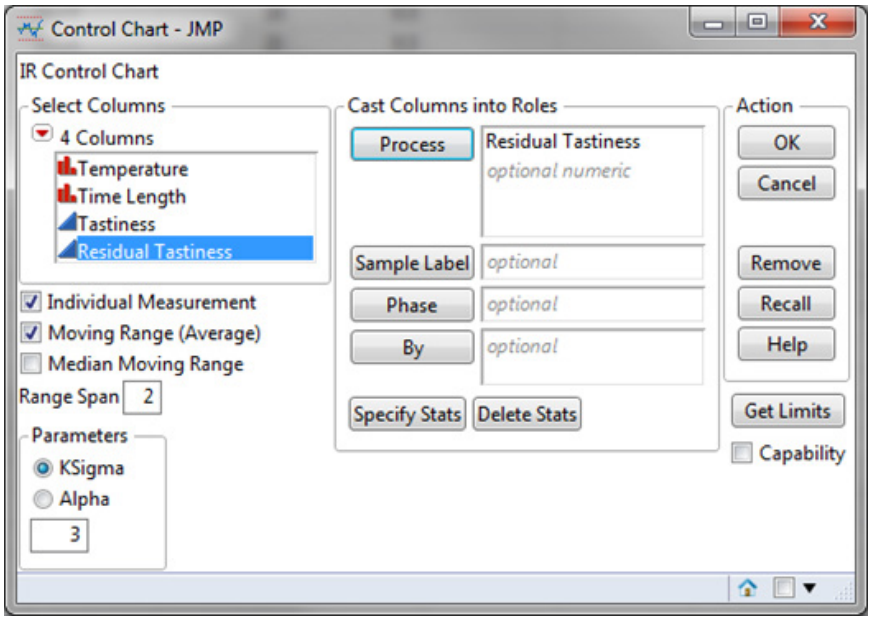


Fig. 4.96 Control Chart Residual selections

3. Click “OK”
4. I-MR chart of residuals are generated.
5. Click on the red triangle button next to “Individual Measurement of Residual Tastiness”
6. Click Tests -> All Tests
7. If no data points on the control charts fail any tests, the residuals are in control and independent of each other.

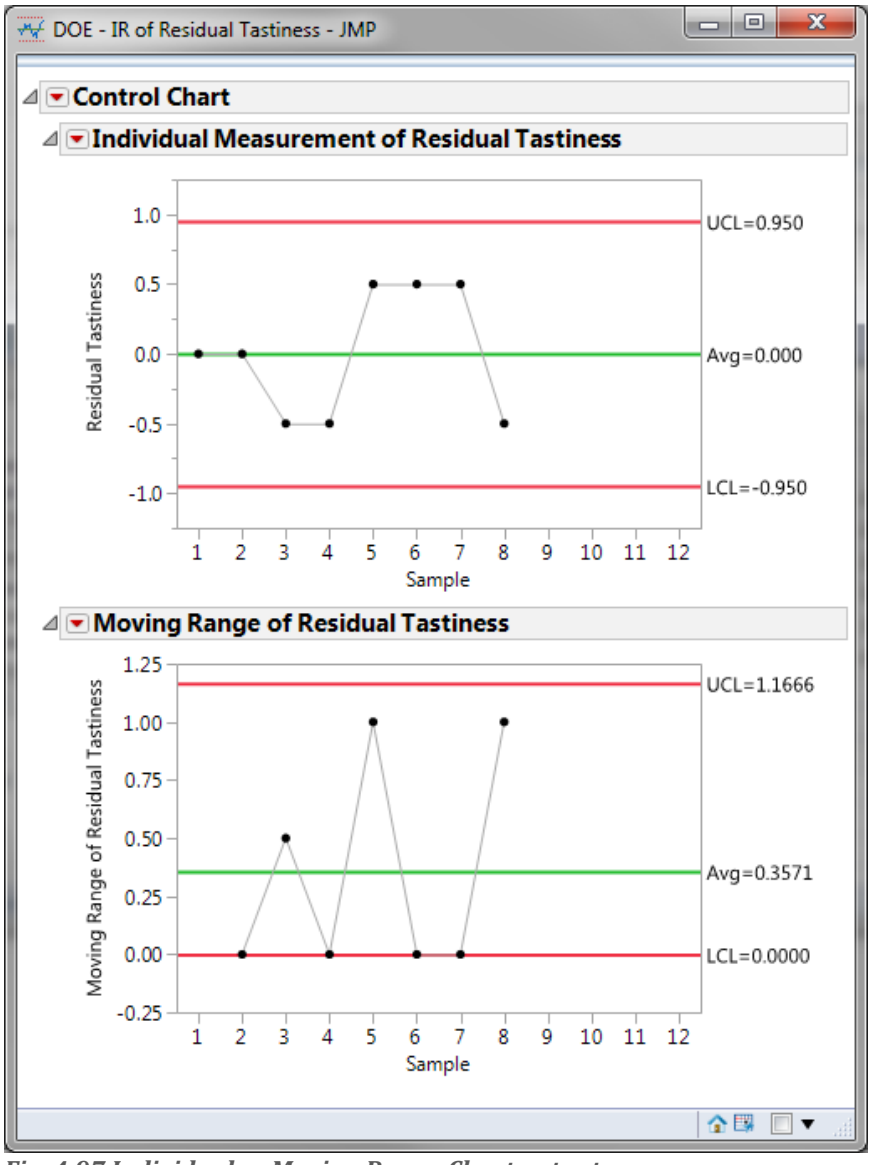


Fig. 4.97 Individuals – Moving Range Chart output

Note: The prerequisite of plotting IR chart for residuals is that the residuals are in the time order.

Step 4: Check whether the residuals have equal variance across the predicted response values.

- At the bottom of the analysis results page of fitting the DOE model is the residual by predicted plot.
- We look for patterns in which the residuals tend to have even variation across the entire range of the fitted response values.

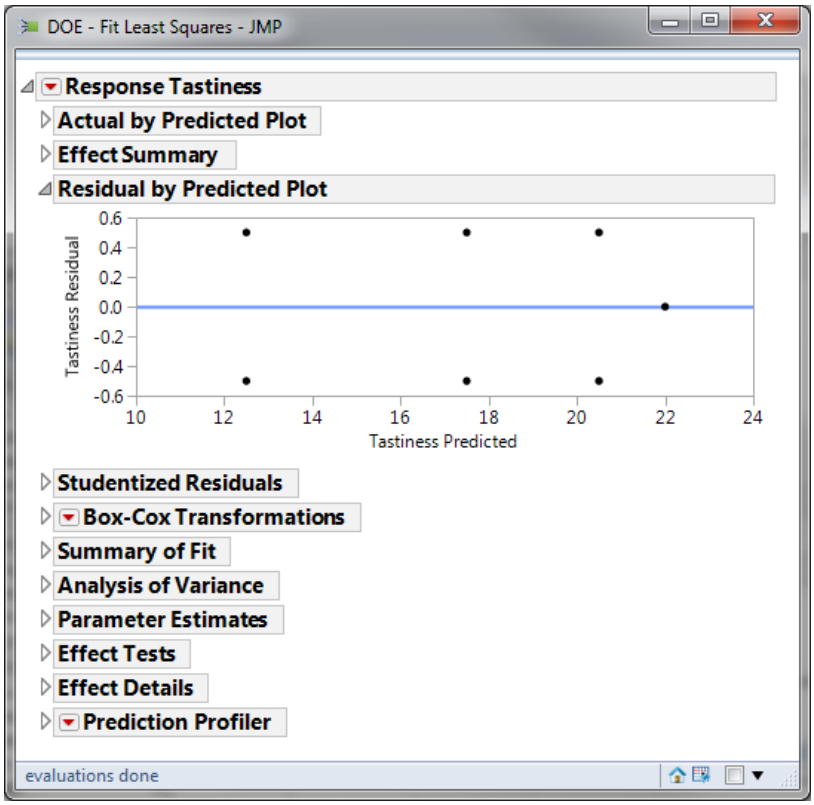


Fig. 4.98 Regular Residuals vs Predicted Values for: Tastiness

Step 5: If the residuals are performing well, we can use the model to find the best settings of factors to maximize the tastiness of the cake. The "Prediction Profiler" already appears in the analysis results page.

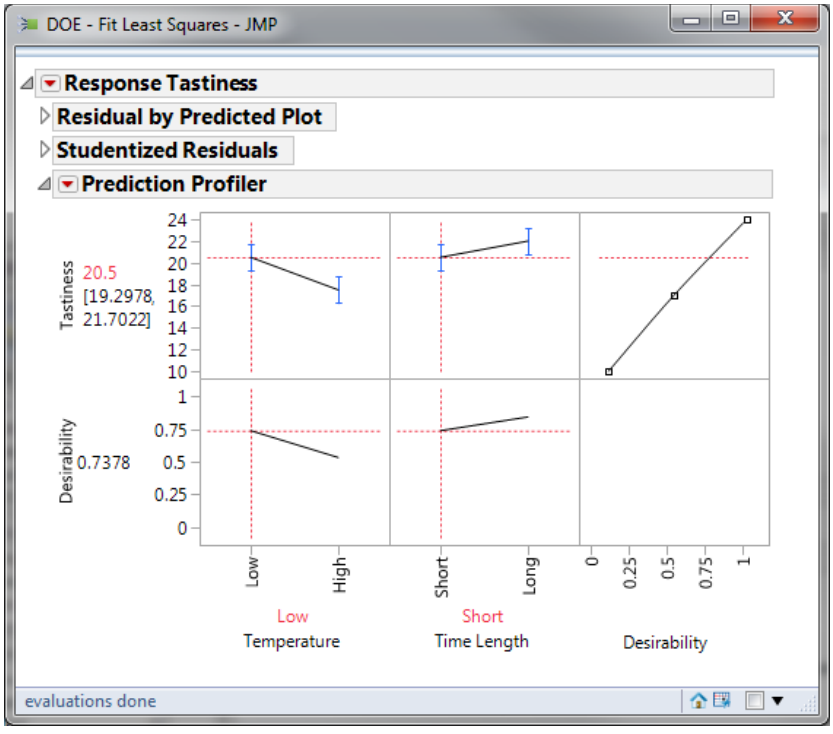


Fig. 4.99 Prediction Profiler output

Center Points in DOE

The two-level design of experiment assumes that the relationship between the response and the factors is linear. We can use center points to check whether the assumption is true by adding center point runs to the experiment. In each center point run, the factors are all set to be in the center setting (zero) between high (+1) and low (−1) settings.

- Center points do not change the model to quadratic.
- They allow a check for adequacy of linear model.
- If a line connecting factor settings passes through the center of the design, the model is adequate to predict within the inference space.

4.5 FRACTIONAL FACTORIAL EXPERIMENTS

4.5.1 FRACTIONAL DESIGNS

What Are Fractional Factorial Experiments?

In simple terms, a *fractional factorial experiment* is a subset of a full factorial experiment.

- Fractional factorials use fewer treatment combinations and runs.
- Fractional factorials are less able to determine effects because of fewer degrees of freedom available to evaluate higher order interactions.
- Fractional factorials can be used to screen a larger number of factors.
- Fractional factorials can also be used for optimization.

Why Fractional Factorial Experiments?

To run a full factorial experiment for k factors, we need 2^k unique treatments. In other words, we need resources that can afford at least 2^k runs.

With k increasing, the number of runs required in full factorial experiments rises dramatically even without any replications, and the percentage of degrees of freedom spent on the main effects decreases. However, the higher order interactions (3 or 4 factor interactions) can typically be ignored, which allows us to run fewer trials to understand the main effects and two-way interactions.

The main effects and two-way interaction are the key effects we need to evaluate. The higher order the interaction is, the more we can ignore it.

Number of Factors	Number of Treatments
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024

Fig. 4.100 Fractional Factorial Experiment

Notice the number of treatments increases dramatically as factors are added.

How Does a Fractional Factorial Work?

We are trying to find the cause-and-effect relationship between a response (Y) and three factors (factor A, B, and C) and their interactions (AB, BC, AC, and ABC). As follows is the 2^3 full factorial design (2 level 3 factor). There are eight treatment combinations (2 * 2 * 2).

Run	Treatment	Factor			2-Way Interaction			3-Way Interaction
		A	B	C	AB	BC	AC	ABC
1	I	-1	-1	-1	+1	+1	+1	-1
2	a	+1	-1	-1	-1	+1	-1	+1
3	b	-1	+1	-1	-1	-1	+1	+1
4	ab	+1	+1	-1	+1	-1	-1	-1
5	c	-1	-1	+1	+1	-1	-1	+1
6	ac	+1	-1	+1	-1	-1	+1	-1
7	bc	-1	+1	+1	-1	+1	-1	-1
8	abc	+1	+1	+1	+1	+1	+1	+1

Fig. 4.101 Fractional Factorial Treatment Combinations

To perform a 2^3 full factorial experiment, we need to run at least eight unique treatments (2 * 2 * 2).

What if we only have enough resources to run four treatments? If this is the case, we need to carefully select a subset from the eight treatments so that all of our main effects can be evaluated and the design can be kept balanced and orthogonal.

Example of an invalid design
This design is invalid because only the low setting of factor C is tested. We cannot evaluate the main effect of factor C using this design. Remember orthogonality?

Factor		
A	B	C
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1

Fig. 4.102 Example of Invalid Design

This design is also invalid because it is neither balanced nor orthogonal. Checking orthogonality: the sum of AC interaction signs should equal zero (0).

- Run 1 (−)
- Run 2 (−)
- Run 3 (−)
- Run 4 (+)
- Sum (−1)

Factor		
A	B	C
+1	+1	-1
+1	-1	-1
-1	+1	+1
+1	+1	+1

Fig. 4.103 Non-Orthogonal design

This design has a low and high setting for each factor, but is not orthogonal. To select the four treatments run in the 2^{3-1} fractional factorial experiment, we start from the 2^2 full factorial design of experiment. If we replace the two-way interaction (AB) column with the factor C column, the design will be valid.

A	B	AB
-1	-1	+1
+1	-1	-1
-1	+1	-1
+1	+1	+1

A	B	C
-1	-1	+1
+1	-1	-1
-1	+1	-1
+1	+1	+1

Fig. 4.104 Valid Design

Imagine a two-factor full factorial with factors A and B. We also learn about the interaction of A and B. In a fractional factorial, we sacrifice learning about the two-way interaction between A and B, and substitute factor C.

2^{3-1} Fractional Factorial Design Pattern
This pattern implies three factors and four treatments.

Run	Treatment	Factor			2-Way Interaction			3-Way Interaction
		A	B	C	AB	BC	AC	ABC
1	c	-1	-1	+1	+1	-1	-1	+1
2	a	+1	-1	-1	-1	+1	-1	+1
3	b	-1	+1	-1	-1	-1	+1	+1
4	abc	+1	+1	+1	+1	+1	+1	+1

Fig. 4.105 2^{3-1} Fractional Factorial Design Pattern

Note: We also call this kind of design a *half-factorial design* since we only have half of the treatments that we would have in a full factorial design. In 2^{3-1} fractional factorial design of experiment, the effect of three-way interaction (ABC) is not measurable since it only has "+1".

In four runs, we are able to run high and low settings for each of the three factors. The three-way interaction ABC is only at the high setting. In the 2^{3-1} fractional factorial design, we notice that the column of each main effect has identical "+1" and "−1" values with one two-way interaction column.

- A and BC
- B and AC
- C and AB

In this situation, we say that A is *aliased* with BC or A is the *alias* of BC. By multiplying any column with itself, we obtain the *identity (I)*.

$$A * A = I$$

The product of any column and the identity is the column itself.

$$A * I = A$$

Column ABC is called the *generator*. By multiplying any column with the generator, we obtain its *alias*.

$$A * ABC = (A * A) * BC = I * BC = BC$$

Use JMP to Run a Fractional Factorial Experiment

Step 1: Initiate the experiment design

1. Click DOE -> Classical->Screening Design

2. A window named “DOE – Screening Design” pops up.
3. One response: Y
4. Add Three factors: A, B and C by clicking the “Categorical” button-> 2-Level, rename the factors, A, B and C
5. Two-level design: each factor has two settings, -1 and +1
6. Click “Continue”
7. Select “Choose from a list of Fractional Factorial Designs”, click “Continue”.
8. Select “Fractional Factorial” and click “Continue”

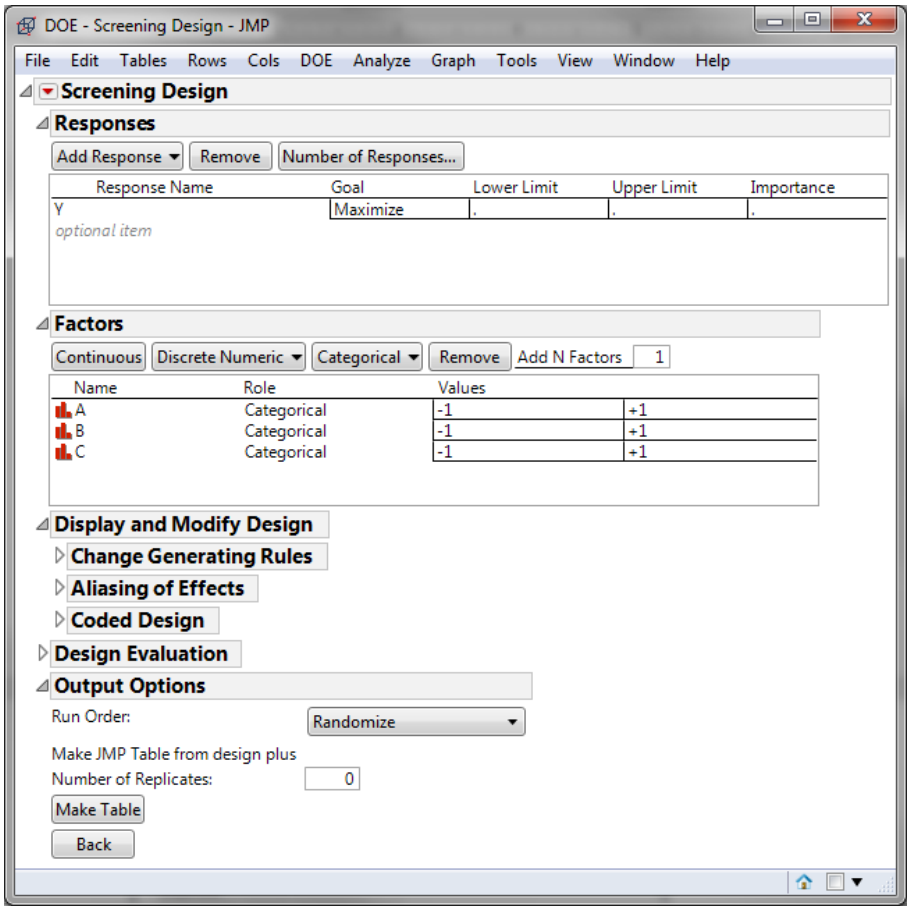


Fig. 4.106 Creating a Table

Step 2: Select the number of replications, make design

1. In this case, we want each treatment to be run twice so that there are enough degrees of freedom to analyze the effects.
2. Enter “1” in the box for “Number of Replicates”
3. Click “Make Table” button

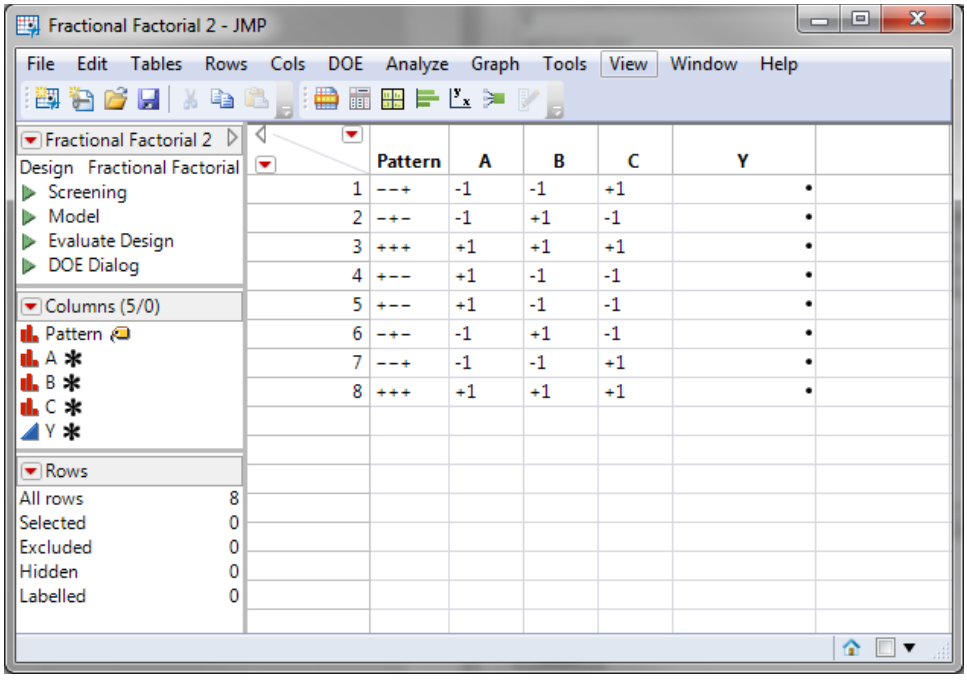


Fig. 4.107 JMP DOE Design and Data Collection Table

Step 3: Implement the experiment and record the results in the table. For this lesson, use the data file "Fractional Factorial.jmp" with the 'Y' data already entered. If you do not use this file then you may not get the same results displayed during this lesson.

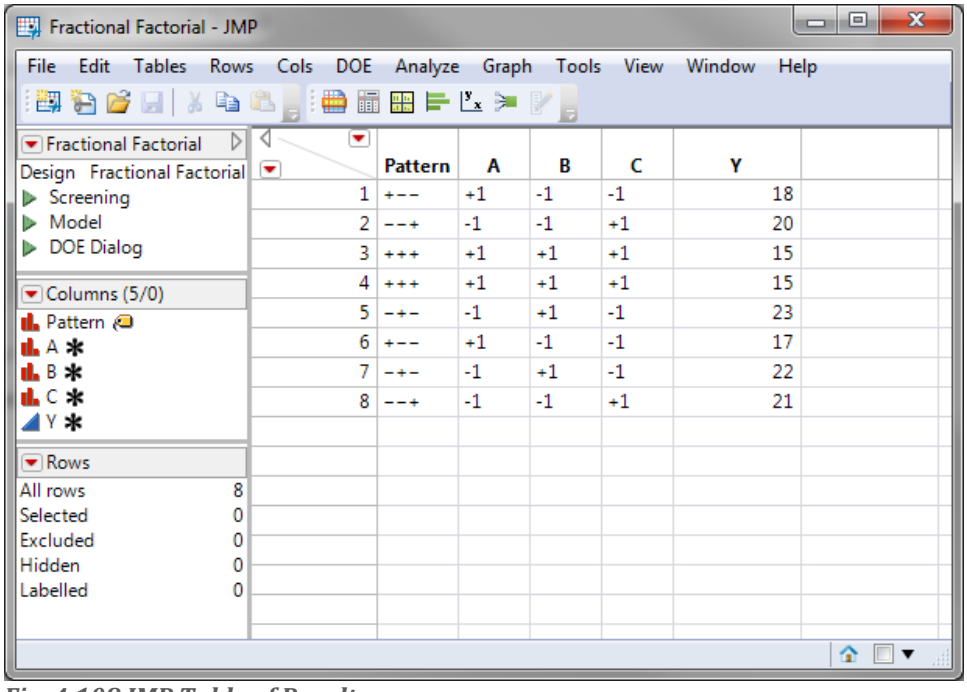


Fig. 4.108 JMP Table of Results

Step 4: Fit the model using the experiment results

1. Click Analyze-> Fit Model
2. The Y and the Xs have been pre-populated into their boxes

3. Click “Run Model”

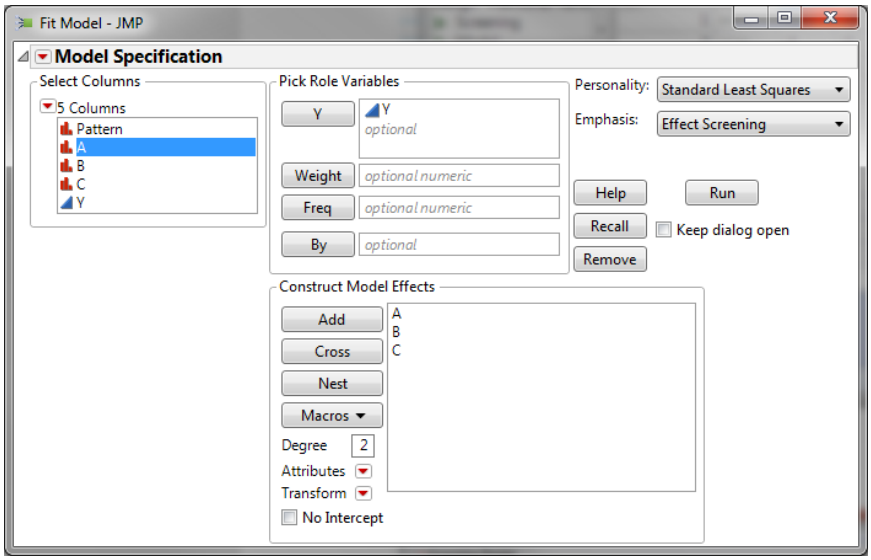


Fig. 4.109 Fit and Run the Model

Step 5: Analyze the model results

1. Click on the red triangle next to “Response Y”
2. Select “Regression Reports -> “Summary or Fit”
3. Select “Regression Reports -> “Analysis of Variance”
4. Check whether the model is statistically significant.
5. Check which factors are insignificant.
6. If any independent variables are not significant, remove them one at a time and rerun the model until all the independent variables in the model are significant.

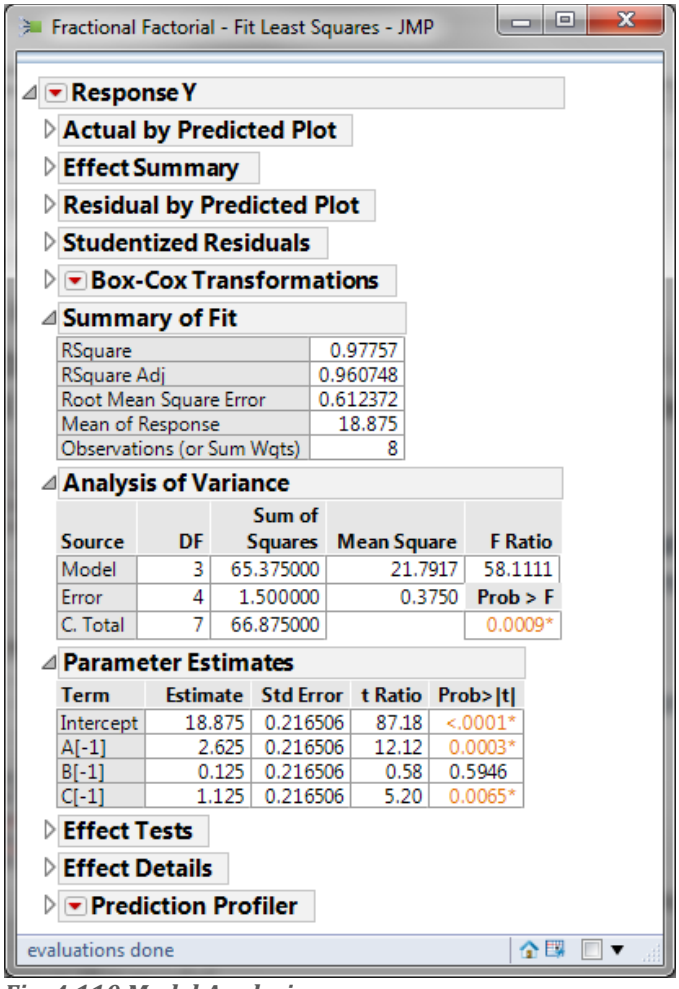


Fig. 4.110 Model Analysis

Step 6: The p-value of factor B is greater than the alpha level (0.05), so it is not statistically significant. Next we need to remove factor B and re-run the model.

1. Click the red triangle next to “Response Y”
2. Select “Model Dialog”
3. Remove Factor B from the model
4. Click “Run”

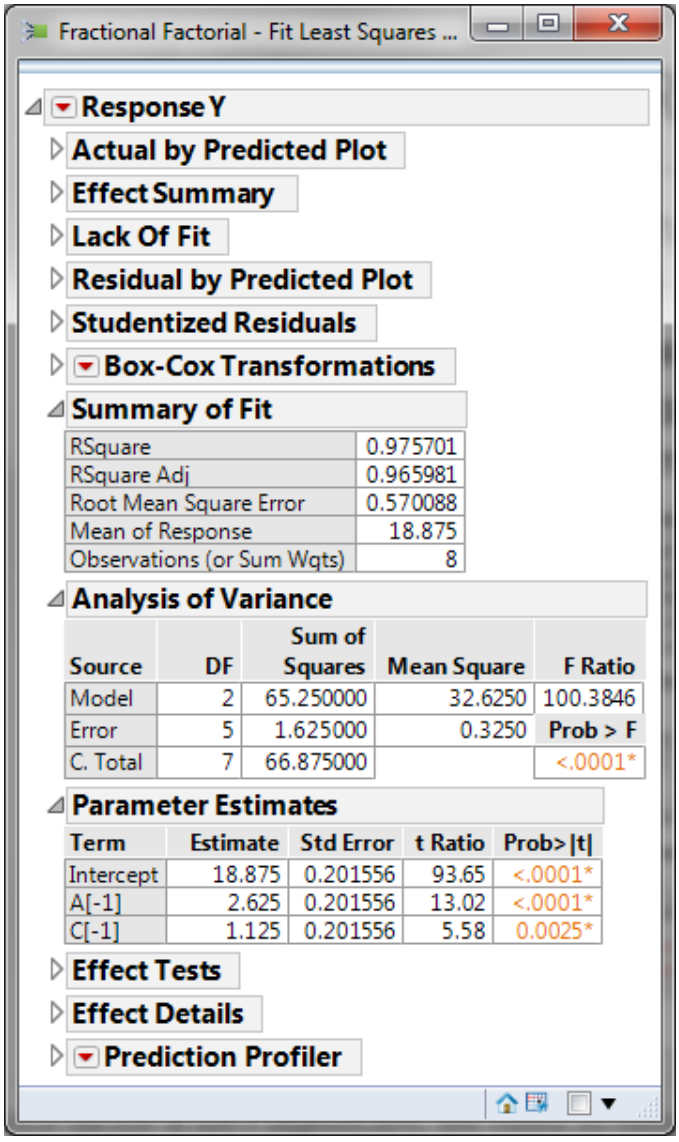


Fig. 4.111 Model Results with B removed

The p-values of all the independent variables are smaller than 0.05. There is no need to remove any independent variables from the model.

Perform residual analysis to ensure that the residuals of the model satisfy the following criteria.

- Mean is equal to zero.
- Normally distributed
- Independent
- Equal variance across the fitted response values

Step 7: Save residuals generated by the model

1. Click on the red triangle button next to “Response Y”
2. Click Save Columns -> Residuals

3. A new column of residuals will appear in the data table

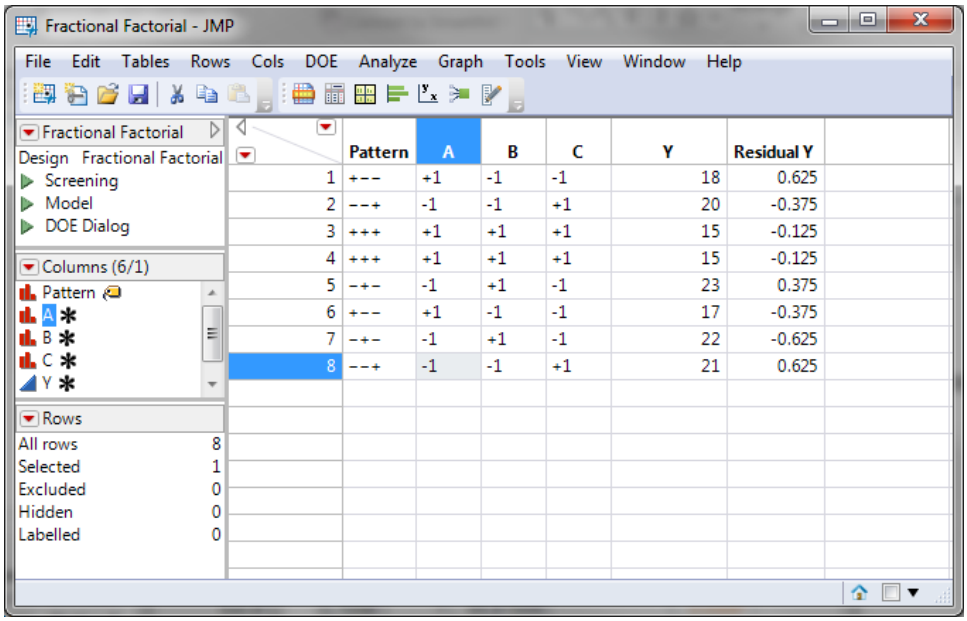


Fig. 4.112 Analyze 2-Level Factorial Design with Plots Selection

Step 8: Check whether residuals are normally distributed with mean equal to zero.

1. Click Analyze -> Distribution
2. Select "Residual Y" in the "Y, Columns"

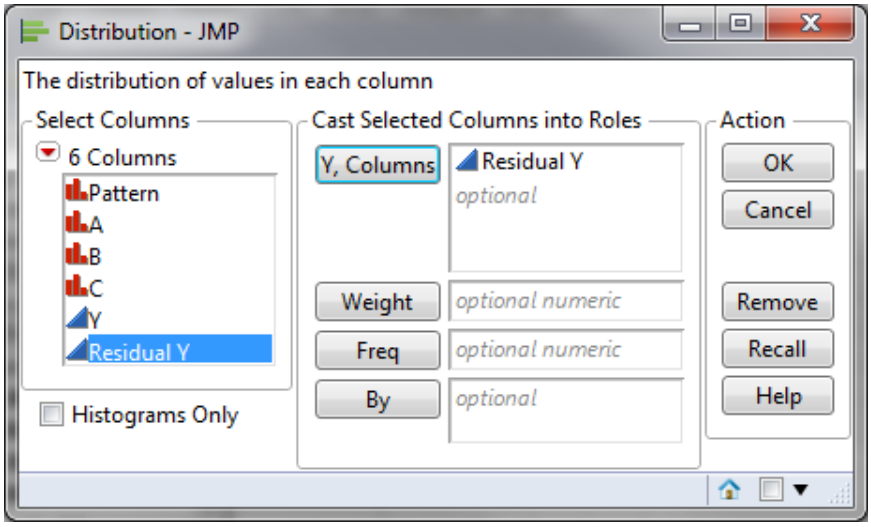


Fig. 4.113 Select Residuals

3. Click "OK"
4. Click on the red triangle button next to "Residual Y"
5. Click Continuous Fit -> Normal
6. Click on the red triangle button next to "Fitted Normal"
7. Click "Goodness of Fit"

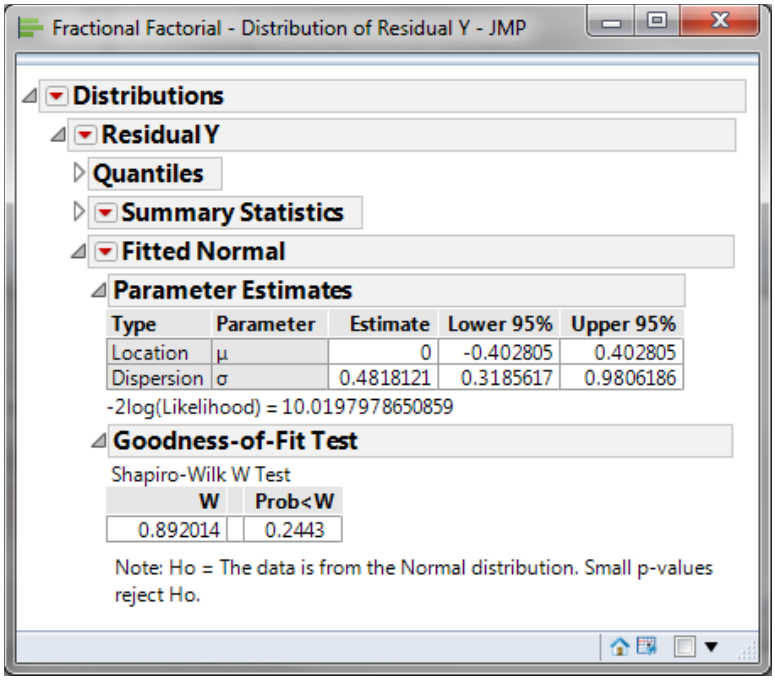


Fig. 4.114 Data output for Fractional Factorial Design

If the p-value of the normality test is greater than the alpha level (0.05), the residuals are normally distributed.

Step 8: Check if the residuals are independent.

1. Click Analyze -> Quality & Process -> Control Chart -> IR
2. Select "Residual Y" as "Process"

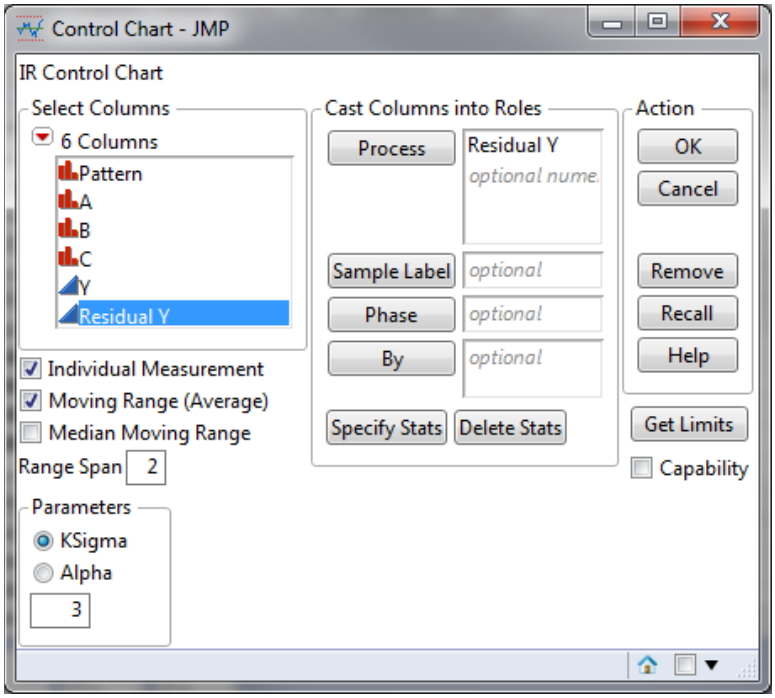


Fig. 4.115 Residual Selection

3. Click “OK”
4. I-MR chart of residuals are generated.
5. Click on the red triangle button next to “Individual Measurement of Residual Tastiness”
6. Click Tests -> All Tests
7. If no data points on the control charts fail any tests, the residuals are in control and independent of each other.

Note: The prerequisite of plotting an IR chart for residuals is that the residuals are in time order.

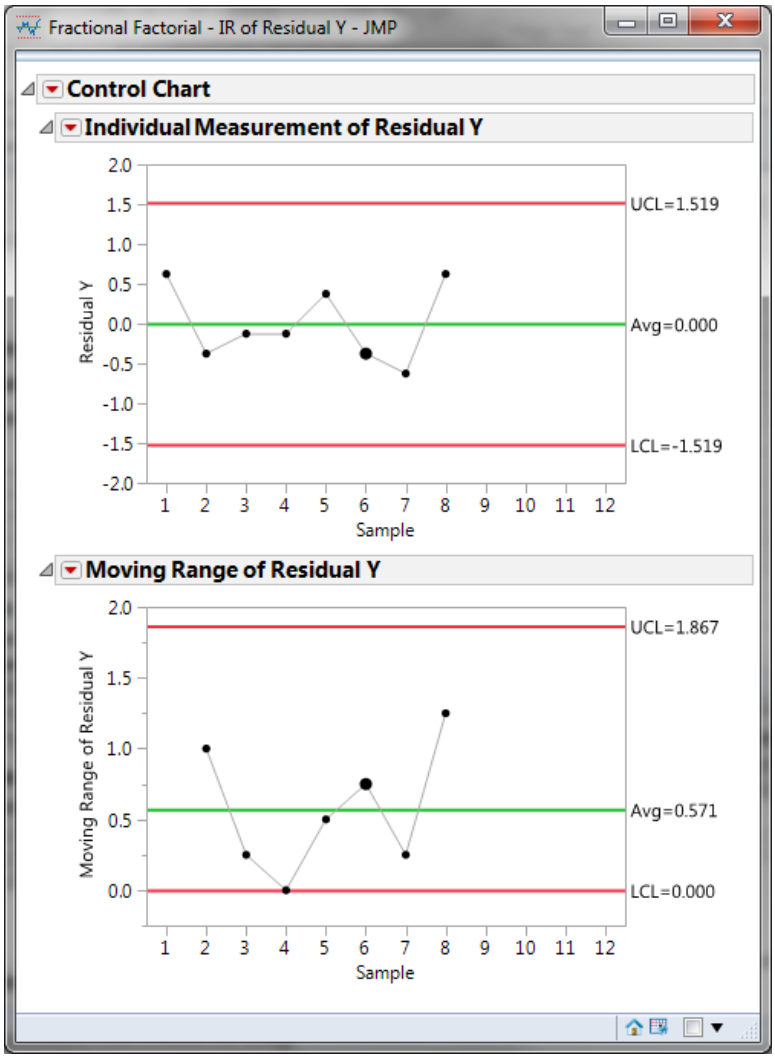


Fig. 4.116 Individuals- Moving Range Chart output

Step 9: Check whether the residuals have equal variance across the predicted response values.

- Within the analysis results page of fitting the DOE model is the “Residual by Predicted Plot.”
- We look for patterns in which the residuals tend to have even variation across the entire range of the fitted response values.

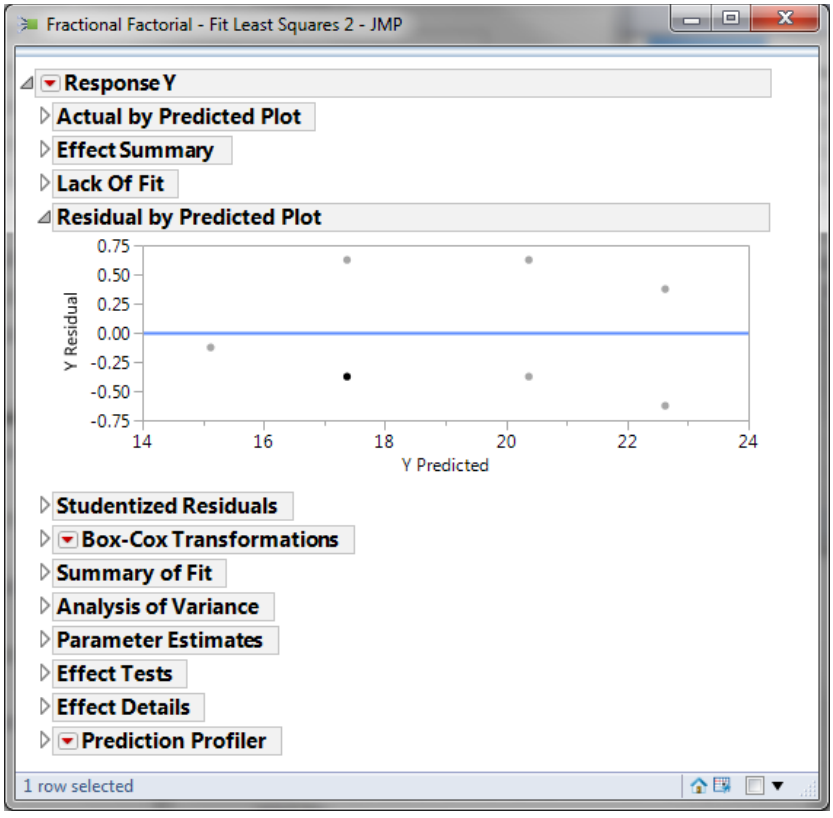


Fig. 4.117 Regular Residuals vs Predicted Values for: Y

4.5.2 CONFOUNDING EFFECTS

Confounding Effects

In full factorial experiments the effects of every factor and their interactions can be calculated because there is a degree of freedom for every treatment combination as well as the error term. Fractional factorial experiments, however, are designed with fewer runs or treatment combinations but will have the same number of input factors. As a result, the experimenter must understand the impact of these consequences.

Confounding (or aliased factors) is that consequence.

Resolution is a quantification or degree of confounding.

Confounding can be considered alternatively as "confusing". If you think of it this way, confounding is when two factors or factor interactions "confuse" our ability to know which factor has the effect on 'Y'. When two input factors are aliases of each other, the effects they each have on the response cannot be separated.

In the 2^{3-1} fractional factorial design, A and BC are aliases. The main effect that A has on the response would be exactly the same as the interaction effect of BC. We cannot tell which one is truly significant. The effects of aliased factors are what is referred to as *confounded*. Due to the

confounding nature of aliased factors, the effect we observe is the mixed effect of aliased factors.

- A + BC
- B + AC
- C + AB

Resolution

Resolution is the measure or degree of confounding. Higher resolution means less confounding or merely confounding main effects with higher order interactions. The resolution of a fractional factorial experiment will always be one number higher than the order of interactions that are confounded with main effects.

- Resolution III: Main effects are confounded with two-way interactions. If you are only interested in main effects, then resolution III is acceptable. Otherwise, resolution III is typically undesirable.
- Resolution IV: Main effect are confounded with three-way interactions.
- Resolution V: Main effects are confounded with four-way interactions.

Consequences of Fractional Factorials

Benefits:

- The primary benefits of properly designed fractional factorial experiments are achieved because of fewer runs.
- You may have limited time, resources, or capital and cannot run a full factorial.

Disadvantages:

- Confounding (a.k.a. aliasing)
- Mitigation Option

Replications: By adding replicates you can introduce more study runs and increase the resolution of your experiment. Replications, however, will also increase the number of trials or runs sometimes, defeating the purpose of running a fractional factorial experiment.

4.5.3 Experimental Resolution

Alias in 2^{3-1} Fractional Factorial Experiment

A 2^{3-1} is a resolution III design. In the 2^{3-1} fractional factorial design of experiment, the main factor is *aliased* or *confounded* with two-way interactions.

- A is the alias of BC.
- B is the alias of AC.
- C is the alias of AB.

Alias in 2^{4-1} Fractional Factorial Experiment

A 2^{4-1} is a resolution IV design. In the 2^{4-1} fractional factorial design of experiment, the main factor is aliased with three-way interaction and two-way interaction is aliased with two-way interaction.

- A is the alias of BCD
- B is the alias of ACD
- C is the alias of ABD
- D is the alias of ABC
- AB is the alias of CD
- AC is the alias of BD
- AD is the alias of BC

Design Resolution

Design resolution (i.e., experimental resolution) refers to the confounding patterns indicating how the effects are confounded with others. If two factors are aliases, they have confounded effect on the response. There is no need to consider both factors in the model due to the redundancy. Only one factor needs to be included in the model for simplicity.

Most popular resolutions are resolution III, IV, and V.

- Resolution III: Main effects are aliased with two-way interactions.
- Resolution IV: Main effects are aliased with three-way interactions. Two-way interactions are aliased with other two-way interactions.
- Resolution V: Main effects are aliased with four-way interactions. Two-way interactions are aliased with three-way interactions.

Determine Aliases

To find the alias of a particular factor, we only need to multiply the factor of interest with the *identity (I)* of the design.

- In 2^{3-1} fractional factorial design, ABC is the identity. To find the alias of factor A, we use A * ABC = BC.
- In 2^{4-1} fractional factorial design, ABCD is the identity. To find the alias of factor A, we use A * ABCD = BCD.
- In 2^{5-1} fractional factorial design, ABCDE is the identity. To find the alias of factor A, we use A * ABCDE = BCDE.

The identity of a design is the same number of letters in the alphabet starting from A out to the same number factors in the design. For example, a three-factor design has an identity of ABC, a four-factor design has an identity of ABCD. To find the alias of a factor, multiply the factor by the identity of the design. A multiplied by ABC leaves BC. Therefore, in a resolution III design, factor A is aliased with BC.

5.0 CONTROL PHASE

5.1 LEAN CONTROLS

5.1.1 CONTROL METHODS FOR 5S

What is 5S?

5S is a methodology that is applied in a Lean workplace. It is a systematic method to organize, order, clean, and standardize a workplace . . . and keep it that way! 5S is summarized in five Japanese words, all starting with the letter S:

- *Seiri* (sorting)
- *Seiton* (straightening)
- *Seiso* (shining)
- *Seiketsu* (standardizing)
- *Shisuke* (sustaining)

Originally developed in Japan, 5S is widely used to optimize the workplace to increase productivity and efficiency.

5S Goals

1. Reduced waste
2. Reduced cost
3. Establish a work environment that is:
   - Self-explaining
   - Self-ordering
   - Self-regulating
   - Self-improving.

   Where there is/are no more:

   - Wandering and/or searching
   - Waiting or delays
   - Secrets hiding spots for tools
   - Obstacles or detours
   - Extra pieces, parts, materials, etc.
   - Injuries
   - Waste

5S Benefits

By meeting the goals of 5S, the following benefits are achieved.

- Reduced changeovers
- Reduced defects

- Reduced waste
- Reduced delays
- Reduced injuries
- Reduced breakdowns
- Reduced complaints
- Reduced red ink
- Higher quality
- Lower costs
- Safer work environment
- Greater associate and equipment capacity

5S Systems Reported Results

Organizations that have implemented 5S systems have reported some very impressive results: reduction in waste, improved safety, reduced costs, and improved performance.

Cut in floor space:	60%
Cut in flow distance:	80%
Cut in accidents:	70%
Cut in rack storage:	68%
Cut in number of forklifts:	45%
Cut in machine changeover time:	62%
Cut in annual physical inventory time:	50%
Cut in classroom training requirements:	55%
Cut in nonconformance in assembly:	96%
Increase in test yields:	50%
Late deliveries:	0%
Increase in throughput:	15%

Sorting (Seiri)

- Go through all the tools, parts, equipment, supply, and material in the workplace.
- Categorize them into two major groups: needed and unneeded.
- Eliminate the unneeded items from the workplace. Dispose of or recycle those items.
- Keep the needed items and sort them in the order of priority. When in doubt . . . throw it out!

Straightening (Seiton)

Straightening in 5S is also called setting in order.

- Label each needed item.
- Store items at their best locations so that the workers can find them easily whenever they needed any item.
- Reduce the motion and time required to locate and obtain any item whenever it is needed.
- Promote an efficient work flow path.

- Use visual aids like the tool board image on this page to remind employees where things belong so the order is sustained.

Shining (Seiso)

Shining in 5S is also called *sweeping*.

- Clean the workplace thoroughly.
- Maintain the tidiness of the workplace.
- Make sure every item is located at the specific location where it should be.
- Create the ownership in the team to keep the work area clean and organized.

Standardizing (Seiketsu)

Standardize the workstation and the layout of tools, equipment, and parts.

Create identical workstations with a consistent way of storing the items at their specific locations so that workers can be moved around to any workstation any time and perform the same task.

Sustaining (Shisuke)

Sustaining in 5S is also called self-discipline.

- Create the culture in the team to follow the first four S's consistently.
- Avoid falling back to the old ways of cluttered and unorganized work environment.
- Keep the momentum of optimizing the workplace.
- Promote innovations of workplace improvement.
- Sustain the first fours S's using:
    - 5S Maps
    - 5S Schedules
    - 5S Job cycle charts
    - Integration of regular work duties
    - 5S Blitz schedules
    - Daily workplace scans

Simplified Summary of 5S

1. *Sort* - "When in doubt, move it out"
2. *Set in Order*—Organize all necessary tools, parts, and components of production. Use visual ordering techniques wherever possible.
3. *Shine*—Clean machines and/or work areas. Set regular cleaning schedules and responsibilities.
4. *Standardize*—Solidify previous three steps, make 5S a regular part of the work environment and everyday life

5. *Sustain*—Audit, manage, and comply with established five-s guidelines for your business or facility

5.1.2 KANBAN

What is Kanban?

The Japanese word *Kanban* means signboard. *Kanban system* is a "pull" production scheduling system to determine when to produce, what to produce, and how much to produce based on the demand. It was originally developed by Taiichi Ohno in order to reduce the waste in inventory and increase the speed of responding to the immediate demand.

Kanban system is a demand-driven system.

- The customer demand is the signal to trigger or pull the production.
- Products are made only to meet the immediate demand. When there is no demand, there is no production.

It is designed to minimize the in-process inventory and to have the right material with the right amount at the right location at the right time.

Principles of the Kanban System:

- Only produce products with exactly the same amount that customers consume.
- Only produce products when customers consume.

The production is driven by the *actual* demand from the customer side instead of the *forecasted* demand planned by the staff. The goal is to match production to demand.

Kanban Card

The *Kanban card* is the ticket or signal to authorize the production or movement of materials. It is the message of asking for more. The Kanban card is sent from the end customer up to the chain of production. Upon receiving of a Kanban card, the production station would start to produce goods. The Kanban card can be a physical card or an electronic signal.

Kanban System Example

The simplest example of a Kanban system is the supermarket operation.

- Customers visit the supermarkets and buy what they need.
- The checkout scanners send electronic Kanban cards to the local warehouse asking for more when the items are sold to customers.
- When the warehouse receives the Kanban cards, it starts to replenish the exact goods being sold.

It the warehouse prepares more than what Kanban cards require, the goods would become obsolete. If it prepares less, the supermarket would not have the goods available when customers need them.

Kanban System Benefits

- Minimize in-process inventory
- Free up space occupied by unnecessary inventory
- Prevent overproduction
- Improve responsiveness to dynamic demand
- Avoid the risk of inaccurate demand forecast
- Streamline the production flow
- Visualize the work flow

5.1.3 Poka-Yoke

What is Poka-Yoke?

The Japanese term *poka-yoke* means mistake-proofing. It is a mechanism to eliminate defects as early as possible in the process. It was originally developed by Shigeo Shingo and was initially called "baka-yoke" (fool-proofing). Poka-yoke is an extremely effective control method because it means to prevent, correct, or draw attention to a mistake as it occurs.

Two Types of Poka-Yoke

- Prevention
    - Preventing defects from occurring
    - Removing the possibility that an error could occur
    - Making the occurrence of an error impossible
- Detection
    - Detecting defects once they occur
    - Highlighting defects to draw workers' attention immediately
    - Correcting defects so that they would not reach the next stage

The best way to drive quality is to prevent defects before they occur. The next best solution is to detect defects as they occur so they can be corrected. Poka-yoke solutions can aim to either prevent or detect defects.

Three Methods of Poka-Yoke

Shigeo Shingo identified three methods of poka-yoke for detecting and preventing errors in a mass production system.

1. Contact Method—Use of shape, color, size, or any other physical attributes of the items

2. Constant Number Method—Use of a fixed number to make sure a certain number of motions are completed
3. Sequence Method—Use of a checklist to make sure all the prescribed process steps are followed in the right order

Poka-Yoke Devices

We are surrounded by poka-yoke devices daily. Prevention Devices (Example: The dishwasher does not start to run when the door is open). Detection Devices (Example: The car starts to beep when the passengers do not buckle their seatbelts)

Poka-yoke devices can be in any format that can quickly and effectively prevent or detect mistakes: visual, electrical, mechanical, procedural, human etc.

Other examples:

- Tether on a car's gas cap to keep driver from leaving it behind.
- Inability to remove the key from car ignition until transmission is in park.
- Safety bar on lawn mower must be engaged to start engine.
- Hole near the top of a bathroom sink to keep it from overflowing.
- Clothes dryer stops when the door is opened.

Steps to Apply Poka-Yoke

Step 1: Identify the process steps in need of mistake proofing.
Step 2: Use the 5-why's to analyze the possible mistakes or failures for the process step.
Step 3: Determine the type of poka-yoke: prevention or detection.
Step 4: Determine the method of poka-yoke: contact, constant number, or sequence.
Step 5: Pilot the poka-yoke approach and make any adjustments if needed.
Step 6: Implement poka-yoke in the operating process and maintain the performance.

5.2 STATISTICAL PROCESS CONTROL

5.2.1 DATA COLLECTION FOR SPC

What is SPC?

Statistical process control (SPC) is a statistical method to monitor the performance of a process using control charts in order to keep the process in statistical control. It can be used to distinguish between special cause variation and common cause variation in the process. SPC presents the voice of the process.

Common Cause Variation

Common cause variation (also called chance variation) is the inherent natural variation in any processes. It is the random background noise, which cannot be controlled or eliminated from the process. Its presence in the process is expected and acceptable due to its relatively small influence on the process. The goal of SPC is to distinguish between *common cause* variation and *special cause* variation to minimize tampering or inaction to real process changes.

Special Cause Variation

Special cause variation (also called assignable cause variation) is the unnatural variation in the process. It is the cause of process instability and leads to defects of the products or services. It is also the signal of unanticipated change (either positive or negative) in the process. It is possible to eliminate the special cause variation from the process.

Special cause variation is unnatural variation that has an assignable cause. In some cases, special cause variation is unwanted, but in other cases it can represent something desirable to replicate.

Process Stability

A process is *stable* when:

- There is not any special cause variation involved in the process
- The process is in statistical control
- The future performance of the process is predictable within certain limits
- The changes happening in the process are all due to random inherent variation
- There are not any trends, unnatural patterns, and outliers in the control chart of the process.

When a process performs predictably within certain limits, we say it is "in control." A process that is in control is stable.

SPC Benefits

- Statistical process control can be used in different phases of Six Sigma projects to:
- Understand the stability of a process
- Detect the special cause variation in the process
- Identify the statistical difference between two phases
- Eliminate or apply the unnatural change in the process
- Improve the quality and productivity.

Control Charts

Control charts are graphical tools to present and analyze the process performance in statistical process control. Control charts are used to detect special cause variation and determine

whether the process is in statistical control (stable). They can differentiate *assignable* (special) sources of variation from *common* sources.

Variation solutions:

- Minimize the common cause variation
- Eliminate the special cause variation when it leads to unanticipated negative changes in the outcome
- Implement the special cause variation when it leads to unanticipated positive changes in the outcome.

Control Charts Elements

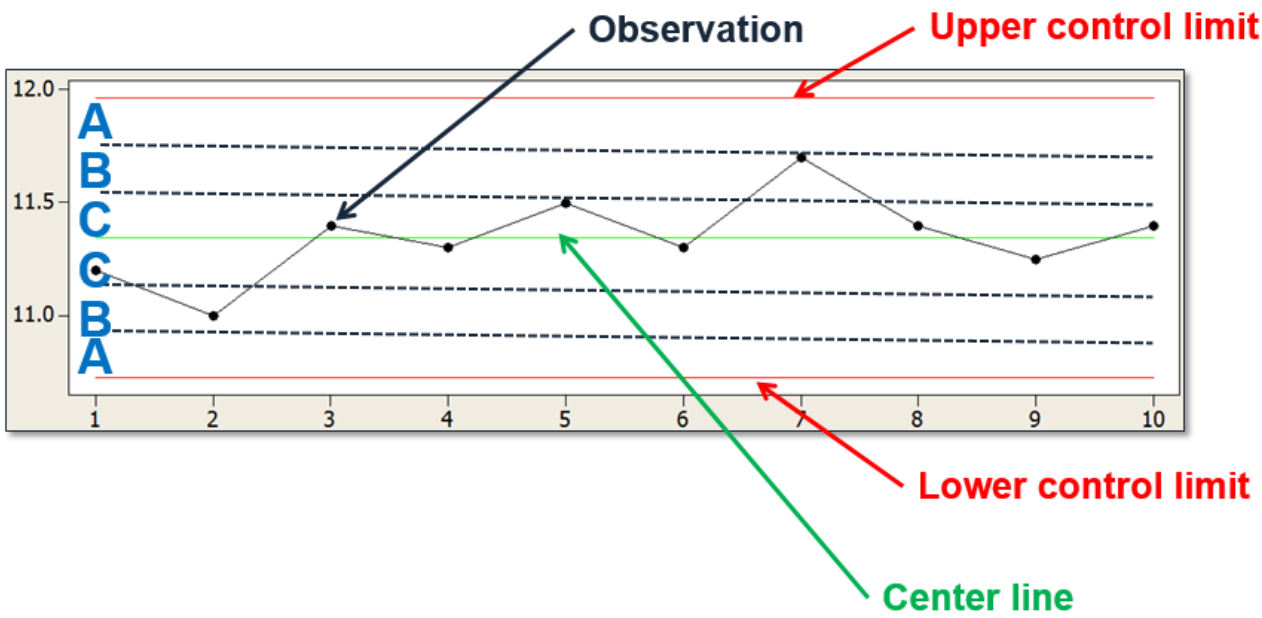


Fig. 5.1 Control Chart Elements

At its basic level, the control chart is a visual tool to illustrate when a process is stable or not. While there are other rules that will tell us whether or not a process is in control, the basic rule is to observe if the data points fall between the upper and lower control limits.

Control Charts Elements

Control charts can work for both continuous data and discrete or count data. Control limits are approximately three-sigma away from the process mean. A process is in statistical control when all the data points on the control charts fall within the control limits and have random patterns only. Otherwise, the process is out of control and we need to investigate the special cause variation in the process.

Possible Errors in SPC

Two types of errors can be made when interpreting and acting (or not acting) on a control chart.

- *Type I (False Positive)*—Tampering, acting to correct variation when you should not, because it is common cause variation.
- *Type II (False Negative)*—Not acting to correct variation when you should, because it is special cause variation.

		Interpretation	
		Common Cause	Special Cause Variation
Truth	Common Cause Variation		Type I Error (False Positive)
	Special Cause Variation	Type II Error (False Negative)	

Fig. 5.2 Types of Errors

A control chart is a form of a hypothesis test where the null hypothesis is that the process is stable, and the alternative is that the process is unstable.

- Null Hypothesis (H_0): The process is stable (i.e., in statistical control).
- Alternative Hypothesis (H_a): The process is unstable (i.e., out of statistical control).

Type I Error

- False positive
- False alarm
- Considering true common cause variation as special cause variation

Type I errors waste resources spent on investigation.

Type II Error

- False negative
- Miss
- Considering true special cause variation as common cause variation

Type II errors neglect the need to investigate critical changes in the process.

Data Collection Considerations

To collect data for plotting control charts, we need to consider:

1. What is the measurement of interest?
2. Are the data discrete or continuous?

3. How many samples do we need?
4. How often do we sample?
5. Where do we sample?
6. What is the sampling strategy?
7. Do we use the raw data collected or transfer them to percentages, proportions, rates, etc.?

Subgroups and Rational Subgrouping

When sampling, we randomly select a group of items (i.e., a subgroup) from the population of interest. The *subgroup size* is the count of samples in a subgroup. It can be constant or variable. Depending on the subgroup sizes, we select different control charts accordingly.

Rational subgrouping is the basic sampling scheme in SPC. The goal of rational subgrouping is to maximize the likelihood of detecting special cause variation. In other words, the control limits should only reflect the variation *between* subgroups.

The numbers of subgroups, subgroup size, and frequency of sampling have great impact on the quality of control charts.

Sampling is randomly selecting a group of items from a population in which all items are produced under the conditions where only random effects are responsible for the observed variation. This maximizes the likelihood of detecting a special cause when it occurs.

Impact of Variation

The rational subgrouping strategy is designed to minimize the opportunity of having special cause variation *within* subgroups. If there is only random variation (background noise) within subgroups, all the special cause variation would be reflected between subgroups. It is easier to detect an out-of-control situation. Random variation is inherent in any process. We are more interested in identifying and taking actions on special cause variation.

Frequency of Sampling

The *frequency of sampling* in SPC depends on whether we have sufficient data to signal the changes in a process with reasonable time and costs. The more frequently we sample, the higher costs we incur. We need the subject matter experts' knowledge on the nature and characteristics of the process to make good decisions on sampling frequency.

5.2.2 I-MR CHART

I-MR Chart

The *I-MR chart* (also called individual-moving range chart or IR chart) is a popular control chart for continuous data with subgroup size equal to one.

- The I chart plots an individual observation as a data point.
- The MR chart plots the absolute value of the difference between two consecutive observations in individual charts as a data point.

If there are *n* data points in the I chart, there are *n –1* data points in the MR chart. The I chart is valid only if the MR chart is in control. The underlying distribution of the I-MR chart is normal distribution.

I Chart Equations

I Chart (Individuals Chart)

Data Point: x_i

Center Line: $\frac{\sum_{i=1}^{n} x_i}{n}$

Control Limits: $\frac{\sum_{i=1}^{n} x_i}{n} \pm 2.66 \times \overline{MR}$

Where: *n* is the number of observations.

MR-Chart Equations

MR Chart (Moving Range Chart)

Data Point: $|x_{i+1} - x_i|$

Center Line: $\frac{\sum |x_{i+1} - x_i|}{n-1}$

Upper Control Limit: $3.267 \times \frac{\sum |x_{i+1} - x_i|}{n-1}$

Lower Control Limit: 0

Where: *n* is the number of observations.

Use JMP to Plot I-MR Charts

 ***Data File: "IR.jmp"***

Steps to plot IR charts in JMP:

1. JMP Command: Analyze -> Quality & Process -> Control Chart -> IR
2. Select “Measurement” as “Process”
3. Click “OK”
4. Click the red triangle next to “Individual Measurement of Measurement”
5. Select “Tests”-> “All Tests”

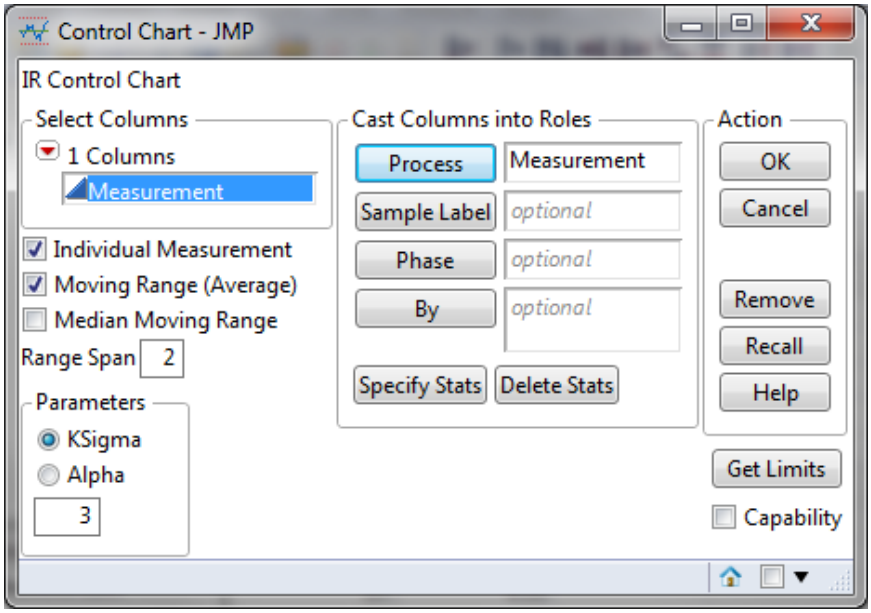


Fig. 5.3 Control Chart dialog box with selections

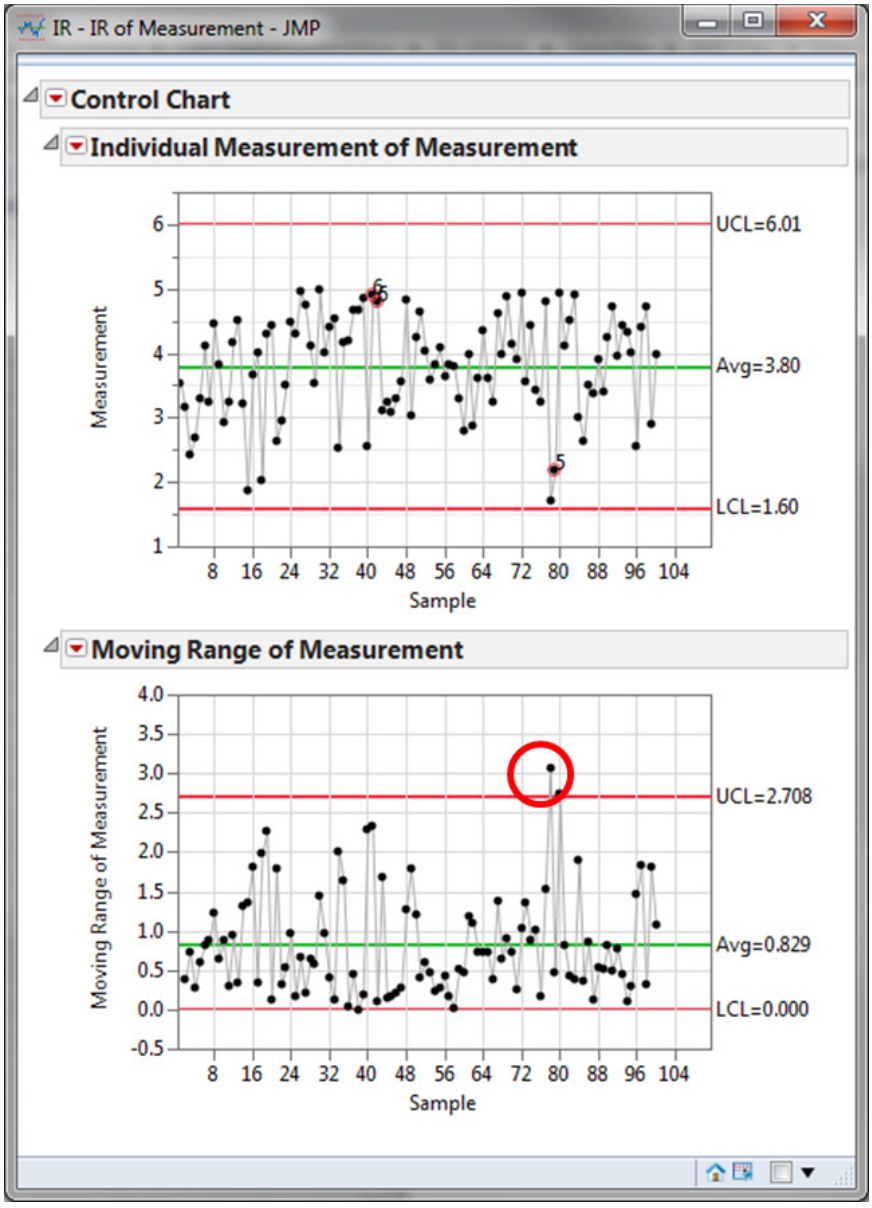


Fig. 5.4 I-MR Charts Diagnosis

I Chart (Individuals' Chart): Since the MR chart is out of control, the I chart is invalid. This is because the control limits on the 'I' chart are calculated using the mean of the Moving Range. If the moving range chart is out of control then the control limits on the 'I'chart can't be trusted.

MR Chart (Moving Range Chart): Two data points fall beyond the upper control limit. This indicates the MR chart is out of control (i.e., the variations between every two contiguous individual samples are not stable over time). We need to further investigate the process, identify the root causes that trigger the outliers, and correct them to bring the process back in control.

5.2.3 XBAR-R CHART

Xbar-R Chart

The *Xbar-R chart* is a control chart for continuous data with a constant subgroup size between two and ten.

- The Xbar chart plots the average of a subgroup as a data point.
- The R chart plots the difference between the highest and lowest values within a subgroup as a data point.

The Xbar chart monitors the process mean and the R chart monitors the variation within subgroups. The Xbar is valid only if the R chart is in control.

The underlying distribution of the Xbar-R chart is normal distribution.

Xbar Chart Equations

Xbar chart

Data Point: $\overline{X_i} = \frac{\sum_{j=1}^{m} x_{ij}}{m}$

Center Line: $\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k}$

Control Limits: $\overline{\overline{X}} \pm A_2 \overline{R}$

Where:

- *m* is the subgroup size
- *k* is the number of subgroups.

A_2 is a constant depending on the subgroup size.

R Chart Equations

R chart (Rage Chart)

Data Point: $R_i = \underset{j \in [1,m]}{Max}(x_{ij}) - \underset{j \in [1,m]}{Min}(x_{ij})$

Center Line: $\overline{R} = \frac{\sum_{i=1}^{k} R_i}{k}$

Upper Control Limit: $D_4 \times \overline{R}$

Lower Control Limit: $D_3 \times \overline{R}$

Where:

- *m* is the subgroup size and *k* is the number of subgroups
- D_3 and D_4 are constants depending on the subgroup size.

Use JMP to Plot Xbar-R Charts

Data File: "XbarR.jmp"

Steps to plot Xbar-R charts in JMP:

1. Analyze -> Quality & Process -> Control Chart -> Xbar
2. Select Measurement in Process
3. Select Subgroup ID in Sample Label
4. Be sure the check Xbar and R check boxes
5. Click OK

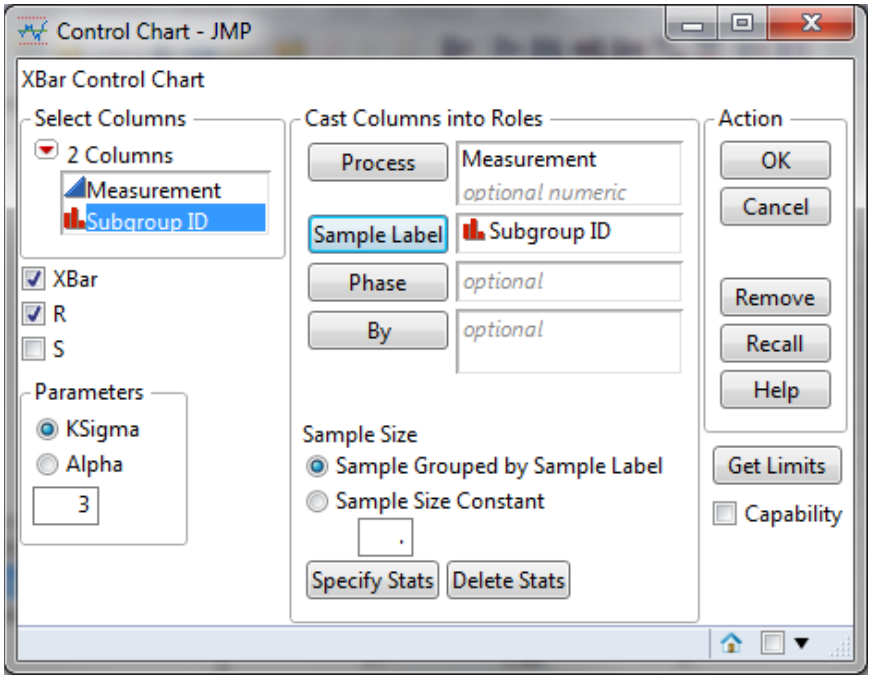


Fig. 5.5 Control Chart selections

Xbar-R Charts Diagnosis

Xbar-R Charts:
Since the R chart is in control, the Xbar chart is valid. In both charts, there are not any data points failing any tests for special causes (i.e., all the data points fall between the control limits and spread around the center line with a random pattern). We conclude that the process is in control.

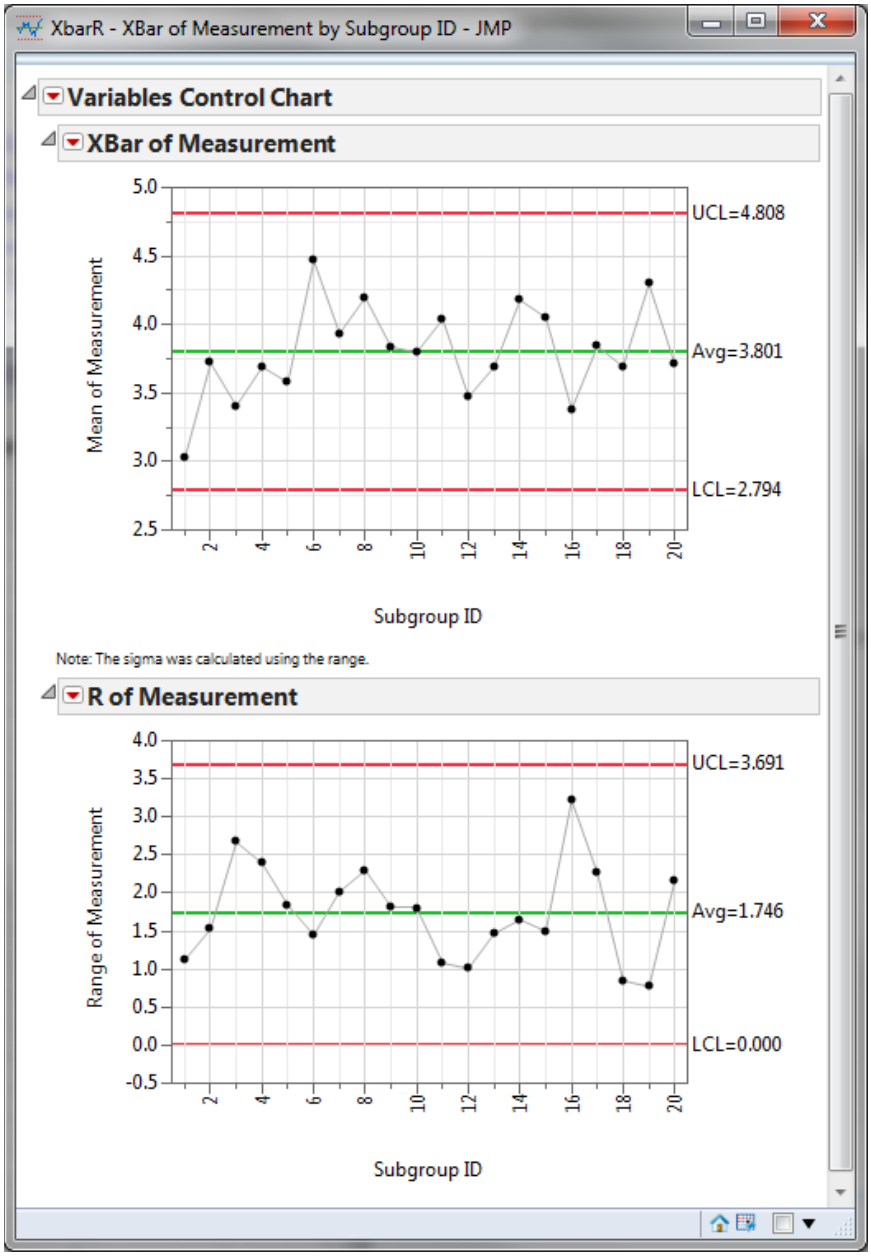


Fig. 5.6 Xbar-R Chart Diagnosis

5.2.4 U CHART

The *U chart* is a type of control chart used to monitor discrete (count) data where the sample size is greater than one, typically the average number of defects per unit.

Defect vs. Defective

Remember the difference between defect and defective?

A *defect* of a unit is the unit's characteristic that does not meet the customers' requirements.

A *defective* is a unit that is not acceptable to the customers.

One defective might have multiple defects. One unit might have multiple defects but still be usable to the customers.

U Chart

The *U chart* is a control chart monitoring the average defects per unit. It plots the count of defects per unit of a subgroup as a data point. It considers the situation when the subgroup size of inspected units for which the defects would be counted is not constant. The underlying distribution of the U chart is Poisson distribution.

U Chart Equations

U chart

Data Point: $u_i = \frac{x_i}{n_i}$

Center Line: $\bar{u} = \frac{\sum_{i=1}^{k} u_i}{k}$

Control Limits: $\bar{u} \pm 3 \times \sqrt{\frac{\bar{u}}{n_i}}$

Where:

- n_i is the subgroup size for the i^{th} subgroup
- k is the number of subgroups
- x_i is the number of defects in the i^{th} subgroup.

Use JMP to Plot a U Chart

 ***Data File: "U.jmp"***

Steps to plot a U chart in JMP:

1. Analyze -> Quality & Process -> Control Chart -> U
2. Select Count of Defects in Process
3. Select Count of Units Inspected in Unit Size
4. Click OK

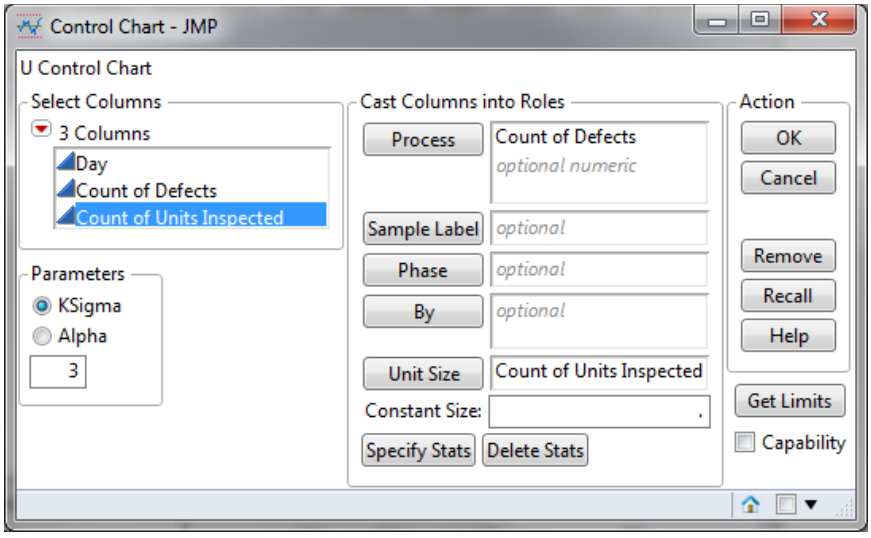


Fig. 5.7 Control Chart Selections

U Chart Diagnosis

Since the sample sizes are not constant over time, the control limits are adjusted to different values accordingly. The data point circled in above falls beyond the upper control limit. We conclude that the process is out of control. Further investigation is needed to determine the special causes that triggered the unnatural pattern of the process.

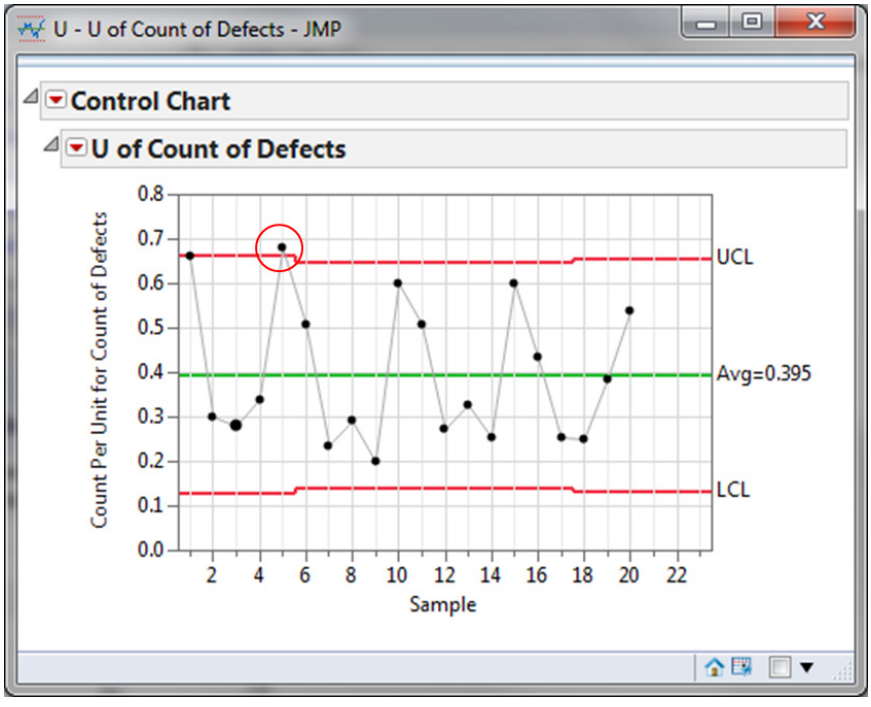


Fig. 5.8 U Chart analysis

5.2.5 P CHART

P Chart

The *P chart* is a control chart monitoring the percentages of defectives. The P chart plots the percentage of defectives in one subgroup as a data point. It considers the situation when the subgroup size of inspected units is not constant. The underlying distribution of the P chart is binomial distribution.

P Chart Equations

P chart

Data Point: $p_i = \frac{x_i}{n_i}$

Center Line: $\bar{p} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}$

Control Limits: $\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}$

Where:

- n_i is the subgroup size for the i^{th} subgroup
- k is the number of subgroups
- x_i is the number of defectives in the i^{th} subgroup.

Use JMP to Plot a P Chart

Data File: "P.jmp"

Steps to plot a P chart in JMP:

1. Analyze -> Quality & Process -> Control Chart -> P
2. Select Fail in the Process Field
3. Select N in the Sample Size Field

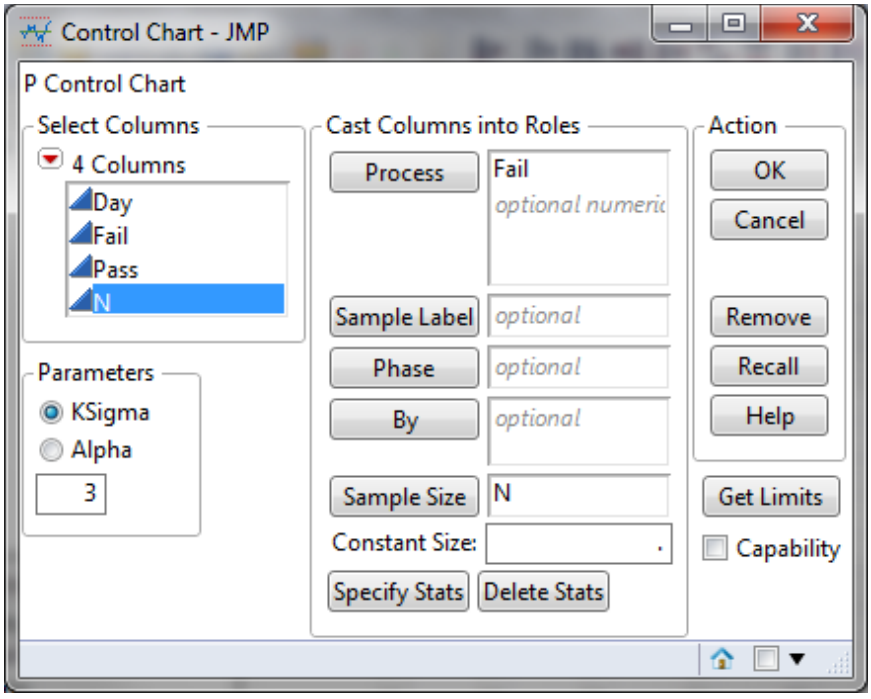


Fig. 5.9 Control Chart Selections

4. Click Ok

P Chart Diagnosis

Since the sample sizes are not constant over time, the control limits are adjusted to different values accordingly. All the data points fall within the control limits and spread randomly around the mean. We conclude that the process is in control.

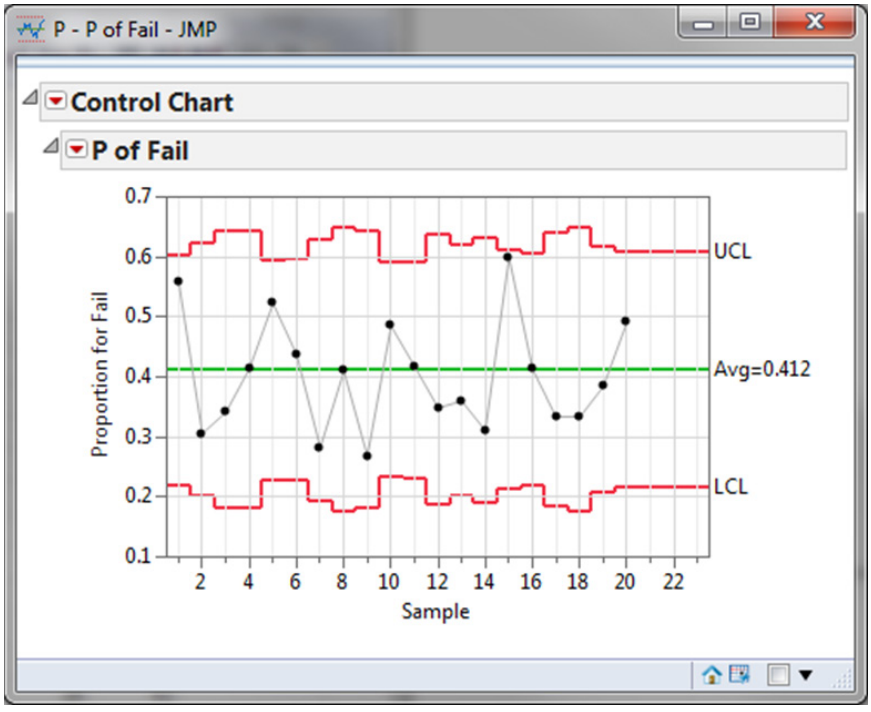


Fig. 5.10 P Chart output

5.2.6 NP CHART

NP Chart

The *NP chart* is a control chart monitoring the count of defectives. It plots the number of defectives in one subgroup as a data point. The subgroup size of the NP chart is constant. The underlying distribution of the NP chart is binomial distribution.

NP Chart Equations

NP chart

Data Point: x_i

Center Line: $\overline{np} = \frac{\sum_{i=1}^{k} x_i}{k}$

Control Limits: $\overline{np} \pm 3 \times \sqrt{\overline{np}(1-\overline{p})}$

Where:

- n is the constant subgroup size
- k is the number of subgroups
- x_i is the number of defectives in the i^{th} subgroup.

Use JMP to Plot an NP Chart

Data File: "NP.jmp"

Steps to plot a NP chart in JMP:

1. Analyze -> Quality & Process -> Control Chart -> NP
2. Select Fail in the Process Field
3. Select N in the Sample Size Field

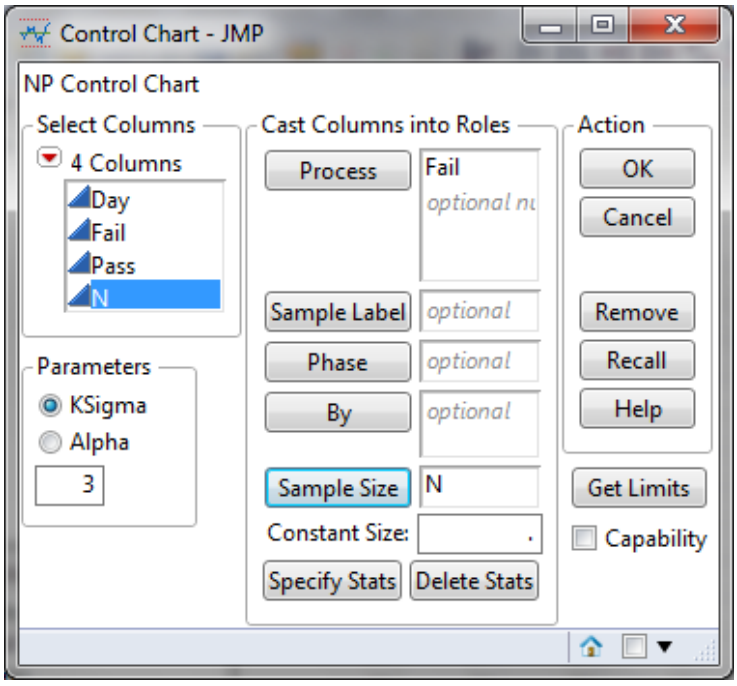


Fig. 5.11 Control Chart selections

4. Click Ok

NP Chart Diagnosis

Four data points, circled in red, fall beyond the upper control limit. We conclude that the NP chart is out of control. Further investigation is needed to determine the special causes that triggered the unnatural pattern of the process.

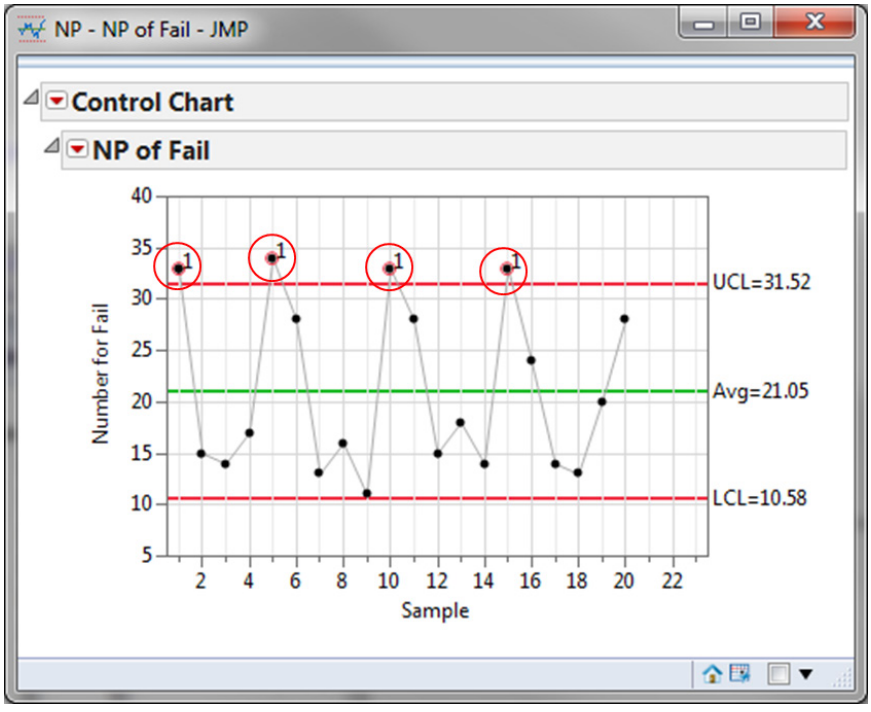


Fig. 5.12 NP Chart output

5.2.7 X-S CHART

X-S Chart

The *X-S chart* (also called Xbar-S chart) is a control chart for continuous data with a constant subgroup size greater than ten. The Xbar chart plots the average of a subgroup as a data point. The S chart plots the standard deviation within a subgroup as a data point.

The Xbar chart monitors the process mean and the S chart monitors the variability within subgroups. The Xbar is valid only if the S chart is in control. The underlying distribution of the Xbar-S chart is normal distribution.

Xbar Chart Equations

Xbar Chart

Data Point: $\overline{X_i} = \frac{\sum_{j=1}^{m} x_{ij}}{m}$

Center Line: $\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k}$

Control Limits: $\overline{\overline{X}} \pm A_3 \overline{s}$

Where:

- m is the subgroup size
- k is the number of subgroups
- A_3 is a constant depending on the subgroup size.

S Chart Equations

S chart

Data Point: $s_i = \sqrt{\frac{\sum_{j=1}^{m} (x_{ij} - \overline{x}_i)^2}{m-1}}$

Center Line: $\bar{s} = \frac{\sum_{i=1}^{k} s_i}{k}$

Upper Control Limit: $B_4 \times \bar{s}$

Lower Control Limit: $B_3 \times \bar{s}$

Where:

- m is the subgroup size
- k is the number of subgroups
- B_3 and B_4 are constants depending on the subgroup size.

Use JMP to Plot Xbar-S Charts

Data File: "XbarS.jmp"

Steps to plot Xbar-S charts in JMP:

1. Analyze -> Quality & Process -> Control Chart -> XBar
2. Select Measurement in the Process Field
3. Select Subgroup ID in the Sample Label Field
4. Be sure the check the Xbar and S check boxes

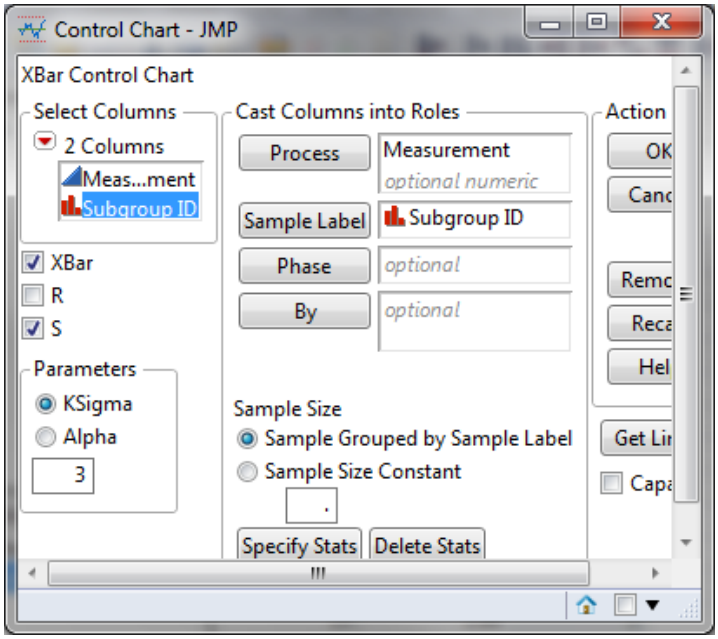


Fig. 5.13 Control Chart selections

5. Click Ok

Xbar-S Charts Diagnosis

Since the S chart is in control, the Xbar chart is valid. In both charts, there are not any data points failing any tests for special causes (i.e., all the data points fall between the control limits and spread around the center line with a random pattern). We conclude that the process is in control.

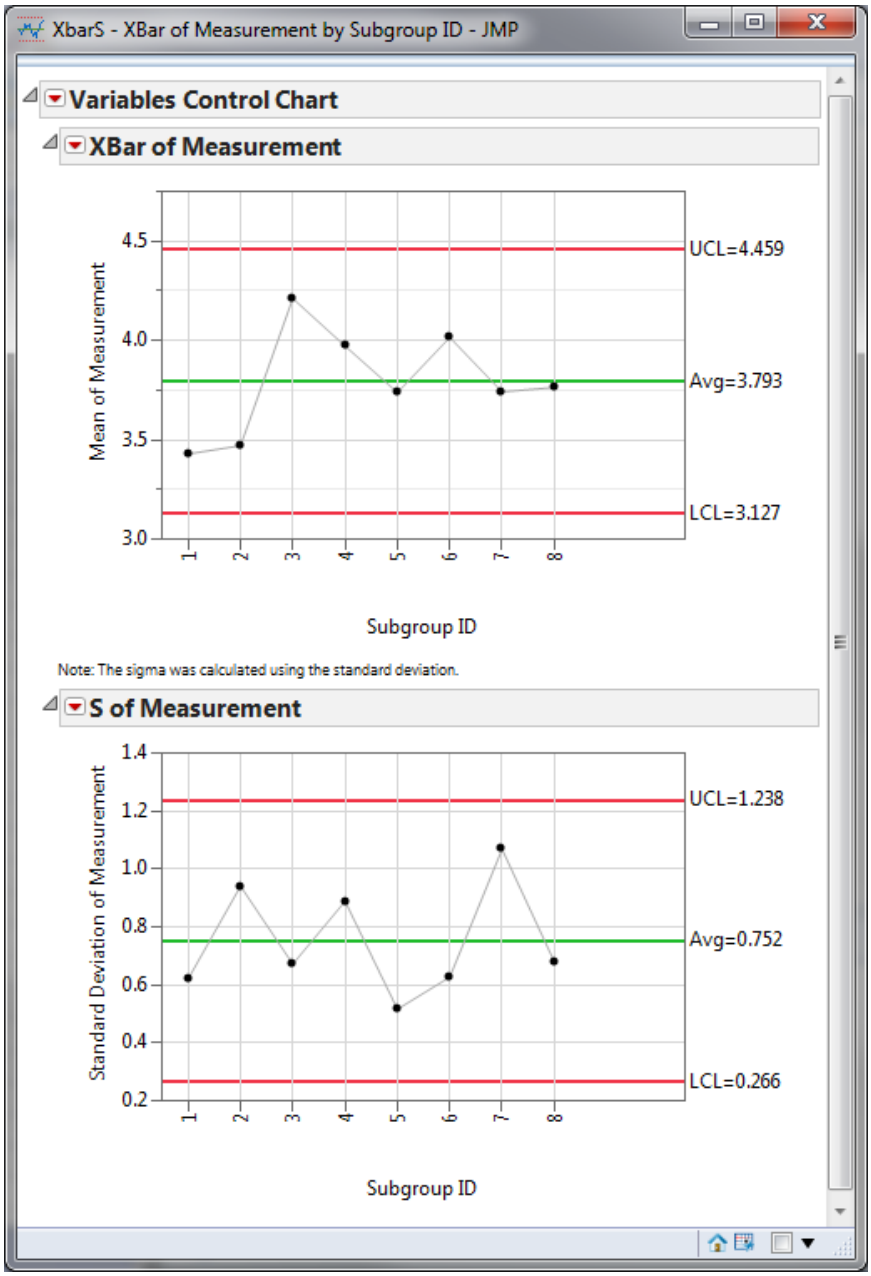


Fig. 5.14 Xbar-S Chart output

5.2.8 CumSum Chart

CumSum Chart

The *CumSum chart* (also called cumulative sum control chart or CUSUM chart) is a control chart of monitoring the cumulative sum of the subgroup mean deviations from the process target. It detects the shift of the process mean from the process target over time. The underlying distribution of the CumSum chart is normal distribution.

There are two types of CumSum charts:

1. One two-sided CumSum charts
2. Two one-sided CumSum charts

Two-Sided CumSum

In two-sided CumSum charts, we use a V-mask to identify out-of-control data points. The V-mask is a transparent overlay shape of a horizontal "V" applied on the top of the CumSum chart. Its origin point is placed on the last data point of the CumSum chart and its center line is horizontal. If all of the data points stay inside the V-mask, we consider the process is in statistical control.

V-Mask

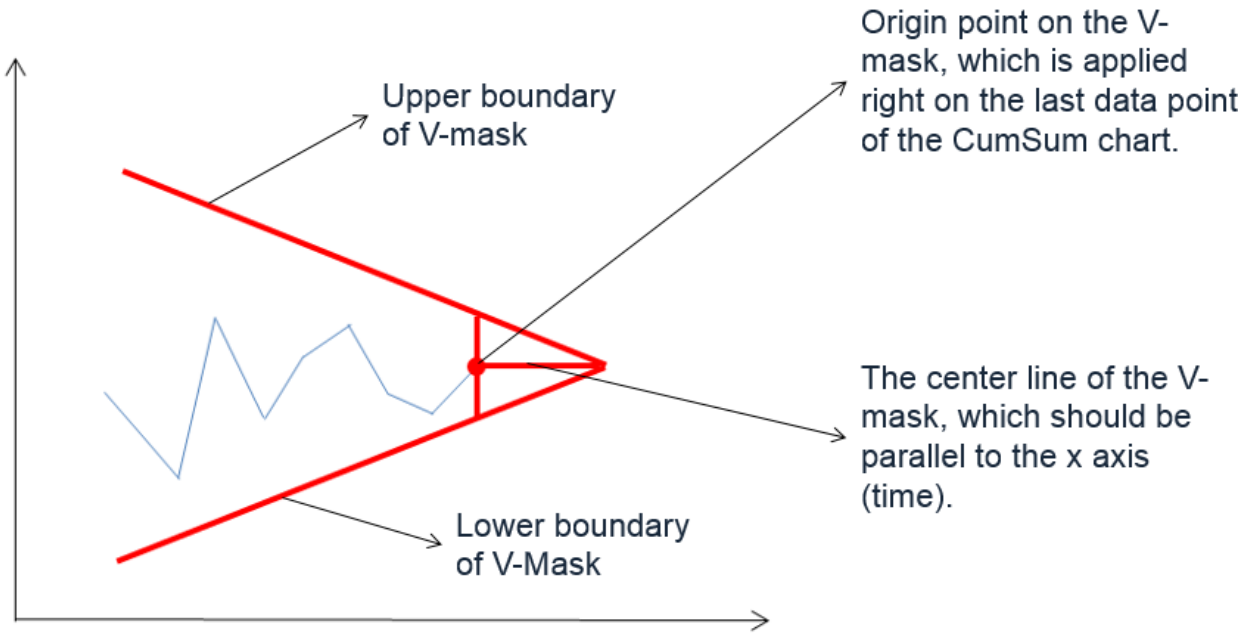


Fig. 5.15 V-Mask Explanation

The origin of the V-mask is the most recent data point, and the arms extend backwards. The process is out of control if any of the data points fall above or below the limits.

Two-Sided CumSum Equations

Two-Sided CumSum

Data Point: $\begin{cases} c_i = c_{i-1} + (\bar{x}_i - T) & i > 0 \\ c_i = 0 & i = 0 \end{cases}$

V-Mask Slope: $k\frac{\sigma}{\sqrt{m}}$

V-Mask Width at the Origin Point: $2h\frac{\sigma}{\sqrt{m}}$

Where:

- $\bar{x}_i$ is the mean of the i[th] subgroup
- T is the process target
- σ is the estimation of process standard deviation
- m is the subgroup size.

One-Sided CumSum

We can also use two one-sided CumSum charts to detect the shift of the process mean from the process target. The upper one-sided CumSum detects the upward shifts of the process mean. The lower one-sided CumSum detects the downward shifts of the process mean.

One-Sided CumSum Equations

One-Sided CumSum

Data Point: $\begin{cases} c_i = c_{i-1} + (\bar{x}_i - T) & i > 0 \\ c_i = 0 & i = 0 \end{cases}$

Center Line: T

Upper Control Limit: $c_i^+ = \max\left(0, \bar{x}_i - (T + k) + c_{i-1}^+\right)$

Lower Control Limit: $c_i^- = \max\left(0, (T - k) - \bar{x}_i + c_{i-1}^-\right)$

Where:

- $\bar{x}_i$ is the mean of the i[th] subgroup
- T is the process target
- k is the slope of the lower boundary of the V-mask.

Use JMP to Plot a CumSum Chart

Data File: "CumSum.jmp"

Steps in JMP to plot CumSum charts:

1. Analyze -> Quality & Process -> Control Chart -> CuSum
2. Select Weight in the Process Field
3. Select hour in the Sample Label Field
4. Be sure the check “H” radio button
5. Enter 2 for the H parameter
6. Specify Stats
   a) Enter 8.1 for Target
   b) Enter 1 for Delta
   c) Enter 0.05 for Sigma

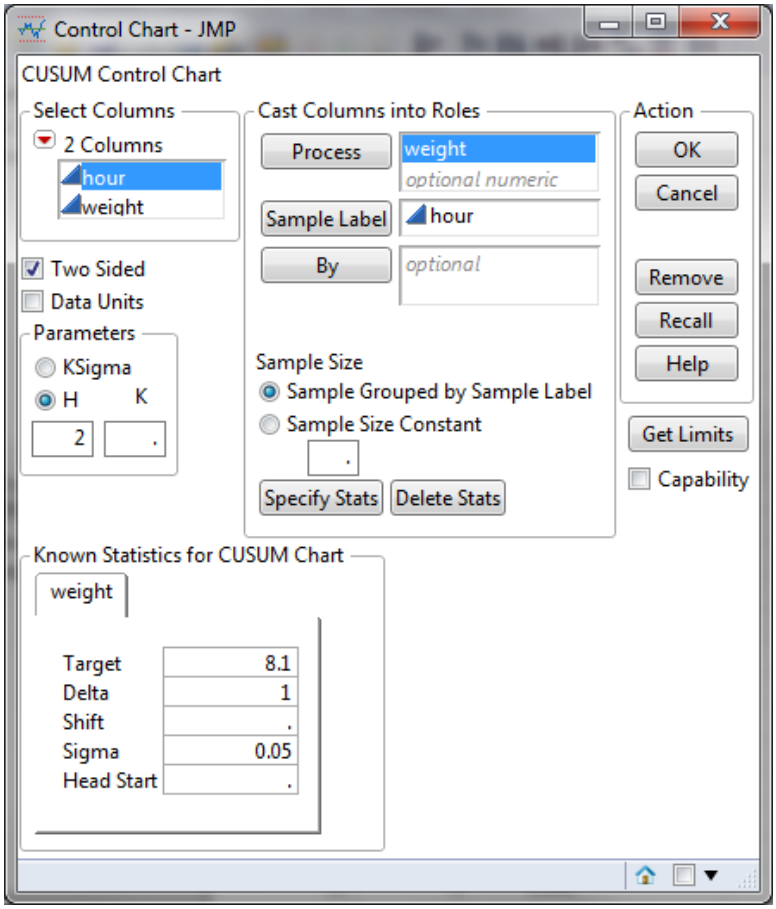


Fig. 5.16 Control Chart selections

7. Click Ok

CumSum Chart Diagnosis

The process is in control since all of the data points fall between the two arms of the V-mask.

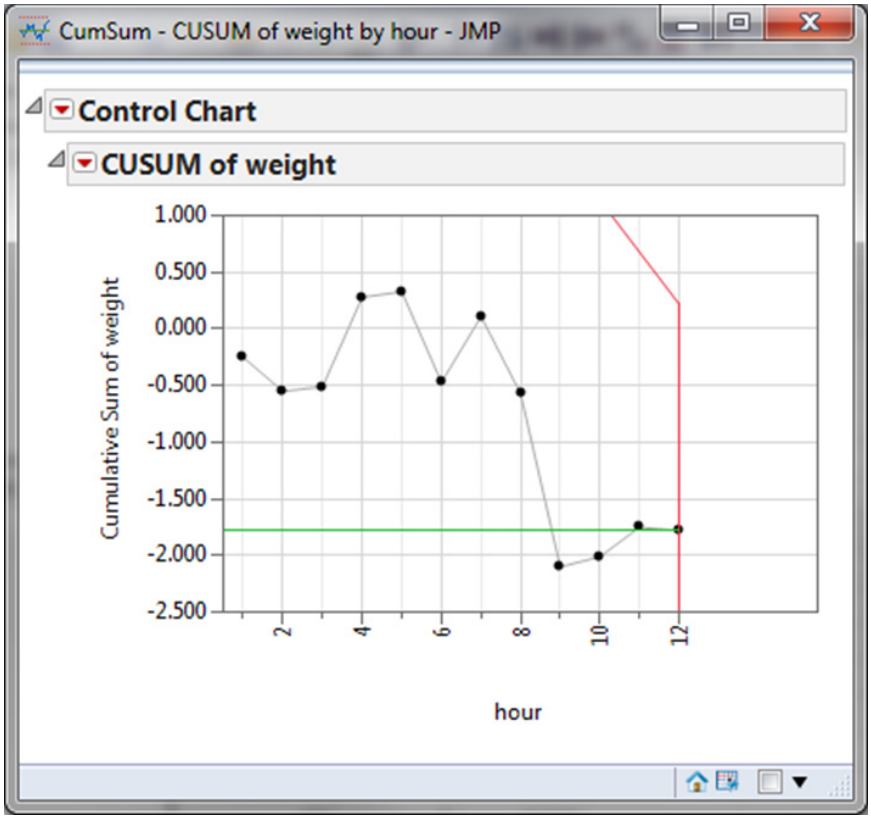


Fig. 5.17 CumSum chart analysis

5.2.9 EWMA CHART

EWMA Chart

The *EWMA chart* (Exponentially-Weighted Moving Average Chart) is a control chart monitoring the exponentially-weighted average of previous and present subgroup means. The more recent data get more weight than older data. It detects the shift of the process mean from the process target over time. The underlying distribution of the EWMA chart is normal distribution.

EWMA Chart Equations

EWMA Chart

Data Point: $z_i = \lambda \bar{x}_i + (1-\lambda) z_{i-1}$ where $0 < \lambda < 1$

Center Line: $\overline{X}$

Control Limits: $\overline{X} \pm k \cdot \frac{s}{\sqrt{n}} \sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left[1-(1-\lambda)^{2i}\right]}$

Where:

- $\bar{x}_i$ is the mean of the i[th] subgroup
- λ and k are user-defined parameters to calculate the EWMA data points and the control limits.

Use JMP to Plot an EWMA Chart

Data File: "EWMA.jmp"

Steps in JMP to plot EWMA charts:

1. Analyze -> Quality & Process -> Control Chart -> EWMA
2. Select Gap in the Process Field
3. Enter 0.5 in the Weight Field
4. Enter 5 as the Sample Size Constant

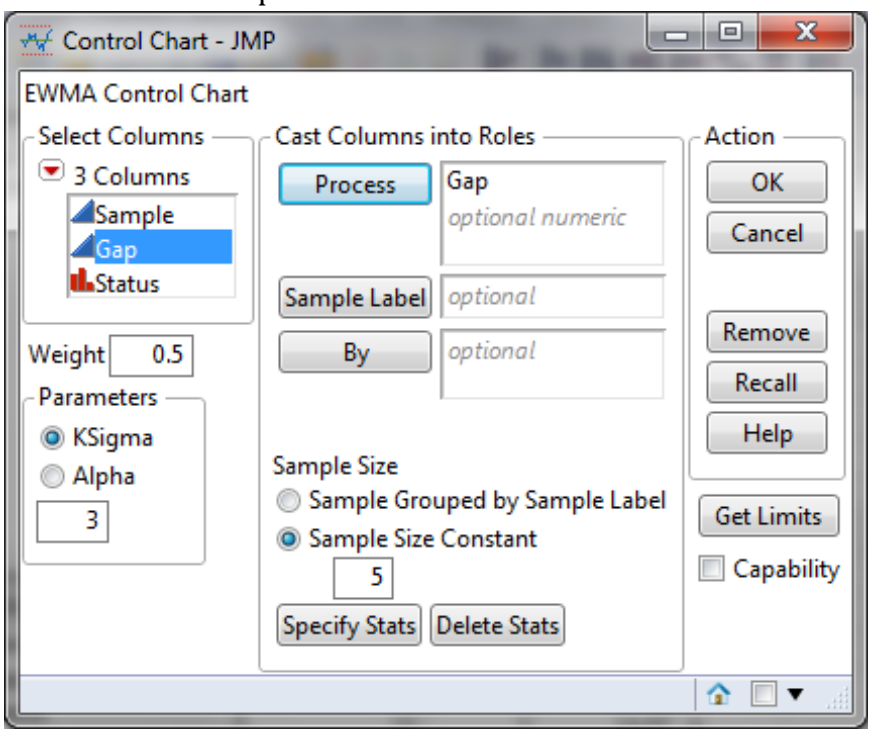


Fig. 5.18 Control Chart selections

5. Click OK

EWMA Chart Diagnosis

One data point falls beyond the upper control limit and we conclude that the process is out of control. Further investigation is needed to discover the root cause for the outlier.

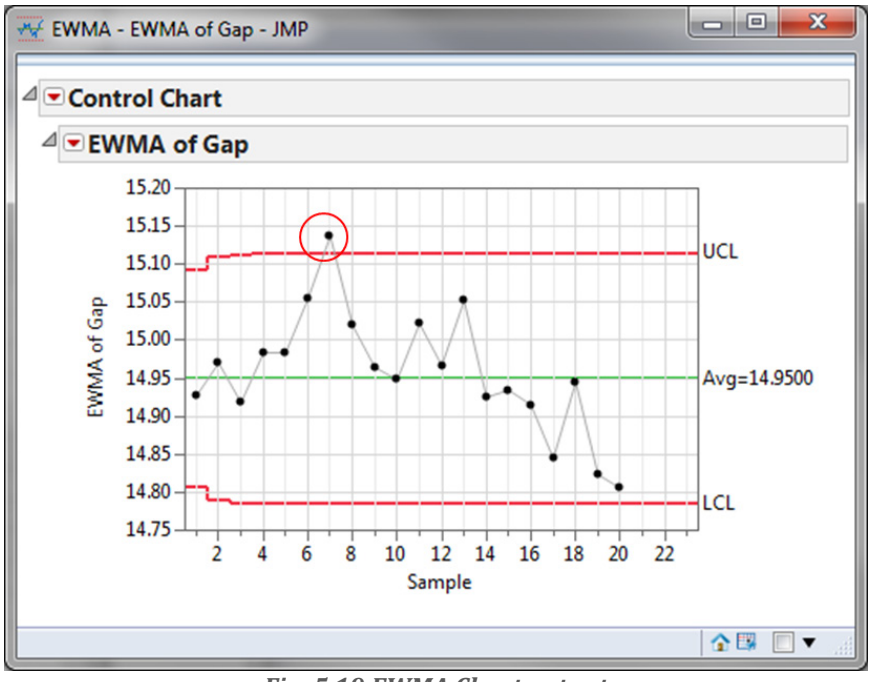


Fig. 5.19 EWMA Chart output

5.2.10 CONTROL METHODS

Control Methods

The fundamental reason for the Control phase is to sustain improvements made by the project and ensure that the process remains stable with minimal variation. There are many control methods available to keep the process stable and minimize the variation. The most common control methods are:

- SPC (statistical process control)—using control charts to monitor process quality over time
- 5S method—to keep the workplace organized and efficient
- Kanban—for visual inventory management
- Poka-Yoke (mistake proofing)

SPC

SPC (Statistical Process Control) is a quantitative control method to monitor the stability of the process performance by identifying the special cause variation in the process. It uses control charts to detect the unanticipated changes in the process.

Which control chart to use depends on:

- Whether the data are continuous or discrete
- How large the subgroup size is
- Whether the subgroup size is constant

- Whether we are interested in measuring defects or defectives
- Whether we are interested in detecting the shifts in the process mean

5S

The *5S* methodology is a systematic approach of cleaning and organizing the workplace through:

1. Seiri (sorting)
2. Seiton (straightening)
3. Seiso (shining)
4. Seiketsu (standardizing)
5. Shisuke (sustaining)

In other words, 5S describes how to organize a workspace for efficiency and effectiveness by identifying and sorting tools and materials used for a process, and maintaining and sustaining the order of the area and items.

Kanban

A *Kanban* system is a demand-driven system. The customer demand is the signal to trigger or pull the production. Products are made only to meet the immediate demand. When there is no demand, there is no production. Kanban was designed to reduce the waste in inventory and increase the speed of responding to immediate demand.

Poka-Yoke

Poka-yoke means "mistake proofing" and is a mechanism to eliminate the defects as early as possible in the process.

Contact Method
The contact method uses physical characteristics (e.g., shape, color, size) to identify defects.

Constant Number Method
The constant number method alerts an operator about a defect when a certain number of motions are not completed.

Sequence Method
The sequence method (or motion-step) identifies a defect when the steps of a process are not performed correctly or in the right order.

5.2.11 CONTROL CHART ANATOMY

Control Chart Calculations Summary

Chart	Center Line	Control Limits	σ_x
I Chart	$\frac{\sum_{i=1}^{n} x_i}{n}$	$\frac{\sum_{i=1}^{n} x_i}{n} \pm 3 \times \frac{\overline{MR}}{d_2}$	$\frac{\overline{MR}}{d_2}$
MR Chart	$\overline{MR} = \frac{\sum_{i=1}^{n-1} \lvert x_{i+1} - x_i \rvert}{n-1}$	$UCL = D_4 \times \overline{MR}$ $LCL = D_3 \times \overline{MR}$	
Xbar Chart (Xbar-R)	$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k}$	$\overline{\overline{X}} \pm A_2 \overline{R}$	$\frac{\overline{R}}{d_2}$
Xbar Chart (Xbar-S)	$\overline{\overline{X}} = \frac{\sum_{i=1}^{k} \overline{X}_i}{k}$	$\overline{\overline{X}} \pm A_3 \overline{s}$	$\frac{\overline{s}}{c_4}$
R Chart	$\overline{R} = \frac{\sum_{i=1}^{k} R_i}{k}$	$UCL = D_4 \times \overline{R}$ $LCL = D_3 \times \overline{R}$	
S Chart	$\overline{s} = \frac{\sum_{i=1}^{k} s_i}{k}$	$UCL = B_4 \times \overline{s}$ $LCL = B_3 \times \overline{s}$	
U Chart	$\overline{u} = \frac{\sum_{i=1}^{k} u_i}{k}$	$\overline{u} \pm 3 \times \sqrt{\frac{\overline{u}}{n_i}}$	$\sqrt{\frac{\overline{u}}{n_i}}$
P Chart	$\overline{p} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}$	$\overline{p} \pm 3 \sqrt{\frac{\overline{p}(1-\overline{p})}{n_i}}$	$\sqrt{\frac{\overline{p}(1-\overline{p})}{n_i}}$
NP Chart	$n\overline{p} = \frac{\sum_{i=1}^{k} x_i}{k}$	$n\overline{p} \pm 3 \times \sqrt{n\overline{p}(1-\overline{p})}$	$\sqrt{n\overline{p}(1-\overline{p})}$

Table 5.1 Control Chart Calculations

This is a helpful reference page to understand how the parameters of various control charts are calculated.

Control Chart Constants

Subgroup Size	A2	A3	B3	B4	c4	d2	D3	D4
2	1.88	2.659	-	3.267	0.7979	1.128	-	3.267
3	1.023	1.954	-	2.568	0.8862	1.693	-	2.574
4	0.729	1.628	-	2.266	0.9213	2.059	-	2.282
5	0.577	1.427	-	2.089	0.94	2.326	-	2.114
6	0.483	1.287	0.03	1.97	0.9515	2.534	-	2.004
7	0.419	1.182	0.118	1.882	0.9594	2.704	0.076	1.924
8	0.373	1.099	0.185	1.815	0.965	2.847	0.136	1.864
9	0.337	1.032	0.239	1.761	0.9693	2.97	0.184	1.816
10	0.308	0.975	0.284	1.716	0.9727	3.078	0.223	1.777
15	0.223	0.789	0.428	1.572	0.9823	3.472	0.347	1.653
25	0.153	0.606	0.565	1.435	0.9896	3.931	0.459	1.541

Table 5.2 Control Chart Constants

This is another helpful reference page. Many control charts use some of these constants in the calculation of parameters. They vary depending on the subgroup size.

Unnatural Patterns

If there are *unnatural patterns* in the control chart of a process, we consider the process out of statistical control. Control charts are designed to help us decipher special cause variation from common cause variation. Typical unnatural patterns in control charts are:

- Outliers (data points that fall outside of the upper or lower control limits)
- Trending
- Cycling
- Auto-correlative
- Mixture

A process is *in control* if all the data points on the control chart are randomly spread out within the control limits.

Western Electric Rules

Western Electric Rules are the most popular decision rules to detect unnatural patterns in the control charts. They are a group of tests for special causes in a process. The Western Electric Rules were developed by the manufacturing division of the Western Electric Company in 1956.

The area between the upper and lower control limits is separated into six subzones defined by the number of standard deviations from the center line.

- Zone A: between two and three standard deviations from the center line
- Zone B: between one and two standard deviations from the center line
- Zone C: within one standard deviation from the center line.

Western Electric Rules

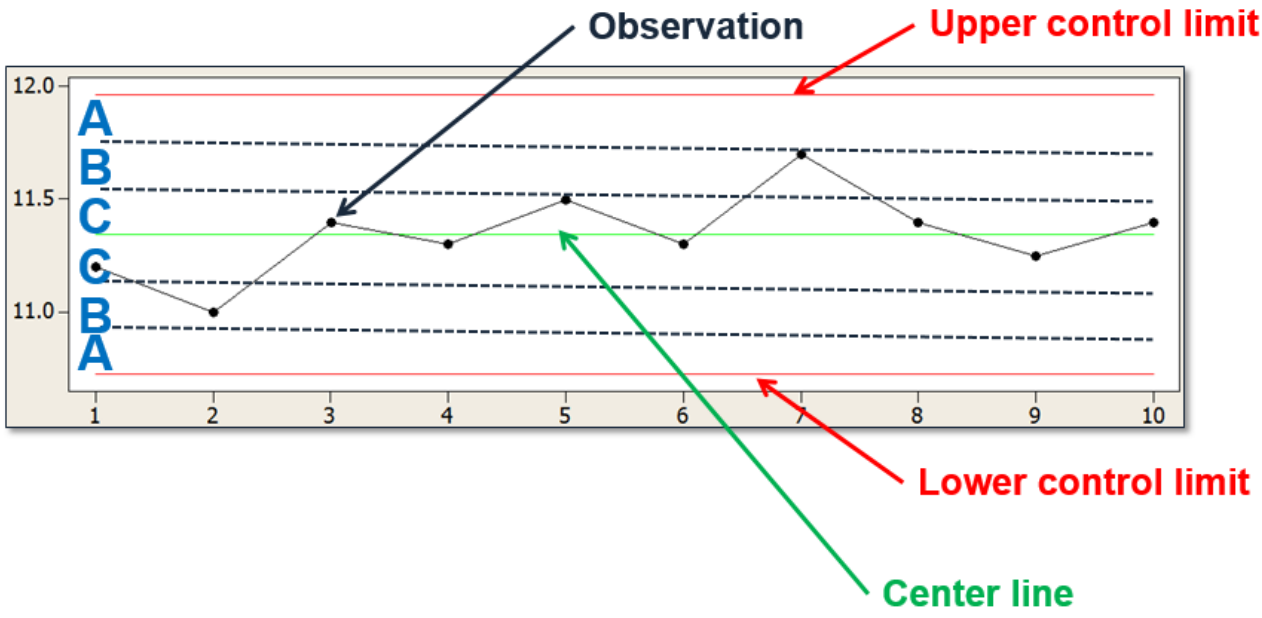


Fig. 5.20 Western Electric Rules

- Zone A: three sigma zone
- Zone B: two sigma zone
- Zone C: one sigma zone

If a data point falls onto the dividing line of two consecutive zones, the point belongs to the outer zone.

Test 1: One point more than three standard deviations from the center line (i.e., one point beyond zone A). Such a point is typically known as an outlier.

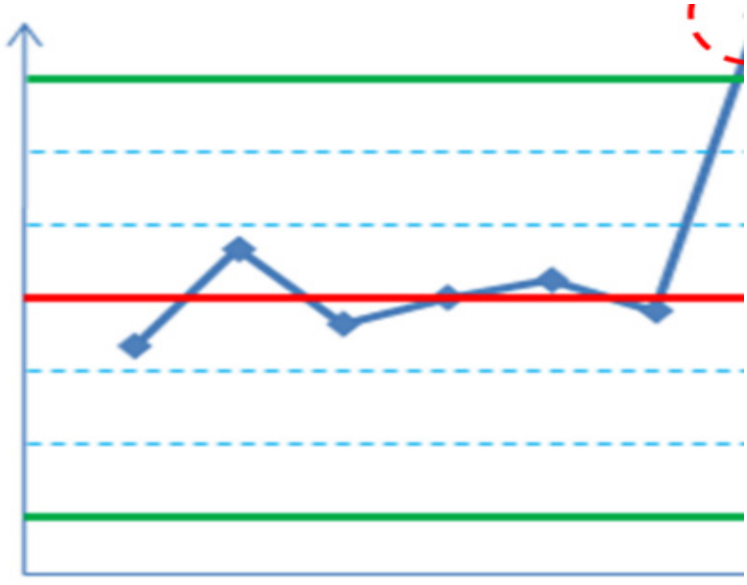


Fig. 5.21 Western Electric Test 1

Test 2: Nine points in a row on the same side of the center line.

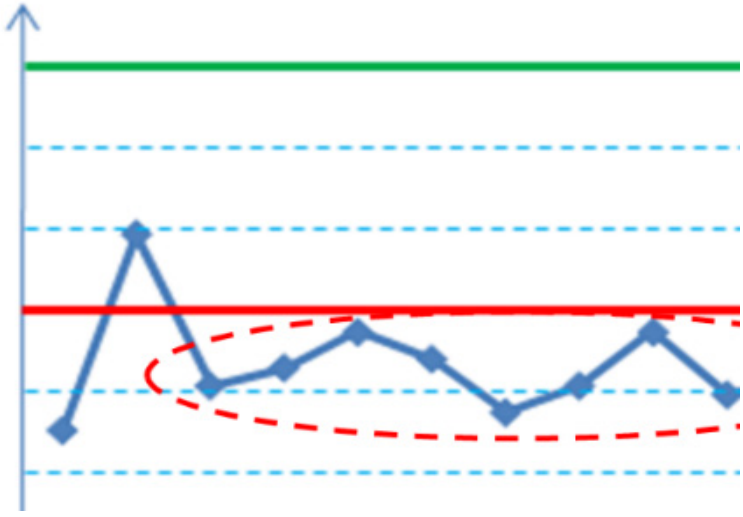


Fig. 5.22 Western Electric Test 2

Test 2 identifies situations where the process mean has temporarily shifts in the process.

Test 3: Six points in a row steadily increasing or steadily decreasing. Test 3 identifies significant trends in performance.

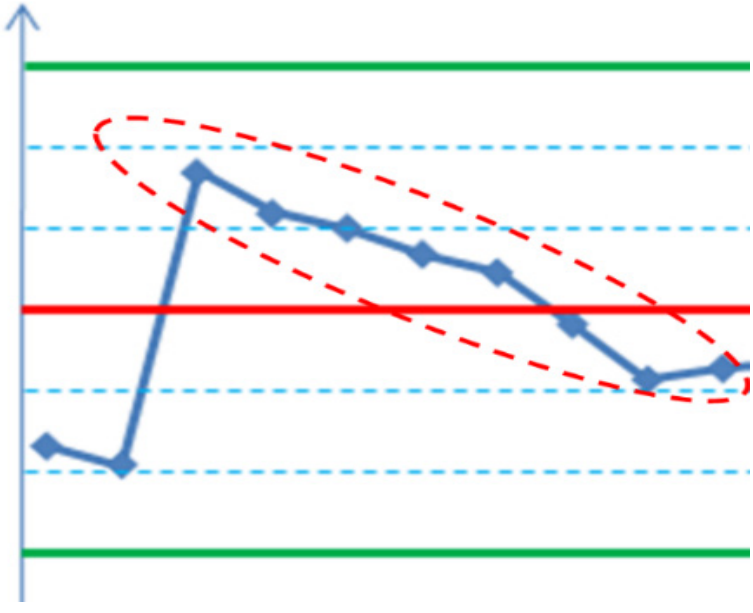


Fig. 5.23 Western Electric Test 3

Test 4: Fourteen points in a row alternating up and down. Test 4 indicates a cycle.

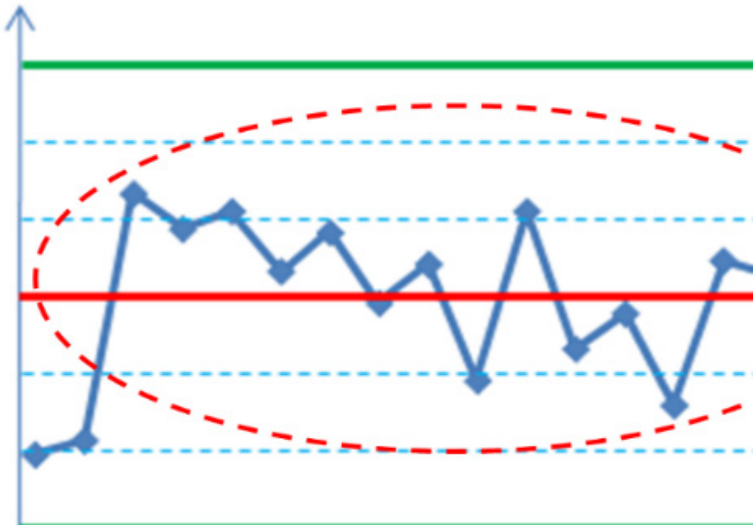


Fig. 5.24 Western Electric Test 4

Test 5: Two out of three points in a row at least two standard deviations from the center line (in zone A or beyond) on the same side of the center line

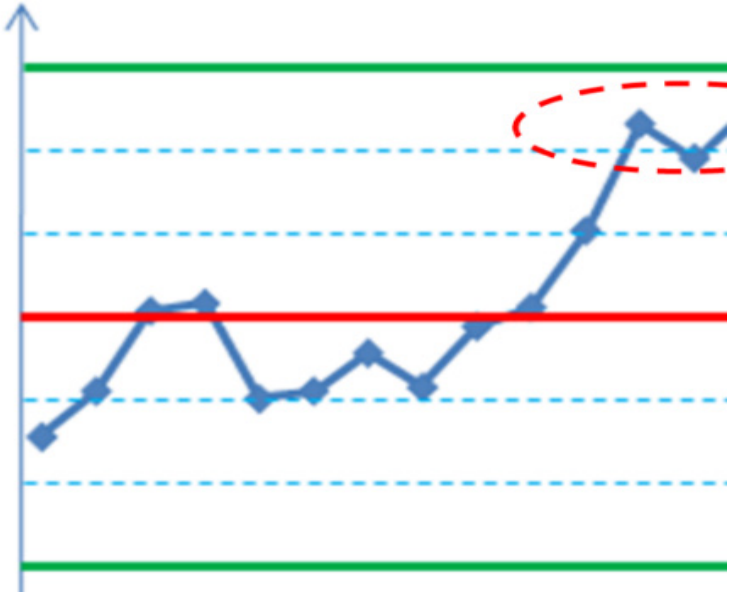


Fig. 5.25 Western Electric Test 5

Test 5 is similar to test 2, but represents a more acute shift in the process mean.

Test 6: Four out of five points in a row at least one standard deviation from the center line (in zone B or beyond) on the same side of the center line

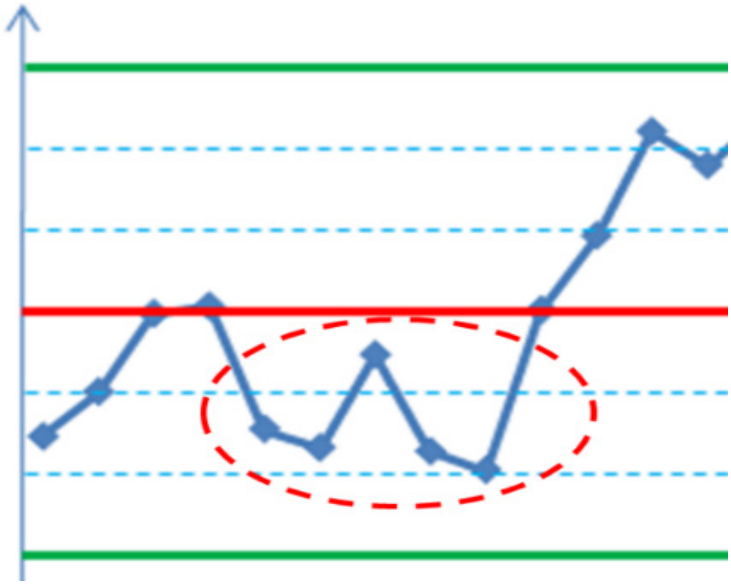


Fig. 5.26 Western Electric Test 6

Test 6 is similar to test 2, but represents a more acute shift in the process mean.

Test 7: Fifteen points in a row within one standard deviation from the center line (in zone C) on either side of the center line

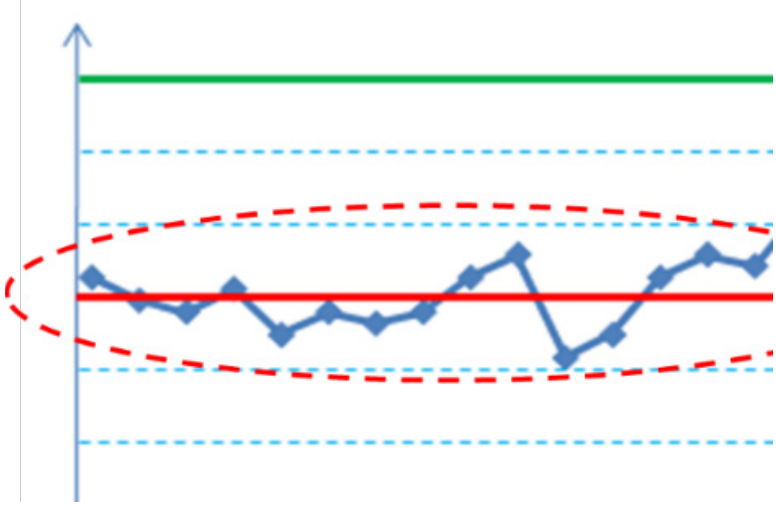


Fig. 5.27 Western Electric Test 7

Test 7 indicates a change in the level of variation around the process mean.

Test 8: Eight points in a row beyond one standard deviation from the center line (beyond zone C) on either side of the center line

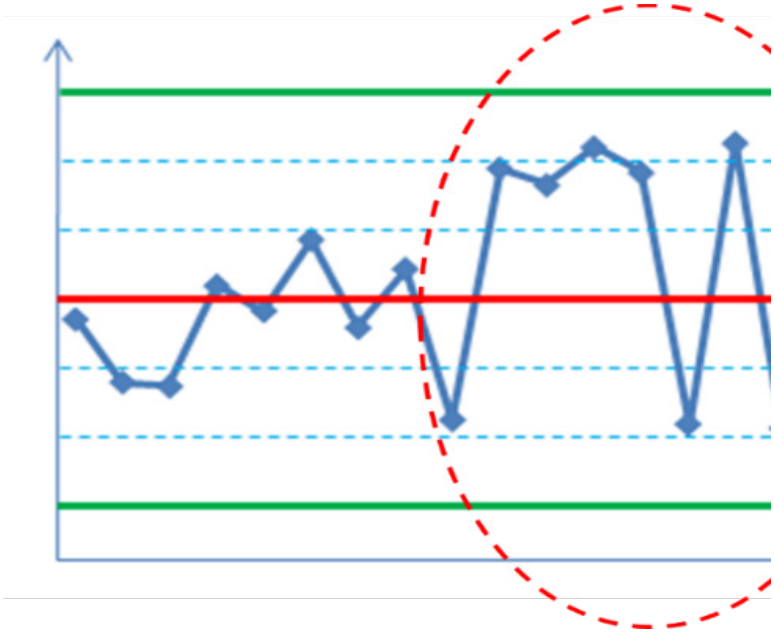


Fig. 5.28 Western Electric Test 8

While test 7 indicates a narrowing of variation, test 8 indicates an increase in variation.

Next Steps
If no data points fail any tests for special causes (Western Electric rules), the process is in *statistical control*. If any data point fails any tests for special causes, the process is *unstable* and we will need to investigate the observation thoroughly to discover and take actions on the special causes leading to the changes. Special cause variation must be addressed and the process must be stable before process capability analysis can be done. Process stability is the prerequisite of process capability analysis.

5.2.12 SUBGROUPS AND SAMPLING

Subgroups

Rational subgrouping is the basic sampling scheme in SPC (Statistical Process Control). When sampling, we randomly select a group (i.e., a subgroup) of items from the population of interest. The subgroup size is the count of samples in a subgroup. It can be constant or variable. Depending on the subgroup sizes, we select different control charts accordingly.

Impact of Variation

The *rational subgrouping* strategy is designed to minimize the opportunity of having special cause variation within subgroups, making it easier to identify special cause between subgroups. If there is only random variation (background noise) within subgroups, all the special cause variation would be reflected between subgroups. It is easier to detect the out-of-control situation. Random variation is inherent and indelible in the process. We are more interested in identifying and taking actions on special cause variation.

Frequency of Sampling

The *frequency of sampling* in SPC depends on whether we have sufficient data to signal the changes in a process with reasonable time and costs. The more frequently we sample, the higher the costs sampling may trigger. We need the subject matter experts' knowledge on the nature and characteristics of the process to make good decisions on sampling frequency.

5.3 Six Sigma Control Plans

5.3.1 Cost Benefit Analysis

What is Cost-Benefit Analysis?

For a project to be feasible, the benefits must outweigh the costs. The *cost-benefit analysis* is a systematic method to assess and compare the financial costs and benefits of multiple scenarios in order to make sound economic decisions.

A cost-benefit analysis is recommended to be done at the beginning of the project based on estimations of the experts from the finance team in order to determine whether the project is financially feasible. It is recommended to update the cost-benefit analysis at each DMAIC phase of the project.

Why Cost-Benefit Analysis?

In the Define phase of the project, the cost-benefit analysis helps us understand the financial feasibility of the project. In the middle phases of the project, updating and reviewing the cost-benefit analysis helps us compare potential solutions and make robust data-driven decisions. In the Control phase of the project, the cost-benefit analysis helps us track the project's profitability.

Return on Investment

A common financial measure of a project's impact to profitability, *return on investment* (also called ROI, rate of return, or ROR) is the ratio of the net financial benefits (either gain or loss) of a project or investment to the financial costs.

$$ROI = \frac{TotalNetBenefits}{TotalCosts} \times 100\%$$

Where:

$$TotalNetBenefits = TotalBenefits - TotalCosts$$

Return on Investment (ROI)

The return on investment is used to evaluate the financial feasibility and profitability of a project or investment.

- If ROI < 0, the investment is not financially viable.
- If ROI = 0, the investment has neither gain nor loss.
- If ROI > 0, the investment has financial gains.

The higher the ROI, the more profitable the project. Any value over 100 percent means the project has a return greater than the cost, and the higher the ROI, the more profitable the project.

Net Present Value (NPV)

The *net present value* (also called NPV, net present worth, or NPW) is the total present value of the cash flows calculated using a discount rate. The NPV accounts for the time value of money.

$$NPV = \frac{NetCashFlow_t}{(1+r)^t}$$

Where:

- $NetCashFlow_t$ is the net cash flow happening at time t
- r is the discount rate
- t is the time of the cash flow.

Often projects do not have immediate returns and it takes time to recover the costs incurred during the project.

Cost Estimation

There are many types of costs that can be incurred during a project. Some are one-time investments, such as equipment, consulting, or other assets, while others are incremental costs to be incurred going forward.

Examples of costs triggered by the project:

- Administration
- Asset
- Equipment
- Material
- Delivery
- Real estate
- Labor
- Training

- Consulting

Benefits Estimation

Benefits can take many forms: additional revenue, reduced wasted, operational resource costs, productivity improvements, or market share increases. Other types of benefits that are harder to qualify include customer satisfaction and associate satisfaction.

Examples of benefits generated by the project:

- Direct revenue increase
- Waste reduction
- Operation cost reduction
- Quality and productivity improvement
- Market share increase
- Cost avoidance
- Customer satisfaction improvement
- Associate satisfaction improvement

Challenges in Cost and Benefit Estimation

Different analysts might come up with different cost and benefit estimations due to their subjectivity in determining:

- The discount rate
- The time length of the project and its impact
- Potential costs of the project
- The tangible/intangible benefits of the project
- The specific contribution of the project to the relevant financial gains/loss

5.3.2 Elements of Control Plans

Control Plans

A key component of a solid DMAIC project, the *control plan* ensures that the changes introduced by a Six Sigma project are sustained over time.

Benefits of the Control phase:

- Methodical roll-out of changes including standardization of processes and work procedures
- Ensure compliance with changes through methods like auditing and corrective actions
- Transfer solutions and learning across the enterprise
- Plan and communicate standardized work procedures
- Coordinate ongoing team and individual involvement

- Standardize data collection and procedures
- Measure process performance, stability, and capability
- Plan actions that mitigate possible out-of-control conditions
- Sustain changes over time

The Control phase ensures new processes and procedures are standardized and compliance is assured.

What is a Control Plan?

A *control plan* is a management planning tool to identify, describe, and monitor the process performance metrics in order to meet the customer specifications steadily. It proposes the plan of monitoring the stability and capability of inputs and outputs of critical process steps in the Control phase of a project. It covers the data collection plan of gathering the process performance measurements. Control plans are the most overlooked element of most projects. It is critical that a good solution be solidified with a great control plan!

Control Plan Elements

There are many possible elements within a control plan. All of these need to be agreed to and owned by the process owner—the person who is responsible for the process performance.

Control Plan—The clear and concise summary document that details key process steps, CTQs metrics, measurements, and corrective actions.

Standard Operating Procedures (SOPs) —Supporting documentation showing the "who does what, when, and how" in completing the tasks.

Communication Plan—Document outlining messages to be delivered and the target audience.

Training Plan—Document outlining the necessary training for employees to successfully perform new processes and procedures.

Audit Checklists—Document that provides auditors with the audit questions they need to ask.

Corrective Actions—Activities that need to be conducted when an audit fails.

Control Plan

The control plan identifies critical process steps that have significant impact on the products or services and the appropriate control mechanisms. It includes measurement systems that monitor and help manage key process performance measures. Specified limits and targets for these performance metrics should be clearly defined and communicated. Sampling plans should be used to collect the data for the measurements.

Lean Six Sigma Control Plan

Process: _____ Preparer: _____ Page: _____ of ____
Customer: _____ Email: _____ Reference No: _____
Stakeholder: _____ Phone: _____ Revision Date: _____
Business: _____ Owner: _____ Approval: _____

Process	Process Step	CTQ/Metric	CTQ / Metric Equation	Specification/ Requirement		Measurement Method	Sample Size	Measure Frequency	Responsible for Metric	Link or Report Name	Corrective Action	Responsible for Action
				LSL	USL							

Fig. 5.29 Control Plan Key Process and Critical Measurements

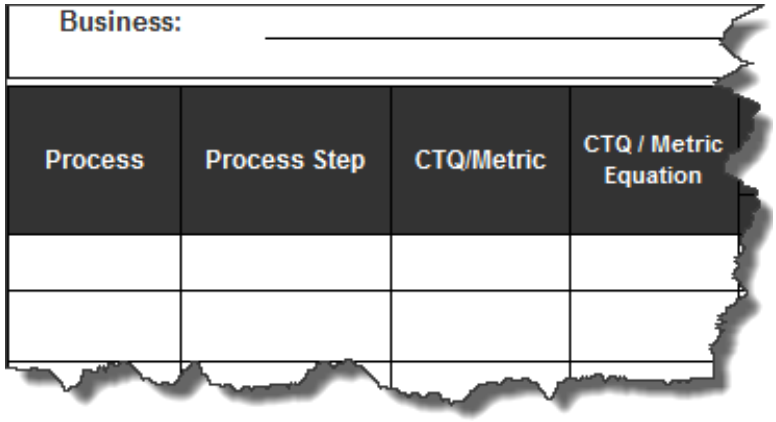


Fig. 5.30 Process and Metrics portion of Control Plan

- Key processes and process steps are identified.
- Critical information regarding key measurements is documented and clarified.

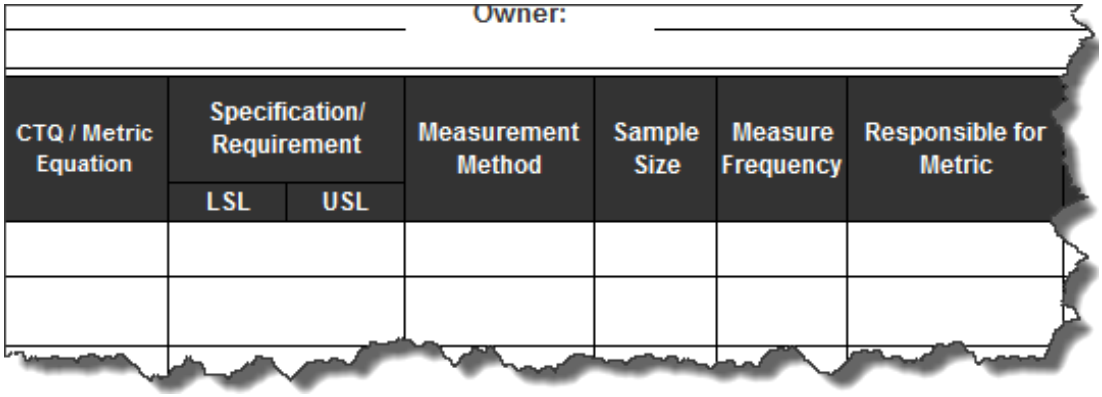


Fig. 5.31 Measurements portion of Control Plan

- Measurements are clearly defined with equations.
- Customer specifications are declared.
- Other key measurement information is documented: sample size, measurement frequency, people responsible for the measurement, etc.

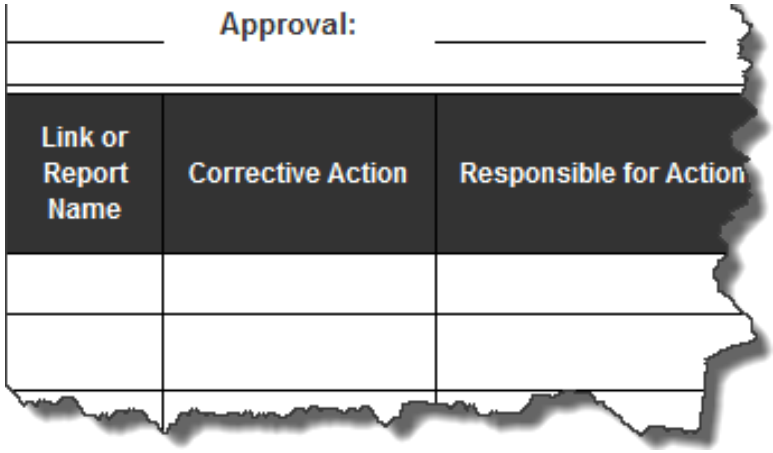


Fig. 5.32 Corrective Actions and Responsible Parties Portion of Control Plan

- Where will this measurement or report be found? Good control plans provide linking information or other report reference information.
- Control plans identify the mitigating action or corrective actions required in the event the measurement falls out of spec or control. Responsible parties are also declared.

Not only does a control plan spell out what is measured and what the limits are around the measurement, but it also details actions that must be taken when out-of-control or out-of-spec conditions occur.

Standard Operating Procedures (SOPs)

Standard Operating Procedures (SOPs) are documents that focus on process steps, activities, and specific tasks required to complete an operation. SOPs should not be much more than two to four pages and should be written to the user's level of required detail and information. The level of detail is dependent on the position's required skills and training. Good SOPs are auditable, easy to follow, and not difficult to find. Auditable characteristics are: observable actions and countable frequencies. Results should be evident to a third party (*compliance to the SOP must be measurable*). Standard operating procedures are an important element in a control plan when a specific process is being prescribed to achieve quality output.

SOP Elements

SOPs are intended to impart high value information in concise and well-documented manner. There are many components to maintain currency, ensure that employees understand why the SOP is necessary and important, and other supporting information to make scope very clear for the SOP.

- SOP Title and Version Number—Provide a title and unique identification number with version information.
- Date—List the original creation date; add all revision dates.
- Purpose—State the reason for the SOP and what it intends to accomplish.

- Scope—Identify all functions, jobs, positions, and/or processes governed or affected by the SOP.
- Responsibilities—Identify job functions and positions (not people) responsible for carrying out activities listed in the SOP.
- Materials—List all material inputs: parts, files, data, information, instruments, etc.
- Process Map—Show high level or level two to three process maps or other graphical representations of operating steps.
- Process Metrics—Declare all process metrics and targets or specifications.
- Procedures—List actual steps required to perform the function.
- References—List any documents that support the SOP.

SOP Template

Standard Operating Procedures

SOP Name/Title:		
Document Storage Location/Source:		Document No:
SOP Originator:	Approving Position:	Effective Date:
Name:	Name:	Last Edited Date:
Signature:	Signature:	Other:

1. Purpose

"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut

2. Scope

"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut

3. Responsibilities (RACI)

Responsible	Accountable	Consulted	Informed
John Doe	Jane Doe	Jack Doe	Jill Doe
Pam Doe	Paul Doe	Phil Doe	Peggy Doe

4. Materials

"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut

5. Related Documents

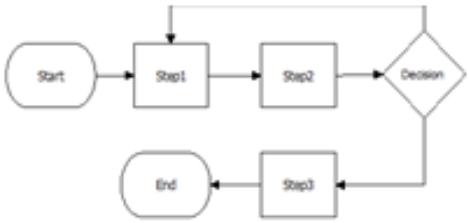
"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut

6. Definitions

"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut

Fig. 5.33 SOP Example part 1

7. Process Map



8. Procedures

Step	Action	Responsible
1		
2		
3		

9. Process Metrics

- "Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut
- "Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut

10. Resources

- "Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut
- "Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut

Fig. 5.34 SOP Example part 2

Communication Plans

Communication plans are documents that focus on planning and preparing for the dissemination of information. They organize messages and ensure that the proper audiences receive the correct message at the right time. A good communication plan identifies:

- Audience
- Key points/message
- Medium (how the message is to be delivered)
- Delivery schedule

- Messenger
- Dependencies and escalation points
- Follow-up messages and delivery mediums

Communication plans help develop and execute strategies for delivering changes to an organization.

Communication Plan Template

Communication Plan Template

Process/Function Name	Project/Program Name	Project Lead	Project Sponsor/Champion
Communication Purpose:			

Target Audience	Key Message	Message Dependencies	Delivery Date	Location	Medium	Follow up Medium	Messenger	Escalation Path	Contact Information

Fig. 5.35 Communication Plan Template

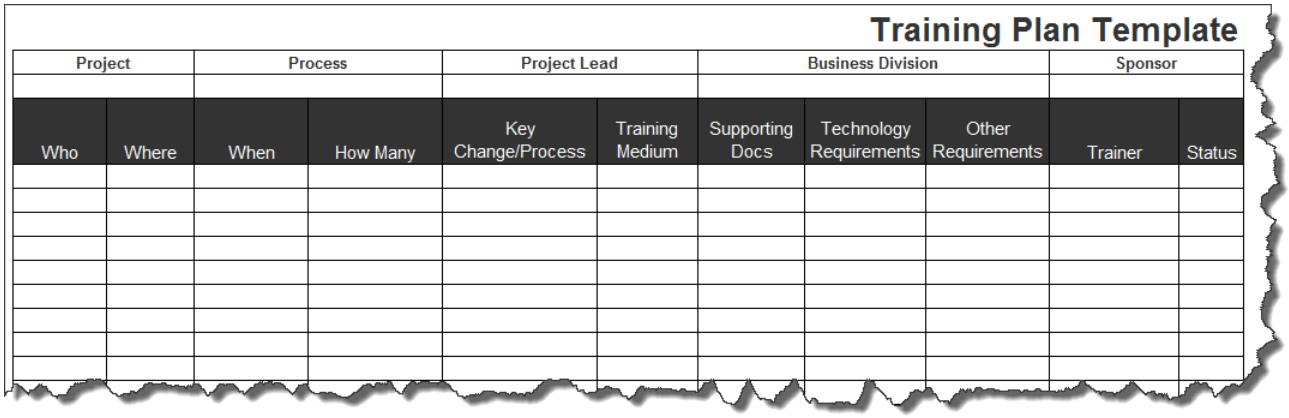
Training Plans

Training plans are used to manage the delivery of training for new processes and procedures. Most GB or BB projects will require changes to processes and/or procedures that must be executed or followed by various employees. Training plans should:

- Incorporate all SOPs related to performing new or modified tasks
- Use and support existing SOPs and do not supersede them
- Include logistics
    - One-on-one or classroom
    - Instruction time
    - Location of training materials
    - Master training reference materials
    - Instructors and intended audience
    - Trainee names

Different components of training plan include: training delivery method, training duration, location of materials and references, instructor names, and intended audience.

Training Plan Template



Training Plan Template

Project		Process		Project Lead		Business Division			Sponsor	
Who	Where	When	How Many	Key Change/Process	Training Medium	Supporting Docs	Technology Requirements	Other Requirements	Trainer	Status

Fig. 5.36 Training Plan Template

Audits

What is an *audit*?

ISO 9000 defines an audit as "a systematic and independent examination to determine whether quality activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve objectives."

Audits are used to ensure actions, processes, procedures, and other tasks are performed as expected. "Say what you do and do what you say."—this is fundamentally what audits are checking for.

Audit Guidelines

Audits must be done carefully to achieve the desired result. Audits should be directed by managers, supervisors, and other accountable positions. An audit's purpose must be well-defined and executed by independent unbiased personnel.

Auditors must:

- Be qualified to perform their tasks
- Attend and successfully complete an internal auditing training session
- Be able to identify whether or not activities are being followed according to the defined SOP
- Base conclusions on facts and objective evidence
- Use a well-documented audit checklist
- Audits should confirm compliance or declare non-compliance

Audit Checklists

Auditors should review the SOPs before preparing checklists or ensure that existing checklists properly reference SOPs.

Audit checklists:

- Serve as guides for identifying items to be examined
- Are used in conjunction with understanding of the procedure
- Ensure a well-defined audit scope
- Identify needed facts during audits
- Provide places to record gathered facts.
- Checklists should include:
- A review of training records
- A review of maintenance records
- Questions or observations that focus on expected behaviors
- Questions should be open-ended where possible
- Definitive observations yes/no, true/false, present/absent, etc.

Audit Checklist Template

Audit Checklist

Target Area:	Statement of Audit Objective:	Auditor:	Audit Date:
Audit Technique	Auditable Item, Observation, Procedure etc.	Individual Auditor Rating (Circle Rating)	
Observation	Have all associates been trained?	YES	NO
Observation	Is training documentation available?	YES	NO
Observation	Is training documentation current?	YES	NO
Observation	Are associates wearing proper safety gear?	YES	NO
Observation	Are SOP's available?	YES	NO
Observation	Are SOP's current?	YES	NO
Observation	Is quality being measured	YES	NO
Observation	Is sampling being conducted in random fashion	YES	NO
Observation	Is sampling meeting it's sample size target?	YES	NO
Observation	Are control charts in control	YES	NO
Observation	Are control charts current?	YES	NO
Observation	Is the process capability index >1.0?	YES	NO
Number of Out of Compliance Observations			
Total Observations			
Audit Yield			#DIV/0!
Corrective Actions Required			
Auditor Comments			

Fig. 5.37 Audit Checklist

5.3.3 Response Plan Elements

What is a Response Plan?

A *response plan* should be a component of as many control plan elements as possible. Response plans are a management planning tool to describe corrective actions necessary in the event of out-of-control situations. There is never any guarantee that processes will always perform as designed. Therefore, it is wise to prepare for occasions when special causes are present. Response plans help us mitigate risks and, as already mentioned, should be part of several control plan elements.

Response Plan Elements

- Action triggers
    - When do we need to take actions to correct a problem or issue?
- Action recommendation
    - What activities are required in order to solve the problem in the process? The action recommended can be short-term (quick fix) or long-term (true process improvement).
- Action respondent
    - Who is responsible for taking actions?
- Action date
    - When did the actions happen?
- Action results
    - What actions have been taken?
    - When were actions taken?
    - What are the outcomes of the actions taken?

A good response plan clearly describes when the response plan is acted upon, what the action is, who is responsible for taking action and when, and what the results are after the action is taken.

Fig. 5.38 Corrective Action and Responsible Party portion of Control Plan

A response plan denotes what the action is to be taken and who is responsible. Note the response plan element in this control plan template.

Escalation Path	Contact Information

Fig. 5.39 Escalation contact information for communication plan

The communication plan has a section denoting how information is disseminated. Note the response element by the identification of escalations and contact information.

Audit Checklist

Target Area:	Statement of Audit Objective:	Auditor:	Audit Date:
Audit Technique	Auditable Item, Observation, Procedure etc.	Individual Auditor Rating (Circle Rating)	
Observation	Have all associates been trained?	YES	NO
Observation	Is training documentation available?	YES	NO
Observation	Is training documentation current?	YES	NO
Observation	Are associates wearing proper safety gear?	YES	NO
Observation	Are SOP's available?	YES	NO
Observation	Are SOP's current?	YES	NO
Observation	Is quality being measured	YES	NO
Observation	Is sampling being conducted in random fashion	YES	NO
Observation	Is sampling meeting it's sample size target?	YES	NO
Observation	Are control charts in control	YES	NO
Observation	Are control charts current?	YES	NO
Observation	Is the process capability index >1.0?	YES	NO
Number of Out of Compliance Observations			
Total Observations			
Audit Yield			#DIV/0!
Corrective Actions Required			
Auditor Comments			

Fig. 5.40 Audit Checklist

An audit checklist denotes corrective actions necessary as a result of the audit. The is a form of response plan.

6.0 INDEX