Section IV

Analyze

Agenda

- ✓ Pre-Analyze Considerations
- ✓ Value Stream Analysis
- ✓ Sources of Variation
- ✓ Regression
- ✓ Confidence Intervals
- ✓ Parametric Hypothesis Testing
- ✓ Non Parametric Hypothesis Testing
- ✓ Analysis of Categorical Data and Current Reality Tree

Section IV, Lesson 1

Pre-Analyze Considerations

Agenda

- ✓Analyze Phase – Introduction
- ✓Pre-Analyze Considerations
- ✓Objectives of Analyze
- ✓Visually Displaying Data (Histogram, Run Chart, Pareto Chart, Scatter Diagram)

Analyze Phase – Introduction

- ✓ Analyze is a key part of the Six Sigma project
- ✓ All the data collected are reviewed here
- ✓ Major analysis for both inputs and outputs are done here

Pre-Analyze Considerations

- ✓The Six Sigma team before stepping into the Analyze phase must have collected the baseline data and process conditions inference, i.e., normality, stability, etc.
- ✓The baseline data along with process conditions need to be presented in the tollgate meeting.
- ✓The Six Sigma team may also decide if they wish to continue with the DMAIC process, especially, if all efforts to bring the process in stable status continue to fail.
- ✓In such a scenario, a DFSS project could be initiated at the end of the Measure phase.

Pre-Analyze Considerations (Contd.)

Sample baseline data

- ✓Measurement system analysis status = Validated
- ✓Stability = Validated
- ✓Normality = Validated
- ✓Project Y = Delivery hours
- ✓Project Y type = Continuous
- ✓Cp = 0.9
- ✓Cpk = 0.8
- ✓DPMO = 67,000 (approx .)
- ✓Baseline sigma = 2.4

A sample baseline document has been provided as part of the toolkit.

Objectives of Analyze

- ✓Analyze the value stream to identify gaps to be fixed.
- ✓Analyze what is causing the variation for the gap.
- ✓Determine the key driver, key process input variable (KPIV or X), that impacts the variations in key process output variables (KPOV).
- ✓Validate the relationship between KPIV and KPOV.
- ✓Test hypothesis or confidence intervals to validate assumptions.

Visually Displaying Data

Data can be represented in various forms, example:

✓**Histogram** – Used for displaying volume of data in each category or range;

✓**Run chart** – Typically used for plotting data on a time scale plot;

✓**Scatter diagram** – Used for finding correlation between 2 data points; and

✓**Pareto Chart** – A modified version of histogram to display data in descending order of volume. A cumulative percentage line graph is plotted to aid 80-20 rule representation.

Depending on the data and purpose of the representation, the tool and chart need to be decided.

Summary

✓Introduction

✓Pre-analyze considerations

✓Objectives of analyze

✓Visually displaying data (histogram, run chart, pareto chart, scatter diagram)

Section IV, Lesson 2

Value Stream Analysis

Agenda

- ✓Value, Waste, and Non-Value Add (NVA) Activities
- ✓What is Value Stream
- ✓Value Stream Example
- ✓Value Stream Analysis – Muda
- ✓Value Stream Map
- ✓Spaghetti Charts

Value, Waste, and NVA Activities

- ✓ Value defined by the customer is any activity or thing in the product or the process for which the customer is willing to pay for.
- ✓ All activities that customers are not willing to pay for, are considered non-value add (NVA) activities.
- ✓ Waste is anything unnecessary for the product or the process. Waste can be in the form of productivity, quality, time, or cost. A waste activity consumes resources but doesn't deliver any value to the customer.
- ✓ Some NVAs could be mandatory for the business.
  - For example, training of resources may not be considered value-adding by the customer as it doesn't add value to the product directly, but is something a company would invest for long-term benefits.

What is a Value Stream?

- ✓A value stream is the flow of activities that consists value add as well as non-value add activities, which takes the raw material from the supplier, and deliver the end product to the customer.
- ✓Analysis of value stream is important to determine value add and non-value add activities in any process.
- ✓The customer defines what is the “value” in a product and not the business. So, when value is analyzed, it should be kept in mind whether the customer is perceiving it as value.
- ✓All warusa kagen conditions, namely, muda (waste), mura (unevenness), and muri (overburden) are identified in the analysis of a value stream.

Value Stream – Example

- ✓A company diagnosed that dents in their shelf welds was the key area to be rectified. They spent a lot of money in rectifying the weld dents, so their customers could be happy. Then they took the shelves to their customers to identify if rectification of weld dents added value to them.
- ✓The customers said, “No”. They flatly refused to identify the weld dents, as their primary need was to have shelves that were straight in angle.
- ✓The voice of customer was – “They wanted shelves in straight angle.”
- ✓The voice of business focused on weld dents.
- ✓This disconnect lead to defining wrong **value** for customers.

Important – The VOC often gives the insight on what value is as per the customer.

Value Stream Analysis – Muda

✓When analyzing the value stream, the Black Belt needs to first identify the muda in the system.

✓In typical cases, muda accounts for 90% of the time spent in operations.

✓The Black Belt needs to identify these muda activities and to start with, has to make it transparent to the Six Sigma team, and discuss with them.

✓The Black Belt can expect tremendous change resistance here.

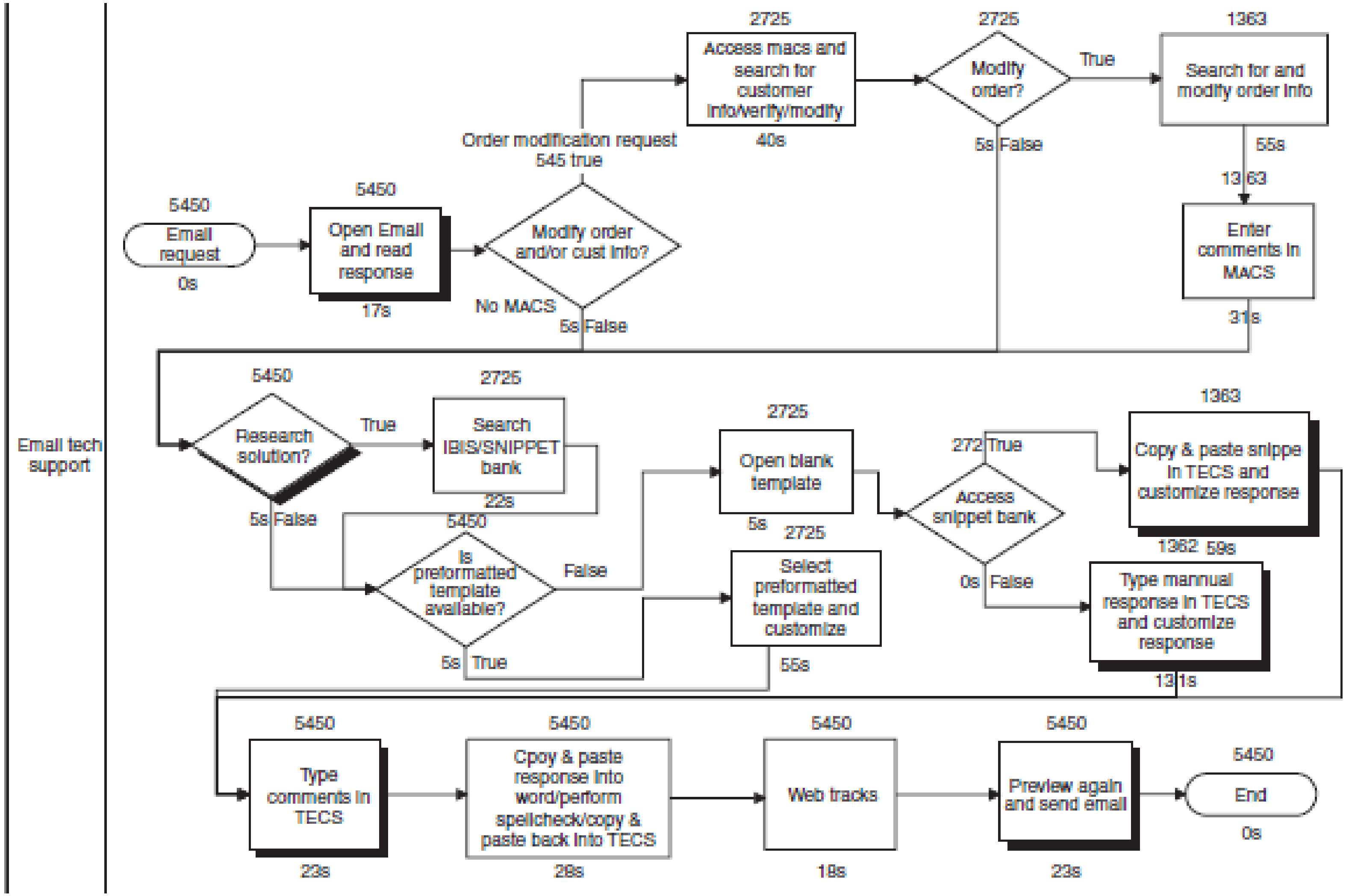
Value Stream Analysis – Muda (Contd.)

✓When identifying muda, the Black Belt should look out for CLOSEDMITTS:

- C – Complexity → Unnecessary steps and excessive documentation;
- L – Labor → Inefficient operations and excess headcount;
- O – Overproduction → Producing more than customer needs or before;
- S – Space → Storage space for inventory;
- E – Energy → Wasted human energy;
- D – Defects → Repair or rework in products;
- M – Materials → Scrap or ordering more than needed;
- I – Idle time → Material sits for time;
- T – Time → Time waste;
- T – Transportation → Movement adding no value; and
- S – Safety hazards → Unsafe environments.

✓CLOSEDMITTS loosely translates to the popular TIMWOODS acronym for wastes.

The value stream map for a technical support process.



Value Stream Map (Contd.)

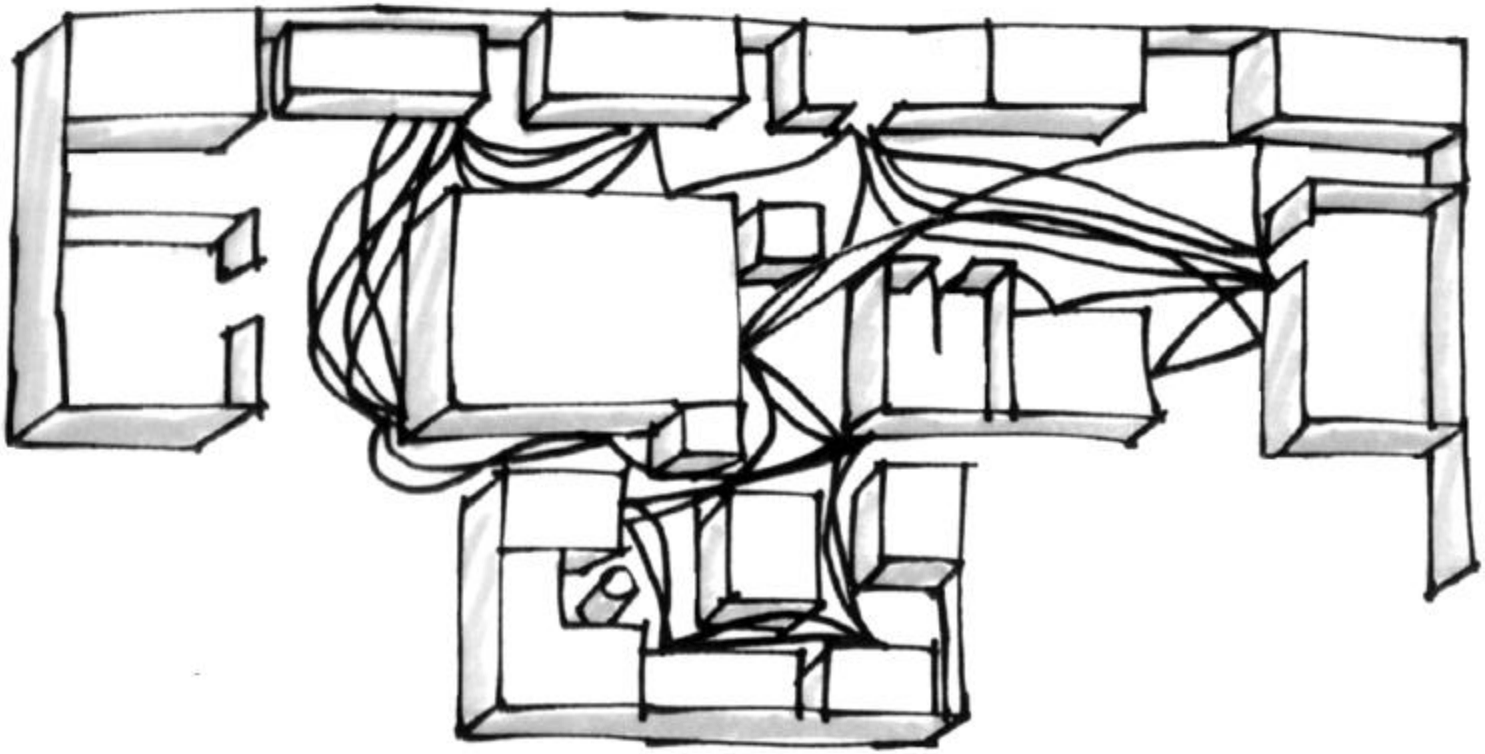
Inferences from value stream map

1. Only about 40% of the time spent on the support is value-added. Most others are non-value added.
2. Some activities in the NVA space could be considered mandatory for businesses. These are necessary NVAs.
3. Organizations have necessary NVAs due to poor quality levels in the first pass of the process (without any rework).
   3.1. Examples of necessary NVAs – Quality checks, management approvals, etc.
4. The Six Sigma team can begin with reviewing necessary NVAs and explore options to eliminate or reduce them.
5. Then the variation and other NVAs in the process should be analyzed.

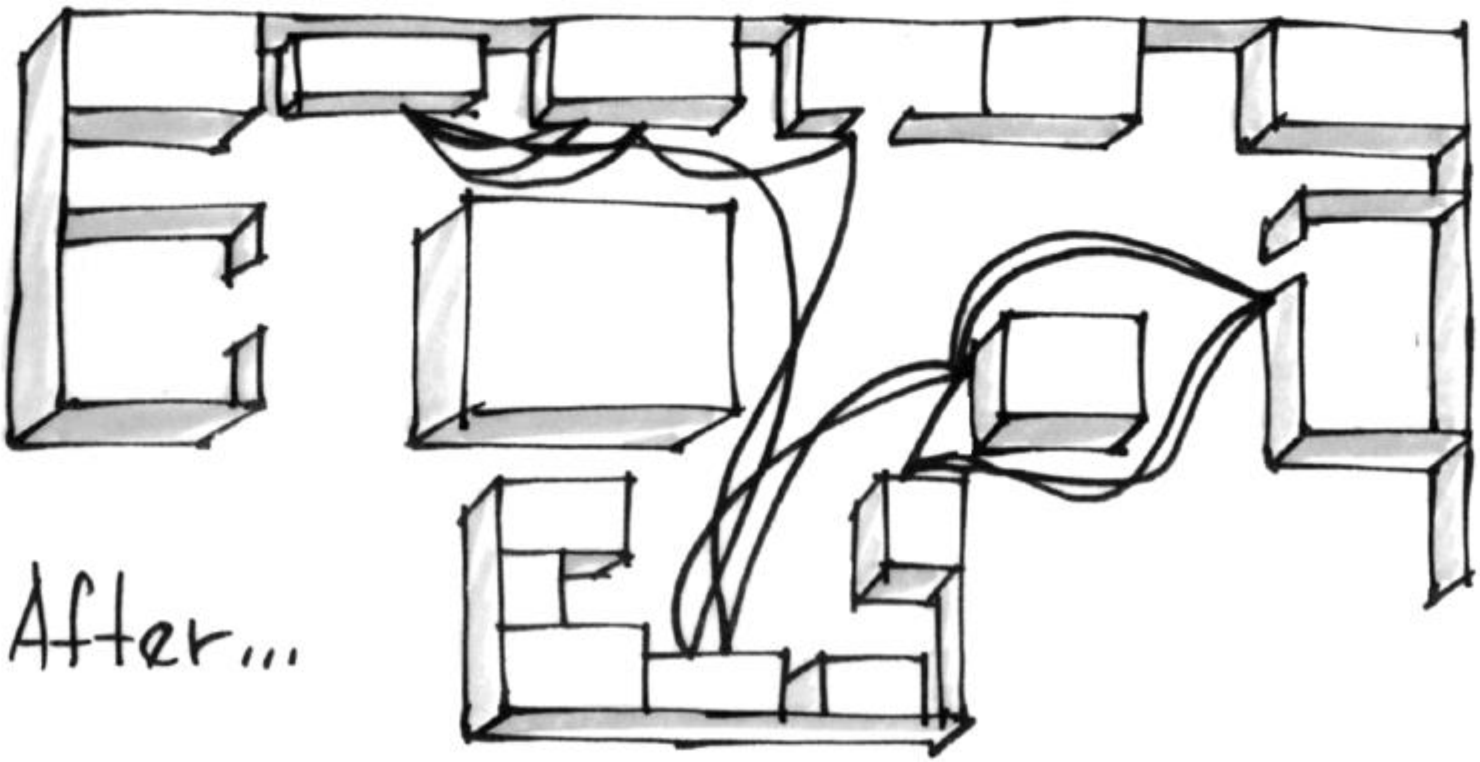
Spaghetti Charts

- ✓ Every Black Belt must learn how to draw a spaghetti chart for a process.
- ✓ The spaghetti chart will help understand the **physical process map** (where, things actually move in the system physically) and not necessarily just the value map (where, the focus is only on value add activities).
- ✓ These charts graphically depict movement of products and people in a process.
- ✓ A populated spaghetti chart with a lot of lines and curves will indicate that product and people move up and down a lot in the process – this is typically called as waste (Japanese term for these type of waste is **muda**).
- ✓ A densely populated spaghetti chart indicates unnecessary redundancies. That is things move from one place to another and back and forth, instead of moving in a direction with minimum repetition.

Spaghetti Chart – As Is



Spaghetti Chart – Should Be



Spaghetti Charts (Contd.)

- ✓Spaghetti charts are excellent process mapping tools that make wastes and redundant activities obvious to the naked eye.
- ✓If used properly, spaghetti charts could also be used to show the to-be or the future state process.
- ✓Spaghetti charts are also known as standardized work charts.

Summary

Topics learned in this lesson:

- ✓How to identify wastes and NVAs in a process?
- ✓How to perform value stream analysis?
- ✓The use of a value stream map.
- ✓The use of spaghetti chart in process setting.

Section IV, Lesson 3

Sources of Variation

Agenda

- ✓Sources of Variation – Common Cause and Special Cause
- ✓Cause and Effect Diagram
- ✓Affinity Diagram
- ✓Box Plot

Sources of Variation

Broadly, a process output varies due to the two reasons or sources of variation mentioned below:

1. **Common causes of variation** – These reasons happen frequently, come from within a process; and by principles of statistical control, they can only be reduced and not eliminated. Common causes of variation are often unassigned to their origins and are also known as chance causes, random causes, and noise.
2. **Special causes of variation** – These variations happen once in a while, unusual, not previously observed, and non-quantifiable; by principles of statistical control, they should be eliminated if found undesirable for the process. Special causes of variation can be tracked to a reason and are also known as non-random causes, assignable causes, and signals.

Sources of Variation (Contd.)

- ✓ A Black Belt must be able to understand all the reasons that contribute to the variation in the data, which results in the process output to vary.
- ✓ For example,
  - A train on an average takes 30 minutes to reach from point A to point B in a city. Typically it takes around 28 to 35 minutes. The variation is caused due to various reasons. This can be called common cause of variation.
  - On the other hand, a train gets delayed by 10 minutes due to signal failure, which is a very rare case. In such a scenario, signal failure due to one of the electrical equipment malfunctioning is a special cause of variation.

Cause and Effect Diagram

- ✓Cause and effect diagram (CE Diagram), also known as the fishbone diagram, is a very popular and an important tool that gives a list of all the issues due to which a problem occurs.
- ✓The CE diagram will give an insight if the source of variation is traceable or not, further concludes if it is a common or a special cause of variation.
- ✓The CE diagram allows detailed investigation (non-numerical) into what is causing the end situation to vary. The end situation here is known as effect, and the issues that cause the effect to vary are known as causes.
- ✓Causes could be level 1. These are causes which are visible. If these causes cannot explain why they are happening, they are common causes of variation.
- ✓After level 1 causes, level 2 causes need to be explored until a dead end is reached.

Cause and Effect Diagram (Contd.)

How to draw a cause and effect diagram:

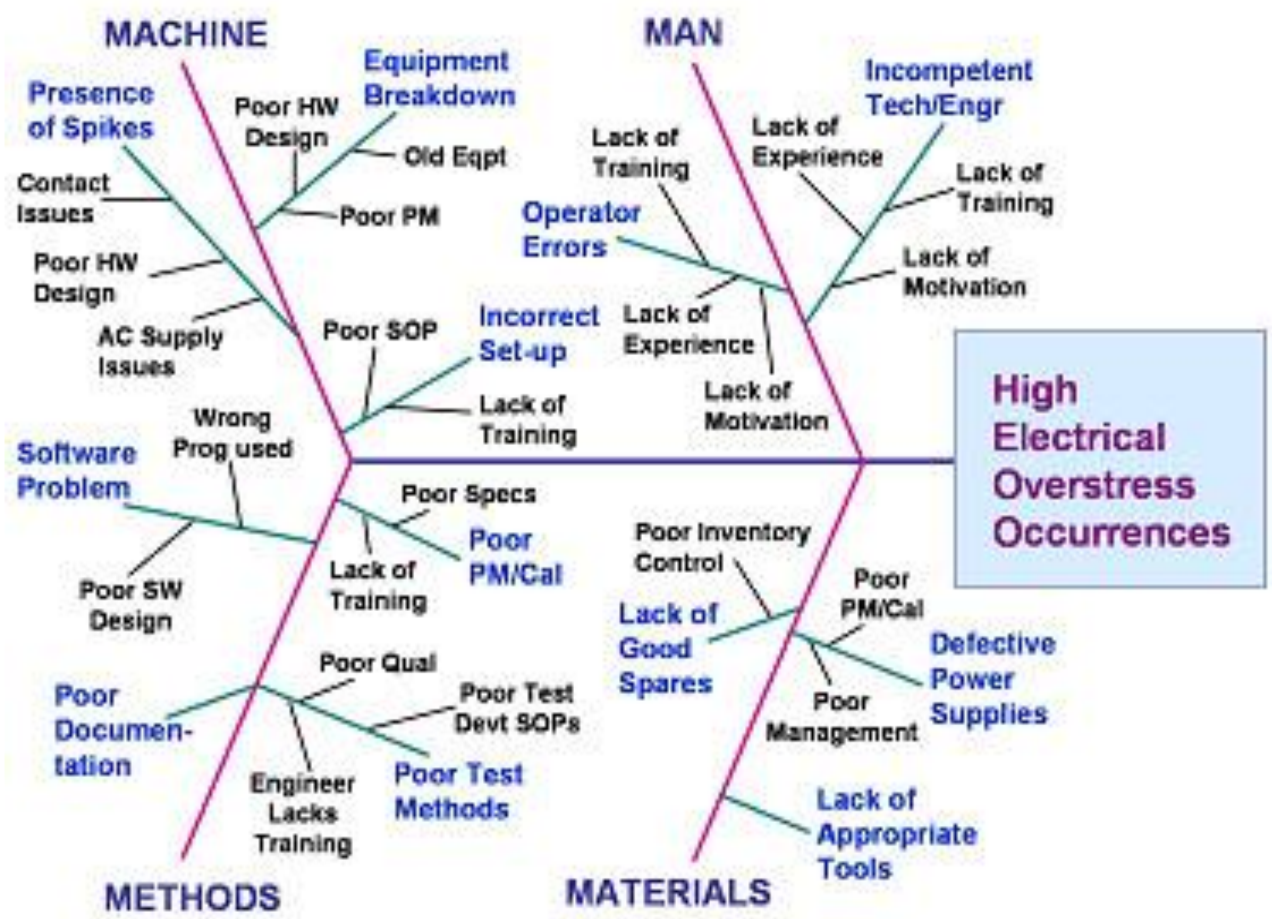
1. Use the tool, cause and effect diagram provided in the toolkit.
2. Draw the horizontal line and name the effect.
3. Brainstorm for causes contributing to the effect.
4. Correlate the causes to each of the 6M categories.
5. Ask questions on why each cause happens.
6. Add the causes as twigs to the level 1 cause.

Cause and Effect Diagram (Contd.)

How to analyze a cause and effect diagram:

1. CE diagrams help identify possible causes. Use a Pareto chart to prioritize the causes.
2. Thick cluster of causes in one area will guide to investigate further on that area.
3. Only a few specific causes in the main category may indicate a need for further identification.
4. If there are only a few branches for several major branches, one can choose to combine them.
5. Look for causes that repeat. These could be root causes.
6. For each of the cause, find out what needs to be measured to understand the cause completely. This will allow one to validate the cause correlation.

- A cause and effect diagram tool is added to the toolkit. A snapshot of the tool is given below:



Affinity Diagram

- ✓Tool to help organize ideas or items into groups and categories
- ✓Helps large teams in brainstorming to have focused idea generation on particular category or group
- ✓Helps discover connections between various pieces of information
- ✓Brainstorm root causes and solutions to a problem
- ✓Process
  - Record each idea on cards or notes
  - Look for ideas that seem to be related
  - Sort cards into groups until all cards have been used

Box Plot

- ✓Box plot is another useful tool that helps know the nature of variability in a process.
- ✓Also known as whiskers plot, candlestick plot, and box n whiskers plot.
- ✓The box plot drawn on Minitab as well as in Excel will divide the entire data set into four parts, known as quartiles. Quartiles consist or expect to consist of at least 25% of the data set.
- ✓The box is plotted for data between 25 percentile to 75 percentile with a line in between for 50th percentile. There is a line drawn for the top 25% and bottom 25%, and the outlier is highlighted separately.
- ✓These plots are often used in the analyze phase to understand the process' variability. It helps see distribution of values in several groups.
- ✓Box plot gives information about location, spread, and variability of data.

Box Plot (Contd.)

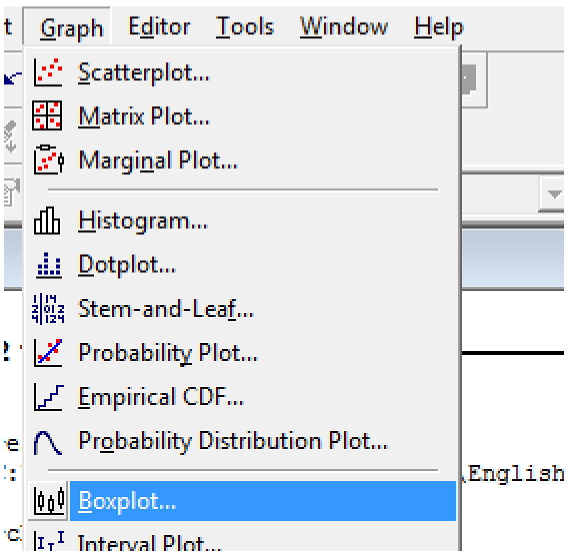
✓Data shown below is collected for delivery hours, assuming it is one of the key input variables, which can impact the performance of output variable.

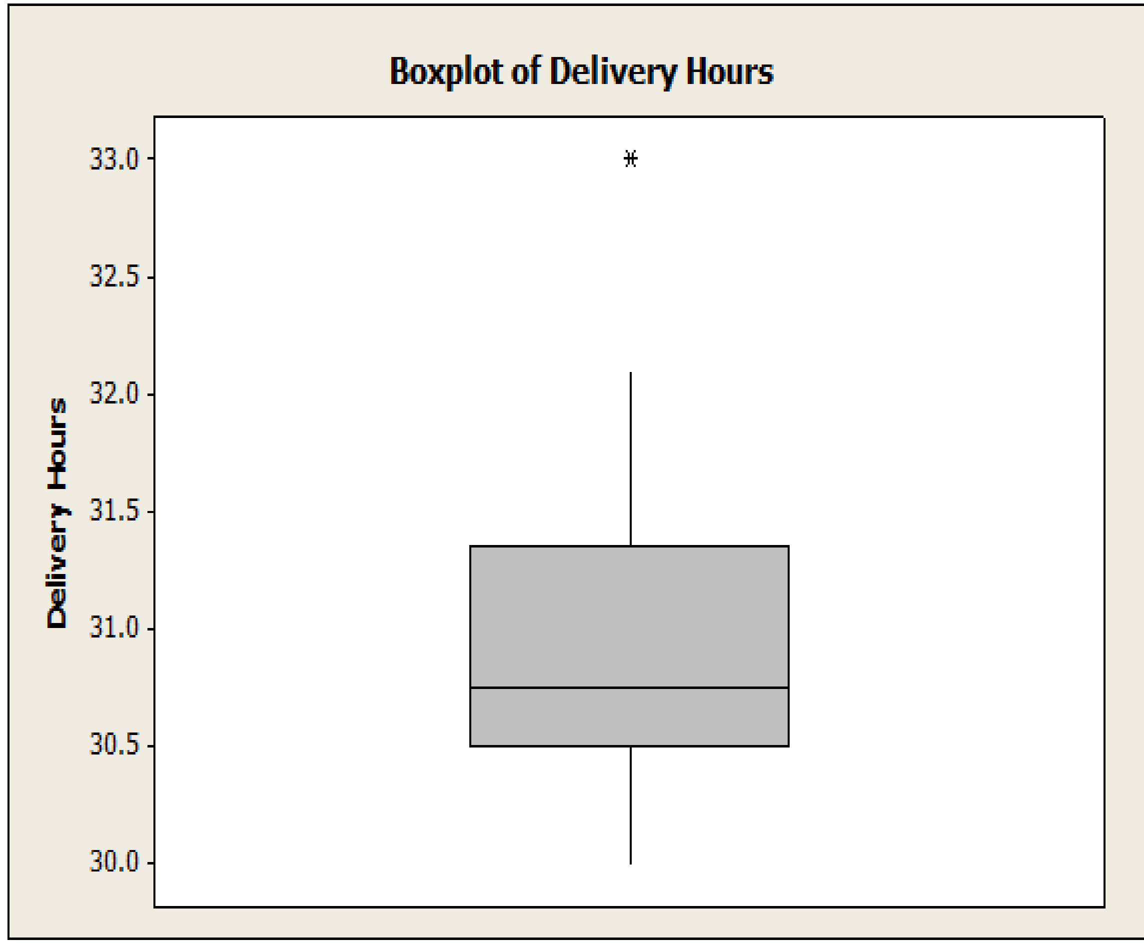
Day	Delivery Hours	Day	Delivery Hours
1	30	10	33
2	31	11	31.5
3	30.5	12	30.7
4	32	13	30.6
5	31.75	14	30.5
6	30.5	15	30.8
7	30.6	16	31.3
8	30.7	17	31
9	32.1	18	30.5

Next page provides steps on how to construct a box plot in Minitab and interpret.

Box Plot (Contd.)

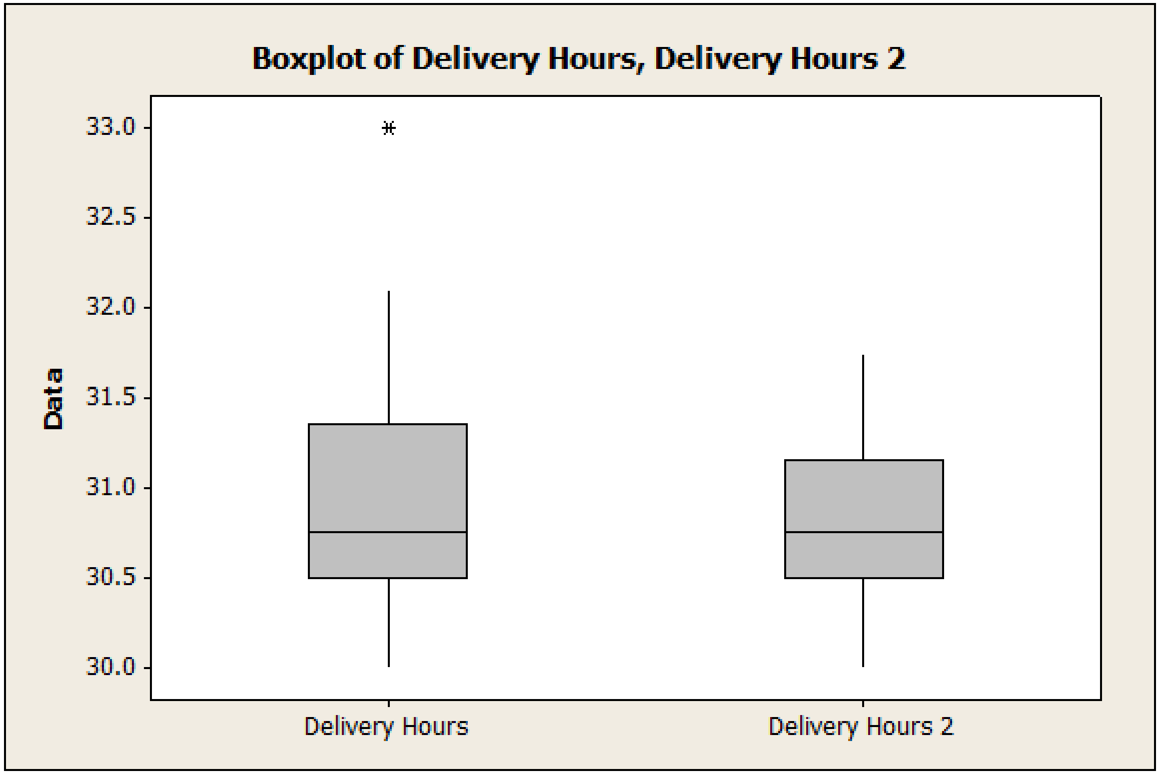
✓ Click on graph, and then on box plot. Select the variable “Delivery Hours” and a simple box plot option.





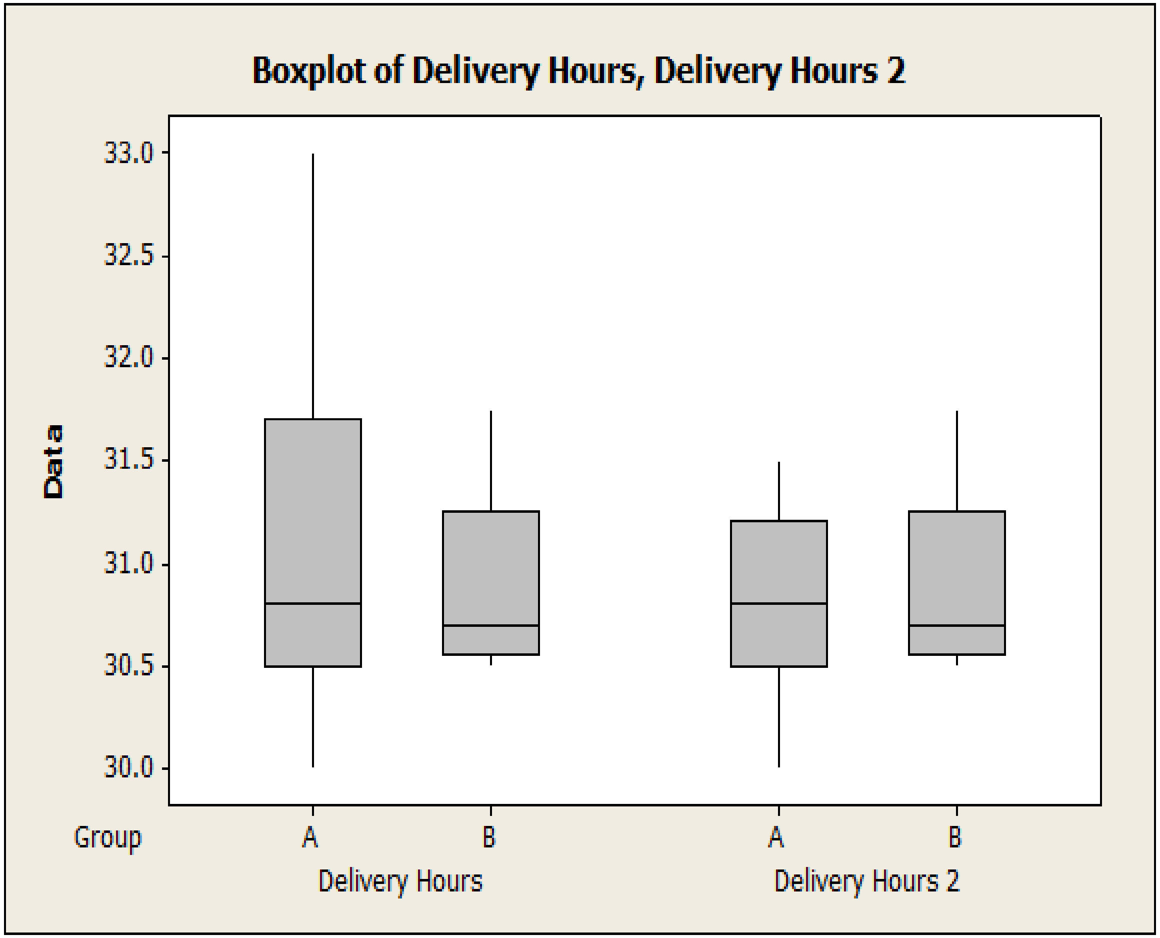
Box Plot (Contd.)

Use box plots to compare between two samples' data and their variability. Plot below shows some special cause variability in delivery hours, and no special causes in delivery hours 2. Also, the degree of skewness is less in delivery hours 2.



Box Plot (Contd.)

A box plot could also be used when data is stratified. This helps us understand variability within and between groups.



Summary

Topics discussed in this lesson:

- ✓ Sources of variation – Common and special causes of variation
- ✓ Cause and effect diagram
- ✓ Box plot and its use in determining variability in a process

Section IV, Lesson 4

Regression

Agenda

- ✓Objectives of Regression Analysis
- ✓Concepts of Regression Analysis
- ✓Simple Linear Regression
- ✓Multiple Linear Regression
- ✓An Introduction to Best Subsets Regression and Stepwise Regression

Objectives of Regression Analysis

- ✓ Regression analysis is a statistical technique for determining the relationships between a dependent variable and one or more independent variables.
- ✓ To find details of correlation between KPIV and KPOV.
- ✓ Regression analysis helps to determine how much variability in KPOV is explained by the selected KPIV.
- ✓ Regression analysis helps with a best fit line, it can be used to model data for estimation, prediction, and future use.

Concepts of Regression Analysis

- ✓Response variables –These are the output variables, which are been tested in the regression analysis.
- ✓Predictor variables – These are the input variables, which impacts the performance of the output variables.
- ✓p-value – It provides the probability value for significance testing. In terms of use, if p-value is less than 0.05, it is considered significant, which means a relationship can be concluded between the response and the predictor variables.
- ✓Multicollinearity – It is a statistical phenomenon in which, two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree of accuracy.

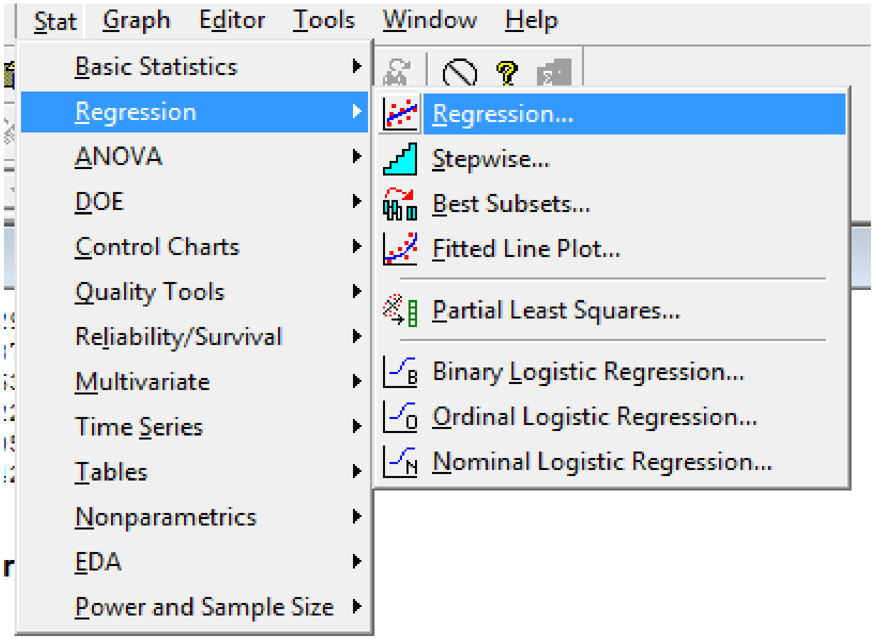
Concepts of Regression Analysis (Contd.)

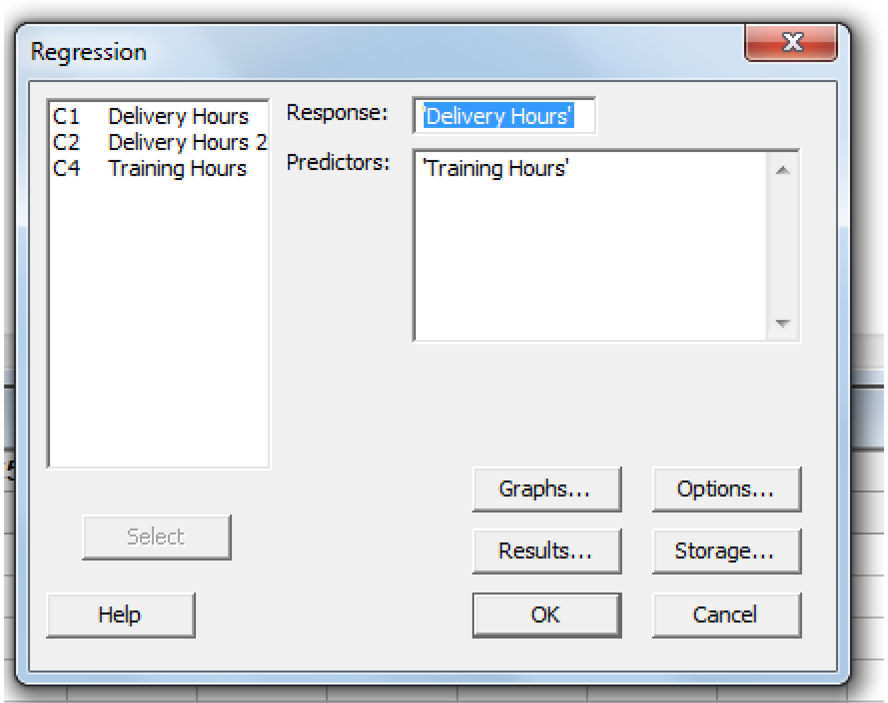
- ✓ Mallow's Cp – A value that compares the precision and bias of full models to the models of predictors resulting from best subsets.
- ✓ Durbin Watson statistic – An important statistic which needs to be selected in multiple regression which tests for presence of auto-correlation between residuals.
- ✓ Adjusted R^2 – The coefficient of determination to be used instead of R^2 , when multiple predictor variables are regressed against a response variable.
- ✓ Predicted R^2 – R^2 shows how well the model fits the data while predicted R^2 can be used to predict responses for new observations (predictor values.)
- ✓ Least squares method – The method used to calculate estimates by fitting a regression line to the points from a data set in such a way that the distance from the data point to the line square is the minimum.

Simple Linear Regression

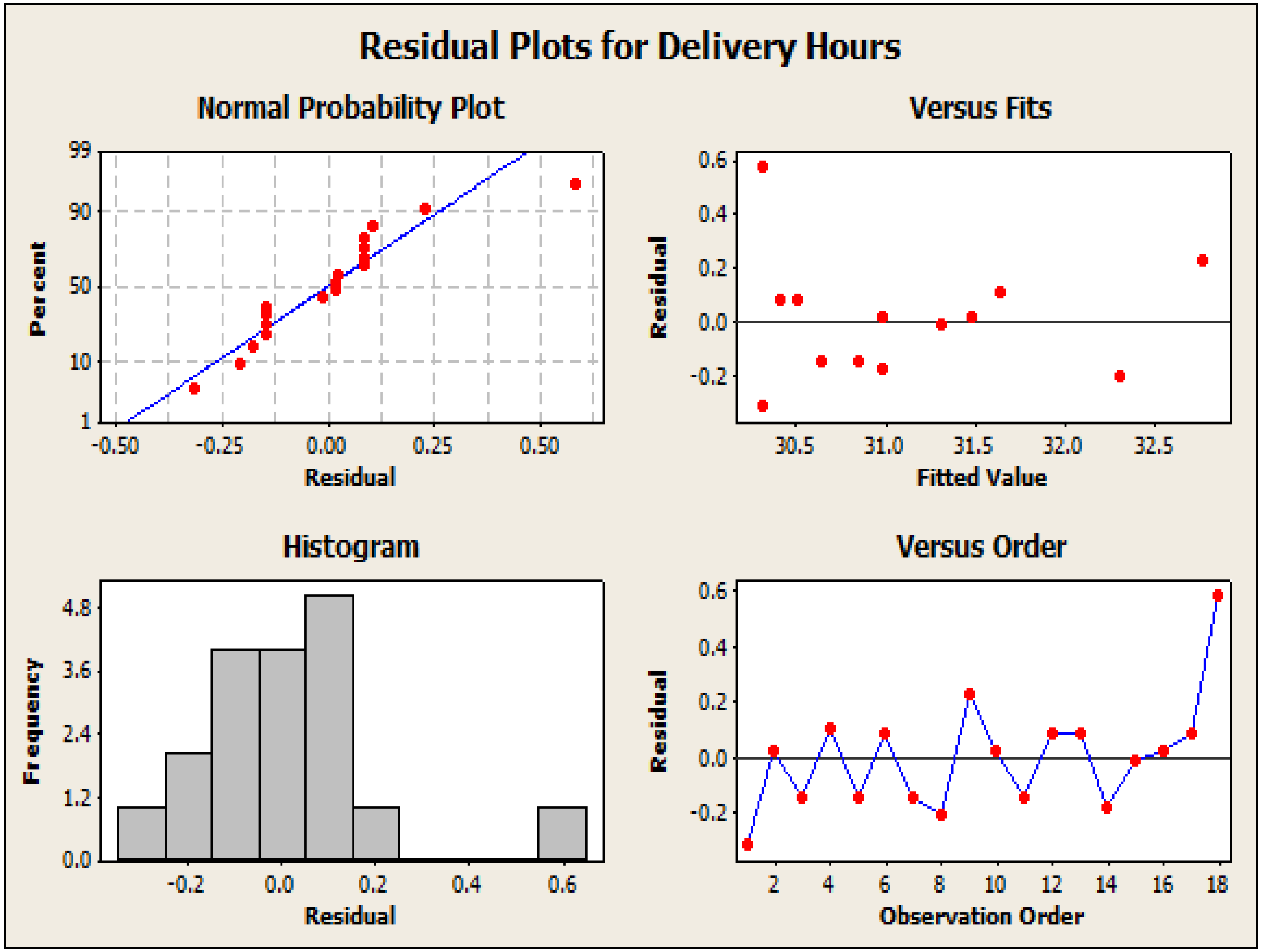
- ✓Simple linear regression to be used when there is one response variable (KPOV) and one predictor variable (KPIV) to test. The decision on how many predictor variables to test should be based on the correlation and other factors in consultation with Black Belt and process owner.
- ✓The sample case tested here is delivery hours (KPOV) and training hours (KPIV) to check if there is any correlation between the two.
- ✓The objective here is to find a correlation and percentage variability in delivery hours explained by training hours.
- ✓Training hours could have been a key take away from the fishbone diagram conducted earlier.

- ✓ Click on Stat → Regression → Regression





Graphical analysis



Simple Linear Regression (Contd.)

Results analysis

✓ **Minitab output table:**

- **Regression analysis: delivery hours versus training hours**

The regression equation is delivery hours = 35.6 - 0.663 training hours

Predictor	Coef	SE Coef	T	P
Constant	35.6249	0.3456	103.07	0.000
Training Hours	-0.66347	0.04905	-13.53	0.000

S = 0.208659 R-Sq = 92.0% R-Sq(adj) = 91.5%

PRESS = 1.00075 R-Sq(pred) = 88.45%

✓Results analysis (continued)

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	**1**	**7.9657**	**7.9657**	**182.96**	**0.000**
Residual Error	16	0.6966	0.0435		
Total	17	8.6624			

Unusual Observations

Obs	Training Hours	Delivery Hours	Fit	SE Fit	Residual	St Resid
9	4.30	33.0000	32.7720	0.1401	0.2280	1.47 X
18	8.00	30.9000	30.3172	0.0703	0.5828	2.97R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it a large leverage.

Durbin-Watson statistic = 1.50753

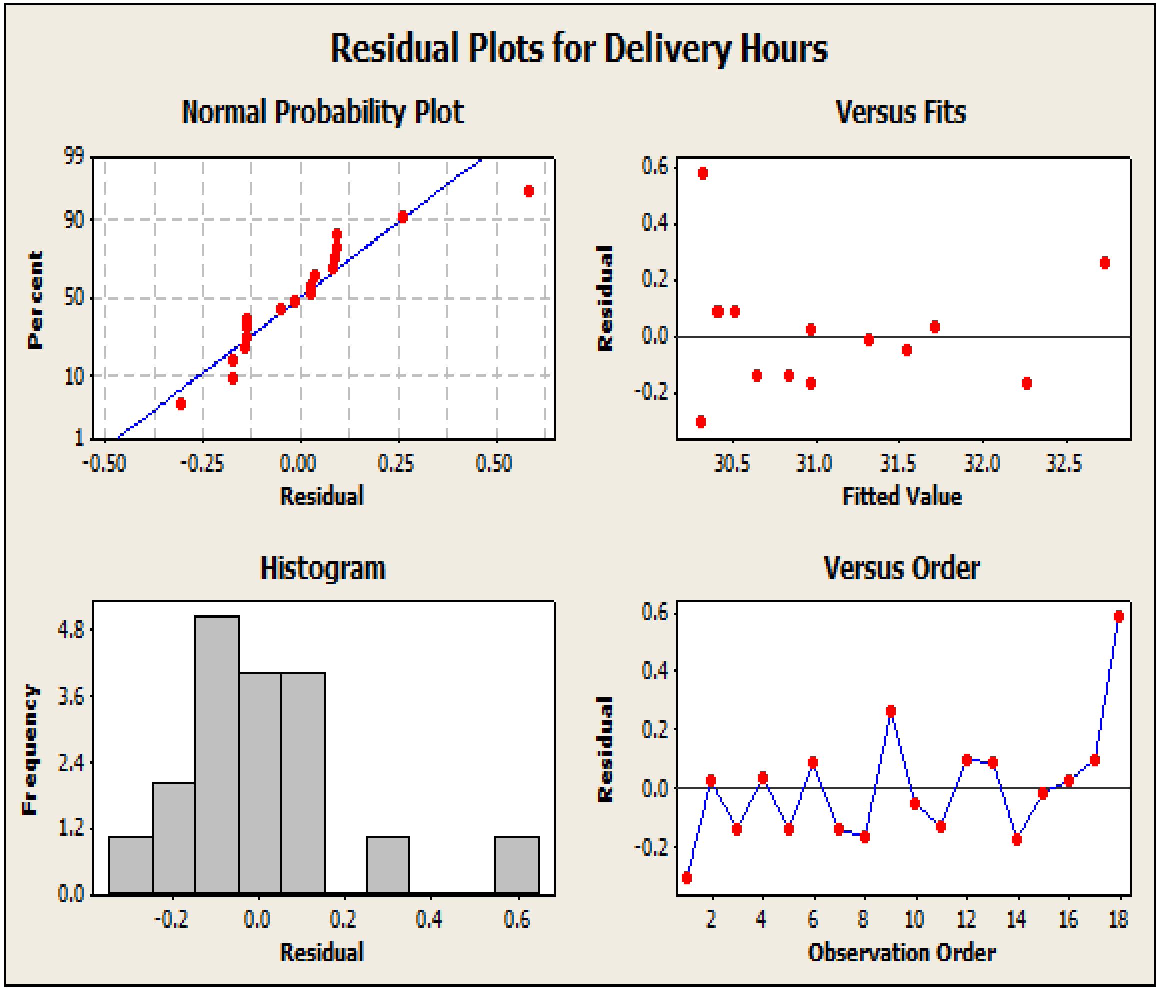
Interpreting the results

- ✓ R-square value is 92%. That is good enough R-squared value to use for regression model.
- ✓ The predicted R-squared is 88% approximately. The model can predict new responses for observations in future as well.
- ✓ The equation, **delivery hours = 35.6 - 0.663 training hours** can be used as a predicting model.
- ✓ If the current average for delivery hours is 32 hours, the current training hours average would be 5.5 hours.
- ✓ If the desired delivery hours target is 30 hours, the training hours need to be hiked up to 8.44 hours. This gives the controllable measure of where the input variable needs to be in terms of giving a controlled and desirable output.

Multiple Linear Regression

- ✓Multiple linear regression is to be used when there is one response variable (KPOV) and multiple predictor variables (KPIV) to test. Use the right set of KPIV to ensure if multiple regression analysis is effectively done by Black Belt.
- ✓The sample case tested here is delivery hours (KPOV), training hours (KPIV), and packaging weight (KPIV).
- ✓The objective here is to find a correlation and percentage variability in delivery hours explained by training hours and packaging weight.
- ✓Training hours and packaging weight could have been a key take away from the fishbone diagram conducted earlier.
- ✓The steps to do a multiple linear regression remains the same as the simple linear regression.

Output graphs



Graphical analysis

- ✓Normal probability plot → Curvature in the tails of the residuals is acceptable because the sample size is small. The residuals are thus **normally distributed**.
- ✓Residuals versus fits → The residuals versus fits graph indicates a regular up-down pattern. The residuals fall on both sides of 0. No sign of non-random variation.
- ✓Histogram → It will plot the frequency of occurrence for each of the values. Most of the values are concentrated between -0.1 to 0.1 with highest concentration around -0.1.
- ✓Residual versus order graph → This shows the data for each of the order that was observed. Most of the values are around the center and shows a random behavior, and towards the end it goes to a value of 0.6.
- ✓Reading both the normal and fit plots, one observation can be classified as an outlier. This is a red-flag, which tells us it is a possible special cause of variation and further investigation is needed.

Multiple Linear Regression (Contd.)

Output results

Regression analysis: delivery hour versus training hour and packaging weight

```
The regression equation is
Delivery Hours = 34.8 - 0.646 Training Hours + 0.063 Packaging weight in lbs
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	34.835	1.489	23.39		0.000
Training Hours	**-0.64614**	**0.05936**	**-10.89**	**0.000**	**1.400**
Packaging weight in lbs	**0.0630**	**0.1154**	**0.55**	**0.593**	**1.400**

```
S = 0.213391   R-Sq = 92.1%   R-Sq(adj) = 91.1%

PRESS = 1.13802   R-Sq(pred) = 86.86%
```

Analysis of variance

Source	DF	SS	MS	F	P
Regression	2	7.9793	3.9897	87.62	0.000
Residual Error	15	0.6830	0.0455		
Total	17	8.6624			

Source	DF	Seq SS
Training Hours	1	7.9657
Packaging weight in lbs	1	0.0136

Unusual Observations

Obs	Training Hours	Delivery Hours	Fit	SE Fit	Residual	St Resid
9	4.30	33.0000	32.7374	0.1567	0.2626	1.81X
18	8.00	30.9000	30.3151	0.0720	0.5849	2.91R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it a large leverage.

Durbin-Watson statistic = 1.46243

Multiple Linear Regression (Contd.)

Results Analysis

1. The variance inflation factor (VIF) provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity. In this example, the value is 1.4. By standard convention, this would show moderate collinearity between variables. If VIF > 10, collinearity may adversely affect the model. Right now, it is acceptable, so there is no need to remove any predictors from the model.
2. P-value for training hours is 0.0 (it is significant as it is less than 0.05) but for packaging weight it is 0.0594, which is considered non-significant (as it is more than 0.05). Packaging weight doesn't significantly impact the result on delivery hours. The Black Belt may decide to drop the packaging weight from the model.
3. The regression equation, delivery hours = 34.8 - 0.646 training hours, is modified post removing the packaging weight variable and also, by looking at the adjusted R-square of 91.1% and predicted R-square of 86.5%.

New and important terms explained

✓ Collinearity

- Collinearity refers to an exact or approximate linear relationship between two explanatory variables.
- Means some predictor variables are correlated to other predictors.
- With Minitab, results would still have high level of statistical accuracy.
- The standard errors of coefficients are relatively high for predictors and residuals.
- For multi-collinearity, identify VIF.
- Multi-collinearity is an issue because it increases the variance of regression terms and makes it unstable.

Multiple Linear Regression (Contd.)

✓**VIF (variance inflation factor)**

- VIF shows the extent to which multi-collinearity exists amongst predictors.
- If VIF <1, predictors are not correlated.
- If VIF is between 1 to 5, predictors are moderately correlated.
- If VIF is greater than 5, predictors are highly correlated.
- On detection of high correlation, the Black Belt should regress one predictor to another.

✓**Durbin Watson statistic**

- Tests for the presence of auto-correlation in residuals.
- When adjacent observations are correlated, the OLS method in regression underestimates the SE of coefficients.
- If D < Lower bound, positive correlation exists. If D > Upper bound, no correlation exists. If D is range bound, the test is inconclusive.

Multiple Linear Regression (Contd.)

✓**PRESS statistic**

- PRESS stands for predicted residuals sum of squares.
- PRESS statistic is used as an indication of the predicting power of the model, i.e., how well the model could fit new observations.
- PRESS statistic is the sum of the squared external residuals. External residuals are calculated by finding the predicted value for an observation by leaving out the observation.
- SSE (Sum Squares of Errors) explains the **quality of fit**, while PRESS statistic explains the **predicting quality**.

Best Subsets Regression and Stepwise Regression

Best subsets regression

- ✓ Identifies the best fitting regression model that can be constructed with the specified predictor variables.
- ✓ The goal of regression is achieved with the fewest predictors possible.
  - For example, if we have three predictors, Minitab will be able to show best and second best one-predictor models, and then show the full predictor model.

Stepwise regression

- ✓ In stepwise regression, the most significant variable is added or removed from the regression model.
- ✓ Three common stepwise regression procedures are standard stepwise, forward selection, and backward elimination. Use of Mallow's Cp statistic in stepwise regression is popular.

Summary

We have learned the concepts and applications of:

- ✓ Objectives of regression analysis
- ✓ Concepts of regression analysis
- ✓ Simple linear regression
- ✓ Multiple linear regression
- ✓ An introduction to best subsets regression and stepwise regression

Section IV, Lesson 5

Confidence Intervals

Agenda

- ✓Concepts of Confidence Intervals and Confidence Intervals Testing
- ✓Confidence Intervals for Difference between Two Means
- ✓Confidence Intervals Working
- ✓Confidence Intervals Impactors
- ✓Chi-Square Confidence Intervals for Variances
- ✓Z Confidence Intervals for Proportions
- ✓Chi-Square and Probability
- ✓T Distribution Confidence Intervals

Concepts of Confidence Intervals and Confidence Intervals Testing

- ✓ Confidence interval (CI) is a type of interval estimate of a population parameter and is used to indicate the reliability of an estimate.
- ✓ Population parameters are µ (population mean) and σ (population standard deviation). When we collect data as a sample, we represent the same parameters as sample statistics (Xbar as sample mean and s as sample standard deviation).
- ✓ As there is variation in every process, we will find uncertainty in these sample estimates. These sample statistics are called sample estimates because they are used to estimate the population parameter.
- ✓ Confidence intervals are known as **interval estimates** while mean, standard deviation, and other measures of descriptive statistics are known as point estimates.

Confidence intervals

- ✓Confidence intervals give an estimated range of values, known as margins or confidence interval widths, which is likely to include the parameter of the population. The likelihood is determined by the confidence levels, which is maintained the standard at 95%, but can also be tailored to 90%, or 99%.
- ✓If independent samples are taken from the population and confidence intervals are computed for them, 95% of the time these intervals will include the unknown population parameter.

Confidence intervals example

- ✓A company conducted a survey amongst a sample of 1000 households in a society of 10,000 to know what percentage of people drank cola. When the company was asked the percentage, they replied 25% ± 2%. That means, on the basis of the survey, they thought the entire population could drink cola in the range of 23% to 27% and they were 95% confident in their findings.
- ✓95% here is known as the confidence levels, and is always assumed before the sampling activity.
- ✓The range between 23% - 27% is known as margin of error, indicating the population result will lie in this range. If the sample survey result for example lies at 22%, this could be rejected as it doesn't belong to the population.

- **Statistical significance of confidence intervals**
  - A process improvement effort shows that the process yield has improved from 78% to 82%. The population yield is in the confidence limits of 84% to 88%. Has the process really made any significant improvements?
- **Solution**
  - As the sample yield is outside the range of the confidence limits defined by the intervals, the process has not made any significant improvement whatsoever. Although some improvement can be seen, the percentage needed to quantify the statistical significance of improvement is still to be reached.

Confidence Intervals for Difference between Two Means

- ✓Confidence intervals for difference between two means establishes limits in which the difference between the two means could assume values.
- ✓The sample statistic and the population parameter could be different, although central limit theorem (CLT) simplifies this analogy, but in reality, there would be an estimation or a sampling error.
- ✓This error is known as error of estimation, margin of error, or standard error.

Confidence Intervals Working

✓ **Use z-distribution confidence intervals (**For known variance and n >=30)

- Confidence Intervals = Xbar ± $Z_{\alpha/2}$ * σ/√n, where n is the sample size
- Example – At 95% confidence levels, calculate confidence intervals for sample mean of 15 and a known standard deviation of 1, with sample data collected on 100 samples.
- **Solution:** Confidence level = 95%, α = 5%, α/2 = 2.5%. (Use the Z-table provided in the toolkit.) $Z_{\alpha/2}$ = 1.96.
- Confidence intervals = 15 ± 1.96 * 1/ √100 = (15 – 0.196, 15 + 0.196) = (14.804, 15.196)
- There is a 2.5% probability each that the population mean could be less than 14.81 and more than 15.196.

Confidence Intervals Working (Contd.)

✓ **Use z-distribution confidence interval** (For known variance and n >=30)

- Confidence Interval = Xbar ± $Z\alpha/2 * \sigma/\sqrt{n}$, where n is the sample size.
- Example – The same as the previous example but with a sample size of 36.
- **Solution:** Confidence level = 95%, α = 5%, $\alpha/2$ = 2.5%. (Use the Z-table provided in the toolkit.) $Z_{\alpha/2} = 1.96$.
- Confidence intervals = 15 ± 1.96 * 1/ $\sqrt{36}$ = (15 – 0.33, 15 + 0.33) = (14.67, 15.33).
- There is a 2.5% probability each that the population mean could be less than 14.66 and more than 15.33.

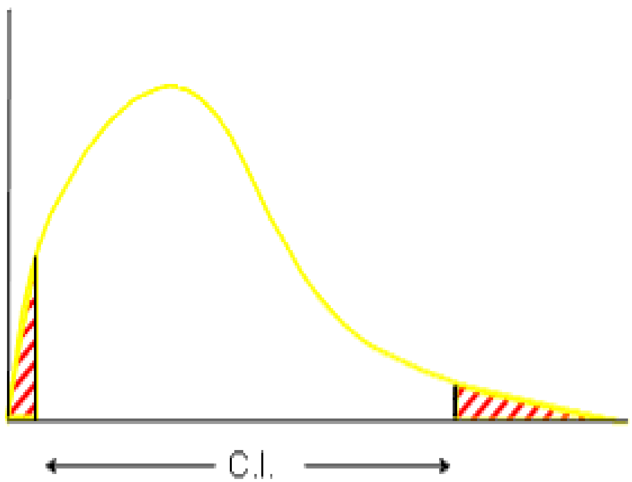
Confidence Intervals Impactors

✓Factors impacting the confidence intervals for a sample data:

- Significance level – The choice of α for the confidence interval dictates the margin of error or confidence interval width. At 99% confidence interval or α = 1%, $Z\ \alpha/2$ = 2.576. The same confidence intervals for 100 samples will now move to 14.724 and 15.296.
- Sample size – With increase in sample size at the same significance level, the confidence intervals width keep decreasing. In other words, the accuracy of the confidence intervals increases with the increase in the sample size.

Chi-Square Confidence Intervals for Variances

- ✓Chi-square confidence intervals are used to determine the confidence intervals for variances.
- ✓The interval represents the likely place where population variances will fall, given a confidence level of 95%, 90%, or 99%.
- ✓Chi-square confidence intervals = $(n-1)s^2/\chi^2_{\alpha/2,n-1} < \sigma 2 < (n-1)s^2/\chi^2_{1-\alpha/2,n-1}$



Chi-Square Confidence Intervals for Variances (Contd.)

✓ Example

- Calculate the confidence intervals for variance for sample size = 35, and a known variance of 2.5.

✓ Solution

- Assume $\alpha = 0.05 \rightarrow \alpha/2 = 0.025 \rightarrow 1- \alpha/2 = 0.975$
- Determine χ^2 values at significance levels. (Use Chi-Square table provided)
- At $\alpha/2$, χ^2 value = 51.96, and at 1- $\alpha/2$, χ^2 value = 19.806
- CI = 34 * 2.5^2/51.96, 34 * 2.5 ^2/19.806 = 4.08, 10.72
- Thus, the confidence intervals for variance are 4.08 and 10.72.

✓ Confidence intervals for proportions

- CI = $p_{bar} \pm Z_{\alpha/2} \sqrt{p_{bar} * q_{bar}/n}$, where p_{bar} is the proportion of successful outcomes.
- **Example:** Calculate the 95% confidence intervals for a class of 70, reporting 30% absenteeism.
- **Solution:** $\alpha = 5\%$, $z_{\alpha/2} = 1.96$, $p_{bar} = 0.3$, $q_{bar} = 0.7$
- CI = $0.3 \pm (1.96 * \sqrt{0.21/70}) = 0.3 \pm (1.96 * 0.05) = 0.3 \pm 0.098 = (0.2, 0.4)$
- The confidence intervals for 30% absenteeism in a class of 70 students are 20%, 40%.

Chi-Square and Probability

- ✓ The chi-square statistic can be used to determine probability of variances occurring.
- ✓ $\chi^2 = (n-1)s^2/\sigma^2$
- ✓ **Example:** For a population variance of 1 and a sample variance on data collected 0.6, determine the probability that sample variance can exceed 0.6. Sample size is 30.
- ✓ **Solution:** $\chi^2 = (n-1)s^2/\sigma^2$

  $= (30-1)0.6/1 = 29 * 0.6 = 17.4$
- ✓ For a degree of freedom 29, and calculated chi-square statistic of 17.4, the probability of having the variance greater than 0.6 is 95% - 97.5% (using chi-square table).

T Distribution Confidence Intervals

- Student's t confidence intervals are to be used when sample size is less than 30 and when the population standard deviation or population variance is unknown.
- **Example:** For a **s**ample **s**ize of 25 and a **s**ample **v**ariance of 2, calculate **c**onfidence **i**ntervals for mean of 14.
- **Solution:** Use the formula, CI = Xbar ± $t_{\alpha/2}$ * s/√n, where α is 5%. Looking up the t distribution table, t value is 2.131.
  - CI = 14 ± 2.131 * 2/ √25 = 14 ± 0.85 = (13.15, 14.85)

Summary

In this lesson we have learned:

- ✓ Concepts of confidence intervals and confidence intervals testing
- ✓ Chi-square probability distribution
- ✓ Student's t distribution
- ✓ Confidence intervals for μ when σ is known
- ✓ Confidence intervals for μ when σ is unknown

Section IV, Lesson 6

Parametric Hypothesis Testing

Agenda

- ✓ Hypothesis Testing Objectives
- ✓ Hypothesis Testing Concepts
- ✓ Null and Alternate Hypothesis
- ✓ Type I Error
- ✓ Type II Error
- ✓ Significance Level
- ✓ β and Power
- ✓ P-value, and Acceptance and Rejection Conditions
- ✓ Sample Size Determination for Tests
- ✓ 1 Sample z Test

Agenda (Contd.)

- ✓2 Sample z Test
- ✓ f-Test of Equality of Variances
- ✓1 Sample t Test
- ✓2 Sample t Test
- ✓Paired t Test
- ✓Paired t Test Interpretation
- ✓ANOVA
- ✓One Way ANOVA
- ✓Two Way ANOVA with Replication

Hypothesis Testing Objectives

- ✓ Hypothesis testing is a form of decision making based on statistical inference that uses data from a sample to draw conclusions about a population parameter.
- ✓ Hypothesis testing is usually done to:
  - Statistically validate if a sample mean does belong to the population;
  - Statistically validate if means of two groups are the same or if they are significantly different; and
  - Statistically validate if variances of two groups are the same or if they are different.
- ✓ Parametric hypothesis testing is testing on the groups of data that come from a normal distribution.

Hypothesis Testing Concepts

- ✓ Null and alternate hypothesis
- ✓ Type I error
- ✓ Type II error
- ✓ Significance level (α)
- ✓ β and power
- ✓ p – Value and acceptance and rejection conditions
- ✓ Sample size determination for tests

Null and Alternate Hypothesis

- ✓ Null hypothesis → Ho = The basic assumption behind doing any activity. For example, we go to a movie assuming it is good.
  - Null hypothesis, Ho is, movie is good.
- ✓ Alternate hypothesis → Ha = The exact opposite of the null hypothesis.
  - Alternate hypothesis, Ha is, movie is not good.
- ✓ The main objective of conducting hypothesis tests is to reject the null or alternate hypothesis. Often, rejecting the null or alternate hypothesis has practical connotations, which will be discussed when the tests are done.

Type 1 Error

- ✓Type I error is said to be committed when the null hypothesis is rejected, when it was actually true. For example, though the movie was good, we said it was bad.
- ✓The mistake is intentionally committed and hence a type 1 error.
- ✓A type 1 error is also known as "false positive". For example, a patient who undergoes an HIV test, is concluded that he carries the HIV virus, and is declared HIV +ve. But in reality, he doesn't carry the virus. That's the reason type I error is known as false positive.
- ✓Type I error is also referred to as 'producer's risk', i.e., even though the product is good, it is getting rejected.

Type II Error

✓Another error experimenters commit is known as type II error. This occurs when someone rejects the alternate when it was actually true.

✓For example, a person is declared HIV -ve, meaning he doesn't carry the virus. This is done despite the person actually carrying the virus. In reality, he should have been declared positive, but the doctor rejected the alternate and declared the patient to be free from the HIV virus.

✓**Important**

- Type I error is considered serious, because the null hypothesis is wrongly rejected. Here, the chance of going with the assumption was missed.
- In this case of the HIV virus incident, the type II error could be even more dangerous, as the patient is declared free from the disease when he is not.

Type II Error (Contd.)

- **Example:** If men having high blood sugar problems are diagnosed with diabetes with the mean blood sugar level to be at 150, and a standard deviation of 10. Any individual having blood sugar level greater than 125 can be diagnosed with diabetes, what is the probability of committing a type II error?
- **Solution:**
  - Z = (150-125)/10 = 25/10 = 2.5
  - Please refer to the Z-Table. Z=2.5 corresponds to 0.0062. Thus, the probability of committing a type II error, β, is 0.62%.

Significance Level (α)

- ✓The probability of committing a type I error is known as the level of significance or the significance level, represented by α.
- ✓In the case of α, one-tailed probability (only one direction is considered extreme) is used to reject the hypothesis.
- ✓α = 100% - confidence level. If confidence level is 95%, level of significance is 5%. 95% confidence level is often assumed as the default confidence level.
- ✓The confidence level could be changed depending on the type of experiment done, in which case possible α values could be 1% or 10%.

Significance Level (α) (Contd.)

- **Example:** If the sample data on delivery hours has a mean of 36 hours and a standard deviation of 2, with delivery hours being normally distributed, what is the probability of a type I error with delivery hours over 40 hours being diagnosed as defective?
- **Solution:**
  - $Z = (x - \mu)/\sigma = (40-36)/2 = 2$
  - Area under 2 with the help of the z table is 0.0228, which corresponds to 2.28%.
  - Thus, probability of a type I error, level of significance is 2.28%.

β and Power

- ✓β is the probability of committing a type II error by rejecting the alternate when it was actually true. It is like concluding a test on two group means as producing non-significant results, when the results were actually significant.
- ✓Power is the fraction of experiments which is expected to **yield a statistically significant p-value.** In other words, power shows the confidence in the test and its results.
- ✓Usual power value is 80%, i.e., out of the 100 experiments 80 of them can be expected to show statistically significant p-values.
- ✓β = 100% - power%.

P-Value, and Acceptance and Rejection Conditions

- ✓ If p-value is less than α, reject the null hypothesis, i.e., conclude with 1-p value level of confidence that there is a significant statistical difference in the two groups.
- ✓ If p-value is greater than α, reject the alternate hypothesis, i.e., conclude that there is no significant statistical difference in the two groups with p-value level of confidence.
- ✓ By accepting or rejecting the hypothesis, it can be found whether the mean has shifted or if the variations really belong to the group or not.

Sample Size Determination for Tests

Sample size for proportion based tests

✓ $n = (t^2 * p * (1-p))/ME^2$

✓ t = Standard value at 5% significance → 1.96

✓ p = Given proportion

✓ ME = Margin of error → 5% for 95% confidence

✓ For a given proportion of 30%, the sample size is

- $n = ((1.96)^2 * 0.3 * (0.7)^2)/(0.05)^2$
- n = 322.72 = 323

✓ This sample size consideration and calculations can be used by the Black Belt also in the measure phase.

1 Sample z Test

- ✓ Used when a hypothecated mean has to be compared with the mean of a sample or population.
    - Ho = Mean belongs to the sample.
    - Ha = Mean does not belong to the sample.
- ✓ To be used when sample size > 30 and when the standard deviation of population is known.
- ✓ **Example:** A sample of 35 was taken for measuring delivery hours. The mean Xbar was found to be 32.5 and population standard deviation was found at 2.2. What is the probability that a mean of 34 could belong to this sample?

1 Sample z Test (Contd.)

- ✓ A straightforward confidence interval testing could have given the result as (31.77, 33.229). As the hypothecated mean is outside the interval range, the null hypothesis should be rejected.

- ✓ **Minitab results**
  - Test of mu = 34 vs. not = 34
  - The assumed standard deviation = 2.2

N	Mean	SE Mean	95% CI	Z	P
35	32.500	0.372	(31.771, 33.229)	-4.03	0.000

- ✓ **Interpretations**
  - p-value of 0 indicates the rejection of the null with 100% confidence → The hypothecated mean of 34 just cannot belong to the sample → There is no way that the mean of 34 happened by chance, as something special might have contributed to it.

1 Sample z Test (Contd.)

- Testing for an upper-bound confidence interval for the same example. The confidence intervals have now been changed to 32.97.

- **Minitab results**
  - Test of mu = 34 vs. < 34
  - The assumed standard deviation = 2.2

			90% Upper		
N	Mean	SE Mean	Bound	Z	P
35	32.500	0.372	32.977	- 4.03	0.000

- **Interpretations**
  - p-value of 0 indicates the rejection of the null with 100% confidence → The hypothecated mean of 34 just cannot belong to the sample → There is no way that the mean of 34 happened by chance, as something special might have contributed to it.

2 Sample z test

- ✓Same conditions apply as of 1 sample z test. The difference is that a 2 sample z test, tests the means of two groups, and returns a significant or a non-significant p-value.
- ✓With a 2-sample z test we can compare the means of two groups and conclude if they are statistically the same.
- ✓Ho → $\mu_a = \mu_b$
- ✓Ha → $\mu_a \neq \mu_b$
- ✓The above hypothesis is assumed for a 2-tailed probability testing.

f-Test of Equality of Variances

Using QI macros (free version), an f-test equal variances check is done on two groups of delivery hours data.

Data1	Data2
32	32
33	33
33.5	34
33.6	33
32.6	33.5
36	35.5

F-Test Two-Sample for Variances	a	0.05					
	Data1	Data2					
Mean	33.45	33.5					
Variance	1.911	1.4					
Observations	6	6					
df	5	5					
F	1.37						
P(F<=f) one-tail	0.371	0.741	Two-tail				
F Critical one-tail	5.05	7.15	Two-tail				
One-tail	Accept Null Hypothesis because p > 0.05 (Variances are the same)						
Two-tail	Accept Null Hypothesis because p > 0.05 (Variances are the same)						

1 Sample t Test

- ✓Used when a hypothecated mean can be compared with the mean of a sample or population.
- ✓Ho = Mean belongs to the sample
- ✓Ha = Mean does not belong to the sample
- ✓To be used when sample size < 30 and standard deviation of population is unknown.
- ✓**Example:** A sample of 25 was taken for measuring delivery hours. The mean Xbar was found to be 32.5 and sample standard deviation was found at 2. What is the probability that a mean of 34 could belong to this sample?

1 Sample t Test (Contd.)

Minitab results:

✓ **One-sample T**

- Test of mu = 34 vs. not = 34

N	Mean	StDev	SE Mean	95% CI	T	P
25	32.500	2.000	0.400	(31.674, 33.326)	-3.75	0.001

✓The confidence intervals will reject the null, and the p-value of 0.001 will reject the null hypothesis.

✓This mean of 34 doesn’t represent the population drawn → This couldn’t be by chance.

2 Sample t Test

- Same conditions of 1 sample t test apply. The difference is that a 2 sample t test tests the means of two groups, and returns a significant or a non-significant p-value.
- With a 2-sample t test the means of two groups are compared and concluded if they are statistically the same.
- Ho → $\mu_a = \mu_b$
- Ha → $\mu_a \neq \mu_b$
- The above hypothesis is assumed for a 2-tailed probability testing.
- An assumption for conducting a 2-sample t test is whether the two groups have equal variances or not. To do this, a **homogeneity of variance** test has to be conducted.

2 Sample t Test (Contd.)

- ✓The two-tailed probability value is 0.741.
- ✓According to the interpretation, since the p-value is non-significant it has to be accepted that two groups have equal variances. So work with the t-test assuming equal variances. For the same group of data, the t-test results are shown below:

t-test: Two-sample assuming equal variances	a	0.05
Equal sample sizes		
	Data1	**Data2**
Mean	33.45	33.5
Variance	1.911	1.4
Observations	6	6
Pooled variance	1.6555	
Hypothesized mean difference	0	
df	10	
t Stat	-0.067	
P(T<=t) one-tail	0.474	
T Critical one-tail	1.812	
P(T<=t) two-tail	0.948	
T critical two-tail	2.228	

2 Sample t Test (Contd.)

- ✓Assuming equal variances, with the two tailed probability value showing as 0.948, the alternate can be rejected.
- ✓With 94.8% confidence, it can be said that the means of the two groups are equal.
- ✓This means there is no statistical difference between the means of the two samples → Both the means seem to come from the same population → There is no indication of non-random variation in the two groups.
- ✓The df, degree of freedom indicated here is 10, $n_1 + n_2 - 2$.

Paired t Test

- ✓The paired t test is a useful 'before-after' test, which is used to validate statistical improvements.
- ✓This test is often used in the improve stage to statistically validate improvements.
- ✓The groups of data must be correlated. Paired t test is not done on independent samples.
- ✓The degrees of freedom would now be computed on one group and not two.

Paired t Test (Contd.)

Before DH	After DH
32	26
33	27
33.5	26.5
33.6	28
32.6	29
36	28

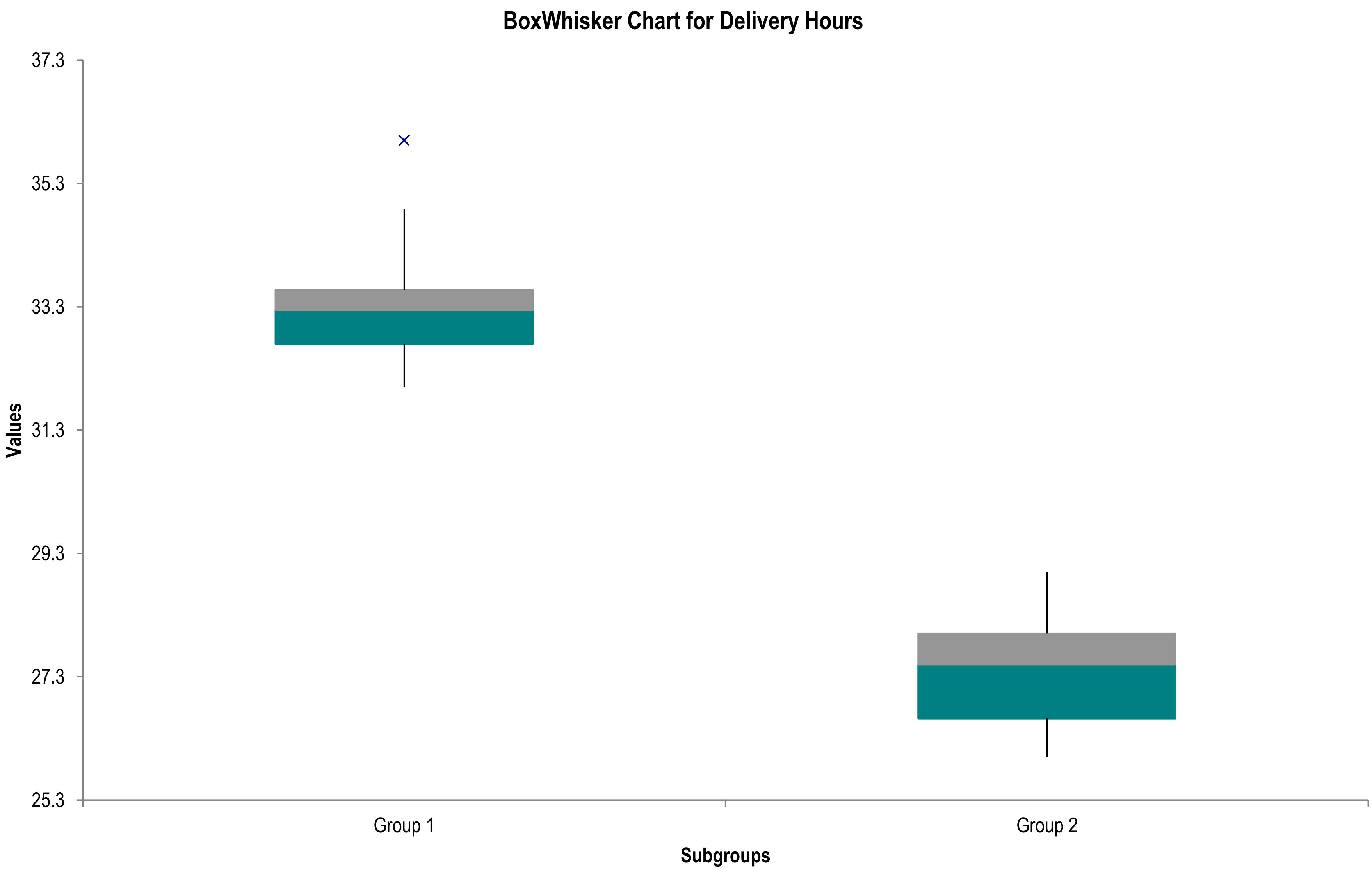
t-Test: Paired Two Sample for Means	a	0.05
	Data1	Data2
Mean	33.45	27.41667
Variance	1.911	1.241667
Observations	6	6
Pearson Correlation	0.314854	
Hypothesized Mean Difference	0	
df	5	
t Stat	10.003	
P(T<=t) one-tail	0.000	
T Critical one-tail	2.015	
P(T<=t) two-tail	0.000	
T Critical Two-tail	2.571	

We can reject the null hypothesis because the p-value is significant. The confidence in rejecting the null is 100%.

Paired t Test Interpretation

- ✓Ho = There is no significant statistical difference between the means of the two groups of data.
- ✓Ha = There is a significant difference between the means of the two groups of data.
- ✓As $p < 0.05$, reject the null.
- ✓The means are significantly different. Group before could be less or greater than group after.
- ✓To check if the improvement has indeed worked, after group mean has to be significantly less than the before group data.
- ✓Click on box plots on QI macros option for a graphical display.

Paired t Test (Contd.)



Paired t Test (Contd.)

Interpretations:

- ✓The improvement is positive.
- ✓The improvement given to the group has worked.
- ✓The working of the improvement has been statistically validated.
- ✓These improvements that have been implemented on a pilot could now be implemented across the organization.
- ✓The degree of freedom here for a paired t test is 5, while in the 2-sample t test, the degree of freedom was 10.

ANOVA

- ✓ANOVA stands for analysis of variance.
- ✓ANOVA could be an excellent testing procedure to test more than 2 groups at the same significance level.
- ✓Using ANOVA, the Black Belt would be able to accept or reject the null hypothesis just by doing one test.
- ✓The three or more groups of data should be independent. Their value should not be dependent on other group's value.
- ✓Although ANOVA analysis is for variance, it also tests for means.
- ✓Ho = There is no significant statistical difference between the means of the three groups.

One-Way ANOVA

We will test the delivery hours metric across three independent samples of data using a one-way ANOVA test in QI macros.

Data1	Data2	Data3
32	32	33
33	34	33.1
33.5	33.5	33.2
33.6	33.6	33.5
32.6	32.6	35
36	36	35.5

Anova: Single Factor	a	0.05						
SUMMARY								
Groups	Count	Sum	Average	Variance				
Data1	6	200.7	33.45	1.911				
Data2	6	201.7	33.61667	1.897667				
Data3	6	203.3	33.88333	1.173667				
ANOVA					Accept Null Hypothesis because p > 0.05 (Means are the same)			
Source of Variation	SS	df	MS	F	P-Value	F crit		
Between Groups	0.573333	2	0.286667	0.17261	0.843	3.68232		
Within Groups	24.91167	15	1.660778					
Total	25.485	17						

The p-value from the one way ANOVA is 0.843, non significant, thus, we will reject alternate hypothesis. We accept the fact that there is no significant statistical difference between the means of the three groups.

Two-Way ANOVA with Replication

- ✓The two-way ANOVA test is used to test interactions in an experiment.
- ✓The significance level is maintained at $\alpha = 0.05$.
- ✓Data formatting needs to be done properly to use this test and present interpretations.
- ✓QI macros or Minitab can be used to present results.

✓**Example:** A company studies the interaction between multiple drugs by using different quantity of each in the category. The readings are presented below in the excel sheet. Conduct a two-way ANOVA with replication and interpret.

Category	Drug 1	Drug 2	Drug 3
Category 1	8	10	8
	4	8	6
	0	6	4
Category 2	14	4	15
	10	2	12
	6	0	9

Two-Way ANOVA with Replication (Contd.)

Anova: Two Factor With Replication	a	0.05				
SUMMARY	Drug 1	Drug 2	Drug 3	Total		
Category 1						
Count	3	3	3	9		
Sum	12	24	18	54		
Average	4	8	6	6		
Variance	16	4	4	9		
Category 2						
Count	3	3	3	9		
Sum	30	6	36	72		
Average	10	2	12	8		
Variance	16	4	9	28.25		
Total						
Count	6	6	6	18		
Sum	42	30	54	126		
Average	7	5	9	7		
Variance	23.6	14	16	18.58824		
ANOVA						
Source of Variation	SS	df	MS	F	P-Value	F crit
Sample	18	1	18	2.037736	0.179	4.747225
Columns	48	2	24	2.716981	0.106	3.885294
Interaction	144	2	72	8.150943	0.006	3.885294
Within	106	12	8.833333			
Total	316	17				

1

2

3

Two-Way ANOVA with Replication (Contd.)

✓ **Interpretations:**

1. There is no effect of the category of patients who use the drug on the diff. rates.
2. There is no effect of the type of drug used by patients on the diff. rates.
3. The interaction between the type of drug used and the category of the patients have a significant effect, shown by the low p-value of 0.006.

Summary

In this lesson we have learned:

- ✓Hypothesis testing objectives
- ✓Hypothesis testing concepts
- ✓1 Sample z test and 2 sample z test
- ✓f-Test of equality of variances
- ✓1 sample t test, 2 sample t test
- ✓Paired t test
- ✓One way ANOVA, Two way ANOVA

In the toolkit, the file name, Hypothesis Tests ,will have the workings of all these tests. Use QI macros to conduct these tests.

Section IV, Lesson 7

Nonparametric Hypothesis Testing

Agenda

- ✓ Nonparametric Testing Conditions
- ✓ Mann-Whitney Test
- ✓ 1 Sample Sign
- ✓ Wilcoxon Test
- ✓ Kruskal Wallis
- ✓ Mood's Median
- ✓ Friedman ANOVA

Nonparametric Testing Conditions

✓Use nonparametric tests instead of parametric tests when data is:

- Counts or frequencies of different types
- Measured on nominal or ordinal scale
- Not meeting assumptions of parametric test
- A small sample

Nonparametric tests correspond to parametric tests, which are easier to comprehend.

Nonparametric (distribution free)	Parametric (normal distribution)
1-sample sign	1-sample z-test, 1-sample t-test
Wilcoxon	1-sample, 2-sample t-test
Mann-Whitney	2-sample t-test
Kruskal-Wallis	one-way ANOVA
Mood's Median	one-way ANOVA
Friedman	2-way ANOVA paired sign test

Mann-Whitney Test

- ✓ Use the Mann-Whitney test, if the data meets nonparametric testing conditions and is divided into two independent samples.
- ✓ Rejection and acceptance conditions remain the same, i.e., $p < \alpha$, reject null, else reject alternate.
- ✓ α is set by default at 0.05.

Use the tool nonparametric testing in the toolkit to know how nonparametric tests work.

Mann-Whitney Test (Contd.)

Sample1	Sample2
31	30.5
31	30.75
31	30.85
31	31
31	31
31	31
31	31
31	31
31	31
31	31
30.5	31

Rank1	Rank2
13.50	1.50
13.50	3.00
13.50	4.00
13.50	13.50
13.50	13.50
13.50	13.50
13.50	13.50
13.50	13.50
13.50	13.50
13.50	13.50
1.50	13.50

50.5	**U1**
70.5	**U2**
50.5	**U**
126.5	**E(U1)**
126.5	**E(U2)**
60.5	**E(U)**
15.22881	**s**
96.65208	**Action(L)**
156.34792	**Action(U)**
0.05	**a**
0.6566501	**z**
0.51	*p*

Accept Null Hypothesis at alpha=0.05

As seen, on updating the data sheet with values, test results are done automatically. Here, null has to be accepted, which means, no differences in the medians of the two groups.

- Use this test instead of a 1 sample t test. Ho = The hypothecated or assumed median of sample belongs to the population by testing its median.

31.0
31.0
30.5
30.8
31.0
30.8
30.9
30.8

H0 - Median	30.9
Sample Size (n=)	8
Number of +	3
p (equal)	0.5
α	0.05
95% Confidence	1
Lower Limit	30.75
Upper Limit	31
30.75<=30.9<=31Accept Null	

95% CI

Hypothecated median of 30.9 is a part of the population indeed.

Wilcoxon Sign Rank Test

- ✓The Wilcoxon sign rank test can be used as a substitute for the 2-sample t test or even in situations where the Black Belt wishes to regress an input variable and an output variable.
- ✓For the data presented in the nonparametric testing sheet (part of the toolkit), the interpretation is again done on p-value.

T	184
n=	20
σ{T}	53.57238
α	0.05
Action(L)	-105.0
Action(U)	105.0
z	3.4
Reject Null at 0.05	
p	0.001

Reject Null

Kruskal Wallis

- ✓Kruskal Wallis test should be used when it is to be checked and understood whether the groups of data have the same variance.
- ✓Chi-square distribution is used in Kruskal Wallis to understand the variances, unlike other tests, which used z-distributions to calculate the test statistic.
- ✓For the data in the Nonparametric Testing Sheet, the results are shown below:

Total A	Total B	Total C	H	*p*
28	37	55	3.78	0.151
253.5	254.4	265.5	**Median**	

Reject the alternate

Mood's Median

- The Mood's median test is a nonparametric test that is used to test the equality of medians from two or more different populations.
- The Mood's median test works when the Y variable is continuous, discrete-ordinal or discrete-count, and when the X variable is discrete with two or more attributes.
- The test involves:
  - Finding the median of the combined data set;
  - Finding the number of values in each sample greater than the median and form a contingency table:

	A	B	C	Total
Greater than the median				
Less than or equal to the median				

  - Finding the expected value for each cell: $\text{expected} = \dfrac{\text{row total} \times \text{column total}}{\text{grand total}}$
  - Finding the chi-square value from: $\chi^2_{\text{contribution}} = \dfrac{(\text{actual} - \text{expected})^2}{\text{expected}}$

Friedman ANOVA

- ✓The Friedman test is a form of nonparametric test, that makes no assumptions about the specific shape of the population from which the sample is drawn, allowing for smaller sample data sets to be analyzed.
- ✓Unlike ANOVA, the Friedman test does not require the dataset be randomly sampled from normally distributed populations with equal variances.
- ✓The Friedman test uses a two-tailed hypothesis test where the null hypothesis is such that the population medians of each treatment are statistically identical to the rest of the group.

Friedman ANOVA (Contd.)

- ✓ Consider the example where three treatments are evaluated on four patients.

Therapy	Andrew	Belinda	Chris	Dave
Relaxed	110	140	100	130
Normal	115	150	105	135
High Intensity	117	155	100	135

- ✓ The test involves:
  - Impose ranks on each of the columns. If values are equal, average the ranks they would have got if they were slightly different:

Therapy	Andrew	Belinda	Chris	Dave
Relaxed	1	1	1.5	1
Normal	2	2	3	2.5
High Intensity	3	3	1.5	2.5

  - Calculate the F_r statistic using the formula:
  - Where:
    - I is the number of samples
    - J is the number of blocks
    - R_i is the sum of the ranks in row 'I'

$$F_r = \frac{12}{IJ(I+1)} \sum_{i=1}^{I} R_i^2 - 3J(I+1)$$

Summary

In this lesson, we have learned how to do:

✓Mann-Whitney test

✓1 sample sign test

✓Wilcoxon sign rank test

✓Kruskal Wallis test

✓Mood's median

✓Friedman ANOVA

Important : A Black Belt need not know the formulas that are used to calculate the test statistic for each of the non-parametric tests. But the Black Belt needs to know which test to use when, and how to use it.

Section IV, Lesson 8

Analyze Additionals – Categorical Data and Current Reality Tree

Agenda

- ✓Categorical Data Analysis
- ✓Current Reality Tree

Categorical Data Analysis

- ✓Analyzing continuous or discrete data is not a major issue as the Black Belt has a wide variety of hypothesis tests and confidence intervals to determine the test values.
- ✓With data that can be represented in the nominal scale, or which can be represented as a fixed number of nominal categories, categorical data analysis needs to be done differently.
- ✓The most popular form of analyzing categorical data is with the help of 2 * 2 contingency table.
- ✓Chi-square distribution is popularly used in analyzing categorical data.

Example:

A sample of teenage dieting and non-dieting folks were studied across men and women population. Rather than representing the data in form of frequency on classes, the frequency was represented on categories. Data representation is as below:

	Men	Women
Dieting	10	27
Non-dieting	20	19

Test to see if the difference in proportions of dieting/non-dieting folks is significant.

Categorical Data Analysis (Contd.)

The test we use is known as Fisher's exact test.

	Men	Women	Total
Dieting	10	27	37
Non-Dieting	20	19	39
Total	30	46	76

Fishers			Men Expected	Women Expected	
p 2-Tail	0.036935		14.60526	22.39474	
Chi-Sq	4.675244		15.39474	23.60526	
p	0.0306				

The p-value is significant at α = 0.05. Thus, the Black Belt will have to reject the null hypothesis. The Black Belt will then conclude that there is a significant difference in the proportions of dieting people in men versus women.

Categorical Data Analysis (Contd.)

✓Fisher's test done in the previous slide is always preferred when the degree of freedom is 1. Fishers test sheet is presented in the toolkit.

✓In most other cases, the chi-square goodness of fit test is used.

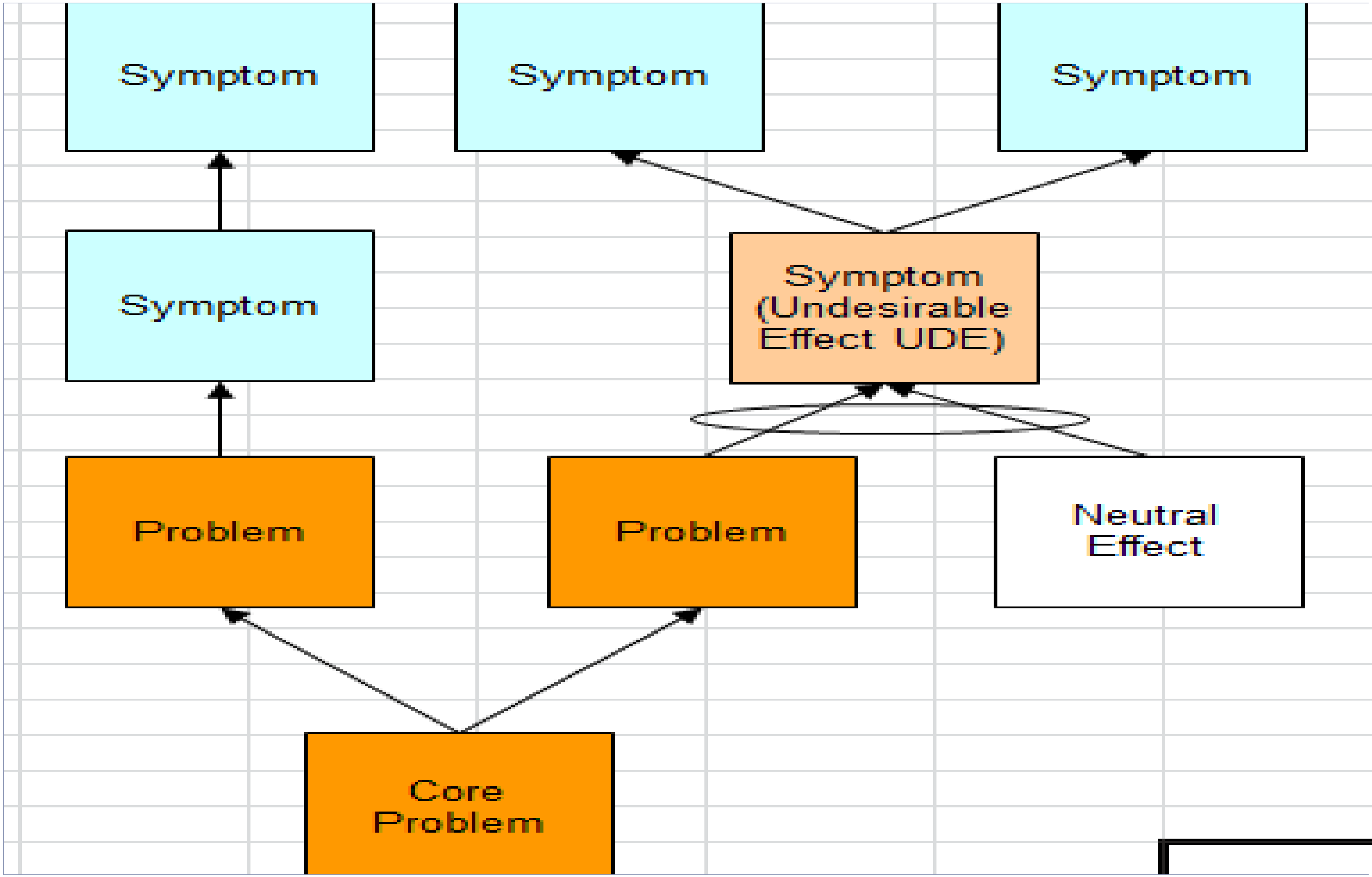
✓Other tests used are:

- Yates correction for continuity;
- Cochran-Mantel-Haenzel test;
- McNemar's test;
- Portmanteau test; and
- Likelihood test for statistical modeling.

Current Reality Tree

- ✓ Current reality is an excellent root cause identification tool.
- ✓ This tree helps plot the problem and why the problem really happens.
- ✓ The tree is categorized into two groups – Symptoms and problems. An undesirable effect (UDE) is a symptom that results in multiple symptoms happening.
- ✓ Symptoms are effects that happen and are visible to the naked eye.
- ✓ A snapshot of the current reality tree is attached in the next slide. The tool for current reality tree is also attached in the toolkit file, TOC tools.

Current Reality Tree (Contd.)



Summary

In this brief lesson, we have learned:

- ✓ How to deal with categorical data.
- ✓ How to do Fisher's Test, one of the tests that can be done to deal with categorical data.
- ✓ A list of other tests that can be used to deal with categorical data. Most of these tests are outside of the Black Belt body of knowledge.

Important: Black Belt should know about these tests.

Activity Summary – Analyze

Below mentioned activities should be performed in chronological order in the analyze phase.

1. Check the lean status of the process.
2. If wastes are analyzed, eliminate or reduce.
3. Re-check process conditions. If no improvement, proceed.
4. Brainstorm.
5. Map what is causing variation in output variable.
6. Test the relationships between input and output variable.
7. Hypothesis or confidence intervals test the characteristics.
8. Understand the root cause of the variation.

Tools Summary – Analyze

Below mentioned tools should be performed in chronological order in the analyze phase:

1. Value stream map
2. Spaghetti chart
3. Lean tools
4. Fishbone diagram
5. Cause and effect matrix
6. Regression
7. Confidence intervals
8. Hypothesis tests
9. Current reality tree and fishbone

Quiz

1. What cause of variation are unusual and happens once in a while?

a) Special cause of variation
b) Common cause of variation
c) Rare cause of variation
d) Non-value add variation

2. What process mapping tool can be used to make wastes and redundant activities obvious to the naked eye?

a) Cause and effect diagram
b) Spaghetti charts
c) Waste reduction tool
d) FMEA

Quiz

3. The tool can be used to find all the probable cause of a problem.

a) FMEA
b) Process map
c) Scatter plot
d) Cause and effect diagram

4. What type of Error is said to be committed when you reject the null hypothesis when it was actually true?

a) Human error
b) Special cause of variation
c) Type I error
d) Type II error

Answers

1. a) Special cause of variation. It occurs once in a while and cannot be predicted.
2. b) Spaghetti charts. When you do any process mapping using spaghetti charts, it visually gives you a good indication if the process flow is messy or takes unnecessary cycles, etc.
3. d) Cause and effect diagram a.k.a fishbone diagram. It helps you map all the probable causes which leads to an effect or an problem.
4. c) Type I error. You commit a **Type I Error**, when you reject a null hypothesis when it is actually true. You commit a **Type II Error,** when you reject the alternate hypotheses when it was actually true.

Thank you