

MINITAB 17 ESSENTIAL TRAINING





MINITAB 17 Essential contents

DAY 1:

- Background
- Hypothesis test
- ANOVA

DAY 1

- Regression
- Design of Experiments (DOE)
- Taguchi Method





- Minitab® 17 is the leading statistical software for quality improvement and statistics education worldwide. Thousands of distinguished organizations rely on Minitab 17's powerful capabilities and ease of use, including American Express, Pfizer, Coca-Cola, Microsoft, eBay and IBM. The world trusts Minitab for quality.
- For more information on Minitab Statistical Software, visit: www.minitab.com.





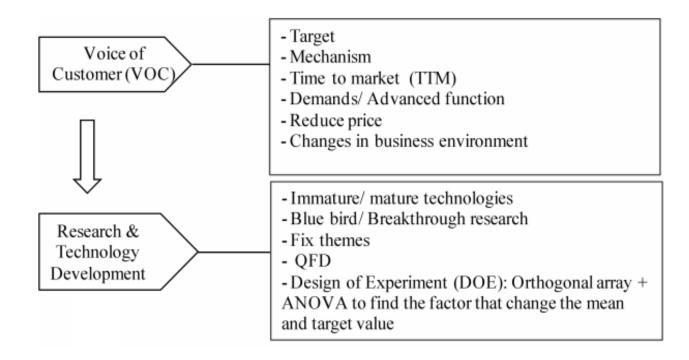
- Pennsylvania State University 1972
- To help professors teach statistics using computers so students could focus on learning statistical concepts rather than on performing manual calculations



- Minitab is designed for statistical analysis and Minitab vary in user friendliness, output presentations, flexibility, and types of analysis available.
- Quality Improvement Expertise
 - Free technical support
 - Convenient training (On-site training & Public training)
 - Resources for success



Eg.: Tools' Flow





- Strategy - Failure Mode and Effects Analysis (FMEA) Product - Product and Design characteristic Development - Reduce cost and TTM - Technology readiness Developmen Robust Design: Design Optimization to diminish t Stage variation and degradation of output - Off-line Quality Control - Taguchi Method to reduce variability around target value and quantify the loss - Test and validation - Reliability method - QE promotions - Eight KFS: a. Involvement of top management b. Responsibility c. Promotion activities d. Engineers' training e. Clarify themes f. Continuous meeting g. Internal / External consultant / expert h. Result clarification (Barometer of success) Design of Experiment (DOE): Orthogonal array + ANOVA - Problem occurence - to find interactions between designed factors - Statistical Process Control (SPC) Manufacturing - To find the factors that change the mean value - Total Productive Maintenance (TPM) After-sales Sales service

Rozzeta Dolah, Zenichi Miyagi, Kazuo Tatebayashi, Management and Production Engineering Review. Volume 3, Issue 4, Pages 26–34, ISSN (Online),

DOI: <u>10.2478/v10270-012-0031-z</u>, January 2013



AGENDA

Overview of MINITAB

Introduction to Minitab

Hypothesis Tests: Continuous Data

Hypothesis Tests: Attribute Data

Equivalence Testing (Optional)

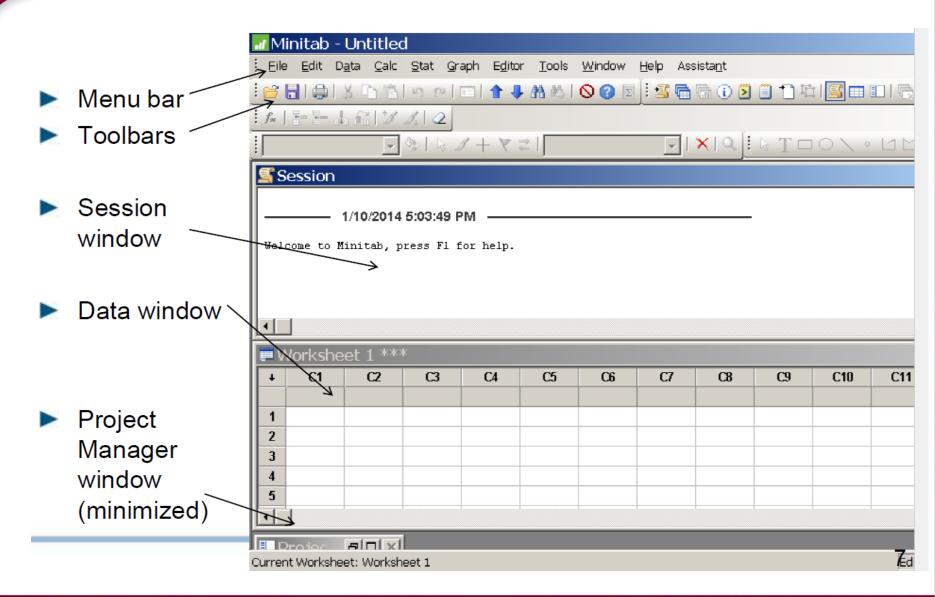
Analysis of Variance

Correlation and Simple Regression

Multiple Correlation and Regression



Opening a New Project

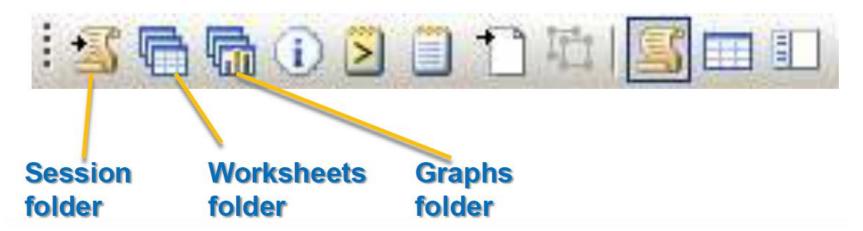




Minitab 16



Minitab 17





Open & Save

Open Project

- Choose File > Open Project > xxx.MPJ
- Click Open.

Open worksheet

- Choose File > Open Worksheet> xxx.MTW
- Click Open.

Save

- File > Save Project as
- File > Save Current worksheet as
- Append To report > ReportPad > Save Report As...
- Send to Microsoft Word
- Send to Microsoft PowerPoint



Column with Column with Column with date/time data numeric data text data Column Name Production.MT

W ***

W *** C1-D C3-T C4-T **C5** DownTime Line Shift Reason Date Row 1/1/2007 40 A Change Over number 1/1/2007 20 IA. Fixture Alignment 1/1/2007 24 A. Recalibrate: 3 <u>1/2/2</u>007 35 A Change Over Cell 1/3/2007 32 | A | 5 Change Over Final Test Missing 6 Missing 17672007 30 I A Change Over data 1/6/2007 40 IA Change Over R



INTRODUCTION TO MINITAB



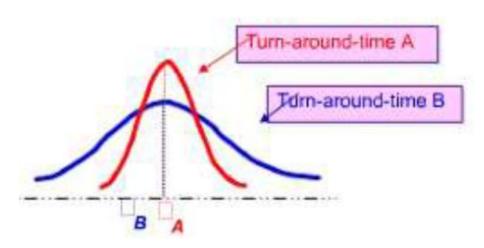
Central Tendency & Dispersion

- We would be studying a few of the important Statistics such as:
 - Statistics which defines the location:
 - Mean, Median, Mode
 - Statistics which defines the dispersion:
 - Range, Variance, Standard Deviation, Quartile



Dispersion

- A measure of dispersion / variability / spread is a numerical value to describe the amount of spread, dispersion, variability or variation in the data
- Data that are grouped closely to the mean will have <u>small</u> <u>value</u> in terms of variation (ex A)
- Data that are distributed more widely about the mean will have <u>large value</u> of variation (ex B)



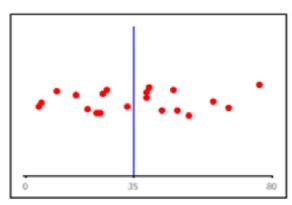


STDEV

- The most common measure of dispersion, or how spread out the data are from the mean.
- Standard deviation roughly estimates the "average" distance of the individual observations from the mean.
- The greater the standard deviation, the greater the spread in the data.

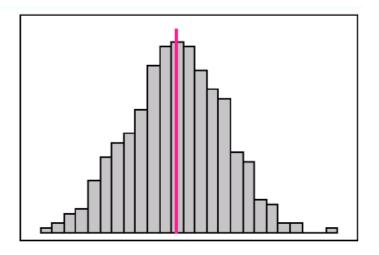
$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$





Histogram

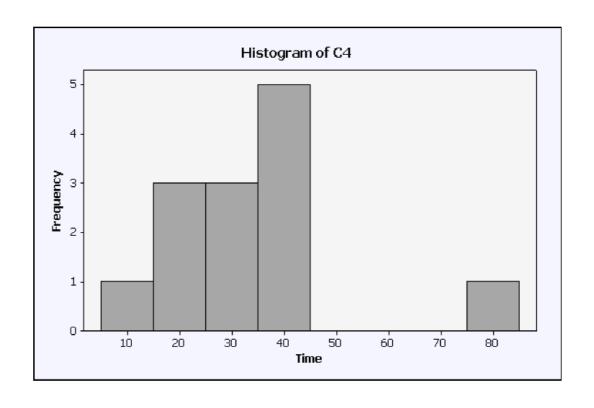


- A histogram is a good way to evaluate the center, spread, and general shape of a distribution – as long as an appropriate number of observations are available.
- Unlike bar graph, it can show distribution of continuous data.
- If you have too few data points, the shape of the histogram may not accurately represent the distribution of the population.

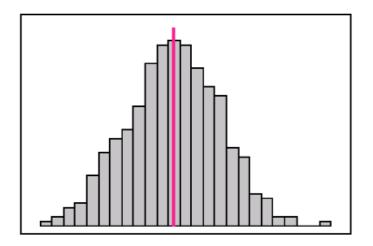


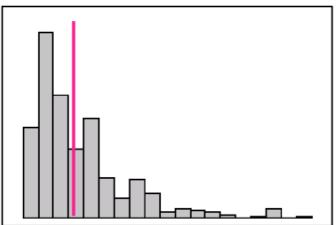
15	18	22	33	36	9	
34	27	41	75	40	39	41

Class Interval Time (min)	Frequency
5-15	I
15-25	III
25-35	III
35-45	TIH
45-55	
55-65	
65-75	
75-85	I







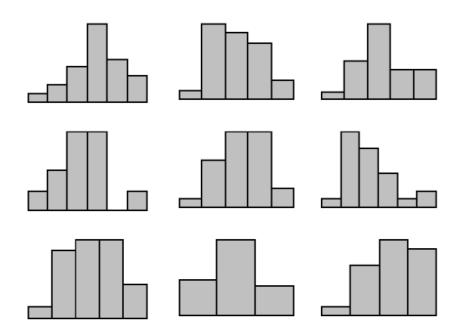


Symmetric data: the mean (pink line) lies near the center of the distribution, making it a good representation of the center.

Skewed data: the mean (pink line) is pulled in the direction of the heavier tail, making it misleading as a representation of the center



Below are nine sets of data sampled from a normal distribution. These samples do not have problems; however, most of these histograms probably do not look bell-shaped to the untrained eye.



Each sample contains 24 observations from a normal distribution

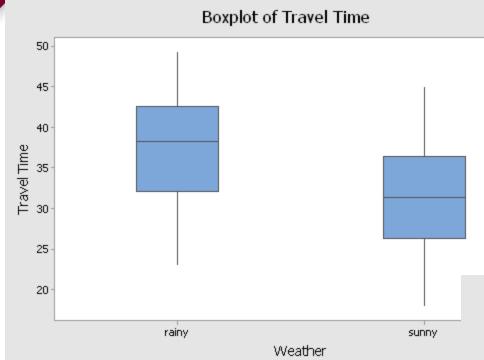


Box Plot

- Use boxplots (also called box-and-whisker plots) to obtain information about the shape, dispersion, and median of a given data.
- Sports outliers
- Used to assess the symmetry of the data
- Used to compare samples of data against each other

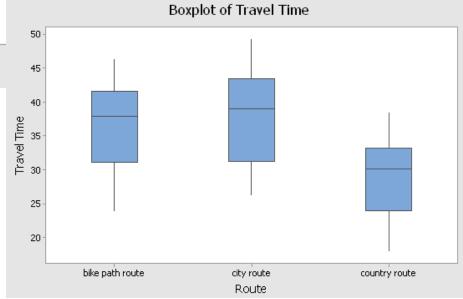
Bike Trip 2.MTW





50% of the data falls within the box

Whiskers extend the whole data





Pareto Chart:

Quality tools / Pareto Chart

X axis: Categorical

Y axis: Continous

Pareto.MTW

Fishbone.MTW

IntSPC.MTW



SPC: Determine the control limit:

Calculate Central line

$$R =$$

- X = avg. of subgroup avg.
- $\overline{x_i}$ = avg. of its subgroup
- g = no. of subgroups
- \overline{R} = avg. of subgroup ranges
- R_i = range of its subgroup
- Where A2, D4, D3 are factors vary according to different n

$$UCL_{\bar{x}} = x + 3\sigma_{\bar{x}} = x + A_{2}\bar{R}$$

$$= x - 3\sigma_{\bar{x}} = x - A_{2}\bar{R}$$

$$LCL_{\bar{x}} = x - 3\sigma_{\bar{x}} = x - A_{2}\bar{R}$$

$$UCL_R = \overline{R} + 3\sigma_{\bar{x}} = D_4\,\overline{R}$$

$$LCL_{R} = \overline{R} - 3\sigma_{\overline{x}} = D_{3}\,\overline{R}$$



AGENDA

Overview of MINITAB

Introduction to Minitab

Hypothesis Tests: Continuous Data

Hypothesis Tests: Attribute Data

Equivalence Testing (Optional)

Analysis of Variance

Correlation and Simple Regression

Multiple Correlation and Regression

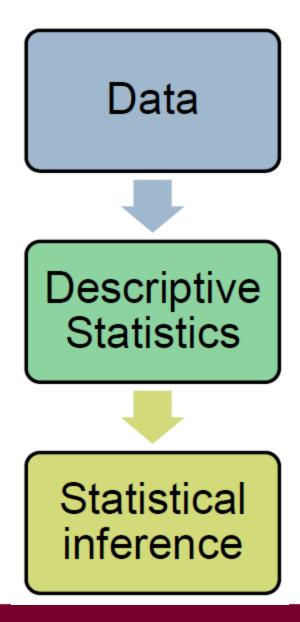


Hypotheses Tests: Continuous Data

- Statistics is the science of making effective use of numerical data.
- Statistics are all around us.
 - Ex: The use of index in newspaper reports, progress reports, profit-loss reports etc.
 - Obtained from research to provide answers to the identified research questions(s).

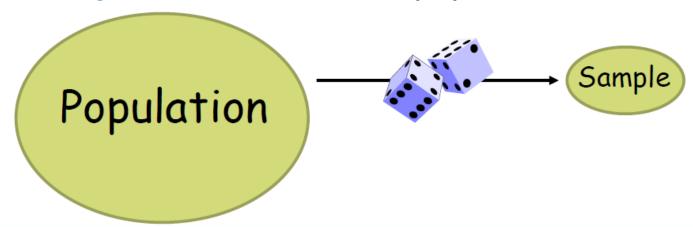


Statistical Inference:





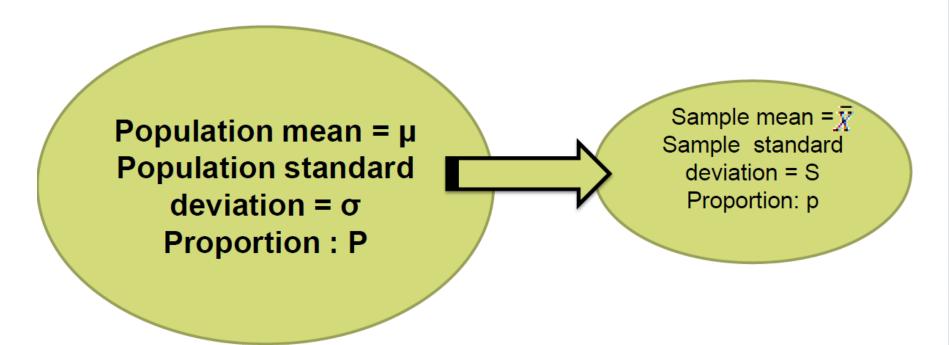
- Statistical inference is a way of using selected group (sample data) to draw conclusions about larger group (population).
- A population is a collect of people, items, or events about which we want to draw conclusions.
- A sample is a subset of the population.





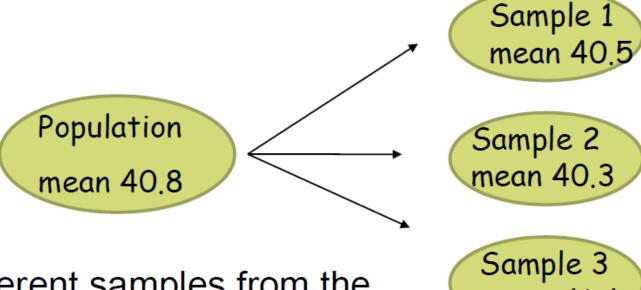
Denotations:

We often use different symbols to distinguish parameters from statistics.





Sampling Error:



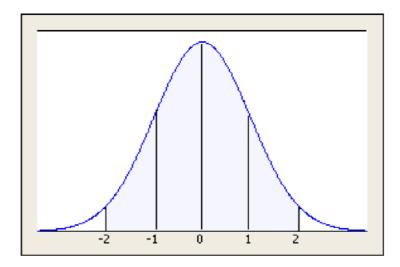
 Different samples from the same population give different results

mean 41.4



- In fact, we cannot be totally sure that the sample is representative of the population. But with a random sample, we can determine how likely it is that the sample represents the population using probabilities.
- Probabilities like this are summarized by the sampling distribution.
- The sampling distribution lists all the possible values of the stat of interest. In this case, the sample % can be from 0 to 100. and for each value, the sampling distribution specifies the probability that the value will occur in a random sample.

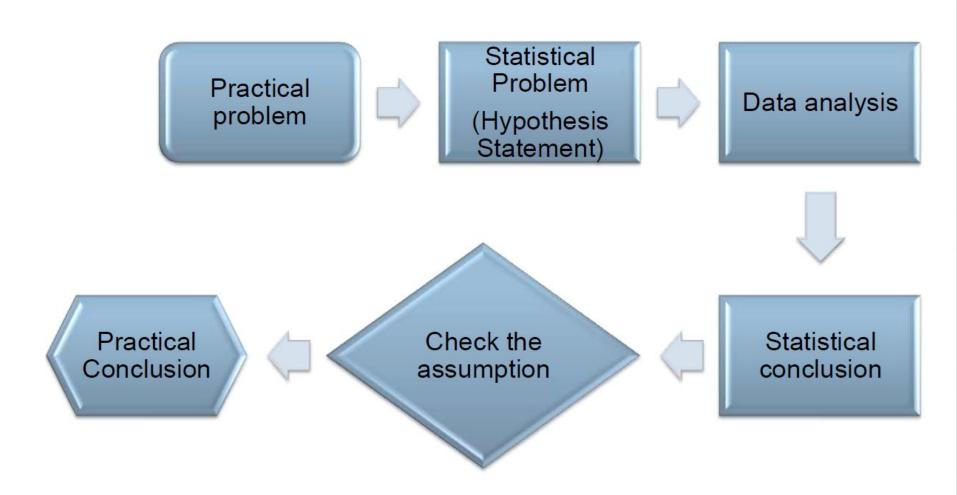




- About 68% of the observations fall within 1 standard deviation from the mean
- About 95% of the observations fall within 2 standard deviation from the mean
- About 99.7% of the observations fall within 3 standard deviation from the mean



Basic Concepts:

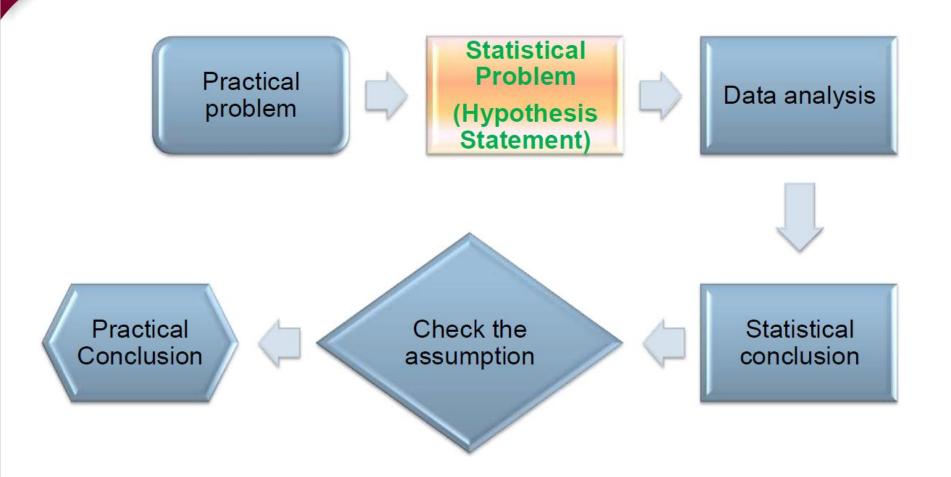




Basic Testing:

- When to use a hypothesis test
 - Use a hypothesis to make inference about one or more population when sample data are available.
- Why use a hypothesis test
 - A hypothesis test answers questions such as:
 - Is a process centered correctly?
 - Is the product from one supplier better than the product from another?
 - Do treatment groups in an experiment differ?







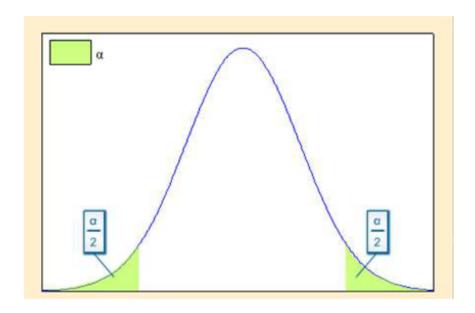
Directional & Non-Directional

- The null hypothesis H₀ usually states that a population parameter is equal to a specified value or a parameter from another population.
- ► The alternative hypothesis H₁ usually states that a population parameter is not equal to a specified value or a parameter from another population.
- Hypothesis tests include two hypotheses: the null hypothesis (denoted by H₀) and the alternative hypothesis (denoted by H₁).
- The null hypothesis is the initial claim and is often specified using previous research or common knowledge.
- The alternative hypothesis is what you may believe to be true.



Directional & Non-Directional

- A nondirectional H₁ simply states that the H₀ is wrong. It does not predict whether the parameter of interest is larger or smaller than the reference value specified in H₀.
- H₀:µ=a vs H₁:µ≠a





Directional & Non-Directional

A directional H₁ states that H₀ is wrong, and also specifies whether the true value of the parameter is greater than or less than the reference value specified in H₀.

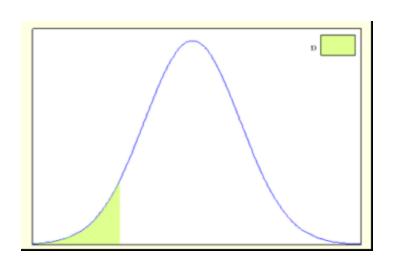
Or

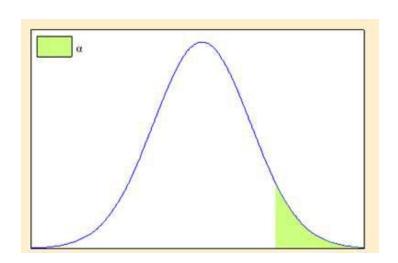
H₀:µ=a vs H₁:µ<a</p>

H₀:µ=a vs H₁:µ>a

H₀:µ≥a vs H₁:µ<a</p>

H₀:µ≤a vs H₁:µ>a

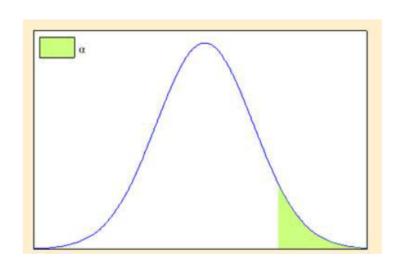






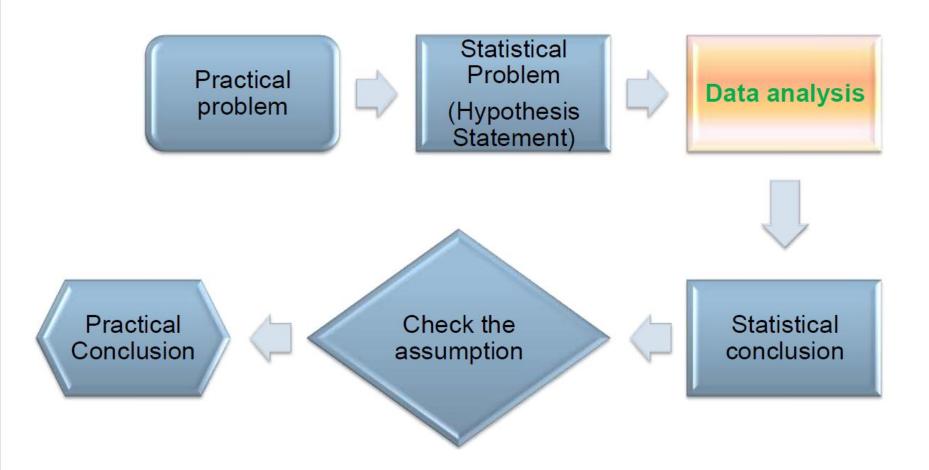
Directional & Non-Directional

- The advantage of using a directional hypothesis is increased power to detect the particular effect you are interested in.
- The disadvantage is that there is no power to detect an effect in the opposite direction.





Basic Testing:



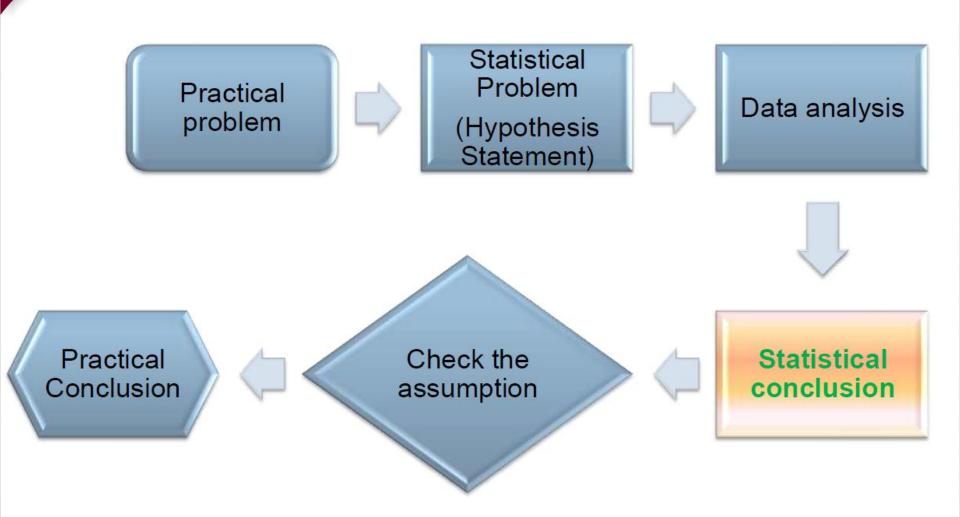


Random Sample

- To draw conclusions from a sample about an entire population, the sample must be <u>random</u>.
- In a random sample, every member of the population has an equal chance of being included in the sample. Members of the sample are selected independently of each other.
- Data from a randomized experiment or stable process may also be treated as a random sample.
- Benefits of Random Samples
 - Random samples are more likely to be representative of the population.
 - They allow us to qualify how well they represent the population



Basic Testing:





Two Possible Conclusions

- Reject the null hypothesis H₀ in favor of the alternative hypothesis. If we reject the null hypothesis, we say that our results is statistically significant.
- Fail to reject the null hypothesis H₀ and conclude that we do not have enough evidence to claim that the alternative hypothesis is true. (That is not the same as saying that the two outcomes are equivalent)



Statistic Conclusion:

- Because statistics are based on a sample of the population, decisions based on those statistics could be wrong. What are the consequences of a wrong decision? How do we minimize the chances of making a wrong decision?
- Alpha (α) is the probability of rejecting a null hypothesis that is true.



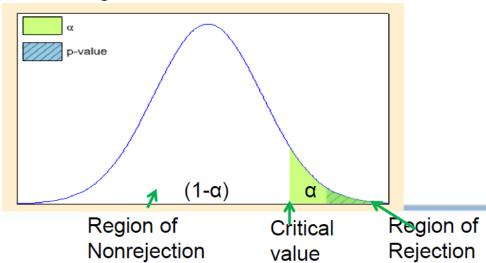
Decision		H₀ Is True	H ₀ Is False
	Fail to reject H ₀	Correct decision p=1-α	Type II error p= β Consumer's risk
	Reject H ₀	Type I error p=α Producer's risk	Correct decision p=1-β (power)

- A Type I error occurs if you reject the null hypothesis H0 when it is true and should not be rejected.
- A Type II error occurs if you do not reject the null hypothesis H0 when it is false and should be rejected.



Alpha & p-value:

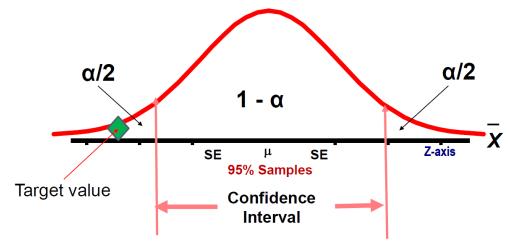
- Alpha (α) is the probability of rejecting a null hypothesis that is true.
- The p-value is the probability of obtaining a test statistic that is at least as extreme as the calculated value if the null hypothesis is true.
- If the p-value ≤ α, reject H₀.
- If the p-value > α , fail to reject H₀.





What is a confidence interval

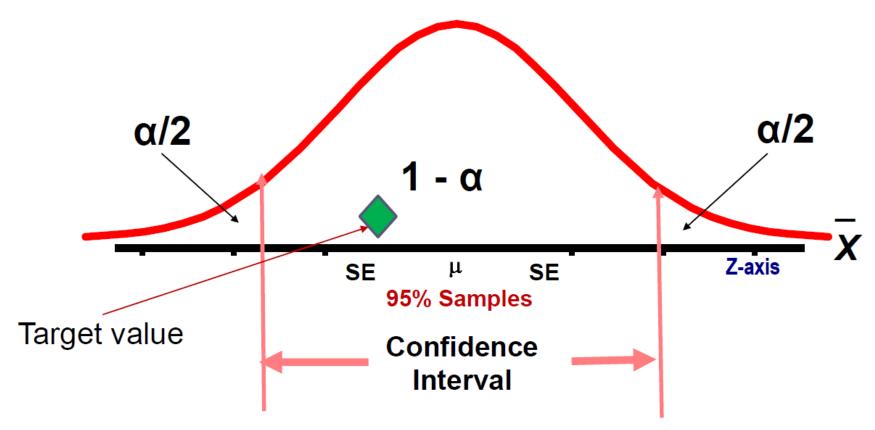
- A confidence interval is a range of likely values for a population parameter (such as μ) that is based on sample data.
- For example, with a 95% confidence interval for μ, you can be 95% confident that the interval contains μ.



The sample is from a population with a mean that is **different** from the target value.



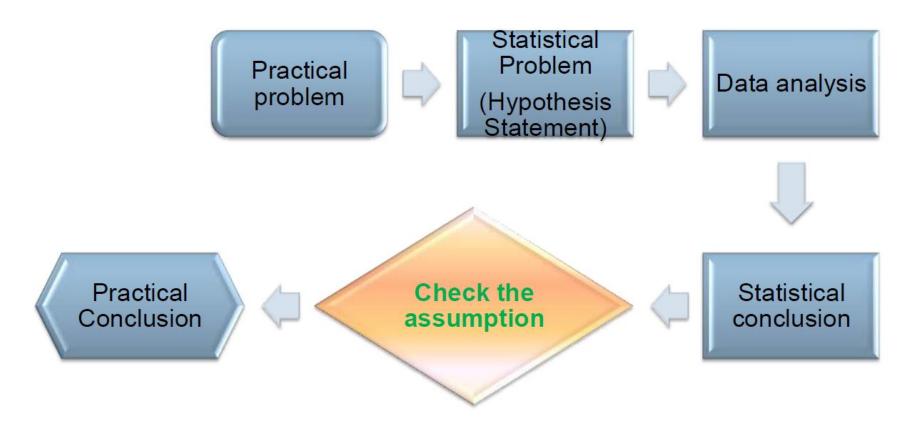
Confidence Interval:



The sample is from a population with a mean that is **not different** from the target value



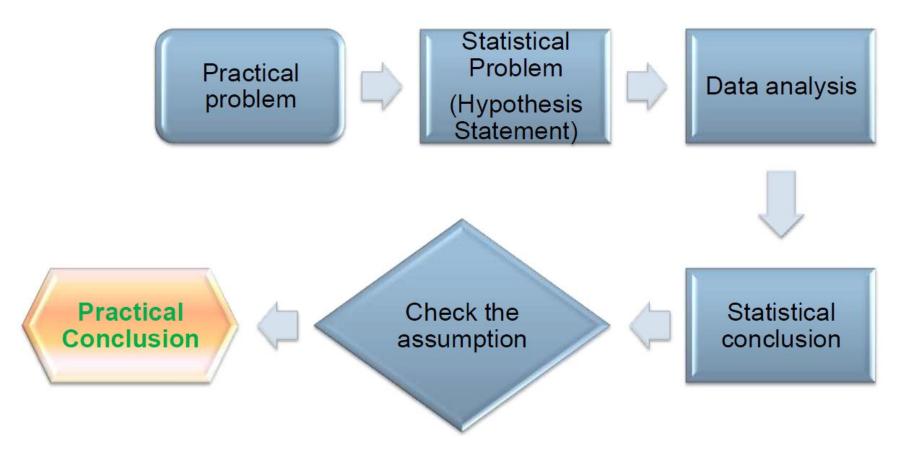
Basic Concepts:



Each hypothesis test is based on one or more assumptions about the data. If these assumptions are not met, the conclusions may not be correct.



Basic Concepts:



If a test is statistically significant, we should use our knowledge of the process to interpret the results.



Power Analysis:

- A power analysis can help answer questions such as:
- How many samples should I collect for the analysis?
- Is the sample size large enough?
- How large a difference can the test detect?
- Are the results of a test trustworthy?
- ▶ Power and Sample Size test can help you to save time and money, while still providing results that you can trust.
- Power is the ability of a test to detect a difference when one exists. A hypothesis test has the following possible outcomes:
 - = chance of detecting a shift



Decision		H₀ Is True	H₀ Is False
	Fail to reject H ₀	Correct decision p=1-α	Type II error p= β Consumer's risk
	Reject H ₀	Type I error p=α Producer's risk	Correct decision p=1-β (power)

- Alpha (α) is the probability of rejecting a null hypothesis is that that is true.
- Beta (β) is the probability of making a Type II error.

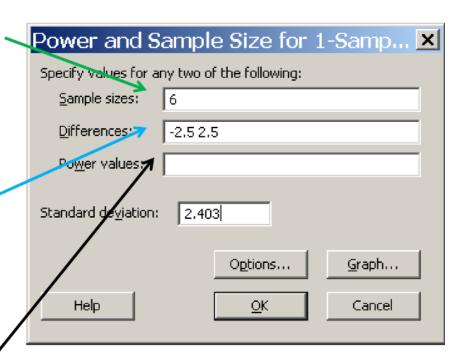


- Use a power analysis:
 - Before collecting data, to determine the sample size
 - After collecting data, to evaluate to detect a difference
- It should be noted at this point that if the power of the statistical test is very high, then the test might be considered over-sensitive from an economic perspective.
- Imagine a sample size that is very large indeed. We may be able to detect a very small difference from the hypothesized value, though that difference may not be of any *practical* significance, especially in relation to the cost of sampling such a large amount of product.



Evaluating Power & sample size:

- Sample sizes—the number of observations in each sample.
- Differences (effects)—the minimum difference between the mean for one population and the mean for the other that you can detect.
- Power—the probability of detecting a significant / difference when one truly exists.





If you enter values for any two of these properties, Minitab will calculate the third.

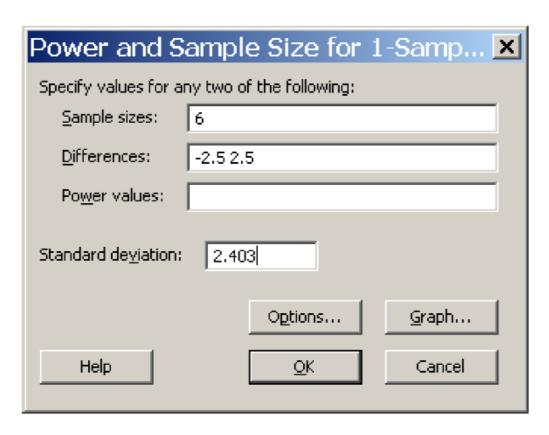
Stat / Power and sample size / 1-sample t

From the cereal data, stdev is 2.403 from t-test results.

CerealBx.MPJ

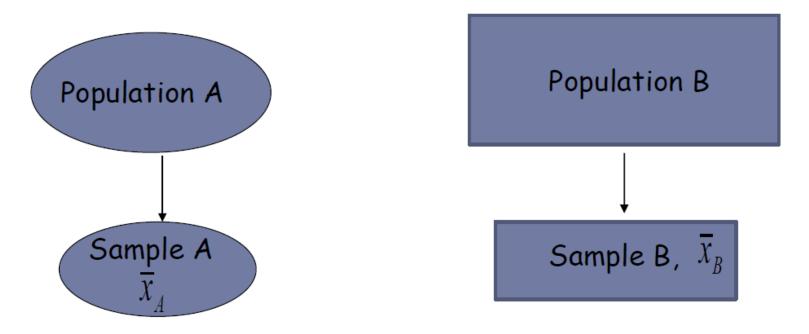
= with 6 observations, stdev of 2.403 and alpha 0.05, the power is only 0.5377. Thus, if mean is off target by 2.5grams, the chance of detecting it with a sample size of 6 is 53.77%

Therefore, if the process mean is off target by 2.5 grams, there is a 46% chance that you will not detect it with a sample size of 6





Two sample t-test:



- Many experiments involve a comparison of two populations.
- In this example, we consider techniques for using two samples to compare the populations from which they were selected



To determines whether two population means are different. The test uses the sample standard deviations to estimate σ for each population. If the difference between the sample means is large relative to the estimated variability of the sample means, then the population means are unlikely to be the same.

$$H_0$$
: $\mu_A - \mu_B = 0$

$$H_1$$
: $\mu_A - \mu_B \neq 0$

To evaluate whether the means of two populations are different by a specific amount.

$$H_0$$
: $\mu_A - \mu_B = \#$

$$H_1$$
: : $\mu_A - \mu_B \neq \#$



When to use Two sample t-test:

- Samples must be independent and random samples.
- Samples data must be continuous.
- Sample data are from normally distributed populations.
- The test also assumes that the data come from normally distributed populations. However, it is fairly robust to violations of this assumption, provided the observations are collected randomly, and the data are continuous, unimodal, and reasonably symmetric.



Why use Two sample t-test:

- An independent two-sample t-test can help answer questions such as:
 - Are the means of same product characteristic between two suppliers comparable?
 - Is one formulation of a product better on average than another?



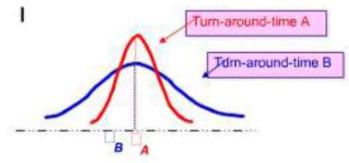
Comparing Variences:

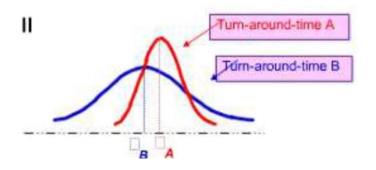
2-variances:

- The 2-sample t-test compares the means of two populations. Often it is of interest to know whether the variances (or standard deviations) of two groups are different.
- Statistics hypothesis

$$H_0$$
: $\sigma_A / \sigma_B = 1$

$$H_1$$
: $\sigma_A / \sigma_B \neq 1$







Eg.: Plastic Strength

A calculator manufacturer is selecting a plastic supplier. Based on the result of power & sample size, he decided to use a sample size of 20 from each supplier to compare the strength of the supplier of plastic pellets.

Plastic.MPJ

- Stat/ Basic/ 2-simple t / each sample is in its own column
- -Graph / Dot plot / Multiple Y's / Simple / ok
- Stat / Basic Statistics / Graphical summary

Two-Sample T-Test and CI: SupplrA, SupplrB

Two-sample T for SupplrA vs Supplr

N Mean StDev SE Mean

SupplrA 20 163.82 5.66 1.3

SupplrB 19 160.14 3.26 0.75

Difference = μ (SupplrA) - μ (SupplrB)

Estimate for difference: 3.67

95% CI for difference: (0.67, 6.68)

T-Test of difference = 0 (vs \neq): T-Value = 2.50 P-Value = 0.018 DF = 30

Ho =
$$\mu$$
A - μ B = 0
H1 = μ A - μ B \neq 0

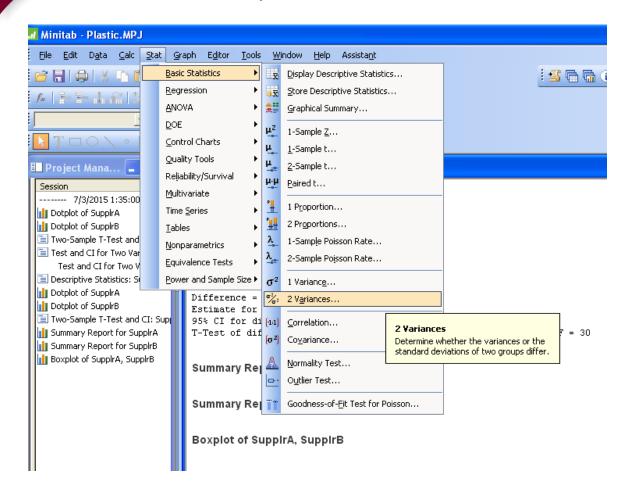
P value < 0.05 → Reject the null. Accept H1.

Conclusion: The mean strengths are different.

Next: compare variance.



Compare variances:

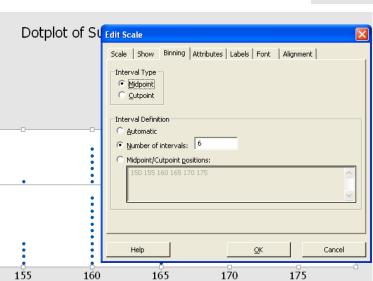


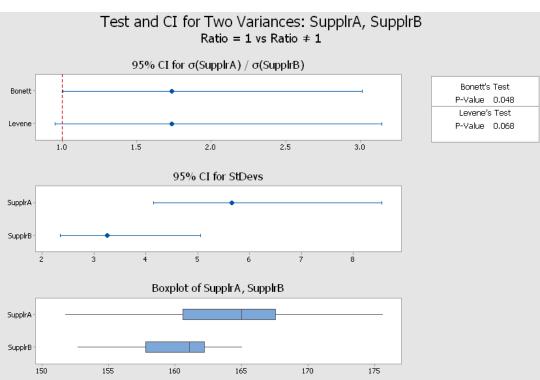
Ho = $\sigma A / \sigma B = 1$ H1 = $\sigma A / \sigma B \neq 1$

P value < 0.10 → Reject the null. Accept H1.



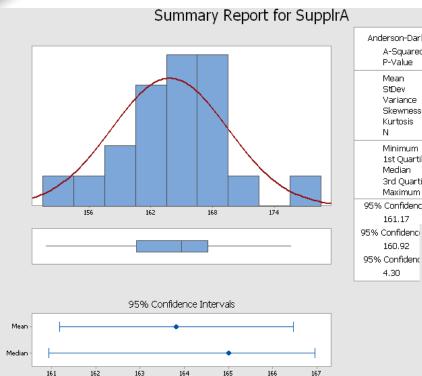
Double click x-axis, Go to Binning Tab







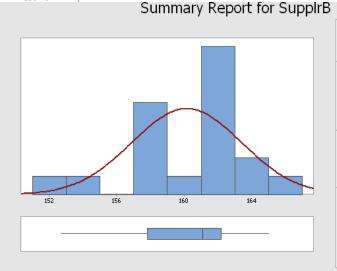
Next, check for Normality: Stat / Basic statistics / Graphical summary



Anderson-Darling Normality Test A-Squared 0.25 P-Value 0.718 163.82 Mean StDev 5.66 32.04 Variance: -0.251625 Skewness Kurtosis 0.235105 Minimum. 151.80 160.60 1st Quartile 165.00 Median 3rd Quartile 167.55 Maximum. 175.60 95% Confidence Interval for Mean 161.17 166.46

160.92 95% Confidence 4.30

P-value 0.718 & 0.164 > 1.0. Conclusion: Data are from Normal distribution.



Anderson-Darling Normality Test			
A-Squared	0.52		
P-Value	0.164		
Mean	160.14		
StDev	3.26		
Variance	10.65		
Skewness	-0.743309		
Kurtosis	0.048938		
N	19		
Minimum	152.70		
1st Quartile	157.80		
Median	161.10		
3rd Quartile	162.20		
Maximum	165.00		
95% Confidence Interval for Mean			
158.57	161.71		
95% Confidence Interval for Median			
157.97	162.20		
95% Confidence Interval for StDev			
2.47	4.83		

Summary Layout: Editor / Layout tool Click Finish.



E.g Anode .MPJ

An electronic manufacturer must ensure that the electrical anode in each capacitor is a certain distance above the surface of the capacitor's ceramic body. Recently, the manufacturer has produced many capacitors with anode heights that violate the lower specification limit.

A change has been implemented which they believe will increase anode height. The engineers must determine how much data to collect, and then compare the height measurements before and after the process change

Process name: Anode insertion process

Hint:

- 1. Determine the required sample size in enabling to detect the process has improved by at least 0.4mm with a power of 0.80. Stdev: 0.55mm.
- 2. Normality ? (graphical summary)
- 3. Equal variance? Use alpha 0.10
- 4. 2-sample t-test to check improvement on After the process change. Use alpha 0.05.



Paired T-Test:

- A paired t-test helps determine whether the mean difference between paired observations is significant.
- Statistically, it is equivalent to performing a one-sample ttest on the difference.
- A paired t-test can also be used to evaluate whether the mean difference is equal to a specific value.

e.g: 20 drivers; Each driver drives Car A and Car B

Parking.xlsx

- -Export to MPJ
- -Assistant / Hypothesis Tests → Paired t



When to use Paired T-Test:

- Use a paired t-test with a random sample of paired observations. The data must be continuous.
- Paired observations are related in some way. Examples include:
 - Weights recorded for individuals before and after an exercise program
 - Measurements of the same part taken with two different measuring devices

Why:

When data are paired, as with before & after measurements, the paired t-test results in a smaller variance and greater power of detecting differences than would the above 2-sample t-test



2-sample t-test:

Ho: $\mu A - \mu B = 0$

 $H1: \mu A - \mu B \neq 0$

2-variance test:

Ho: $\sigma A / \sigma B = 1$

 $H1: \sigma A / \sigma B \neq 1$

1-sample t-test:

Ho: $\mu = \#$

 $H1: \mu \neq \#$

Paired t-test:

Ho: Diff A-B = 0

 $H1: Diff A \neq 0$



Outlier test:

- Many statistical techniques are sensitive to the presence of outliers.
- Checking for outliers should be a routine part of any data analysis. Potential outliers should be examined to see if they are possibly wrong.
- If the data point is in error, it should be corrected if possible and deleted if it is not possible.
- ▶ If there is no reason to believe that the outlying point is in error, it should not be deleted without careful consideration. However, the use of more robust techniques may be warranted. Robust techniques will often downweight the effect of outlying points without deleting them.



- The test is called Grubbs outlier test. It assumes that the data come from a normal population.
- The hypotheses are the following:
 - H₀: All data values come from the same normal population.
 - H₁: The smallest or largest data value is an outlier.



- Choose Stat > Basic Statistics > Outlier Test.
- What do you want to determine
 - Smallest or largest data value is an outlier: Use
 when you are not sure whether the smallest data
 value or the largest data value might be an outlier.
 - Smallest data value is an outlier: Use when you suspect that the smallest data value is an outlier.
 - Largest data value is an outlier: Use when you suspect that the largest data value is an outlier.



You should not use Minitab's outlier tests more than once on the same sample. If you remove an outlier from your sample and then retest, you risk removing values that are not actually outliers.

* Another way to check the data through Editor/ Brush



ANOVA Analysis of Variance

ANOVA with One Factor

ANOVA with Two Factors

ANOVA with Block

General Linear Model



Important terms:

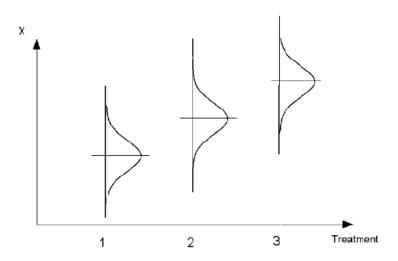
- The explanatory variable, or factor, is categorical, meaning the data classify people, objects, or events.
- The group or categories that comprise a factor are called levels.
- The response is the variable being measured in the study. It is continuous, meaning the data are measured quantities for which fractional values make sense.



- Analysis of Variance = ANOVA
- Developed by Sir R.A Fisher
- Variance in variable is decomposed into different section of variations
- One way ANOVA
- Two way ANOVA
- Factorial ANOVA
- Repeated measures ANOVA



- Residual plots
- To ensure that the results are valid, check whether all the assumptions about the errors have been met.
 - Normally distributed
 - Constant variance for all fitted values
 - Random over time





- The one-way ANOVA (Analysis of Variance) procedure is a generalization of the independent samples t-test. Unlike the t-test, however, you can use one-way ANOVA to analyze the means of more than two groups (samples) at once.
- The basic logic behind ANOVA is that within-group variation is due only to random error.



Basic idea is to partition the total variation in the response into pieces of interest to determine which pieces are large

Total Variation (SST) d.f. = n-1 Between-Group Variation (SSF)

d.f. = c-1

Within-Group Variation (SSE)

d.f. = n-c

$$Variation = \sum_{sum} (\#-mean)^2$$



Eg: Car Seat thread strength

- Problem
- Three quality inspectors who measure car seat thread strengths (in kilograms) participate in a gage reproducibility study. Test for a measurement error bias between operators by comparing both the means and the variances of their measurements.
- Data collection
- The operators, Kevin, Michelle, and Rob, are randomly assigned 25 car seat thread samples to measure. All 75 thread samples come from one batch of product.
- Data set
- CARSEAT.MPJ



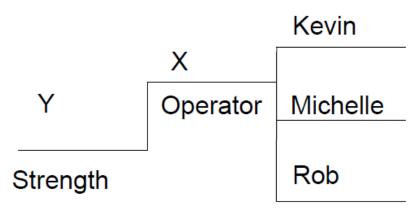
One Way ANOVA: between-inspector differences

Use General Linear Model (GLM) to compare the mean strength measurements across the three inspectors. The hypotheses are:

 H_0 : All inspectors have the same mean (no bias).

$$H_0 = \mu_k = \mu_m = \mu_R$$

H₁: Not all inspectors have the same mean (bias).



Assumptions



Results:

Analysis of Variance

```
      Source
      DF
      Adj SS
      Adj MS
      F-Value
      P-Value

      Inspector
      2
      6.621
      3.3104
      3.85
      0.026

      Error
      72
      61.895
      0.8597

      Total
      74
      68.516
```

1 less than # of groups

(equals df for each group

of data values - # of groups

added together)

1 less than # of individuals (just like other situations)



Analysis of Variance

Source Inspector Error

Total

DF Adj SS Adj MS
2 6.621 3.3104
72 61.895 0.8597
74 68.516

F-Value P-Val 3.85 0.02

SS_E also known as Within Group Sum of Squares

$$\sum\sum(x_{ij}-\overline{x}_{i.})^2$$

SS _{treatment} also known as Between-Group Sum of Square

$$\sum n_i (\overline{x}_i - \overline{x}_i)^2$$

Total Sum of Square

$$\sum \sum (x_{ij} - \overline{x}_{..})^2$$



Probability

Analysis of Variance

3.85 0.026

MS _{treatment} also known as Between Group Mean Square

$$MSF = \frac{SSF}{k-1}$$

MS_F also known as Within Group Mean Squares

$$MSE = \frac{SSE}{n-k}$$

F ratio. Ratio of the Mean Squares

$$F = \frac{MSF}{MSE}$$

(P-values for the F statistic are in Table)



Two way ANOVA

 We have identified 2 product characteristics (factors) that are thought to influence the amount of vibration of running motors. Factor A = brand of bearing used in motors & Factor B = the material used for motor casing. A is 5 brands and B is three types of material: steel, aluminum, plastic. Two motors (r = 2) were constructed and tested for each of the a.b = 5x3 = 15 combination with 2 repetitions = 30 runs of experiments.

innovative • entrepreneurial • global



Two way ANOVA

			FACTOR B (material)					
		Steel		Aluminum		Plastic		
		r1	r2	r1	r2	r1	r2	Avg A
	1	13.1	13.2	1 5	14.8	14	14.3	14.07
FACTOR	2	16.3	15.8	15.7	16.4	17.2	16.7	16.35
Α	3	13.7	14.3	13.9	14.3	12.4	12.3	13.48
(brand)	4	15.7	15.8	13.7	14.2	14.4	13.9	14.62
	5	13.5	12.5	13.4	13.8	13.2	13.1	13.25
	Avg B	14.39		14.52		14.15		

```
SSA = b.r \sum ( average A – Ybar)<sup>2</sup>

SSA = 3.2[ (14.07 – 14.353)<sup>2</sup>+(16.35-14.353)<sup>2</sup>+ (13.48-14.353)<sup>2</sup>+(14.62-14.353)<sup>2</sup>+(13.25-14.353)<sup>2</sup>]

= 6[6.118125] = 36.709
```

```
SSB = a.r \sum ( average B – Ybar)<sup>2</sup>
=5.2[(14.39 – 14.353)2+(14.52 – 14.353)2+(14.15 – 14.353)2]
= 10 [ 0.070467 ] = 0.705
```



Two way ANOVA

 Total sum of squares is the sum of the squared differences of all 30 values from Y

$$SST = [(13.1 - 14.353)2 + ... + (13.1 - 14.353)2 = 50.655$$

 SSE is the sum of squared differences of each response value from its own cell mean

• By subtraction, SS (AB) = SST - SSA - SSB - SSE = 50.655 - 36.709 - 0.705 - 1.670 = 11.571



Two Way ANOVA

Source of variation	DOF	SS	MS	F
Factor A (bearing brand)	5-1 = 4	36.71	9.18	82.43
Factor B (casing material)	3-1 = 2	0.71	0.35	3.17
AB interaction	(5-1)(3-1)=8	11.57	1.45	12.99
Error	5*3(2-1)=15	1.67	0.11	
Total Variation	5*3*2 - 1 = 29	50.66		

- At a significant level level of 0.05, main effect for A and interaction AB are significant. B is not significant factor.
- Meaning: the brand of bearing is very important. Although the casing material does not have significant effect, it does influence the effect of A becaue the interaction AB is significant



Interpreting Results:

F is the ratio of the variability contributed by the factor to the variability contributed by error

$$F = \frac{AdjMS(Operator)}{AdjMS(Error)}$$

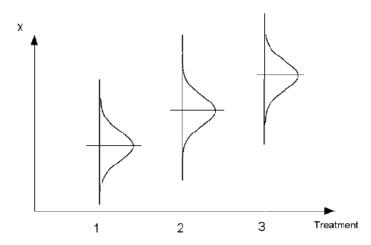
- If the differences between factor level means are similar to what you would expect from random variation, the F ratio should be closed to one.
- If the differences between factor level means are greater than what you would expect from random variation, the F ratio should be greater than one.



- P-value
- P-value is the probability that F would be as large (or larger) if the factor had no effect. A large F suggests that the factor level means are more different than expected by chance, thus the p-value is small.
- Use the p-value to test the following hypotheses:
 - H₀: The factor level means are all the same
 - H₁: At least two of the factor level means are different.



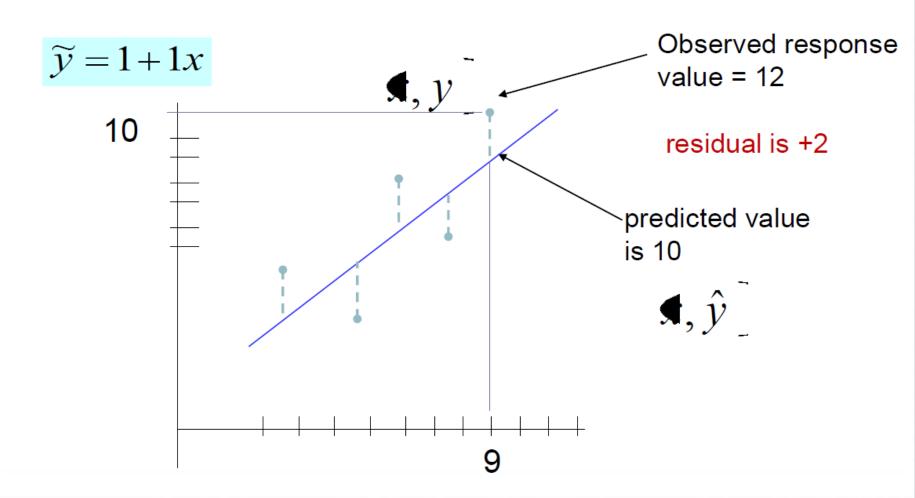
- Residual plots
- To ensure that the results are valid, check whether all the assumptions about the errors have been met.
 - Normally distributed
 - Constant variance for all fitted values
 - Random over time





Residual:

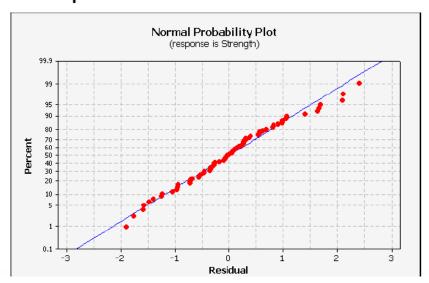
Residual = observed value – predicted value = $y - \hat{y}$





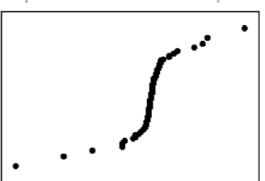
Normal Probability Plot:

► The first assumption states that the errors follow a normal distribution. To test this assumption, we can use a normal probability plot. If the residuals follow a normal distribution, they should approximately follow the straight line on the plot.

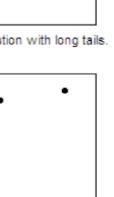




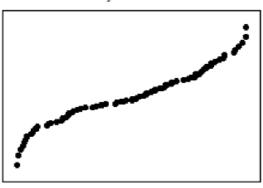
These four plots exhibit patterns that indicate that the residuals do not follow a normal distribution.



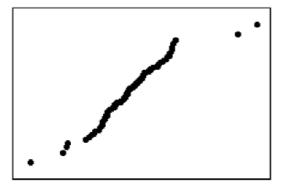
S-curve implies a distribution with long tails.



Downward curve implies an asymmetric distribution.



Inverted S-curve implies a distribution with short



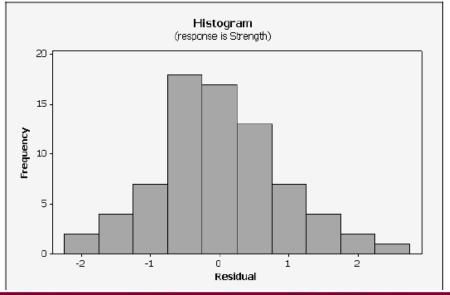
A few points lying away from the line implies a distribution with outliers.



Residuals vs. Fit:

- With a reasonably large sample size, the histogram displays compatible information with normal probability plot.
- The histogram of the residuals should appear approximately bell-shape with no unusual values or

outliers.

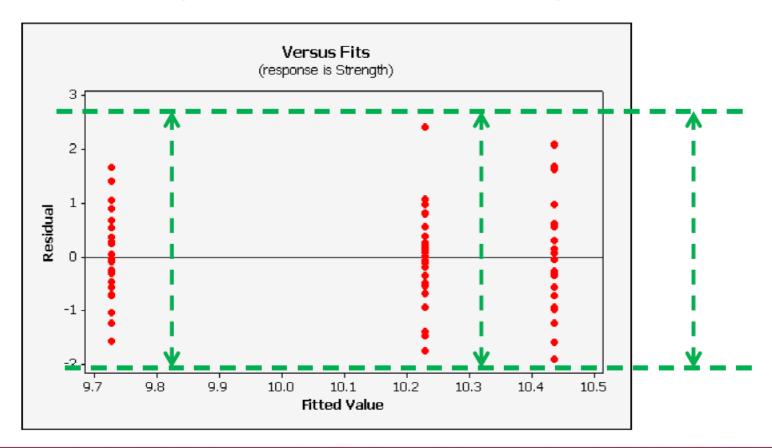




- ➤ The second assumption states that the variance or spread in the errors across the values of x is constant. To test this assumption, we can plot the residuals versus the fitted values are on the vertical axis and the fitted values are on the horizontal axis.
- ▶ Use the plot of the residuals versus fits to verify that the residuals are scattered randomly about zero.

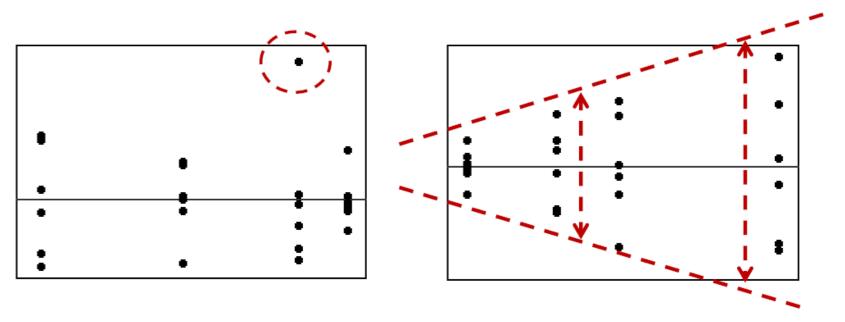


Use the plot of the residuals versus fits to verify that the residuals are scattered randomly about zero, and no unusual patterns in the data are present.





The patterns below show an outlier and a violation of the assumption that the errors are constant.



Plot with outlier

Plot with non-constant variance

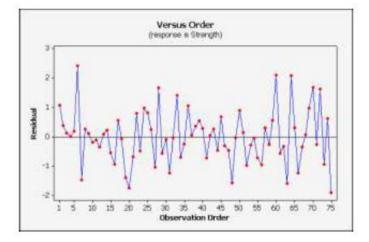


Residuals vs. Order:

 The plot of the residuals versus order displays the residuals in the order of data collection.

 If the data collection order affects the results, residuals near each other may be correlated, and thus, not

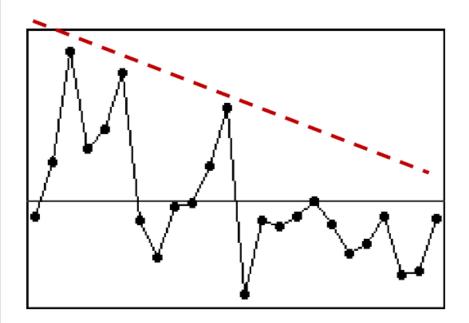
independent.

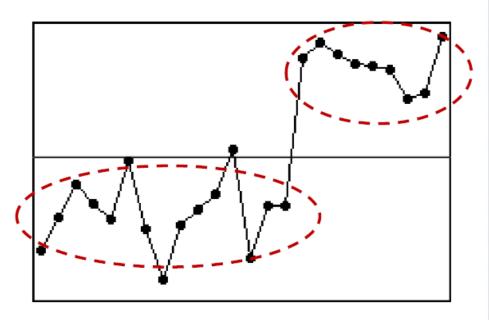


 The Points on this plots fall in a random order and do not exhibit any patterns. The residuals appear to be independent of one another.



 This plot indicates that the residuals may not be independent; the points are gradually decreasing over time







Correlation & Regression

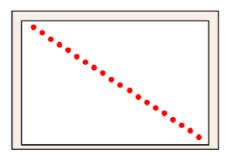
Correlation and Simple Regression

Multiple Correlation and Simple Regression

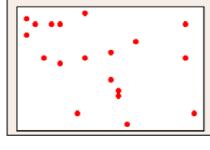


Scatter Plot - Direction and Form:

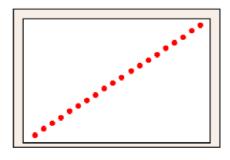
A scatter plot is essentially a plot between the pair of (x,y) values. The purpose of constructing the plot is to examine the relationship between the two variables.



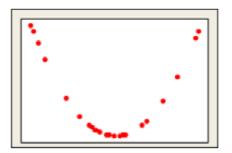
Negative linear association



No association



Positive linear association

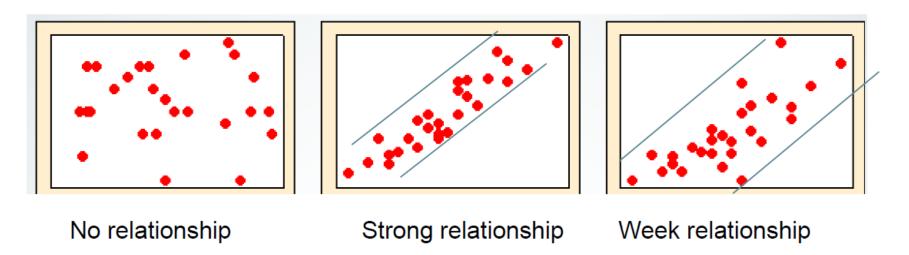


Nonlinear association



Scatter Plot -Strength:

A scatter shows a stronger relationship between variables when the data closely approximate a straight, sloped line.





Correlation:

- The sample correlation coefficient, r, measures the degree of linear association between two variables (the degree to which one variable changes with another).
- Use correlation when you have data for two continuous variables and wish to determine whether there is a linear relationship between them.
- The correlation will not tell you whether the variables are related in a nonlinear fashion.

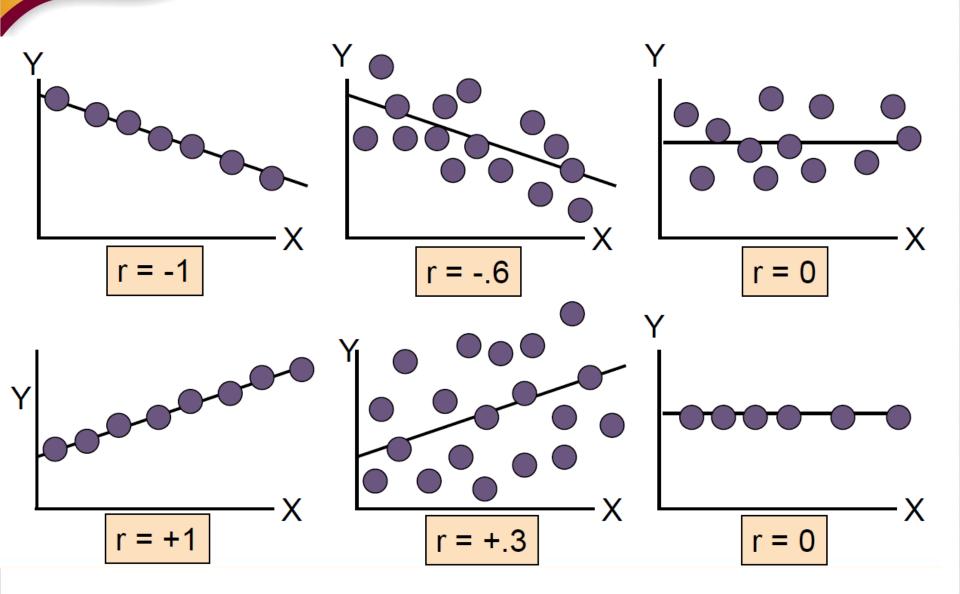


- Calculate the Pearson correlation coefficient to determine the strength of the linear association between two variables.
- Correlation Coefficient: -1≤r≤1
- ► The closer to -1, the stronger the negative linear relationship
- The closer to 1, the stronger the positive linear relationship
- The closer to 0, the weaker any positive linear relationship

Strong	moderate	weak		weak	moderate	Strong
-1	-0.8	-0.5	0	0.5	0.8	1

innovative • entrepreneurial • global







Hypothesis Test:

The p-value tests the following hypotheses:

$$H_0: r = 0$$
 vs $H_1: r \neq 0$

- To test the hypothesis, use the p-value associated with the correlation coefficient.
 - If the p-value ≤ α, reject H₀. (linear association exist)
 - If the p-value > α, fail to reject H₀. (no linear association exist)



Simple Regression:

Simple regression examines the relationship between a continuous response variable (Y) and a predictor variable (X). The general equation for a simple regression model is:

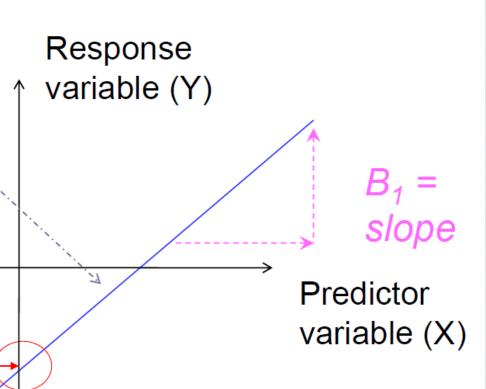
Equation	Type of model
$Y = \beta_0 + \beta_1 X + \epsilon$	Linear

Mhere $β_0$ is the intercept, $β_1$ is the slope, and ε is the random error.



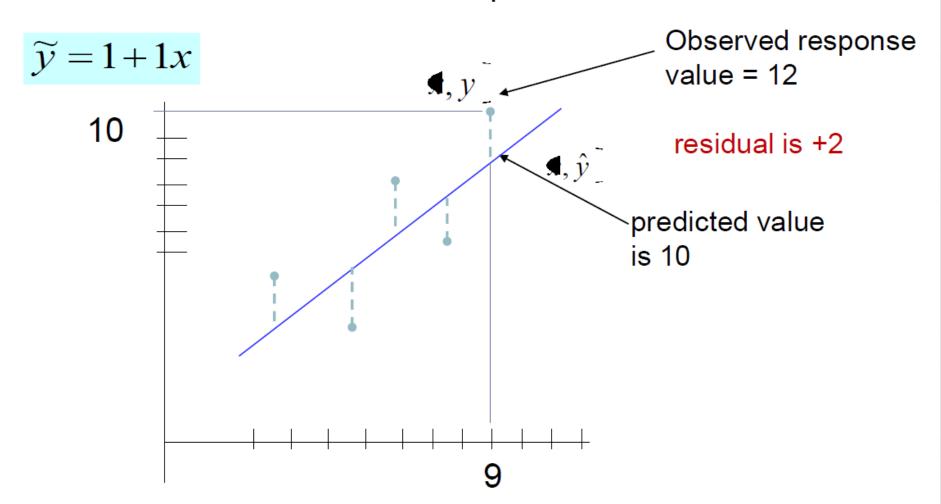
- Remember this:
- $Y = \beta_0 + \beta_1 X + \varepsilon$
- A slope of β_1 means that every 1-unit change in X yields a β_1 -unit change in Y.

 $B_0 = y - intercept$





Residual = observed value – predicted value = $y - \hat{y}$





The least squares regression line:

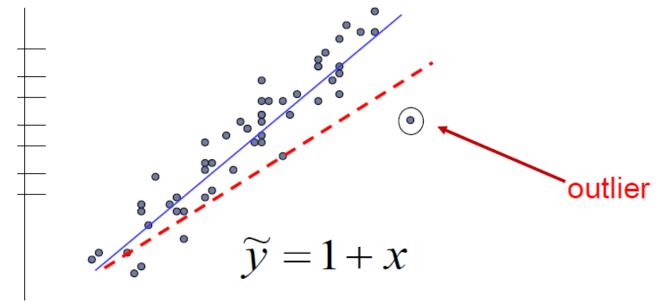
$$Y = \beta_0 + \beta_1 X + \varepsilon$$

The coefficients for the regression equation are chosen to minimize the sum of the squared differences between the response values observed in the sample and those predicted by the equation.

In other words, the least squares regression line minimizes the vertical distances between the points and the line



Be alert for outliers when using regression procedures. Some outliers (called high leverage points) have a large effect on the calculation of the least squares regression line. In such cases, the line may no longer accurately represent the data.





Why use regression?:

- Analysis of variance (ANOVA) is similar to regression in that it is used to investigate and model the relationship between a response variable and one or more independent variables.
- ► However, analysis of variance differs from regression in 2 ways: the independent variables are qualitative (categorical), and no assumptions is made about the nature of the relationship (that is, the model does not include coefficients for variables).
- Several of Minitab's ANOVA procedures, however, allow models with both qualitative and quantitative variables.



Linear Regression Equation:

The true regression line or the probabilistic model is given by

$$Y = \beta_0 + \beta_1 X + \epsilon$$

- Hypothesis test of the slope:
- H_0 : $β_1 = 0$ $Y = β_0 + ε$ (no linear relationship exists between the variables)



- ► If p-value $\leq \alpha$, reject H₀.
 - The regression model explains significantly more variability in the response than does the restricted model. β₁ does not equal zero.
- If p-value > α, fail to reject H_0 .
 - B1 is not significantly different from zero



Common Mistake:

- When using regression procedures, be alert for outliers in both the X and Y variables.
- Outlier in the X variable(s) (high leverage points) can have a large influence on the regression coefficients and p-values. Plot the deleted residuals to detect high leverage points.
- Outliers in the Y variable can also have a large influence. Plot the regular, standardized, or deleted residuals to detect outliers in the Y variable.
- If outliers are present, the regression model may no longer adequately fit the data nor predict future observation.



Common Mistake:

- Extrapolation outside the range
- ▶ Do not apply regression results to values of X that are outside the sample range. For example, you should not use the regression equation derived in this example to predict impurity levels for a stir rate of 100, because the highest between Stirrate and Impurity may be very different for stir rates above 42.



Multiple Regression:

Multiple regression examines the relationship between a continuous response variable (Y) and more than one predictor variable (X). The general equation for a multiple regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \varepsilon$$

where Y is the response, β_0 is the intercept, each X_i is a predictor variable with a model coefficient of β_i , and ϵ is random error.



- Use multiple regression when you have a continuous Y and more than one X.
 - X is ordinal or continuous.
 - In theory, X should be fixed. In practice, however, X often varies.
 - Any uncertainty in the measurement of X is assumed negligible compared to the range in which X is measured.
- To confirm that the analysis is valid, verify all assumptions about the model error term. Use residual plots to check that the errors have the following characteristics:
 - Normally distributed
 - Constant variance for all fitted values
 - Random over time



Best subsets regression:

▶ Best subsets regression evaluates all possible combinations of predictors to help you determine which combination makes the best regression model. Minitab uses a maximum R² criterion to choose the best model. Other criteria may provide a different model.

# of predictors (p-1)	# of regression models
1	$2:()(x_1)$
2	$4:()(x_1)(x_2)(x_1,x_2)$
3	8: () (x_1) (x_2) (x_3) (x_1, x_2) (x_1, x_3) (x_2, x_3) (x_1, x_2, x_3)
4	16: 1 none, 4 one, 6 two, 4 three, 1 four



For every set, one or two models with large R^2 are picked. They are the following:

Sets	Models	R ²
Set B	$Y = \beta_0 + \beta_2 X_2 + \varepsilon$	0.666
	$Y = \beta_0 + \beta_4 X_4 + \varepsilon$	0.675
Set C	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$	0.979
	$Y = \beta_0 + \beta_1 X_1 + \beta_4 X_4 + \varepsilon$	0.972
Set D	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 X_4 + \varepsilon$	0.982
	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$	0.982
Set E	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$	0.982

Use best subsets regression when you have many potential predictors and thus many possible regression models to choose from.



- R2, R2 adjusted, and R2 predicted
- When comparing models:
 - If the number of predictors is the same, look for the model with the highest R2.
 - If the number of predictors is different, look for the model with the highest adjusted R2.
 - Look for the model with the highest predicted R2.
- Mallows' Cp
- Look for models where Cp is small and close to the number of parameters in the model. A small Cp value indicates that the model is relatively precise in estimating the true regression coefficients and predicting future responses.





- Types of Experimentation
- Types of Improvement & Tools
- How DOE Fits Into a Statistical System
- Basic Terminology of DOE
- How DOE is an Improvement Over One-at-a-Time Experimentation
- Fundamental Ideas for Experimentation



Trial and Error

One-Factor-at-a-Time (OFAT)

- DOE
 - Full Factorials
 - Fractional Factorials
 - Response Surface Designs



Trial and Error

- Problem:
 - cakes are not cooked properly
 - (rating high values are best)
- Possible variables to try
 - lower oven temperature
 - shorter cooking time
 - more water in recipe
 - change position on shelf
 - change flour supplier
 - clean the oven
- If it works
 - we adopt the change (for ever?)
- If it doesn't
 - we reject the change (for ever?)





One-Factor-at-a-Time

Cook Rating: High values are best

Oven Temp	Cooking Time	Water %	Rating
180	8	5	46
200	8	5	48
200	10	5	48
200	8	10	32

What settings would you recommend for the cake baking process?



Full Factorial Design

Task: High ratings are best



Oven Temp	Cooking Time	Water %	Rating
180	8	5	46
200	8	5	48,
, 180	10	5	₹78 ₹
200	10	5	48
180	8	10	34
200	8	10	32
180	10	10	₹ ⁸⁰ 3
200	10	10	7824



OFAT Runs



Classification of Designs

Design Type Screening Full Response Design or **Factorial** Surface Fractional Design **Factorial Factors** 3 to 15 2 to 6 2 to 5 Explored: Effects Main Effects Main Effects Main Effects, Estimated: Interaction Interaction, Curvature Result: Identify Understand High Quality important System Prediction, **Factors** Behavior Optimization



2^k Factorial Designs

- The term 2^k factorial refers to:
 - A design with k factors
 - Each factor has 2 levels we are investigating
- A 2² factorial is also represented as a 2x2 factorial
 - This design has two factors with two levels and has 2x2 (4 total) treatment combinations.
- Likewise a 2³ factorial has includes 3 factors, each with two levels
 - This experiment has 2x2x2 (8 total) treatment combinations



Design Matrix of Full 2^k Factorial Designs

	Fact	or
Test	1	
1	•	-
2	+	•
3	-	+
4	4	+
(a)	2 ² desi	gn

Factor				
Test	1	2	3	
1	•	-	•	
2	+	-	-	
3	•	+	-	
1 2 3 4 5	+	+	-	
	-	-	+	
6	+	•	+	
7	•	+	+	
8	+	+	+	
	(b) 2 ³ des	sign		

	Facto	or		
Test 1 2 3 4 5	1	2	. 3	4
1			-	•
2	+	-	-	•
3		+	-	-
4	+	+	-	-
5	-	•	4	•
6	+	•	+	•
7	•	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	•	-	+
11	•	+	-	+
12	+	+	•	+
13	-	•	+	+
14	+	•	+	+
15	•	+	+	+
16	+	+	+	+
	{c) 2 ⁴ desi	gn	



Engineering Tools for Improvement

Type of Improvement

Statistical Tools



Removal of Assignable Causes

Root Cause Analysis/DOE

Optimization of Existing Process

DOE & Response Surface

Methods (RSM)

Modification of Existing Tooling

DOE

Improved Procedures

Simulation/Descriptive

Statistics/Hypothesis Tests

/ DOE

New Product Design

Simulation/DOE

New Process Introduction

Characterization/DOE/RSM

New Technology

Characterization/DOE/RSM

Automation

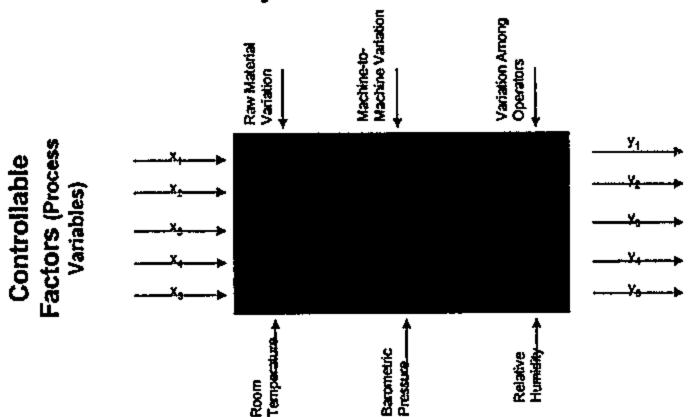
Simulation/Descriptive

Statistics/Hypothesis Tests



Experiment: The act of taking an observation where the outcome is subject to uncertainty, usually with controllable variables and inputs purposely set to note changes in the output variables.

Systematic & Random Noise



Process
Outputs (Quality
Characteristics)



Experimental Unit:

Objects on which the experiment is to be performed. It is the smallest division of material such that any two units may receive different treatments

Responses, Y:

A variable observed or measured in an experiment, sometimes called a dependent variable.

Factors, X:

A variable that is deliberately varied or changed in a controlled manner to observe its impact on the response variable. Sometimes called an independent variable or causal variable.



Level:

A given value of specific setting of a quantitative factor or a specific option of a qualitative factor that is included in the experiment

Treatment:

Specific experimental condition usually given as a combination of levels of the factors.

Noise - Background Variable:

A variable that potentially can affect a response variable in an experiment but is not of interest as a factor. Sometimes called a blocking variable

Noise - Nuisance Variable:

An unknown variable that can affect a response variable in an experiment. Sometimes called a lurking variable or an extraneous variable



Experimental Design:

Ordered set of combination of levels of factors along with a prescription for how the experimental units are assigned to each combination.

Effect:

The change in the response variable that occurs as a factor or background variable is changed from one level to another



Vague Objective

The objective of this experiment is to determine factors that affect the etch process.

The objective of this experiment is to improve the yield of the mold operation.

The objective of this experiment is to characterize the die attach operation.

More Specific Objective

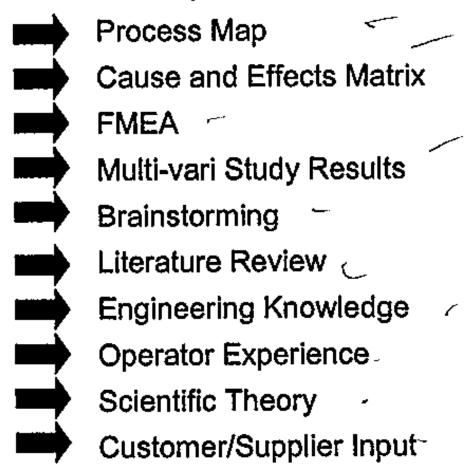
The objective of this experiment is to determine which of the following factors that affect the etch rate by an amount greater than 500 A/min: power, press, temp, gap, spin speed, substrate, clean(Y/N).

The objective of this experiment is find settings for mold pressure, transfer time, clamping pressure, and compound type to reduce the current porosity rate of 600ppm to under 200ppm.

The objective of this experiment is to characterize the effect of, and optimal settings for, die attach cooling rate, mount down speed, collet force, dwell time and over travel on bond line thickness (mean & std), tilt (mean & std) and voids.



Which Factors (KPIV's) do we include in the experiment? Use various sources to prioritize Factor selection:





Example: Characterize and Optimize a Wire Bond Process

Y's: Pull Strength, Shear Strength, Flick Test, Cratering, Ball Size, Ball Thickness, Visual

Factors: Force, Power, Time, Temperature

Questions:

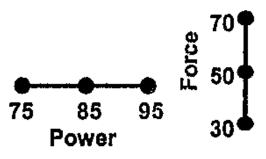
What X's are the drivers of the Y's?
What are the optimum settings of the X's to simultaneously optimize the Y's?

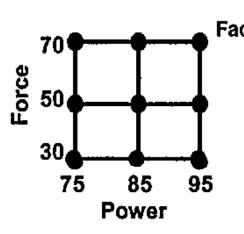
How would you proceed?

UTM UNIVERSITI TEKNOLOGI MALAYSIA

Design of Experiment - DOE

Experimental Design: Arrangement of combinations of factor levels.





Factor Levels: Power: 75, 85, 95

Force: 30, 50, 70

Time: 10, 20, 30

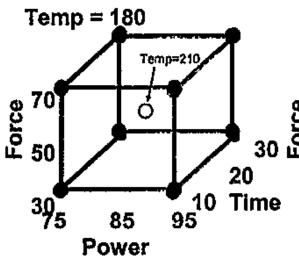
Temp: 180, 210, 240

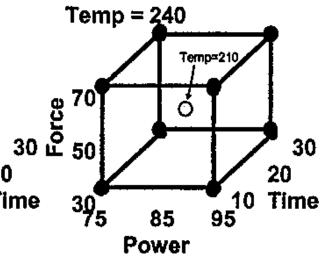
○ Power = 85

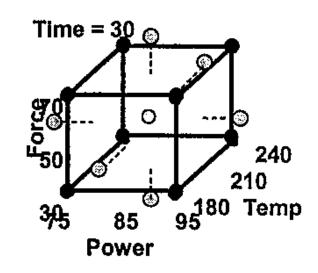
Force = 50

Time = 20

Temp = 210









Chemical Process Control Method Experiment Planning Form

Response variables

1. Sulfate(%)

Water (%)

3. Ingredient X (%)

Factor under Study

1. Control Method

Background Variables

None identified

Measurement Technique

Titration

Titration

Gas chromatograph

Levels

Current, Valve, Meter

Method of Control

Replication

A decision on methodology is needed within about three weeks, so tests will Be conducted on each method for five days (a total of 15 days)

Method of Randomization

Because of the extreme expenses required to change the control method, test cannot be randomized. The current method will be tested for five days, then the valve for five days, then the meter for five days.



Chemical Process Control Method Experimental Data

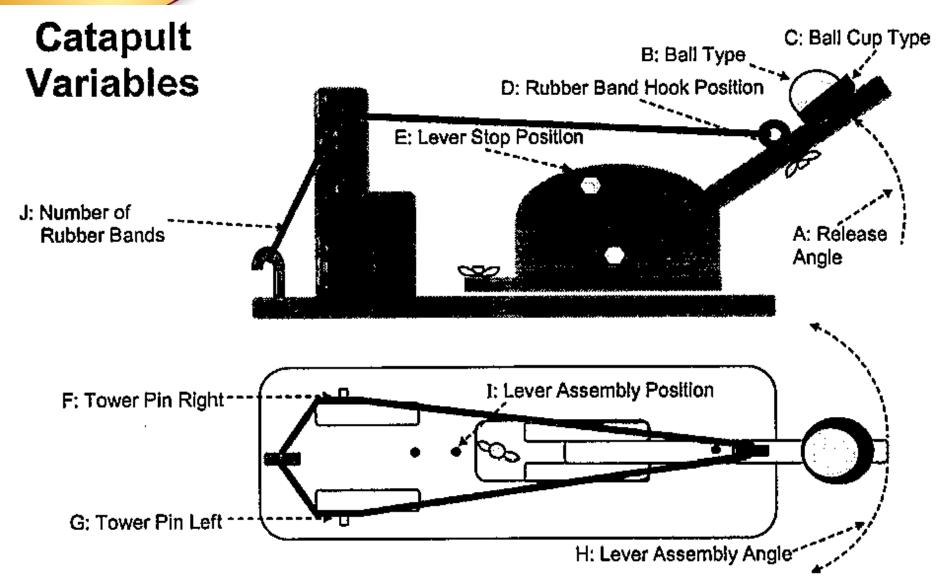
Control Method	Date	Water (%)	Sulfate (%)	Ingredient X (%)
Current	7/2	2.2	0.5	15.7
	7/3	2.0	0.2	17.2
	7/4	2.4	0.6	15.1
	7/5	2.7	0.3	16.5
	7/6	2.9	0.4	14.8
Valve	7/7	3.3	0.2	11.8
	7/8	3.2	0.5	12.6
	7/9	3.4	0.4	13.4
	7/10	3.6	0.3	19.4
	7/11	3.5	0.5	12.3
Meter	7/12	3.7	0.4	18.8
	: 7/13	3.8	0.5	17.6
· .	7/14	4.0	0.3	19.0
•	7/15	4.2	0.5	13.0
·	7/16	4.1	0.6	18.2



Chemical Process Control Method Experiment

Analyze the experiment for all three response variables and make a recommendation to your BOSS as to which control method should be used.







High and Mighty Tower Cake

You have been a cake bake engineer for about 2 years at the High and Mighty Cake Company. The company started in 1924 and has used the same type ovens since that time. Due to age, cost and capacity considerations of the ovens, Purchasing has bought a new oven to replace 10 smaller ones.

Your company produces the famous Tower Cake noted for its height of over 180mm. The first cake that came out of the new oven "soared" to a height of only 95mm. The grandson of the factory founder and now company president who ordered the new ovens to be purchased has given you the task to solve the Tower Cake problem. He has also hinted that if you cannot solve it he will look for a replacement who can. He has given you a budget of \$50.00 for the optimization.

The company who sold you the new oven has sent along some recommendations for usage.

This '	Variable	Knob	Recommended Range	Allowable Settings
	X1	Time	40 to 45 minutes	0 to 100 minutes
	X2	Temp	200 to 225°F	75 to 350°F
100 pm	X3	Rack Height	18 to 24	0 to 30
والمساعدة	X4	Preheat	30 minutes	10 to 40 minutes
	X5	Air Draw	0	-1 to +1
-	X6	Pre-Position	40	12 to 50
	X7	Venting	25 to 35	4 to 44

You can request the batter department to prepare batter for up to 10 cakes at one time. It will cost you one dollar for each batter department request and one dollar for each tower cake baked.

Find a way to make the Tower Cake rise to new heights.



Eg. Etching stain defect

 Three etching factors are varied to observe the three responses, namely total thickness variations (TTV), etching removal, and wafer brightness after etching.

The target for TTV is to achieve the lowest value as possible. Etching removal is aimed at 38 ± 2 um and finally the brightness is approximately 75 ± 2 %.

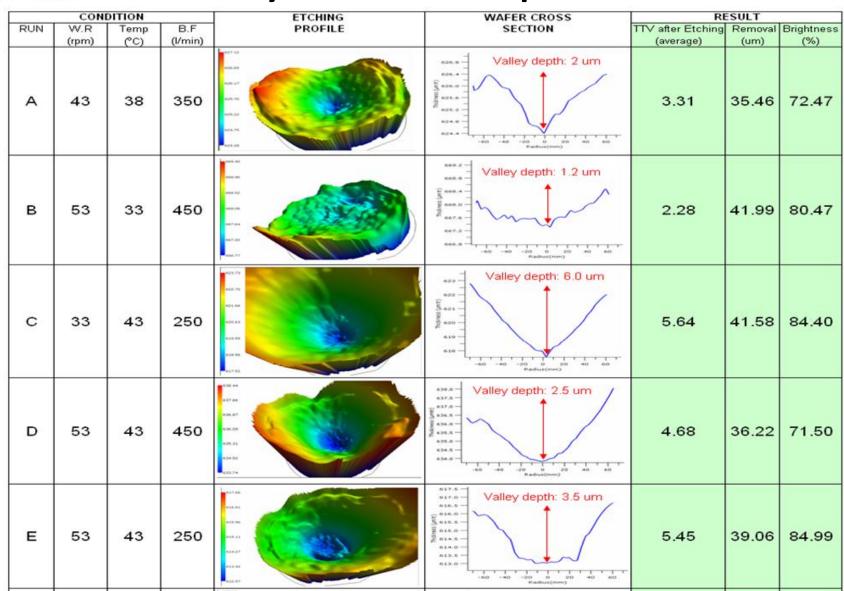


Analyze using DOE 2^3 with cp

RUN	Wafer Rotation at Etching (rpm)	Acid Temperature (°C)	Etching Bubling Flowrate (I/min)	TTV	Removal (um)	Brightness (%)
1	43	38	350	3.31	35.46	72.47
2	53	33	450	2.28	41.99	80.47
3	33	43	250	5.64	41.58	84.40
4	53	43	450	4.68	36.22	71.50
5	53	43	250	5.45	39.06	84.99
6	43	38	350	3.16	35.52	71.25
7	33	33	450	3.11	40.28	86.50
8	33	33	250	2.44	34.05	68.90
9	43	38	350	3.28	35.10	71.56
10	33	43	450	4.24	35.34	64.30
11	53	33	250	2.57	38.48	76.54



Analyze the responses:





Process Robustness



	Classification of Experimental Designs									
Purpose	Screening	Character	rization	Optimization	Robust- ness	Ruggedness Testing	Maintain Optimum			
Results	Identify Important Factors	Compare Process Changes for One Factor	Understand System Behavior	Prediction & Best Performance	Process Robustness	Release for Manufacturing	Keep Process on Target			
Design Type	Fractional Factorials	Comparative: t-tests; ANOVA; Regression	Full Factorials	Central Composite & Box- Behnkens	Factorials Taguchi	Fractional Factorials	EVOP: Replicated CCD & Box- Behnkens			
Resolu- tion	III or IV	NA	Full	V or Higher for Cube Portion	IV or V or Higher	Ш	V or Higher for Cube Portion			
Factors Explored	3 to 15	1 Factor or 1 Combination of Factors	2 to 6	2 to 5	2 to 5	3 to 15	2 to 3			
Type of Factors	Controllable	Controllable	Controllable	Controllable	Controllable & Noise	Noise	Controllable			
Effects Estimated	Main Effects	Main Effects	Main Effects, Interactions & Detection of Curvature	Main Effects, Interactions & Model Curvature	Main Effects, Interactions & Model Curvature	Main Effects	Main Effects, Interactions & Model Curvature			
Order of Applica- tion		\rightarrow \circ	$\circ \rightarrow$		roscope frame Simbal Rotor	xis .				



Three Main Ideas of Taguchi

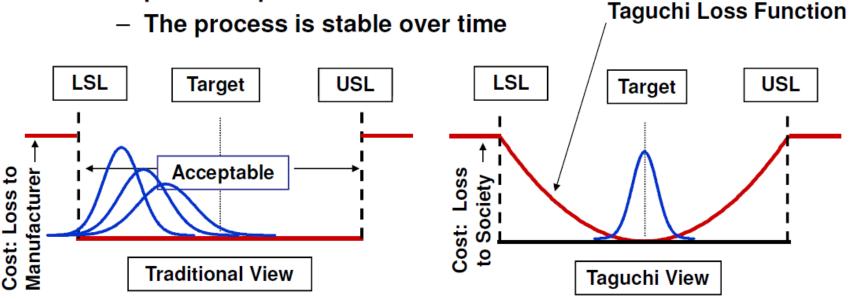
- Strive to reduce societal loss (i.e. products and processes should be on target with minimal variation), rather than simply be within specs.
- Design Products and Process that are Robust to external variation; have minimal variation and are centered on target.
- 3. Use experimental design methods as an engineering tool to improve robustness, decrease variation and center a process.





When Can You Tolerate Variation?

- There will always be variability present in any process.
- We can tolerate variability if:
 - The process is on target
 - The total variability is relatively small compared to the process specifications





Taguchi Loss Function

$$L(y) = L(m) + \frac{L'(m)}{1!} (y - m) + \frac{L''(m)}{2!} (y - m)^2 + \frac{L'''(m)}{3!} (y - m)^3 + \dots + \frac{L^{(n)} [m + \theta(y - m)]}{n!} (y - m)^n$$
for $0 < \theta < 1$

Note L(m)=0 and y=m is the minimum so that L'(m)=0. Neglecting terms of order higher than 2 gives:

$$L(y) = \frac{L''(m)}{2!} (y - m)^2$$
$$= k(y - m)^2$$

If
$$L(LSL) = L(USL) = L$$
 and $m = \frac{USL + LSL}{2}$, then $k = \frac{4L}{(USL - LSL)^2}$.



Taguchi Orthogonal Arrays (OA)

Taguchi uses highly fractionated factorial designs (Resolution III) to assess primarily main effects.

```
Notation: L<sub>a</sub>(b<sup>c</sup>)

a = # experimental runs;

b = # levels of each factor;

c = # of columns in the array.
```

$L_4(2^3)$	$L_9(3^4)$	$L_{16}(4^5)$
$L_8(2^7)$	$L_{18}(2^1 \times 3^7)$	$L_{32}(2^1 \times 4^9)$
$L_{12}(2^{11})$	$L_{27}(3^{13})$	$L_{64}(4^{21})$
$L_{16}(2^{15})$	$L_{36}(2^3x3^{13})$	$L_{25}(5^6)$
$L_{32}(2^{31})$	$L_{54}(2^{1}x3^{25})$	$L_{50}^{-1}(2^{1}x5^{11})$
$L_{64}(2^{63})$	L ₈₁ (3 ⁴⁰)	$L_{36}(2^{11}x3^{12})$

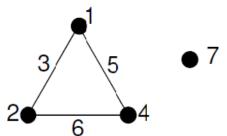


 $L_8(2^7)$

OA Column

ı	·	1	2	3	4	5	6	7
	1	1	1	1	1	1	1	1
	2	1	1	1	2	2	2	2
	3	1	2	2	1	1	2	2
	4	1	2	2	2	2	1	1
	5	2	1	2	1	2	1	2
	6	2	1	2	2	1	2	1
ı	7	2	2	1	1	2	2	1
ı	8	2	2	1	2	1	1	2

Linear Graph



founding Structure

27-4 Fractional Factorial

Factor Label

		Α	В	С	D	Е	F	G
	8	-1	-1	-1	1	1	1	-1
30w	7	1	-1	-1	-1	-1	1	1
JA F	6	-1	1	-1	-1	1	-1	1
ng (5	1	1	-1	1	-1	-1	-1
Corresponding OA Row	4	-1	-1	1	1	-1	-1	1
esbo	3	1	-1	1	-1	1	-1	-1
Sorr	2	-1	1	1	-1	7	1	-1
	1	1	1	1	1	1	1	1
		4	2	1	6	5	3	7

Corresponding OA Column

[Intercept] = Intercept + ABD + ACE + AFG + BCF + BEG (Resolution III) + CDG + DEF [A] = A + BD + CE + FG + BCG + BEF + CDF + DEG

[B] = B + AD + CF + EG + ACG + AEF + CDE + DFG

[C] = C + AE + BF + DG + ABG + ADF + BDE + EFG

[D] = D + AB + CG + EF + ACF + AEG + BCE + BFG [E] = E + AC + BG + DF + ABF + ADG + BCD + CFG



Taguchi Example

D.M. Byrne and S. Taguchi, *Quality Progress*, December 1987, pp. 19-26.

Goal: Improve a process to assemble and elastomeric connector to a nylon tube to optimize (maximize) the pull-off performance for automotive use.

Controllable Factors:

A: Interference	Low	Med	High
B: Connector wall thickness	Thin	Med	Thick
C: Insertion depth	Shallow	Med	Deep
D: % adhesive in connector pre-dip	Low	Med	High

Noise Factors

E: Conditioning time	24h	120h
F: Conditioning temperature	72°F	150°F
G: Conditioning relative humidity	25 %	75%





		Outer Array (L8)								
Ε	1	1	1	1	2	2	2	2		
F	1	1	2	2	1	1	2	2		
G	1	2	1	2	1	2	1	2		

Taguchi Example

Run	Α	В	С	D
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

	Data							Average	SN Ratio
15.6	9.5	16.9	19.9	19.6	19.6	20.0	19.1	17.525	24.025
15.0	16.2	19.4	19.2	19.7	19.8	24.2	21.9	19.425	25.500
16.3	16.7	19.1	15.6	22.6	18.2	23.3	20.4	19.025	25.335
18.3	17.4	18.9	18.6	21.0	18.9	23.2	24.7	20.125	25.904
19.7	18.6	19.4	25.1	25.6	21.4	27.5	25.3	22.825	26.908
16.2	16.3	20.0	19.8	14.7	19.6	22.5	24.7	19.225	25.326
16.4	19.1	18.4	23.6	16.8	18.6	24.3	21.6	19.850	25.711
14.2	15.6	15.1	16.8	17.8	19.6	23.2	24.2	18.313	24.828
16.1	19.9	19.3	17.3	23.1	22.7	22.6	28.6	21.200	26.152



Summary

- Robust Product/Process Design: Introduced Methods to determine settings of combinations of controllable factors that minimize the variation of response in the presence of variation introduced by uncontrollable variables.
- Discussed the Taguchi Approach to Robustness



Questions?



rozzeta.kl@utm.my rozzetadolah@gmail.com +6 017 361 7855

Thank you!