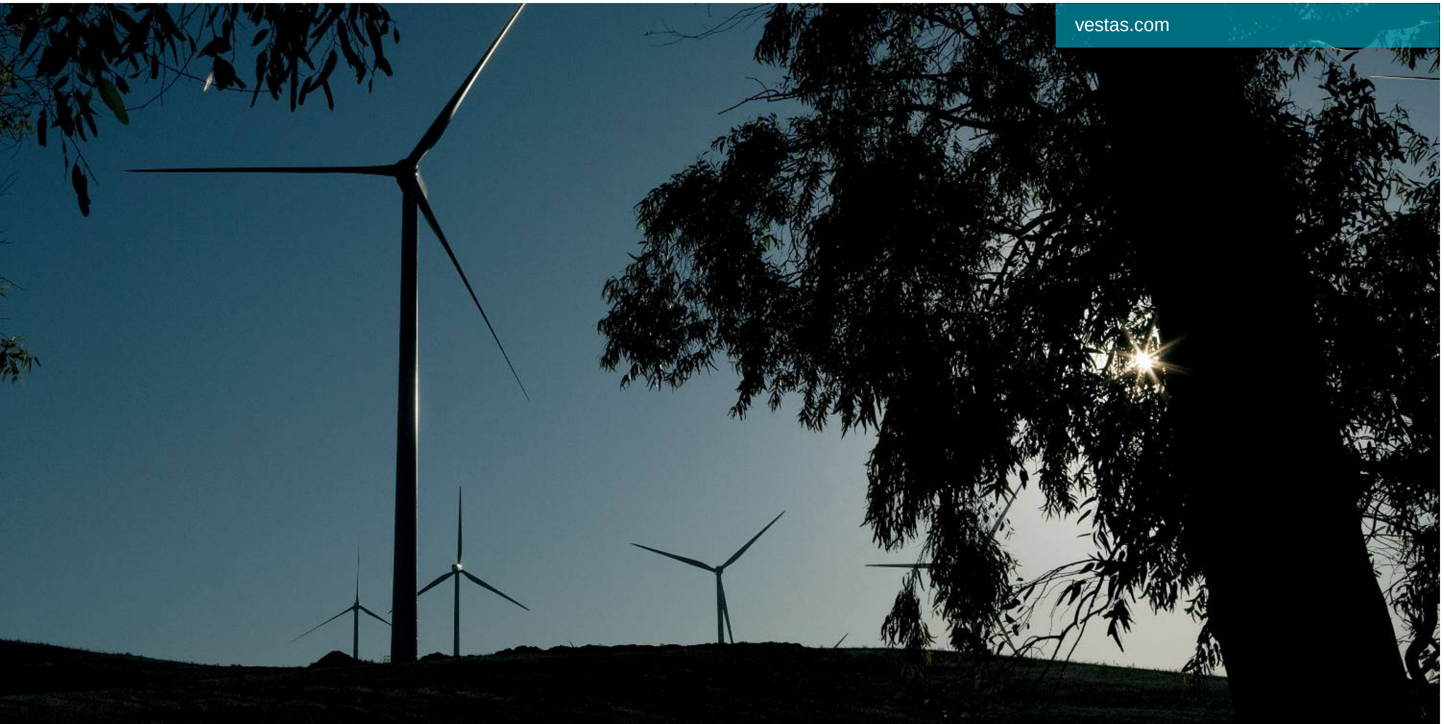
No. 1 in Modern Energy

Hypothesis Testing on Averages, Medians, and Variation



vestas.com

Vestas®

Learning Objectives

At the end of this section delegates will be able to:

- Formulate a Null and Alternative Hypothesis
- Understand Type I and Type II errors
- Understand alpha and beta risk
- Recognise a range of different types of Hypothesis Tests
- Determine the necessary sample size

Agenda

- Introduction to Hypothesis Testing
- Testing between Averages
- Steps in Hypothesis Testing
- T tests
- ANOVA
- Testing between Variances
- Testing between Proportions

DMAIC Improvement Process

Define

- Select Project
- Define Project Objective
- Form the Team

- Map the Process
- Identify Customer Requirements

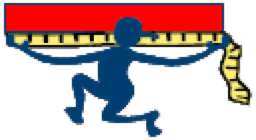
- Identify Priorities
- Update Project File

Phase Review

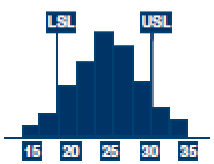
Measure

- Define Measures (y's)

- Evaluate Measurement System



- Determine Process Stability
- Determine Process Capability

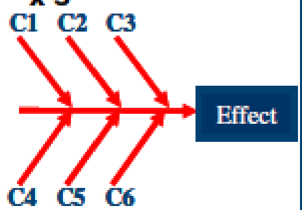


- Set Targets for Measures

Phase Review

Analyse

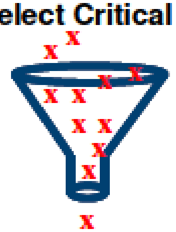
- Identify Potential x's



- Analyse x's

Run	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
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

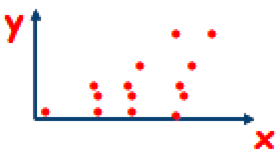
- Select Critical x's



Phase Review

Improve

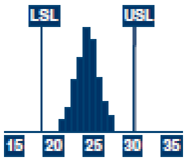
- Characterise x's



$y=f(x_1,x_2,\ldots)$

- Optimise x's

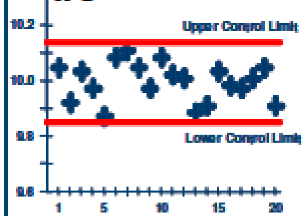
- Set Tolerances for x's
- Verify Improvement



Phase Review

Control

- Control Critical x's



- Monitor y's

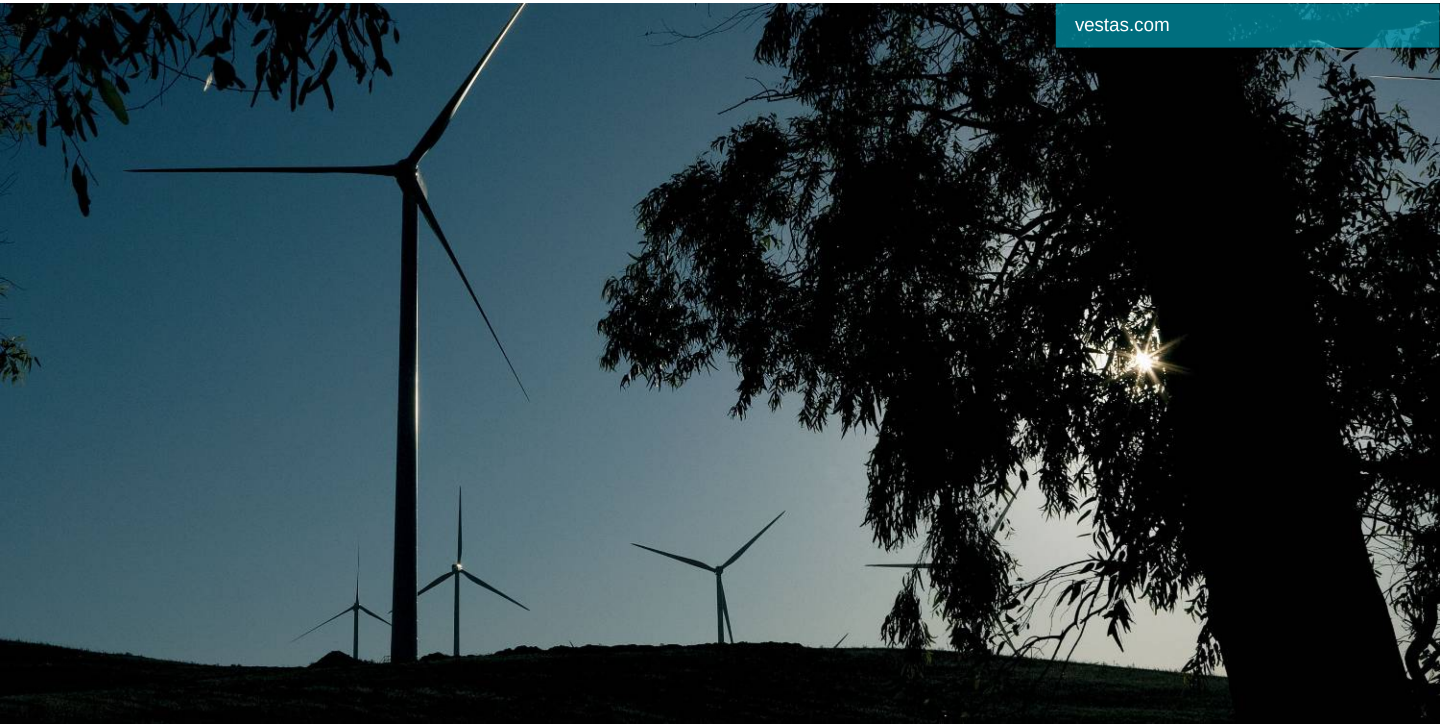
- Validate Control Plan

- Close Project

Phase Review

Vestas

Introduction to Hypothesis Testing



Vestas

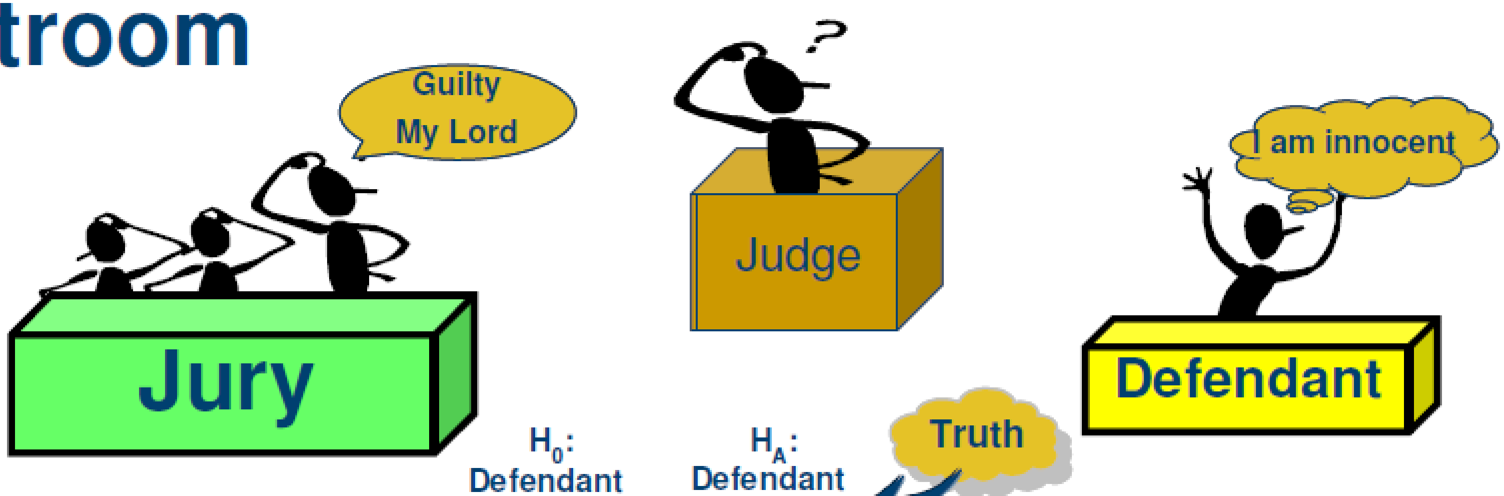
What do Hypothesis Tests do…

Hypothesis tests assess statistical significance, which enables quantitative decisions to be taken.

They add rigour to information suggested by graphs and charts.

- Hypothesis testing is a procedure for detecting differences
- Within the DMAIC we can use it to:
  - Determine if the output (y) meets a standard
  - Determine if two outputs (y's) are significantly different
  - Determine if a change in an input (x), significantly changes the output (y)
  - Determine if the data set follows a particular distribution
  - Determine if there is a "real difference" or just common cause variation

Hypothesis Testing Example: Courtroom



	H_0: Defendant Innocent	H_A: Defendant Guilty
Accept H_0: Innocent	✓ Correct Decision	✗ Wrong Decision (Type II error) -? risk
Reject H_0: Guilty	✗ Wrong Decision (Type I error) α - risk	✓ Correct Decision (Power)

Decision

Hypothesis Testing Can Produce Errors Just Like A Jury Can

How Confident do you want to be…

- The confidence level you choose depends on the risks you are willing to accept, and these are dependant on the situation
- How confident would you want to be if:
  - Large returns were likely for minimal cost
    - Tests indicate a large investment is required
    - A life was at risk (safety critical)

A typical confidence level for non-safety critical situations is 95%

Comments on Hypothesis Testing

Hypotheses represent the translation of a practical question into a statistical question.

In this manner, the “real world” problem is represented in terms which are suitable for scientific examination and testing.

In essence, hypotheses are statements related to the parameters of a given probability distribution; eg the mean and/or variance.

In other words, hypotheses are statements which allow us to represent all possible outcomes prior to conducting an investigation.

We accept or reject each hypothesis which provides a solid foundation for making practical “real world” decisions.

In the null form, the meaning of the hypothesis is associated with the distribution of chance events.

This hypothesis is referred to as the “null hypothesis” and designated as “H_o.”

Its meaning is that:

The parameters under investigation are equal

ie There is no difference in respect to the parameter of interest (mean and/or variance.)

Comments on Hypothesis Testing (continued)

In contrast to the null hypothesis is the alternate hypothesis (Ha).

These hypotheses are generally associated with distributions *other than* chance and, as such, are said to be "statistically significantly different" from the chance distribution.

This means any observed difference in the sample could not have resulted by chance variations inherent to the sample.

Then we conclude that sample is different from the qualifications necessary for the area of interest.

Therefore, we accept the alternate hypothesis of inequality and conclude the sample was from an area different from the one under investigation.

When accepting or rejecting the null and alternate hypotheses, we do so with a known degree of risk and confidence.

To do this, we specify (in advance of the investigation) the magnitude of decision risk (alpha, beta) and test sensitivity (d/s) which is acceptable.

Once this has been accomplished, we have the information necessary to determine a "rational" sample size.

Mathematical equations do exist for this purpose; however, we must balance such computations against the practical limitations of cost, time, and available resources in order to arrive at a "rational" sampling plan.

Types of Hypothesis Tests in Six Sigma…

Those that look at *AVERAGE / MEDIAN*

…….testing the location/centre/position of data

Those that look at *STANDARD DEVIATIONS*

…….testing the spread/variation of data

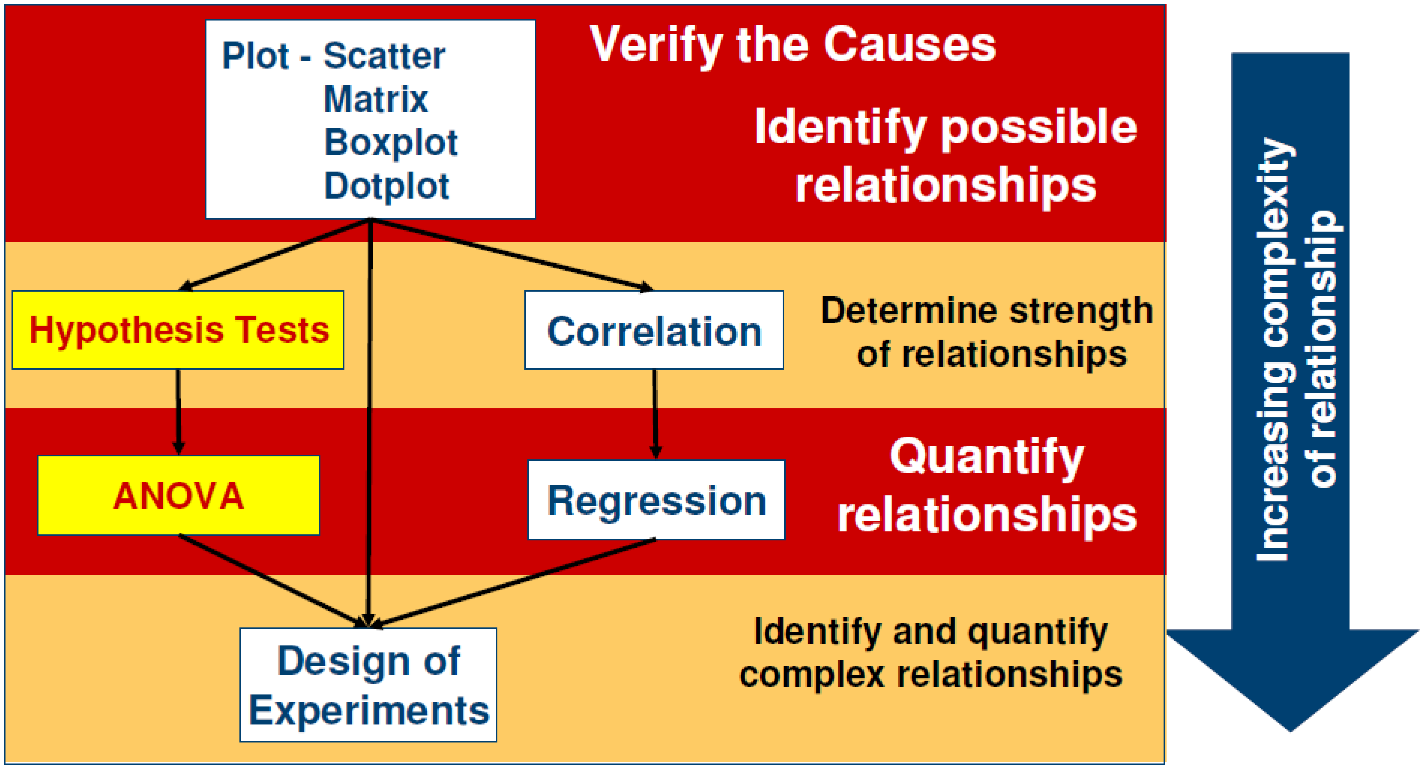
Those that look at *PROPORTIONS (or percentages)*

…….testing the proportions of data

Those that look at *DISTRIBUTIONS*

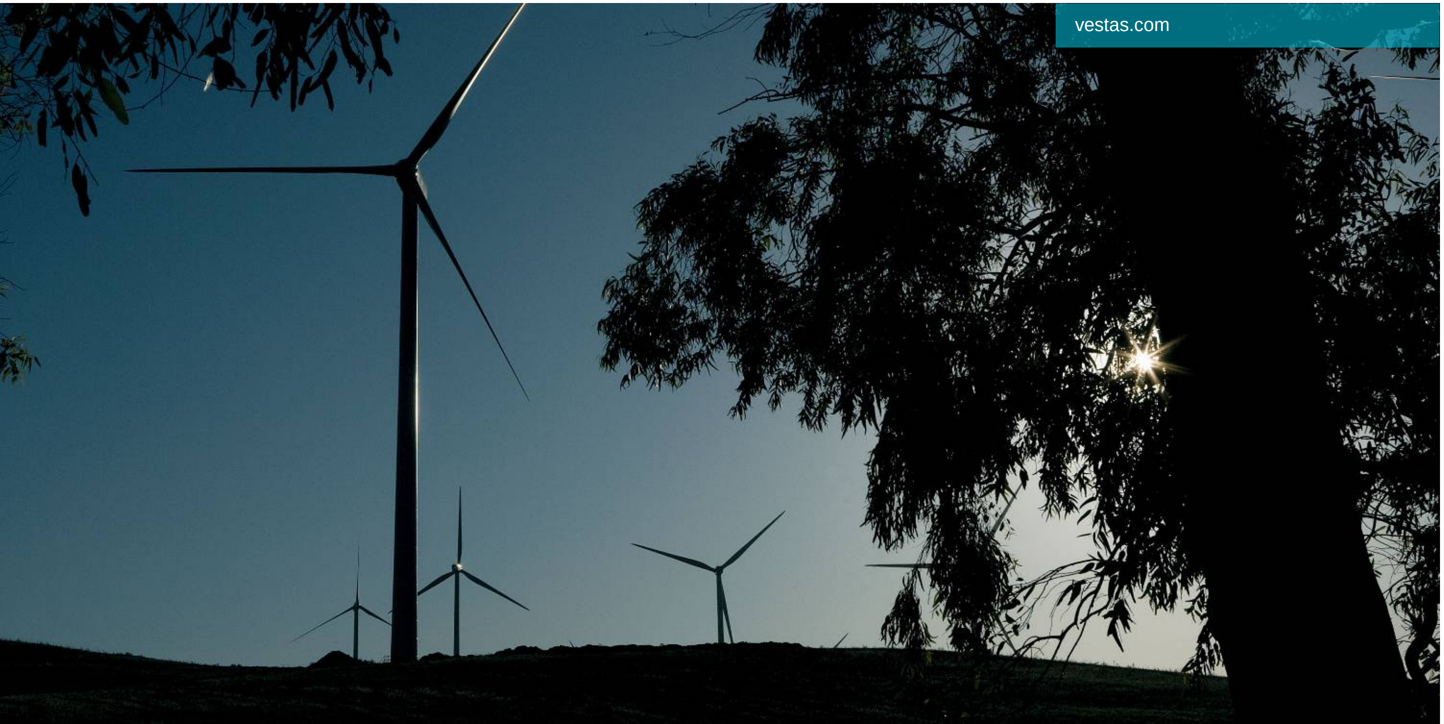
……..testing if data is Normally distributed

Tools for Verifying the Causes



No. 1 in Modern Energy

Testing between Averages



Objectives of this Section

- To introduce a range of hypothesis tests suitable for testing for differences between averages
- Specifically:
  - 1 Sample t-test
  - 2 Sample t-test
  - Paired t-test
  - ANOVA

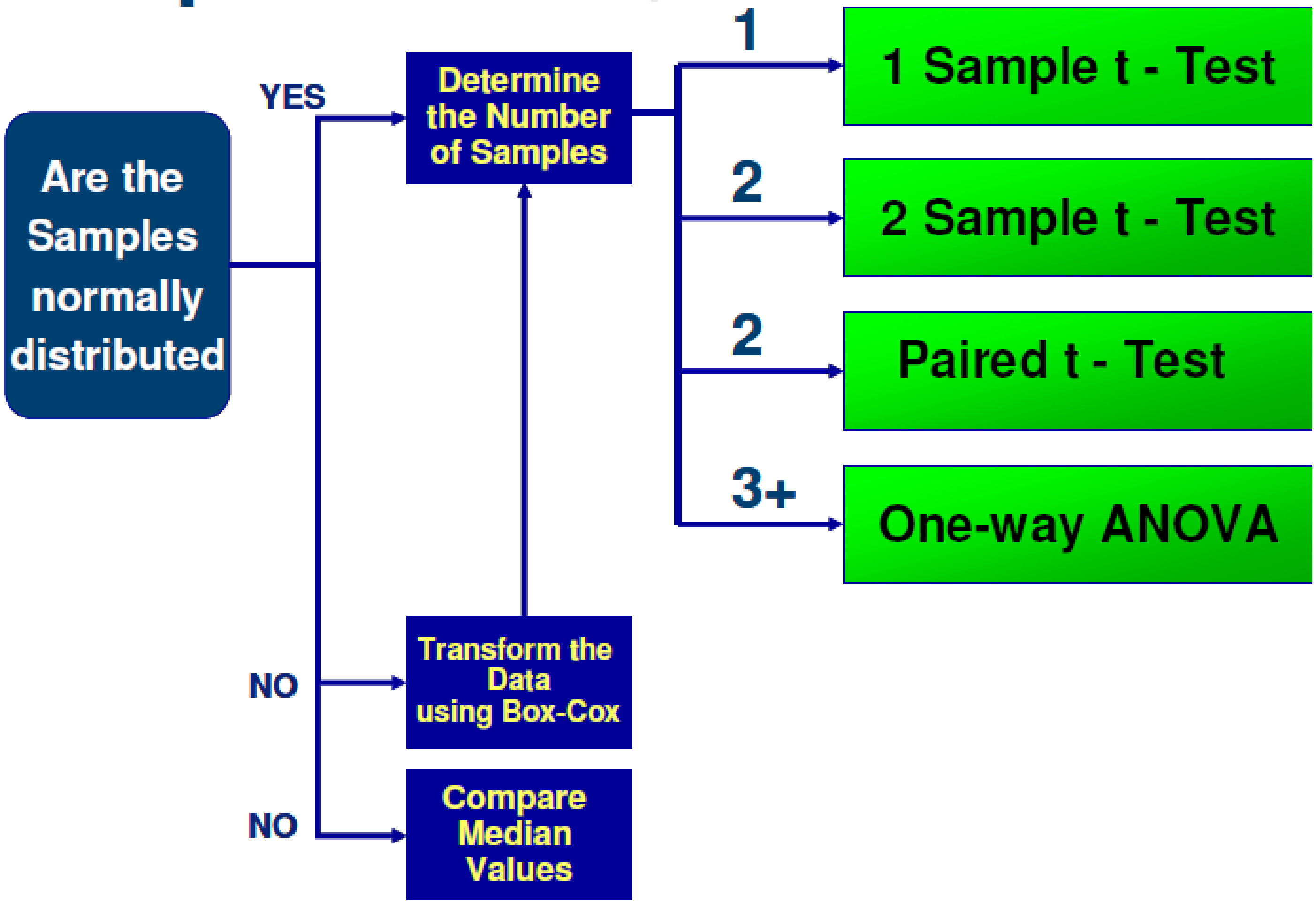
Hypothesis Tests for Average

Before a hypothesis test for “averages” can be selected, there are two preliminary questions:

1. Are the samples Normally distributed?
   Each of the samples that are to be included in the test must contain Normally distributed data in order for the results of the test to be statistically valid.

2. How many samples do you want to compare?
   The “number of samples” refers to the number of different samples (subgroups) that are to be compared, *not* the amount of data that is in each of those samples (that’s sample size).

To Compare the Averages of Samples of Data



An Expensive Situation

- You manage a warranty claims department. A customer claims loss of earnings of $1,245 for an item which usually is about $1,000
- You examine 250 previous claims of the same item to make a comparison and find the average to indeed be $997 with a standard deviation $88
- You want to know if the customer is over-claiming or if it is reasonable

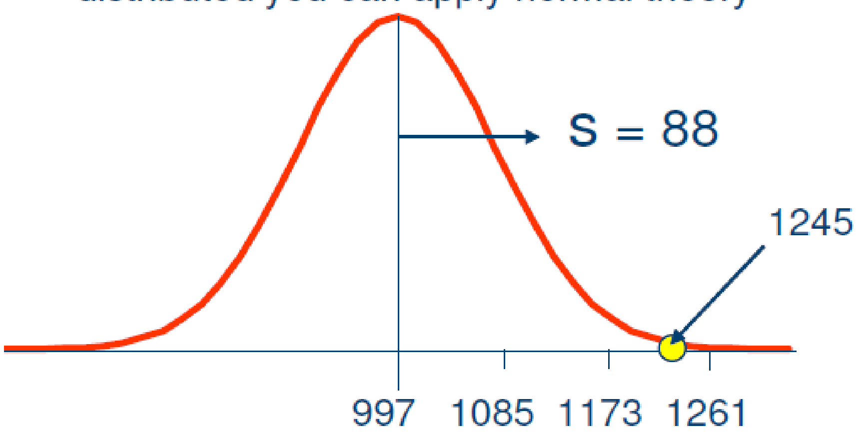
Innocent or Guilty?

- If the customer is not over-claiming (it is a legitimate claim) then we would expect the claim to fit with the pattern of data represented by the previous 250 claims

- If the customer is over-claiming then we would expect the claim to not fit the pattern of data from the previous 250 claims

Does the Claim fit the Pattern?

You remember from previous module that if the data is normally distributed you can apply normal theory



By using standard Normal tables we can calculate the probability of a claim of $1245 based on the historical data. If the probability is low, then we can assume an Illegitimate claim

$$Z = \frac{X - \bar{X}}{s}$$

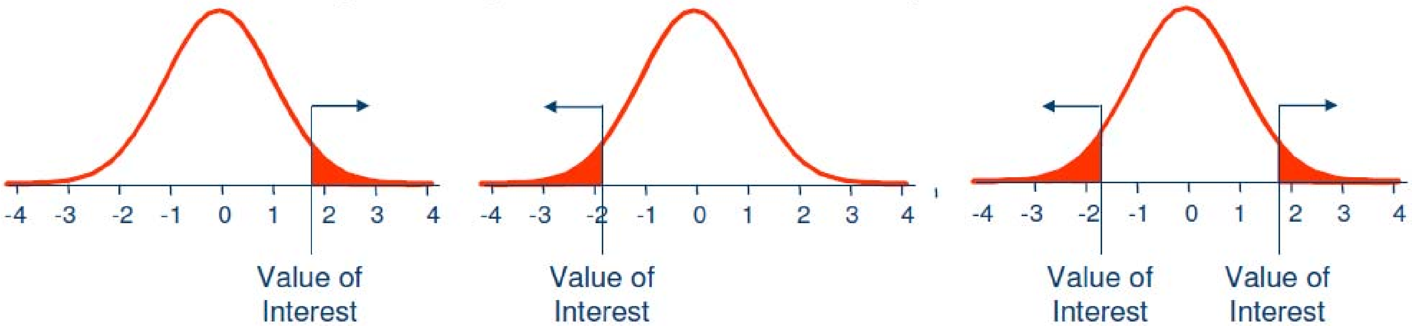
$$Z = \frac{1245 - 997}{88} = 2.82$$

From Standard Normal tables the P -value, the probability of being equal to or greater than 1245 is 0.0024 or 0.24%. In other words we would expect such a claim to happen 1 in 417 claims

P-values are Probabilities of Interest

P-value

- Tail area
- Area under curve beyond value of interest
- Probability of being at value of interest or beyond



"Greater than" **"Less than"** **"Not equal to"**

Making the Decision

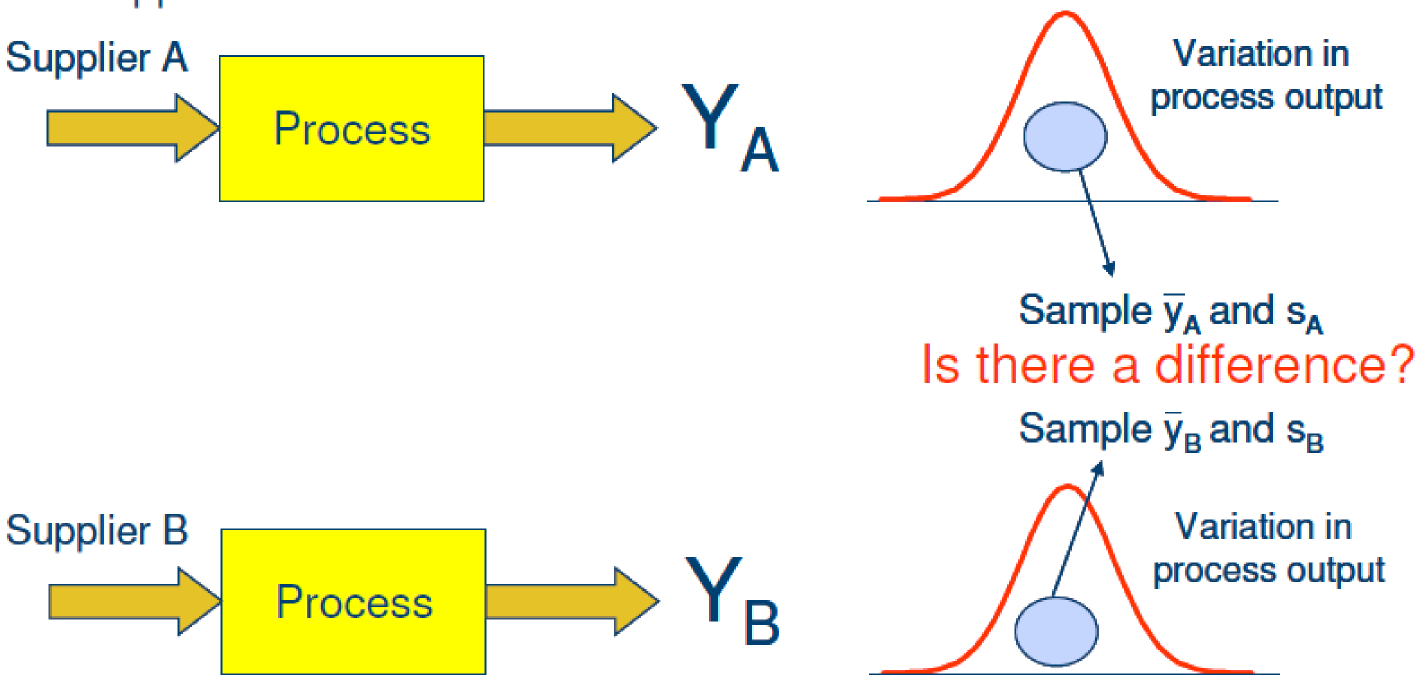
- The Normal theory is telling us that based on the previous 250 claims, we should expect a claim of $1245 or greater every 417 claims
  Hence there are 2 options:
  - This claim is legitimate – it is that 1 in 417
  - It is not legitimate - it does not fit the previous data
- Experience shows a small p-value (0 to 0.05) means
  - The probability is small that the value of interest comes from that distribution by chance therefore something else is going on
  - Since our p-value 0.024 < 0.05 we can conclude that the claim is not legitimate

What have we done?

- We have used the properties of the Normal distribution to test whether the occurrence of an event could have happened:
  - By chance (the data fits the expected pattern)
  - Or there is a real difference (the data does not fit the expected pattern)
- This type of situation occurs frequently during Six Sigma improvement projects, either in the
  - Analyse phase when we are *looking for differences* to identify potential roots causes
  - Improve and Control phases when we are aiming to demonstrate that a real change has been made – *we have made a difference*

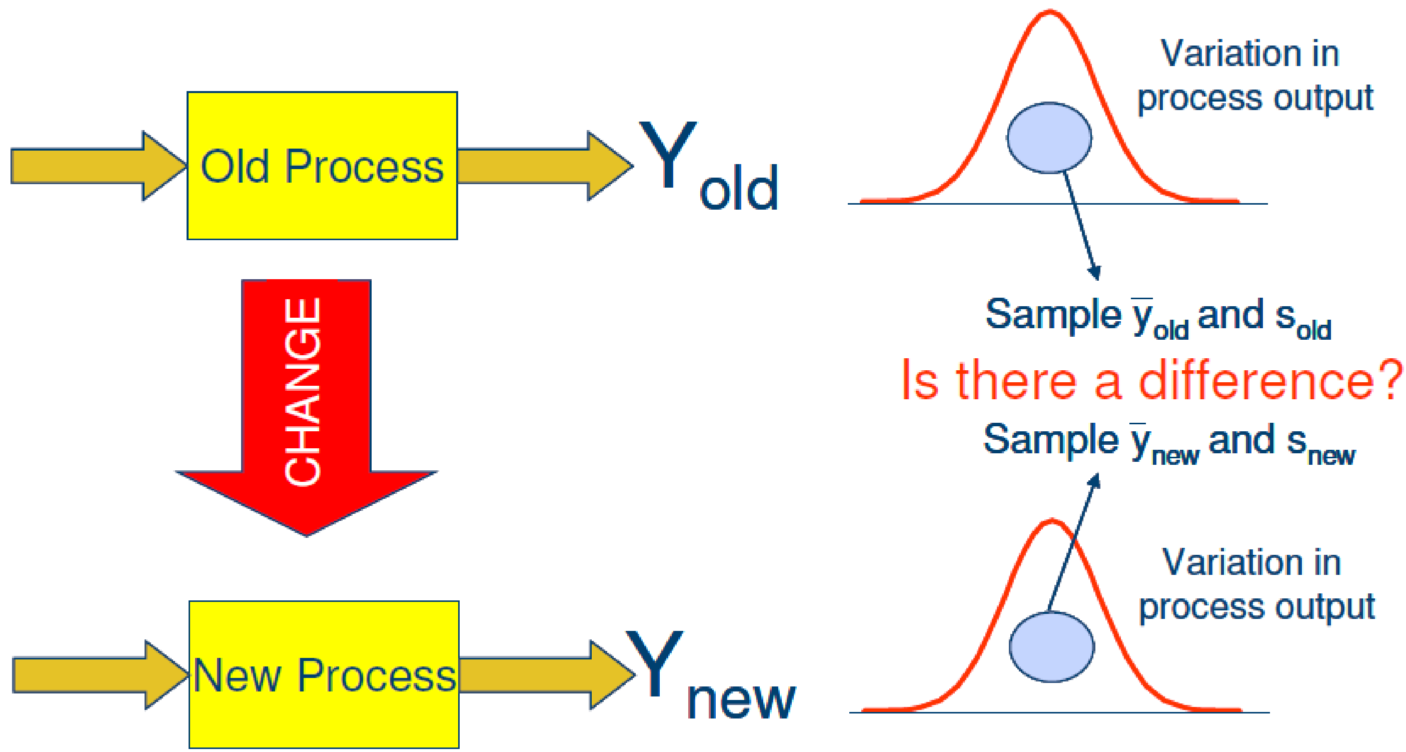
Analyse: Is there a difference?

- A common question during the Analyze phase is "is there really a difference?"
- For example: We suspect the output of a process depends upon the supplier



Improve: Have we made a difference?

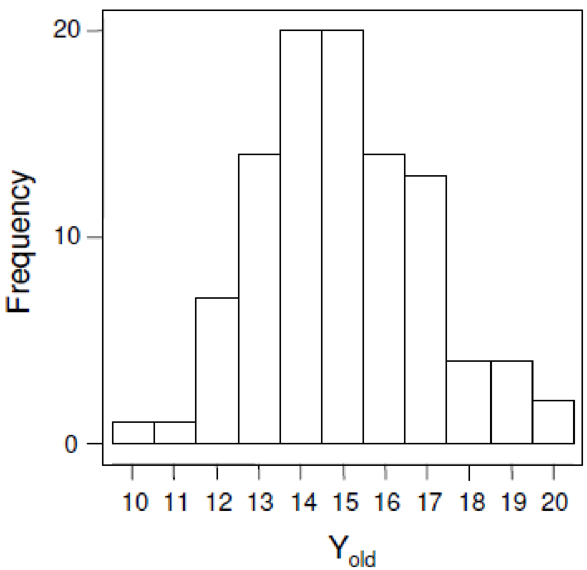
A common question having improved a process is “have we really made a difference?”

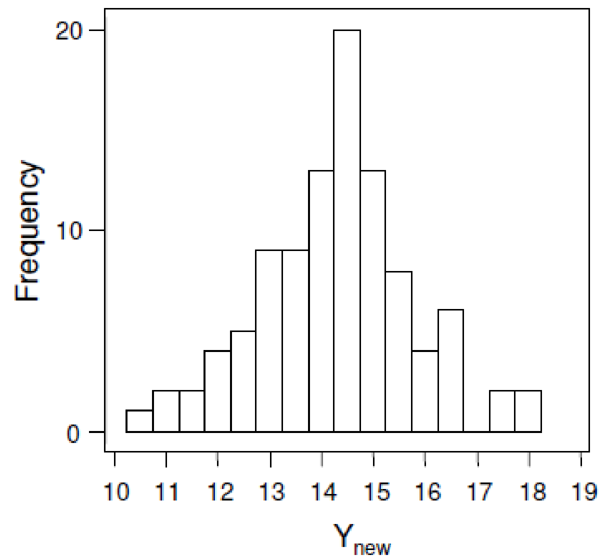


Vestas®

Is there a difference?

Looking at the histograms may give us the idea that there is a difference

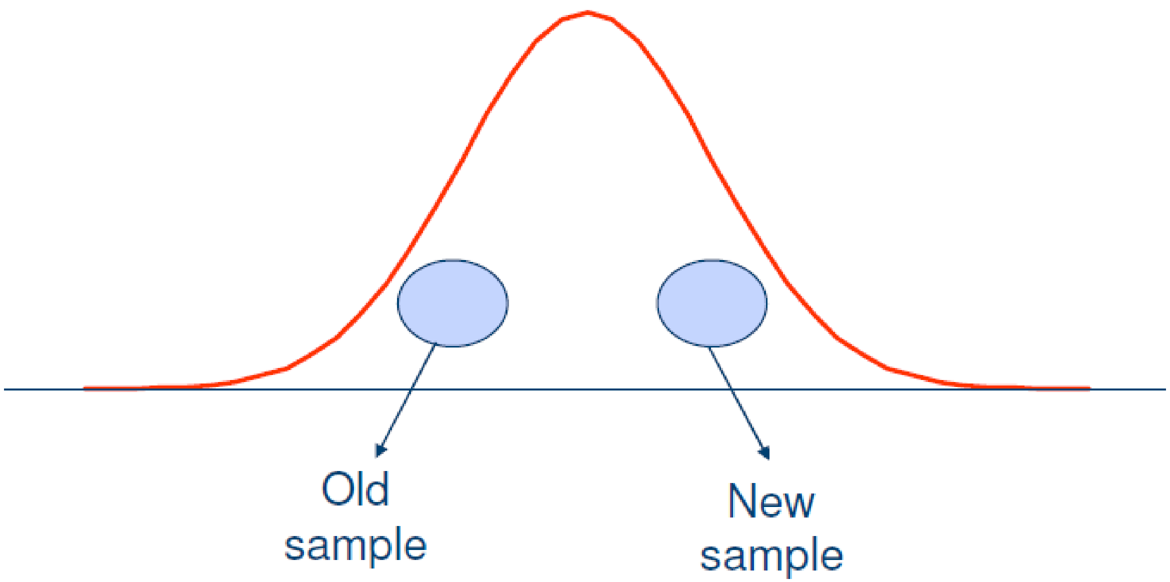




Vestas®

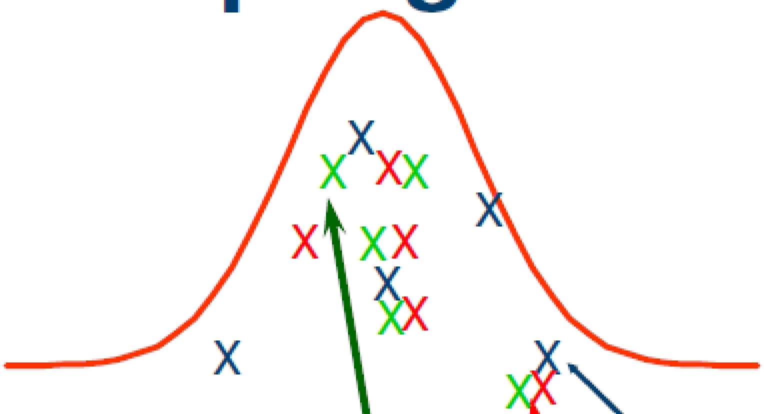
A difference?

It is possible that we have taken two samples from the same distribution



While on face value they look different statistically they are not!! The difference is due to chance (sampling)

Sampling Theory



Consider a population with a mean μ and standard deviation σ.
It may or may not be Normal

If we take a sample it will have a mean $\bar{x}$

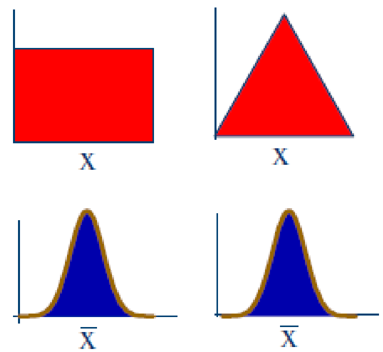
If we take a second sample it will have a mean $\bar{x}$

If we take an j^{th} sample it will have a mean $\bar{x}_j$

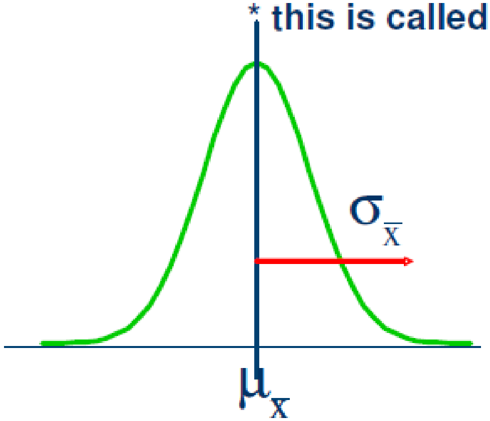
Any one of these sample means is an estimate of the true mean μ

Sampling Distribution of Means

The sample means will all be different and will have their own distribution - this distribution will tend to be normal - irrespective of the original distribution*



*** this is called the Central Limit Theorem**



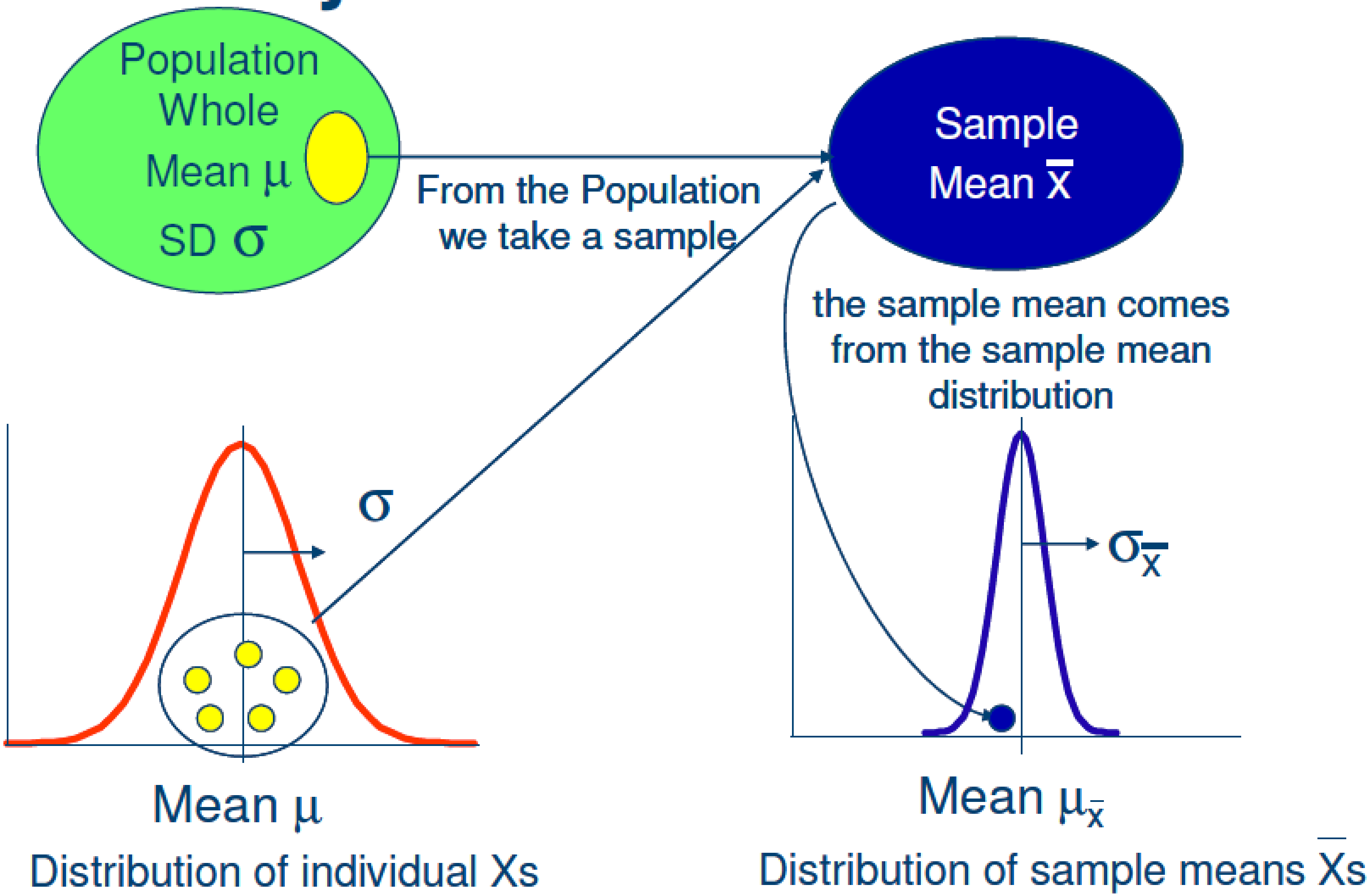
$$\mu = \mu_{\bar{x}}$$

$$\sigma_{\bar{x}} = \sigma/\sqrt{n}$$ **n = sample size**

The sampling distribution will have a mean $\mu_{\bar{x}}$ and standard deviation $\sigma_{\bar{x}}$

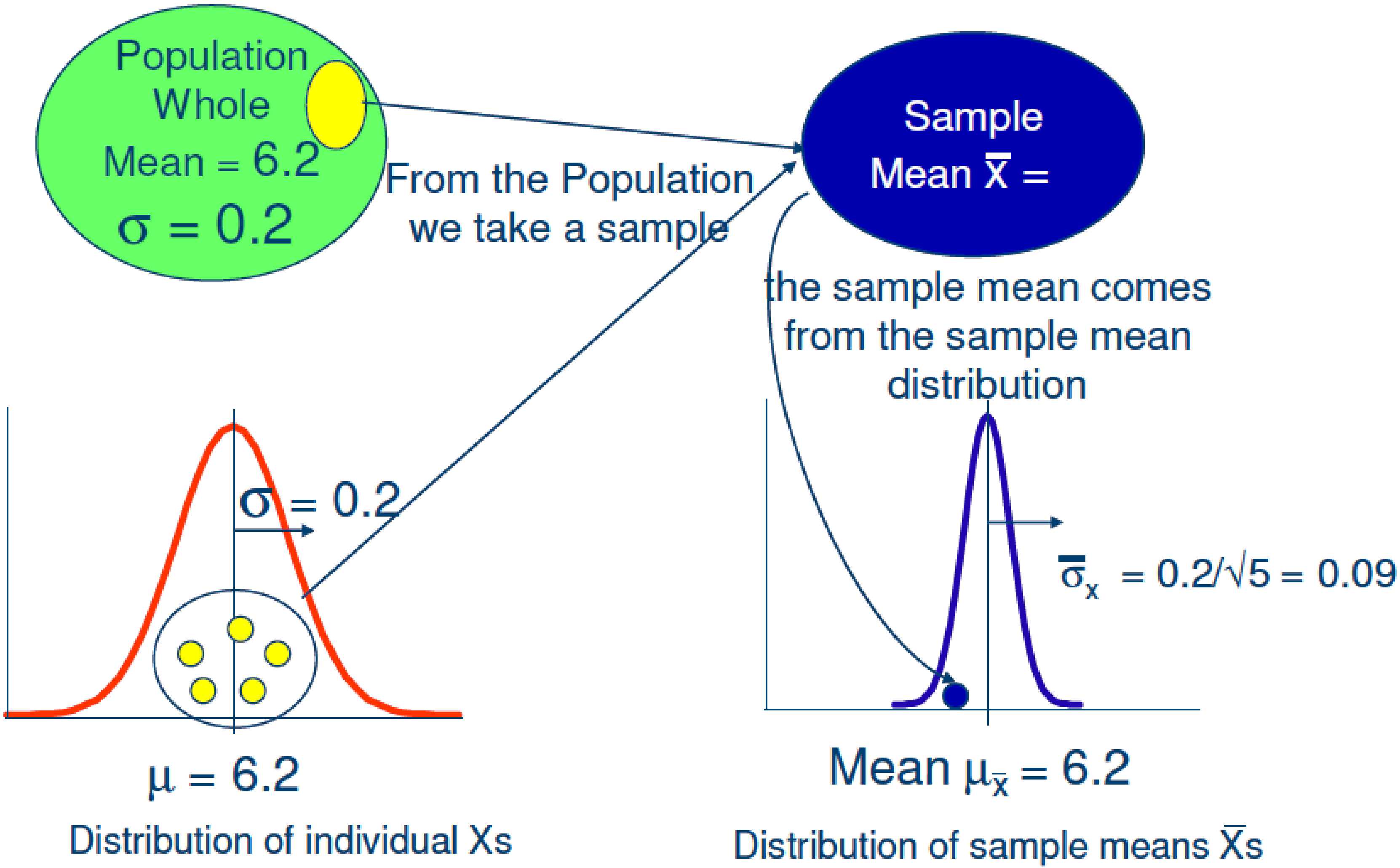
This standard deviation is important since it defines how close our estimates of the mean are to the true mean and is called the standard error.

Summary

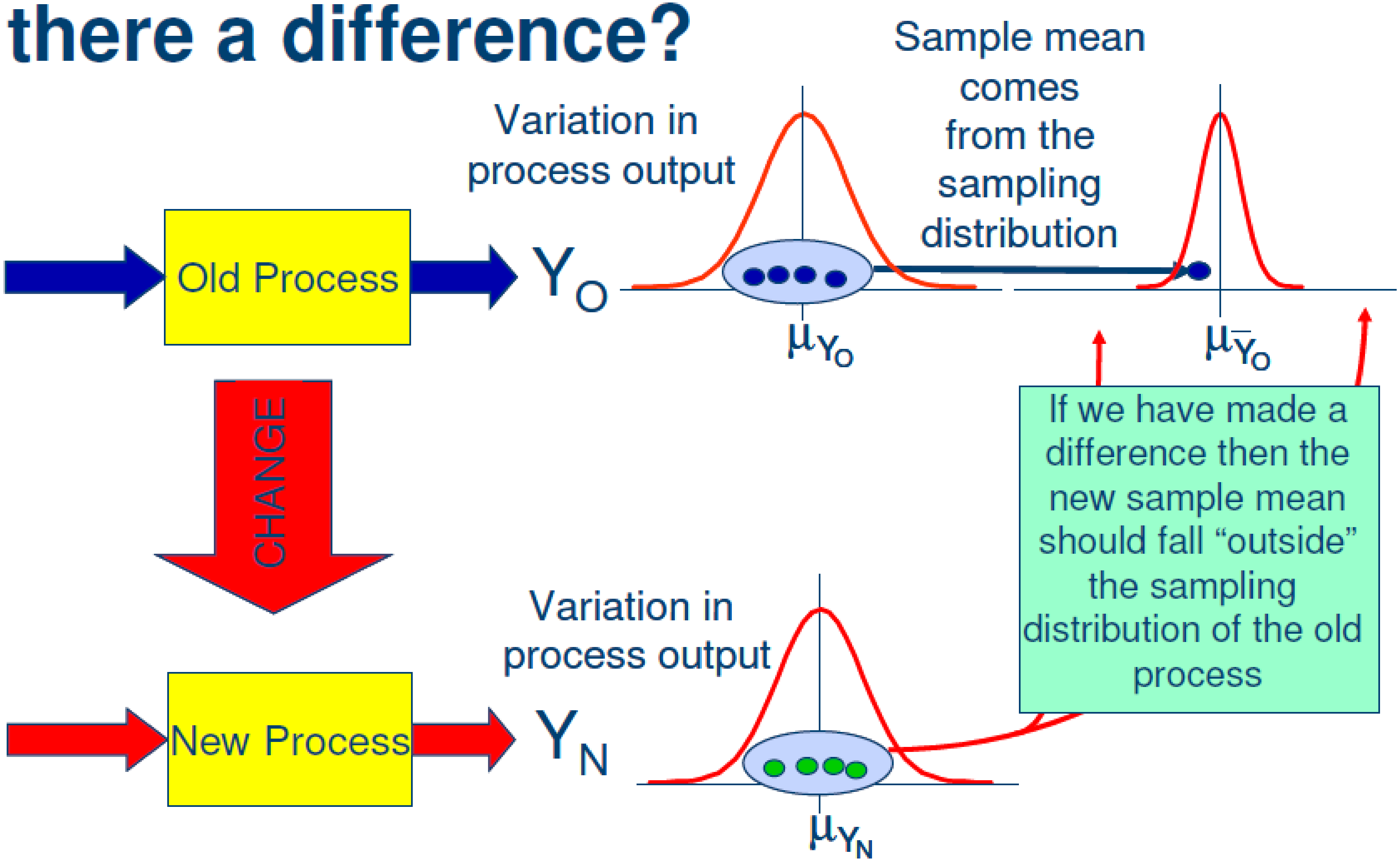


Population
Whole
Mean μ
SD σ
From the Population
we take a sample
Sample
Mean $\bar{x}$
the sample mean comes
from the sample mean
distribution
σ
$\sigma_{\bar{x}}$
Mean μ
Distribution of individual Xs
Mean $\mu_{\bar{x}}$
Distribution of sample means $\bar{X}$s

Example


Population
Whole
Mean = 6.2
σ = 0.2
From the Population
we take a sample
Sample
Mean X̄ =
the sample mean comes
from the sample mean
distribution
σ = 0.2
σ̄x = 0.2/√5 = 0.09
μ = 6.2
Distribution of individual Xs
Mean μx̄ = 6.2
Distribution of sample means X̄s

Using Sampling Distributions is there a difference?



Other Sampling Distributions

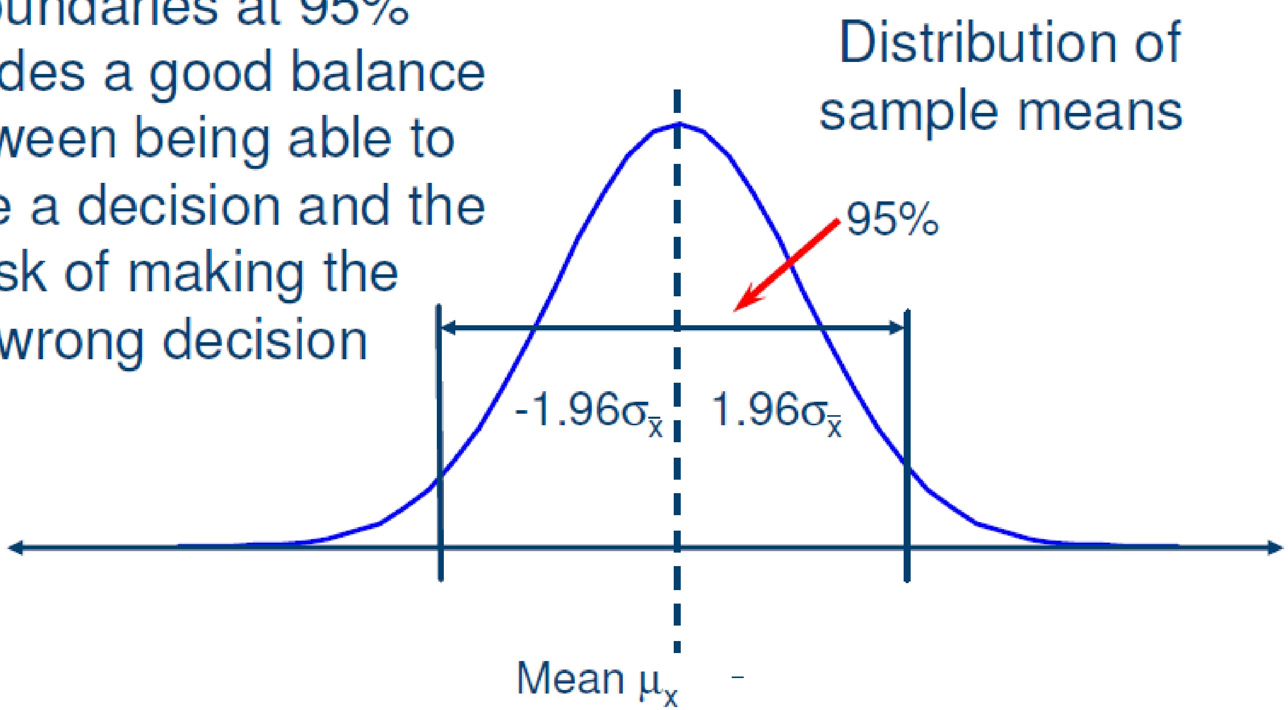
- The ideas developed so far have been based of the distribution of sample means - there are other useful sampling distributions
  - Difference between two means $\mu_1 - \mu_2$
  - Ratio of standard deviations σ_1/σ_2
- We shall be considering these later

Hypothesis Testing Concept

- In order to show that an apparent difference is real or due to chance we start by assuming that there is no difference (Null Hypothesis, H_0)
  - If the observed difference is within that expected by chance then the Null Hypothesis of no difference is correct
  - If the observed difference is greater than that expected by chance then the Null Hypothesis of no difference is not correct

95% Confidence

Experience has shown that setting the decision Boundaries at 95% provides a good balance between being able to make a decision and the risk of making the wrong decision



Risk

A point here could be due to chance or a difference

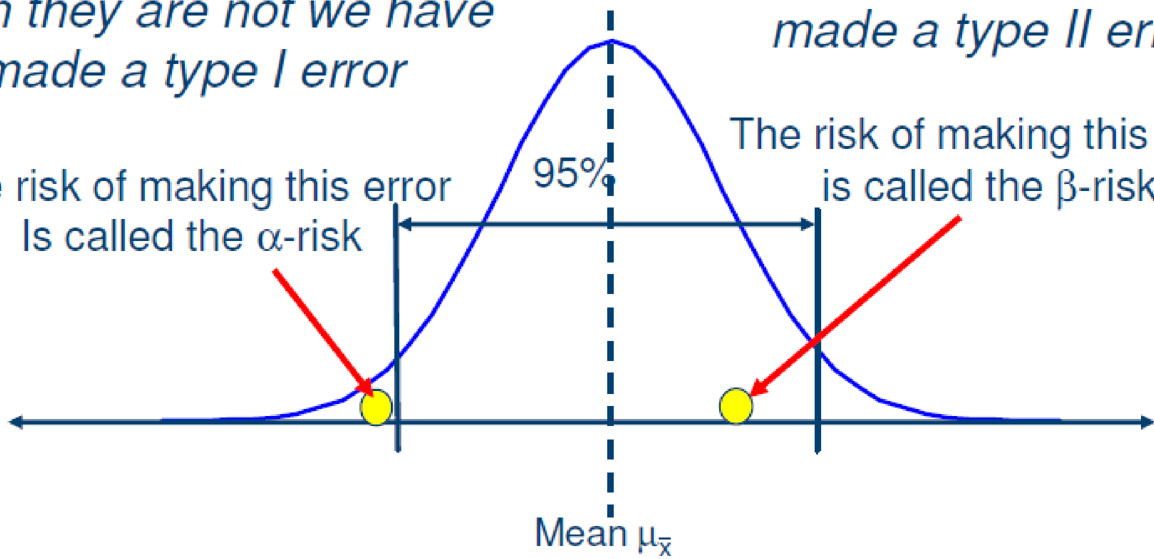
If we decide they are different when they are not we have made a type I error

The risk of making this error Is called the α-risk

A point here could be due to chance or a difference.

If decide they are the same when they are not we have made a type II error

The risk of making this error is called the β-risk



Type I and II Errors

- Type I - Deciding they are different when they are not
- Type II - Not detecting a difference when there is one
- Type I – Producer's risk
- Type II – Consumer's risk
- P-value = the actual probability of making a Type I error
- Both are important
- Guarding too heavily against one increase the risk of the other
- Increasing the sample size
  - Reduces the likelihood of Type II errors
  - Allows you to detect smaller differences

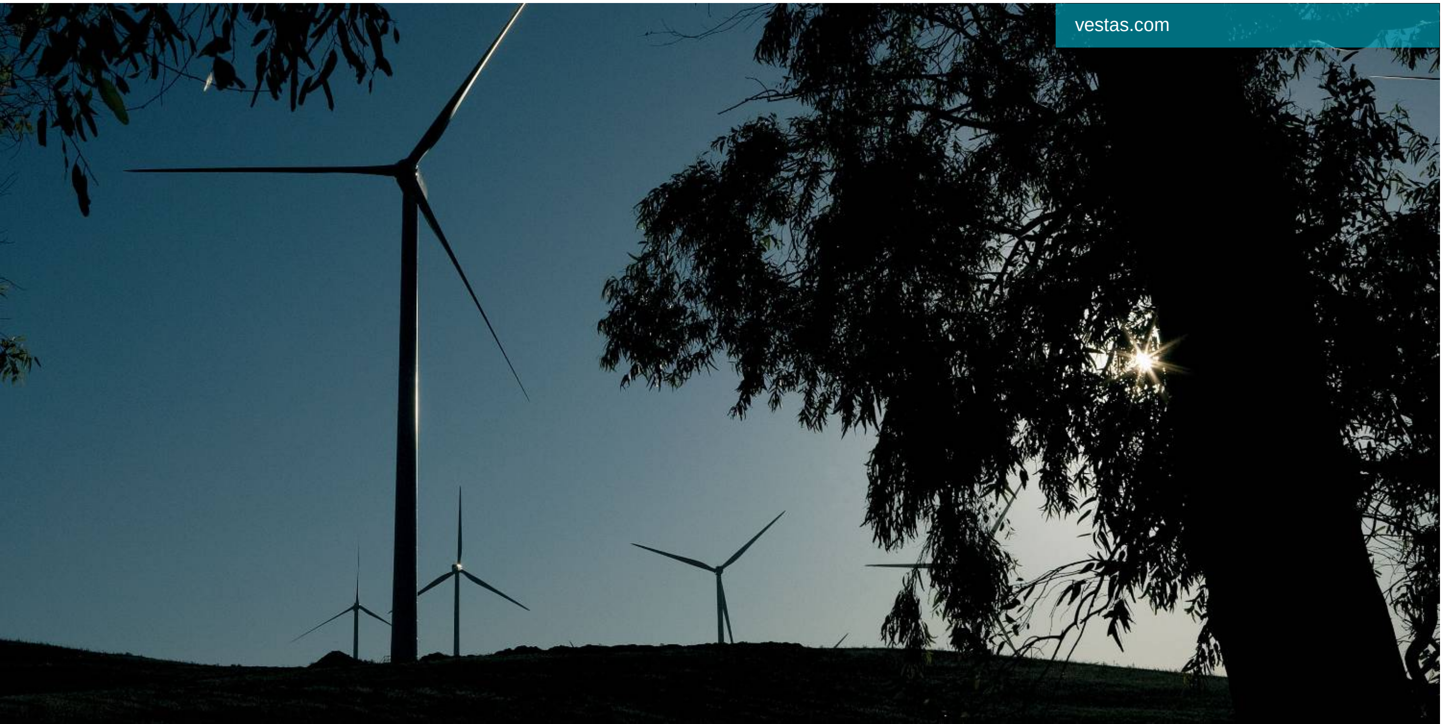
Hypothesis Tests

- Hypothesis tests can be used in a number of situations
  - A sample is taken and found to have a mean $\overline{x}$
    - Is this value "equal" to a target value T?
  - Samples taken before and after a change
    - The two sample means are different - but are they?
    - The standard deviations are different - but are they?
  - Samples from a process for different stratification factors
    - The two sample means are different - but are they?
    - The two sample standard deviations are different - but are they?
  - Samples from a process for different stratification factors,
    - The proportion defective appears to vary with different factors – but does it?

Vestas

No. 1 in Modern Energy

Steps in Hypothesis Testing



Vestas

Steps in Hypothesis Testing

- Hypothesis tests follow the pattern
  - Determine the null hypothesis H_0
  - Determine the alternative hypothesis H_1 or H_a
  - Determine statistical test
  - State level of risk
  - Establish sample size
  - Collect data
  - Analyse data
  - Accept or reject the null hypothesis
  - Determine conclusion

Vestas®

1 The Null Hypothesis

- The null hypothesis is always stated so as to specify that there is no (null) difference
- When comparing means or variances the null hypothesis is:

$$H_0: \mu = \text{Target}$$

$$H_0: \mu_1 = \mu_2$$

$$H_0: \sigma_1 = \sigma_2$$

2 The Alternative Hypothesis

- The alternative hypothesis opposes the null hypothesis and can therefore has several options

For example

$$H_1: \quad \mu_1 < \mu_2$$

or

$$H_1: \quad \mu_1 > \mu_2$$

or

$$H_1: \quad \mu_1 \neq \mu_2$$

The choice depends upon the problem
Minitab gives the option to select these

3 Determine Tests

Hypothesis Test	Can be used to
t-test	Compare a group average against a target. Compare two group averages.
paired t-test	Compare two group averages when data is matched.
ANOVA (Analysis of Variance)	Compare two or more group averages.
Test for equal variances (F-test, Bartlett's test, Levene's test)	Compare two or more group variances.
Chi-square test	Compare two or more group proportions.

Vestas

3 Determine Test

- Hypothesis tests are used when the input or process (X) variable is discrete
- If the X data are continuous, use Regression analysis to judge whether they are related to the output (Y) variable

		X: Discrete ("Groups")	X: Continuous
Y	Continuous	*t*-test Paired *t*-test ANOVA	Regression
Y	Discrete (Proportions)	Chi-Square	Logistic regression

4 State Level of Risk

- It is very important to specify before conducting the test the level of risk that the decision will be based on
- The risk means that there is the chance that we will make the wrong decision
- In assessing the risk level we need to consider the situation and the business – it is a question of significance vs. importance
- Significant means there is a real difference as proven by a hypothesis test
- Important means that the difference observed is important to the business
- Typically the α-risk is set at 5% which provides a 95% confidence

"Important" vs "Significant" Differences

- Significant but not important differences
  - Sometimes you will detect a statistically significant difference - yet it will be too small to be of practical importance to your business
  - Example - A project to reduce machine setup times
    - The average time for the old setup is 45 minutes. The new average setup time is 30 minutes. The observed difference of 15 minutes is statistically significant
    - However, to justify the cost of implementation, a reduction of 25 minutes is necessary

"Important" vs "Significant" Differences

- Important but not significant differences
- Sometimes you cannot claim a difference is statistically significant yet the observed difference is of importance to the business
- Example - Production volumes
  - An increase of 1000 items produced per day is observed during the pilot
  - An increase of 1000 is important to the business
  - However, the difference is not statistically significant
  - Either the observed difference is due to random variation and no true difference exists, or the the variation is too large (or sample size too small) to detect the difference
- You (and your business leaders) need to decide if it is worth the risk to go ahead and implement the new process

5 Establish Sample Size

- The ability to detect differences is related to the sample size
  - A small sample means we cannot detect small differences, but the cost of data collection is low
  - A large sample size will enable the detection of small differences, but the cost of data collection could be high
- Selecting the "right" sample size depends upon the Power = $(1 - \beta)$

6 Collect Data

- Remember all the rules and guidelines about collecting data
  - Sample size
  - Good operational definition
  - Clear measurement procedure
  - Un-biased samples
  - Good Gauge R&R

7 Analyse Data

- All hypothesis tests can be conducted by hand calculation – BUT
- Minitab rules!

8 Acceptance or Rejection

- Method 1

  if the p-value is > or = to α – fail to reject H_0

  if the p-value is < α - reject H_0 and accept H_1
- Method 2

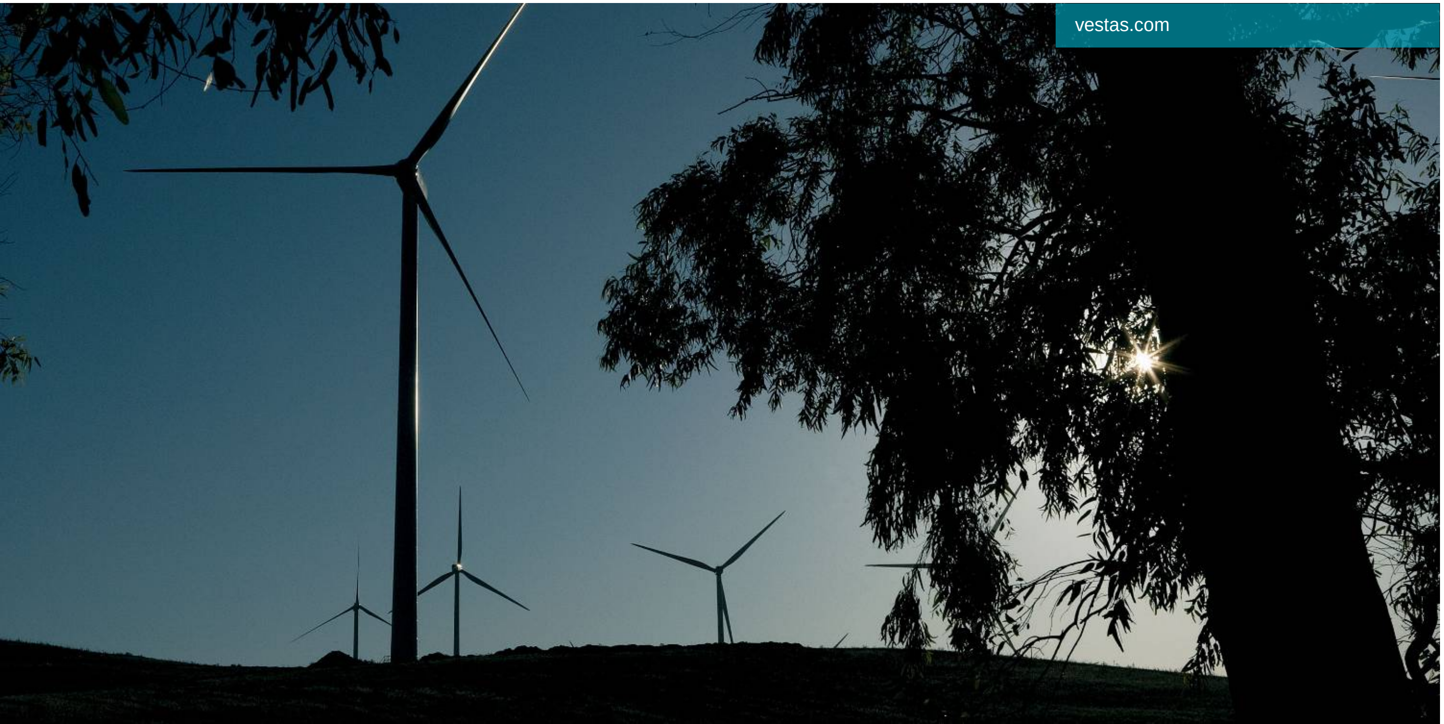
  if "0" falls within the confidence interval - fail to reject H_0

  if "0" falls outside the confidence interval - reject H_0 and accept H_1

- In practice we use Methods 1 and 2 from information provided by Minitab

No. 1 in Modern Energy

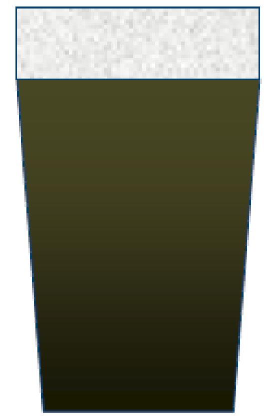
T-tests



Vestas

t- tests – A bit of history

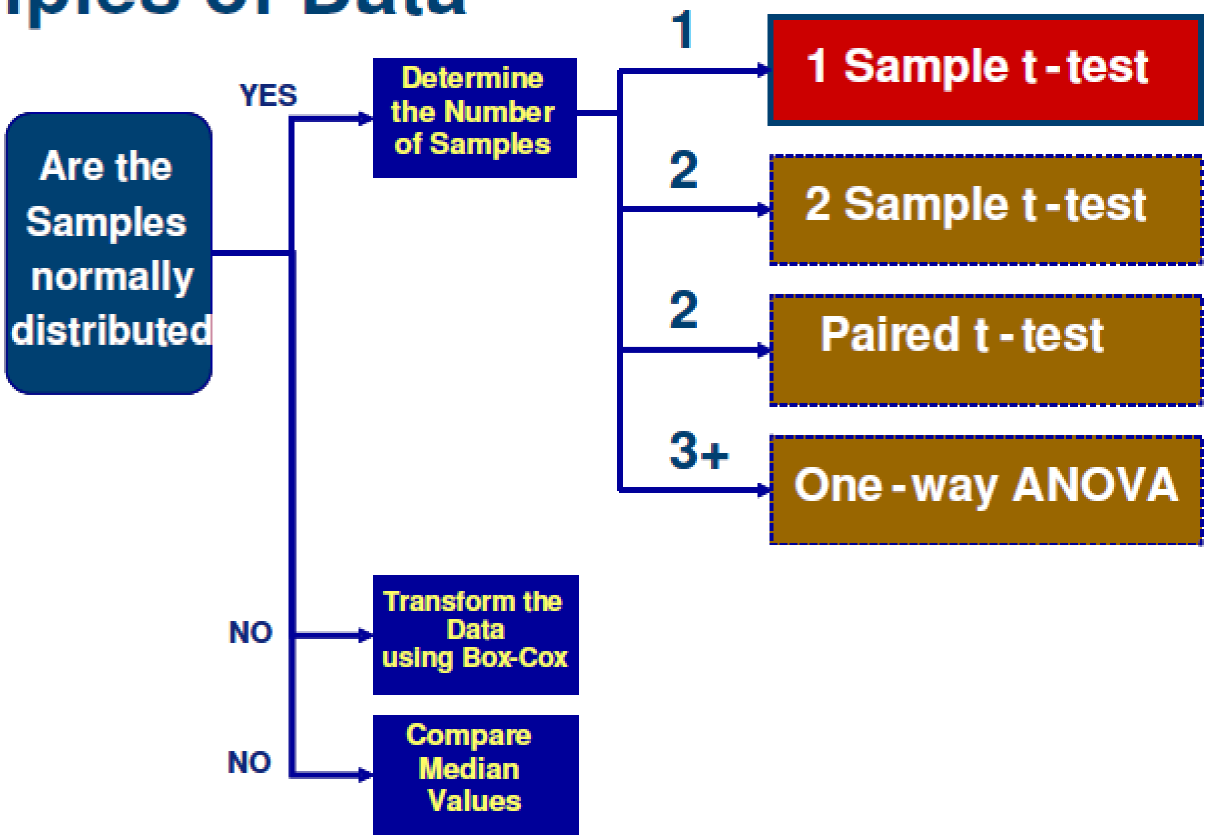
Mr Gossett (1876 – 1936) worked for the Guinness company and published his work under the pen name of Student, hence the “student t-test”.



The t-test measures the significance of shift in the *average.*

Vestas

To Compare the Averages of Samples of Data



1 Sample t-test (1)

- 1 Sample t-tests are used to compare the average of one sample of data against a known average
- The known average may be a historical average or an industry benchmark
- For the purpose of the test, the known average is assumed to be “exact”

1 Sample t-test - Example 1

- A new suppler claims their product will last on average 1000 hours with a standard deviation of 50
- We take a sample of 50 and find the average life is 986.6 hours with a standard deviation of 48.64
- What can we say about the supplier's claim given that our sample is only an estimate?

1 Sample t-test - Example 1

- To find the answer we would set up a hypothesis test

$H_0 : \mu = 1000$

$H_1 : \mu < 1000$

This alternative Hypothesis is chosen since we are interested in whether the product has a lower life than the claim. (less than)

- We use a 1 sample t-test when comparing data against a target where the population standard deviation is unknown

1 Sample t-test – Example 1

- We judge the difference between the group average and the target value using the t-test

$$t = \frac{\overline{X}_A - \mu_{\overline{X}}}{S_{\overline{X}}}$$

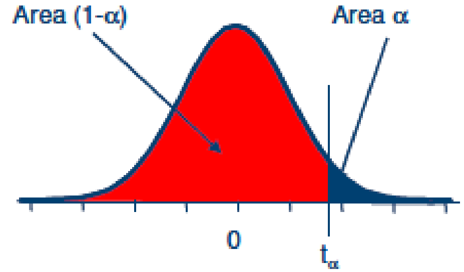
- Formula for t means that the p-values come from a t distribution, and $S_{\overline{X}} = \frac{S}{\sqrt{n}}$

$$t = \frac{986.6 - 1000}{48.64/\sqrt{50}} = -1.95 \text{ or } 1.95$$ **(symmetric distribution)**

t-distribution Tables

The t-distribution provides values of t_{α} and a specified number of degrees of freedom

t-distribution							
Dof/α (1-α)	.4 .600	.3 .700	.2 .800	.1 .900	.05 .950	.025 .975	.001 .990
28							
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462
30	0.256	0.530	0.835	1.310	1.697	2.042	2.457
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423
60	0.254	0.527	0.848	1.296	1.671	2.000	2.30
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326



In this case the sample size was 50, hence the dof = 49

Obtaining a p-value from the table is difficult. In this case since the calculated t value (1.95) lies in the area shown and therefore the p-value is between 0.05 & 0.025

This shows the p-value to be below the 0.05 risk level, hence we must reject the null hypothesis and accept the alternative

1 Sample t-test – Example 2

- We are coating medical implants with a measured weight of an anti-rejection drug
- We want to claim that the *mean* dosage is 1.0mg ± 0.1mg
- A random sample of 50 implants gives a mean weight of 0.98 mg with a standard deviation of 0.242 mg
- Can we meet our product claim with 95% confidence?

1 Sample t-test – Example 2

Open Worksheet "Hypothesis Testing"

Stat> Basic Stats> 1-Sample t…

Enter "Dosage wt", Hypothesised Mean = 1.0

One-Sample T: Dosage wt

Test of mu = 1 vs not = 1

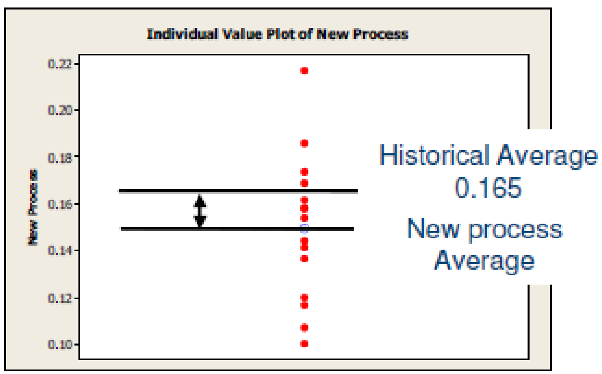
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Dosage wt	50	0.9821	0.2421	0.0342	(0.9132, 1.0509)	- 0.52	0.603

As $p > 0.05$ we cannot reject H_0 therefore the mean dosage weight is 1.0 mg. Also the Confidence Interval includes 1, confirming our decision.

Vestas

1 Sample t-test - Example 3

- A project team has implemented a new process, and wants to compare the results against the historical average of 0.165 mm.
  - Open Worksheet – 1-Sample t Average.MTW



Hypothesis:

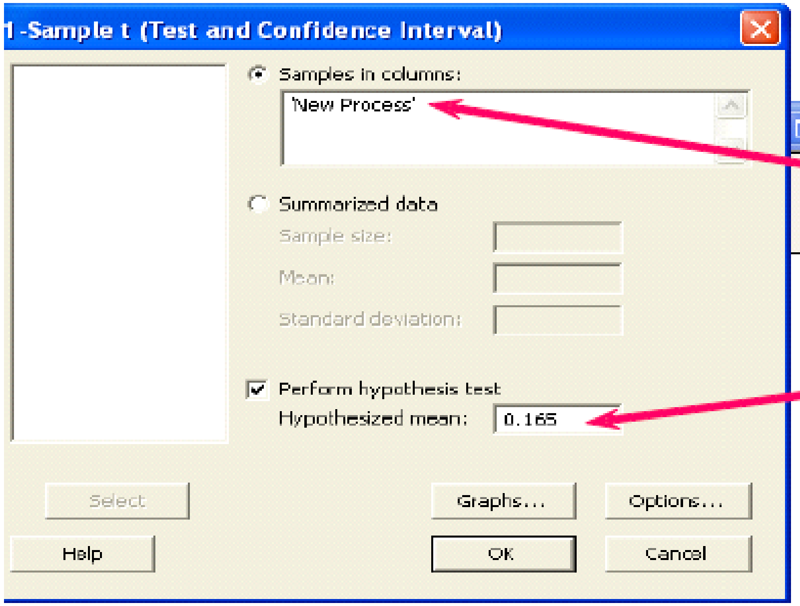
The Individual Value Plot suggests that the average dimension the new process is lower (quicker) than the historical average of 0.165

1 Sample t-test - Example 3

- A 1 Sample t-test can be used to test if the difference between the averages is statistically significant
- The hypotheses for the test would be:
  - Ho: There is no difference between the average invoice time of the new process and 0.165
  - Ha: New process is lower than average invoice time of 0.165
- Because our theory is that there is a difference, we expect to reject the Null hypothesis

1 Sample t-test - Example 3

Minitab: Stat > Basic Statistics > 1-Sample t



Enter the column that contains the data here.

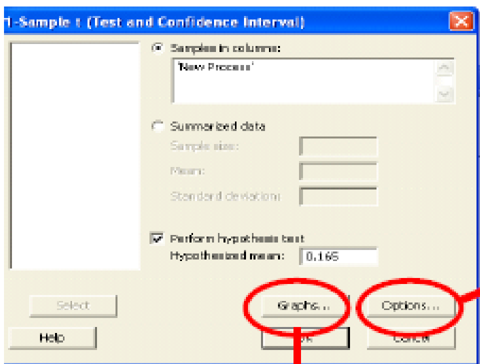
Enter the Historical mean here

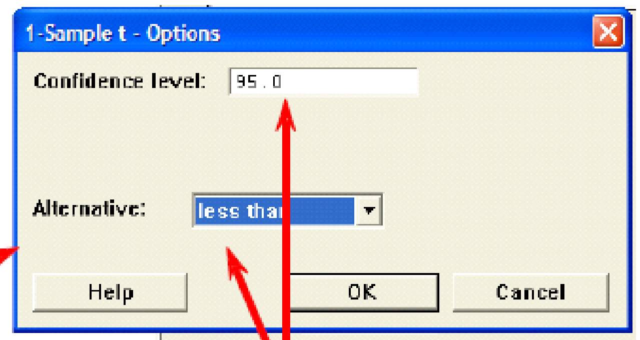
In both cases, the "known average" that the data is to be compared against should be entered here.

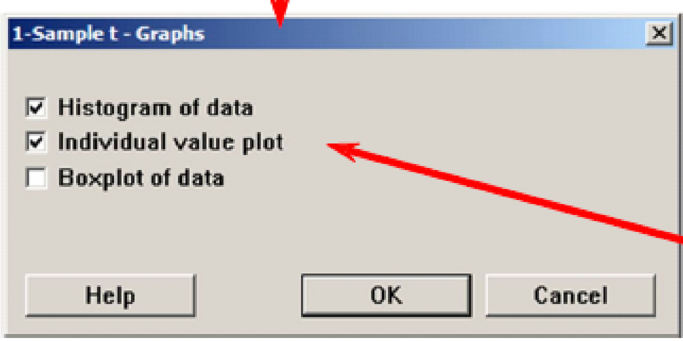
Vestas

1 Sample t-tests – Example 3

Minitab: Stat > Basic Statistics > 1 Sample t







Leave the confidence level of 95%

Alternative: less than

Select the graph most appropriate for the samples sizes involved. If in doubt, check all three.

1 Sample t-tests – Example 3

One-Sample T: New Process

Session Window Output for 1 Sample t-test

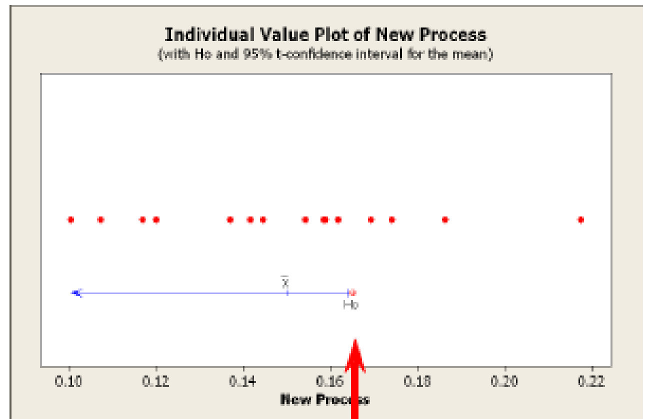
Test of mu = 0.165 vs < 0.165

Variable	N	Mean	StDev	SE Mean	95% Upper Bound	T	P
New Process	15	0.14982	0.03122	0.00806	0.16402	-1.88	0.040

The p-value for the 1 sample t-test is found in the session window output. In this case, the p-value is 0.040

Since the p-value for this test is less than 0.05, we can be 95% confident that there is a difference between the average invoicing time of the new process versus the historical average. We reject the Null Hypothesis

1 Sample t-tests – Example 3



Minitab adds a blue line to the graphs that represents the confidence interval of the sample average, and also indicates the position of H_0 (the known average).

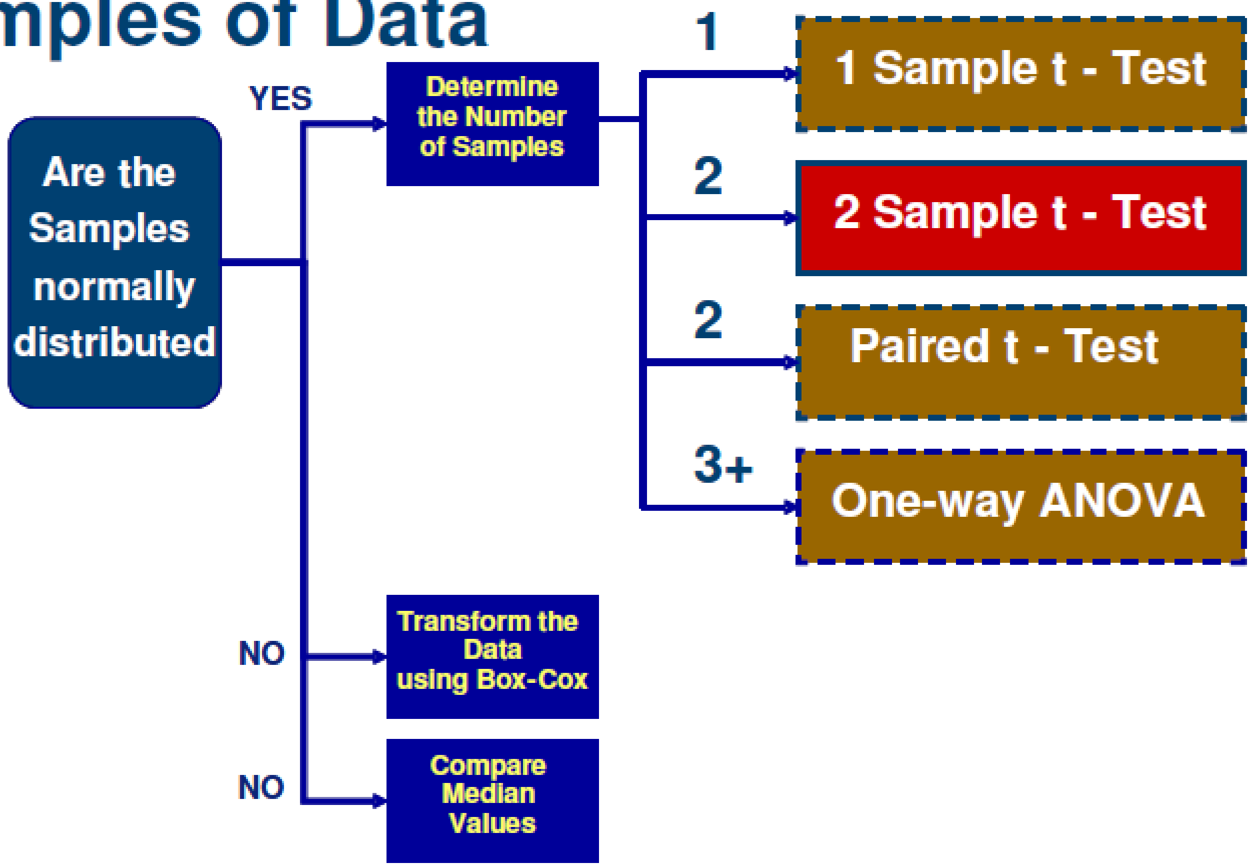
In this case, H_0 (the known average) is outside the confidence interval of the sample average and therefore we can prove a difference between the two.

1 Sample t-test – Whiskey or Water?

- Problem : A cocktail bar suspects that one of its suppliers of whiskey is adding water to increase profits. The freezing temperature of whiskey has a normal distribution with a mean of $\mu = -0.5450$ C. Adding water raises the freezing temperature. The cocktail bar wants to know if its suspicions are correct. Is the supplier adding water to it's whiskey?
- Procedure: A sample from each of ten bottles is obtained from this supplier.
- Experimental Unit : A bottle.
- Measurement : Freezing temperature of the sample.
- Significance Level : $\alpha = 0.05$
- Data Set : Whiskey.MPJ

Vestas

To Compare the Averages of Samples of Data



2 Sample t-test

- The 2 Sample t-test is an extremely useful tool in a 6-Sigma project
- We use a 2 Sample t-test to examine differences between two groups of data
- The 2 Sample t-test will tell you if an observed difference is statistically significant – real
- We use 2 Sample t-tests to help find vital Xs during the analyse phase
- We also use 2 Sample t-tests to prove we have made a real difference

2 Sample t-test Example

- We have collected data from a process that is operated on two lines and we suspect that there is a difference

  ie the lines are an X
- For Line A

the sample mean	=	14.96
standard deviation	=	1.93

- For Line B

the sample mean	=	14.30
standard deviation	=	1.48

Note that the sample sizes do not have to be the same!
Hence we conduct a 2-sample t-test

- Set up the test $H_0 : \mu_A = \mu_B$

  $H_a : \mu_A \neq \mu_B$

Vestas

We Look at the Difference Distribution

- If there was no difference we would expect the difference distribution to be normal and centred at 0
- AveLineA – AveLineB = 14.96 – 14.30 = 0.66
- How far is 0.66 from 0? What is the p-value?
- To calculate the standard deviation of the difference distribution we remember that only variances are additive

$$S_{A\text{-}B} = \sqrt{S^2_A + S^2_B}$$

- We are of course dealing with sampling distributions and therefore

$$S_{A\text{-}B} = \sqrt{\frac{S^2_A}{n_A} + \frac{S^2_B}{n_B}}$$

NB: This equation is used if the sample variances are not equal, if they are assumed to be equal the pooled standard deviation is used

t-test Comparing Two Group Averages

- We judge the difference between two group averages using the t-test

$$t_{A\text{-}B} = \frac{(\overline{X}_A - \overline{X}_B) - 0}{S_{A\text{-}B}}$$

- Formula for t is the same as for Z, but the p-values come from a t – distribution

- In practice we use Minitab

Interpreting the Results...

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE M
1	100	14.96	1.93	0.19
2	100	14.30	1.48	0.15

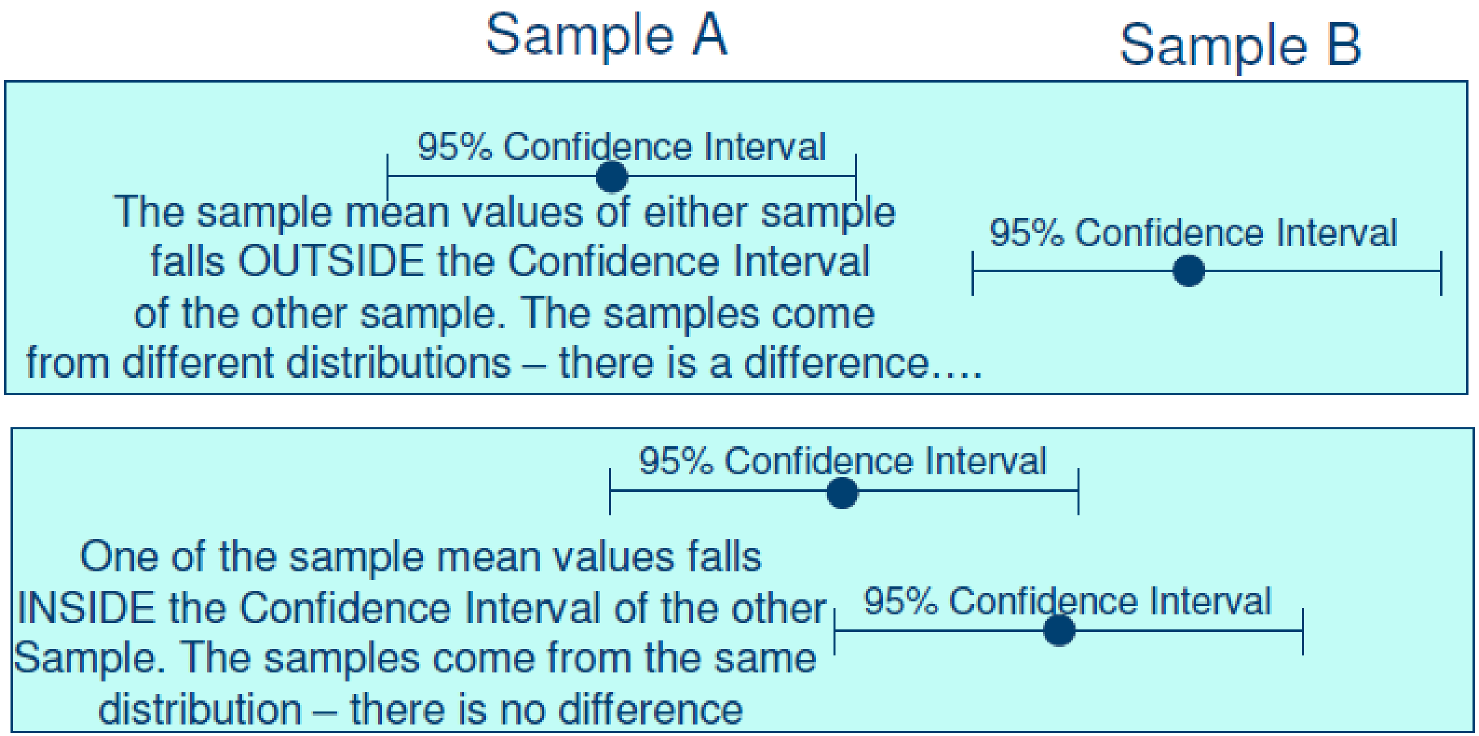
$P = 0.007 < 0.05$

P is low H_0 must go

Difference = mu (1) - mu (2) Estimate for difference: 0.660
95% CI for difference: (0.180, 1.140)
T-Test of difference = 0 (vs not =): T-Value = 2.71 P-Value = 0.007
DF = 185

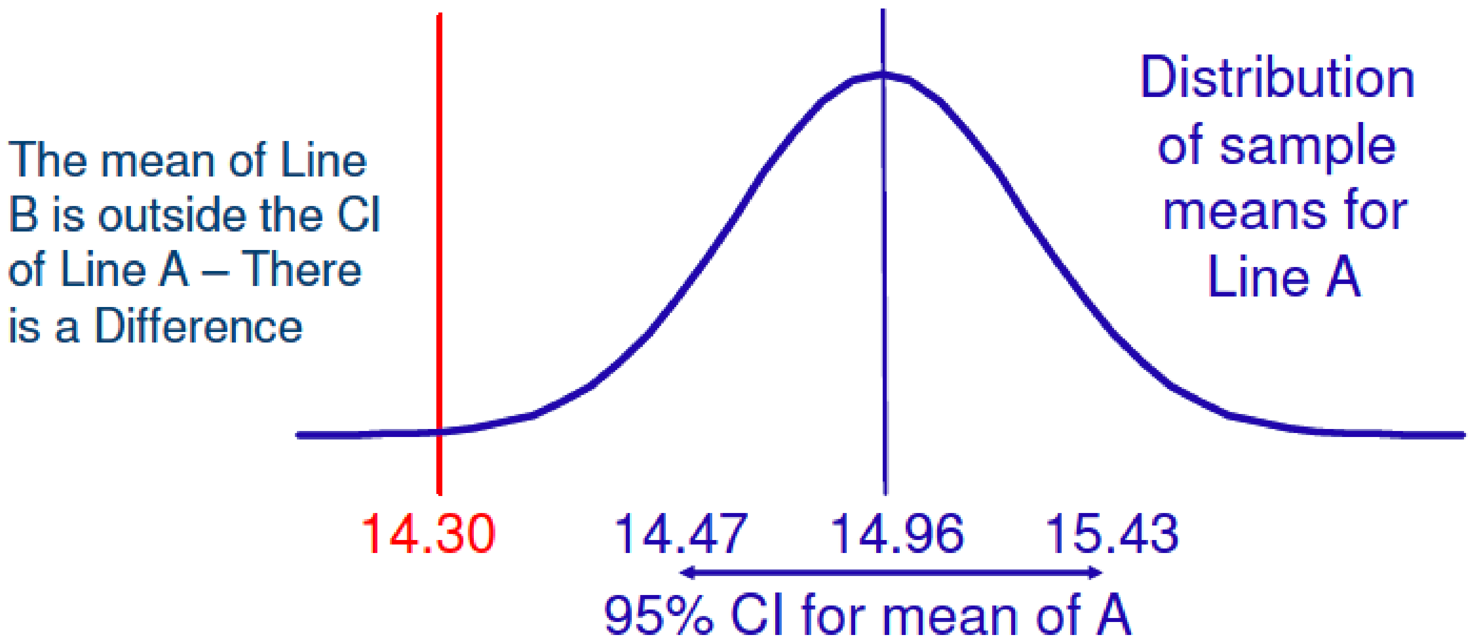
- Test looks at the difference between the means
- Null hypothesis is rejected, hence the Lines are different
- We can also use the fact that the 95% Confidence Interval (see sampling theory section) does not include zero

Interpreting Confidence Intervals



Another approach is to subtract the mean of sample 2 from the confidence interval of sample 1. If this answer contains 0 (ie the range goes from -ve to +ve) then there is no difference

Confidence Intervals



Confidence interval: 95%CI for mean of A - Mean of B
(14.47 -14.30) , (15.43 - 14.30) = 0.17, 1.13

If there is no significant difference between the means, the confidence interval will contain 0 - i.e. ranges from -ve to +ve.

Assumptions t-tests

- If data are continuous we assume the underlying data is normal
  - We may need to transform the data
- When comparing populations
  - Independent random unbiased samples
- When comparing groups from different processes we assume
  - Each process is stable (not many old processes may not be – but we would not necessarily stabilise a very poor process)
  - Samples are independent random unbiased

Pass Fail Data

- 2 Sample t-tests can useful when we have pass/fail data and a number of X variables to search
- Example
  - A product had a high failure rate. The product comprised two key sub-assemblies – the bulb and the thimble masses

↓	C1	C2	C3	C4	C5
	Bulb	Thimble	Summed	Pass/Fail	
115	29.97	53.83	83.80	0	
116	31.04	54.57	85.61	1	
117	30.65	54.15	84.80	0	
118	33.60	54.51	88.11	0	
119	32.81	54.49	87.30	0	
120	34.05	54.55	88.50	1	
121	34.81	54.60	89.41	0	
122	34.65	54.20	88.85	1	
123	28.97	54.52	83.49	1	

Two-Sample T-Test and CI: Bulb, Pass/Fail

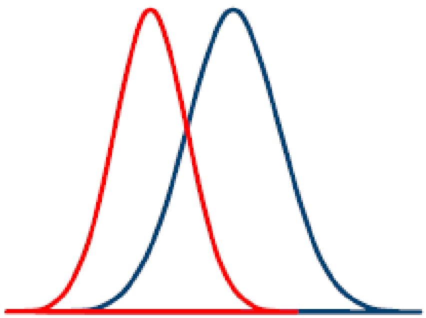
Two-sample T for Bulb

Pass/Fai	N	Mean	StDev	SE Mean
0	105	33.36	2.68	0.26
1	125	34.29	2.76	0.25

Difference = mu (0) - mu (1)
Estimate for difference: -0.926
95% CI for difference: (-1.637, -0.216)
T-Test of difference = 0 (vs not =): T-Value = -2.57 P-Value = 0.011
Both use Pooled StDev = 2.72

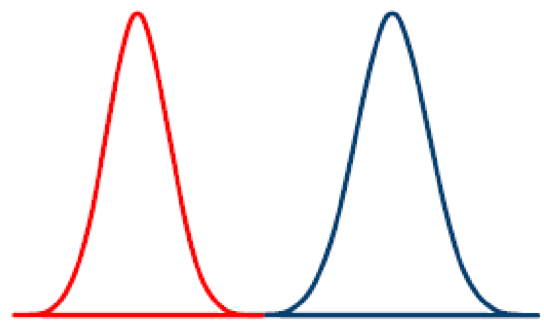
P-value<0.05
Hence Reject H_0
Bulb mass affects pass/fail

The Concept of Two Sample t-tests



You take two samples of data from the same process, but at different times. Has the process mean moved?

Would you think the process mean has moved if the results had been…..



What did you look at to make your decisions?

2 Sample t-tests (1)

Breaking strength of plastic

Problem : You are evaluating the breaking strength of plastic from two potential suppliers for use in your product; mobile phones. Supplier B is the current supplier and has a good working relationship with the company. However, representatives from supplier A claim that their plastic is superior to that of supplier B. You are trying to determine if you should switch suppliers. From your company's perspective, the breaking strength would have to be more than 10 psi higher for a practical difference in strength to exist.

Data Set : Plastic.MPJ

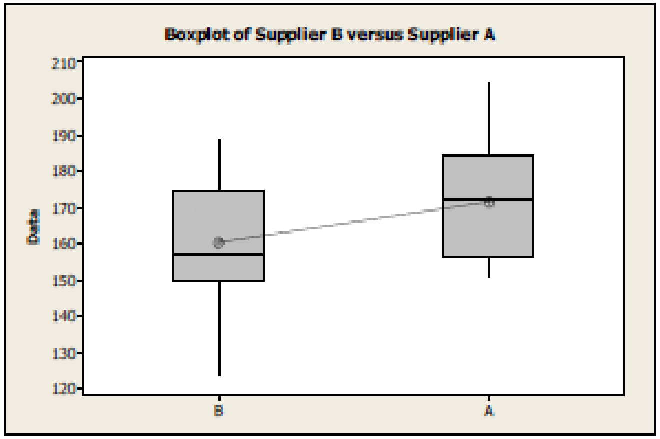
2 Sample t-tests (2)

Procedure:	A random sample of plastic is taken from each supplier and the breaking strength is measured.
Factor levels :	Supplier (A and B)
Experimental Unit :	Plastic samples of the same thickness as the mobile phones.
Measurement :	Breaking strength (psi)
Significance Level :	$\alpha = 0.05$

2 Sample t-tests (3)

Hypothesis:

The Box Plot suggests that the average Breaking strength for Supplier A is greater than for Supplier B. But can we be sure?

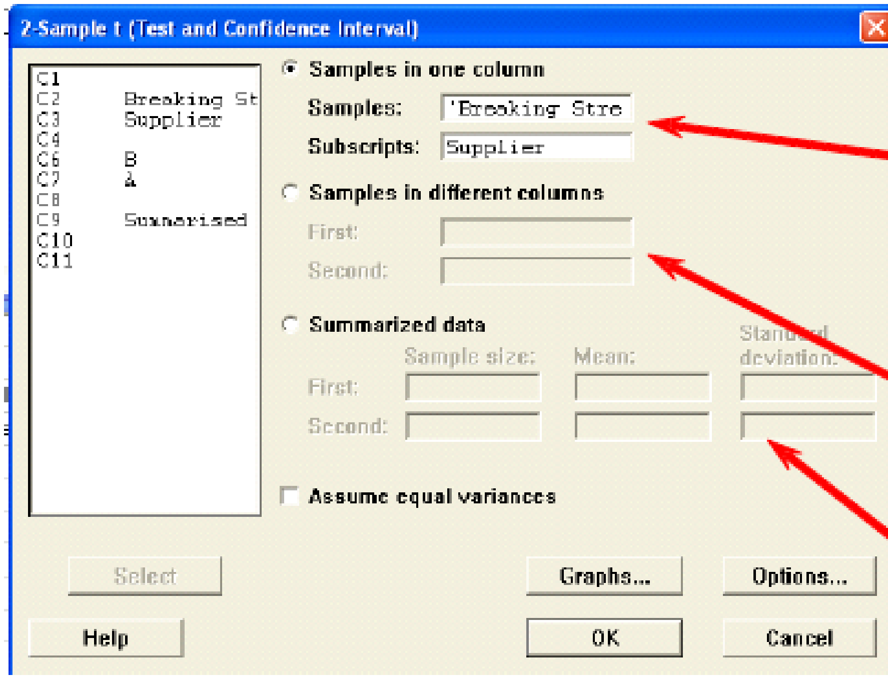


2 Sample t-tests (3)

- A 2 Sample t-test can be used to test if the difference between the average Breaking Strength is statistically significant
- The hypotheses for the test would be:
  - H_o: There is no difference between the average breaking strength for suppliers A and B processes
  - H_a: The breaking strength for supplier A is greater than the breaking strength for supplier B
- Because our theory is that the A is greater, we expect to reject the Null hypothesis

2 Sample t-tests (4)

Minitab: Stat > Basic Statistics > 2 Sample t



There are several options for entering data.

1st Option: If the data is in a single column, with a second column containing subgroup data.

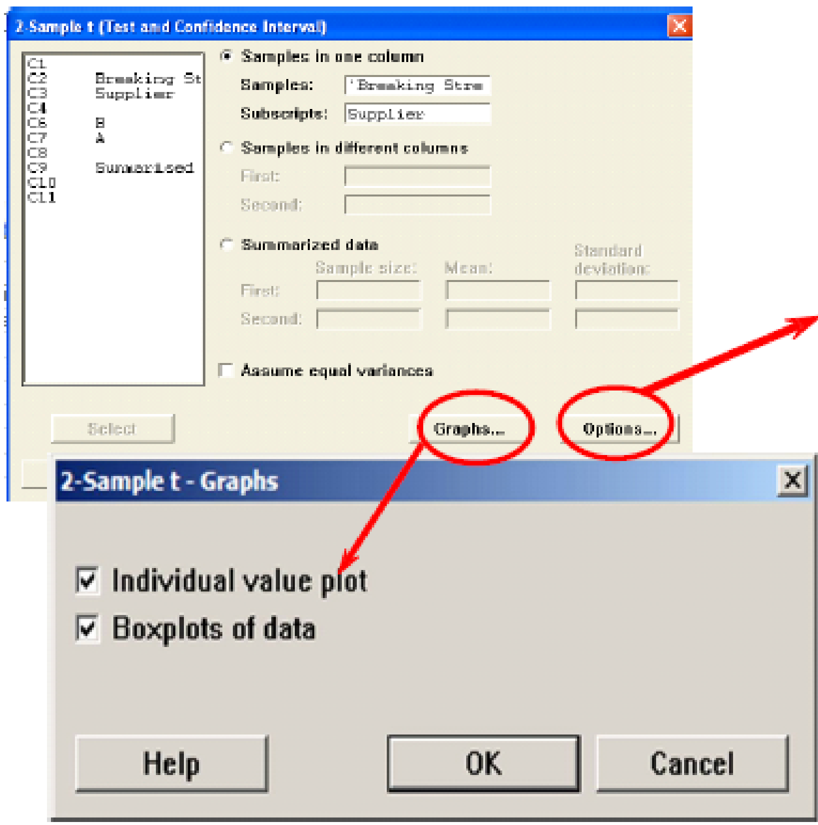
2nd Option: If the two data samples are in two separate columns.

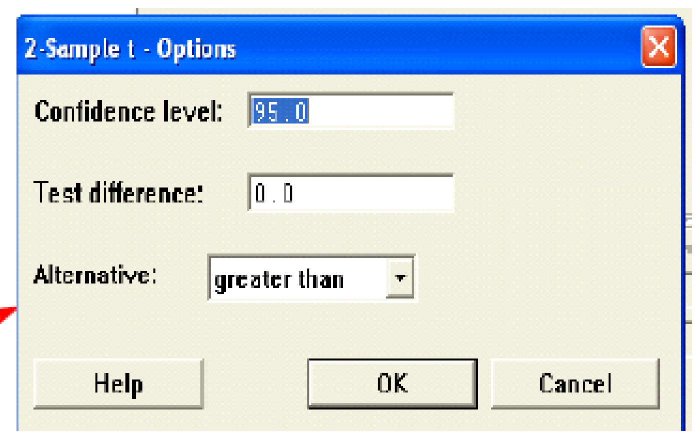
3rd Option: If you know the sample size, mean and standard deviation of both samples, this information can be entered here.

Data Set : Plastic.MPJ

2 Sample t-tests (5)

Minitab: Stat > Basic Statistics > 2 Sample t





Set the options as shown above:

- A confidence of 95%
- A test difference of 0.0
- Alternative: greater than

Vestas

2 Sample t-tests (6)

Two-Sample T-Test and CI: Breaking Strength, Supplier

Session Window Output for 2 Sample t-test

Two-sample T for Breaking Strength

Supplier	N	Mean	StDev	SE Mean
A	25	171.8	15.6	3.1
B	20	160.6	17.3	3.9

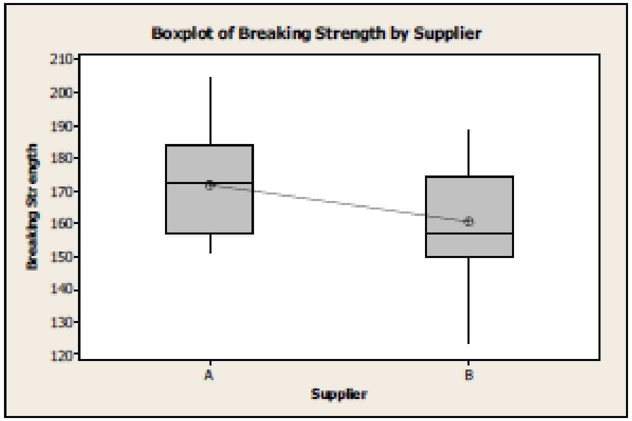
Difference = mu (A) - mu (B)
Estimate for difference: 11.1850
95% lower bound for difference: 2.8041
T-Test of difference = 0 (vs >): T-Value = 2.25 P-Value = 0.015 DF = 38

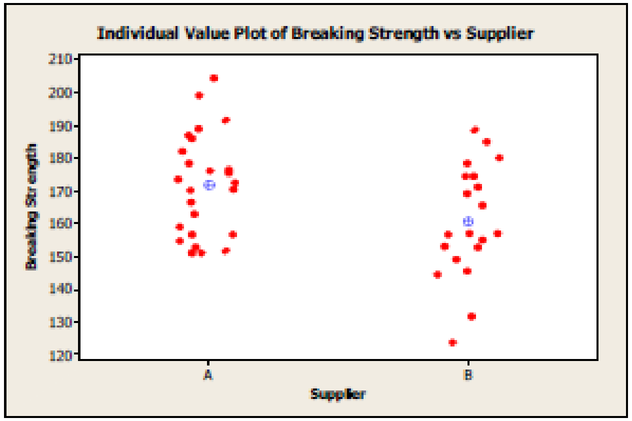
The p-value for the 2 Sample t-test is found in the session window output. In this case, the p-value is 0.015.

Since the p-value for this test is less than 0.05, we can be over 95% confident that there is a difference between the average breaking strengths.

Vestas

2 Sample t-tests (7)





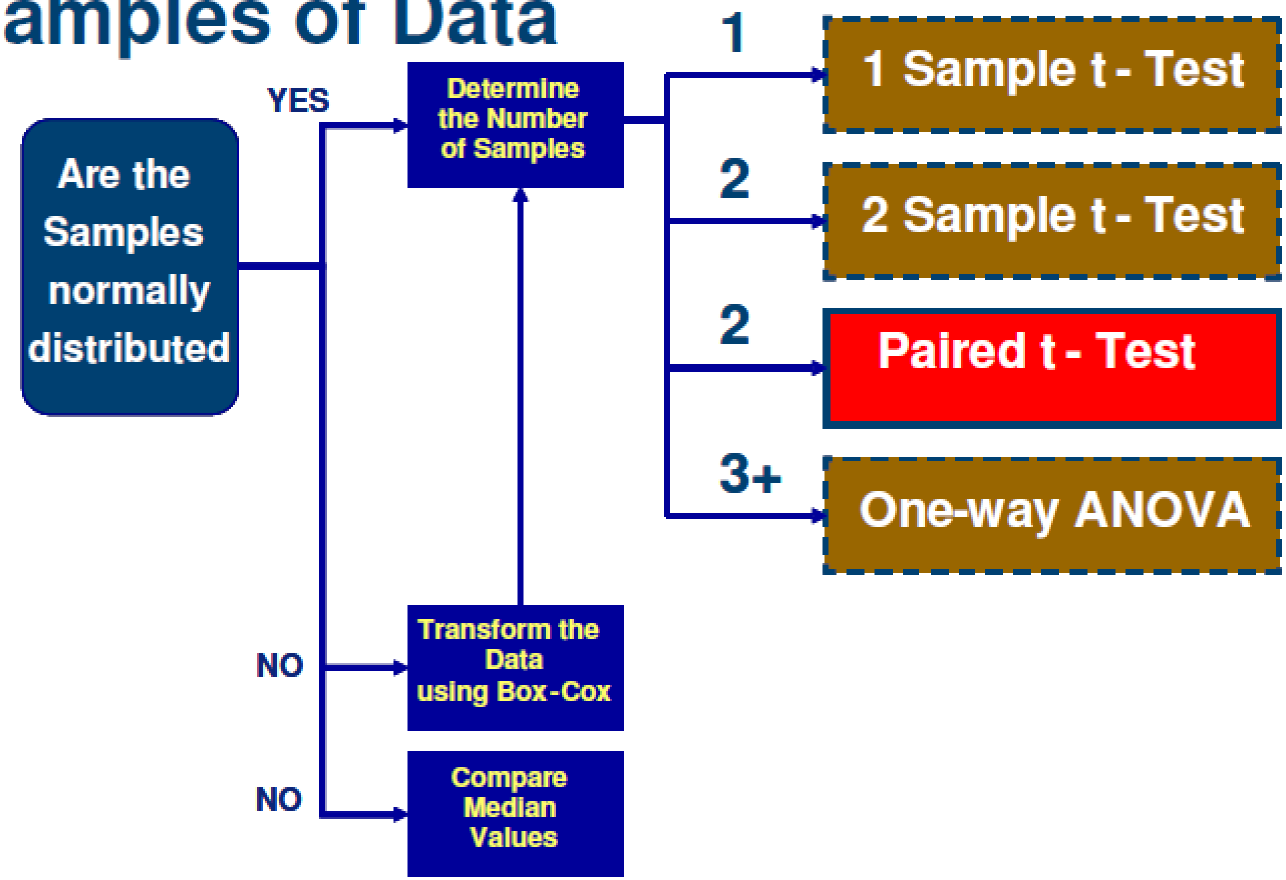
The graphs provided by the 2 Sample t-test are the same as those provided by using the graph menu.

By default, Minitab adds a symbol to show the average of each subgroup.

2 Sample t-tests (8)

- *Exercise:* Using the data file: “Plastic Strength”
- 2) Repeat the test using the *“samples in different columns”* option
- 3) Repeat the test using the *“summarised data”* option

To Compare the Averages of Samples of Data



Paired t-tests (1)

- Paired t-tests are very similar to 2 Sample t-tests, but they are used on samples of data that are linked in pairs.
- The two samples may represent:
  - Two different suppliers
  - Two different processes
  - Two different teams
  - Two different products etc
- The samples may be linked in pairs that were processed under similar conditions, such as:
  - By the same person
  - From the same product
  - By the same supplier etc

Paired t-test

- A variation on the standard t-test when the data are matched

Sampling Unit	Measure 1	Measure 2
1	___	___
2	___	___
3	___	___
.	.	.
.	.	.
.	.	.
n	___	___

Sampling Unit	Right wrist	Left wrist
Bill	17.2	16.9
Fred	19.7	19.2
Jane	15.9	16.8
.	.	.
.	.	.
.	.	.
n		

- Each *sampling unit* has two measurements
- The measurements in the second group are not independent of the first group
  - They are matched or paired
  - Both measurements are taken on the same sampling unit

Paired t-test - Example

Name	Right	Left	Difference
James	17.5	16.8	0.7
John	17.6	16.7	0.9
Jonathan	16.2	16.0	0.2
Mike	19.0	18.9	0.1
Rob	19.7	19.3	0.4
Steve	18.1	18.0	0.1
Bill	20.5	20.0	0.5
Sola	17.9	17.6	0.3
Mean	18.313	17.913	0.4
Std Dev	1.364	1.40	0.288

- A 2 Sample t-test calculates the mean of each group first, then takes the difference
- A paired t-test takes the paired difference first, then takes the mean of the differences

Paired t-test - Example

- Both methods will have the same mean difference

  Difference of means = 0.4

  Mean of differences = 0.4

- The variation, however, is much smaller in a paired t-test – it eliminates the differences between people

  Standard deviation for left and right wrists ≈ 1.4

  Standard deviation for paired difference = 0.288

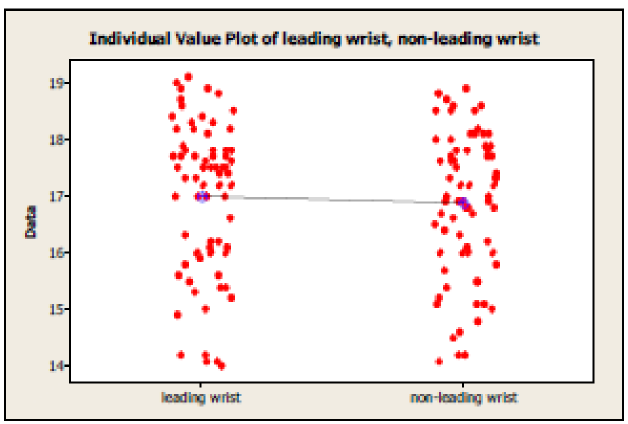
- The smaller standard deviation allows the paired t-test to detect smaller differences
- 2 Sample t-tests and paired t-tests can give quite different results

Paired t-tests (2)

Minitab file BB Wrists.mpj

- The Individual Value Plot shows two samples of wrists of "data capture" process

- One sample was taken of the Leading wrist and one of the Non-Leading wrist



There appears to be a slight reduction in average between the two wrists but only slight in comparison to the process variation.

Vestas

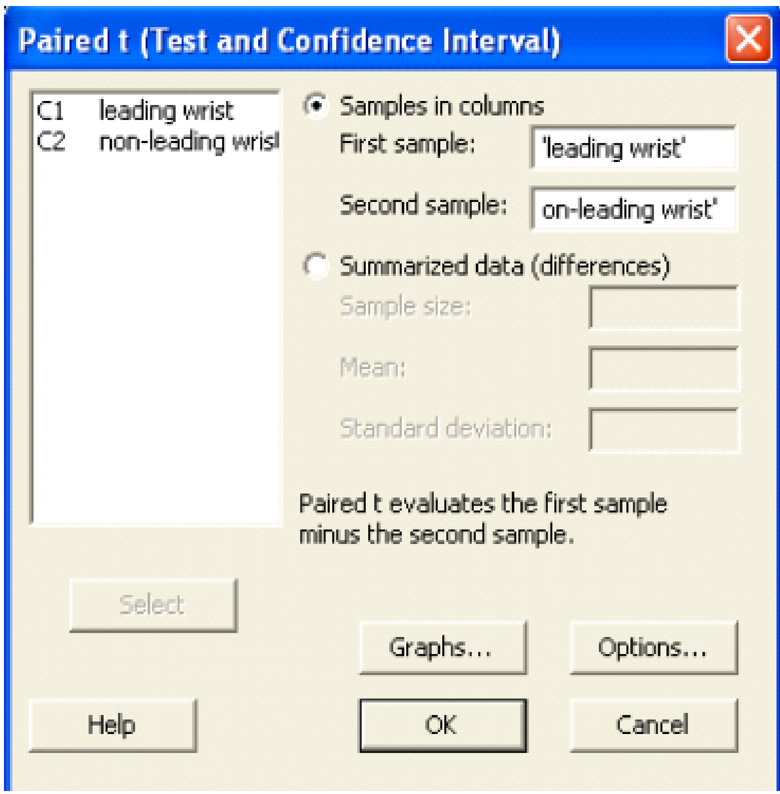
Paired t-tests (3)

- A paired t-test can be used to test if the differences between the pairs of data are statistically significant
- The hypotheses for the test would be:
  - H_o: There is no difference in the wrist size for Leading and Non-leading wrists
  - H_a: The Leading wrist is greater than the non-leading wrist
- Because our theory is that there is a difference, we expect to reject the Null hypothesis

Vestas®

Paired t-tests (4)

Minitab: Stat > Basic Statistics > Paired -t



There are two options for entering data.

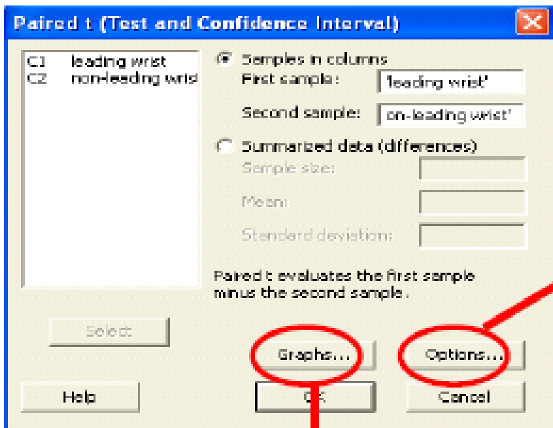
1st Option: Enter the two columns containing the two samples of data here.

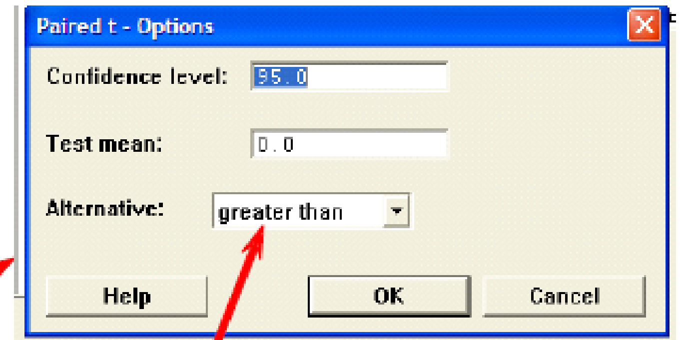
2nd Option: Enter the *summarised* data of the *differences* between the pairs of data here.

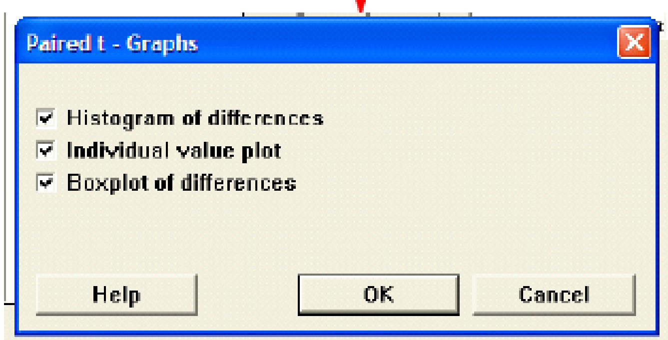
Vestas

Paired t-tests (5)

Minitab: Stat > Basic Statistics > 2 -Sample t







Set the Alternative to greater than

Select the graph most appropriate for the samples sizes involved. Note that all the graphs plot the 'differences'.

Paired t-tests (6)

Paired T-Test and CI: leading wrist, non-leading wrist

Paired T for leading wrist - non-leading wrist

	N	Mean	StDev	SE Mean
leading wrist	72	16.9944	1.3260	0.1563
non-leading wris	72	16.8792	1.2739	0.1501
Difference	72	0.115278	0.362176	0.042683

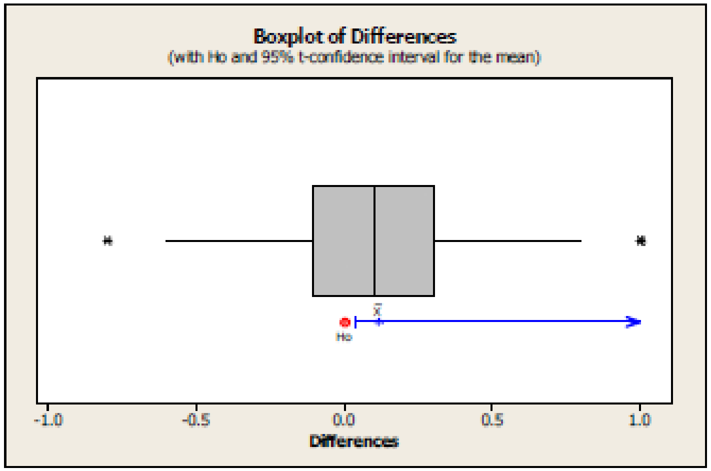
95% lower bound for mean difference: 0.044142
T-Test of mean difference = 0 (vs > 0): T-Value = 2.70 P-Value = 0.004

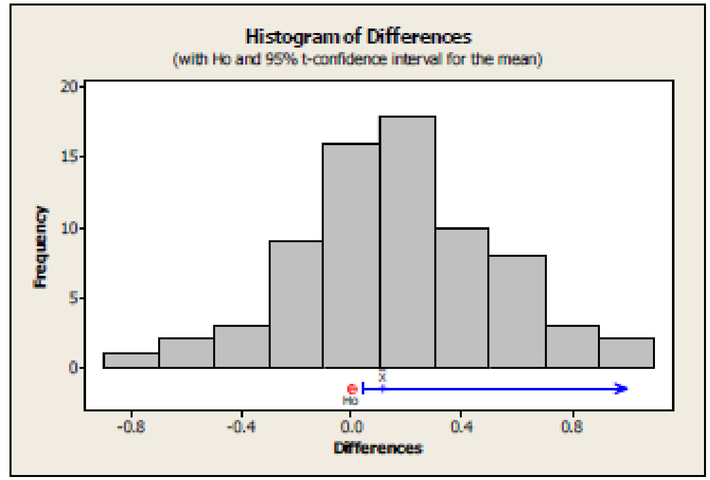
Session Window Output
for Paired t-test

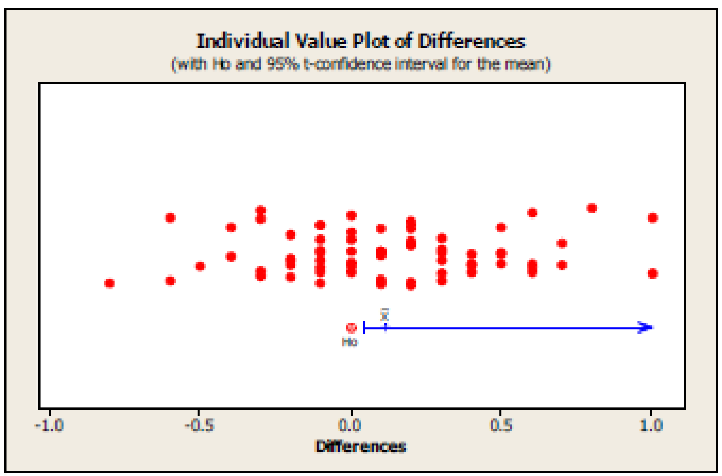
The average difference between the pairs of data is 0.1153

Since the p-value for this test is less than 0.05, we can be over 95% confident that the average difference of 0.1153 is statistically significant.

Paired t-tests (7)







The graphs provided by the paired t-test all plot the *differences* in the pairs of data.

The Null hypothesis (Ho) is shown outside the confidence intervals on each graph to show a difference

Paired t-tests (8)

Exercise:

- Complete a “2 Sample t-test” on the wrists data and explain your findings

Paired t-tests (9)

Example: Paired t-Test – Car Parking

Problem : A consumer group wishes to determine whether or not there is a difference in handling ability between two popular cars currently on the market. To measure the handling ability of the two cars, the time it takes drivers to parallel park each of the cars is recorded.

Procedure: Each driver will park both cars in random order. Thus two times will be recorded for each driver.

Factor levels : Car A and Car B

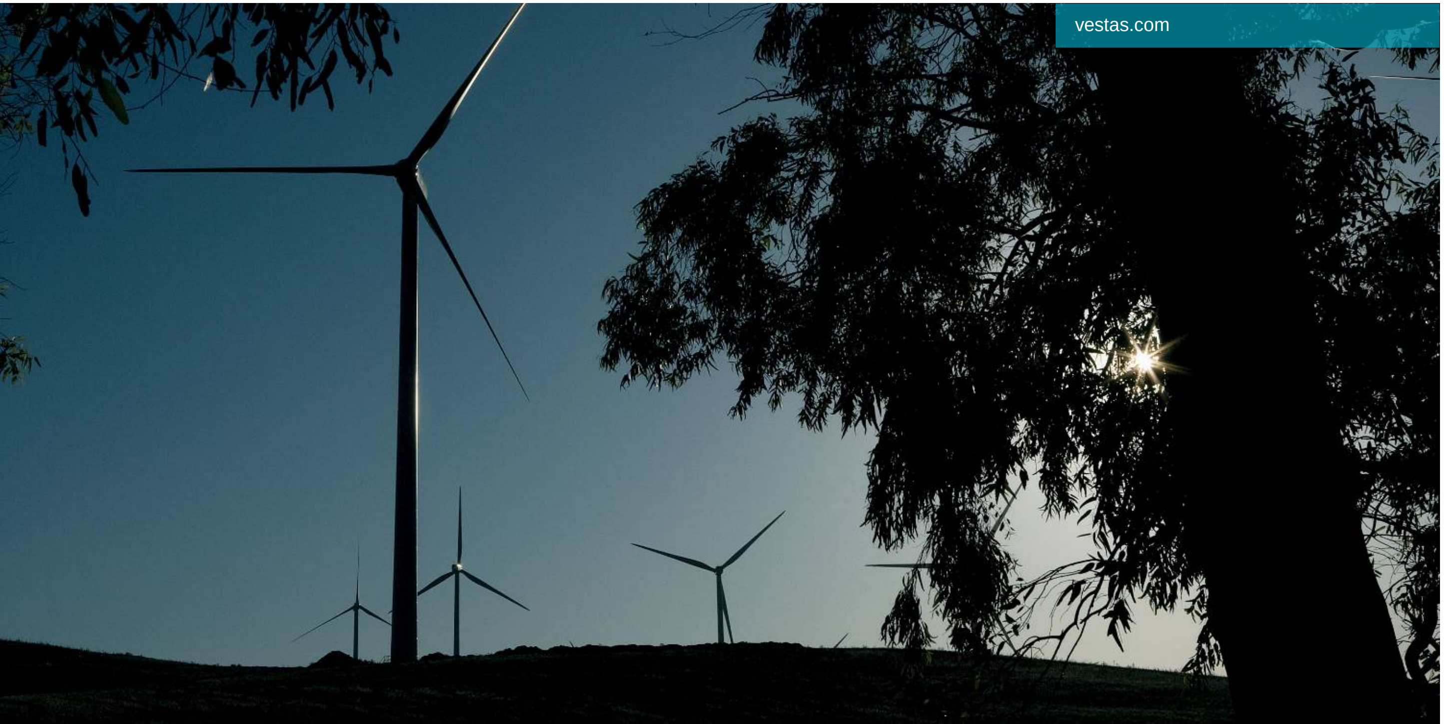
Experimental Unit : A driver parking a particular car.

Measurement : The time (in seconds) that it takes to park the car.

Data Set : Carpark Paired t-test.MPJ

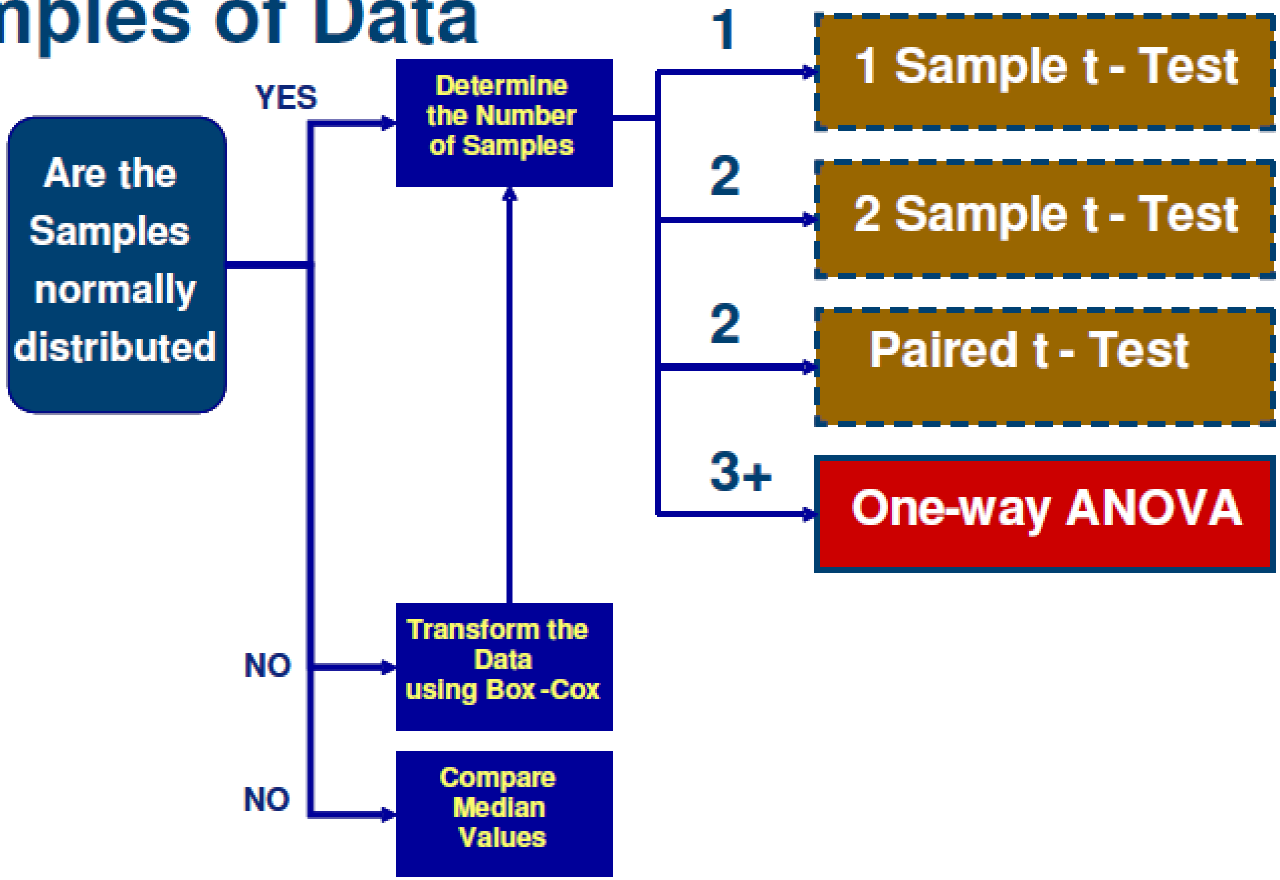
No. 1 in Modern Energy

ANOVA



Vestas

To Compare the Averages of Samples of Data



ANOVA Introduction

The ANOVA technique is used as a hypothesis test for difference in average, where there 3 or more separate samples.

Area of particular interest - are all group means are equal.

If all means are the same in the underlying model, then the groups are identical.

If the model means are all equal, it would be expected that the sample means would be similar.

However they are unlikely to be identical.

Therefore assess whether the variation between the group means is unusually great.

To do this, take account of the variation within the groups.

Exercise – One Way ANOVA

- In small teams of 3
- Determine whether there is a difference in the rebound distance for a range of coin sizes (eg 5p, 10p, 50p) when the coins are thrown against a wall for one thrower

One Way ANOVA (1)

Exercise: Open data file: “Failure Rates.MPJ”

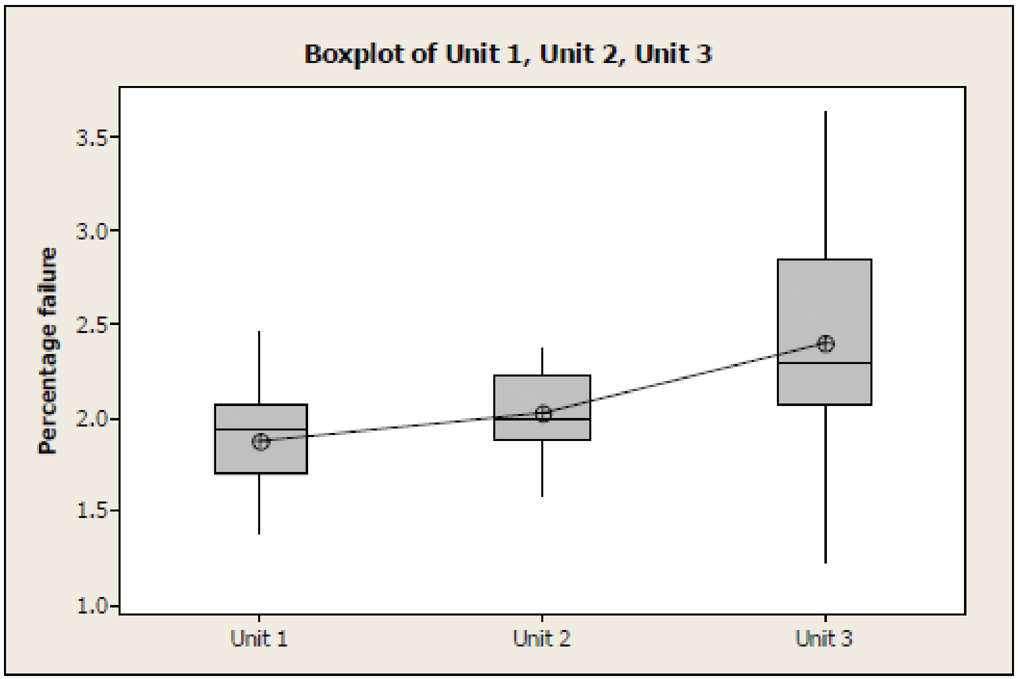
- The data file has the same data in two different worksheets (in two different formats)
- 1) Using worksheet *“Failure Rates Stacked”*, follow the step by step instructions shown on the previous pages
- 2) Repeat the test using the worksheet *“Failure Rates for ANOVA-unstacked”* and using the following Minitab function:

  Minitab: Stat > ANOVA > One Way (unstacked)

Vestas

One Way ANOVA (2)

A project team is investigating the average percentage failure rates between manufacturing units.



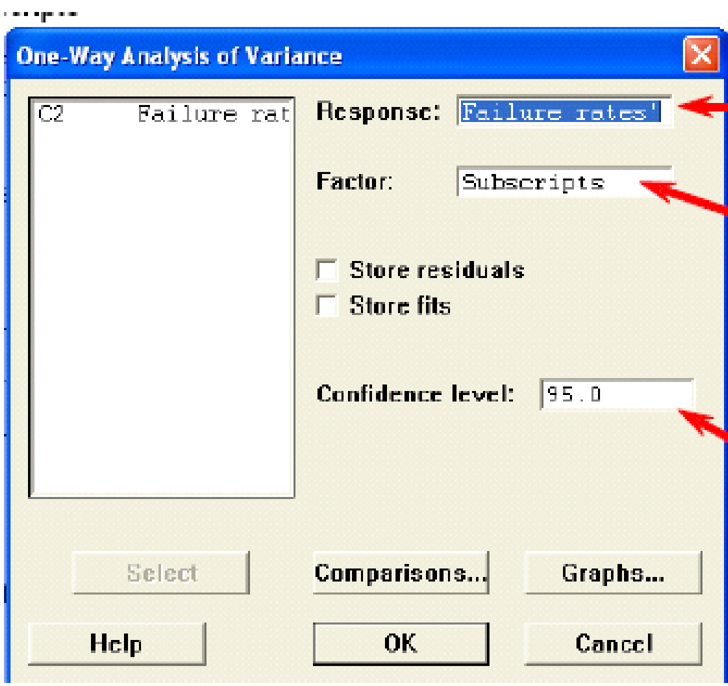
Theory: The Box Plot suggests that the average failure rate is different across the units.

One Way ANOVA (3)

- A One Way ANOVA can be used to test if the difference between the average failure rates is statistically significant
- The hypotheses for the test would be:
  - H_o: There is no difference between the average failure rates between the units
  - H_a: There is a difference between the average failure rates between the units
- Because our theory is that there is a difference, we expect to reject the Null hypothesis

One Way ANOVA (4)

Minitab: Stat > ANOVA > One Way



Enter the column containing the data here.

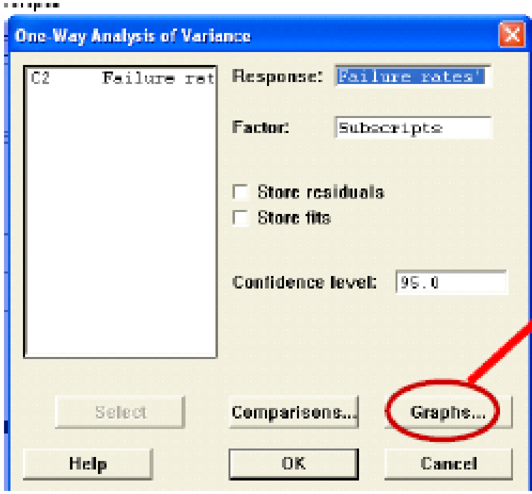
Enter the column containing the details of the samples here (subscripts).

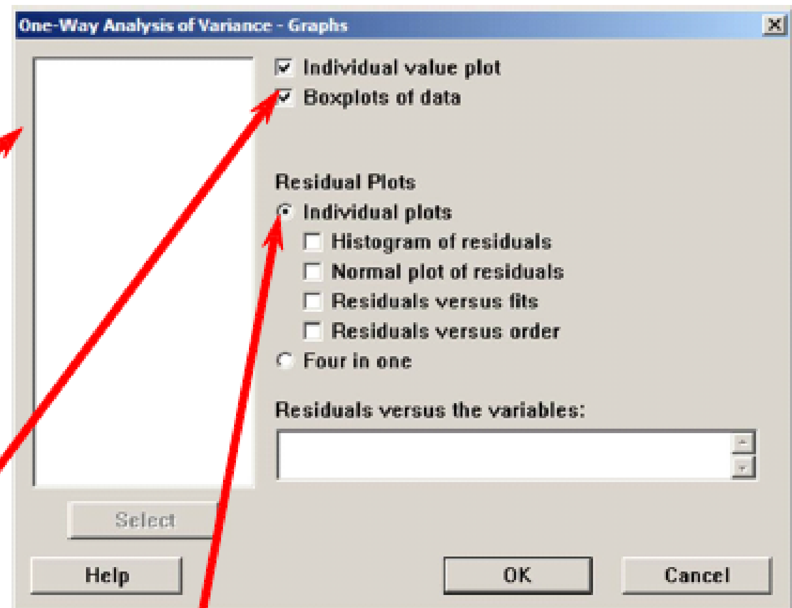
Leave the Confidence level at the default of 95%

If the data samples are in separate columns, then use MINITAB: Stat> ANOVA > One Way (unstacked)

One Way ANOVA (5)

Minitab: Stat > ANOVA > One Way





Select the graph most appropriate for the samples sizes involved. If in doubt, check both.

Do not request Residual Plots (leave as shown).

One Way ANOVA (6)

Session Window Output for One Way ANOVA

One-way ANOVA: Failure rates versus Subscripts

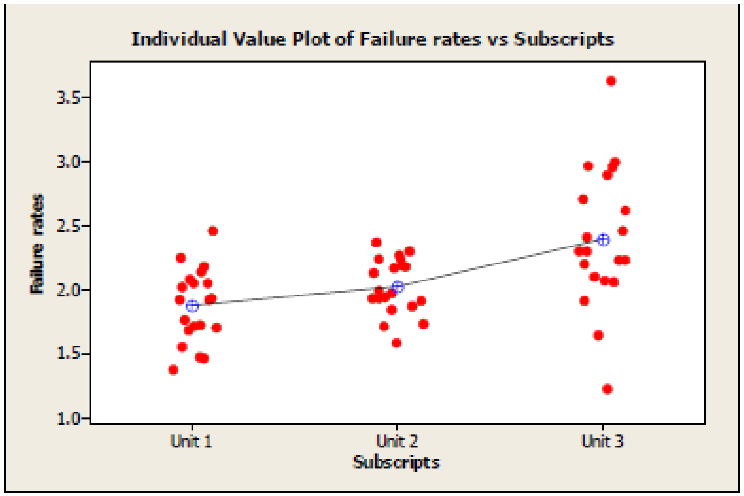
Source	DF	SS	MS	F	P
Subscripts	2	2.858	1.429	10.25	0.000
Error	57	7.948	0.139		
Total	59	10.806			

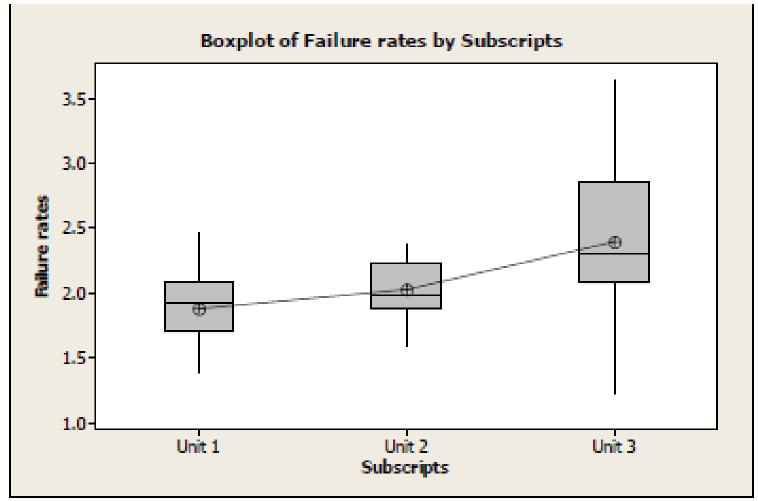
S = 0.3734   R-Sq = 26.45%   R-Sq(adj) = 23.87%

The p-value for the one way ANOVA is found in the session window output. In this case, the p-value is 0.000.

Since the p-value for this test is less than 0.05, we can be over 95% confident that there is a difference between the average transaction values.

One Way ANOVA (7)



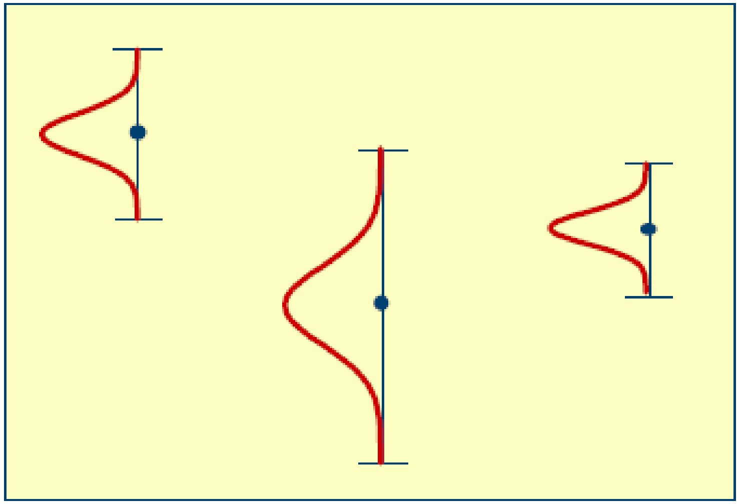


By default, Minitab adds a symbol to the graphs in order to show the average of each subgroup.

The p-value (previous slide) indicates that the averages of the subgroups are not all the same. However, that does not mean that they are all different. This requires further investigation.

Vestas

Comparing Group Variances



Is the variation different?

- Comparing variation is important for two reasons: Check out ANOVA assumption that all group variation is equal
- Demonstrating there is a difference between two groups of data
- When attempting to find root causes
- When wishing to prove a change has made the desired difference

Test for Equal Variances

- A Test for equal Variance is used to compare the variances of two or more distributions

- Two situations:
  - Bartlett's Test - Normal data – replaced by an F-test if only two levels
  - Levene's Test - Non Normal data

- The Null and Alternative Hypotheses are:
  - H_o: $\sigma_1 = \sigma_2 = \ldots = \sigma_n$
  - H_1: at least one variance is different

Vestas®

Test for Equal Variances (2)

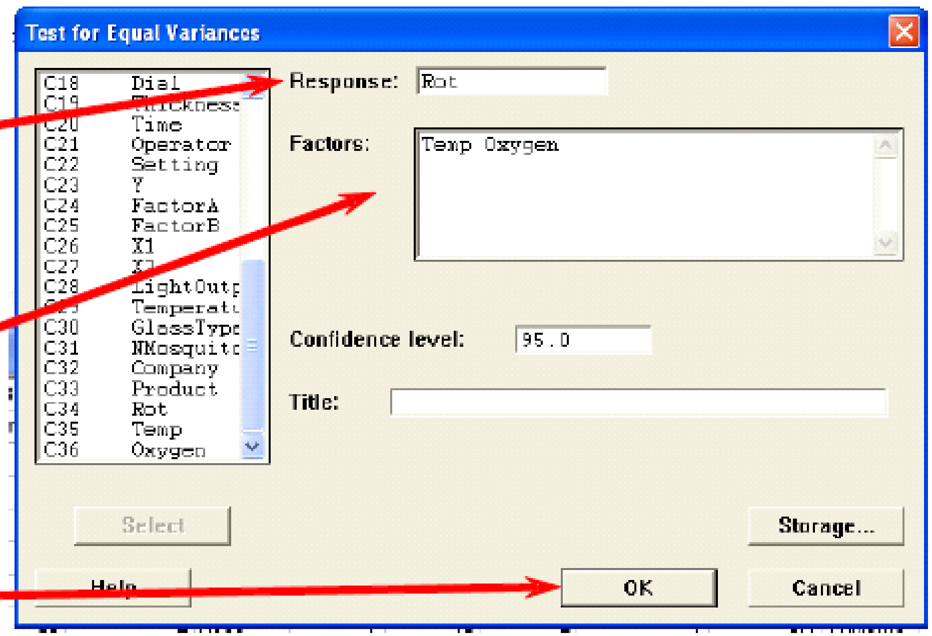
Open data file: Exh_aov.mtw

Minitab: Stat > ANOVA > Test for Equal Variance

In Response enter:
Rot

In factors enter:
Temp
Oxygen

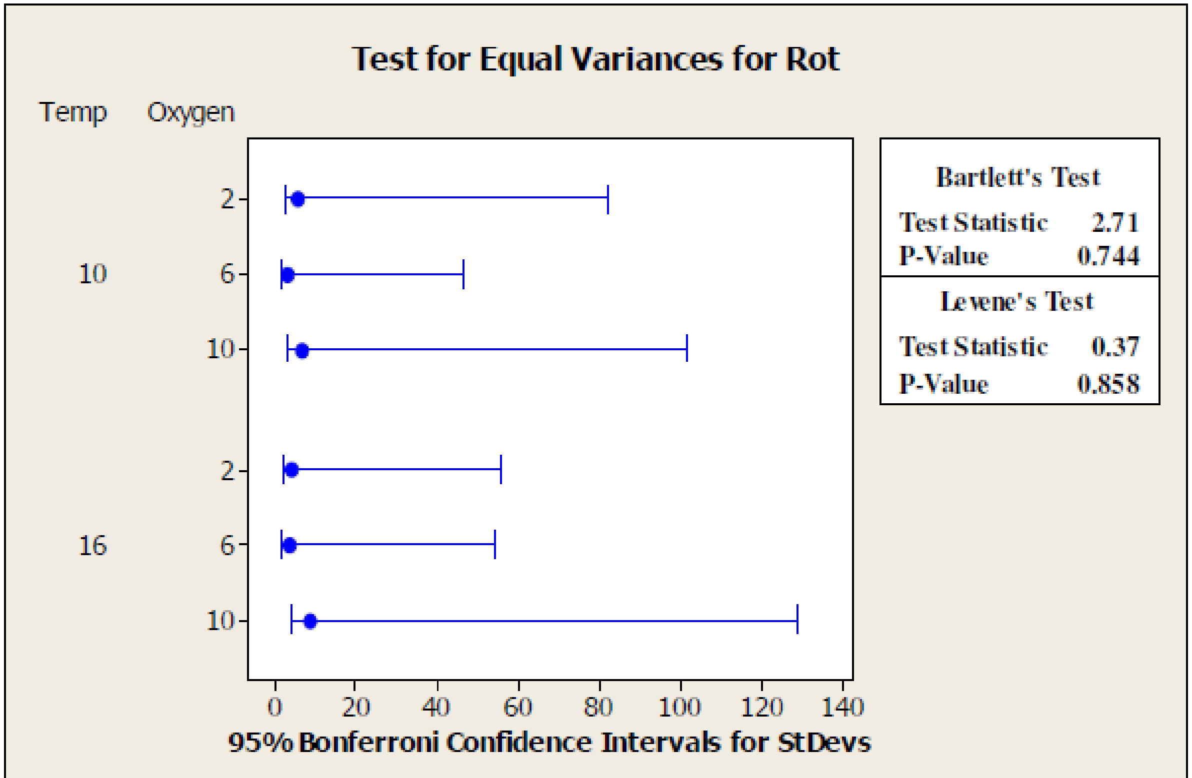
Click OK



Vestas®

Results

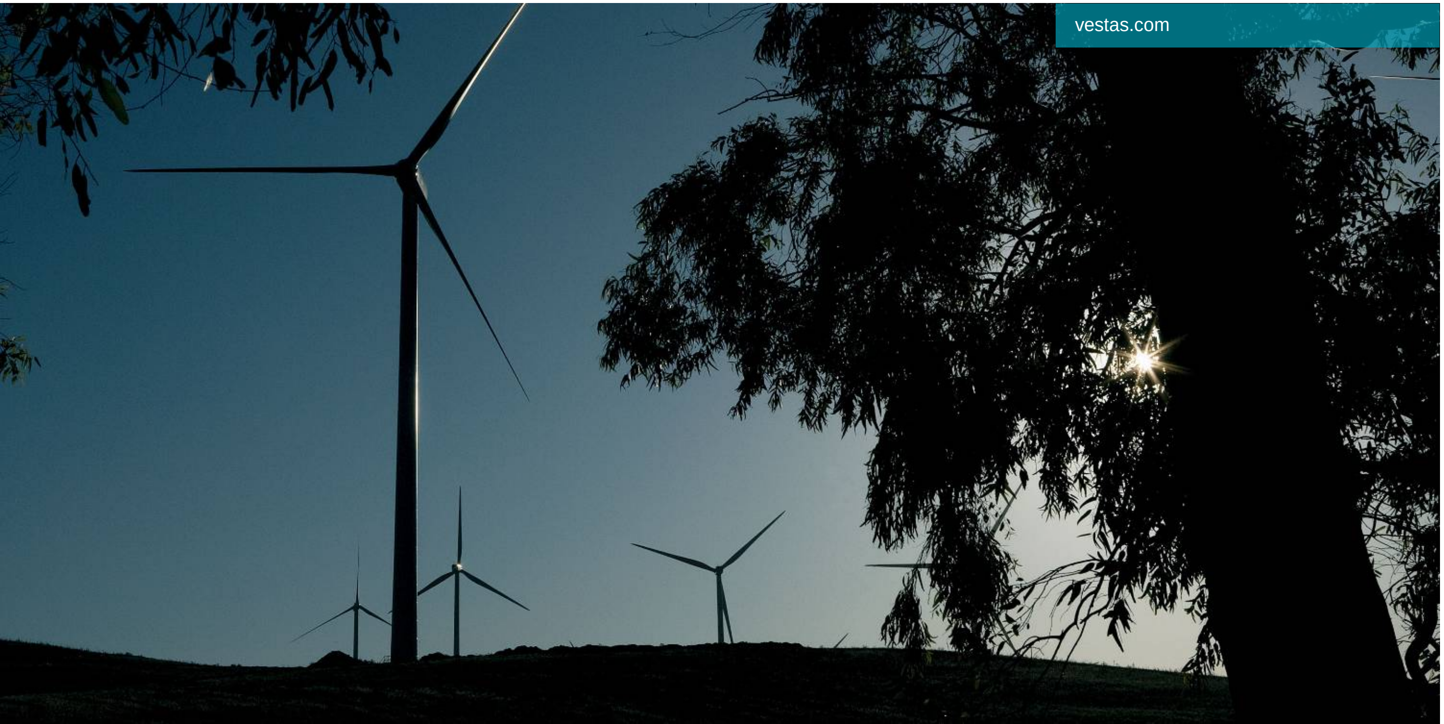
Are the variances equal? H_0 equal variances



P values greater than 0.05 so accept H_0 that the variances are equal

No. 1 in Modern Energy

Hypothesis Testing on Proportions



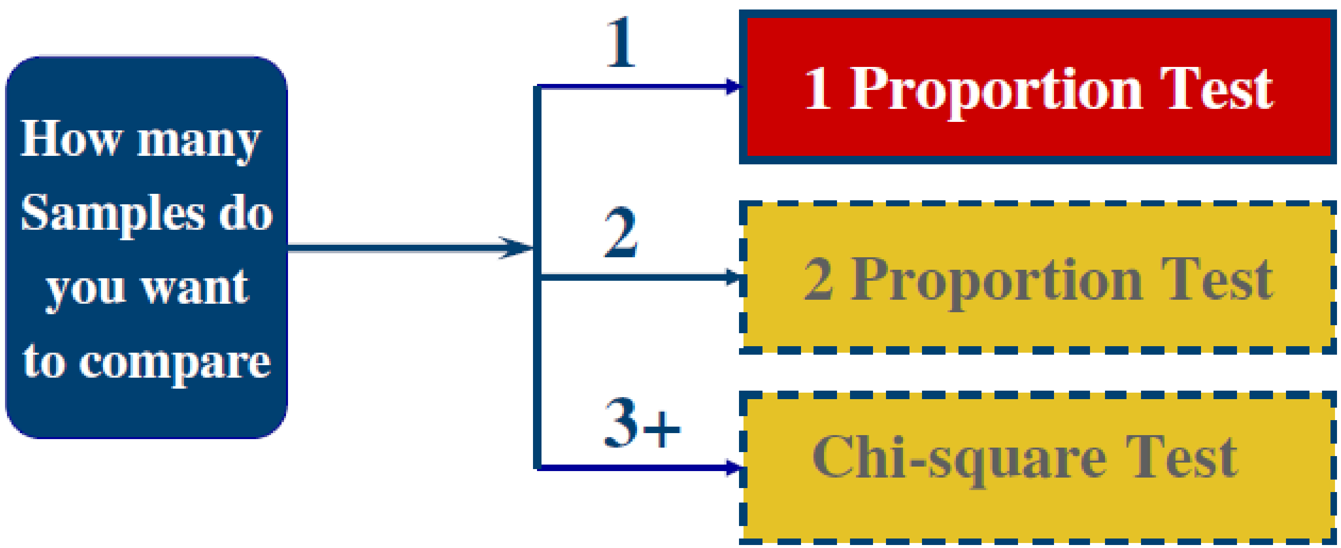
Vestas

Objectives of this Module

- To introduce a range of hypothesis tests suitable for testing for differences between proportions
- Specifically:
  - 1 Proportion test
  - 2 Proportion test
  - Chi-square test

To Compare the Proportions Data Samples

You want to compare proportions or percentages that came from different samples of data to decide if they ‘statistically’ different.



1 Proportion Test (1)

- 1 Proportion tests are used to compare a proportion (%) of one sample of data against a known proportion
- The "known proportion" may be a historical average or an industry benchmark
- For the purpose of the test, the "known proportion" is assumed to be "exact"

1 Proportion Test (2)

- A project team is looking at the proportion of warranty claims on supplied products
- Historically (over the last 5 years) the proportion of claims has been stable at 1.4%
- A sample of 1000 products over the last two months provided 22 claims (2.2%)

Theory: The sample suggests that the proportion of claims during the last two months (2.2%) is higher than the historical rate of 1.4%.

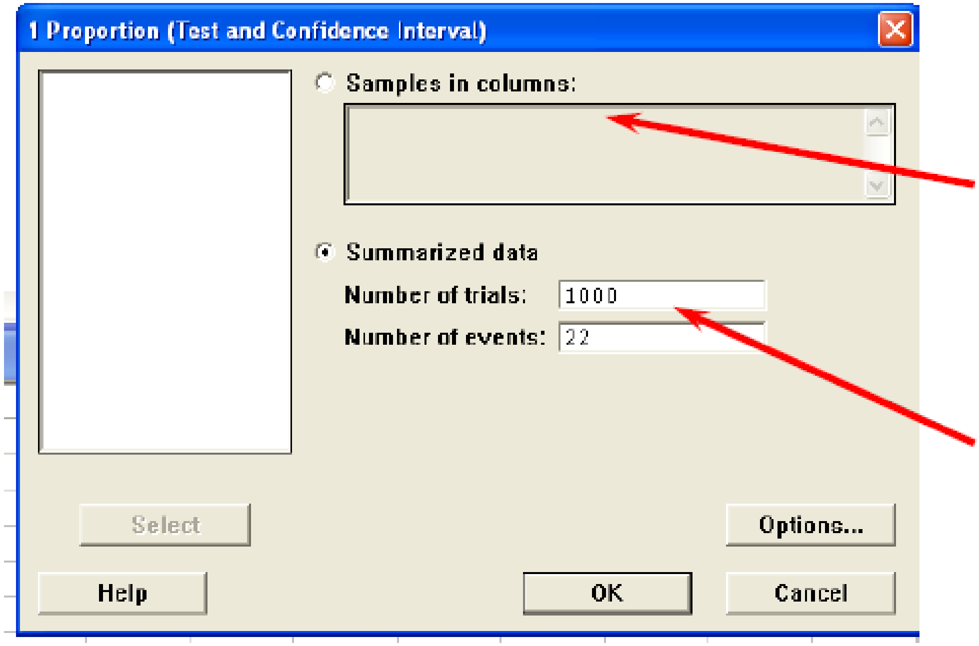
1 Proportion Test (3)

- A 1 Proportion test can be used to test if the difference between the sample and historical proportions is statistically significant
- The hypotheses for the test would be:
  - Ho: There is no difference between the sample proportion and the historical proportion of 1.4%
  - Ha: There is a difference between the sample proportion and historical proportion of 1.4%
- Because our theory is that there is a difference, we expect to reject the Null hypothesis

Vestas

1 Proportion Test (4)

Minitab: Stat > Basic Statistics>1 Proportion



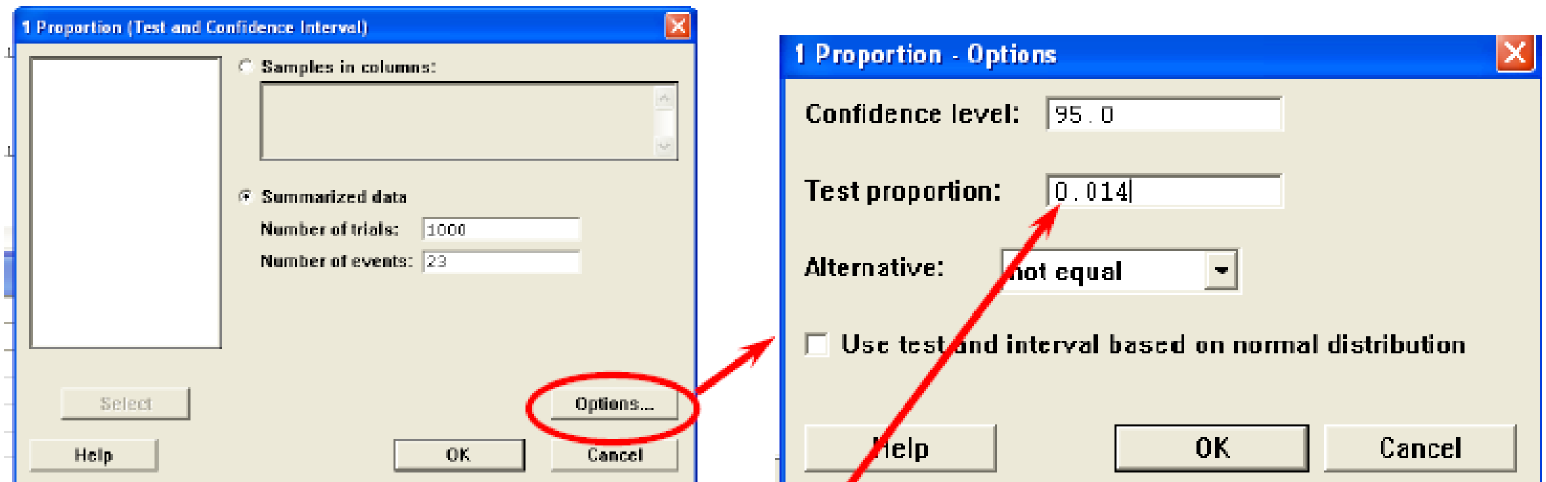
There are two options for entering data.

1st Option: Enter the column (if available) that contains the data here.

2nd Option: Enter the sample size (trials) and number of incorrect assessments (events) here.

Vestas

1 Proportion Test (5)



The “Test Proportion” refers to the known proportion that you are testing against. In this case it is 0.014 (1.4%).

The default is 0.5 (50%).

Leave these options as the defaults shown above:

- A confidence of 95%
- Alternative: not equal

1 Proportion Test (6)

Test and CI for One Proportion

Test of p = 0.014 vs p not = 0.014

Sample	X	N	Sample p	Exact 95% CI	P-Value
1	23	1000	0.023000	(0.014635, 0.034312)	0.030

Session Window Output for 1 Proportion test

The p-value for the 1 proportion test is found in the session window output. In this case, the p-value is 0.030.

Since the p-value for this test is less than 0.05, we can be over 95% confident that there *is* a difference between the sample proportion of 2.3% and the historical proportion of 1.4%.

1 Proportion Test – Television Repair Rate

Problem : The QA dept for a particular television (TV) manufacturer wishes to estimate the proportion of their 35 -inch TV sets that needed to be repaired within 4 years of purchase. The dept is particularly interested in whether the need for repairs for their brand of TV's is different than that for all brands currently on the market. It is known that the overall repair rate for such TV's is 6.8% (0.068).

Procedure: A survey is sent to consumers who purchased a 35-inch TV from this manufacturer during a particular time period. Consumers were asked whether or not the TV they purchased needed repair within 4 years of purchase.

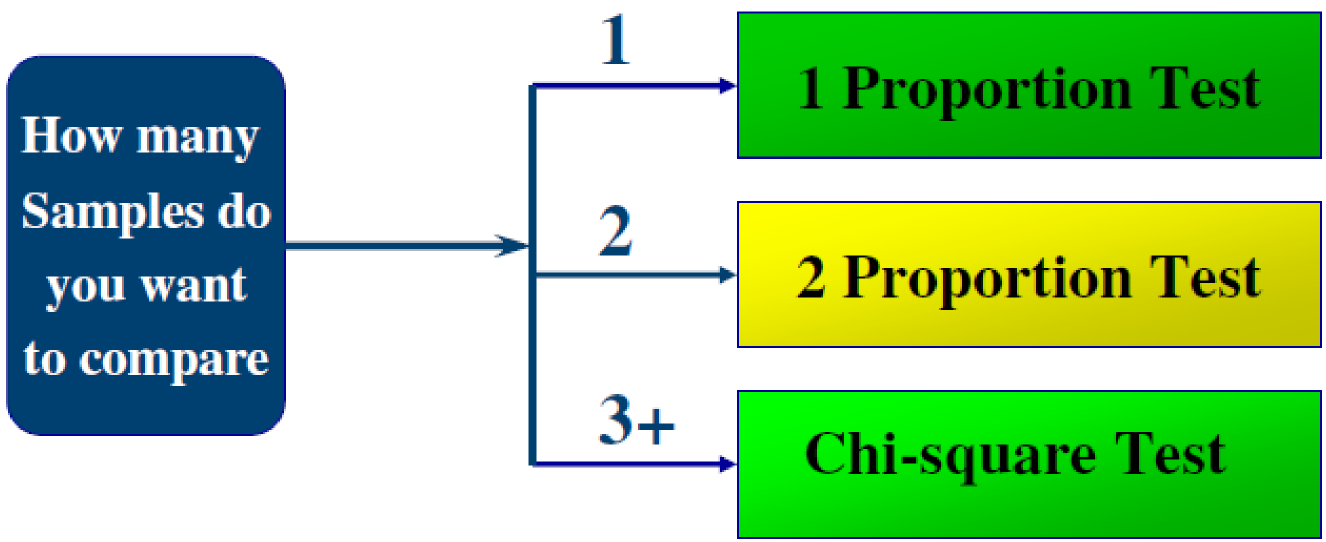
Results : There were 2,856 respondents. Of these, 236 (8.3%) answered "yes".

Experimental Unit : Each 35-inch TV from this manufacturer during a particular period.

Measurement : Whether or not the television needed repair within 4 years of its purchase (Yes or No).

To Compare the Proportions Data Samples

You want to compare proportions or percentages that came from different samples of data to decide if they ‘statistically’ different.



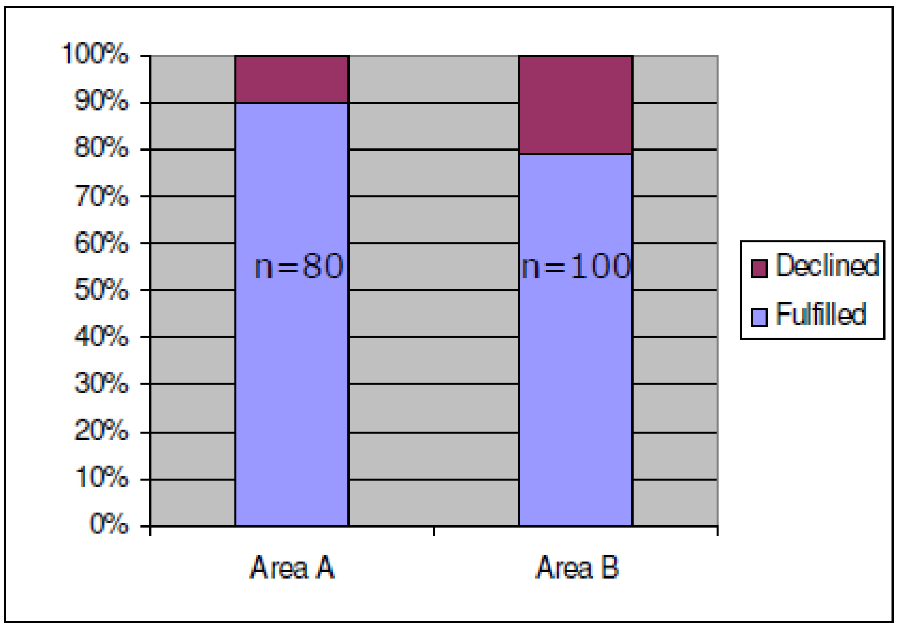
2 Proportion Tests (1)

- 2 Proportion tests are used to compare proportions from two samples of data. The two samples may represent:
  - two different suppliers
  - two different processes
  - two different teams
  - two different products etc
- The two samples can be of different sample sizes, since the 2 Proportion test takes account of both sample sizes.

Vestas®

2 Proportion Tests (2)

- A project team has noticed that the % of applications fulfilled (fulfil rate) is higher in Area A than Area B, and wants to confirm that this is statistically significant before investigating why.



Theory: It *appears* that Area A (90%) is higher than Area B (79%).

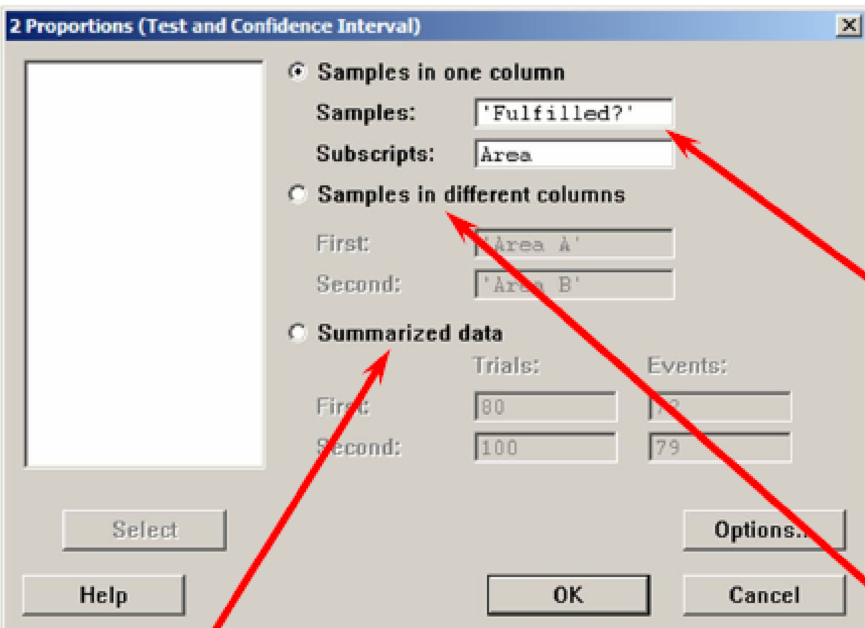
However, this was based on sample sizes of 80 and 100.

2 Proportion Tests (3)

- A 2 Proportion test can be used to test if the difference between the fulfil rates is statistically significant
- The hypotheses for the test would be:
  - Ho: There is no difference between the % fulfil rates of areas A & B
  - Ha: There is a difference between the % fulfil rates of areas A and B
- Because our theory is that there is a difference, we expect to reject the Null hypothesis

2 Proportion Tests (4)

Minitab: Stat > Basic Statistics>2 Proportion



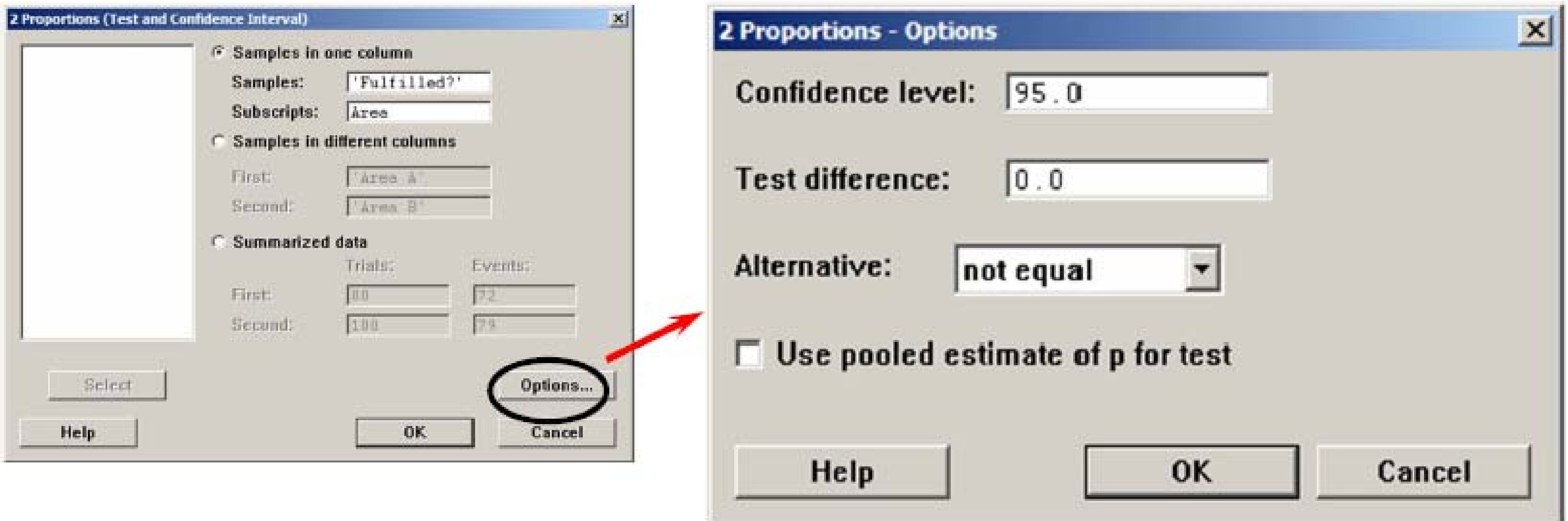
There are several options for entering data, depending on how your worksheet is arranged.

1st Option: If the data is in a single column, with a second column containing subgroup data.

2nd Option: If the two data samples are in two separate columns.

3rd Option: You can enter a summary of the data directly into this function. "Events" refers to either the number of "fulfilled applications" in this case.

2 Proportion Tests (5)



Minitab's

2 Proportion test does not offer any graphs.

Leave the options as the defaults shown above:

- A confidence of 95%
- A test difference of 0.0
- Alternative: not equal

Vestas

2 Proportion Tests (6)

Test and CI for Two Proportions: Fulfilled?, Area

```
Event = Fulfilled

Area      X    N  Sample p
Area A   72   80  0.900000
Area B   79  100  0.790000

Difference = p (Area A) - p (Area B)
Estimate for difference:  0.11
95% CI for difference:  (0.00658520, 0.213415)
Test for difference = 0 (vs not = 0):  Z = 2.08  P-Value = 0.037
```

Session Window Output for 2 Proportion test

The p-value for the 2 Proportion test is found in the session window output. In this case, the p-value is 0.037

Since the p-value for this test is less than 0.05, we can be over 95% confident that there is a difference between the fulfil rates of the two areas.

2 Proportion Tests (7)

Exercise: Open data file: "2 Proportion Test-Fulfil Rates.MPJ"

The data file has the same data in two different formats in two different worksheets.

1) Follow the step by step instructions shown on the previous pages.
2) Repeat the test using the "samples in different columns" option.
3) Repeat the test using the "summarised data" option.

Vestas

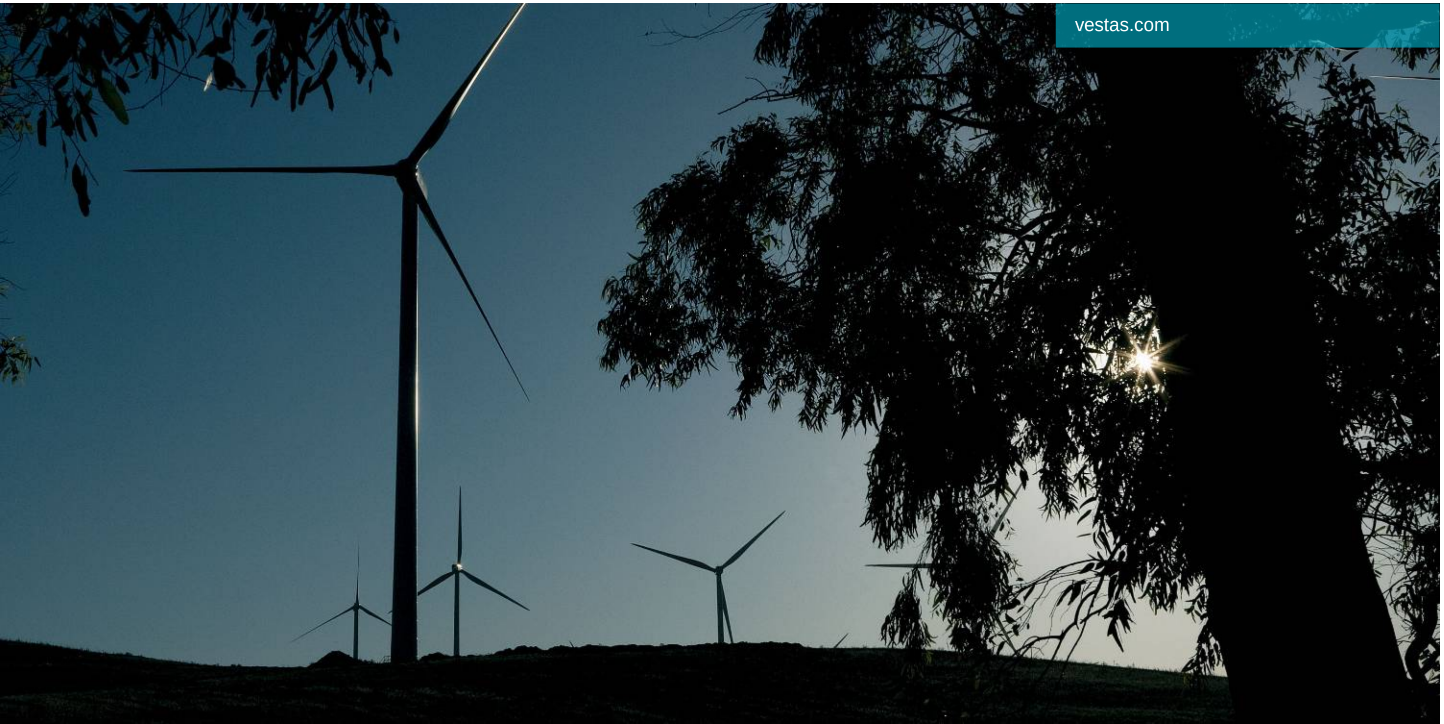
2 Proportion Test – Customer Satisfaction

- Problem:
  - After conducting a customer satisfaction survey, you suspect that the overall percentage of customers that are 'dissatisfied' with calls from home is different in your two key market areas; Central London and the Midlands.
- Results:
  - In Central London, 987 out of 19,554 were dissatisfied
  - In the Midlands, 846 out of 17766 were dissatisfied
- Can you decide if the two areas have different customer satisfaction levels?

Vestas

No. 1 in Modern Energy

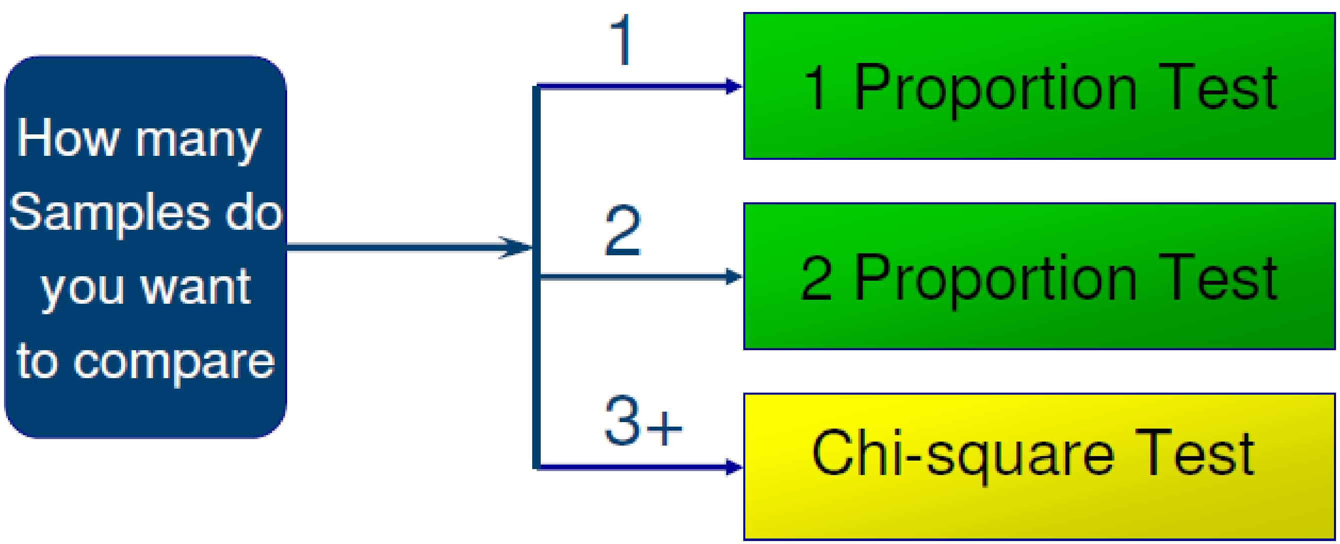
Chi Squared Test



Vestas

To Compare the Proportions Data Samples

You want to compare proportions or percentages that came from different samples of data to decide if they 'statistically' different.



Vestas

Hypothesis Testing For Attributes

- The χ^2 test is used to test hypotheses about the frequency of occurrence of some event happening against the expected frequency this is often called a “goodness of fit test”

$$H_0 : p_1 = p_2 = p_3 = \dots p_n$$

$$H_a : \text{ at least one is not equal}$$

- It can also be used to test hypotheses about the relationships between sources of variation being dependent or independent this is often called a test of association

$$H_0 : \text{independent}$$

$$H_a : \text{dependent}$$

Chi Squared Tests (1)

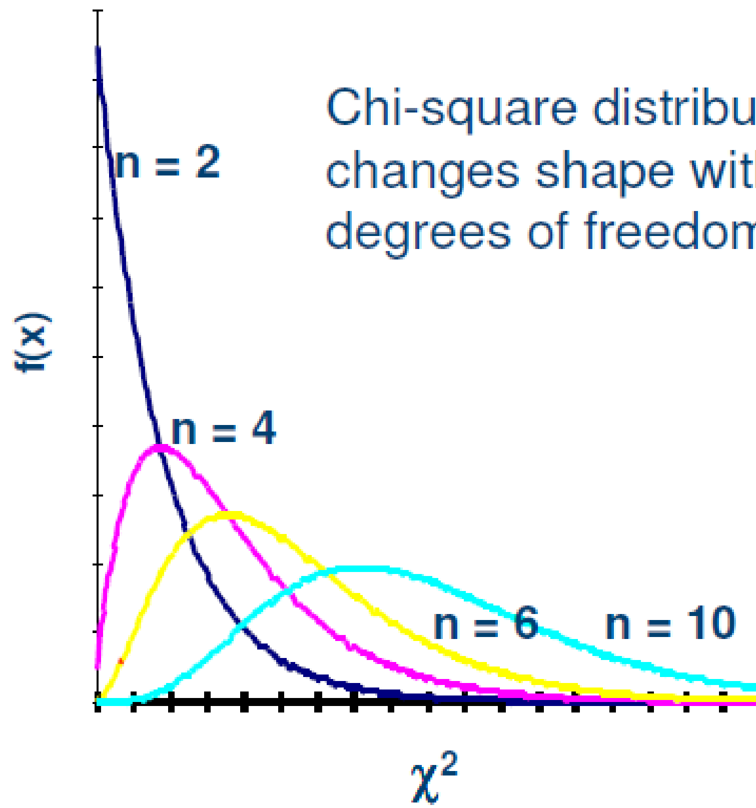
- Chi Squared tests are used to compare proportions from three or more samples of data. The samples may represent:
  - different suppliers
  - different processes
  - different teams
  - different products etc
- The samples can be of different sample sizes, since the Chi Squared test takes account of both sample sizes.

Chi-squared Distribution

Chi-squared is calculated from

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

Where f_o and f_e are the observed and expected frequencies respectively



Chi-square distribution changes shape with the degrees of freedom (n)

Vestas

Chi Squared Tests (2)

- Chi Squared tests focus on the difference between the observed and expected results in each category:

	Observed (fo)	Expected (fe)	$\Sigma \frac{(f_o - f_e)^2}{f_e}$
Heads	63	50	
Tails	37	50	

What kind of χ^2 test is this?

Deg. of Freedom =

$\chi^2_{calc} =$

$\chi^2_{table} =$

Goodness of Fit Example

A coin is flipped 100 times. There are 63 heads and 37 tails! We would expect 50:50 so could the observed have happened by chance or is the coin biased?

	Observed f_o	Expected f_e	$\Sigma(f_o - f_e)^2 / f_e$
H	**63**	**50**	**3.38**
T	**37**	**50**	**3.38**
		χ^2 Calc	**6.76**

We need to calculate the degrees of Freedom =
No. Categories -1
DoF = 2 - 1 = 1

From χ^2 table with $\alpha = 0.05$
$\chi^2 = 3.841$
since $6.76 > 3.841$
we reject H_0 so the coin is biased

Vestas®

Test of Association Example

A improvement project team is working to reduce the number of errors in purchase orders.

	Department				
	Design	Engineering	Test	Stores	Total
With Errors	45	18	6	20	89
No Errors	280	180	105	222	787
Total	325	198	111	242	876

Do the error rates differ between departments?

Vestas

Approach

The approach is to use proportions since this allows comparison of categories even if the sample sizes differ.

	Department				
	Design	Engineering	Test	Stores	Total
With Errors	13.85%	9.1%	5.41%	8.26%	10.16%
No Errors	86.15%	90.9%	94.59%	91.74%	89.84%
Total	100%	100%	100%	100%	100%

χ^2 (CHI)^2 Analysis

- If there were no true difference between the departments, then each would have the same relative percentage of errors

$$\frac{\text{Row Total}}{\text{Grand Total}} = \frac{89}{876} = 10.16\%$$

- We can use this figure to determine the expected performance
- Example

Design: 0.1016x 325 = 33.02

Calculating the χ^2 Value

	Department							
	Design		Engineers		Test		Stores	
	Obs	Exp	Obs	Exp	Obs	Exp	Obs	Exp
With Errors	45	33.02	18	20.12	6	11.28	20	24.59
No Errors	280	292	180	177.9	105	99.72	222	217.14
Total	325	325	198	198	111	111	242	242

χ^2 Measures the Differences between the Observed and Expected

$$\chi^2 = \sum \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$

In this case

$$\chi^2 = (45 - 33.02)^2/33.02 + (18 - 20.12)^2/20.12 + \dots + (105 - 99.72)^2/99.72 + (222 - 217.14)^2/217.14 = 8.788$$

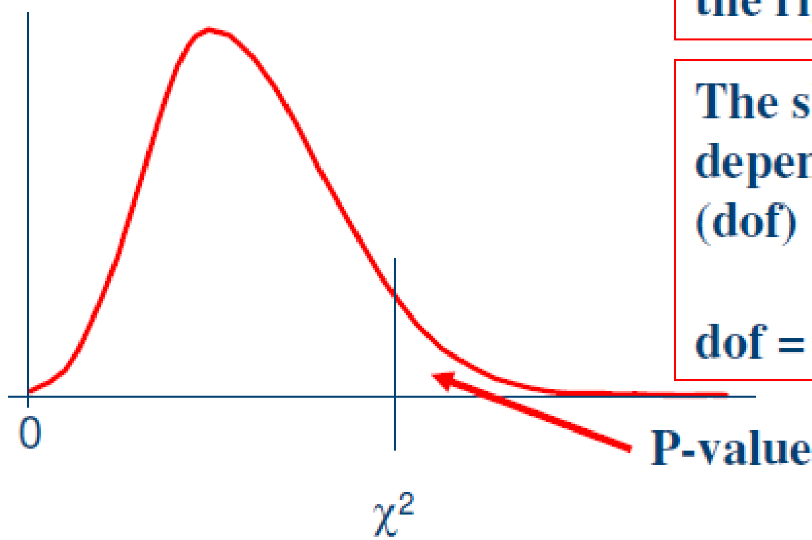
χ^2 Distribution

- The statistic χ^2 has a distribution which we can use to determine a p-value

The χ^2 -distribution is skewed to the right with a lower bound of 0

The shape of the χ^2 -distribution depends upon the degrees of freedom (dof)

dof = (No. Rows -1) x (No. Columns -1)



In this Example dof = (2 -1)(4 – 1) = 3 and χ^2 = 8.788
From the χ^2 table for α = 0.05 and 3 dof = 7.815 (see next slide)

Degrees of Freedom: 3

At an $\alpha = 0.05$

The $\chi 2$ is:

7.815

χ^2 Distribution

Dof/α	.250	.100	.050	.025	.01	.005	.001
1	1.323	2.706	3.841	5.024	6.645	7.879	10.828
2	2.773	4.605	5.991	7.378	9.210	10.597	13.816
3	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	5.385	7.779	9.488	11.143	13.277	14.860	18.467
5	6.626	9.236	11.070	12.832	15.086	16.750	20.515
6	7.841	10.645	12.592	14.449	16.812	18.548	22.458
7	9.037	12.017	14.067	16.013	18.475	20.278	24.322
8	10.219	13.362	15.507	17.535	20.090	21.955	26.125
9	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	12.549	15.987	18.307	20.483	23.209	25.188	29.588
11	13.701	17.275	19.675	21.920	24.725	26.757	31.264
12	14.845	18.549	21.026	23.337	26.217	28.300	32.909
13	15.984	19.812	22.362	24.736	27.688	29.819	34.528
14	17.117	21.064	23.685	26.119	29.141	31.319	36.123
15	18.245	22.307	24.996	27.488	30.578	32.801	37.697
16	19.369	23.542	26.296	28.845	32.000	34.267	39.252
17	20.489	24.769	27.587	30.191	33.409	35.718	40.790
18	21.605	25.989	28.869	31.526	34.805	37.156	43.312
19	22.718	27.204	30.144	32.852	36.191	38.582	43.820

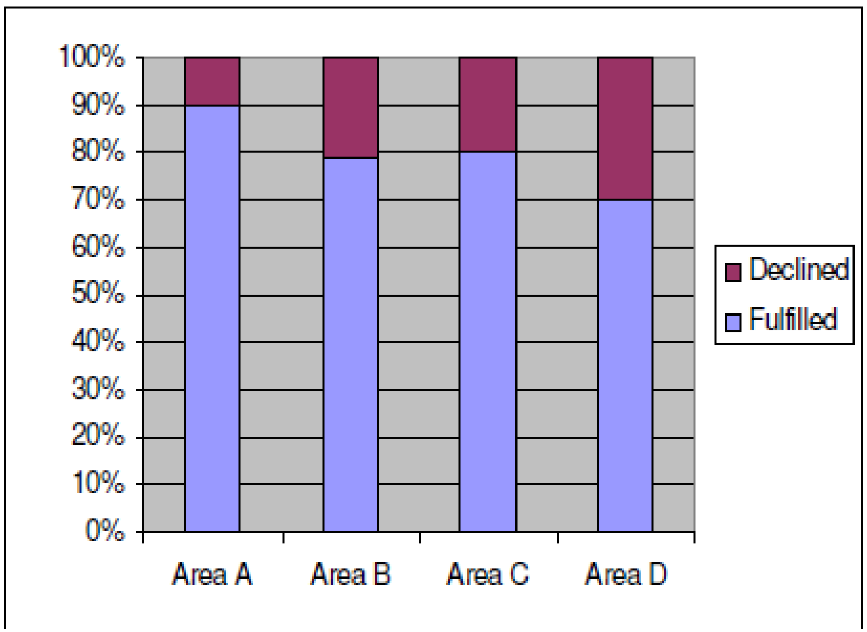
Since the calculated value 8.788 > table value 7.815 we conclude that there is a difference between the different departments.

Assumptions for χ^2 Tests

- Samples are representative
- Underlying distribution is binomial
- The expected count must be greater or equal to 5 for each cell or the test will not perform properly
    - If the expected cell count is less than 5 the only option is to collect more data

Chi Squared Tests (3)

- The project focusing on the % of applications fulfilled (see previous 2 Proportion example) needs to compare the fulfil rates of several more areas



Theory: It *appears* that there are different fulfil rates across areas A-D.

Chi Squared Tests (4)

- A Chi-Squared test can be used to test if the difference between the fulfil rates is statistically significant
- The hypotheses for the test would be:
  - Ho: There is no difference between the % fulfil rates of areas A, B, C and D
  - Ha: There is a difference between the % fulfil rates of areas A, B, C and D
- Because our theory is that there is a difference, we expect to reject the Null hypothesis

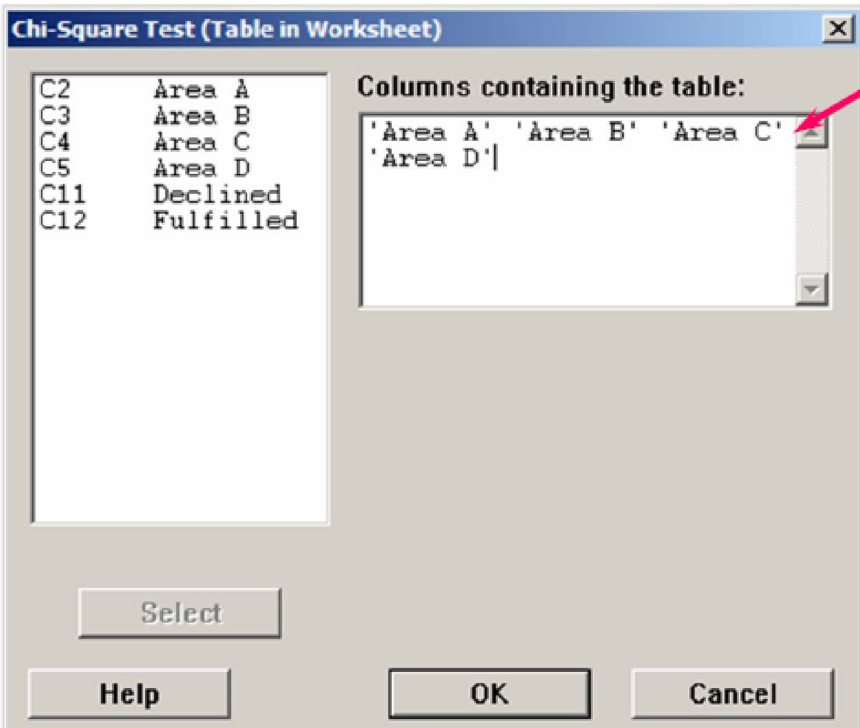
Chi Squared Tests (5)

- Minitab's Chi Squared test requires the data to be summarised in a table format as shown below.
- Note that only the number of results observed in each category are used, and the sample sizes are not added. Minitab calculates the sample sizes from the data provided.

	Area A	Area B	Area C	Area D
Fulfilled	72	79	120	42
Declined	8	21	30	18

Chi Squared Tests (6)

Minitab: Stat >Tables>Chi-square test



Enter the data columns of the table here.

Note: Do not enter the column that contains the data labels.

Vestas

Chi Squared Tests (7)

Chi-Square Test: Area A, Area B, Area C, Area D

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

        Area A  Area B  Area C  Area D  Total
    1       72      79     120      42    313
         64.21   80.26  120.38   48.15
         0.946   0.020   0.001   0.786

    2        8      21      30      18     77
         15.79   19.74   29.62   11.85
         3.847   0.080   0.005   3.197

Total       80     100     150      60    390

Chi-Sq = 8.882, DF = 3, P-Value = 0.031
```

Session Window Output

for Chi- Squared test

The p-value for the Chi-Squared test is found in the session window output. In this case, the p-value is 0.031

Since the p-value for this test is less than 0.05, we can be over 95% confident that there is a difference between the fulfil rates of the four areas.

Vestas

Chi Squared Tests (8)

Exercise: Open data file: “Chi-Squared - Fulfil Rates.MPJ”

The data file has the same data table in two different formats.

1) Follow the step by step instructions shown on the previous pages.

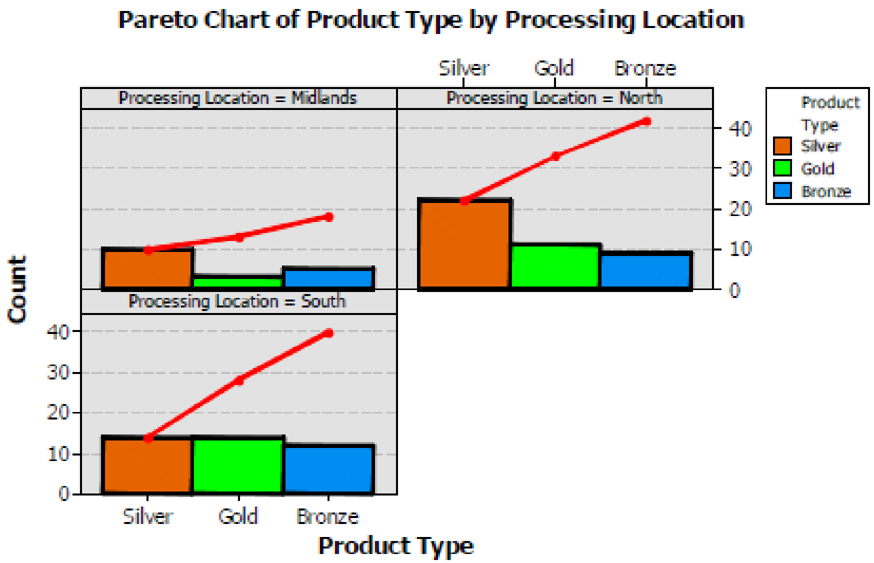
2) Repeat the test using the alternative data layout shown in columns C10 thru C12.

Hypothesis Testing - Case Studies (1)

- Using case study 1 (Cycle Time), test the following theories using an appropriate hypothesis test, and select an appropriate graph to show the results
- Theory: The North processing location has proportionally more Silver product applications than the Midlands and South (combined)

Hypothesis Testing - Case Studies (2)

Theory: The North processing location has proportionally more Silver product applications than the Midlands and South (combined).



P-Value = 0.274

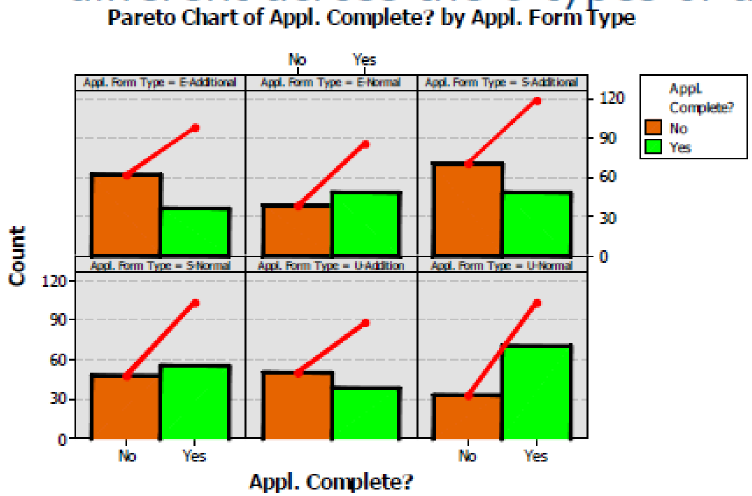
The p-value is ***above*** 0.05, which indicates that there is ***no*** statistically significant difference in the proportion of silver product applications between the North and other locations.

Hypothesis Testing - Case Studies (3)

- Using case study 2 (Incomplete Applications), test the following theory using an appropriate hypothesis test, and select an appropriate graph to show the results
- **Theory**: The proportion of complete applications is different across the 6 types of application form

Hypothesis Testing - Case Studies (4)

Theory: The proportion of complete applications is different across the 6 types of application form.



This can be seen on the Pareto chart above.

P-Value = 0.000

The p-value is below 0.05, which indicates that we can be over 95% confident that there ***is*** a statistically significant difference in the proportion of complete applications across the 6 types of application form.

Vestas

Any questions?