FIREBRAND

Lean Six Sigma

Black Belt

Courseware

Version 2.5

www.firebrandtraining.com

Case Study – Acquanas International

Lean Six Sigma

Black Belt

Case Study – Acquanas International

Congratulations!! You have been recruited as a Black Belt by Vintage Calls, a provider of outsourced call handling solutions to the utilities industry.

Vintage Calls major client is a heavily regulated water supplier named Acquanas International. Vintage have been taking calls on behalf of Acquanas for the last year.

Acquanas have become disappointed with performance at Vintage and the water industry regulator has become increasingly aware of the service Acquanas customers have been receiving. As a result, Acquanas have been considering changing to another local supplier.

In response, Vintage have launched a continuous improvement programme and have prioritised a project to improve the Acquanas call performance as of key strategic importance to the organisation.

You have been recruited as the project team responsible for running this project.

Case Study – Acquanas International

Other information:

- In an effort to improve performance, Vintage have invested heavily in a new Interactive Voice Response (IVR) system to help better route calls but to date they have had little success in optimising this new functionality.
- Vintage have recently recruited significant additional staff to help address the performance issues.
- To promote teamwork, staff are paid a bonus based on overall **contact centre call quality** and **call wait times**.
- All staff work 8am-5pm with 1 hour for lunch (staggered from 11:30am).

Your workshop leader will assume the role of CEO of Acquanas, your major customer as well as your Business Champion and Subject Matter Expert. Your workshop leader has information and data for you but will only provide it if you ask the correct questions!

Please work as a team to prepare and present the Define tollgate by the end of day 1 using the Green Belt and Black Belt knowledge at your disposal – Good Luck!

Welcome

Lean Six Sigma

Black Belt

Welcome to the

Lean Six Sigma

Black Belt

Training

Introductions

Be prepared to share with the class your:

- Name
- Designation
- Organisation
- Location
- Expectations for the course
- Summary of your project

Graphical view

	Key content Indicates key concepts/ content that might be referenced often
	Class activity Prompts to instructor and participants for individual or group activities
	Coach support points Indicates areas where the Coach might be consulted to complete critical project tasks

Software

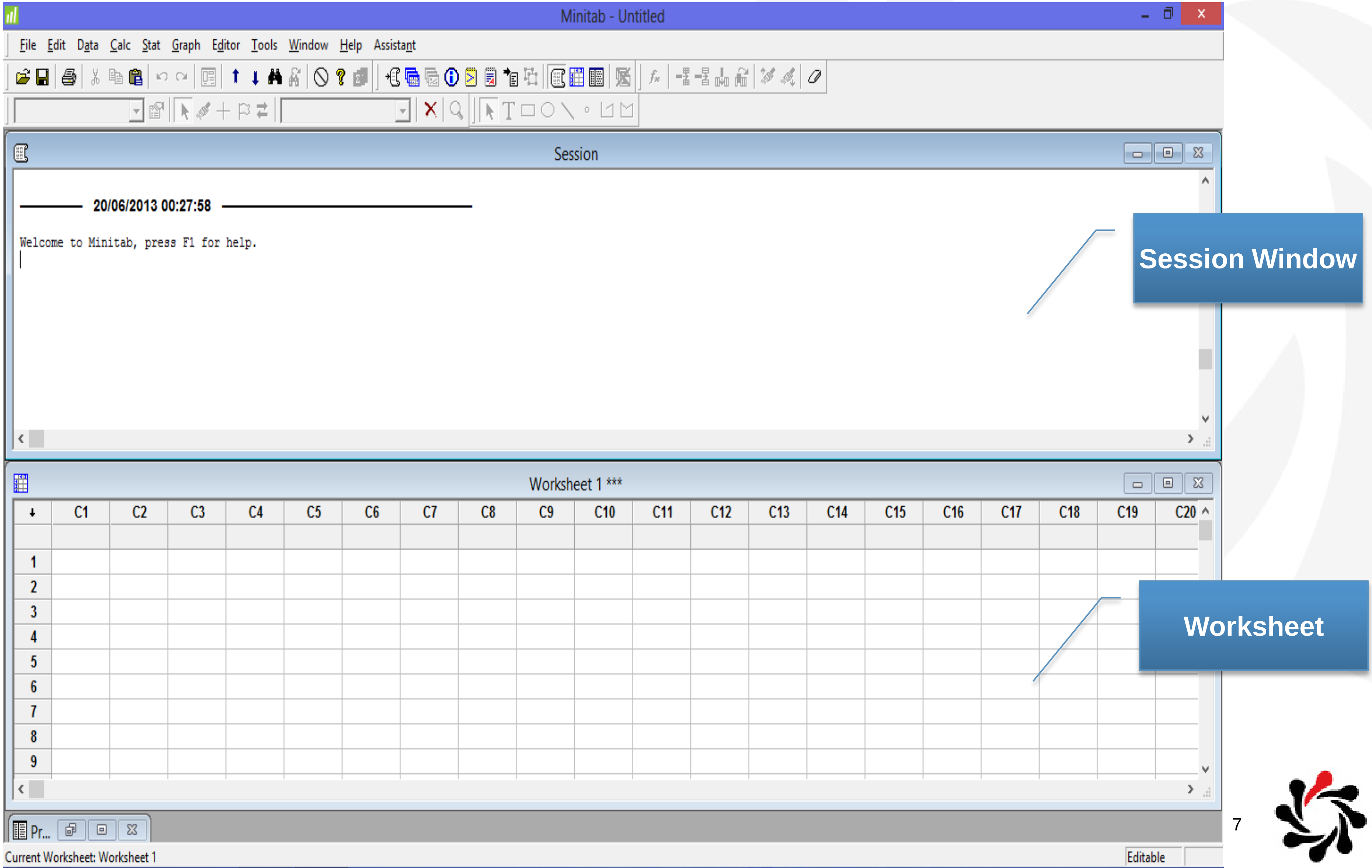
Introduction to Minitab

Objectives of this module

At the end of this module, you will be able to

- Have a basic appreciation of how Minitab v16 can help you generate **information from data**
- Understand where to look for help, should you need it, within the software
- Revise some basic concepts from Green Belt

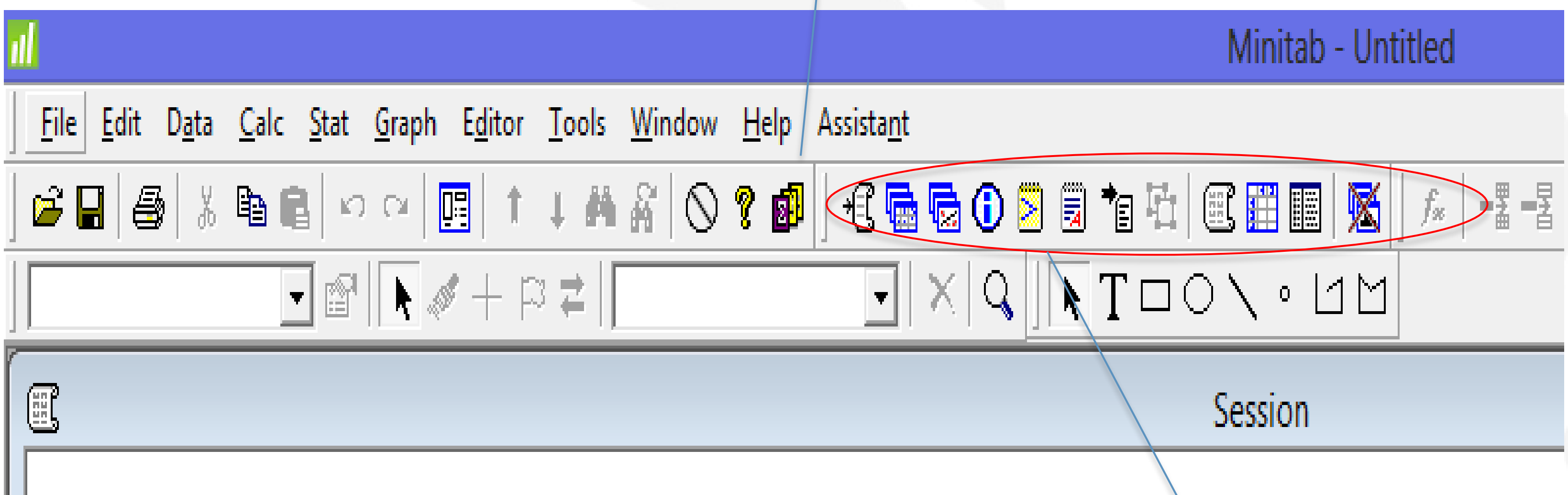
Minitab screen



Toolbar buttons

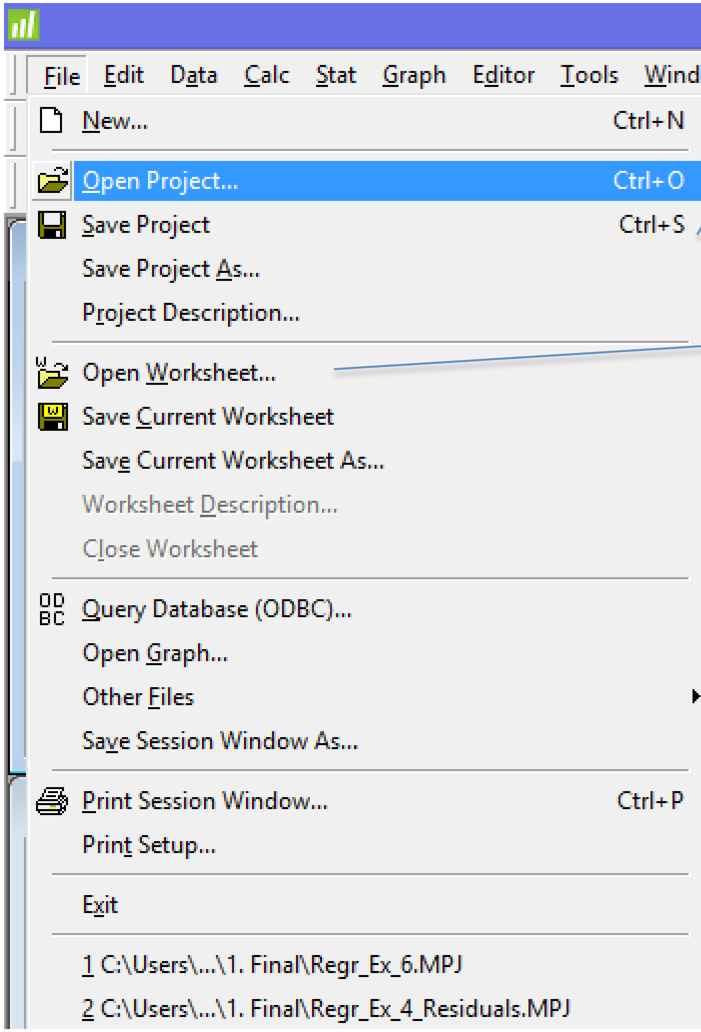
Useful buttons on the toolbar to manage what you see and to get context sensitive help

Stat guide – very useful



Change the windows you see – session, worksheet, graphs etc

Minitab file menu



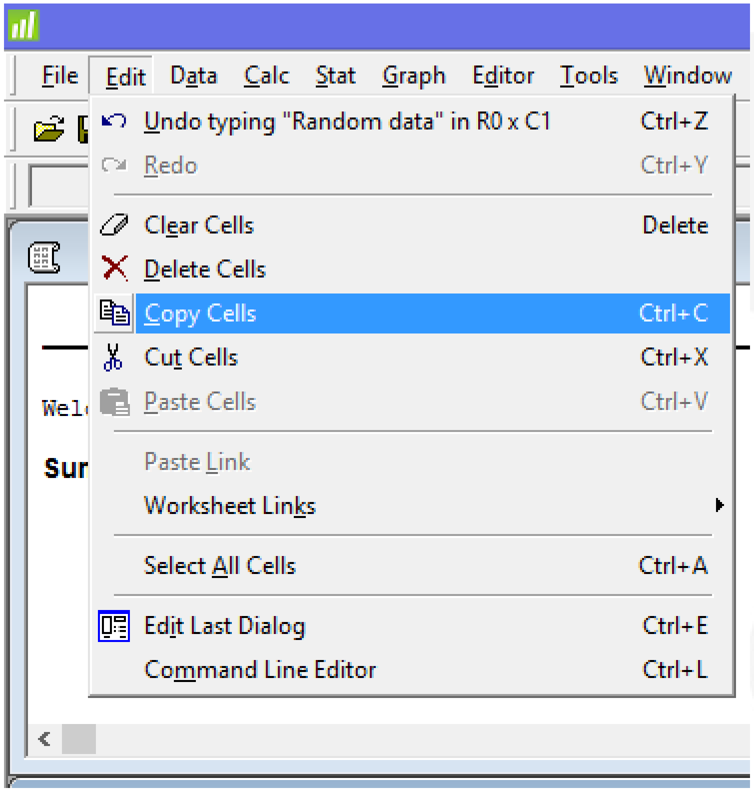
Open project .mpj file

Open worksheet .mtw file

In Minitab, projects and worksheets can be saved. The icon below the menu bar only opens project files.

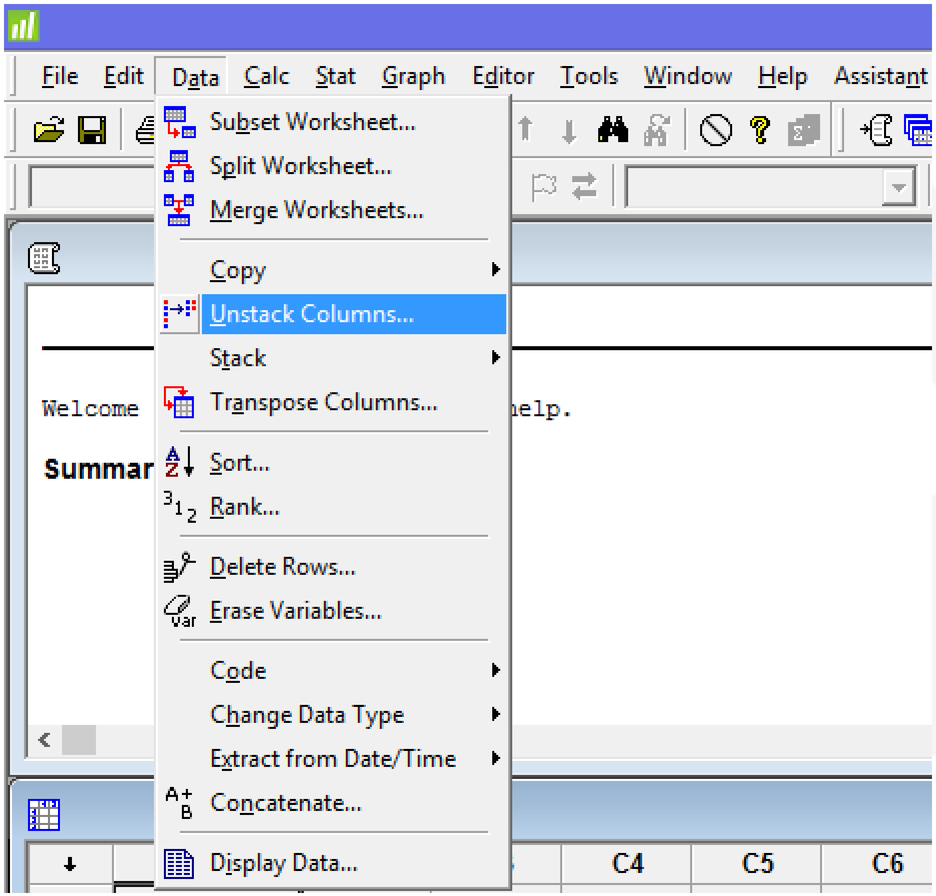
File/open must be used for worksheets

Minitab Edit menu



Works in a similar way to Excel allowing you to manipulate the content of cells

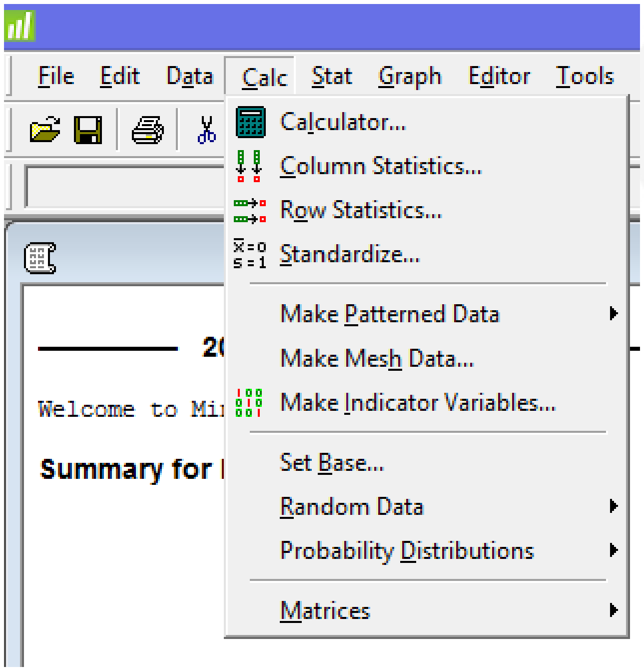
Minitab Data menu



Minitab statistical analysis requires you to present your data for analysis within you worksheets in specific ways.

This functionality allows you to manipulate that data, ready for analysis

Minitab Calc menu



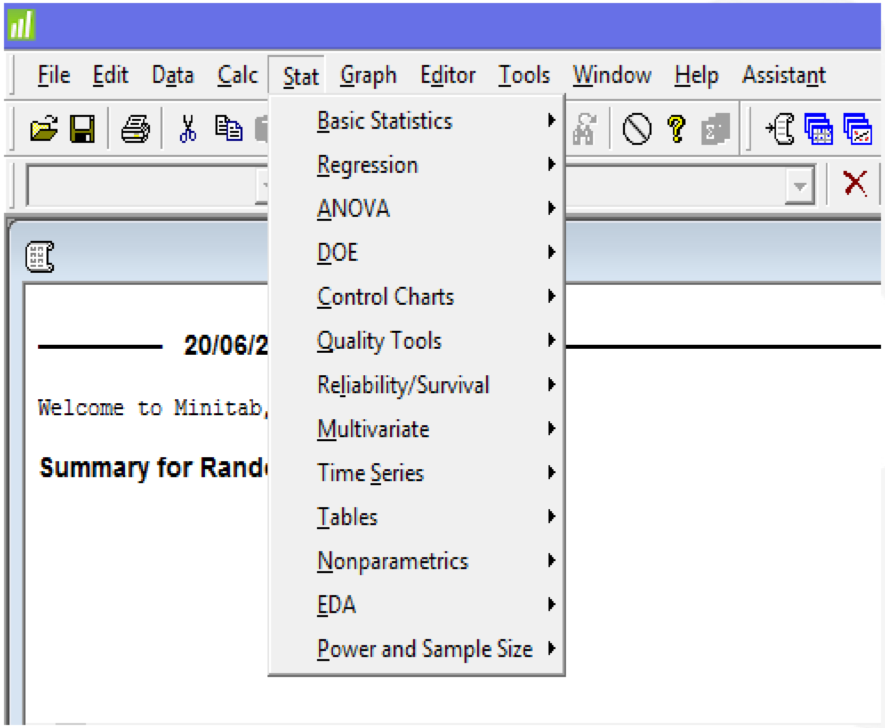
This functionality allows you to create data within your worksheet.

This data could be as a result of a formula created or could simply be randomly generated.

Lets create some Random Data!

- A Normal Distribution
- 100 rows
- Mean of 5
- Standard deviation of 3

Minitab Stat menu

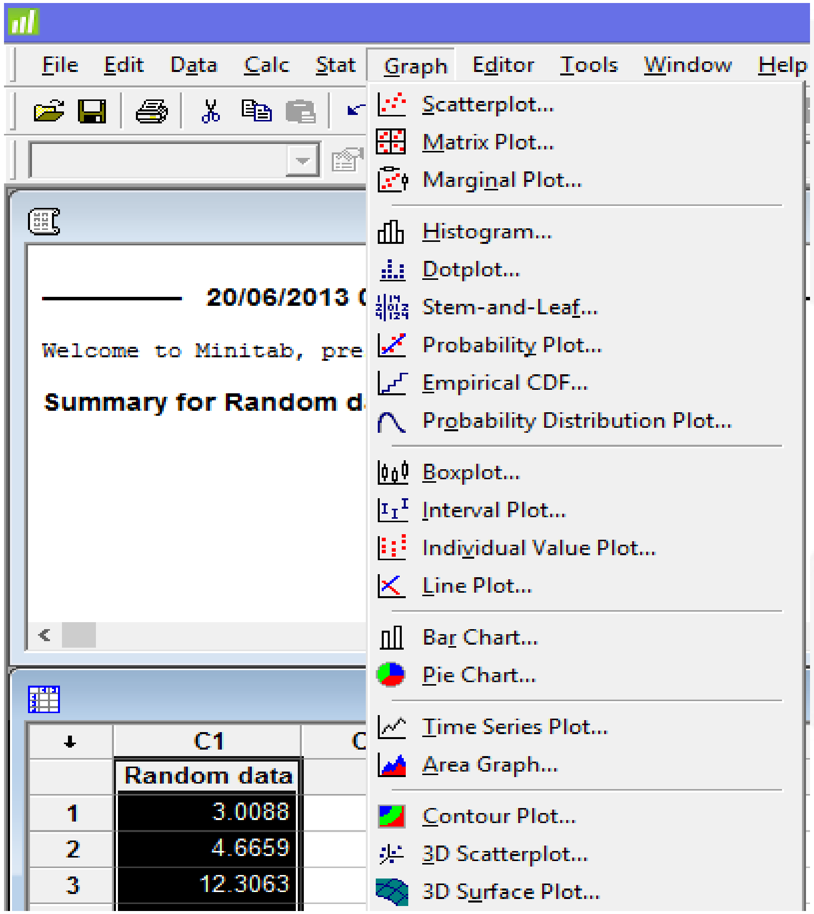


This area is your home for statistical analysis.

As you can see there's plenty of functionality ranging from calculation of sample size through to logistic regression.

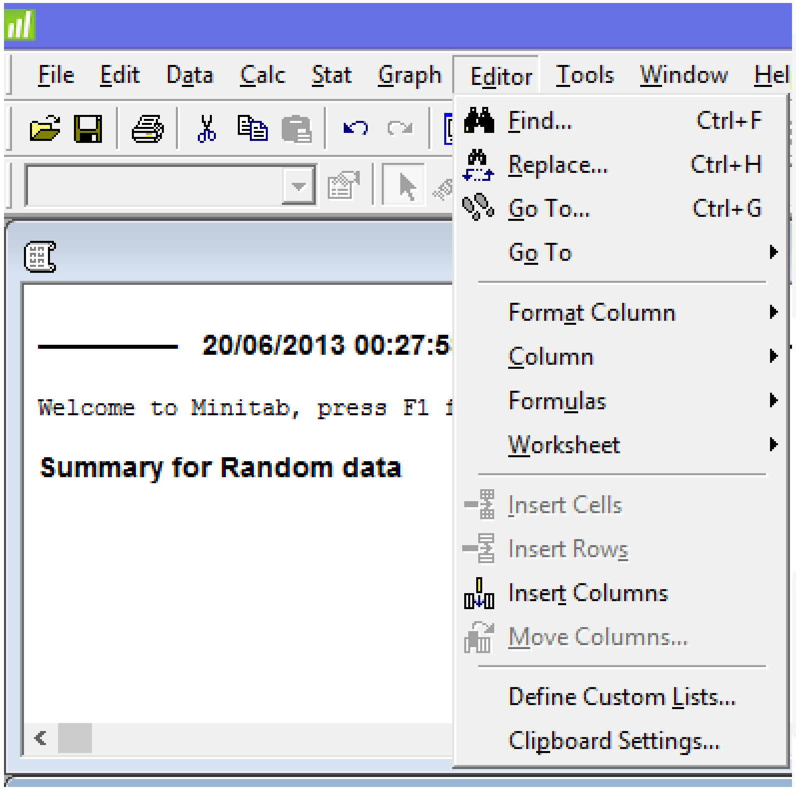
Let's create a graphical summary of the random data we've just created.

Minitab Graph menu



Spot trends, picture variability, some pretty cool stuff in here!

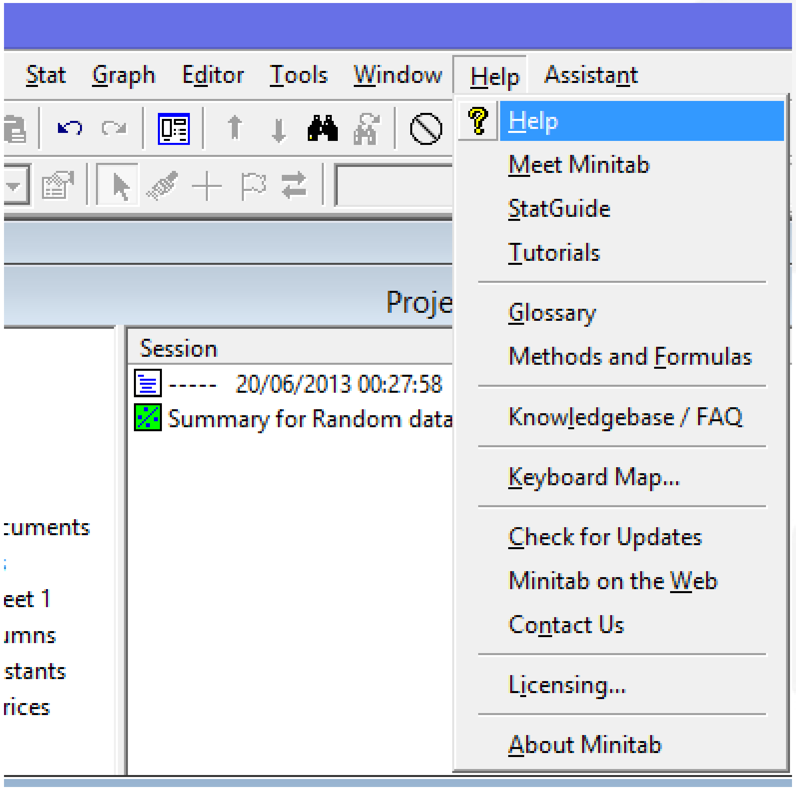
Minitab Editor



A very useful function when dealing with large volumes of data

Some of the functions are similar to Excel including Inserting Columns, Rows, a new worksheet etc.

Help functionality

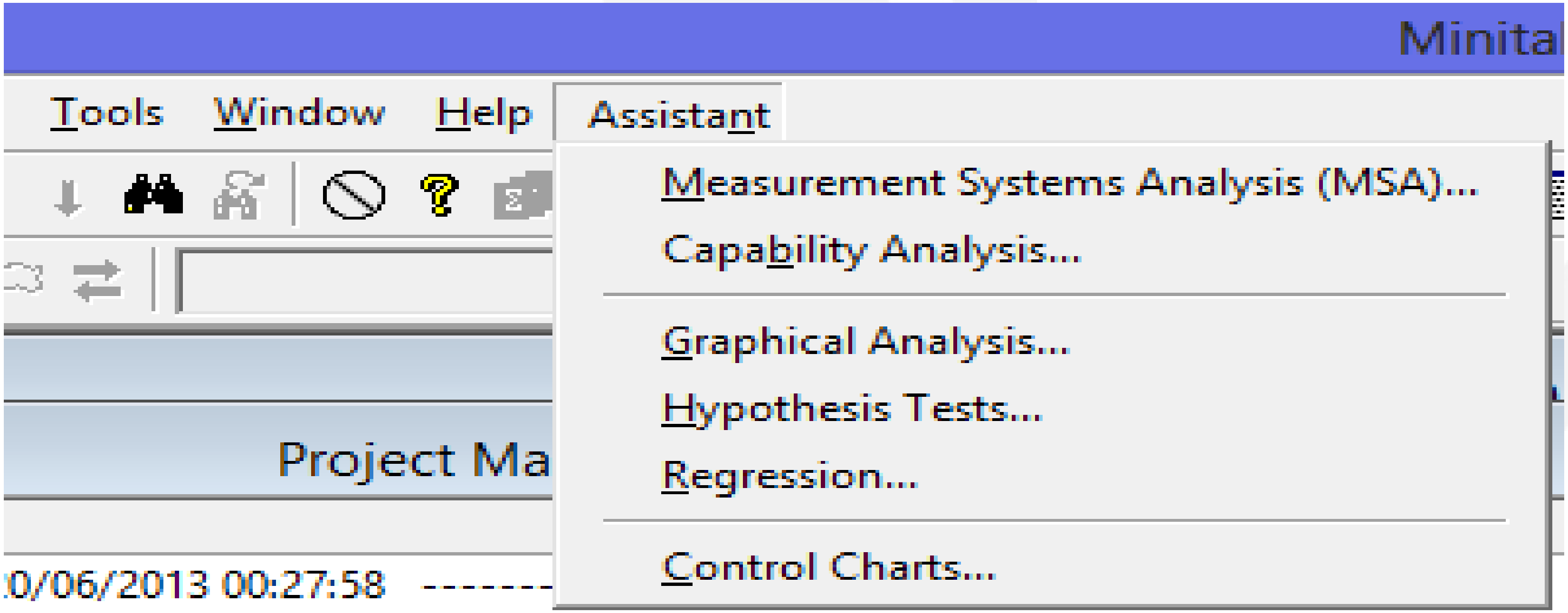


The help includes a full index and description of all its functionality.

Included within this are statistical tutorials, FAQs and even contact details for Technical support from the team of statisticians at Minitab!

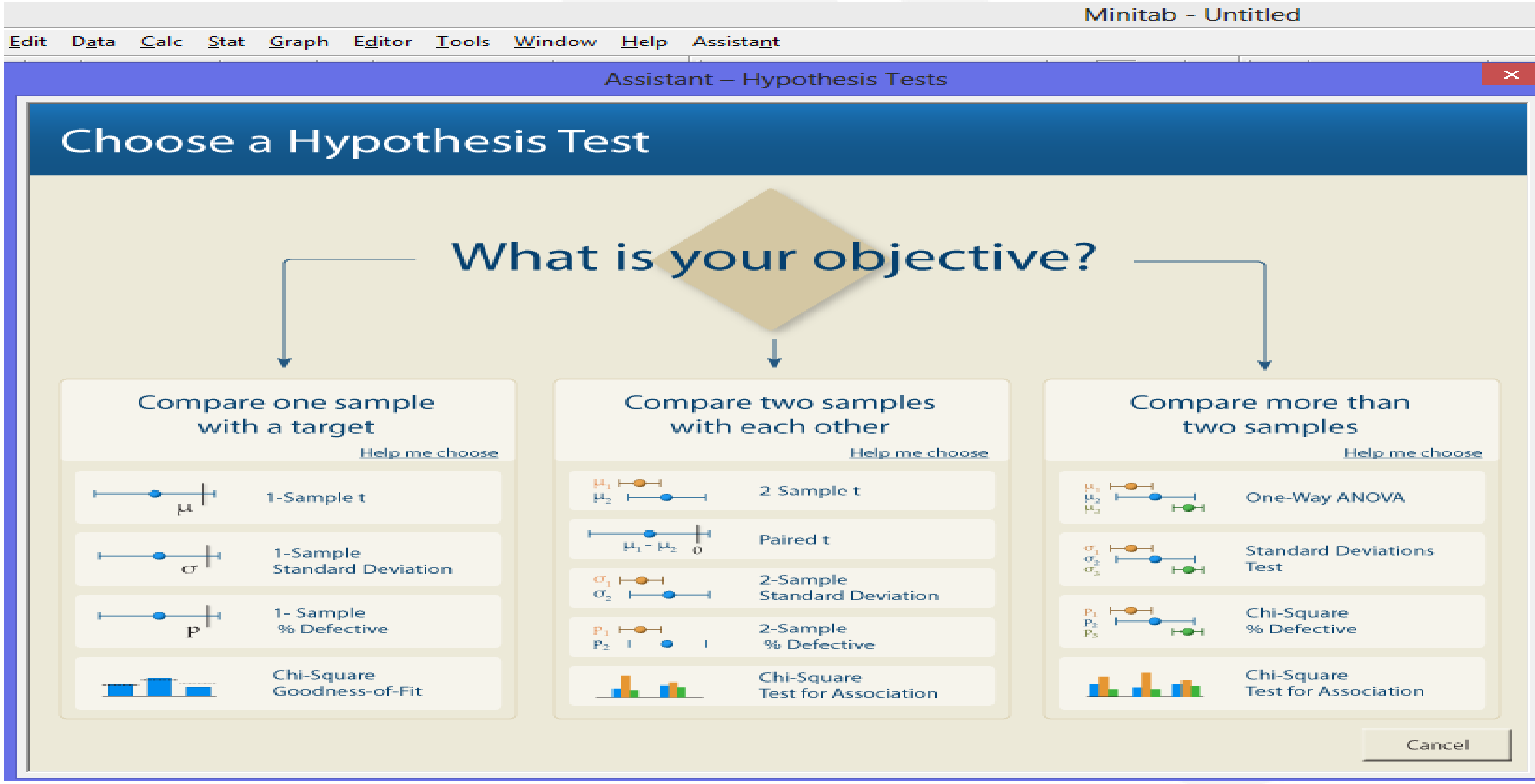
Minitab Assistant

Like many statistical software packages there's normally a few ways to produce the same result, Minitab is no different in that respect.



Minitab Assistant

Within the Assistant menu, it will guide you on MSA, Capability, Graphical Analysis, Hypothesis Test, Regression and Control charts by asking you a series of questions.



Some Green Belt Revision

Now we’ve had a brief introduction to what Minitab can provide us, lets remind ourselves of some of the basic concepts from Green Belt training to ensure we get the most out of the software……..

Some Green Belt Revision

Group 1 – Teach back on hypothesis tests

Please cover:

- P-values
- Alpha and Beta Risk

Some Green Belt Revision

Group 2 – Teach back on SPC

Please cover:

- Discrete v Continuous data
- Defects v Defectives

Some Green Belt Revision

Group 3 – Teach back on MSA

Please cover:

- Sources of measurement system variation

Key learning points

- Minitab is the principal statistical software used in Lean Six Sigma projects
- Minitab can support you analysing and drawing conclusions from data
- The software is user friendly and contains lots of support should you need it
- A reminder of some key Green Belt concepts

Black Belt

Distributions

Objectives of this module

At the end of this module, you will be able to:

- better understand the **different types** of statistical distributions
- identify the **uses** of different distributions
- make **assumptions** given a known distribution

“Questions Lead, Tools Follow”

Basic question: What is the shape of my data?

- Is my data symmetrical or is it skewed out to one side or the other?
- How much of my data is **above the average**?
- How much of my data **is above the customer’s specification**?
- Does it look like **more than one thing is going on**?

Why Distributions?

- Base Line Calculations
  - Normal Data: Z-Score gives Sigma Level.
  - Non-Normal Data: Currently can only use discrete defect counts.
  - Knowing the distribution can give us the same accuracy as for a normal distribution.
- Prediction
  - What percent is above or below spec?
    - “I think 95% of the time we meet customer expectations of 8 hours or less… Can you tell me if the data reflects that?”
- Process Understanding
  - “Customers are complaining that we take too long to resolve issues. Here’s the data, can you tell me how often we take longer than 5 days to resolve issues?”

Exercise 1

Consider the following scenario:

You need to estimate how well your Banking Centre is doing in managing **customer wait time**. You gather a representative sample of **100 customer wait times**. The data are in Distr_Ex_1.MPJ Given an upper specification limit of 5 minutes, calculate the sigma level of your process and the proportion of your customers that you predict will have to wait **more than 5 minutes**.

Take 10 minutes to work on this

What is Probability?

Probability of an outcome is a number that measures the likelihood that the outcome will occur when the process is observed.

Probability—Proportion—Percentage

In toss of a fair coin, there is a probability of 0.5 that it will land “heads” up so the proportion of “heads” expected will be 0.5, said another way, we expect 50% of the tosses to result in “heads”.

What are Distributions?

Probability distribution is a mathematical rule that specifies the possible outcomes and gives the probability of observing each of these values.

Properties of a distribution:

- Probability of observing any particular value must be between 0 and 1.
- Sum of probabilities of all the values must equal 1.

Distribution Properties

The probability that any one of two or more values occurs is equal to the sum of their individual probabilities.

$$P(X=1 \text{ or } X=2) = P(X=1) + P(X=2)$$

P( ) denotes probability of an outcome.

X is known as a *random variable.*

If X can take values 0, 1, 2, ... What is $P(X \leq 2)$? How about $P(1 \leq X \leq 4)$?

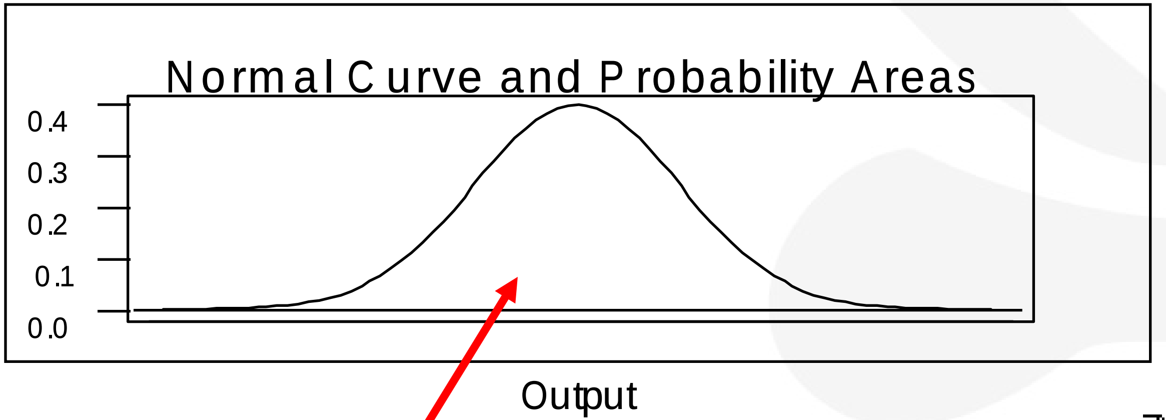
Distribution Properties

- $P(X <= 2) = P(X=0) + P(X=1) + P(X=2)$

- $P(1 <= X <= 4) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$

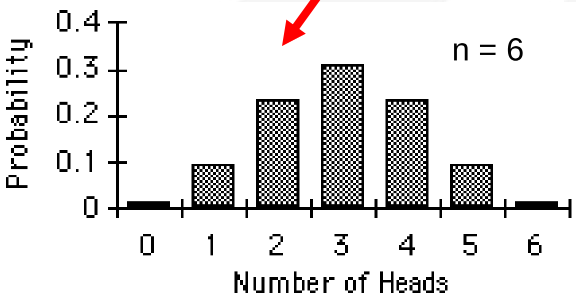
Distributions and Data Types

Distributions for **discrete** random variables look different from distributions for **continuous** random variables.



Continuous

Discrete



Discrete vs. Continuous Distributions

Continuous

Probability Density Function (PDF)

- For example, the probability that x is between 3.15 and 3.25.
- Denoted by $f(x)$

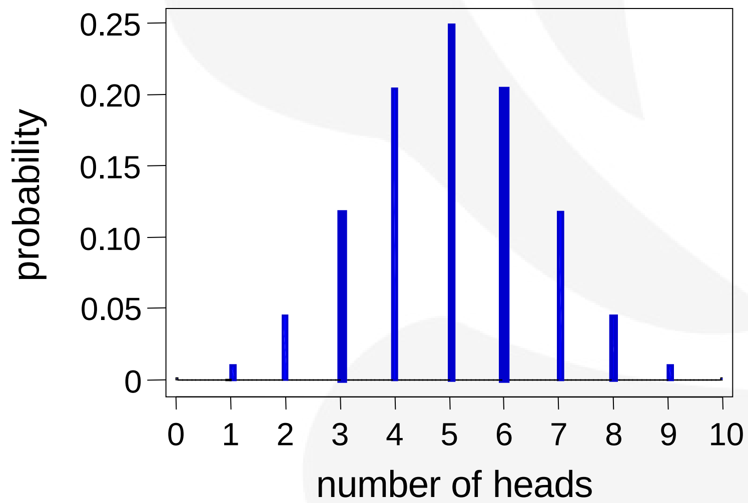
Discrete

Probability Mass Function (PMF)

- For example, the probability that x is 3.
- Denoted by $P(X=x)$

Example
Discrete Probability Distribution

This is the probability distribution for the number of heads expected when flipping a coin 10 times (binomial distribution):

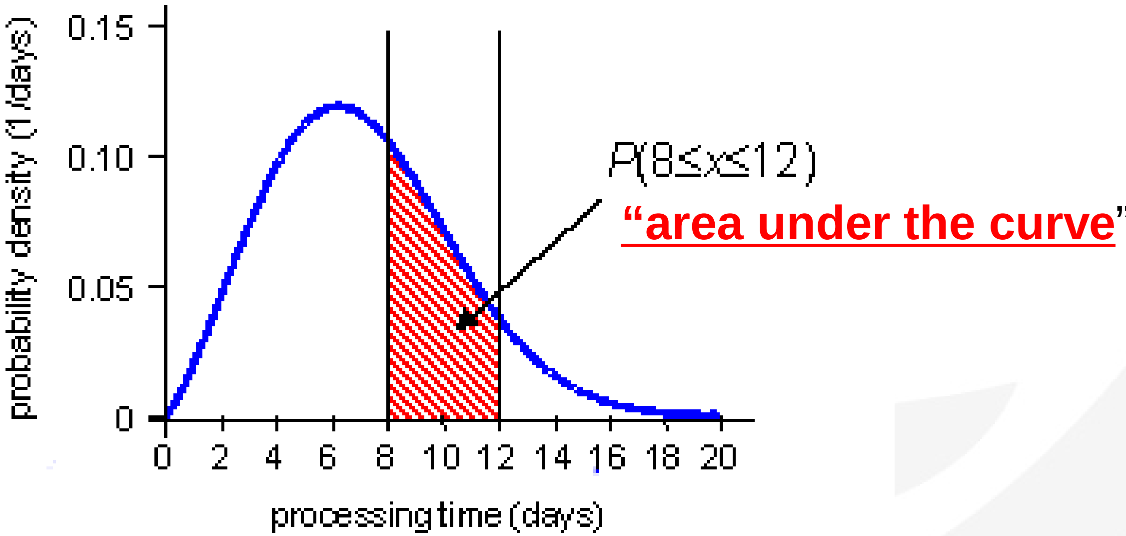


We can use this distribution to make predictions about coin flipping results:

- Probability of getting exactly 3 heads: $P(x=3) = 0.117 = 11.7\%$
- Probability of getting 8 or more heads: $P(x \geq 8) = P(x=8) + P(x=9) + P(x=10)$
  $= 5.5\%$

Example
Continuous Probability Distribution

The time it takes to process a credit card application may have a probability distribution that looks like this (Weibull distribution):



Knowing the probability distribution allows us to make predictions, for example:

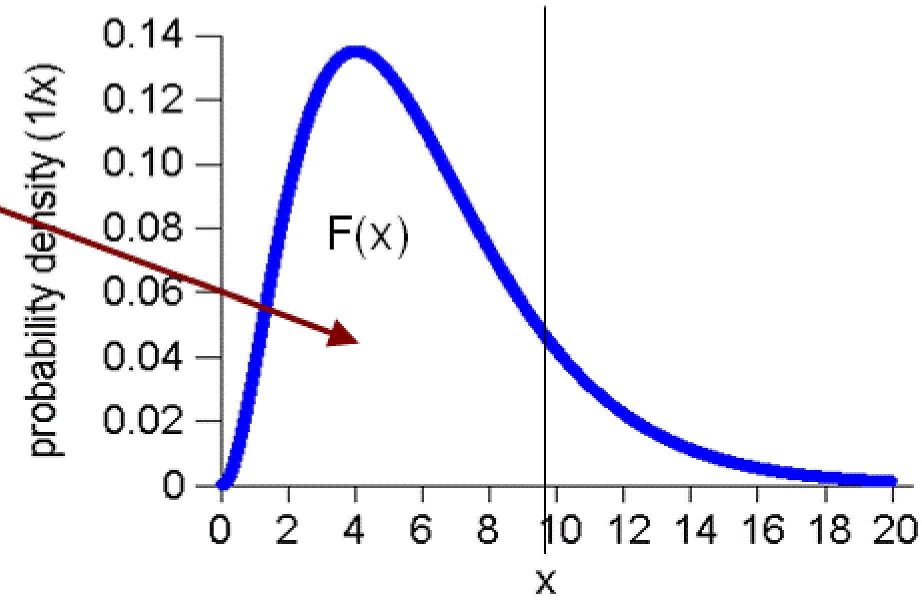
- The probability of taking between 8 and 12 days to process a credit card app is the area under the curve between 8 and 12.
- $P(8 \le x \le 12) = 0.289 = 28.9\%$ for the above distribution.

Cumulative Distribution Function

The probability that a variable takes a value less than or equal to x.

$P(X \leq x)$ for both discrete & continuous.

F(x) is the area to the left of the vertical line at x



Common Probability Distributions

Continuous

- **Normal**
- **Exponential**
- **Weibull**
- **Lognormal**
- **Chi Squared**
- **t**
- **F**

❖ ***Gamma***

❖ ***Beta***

Discrete

- **Binomial**
- **Poisson**

❖ ***Geometric***

❖ ***Negative Binomial***

❖ ***Hypergeometric***

❖ ***Not covered in this class.***

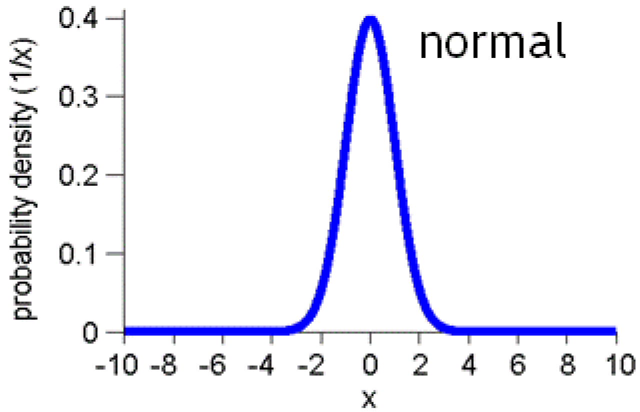
The objective is to gain insight into the process through a distribution analysis. This might mean identifying which distribution best fits our process data.

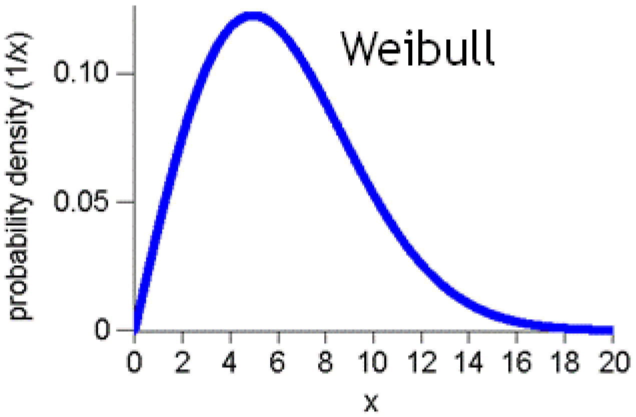
Distribution Descriptions

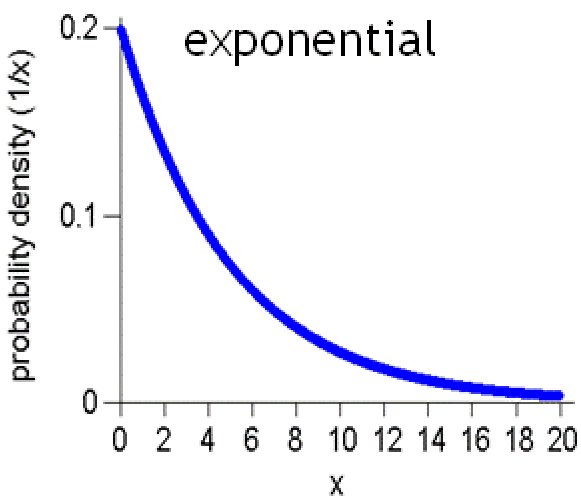
Parameters characterise the distributions.

- Location Parameter
  - The **lower or midpoint** (as prescribed by the distribution) of the range of the random variable. E.g., for a normal distribution, the mean.
- Scale Parameter
  - Determines the scale of measurement for x (magnitude of the x-axis scale). E.g., for a normal distribution, the standard deviation.
- Shape Parameter
  - Defines the PDF shape within a family of shapes. E.g., for a t distribution, the degrees of freedom.

Continuous Distributions

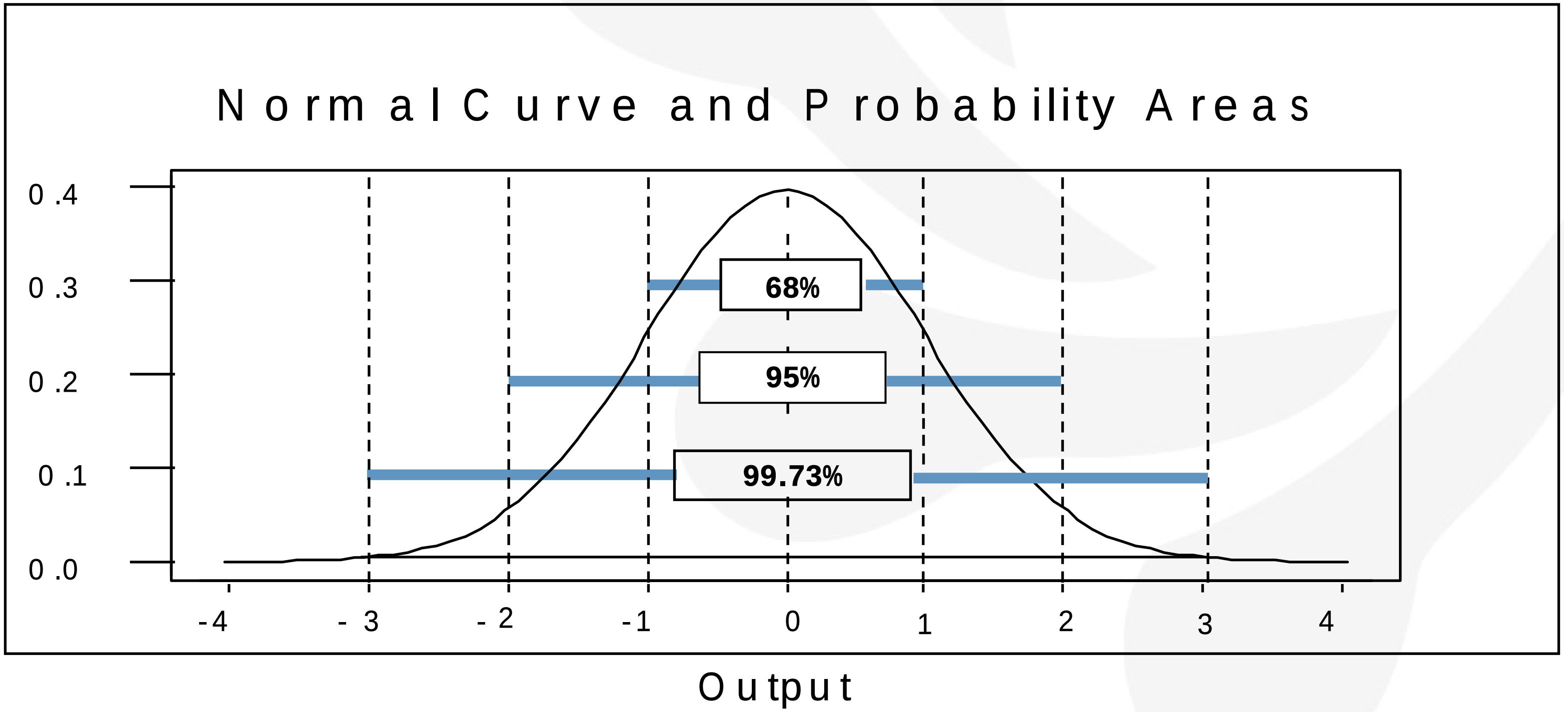






Normal Distribution

GB training: focused on inferential statistics based on a normal distribution.

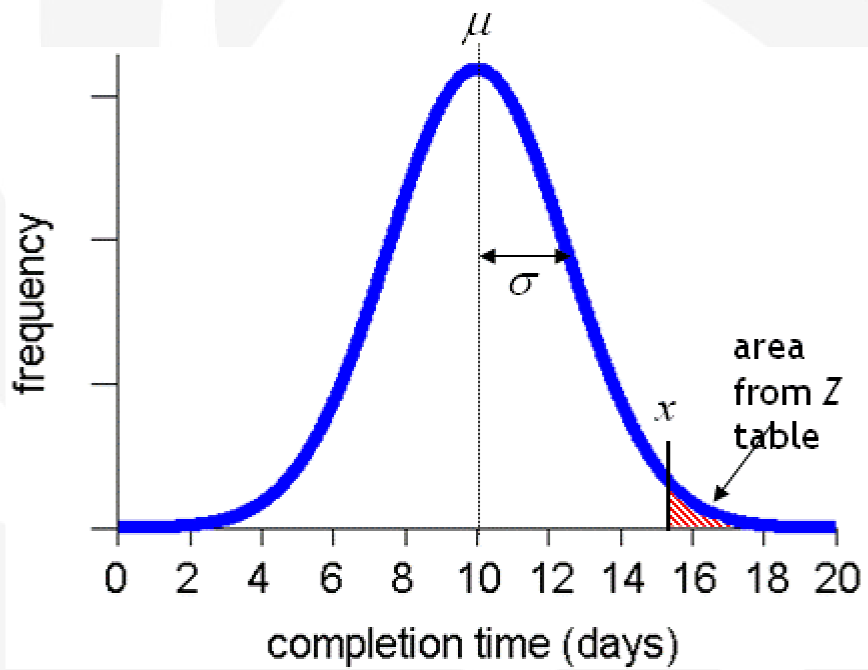


3 Characteristics – helps us predict process response & capability

Normal Distribution

- **Population mean**
  - **denoted by μ**
  - **describes location**
- **Population standard deviation**
  - **denoted by σ**
  - **describes spread and scale**
- ***Z* score at a given value of interest *x***

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



- ***Z* table shows the area under the curve from the *Z* score towards the left tail when *Z* is positive and toward the right tail when *Z* is negative.**

Applications of Normal Distribution

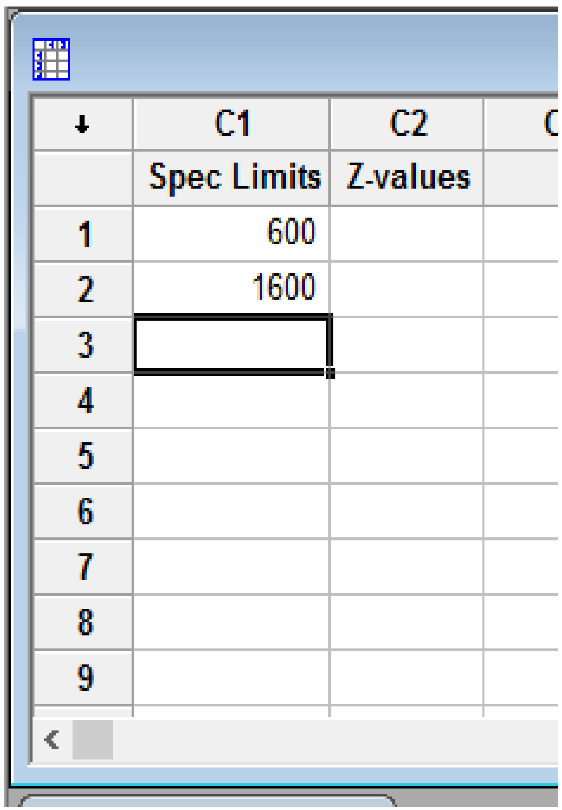
- Processes which have random variation about a target will typically be described mathematically by a normal distribution.
- This is the most commonly used probability distribution in Six Sigma.
- Uses of the normal distribution in Six Sigma include:
  - Capability analyses.
  - Hypothesis testing and confidence intervals.
  - Central Limit Theorem.
  - Understanding out of control conditions on control charts.

Procedure to move from specification limits to probabilities in Minitab

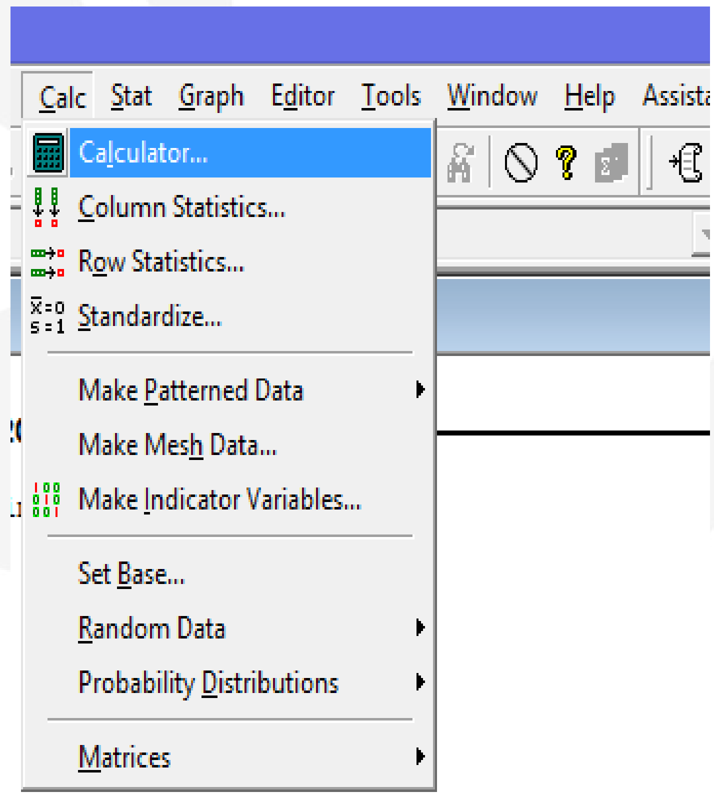
1. Open a new table in Minitab.
2. Name the column and type in the spec limits.
3. Create a new column, title it, select 'Formula' and input the formula for Z score (normal distributions only!).
4. Create a new column, title it, select Calculator > Probability > Normal to do the normal probability lookup.

Probability

Open a new table in Minitab, create two columns, and enter the Spec Limits.

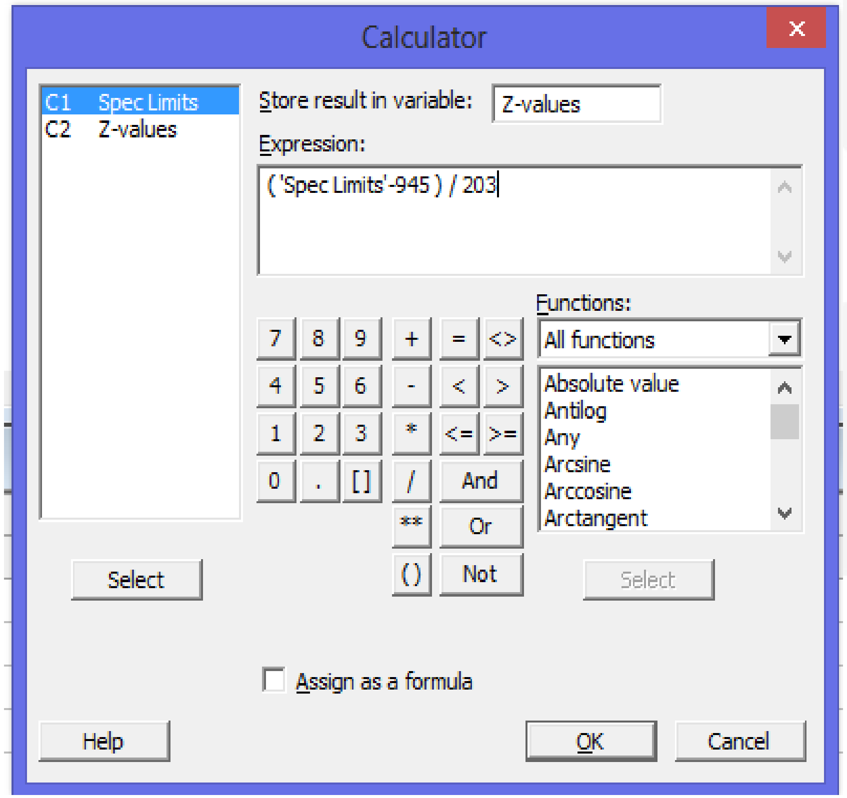


Open the calculator functionality from the toolbar

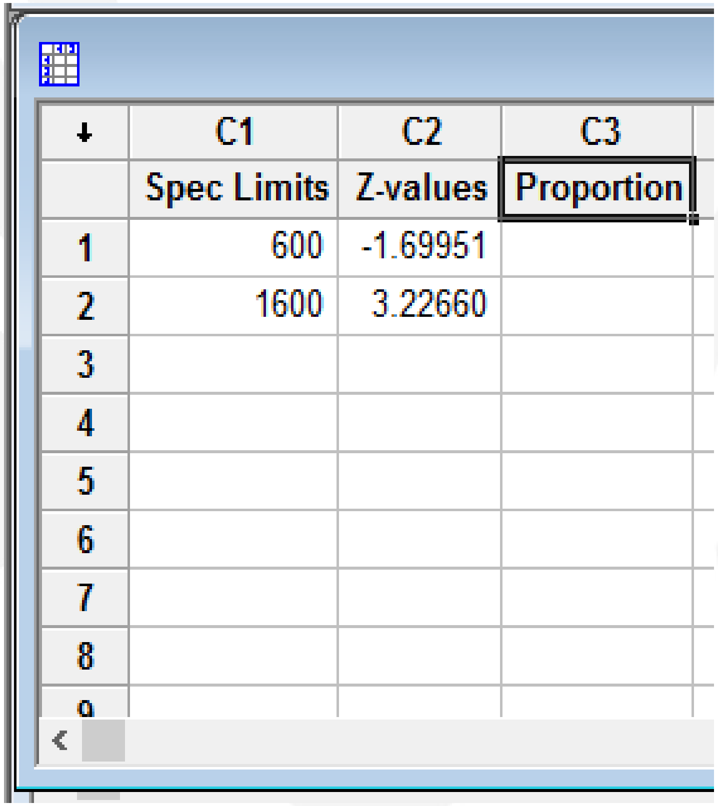


Probability

Type in the Z score formula for your problem, click OK.



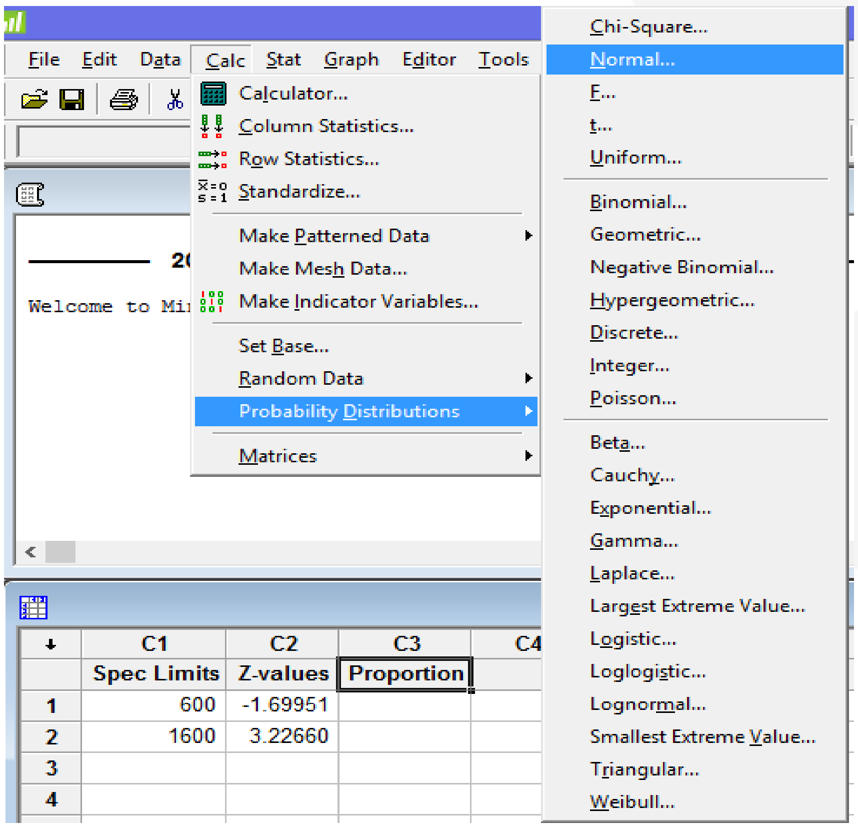
Create another column for the proportion

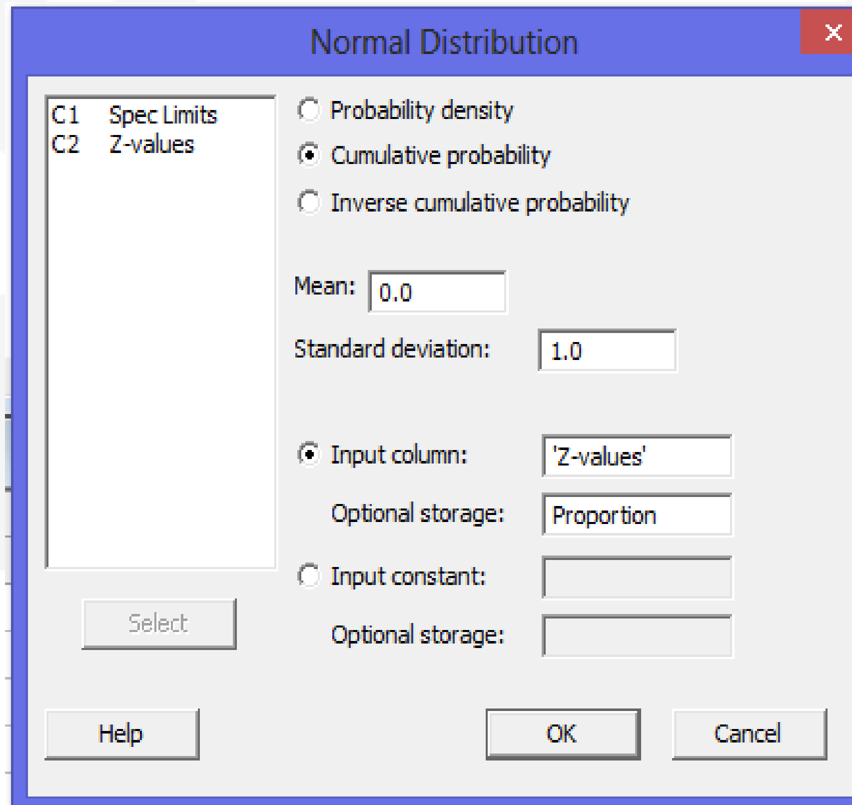


Probability

Select: Calc > Probability distributions > Normal

Select to create the Cumulative probability in the proportion column





Probability

↓	C1	C2	C3
	Spec Limits	Z-values	Proportion
1	600	-1.69951	0.044612
2	1600	3.22660	0.999374
3			
4			
5			
6			
7			
8			

Minitab now displays the cumulative probability that the process will have a value below the calculated Z score. To get the probability for above spec, subtract the cumulative probability from 1.

Normal Distribution

Objective: Practice use of Minitab for Proportions

A. Our teller drawers are supposed to have £1200 on average. If it **goes to £1600**, the teller should send some to the vault, or if it **drops to £600**, they should get currency from the vault. We take some measurements and find that the distribution is **normal** with an **average of £945** and **a standard deviation of £203**.

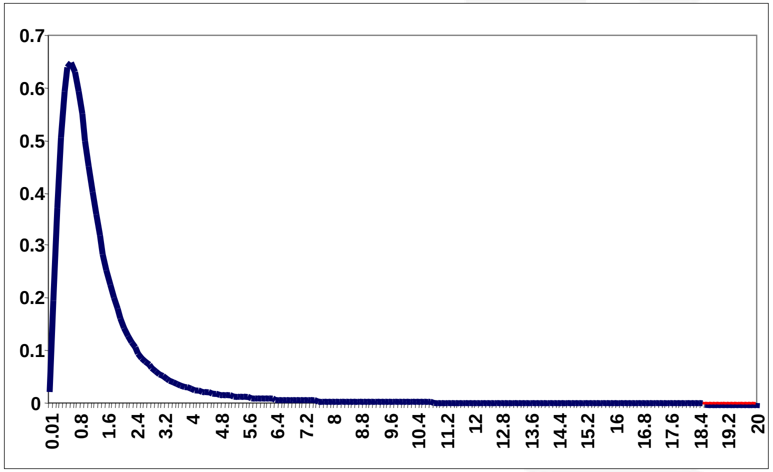
- What is probability that a drawer is out of "spec" at any given time?

25 minutes

B. A sales manager randomly samples the average weekly sales of the sales force. The data form a **normal** distribution with an **average of £6,852** and a **standard deviation of £986**.

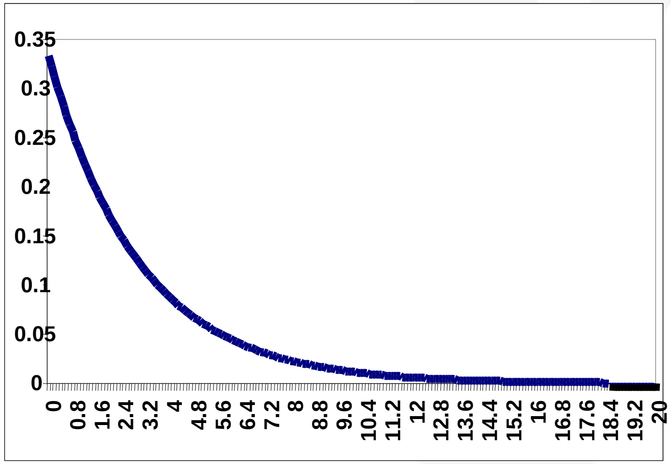
- What percent of the sales force is selling over £8,000 during any given week?
- In a group of 1,500 sales people, how many will sell over £10,000?

Lognormal Distribution



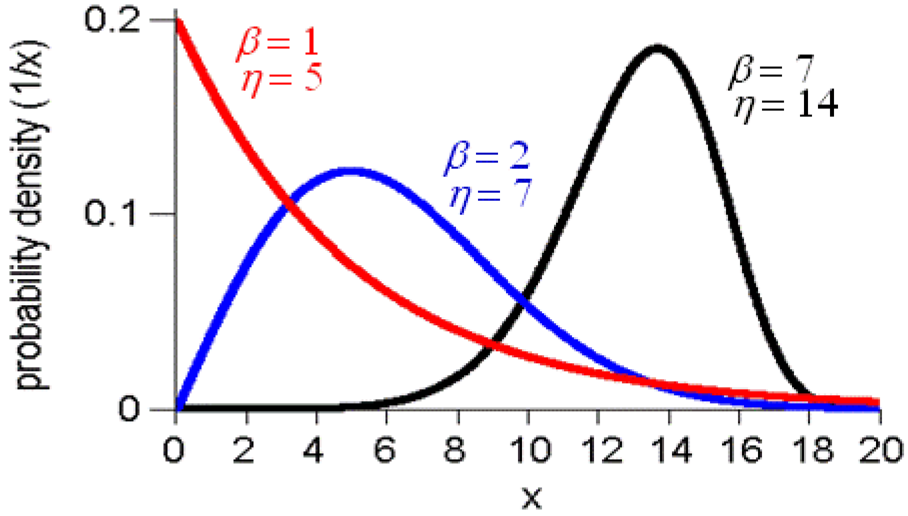
- The logarithms of the X values are normally distributed; no negative Xs.
- Some uses include:
  - **Waiting times**.
  - Distributions of wealth or salaries.

Exponential Distribution



- Commonly seen in processes with any kind of **scheduling or queuing**
- Mean equals the standard deviation.
- Maximum at zero and steadily decreases and approaches zero as X goes to infinity; no negative Xs.
- Some uses include:
  - Time between arrivals.
  - Time between failures (MTBF).

Weibull Distribution



- The Weibull is a family of distributions whose form varies greatly depending on its shape (*b*) and scale (*h*) parameters
- It is a "catch-all" distribution which should be used to describe data that has a **single peak but is skewed** (thus non-normal)
- **Time studies** very often result in a Weibull distribution.

Exercise 3 - Fitting Distributions

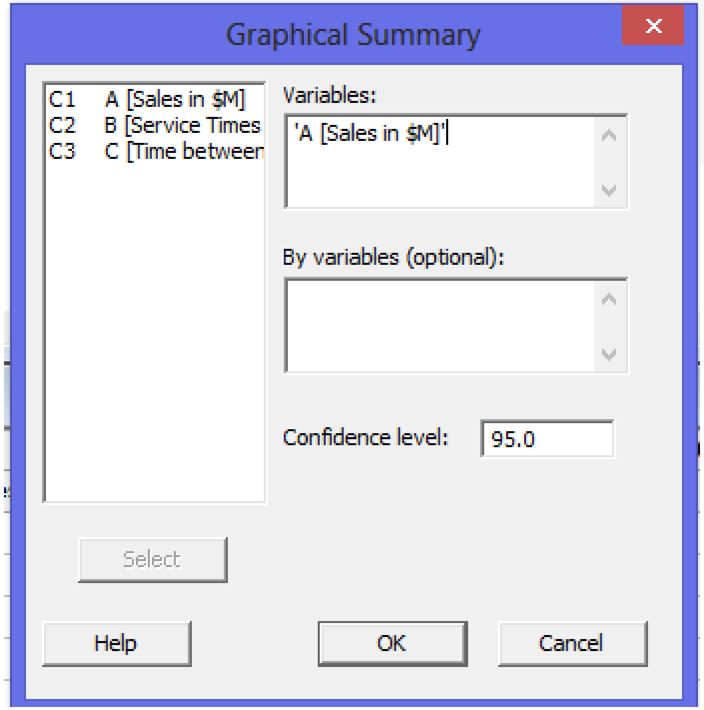
- Minitab file Distr_Ex_3.MPJ has three columns of data Fit the appropriate distribution for A, B and C.

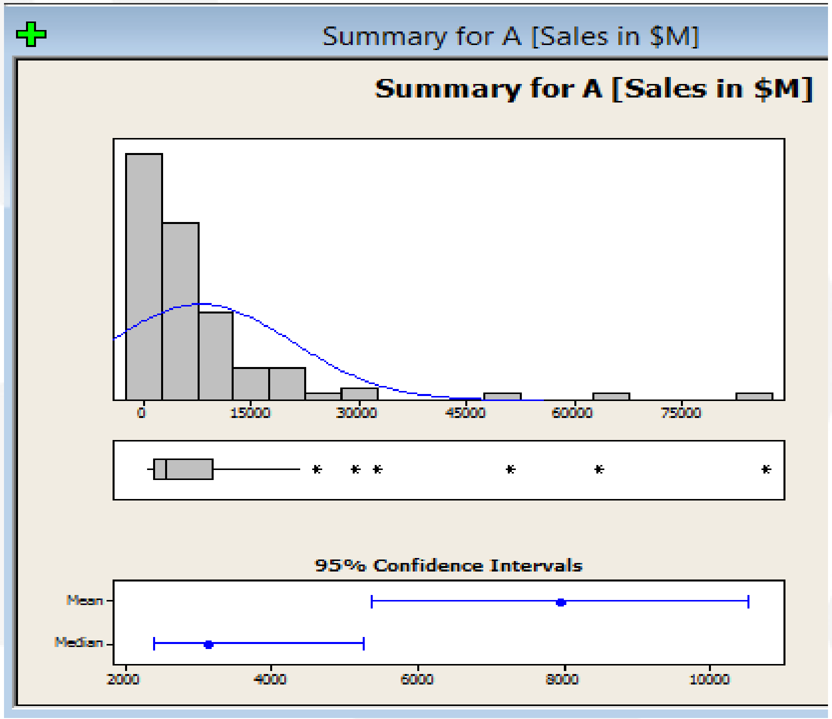
- Once you have found the appropriate distributions, for each of the distributions, determine:

  - A: What proportion of the sales figures are over $5,000?
  - B: What proportion of service times are over 4 minutes?
  - C: What is the probability that we can go for 2½ hours without an outage?

Exercise 3A - Histogram

Stats > Basic Stats > Graphical Summary

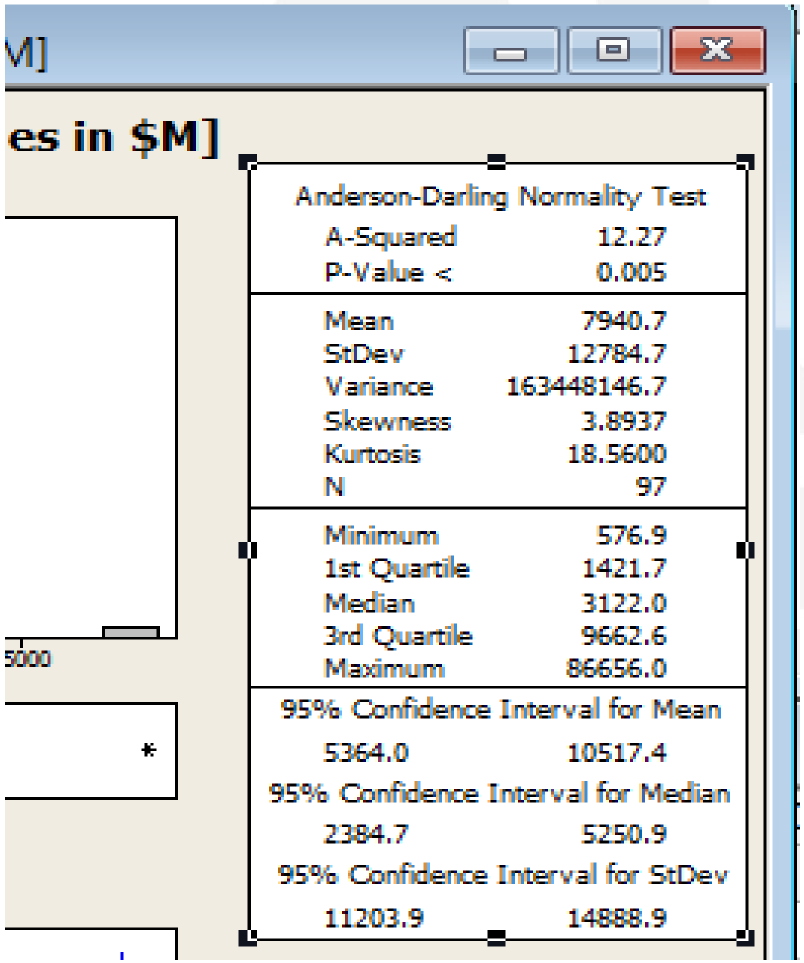




Does it look normally distributed?

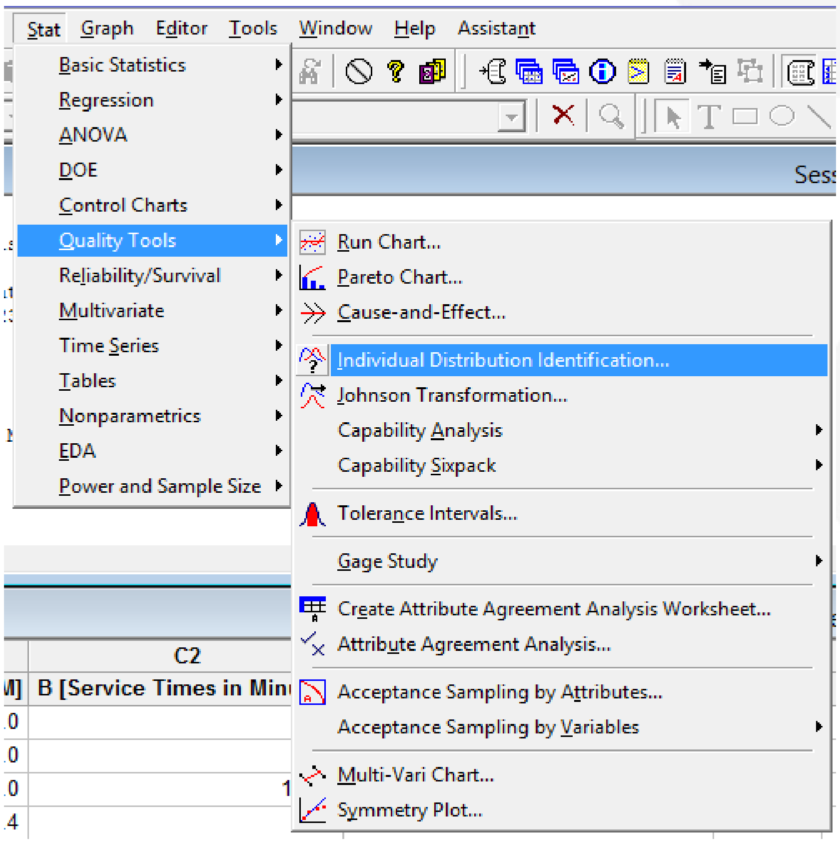
Exercise 3A - Basic Statistics

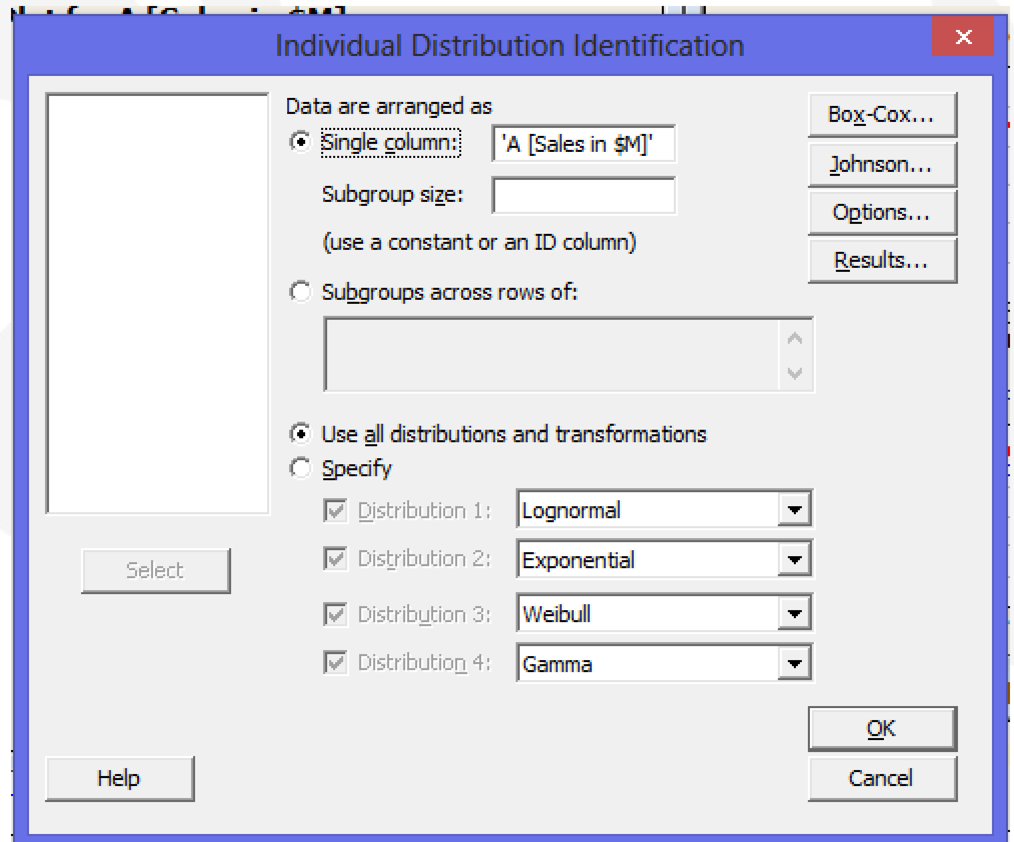
Review the basic statistics.



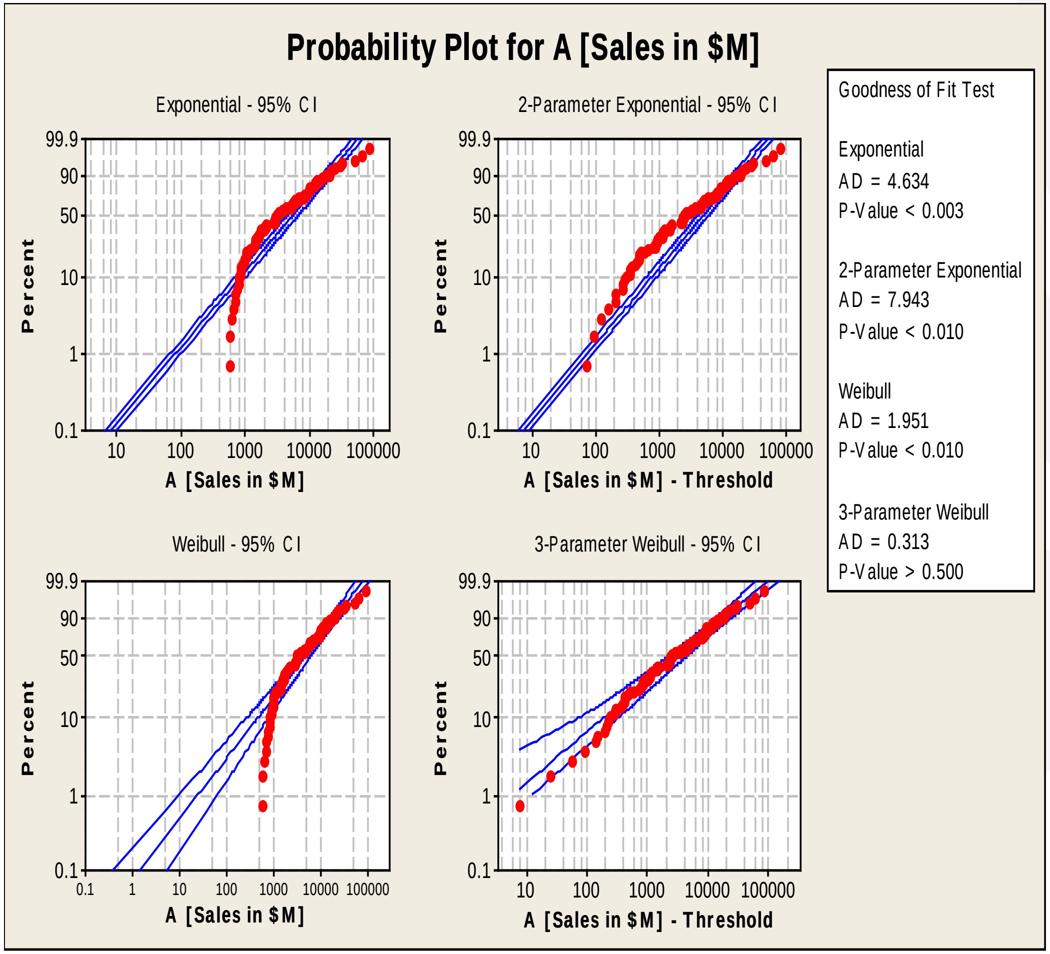
Exercise 3A - Distribution Fitting

Besides the normal distribution, Minitab can help you assess the distribution of your data against specific other distributions, allow you to compare four or run your data against all of its distributions and transformations.





Exercise 3A - Distribution Fitting



Having assessed all Minitab distributions, the 3 parameter Weibull is a good fit because:

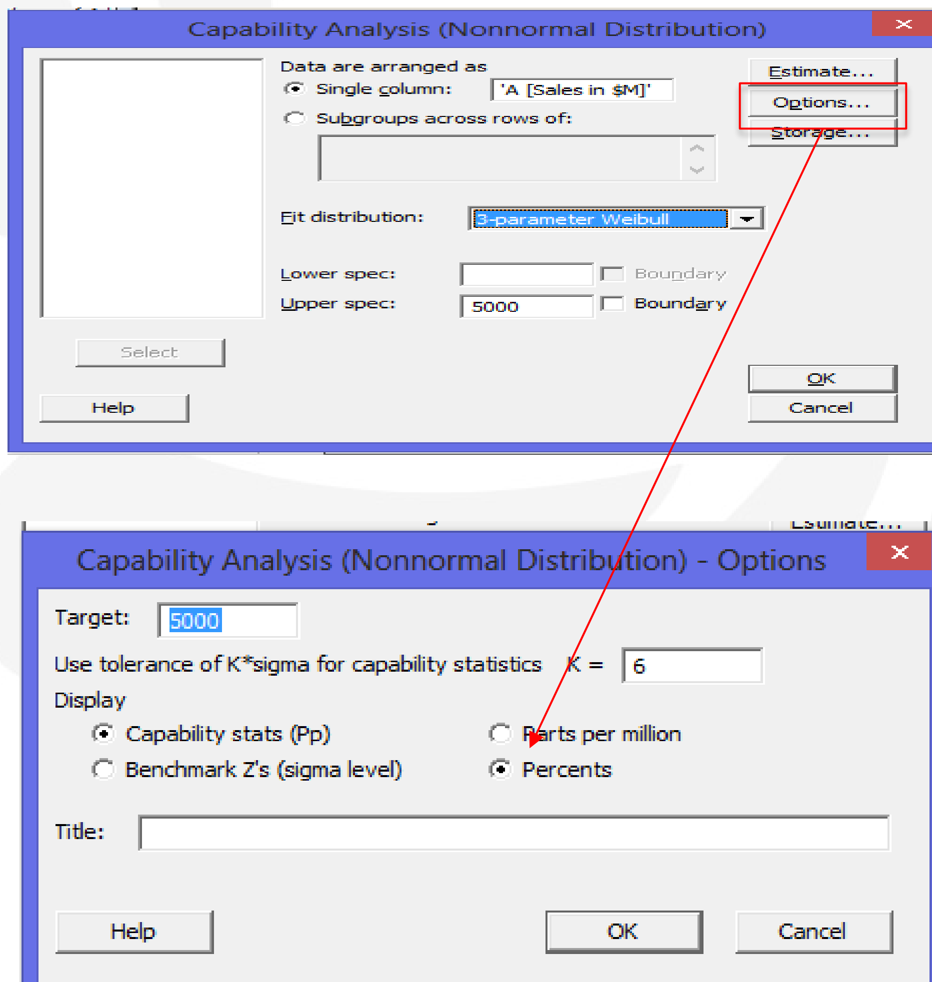
- The data points roughly follow a straight line
- The p-value was the highest of all available options

Calculating proportions

Now it's time to answer the business question: What proportion of sales are over $5,000?

Select:

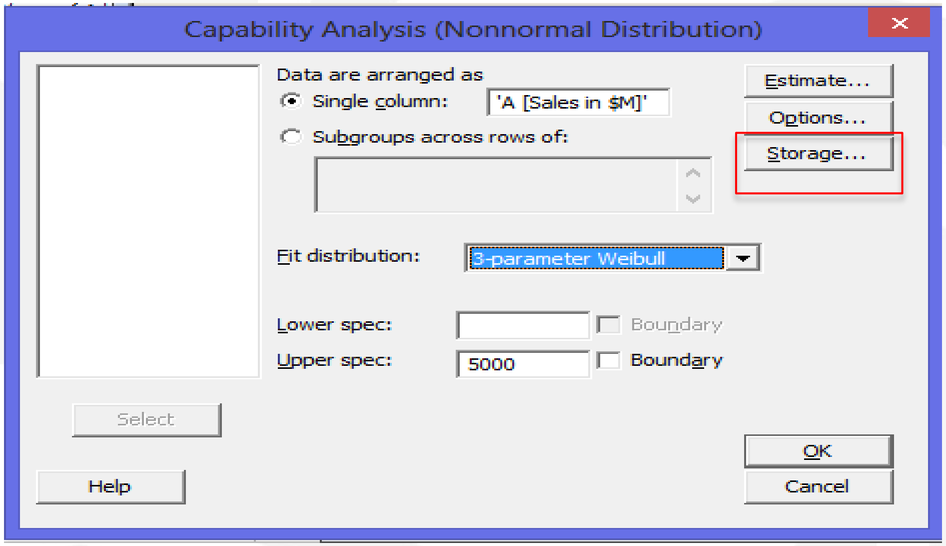
Stat> Quality Tools > Capability Analysis > Non-normal…



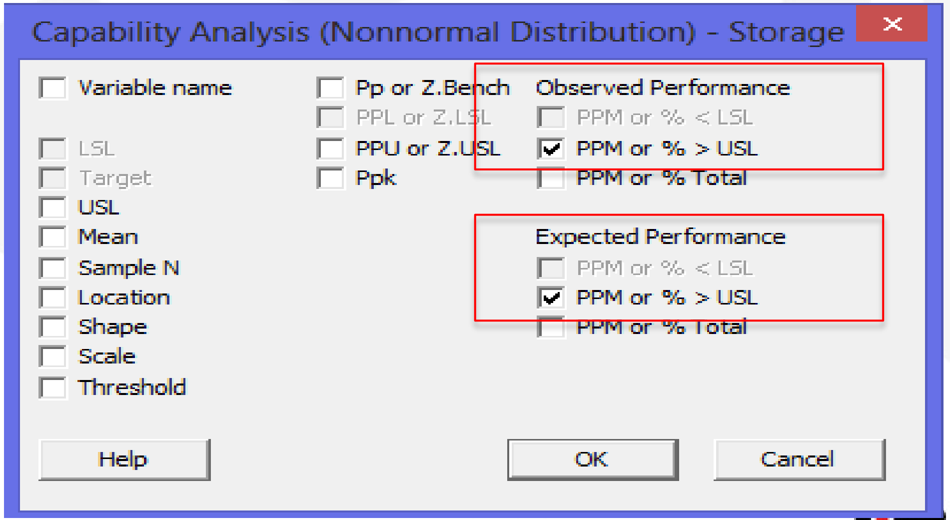
Enter the column to be assessed, Distribution and Specification

Calculating proportions

To display the proportion above the Upper Specification Limit, we need to select this within storage



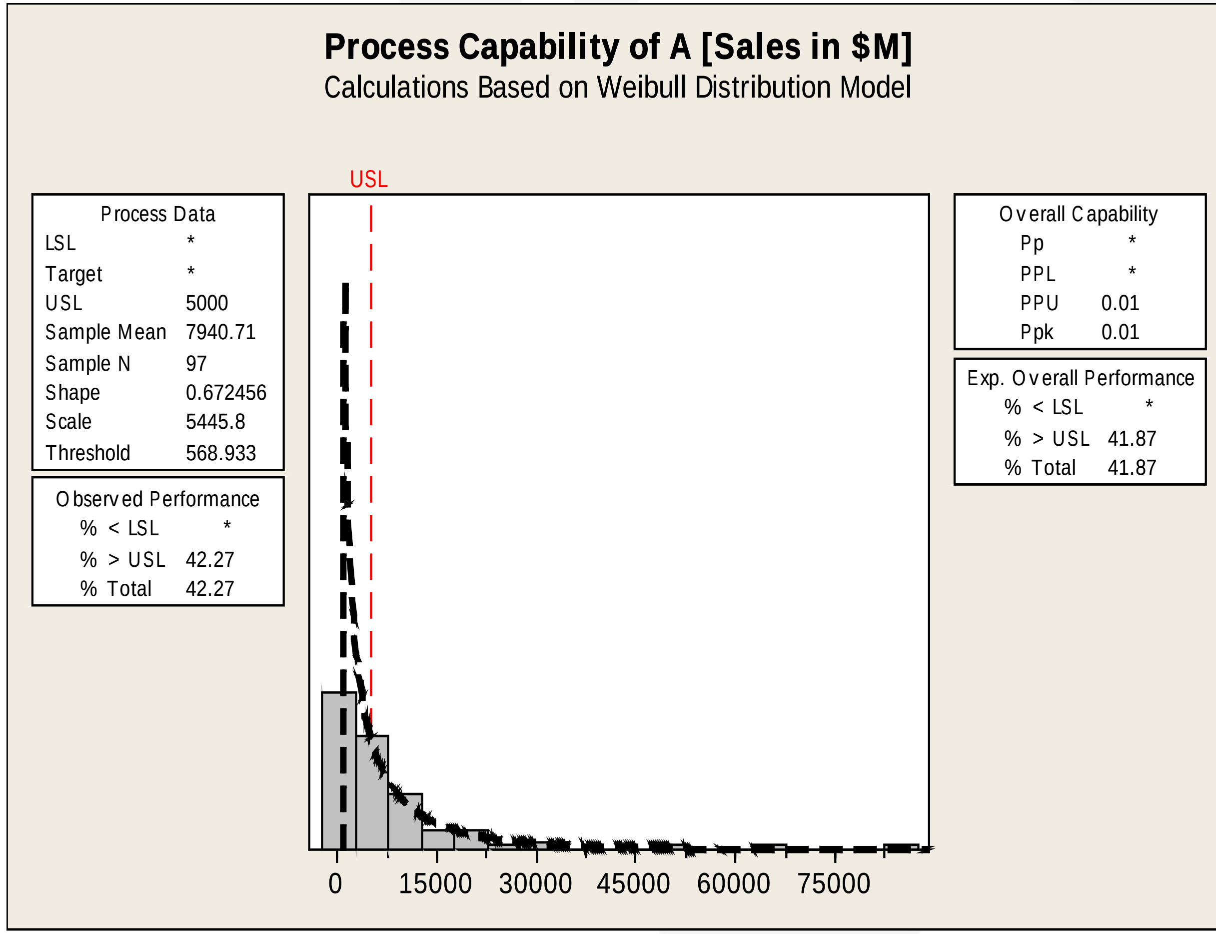
Select Observed and Expected performance greater than the Upper Specification Limit (£5,000)



Calculating proportions

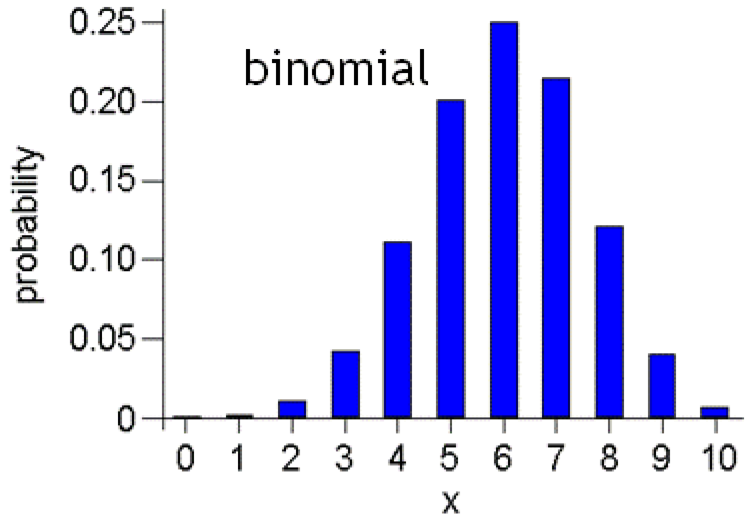
- Based on the Weibull distribution (our best fit) the observed performance within the data collected was 42.27% of sales were greater than £5,000

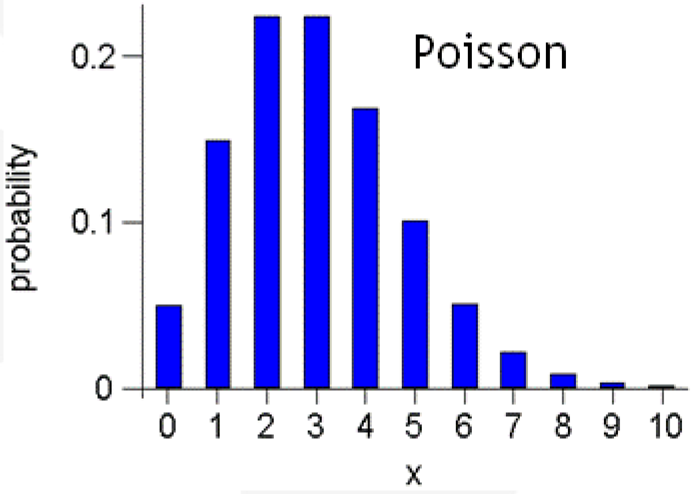
- If the sample is truly representative and if the population really follows this distribution, in the long term we would expect only 41.87% to be over £5,000



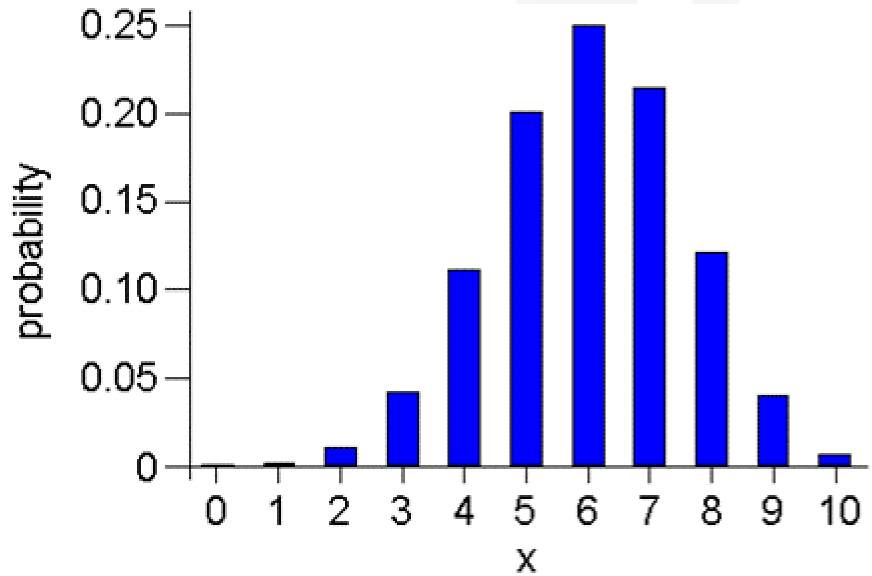
Now you have a go at parts B and C using the same data set (20 minutes)

Discrete Distributions





Binomial Distribution



Assumptions:

- Number of trials are fixed in advance; e.g. 10 sales calls.
- Just 2 outcomes for each trial; e.g. sale or no sale.
- Trials are independent; e.g. making a sale on this attempt does not influence whether a sale is made on the next.
- Probability of an outcome does not change from trial to trial; e.g. same sales person, same pitch, same materials.

Binomial Distribution

Uses include…

- Estimating the probabilities of an outcome in any set of success or failure trials.
- Number of “top two box” ratings on a survey.
- Number of defective items in a same-size sample.

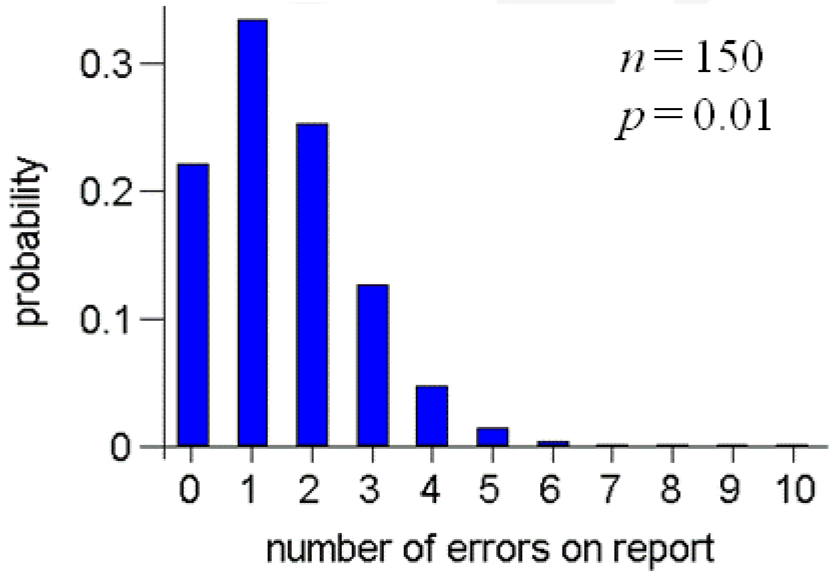
Exercise 4 - Binomial Probabilities

In a month-end report there is 1% probability of making an error in an expense figure. Every month 150 expense figures are presented in the report.

A. You would like to know the probability that there will be any error in a report.

B. An analyst would like to know the probability of 2 or more errors.

Exercise 4 - Binomial Probabilities

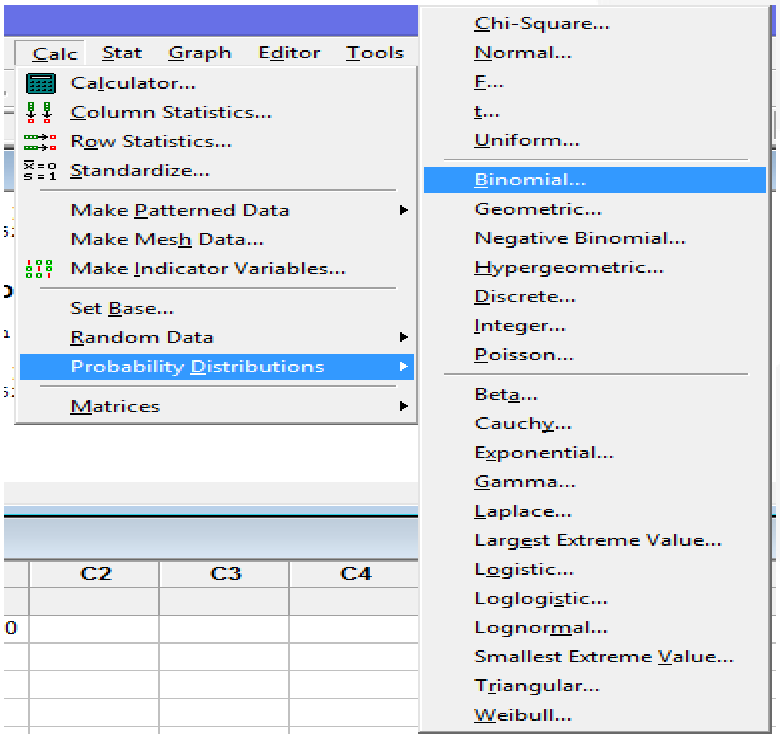


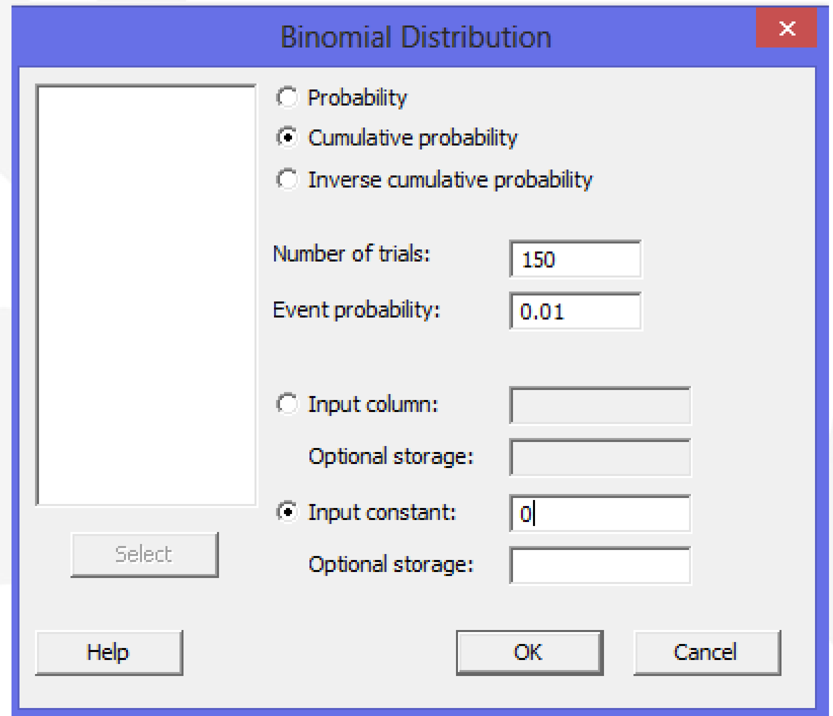
- Probability of no errors: $P(0) = 22.1\%$
- Probability of one or more errors: $P(x \geq 1) = 1 - P(0) = 77.9\%$
- Probability of exactly one error: $P(1) = 33.6\%$
- Probability of two or more errors: $P(x \geq 2) = 1 - [P(0) + P(1)] = 44.3\%$

Exercise 4A, Binomial Probabilities—Probability of a Specific Number

Number of trials, 150. For probability of an error, 0.01.

For the input constant enter the number of interest, in this case 0. Click OK.





Why do we want the probability of 0?

Exercise 4A, Binomial Probabilities—Probability of a Specific Number

Cumulative Distribution Function

```
Binomial with n = 150 and p = 0.01

x   P( X <= x )
0      0.221452
```

This shows the probability of getting zero errors is 22.1%, so the probability of any error is 77.9%.

Theoretically the probability of getting 0 or less than 0 errors!

Exercise 4B, Probability of a Specific Number and Below

Binomial Distribution

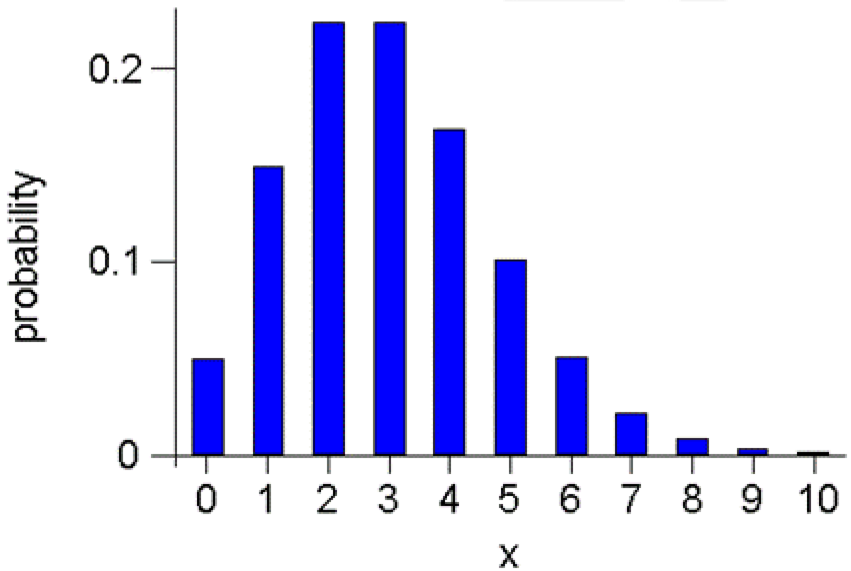
Objective: Practice use of Minitab for Proportions

Try working 4B in the same manner as the last one, but this time select “Binomial Distribution” and enter 1 as the number of interest.

Minitab will return the probability of getting 1 or below. So what is the probability of getting 2 or more?

5 minutes

Poisson Distribution



- Assumptions:
  - Length of the observation period (or area) is fixed in advance; e.g. one day.
  - Events occur at a constant average rate (called lambda); e.g. on average there are 23 ATM customers at this ATM per hour.
  - Occurrences are independent; e.g. one customer at the ATM does not encourage or discourage the next customer.

Poisson Distribution

- Uses include…
  - Number of events in an interval of time (or area) when the events are occurring at a constant rate.
  - Number of customer arrivals at a Banking Centre.
  - Design reliability tests where the failure rate is considered to be constant as a function of usage.

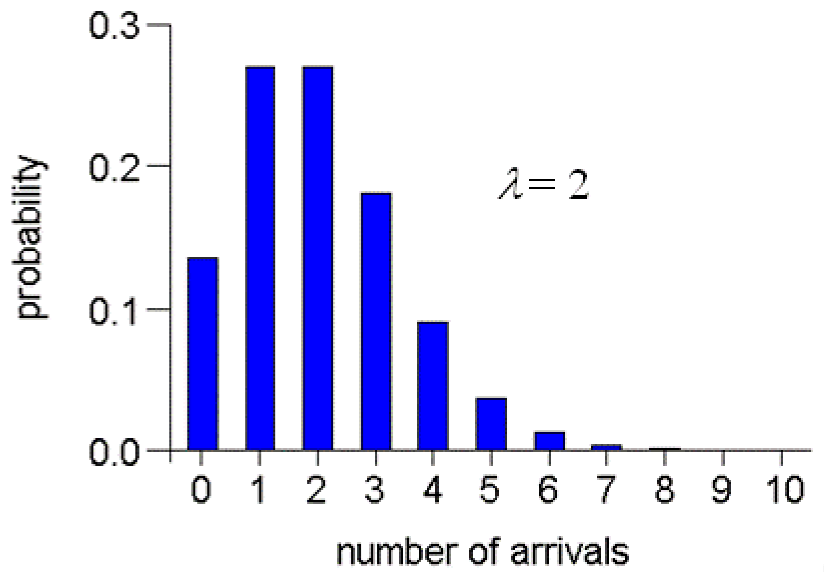
Exercise 5 - Poisson Probabilities

At a hotel reception desk, guests arrive at an average rate of 2 per minute during the noon hour.

The hotel manager needs your help in determining staffing level for this hour. She wants to know:

What is the probability that there will be 4 or more arrivals between 12:00 and 12:01?

Exercise 5 - Poisson Probabilities



Probability of 4 or more arrivals:

$$
\begin{aligned}
P(x \geq 4) &= 1 - P(x < 4) \\
&= 1 - \left[P(0) + P(1) + P(2) + P(3)\right] \\
&= 1 - \left[0.135 + 0.270 + 0.270 + 0.180\right] \\
&= 14.3\%
\end{aligned}
$$

Exercise 5 - Poisson Probabilities

Poisson Distribution

Objective: Practice use of Minitab for Proportions

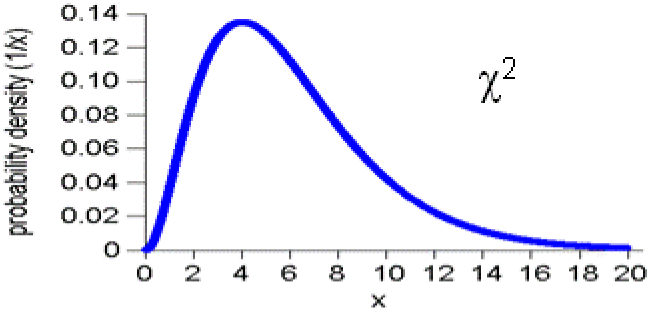
Try working this one in the same manner as the binomial probabilities exercise, but this time select “Poisson Distribution”, lambda is the term used to express the average, and enter 3 as the number of interest.

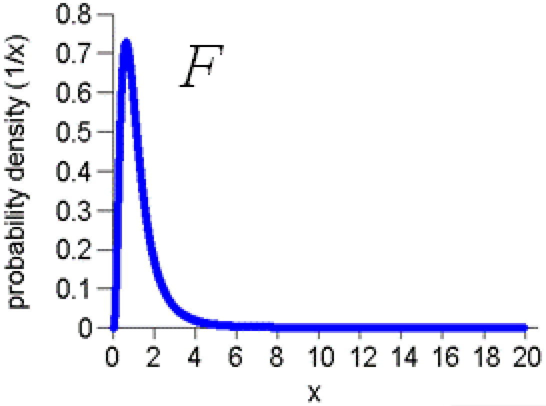
5 minutes

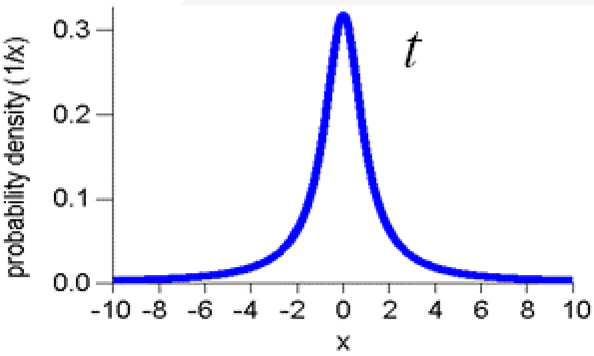
Minitab will return the probability of getting 3 or below.

So what is the probability of getting 4 or more?

χ^2, t, and F Distributions

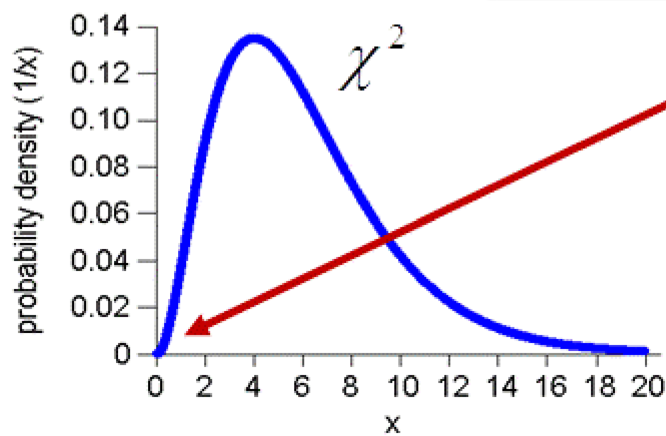






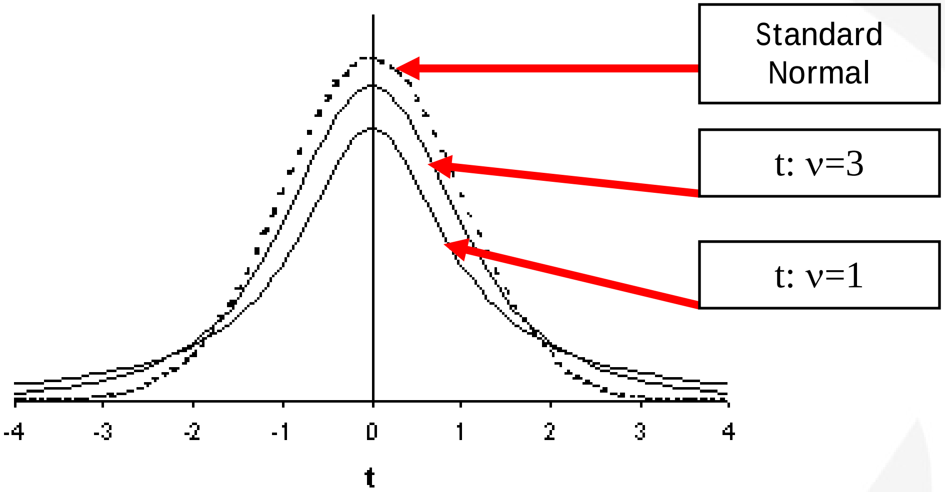
χ^2 Distribution

- The χ^2 distribution has many applications:
  - Looking for differences between what is observed and what is expected (contingency tables from GB training).
  - Judging how well a theoretical curve fits the data (goodness-of-fit test).
  - Determining if there is a relationship between factors.
  - Calculating confidence intervals for sample standard deviations.



Chi-square value is never negative!

t Distribution



- The *t* distributions were developed in 1908 by William Gosset who was a chemist and a statistician employed by the Guinness brewing company.

- Mean = 0, Variance = $\nu/(\nu-2)$ for $\nu>2$

  (ν denotes degrees of freedom)

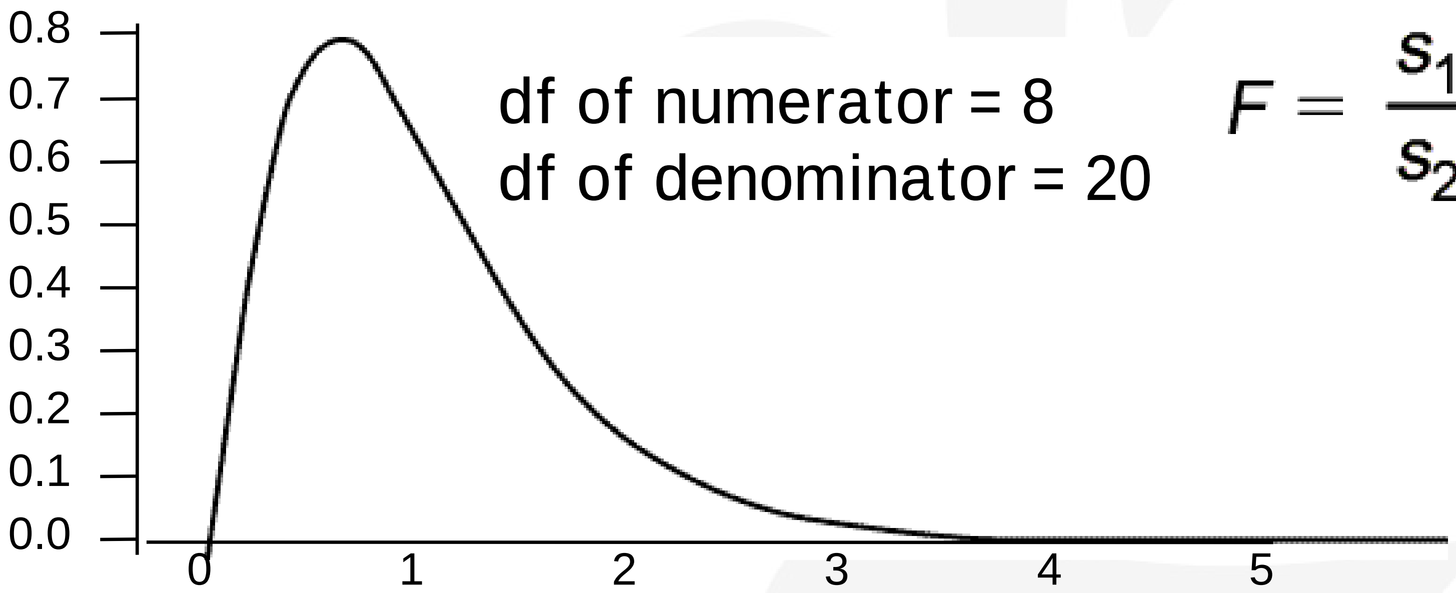
t Distribution

- The *t* distribution is used to describe small sample sizes taken from a normal distribution in which the population standard deviation *s* is unknown.
- The *t* distribution looks similar to the normal distribution but with wider tails.
- The wider tails reflect the greater uncertainty due to the small sample size and the unknown *s.*
- We will explore the *t* distribution in more depth in the hypothesis testing section of the BB materials.

A Note on Degrees of Freedom

- The degrees of freedom in any statistical analysis will vary depending on the data type and form of the analysis.
- In general, DF is a way of determining how many estimates we can make based on the amount of data we have.
  - The fewer the DF, the larger the variation/uncertainty in the estimates.
  - For every data observation we collect, we gain a DF we can use to make an estimate.
  - For every parameter we estimate, we must use a DF; which leaves fewer for the next estimate.
  - Examples: If I have one group of data, and want to estimate the variability, my DF = n-1; if I have two groups of data, the DF = $n_1 + n_2 - 2$

F Distribution



$$F = \frac{s_1^2}{s_2^2}$$

- Named after Sir Ronald A Fisher.
- He made important contributions in the analysis of variance, hypothesis testing, maximum likelihood methods, and the design of experiments.

F Distribution

- *F* distributions are used to determine differences in variances between two data sets (*F*-test).
- *F* distributions are also used in ANOVA hypothesis testing to calculate *p*-values.
- We will look at these usages in more detail in the hypothesis testing section of the BB materials.

Summary of key learning points

- Probability distribution is a rule that specifies the possible outcomes and gives the probability of observing each of these values.
- Distributions for discrete random variables look different from distributions for continuous random variables.
- Cumulative Distribution Function (CDF) denotes the area under the curve to the left of a value of interest.
- CDF a.k.a. F(x) can be calculated for discrete or continuous distributions.
- Probability distributions can be described using location, scale and shape parameters.
- Given a set of data, it is possible to fit distributions with Minitab.
- Minitab will give probabilities related to the various distributions.

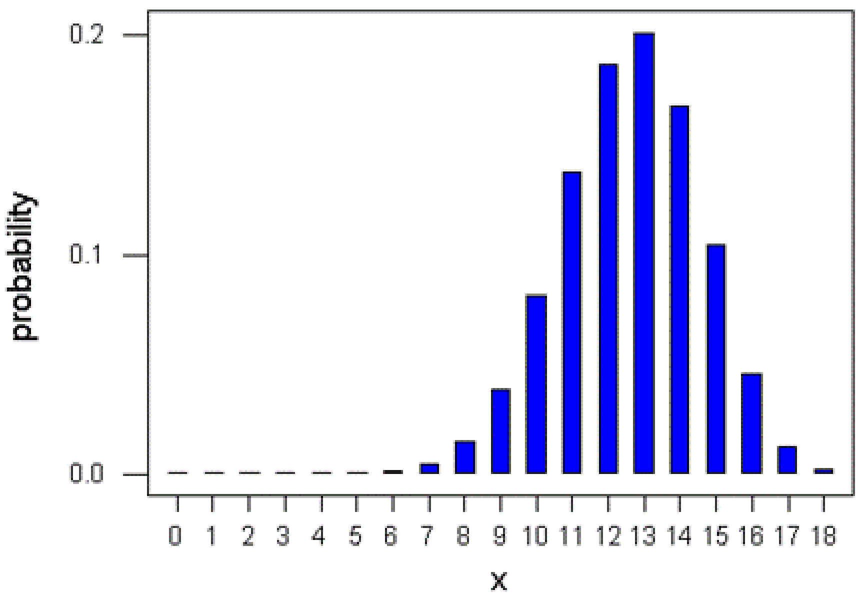
It’s time to play - “Name that Distribution”

CLOSE YOUR BOOKS.

- Class will be broken up into teams.
- A description of a distribution will be shown.
- The chosen team will try to identify the distribution collectively.
- If the answer is correct, the team is awarded one point.
- And now, the first question…

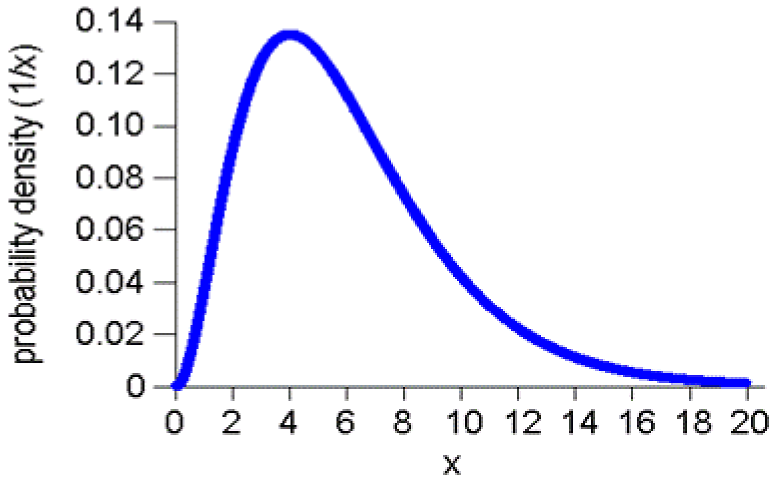
Name that Distribution (cont.)

1. I am useful for describing the probability of getting a number of particular outcomes out of a total number of trials when there are **two possibilities** (pass/fail, yes/no, etc.).



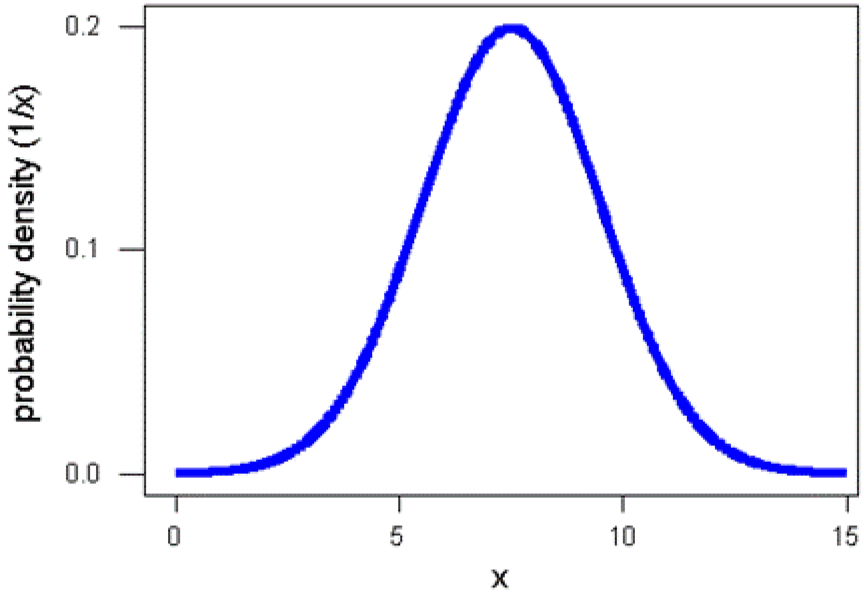
Name that Distribution (cont.)

2. I can be used to determine if an **observed result is significantly different from what is expected**.



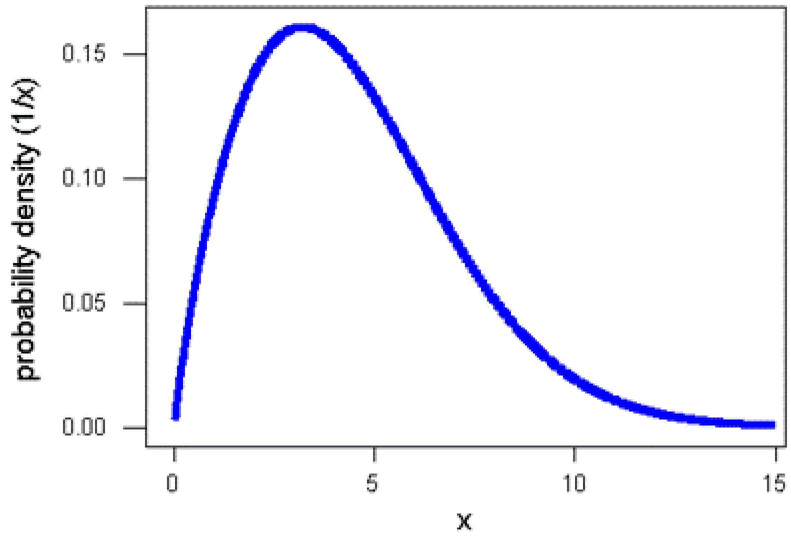
Name that Distribution (cont.)

3. I am the **most commonly** used distribution in describing **random variation about a central point**.



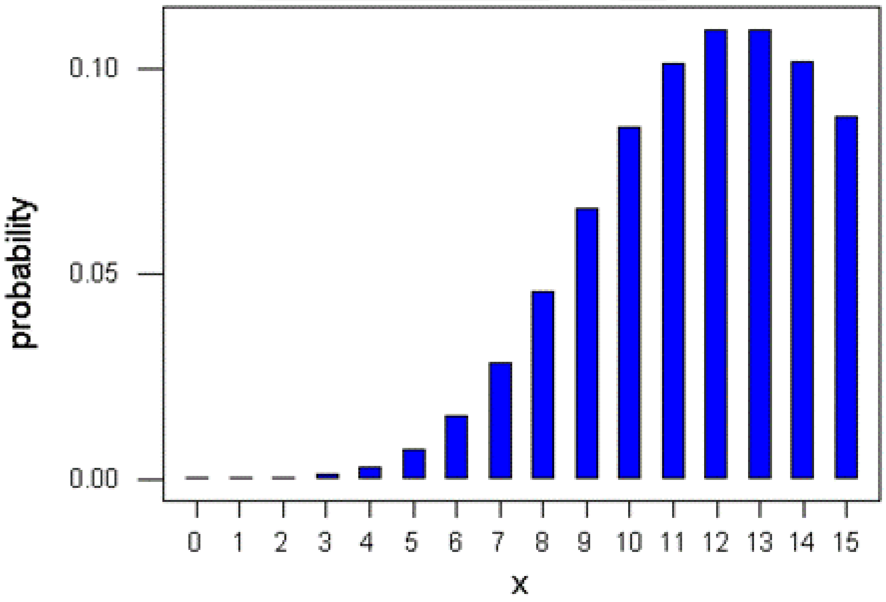
Name that Distribution (cont.)

4. I am a "catch-all" distribution which can be used to fit data which is **single-peaked but NOT normally distributed**.



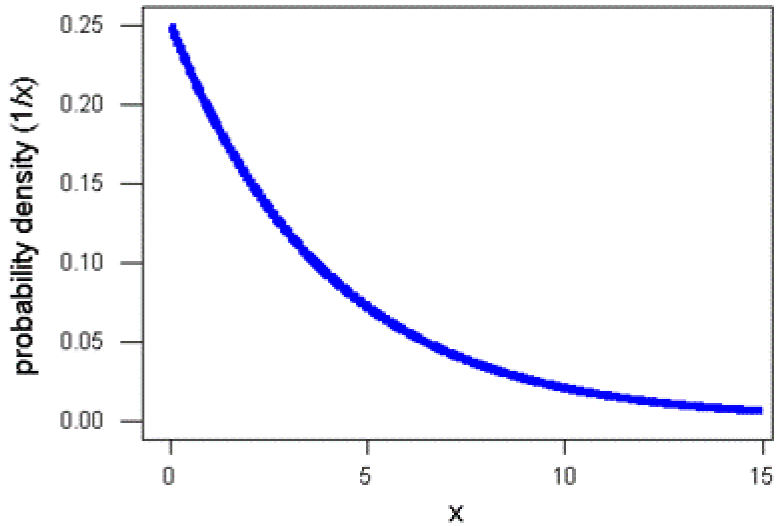
Name that Distribution (cont.)

5. I can be used to describe the **probability of a certain number of events occurring during a particular length of time** (e.g., sales per hour).



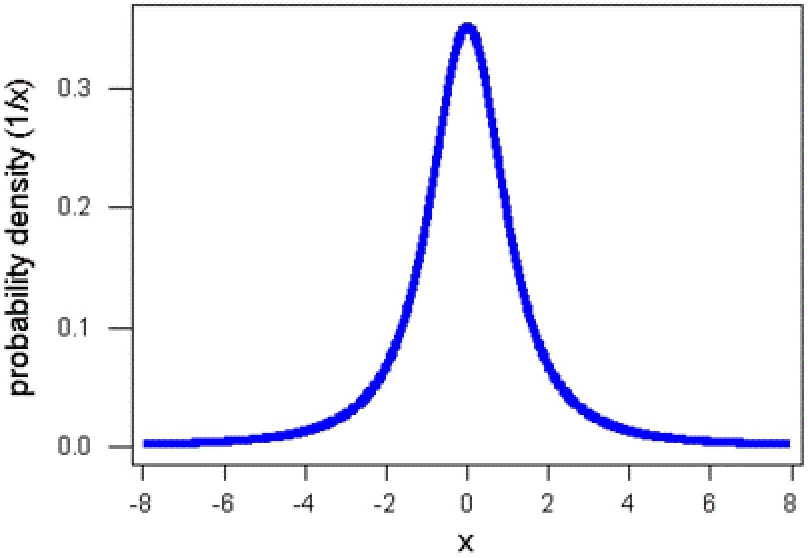
Name that Distribution (cont.)

6. I always have a probability shape which **decays constantly-**-I can be used to describe the probability of uninterrupted service if a mean-time-to-failure is known.



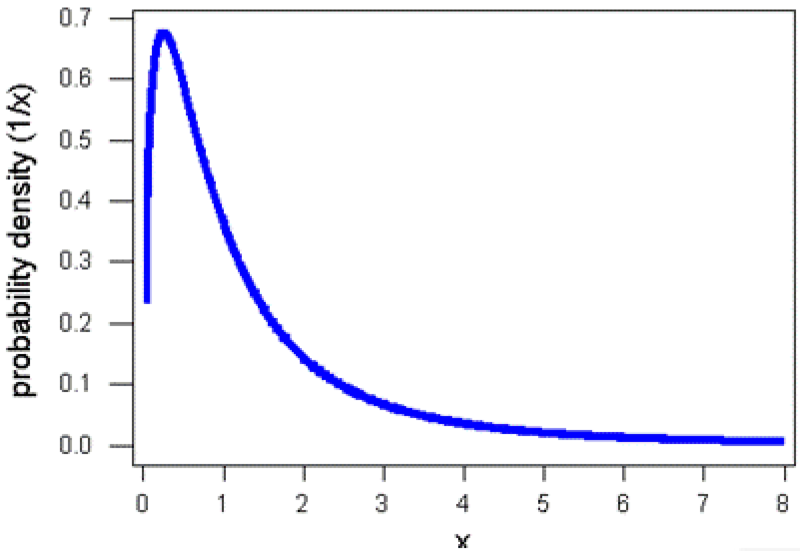
Name that Distribution (cont.)

7. I am used to describe data that is taken from a **normal population** whose **standard deviation s is unknown**.



Name that Distribution (cont.)

8. I can be used to **compare "between" vs. "within" variation** to determine the statistical significance of factors in an ANOVA.



Recommended coach support points

- Identification of your distribution
- Application of probabilities

Appendix – Standard Normal table

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9031	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Black Belt

Central Limit Theorem and Confidence Intervals

Objectives of this module

At the end of this module, you will be able to:

- Understand Central Limit Theorem (CLT) and why it is important to other statistical tools

- Understand and calculate confidence intervals

CLT Exercise 1

- Divide into four teams.
- Each team will be given a process that delivers an output.
- Run the process, measure and record the output until the facilitator says “stop”.
- Calculate the statistics of the process output:
  - Mean
  - Variance
  - Standard Deviation
  - P-value for Normality
  - Show histogram

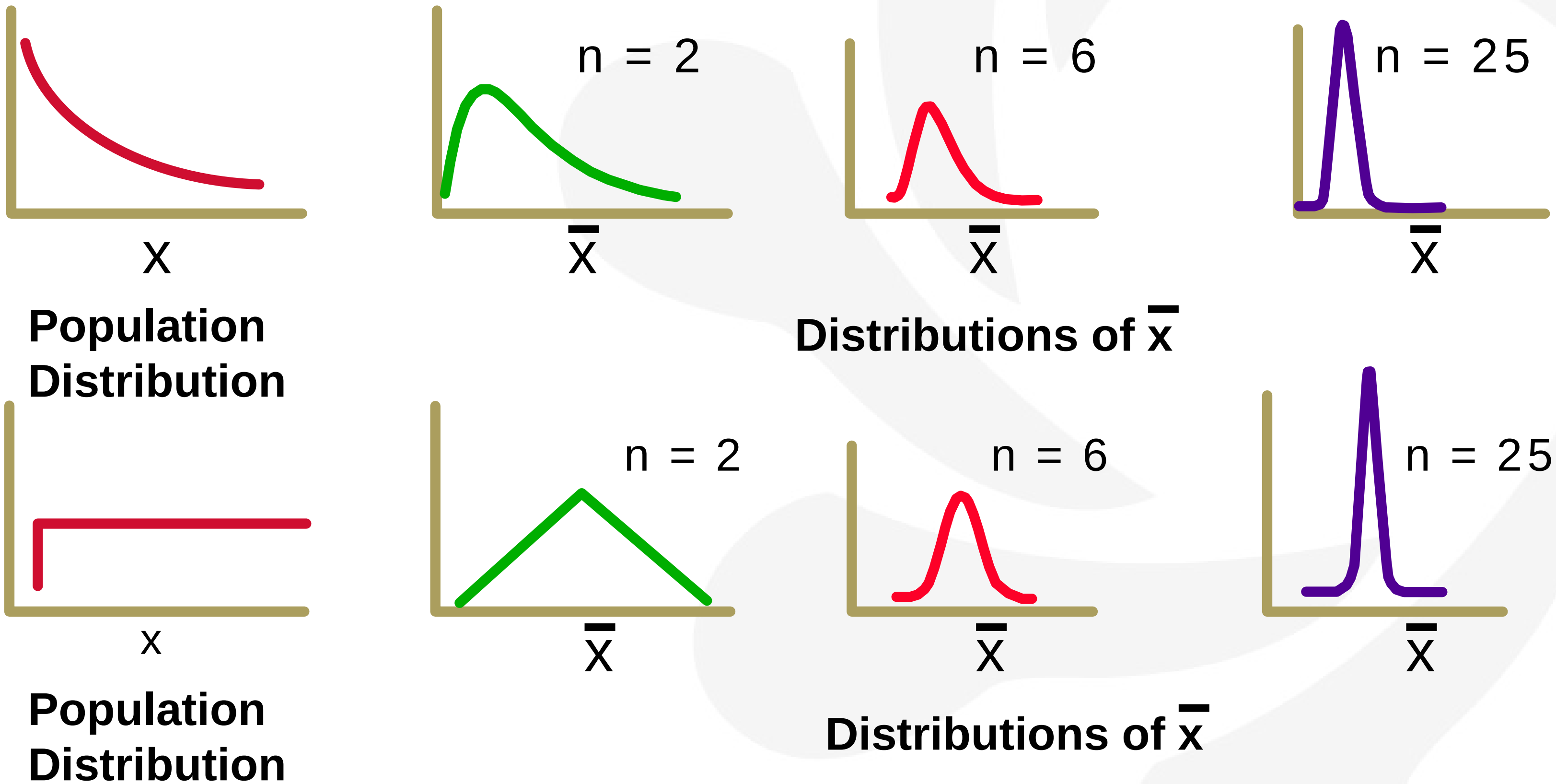
CLT Exercise 1

Team	Sample Size	Sample Mean	Sample Variance	Sample Std. Dev.	P-value for Normality
1					
2					
3					
4					

Central Limit Theorem

- Given a population with a mean of m and a variance of s^2, if we sample that population repeatedly using a sample size of n, and further plot a distribution of the means of those samples, then the following will be true:
  - The mean of the sampling distribution will be m.
  - The variance of the sampling distribution will be s^2 / n.
  - The shape of the sampling distribution will approach a normal distribution as n gets larger regardless of the shape of the original population.

CLT – Demonstrating Normal Sampling Distributions



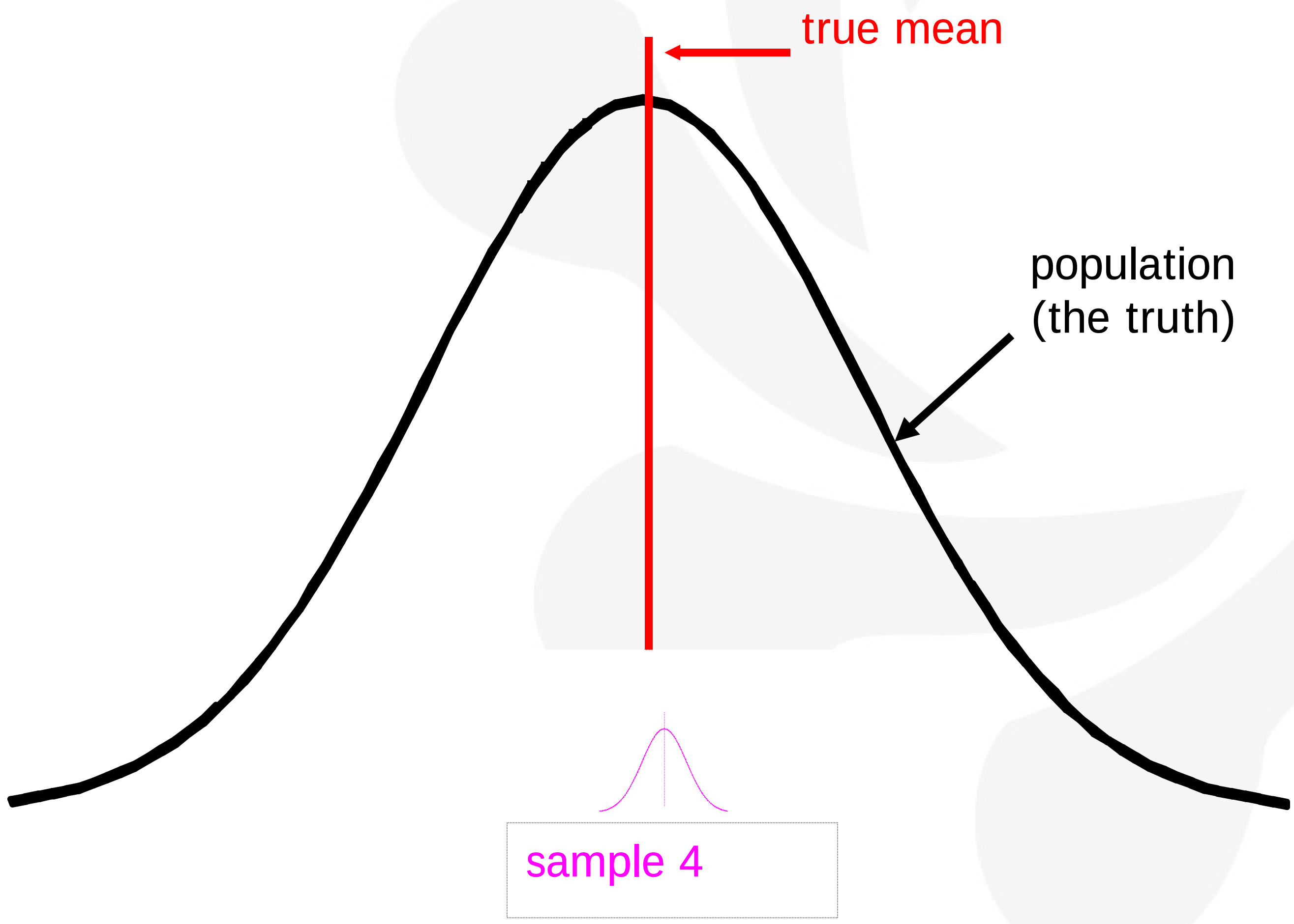
n = sample size used to calculate the x-bars that are plotted in the histograms.

Reference: *The Black Belt Memory Jogger*, p. 139ff.
Dice animation: http://www.stat.sc.edu/~west/javahtml/CLT.html
Simulation with various distributions: http://www.statisticalengineering.com/central_limit_theorem.htm

Applications of CLT

- How can we use the Theorem to our advantage?
  - A sample average is an estimate of the population mean.
  - If we wanted to estimate the population mean of a process, and we could only take one sample, what sample size would you choose: 1, 2, 5, or 10? Why?
  - How could we estimate how far away this sample average might be from the population mean?
  - To the 95% confidence level, how far off might the true mean (population mean) be from the sample mean in your dice exercise?

Confidence Interval



Questions Lead

- **Basic question:** **How sure am I that this is the true population mean (standard deviation, median, proportion)?**
  - **I know what my sample says, but how close is that to the truth about the population?**
  - **How far off might I be?**
  - **How confident can I be when I give my boss the answer?**

Confidence Intervals - Background

- Statistics such as the sample mean and standard deviation are only *estimates* of the population Mu (μ) and Sigma (σ) and are based upon a limited amount of data.
- Because there is *variability* in these estimates from sample to sample, we can quantify our uncertainty using **confidence intervals** based on the Central Limit Theorem.
- Confidence intervals provide a range of *plausible values* for the population parameters (μ and σ).

Confidence Interval Interpretations

- Most of the time, we calculate 95% confidence intervals (CIs) for a parameter (occasionally 90% or 99%).
- The CI is interpreted as:
  - We are 95% certain that our calculated interval surrounds the true population parameter (e.g**., μ , σ or η).**
  - In technical terms it would be more correct to say, the method we use for calculating the interval will yield correct results 95% of the time.

What is a Confidence Interval?

Usually, confidence intervals have a '±' uncertainty:

Estimate ± margin of error

Sample statistic ± [___ X ___]

e.g., $\bar{x}$, s, ...

Confidence factor

Measure of variability

In some cases the uncertainty is not symmetrical and the '+' term is different from the '-' term; e.g., for σ.

CI Exercise 2

- To see the variability in sample estimates, let's define a process that has a normal distribution with:
  - known (true) mean value = 70
  - known (true) standard deviation = 5
- Each member in the class will generate 20 observations from this process, with mean = 70 and standard deviation = 5

In Minitab: Calc > Random Data> Normal...

Use Stat > Basic Stat > Graphical Summary for your column to calculate the 95% confidence interval for the mean and standard deviation based on your sample of 20 data points. Do they contain the true mean 70 and the true standard deviation of 5?

- How many of our confidence intervals would we expect to NOT contain the population parameter?

In a class of 20 students, how many should I expect to call out that the true mean (70) does not lie in their confidence interval?

Luck of the Draw

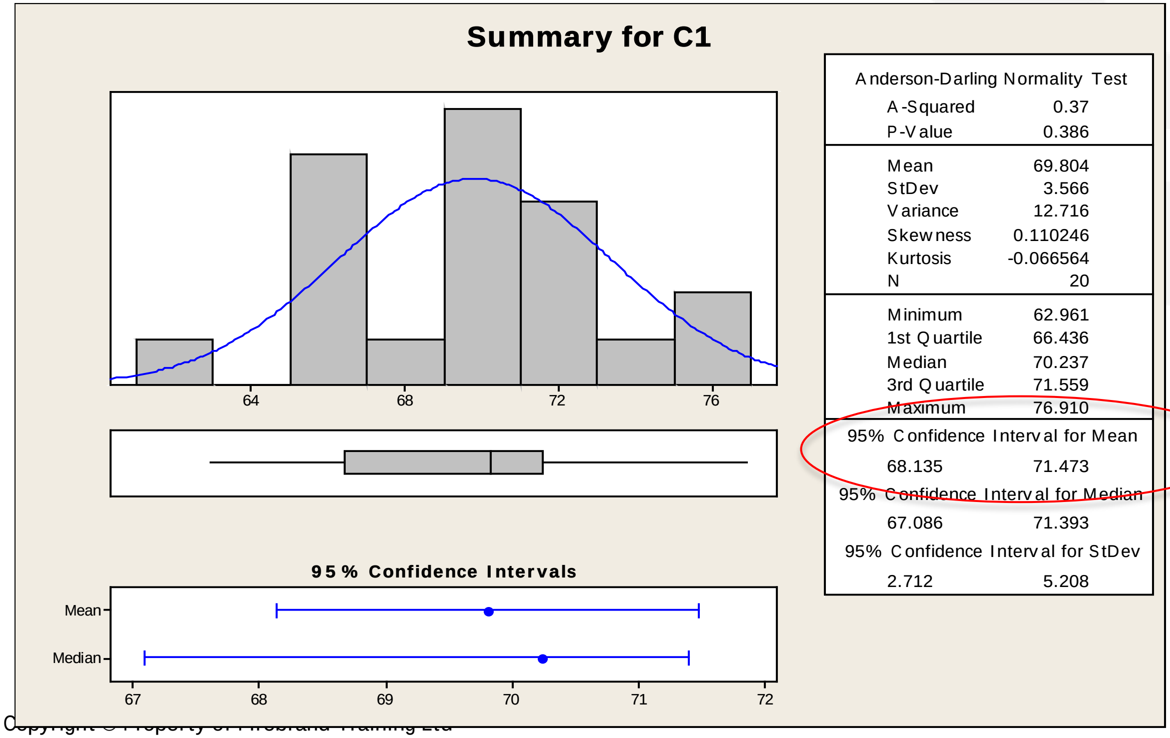
We want to be certain that our confidence interval contains the population parameters. But certainty only comes with measuring the entire population. Therefore, for the vast majority of cases, we have to live with being 95% certain that our sample has captured the population parameters inside the confidence intervals. We say that we are 95% confident.

In reality we will never know whether our sample was one of the lucky 95% that actually contained the true parameter, or one of the unlucky 5% that did not.

CI's from Minitab—Continuous Data

- **When Minitab has raw continuous data, it is simple to calculate the confidence intervals around the population mean and standard deviation like we did in Exercise 2.**

Stat > Basic Statistics > Graphical summary



CIs Without Raw Data

- **Without raw data, Minitab cannot calculate CIs.**
- **For continuous data, given the mean, standard deviation, sample size and confidence level, the formulas are pretty straight forward.**

CI for Mean of Continuous Data (s known or > 50 Samples)

$$CI = \overline{X} \pm Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$Z_{\alpha/2}$ = normal distribution value for a given confidence level

$\overline{X}$ = mean of data

σ = population standard deviation

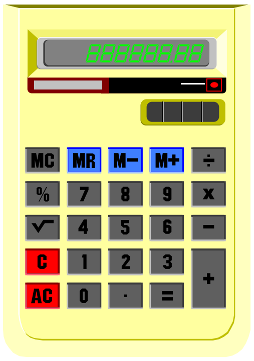
n = sample size

WARNING!!!!

This formula only applies when σ is known, which is rare. If the sample size is large (exceeds 50), it is a good approximation.

Example

- **Mean = 2.15**
- **Standard Deviation = 0.8**
- **Sample Size = 55**
- **α = 0.05**



- **What is the confidence interval of the mean for this situation?**

CI for Mean of Continuous Data - Solution

$\overline{X}$ **= 2.15** $\quad \alpha$ **= 0.05**

σ **= 0.8** $\quad$ **n = 55**

$$\mathbf{CI = 2.15 \pm Z_{0.05/2} \frac{0.8}{\sqrt{55}}} = 2.15 \pm 1.96 \frac{0.8}{\sqrt{55}} = 2.15 \pm 0.211$$

CI for Mean of Continuous Data (σ unknown or <50 Samples*)

$$CI = \overline{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$$

***This formula has no restrictions on sample size. It is based on the Student's t Distribution.**

This t value comes from the t table using the column for the alpha risk divided by 2 (risk divided between each tail) and the row for the degrees of freedom, n-1.

$\overline{x}$ = mean

s = standard deviation

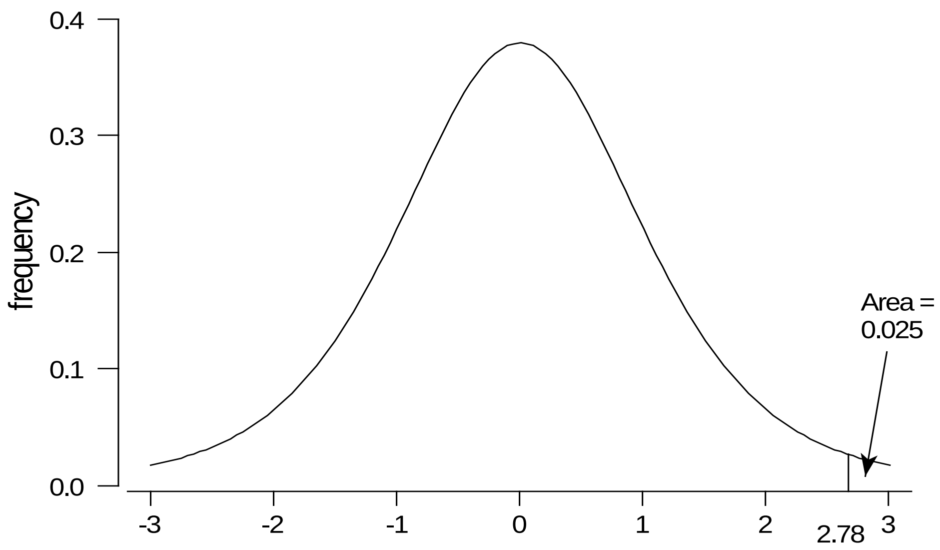
n = sample size

$t_{\alpha/2,n-1}$ = value from t distribution table

Some t tables use υ, lower case Greek letter nu to mean degrees of freedom.

What is this t-distribution?

- The t-distributions comprise a family of distributions with one extra parameter (“degrees of freedom” where df = sample size -1).
- They are similar in shape to the normal distribution (symmetric and bell-shaped), although wider, and flatter in the tails.
- Used for estimating population parameters when the sample size is small (<50).
- The smaller the sample size, the flatter the distribution tails.



t-distribution
with 4 df (n=5)

Selected t-values

- Here are values from the t-distribution for various sample sizes (for 95% confidence intervals):

Sample Size	t-value ($\alpha/2$ =.025)
2	12.71
3	4.30
5	2.78
10	2.26
20	2.09
30	2.05
100	1.98
1000	1.96

Why is a sample size of 1 not in the table?

As the sample size increases, what happens to the t-value?

Example – Small Sample Size

Same example as before, but what if only 25 samples instead of 55:

$$\overline{X} = 2.15 \quad \alpha = 0.05$$

$$s = 0.8 \quad n = 25$$

$$CI = 2.15 \pm t_{0.05/2,24} \frac{0.8}{\sqrt{25}} = 2.15 \pm 2.064 \frac{0.8}{\sqrt{25}} = 2.15 \pm 0.330$$

Another Example

- **Mean = 15.82**
- **Standard Deviation = 6.54**
- **Sample Size = 30**
- **α = 0.01**

- **What is the confidence interval of the mean for this situation?**

CI's for Variance of Continuous Data

$$\frac{(n-1)s^2}{\chi^2_{(\alpha/2),(n-1)}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-[\alpha/2]),(n-1)}}$$

The variance CI is based on the χ^2 Distribution. Since it is based on this distribution, the CI will not be symmetrical!

Where…

n = sample size

s^2 = variance

α = risk

$\chi^2_{(\alpha/2),(n-1)} = \chi^2$ **lookup value**

$\chi^2_{(1-[\alpha/2]),(n-1)} = \chi^2$ **lookup value**

Similar to the t table, the alpha term indicates which column of the table to use and the degrees of freedom term which row.

Example

$$n = 25,\ s = 0.673,\ \alpha = 0.10$$

$$\frac{(25-1)0.673^2}{\chi^2_{(0.10/2),(25-1)}} \le \sigma^2 \le \frac{(25-1)0.673^2}{\chi^2_{(1-0.10/2),(25-1)}}$$

What does this mean?

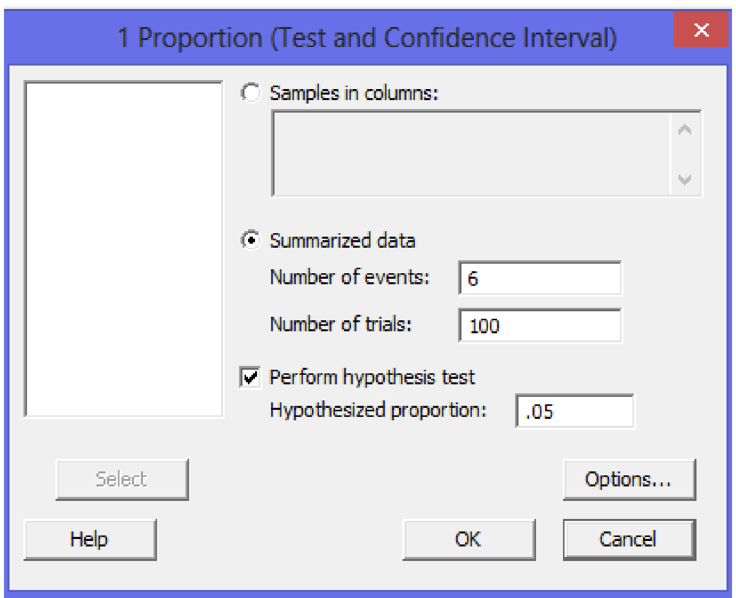
$$0.30 \le \sigma^2 \le 0.78$$

$$0.55 < \sigma < 0.88$$

Exercise 3: CIs from Minitab—Discrete Data

- **When Minitab has summarised discrete data, it is simple to calculate the confidence interval around the population proportion.**

  - **Create a data table with a column for the category and one for the counts.**
  - **Minitab will create confidence intervals for the proportions**



Exact

Sample	X	N	Sample p	95% CI	P-Value
1	6	100	0.060000	(0.022335, 0.126030)	0.820

CI for Proportions

$$\bar{p} \pm Z_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Where…

$\bar{p}$ = average proportion seen in the sample

n = sample size

α = risk

The proportion CI is based on the fact that the Z Distribution is a fairly good approximation of the binomial distribution at reasonable sample sizes. This formula only applies to <u>sample sizes of 30 or more</u>. Minitab does an exact calculation and is a better tool.

CI for Proportions Example

n = 700, # defectives = 16, α = 0.10

$$0.023 \pm Z_{0.10/2}\sqrt{\frac{0.023(1-0.023)}{700}}$$

$$0.023 \pm 1.645\sqrt{\frac{0.023(1-0.023)}{700}}$$

What does this mean?

$$\boxed{0.023 \pm 0.0094}$$

CI Exercise 4

1. Find CI for Mean and Standard Deviation for data in: CI_Ex4.MPJ
   - Repeat at the 90% confidence level.
2. Calculate the 95% CI for the proportion of defects from a process based on a sample of 431 items where 48 were found to be defective.

Summary of key learning points

- **Confidence Intervals provide realistic bounds for parameter estimates, i.e., an interval of "plausible values".**
- **If we have raw continuous data, Minitab will calculate CIs for us.**
- **If we don't have the raw continuous data, by using the mean, standard deviation, sample size, and confidence level we can still calculate confidence intervals for parameters such as the mean and standard deviation of the population.**
- **In Minitab we can use the sample size and number defective to calculate the CI for proportions. (The category of interest doesn't have to be "defective", it can be anything such as number of top two box scores or number of small businesses.)**
- **The factors that affect the width of a CI are:**
  - **Variation**
  - **Sample Size**
  - **Risk**

Black Belt

Sampling

Objectives of this module

At the end of this module, you will be able to:

- Be able to determine the goal of sampling
- B**e able to select the proper sampling strategy.**
- **Calculate the proper sample size for a given confidence level and margin of error.**
- **Understand what drives sample size.**

Questions Lead

Basic question: How can I make sure that the data I get really represents the population?

- Can samples fool me? How will I know that I can trust my samples?
- How many samples do I need?
- Why so many? (How can we get away with so few?)

Why Collect Data?

There are two basic needs for data:

<u>Estimation</u>

- **Your focus is to collect enough representative data to give you a "good" estimate of the parameters of the process. "Good" is defined by confidence level and margin of error.**
- **If historical data are not available, a data collection plan should be instituted to collect the appropriate data.**
- **<u>The rest of this module deals with sampling for estimation.</u>**

<u>Hypothesis Testing</u>

- **Your focus here is to collect data on the Y and different Xs to determine if there are differences in how the Xs affect the Y, or if different values or levels of an X affect the Y.**
- **Here again we need enough representative data to be able to prove the significance of the effects of the Xs.**
- **<u>We will cover sampling for testing in the Hypothesis Testing module.</u>**

Sampling Strategy

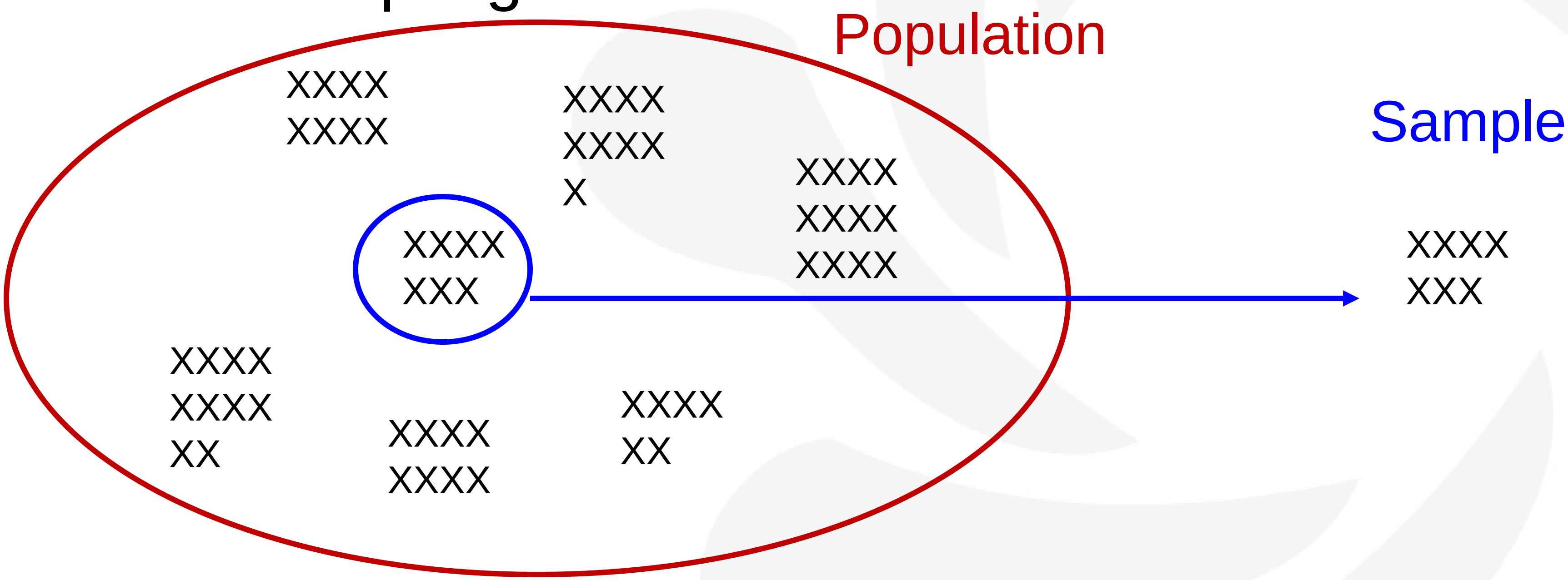
- What must a sample be?
  - Representative.
  - Unbiased.

- What kind of sampling can we do?
  - Population—Items exist and their characteristics are (relatively) stable.
  - Process—Items continue to be produced and their characteristics may change as the process varies.

Four Basic Sampling Strategies

- Random Sample – Population Studies. Each unit has an equal probability of being selected in a sample.
- Stratified Random Sample – Population Studies. Randomly sample within a stratified category or group. Sample sizes for each group are generally proportional to the relative size of the group.
- Systematic Sampling – Population or Process Studies. Sample every n^{th} unit. For example, collecting every 4^{th} unit.
- Rational Subgroup – **Process Studies**. Each unit is collected at point "A" in a process every n^{th} hour. Usually multiple sequential units are collected.

Cluster Sampling



- The population is composed of small groups called clusters.
- There is very little variation in the demographics of the clusters.
- Data is gathered in detail within one (or a few) cluster(s).
  (eg all the doctors in one representative hospital or all doctors in one city)

Sampling Exercise 1

Match the sampling situation with the best sampling strategy. Try to use each strategy only once.

1. We want to monitor our item processing staff on a regular basis to see if the number of errors are going up.
2. Without spending a lot of money, we want to know whether the people in Glasgow prefer Celtic or Rangers.
3. A DFSS team needs to thoroughly understand our affluent urban customers to better develop products for their needs.
4. Our Supply Chain Management needs to forecast how long we expect it will take to enter into contracts. We deal with many types of suppliers in many different areas of the country.
5. We want to know how accurate is our billing to our corporate customers. It will take about 15 minutes to review each bill for accuracy. Bills are issued monthly, but evenly throughout the month (i.e. not all of them are end-of-month billing).

- **Random**
- **Rational Subgroup**
- **Stratified Random**
- **Systematic**
- **Cluster**

How Much Data Do You Need?

- How many actual data points do you need to collect for your sample?
- We give the Green Belts an Excel calculator, but what are the equations behind it?
- Where did the equations come from?
- What are the things that drive higher sample sizes?

Sample Size Equations

Continuous data formula

$$n = \left(\frac{Z_{\alpha/2} s}{\Delta} \right)^2$$

Discrete data formula

$$n = \left(\frac{Z_{\alpha/2}}{\Delta} \right)^2 p(1-p)$$

***n* = sample size**

***s* = standard deviation**

***p* = proportion of observations exhibiting the event. Commonly the proportion 'defective' or proportion 'good'.**

Δ = margin of error, ±, of the estimate of population mean or proportion.

$Z_{\alpha/2}$ = Z score with risk of $\alpha/2$.

Drivers of Sample Size

Continuous data formula

$$n = \left(\frac{Z_{\alpha/2}s}{\Delta}\right)^2$$

Discrete data formula

$$n = \left(\frac{Z_{\alpha/2}}{\Delta}\right)^2 p(1-p)$$

Given these formulas, what are the drivers of sample size?

- **Confidence Level**
- **Variation**
- **Margin of Error**

What isn't here that you would suspect?

- **Population Size**

Finite Population Correction Factor

For both continuous and discrete data, where the population size is small relative to the calculated sample size, use the population correction factor to adjust in down:

$$n_{finite} = n \ / \ (1+ ((n - 1)/N))$$

n	**=**	**calculated sample size**
N	**=**	**population size**
***n*finite**	**=**	**population adjusted sample size**

This population correction factor is applicable for sample sizes calculated using discrete and continuous formula

Finite Population Correction Factor (reorganised)

$$Nfinite = n \;/\; (1 + ((n - 1) / N))$$

Sample Size Formula Examples

Continuous Data Example:
We want to be able to know at what level our call center process is running. If our process has a standard deviation = 5.0, what size sample will we need to be able to estimate the mean within ± 1 minute?

Discrete Data Example:
Based on historical data, our defect rate (p) has been running at approximately 0.05. We want to sample the process occasionally and be able to see if it has gone up or down by 0.02. How large of a sample do we need to take?

$$\mathbf{n} = \left(\frac{1.96\mathrm{x}5}{1}\right)^2 = 97$$

$$\mathbf{n} = \left(\frac{1.96}{.02}\right)^2 \left(.05(1-.05)\right) = 457$$

Other Common Confidence Factors

- For 85% confidence, Z = 1.439
- For 90% confidence, Z = 1.645
- For 99% confidence, Z = 2.575

Sample Size Calculator Continuous

Open the Excel file named: BB_Sample_Size_Calculator.xls

FIREBRAND Immerse. Accelerate. Measure.

Estimated Sample Sizes for Continuous Data at 99%, 95% and 90% Confidence Levels

Enter Population Size Here* **1,000,000,000**

Enter Estimated Standard Deviation Here **20**

(If unknown, use 1/6 of the known range of the data)

The population size is used to adjust the sample size with the Finite Population Correction Factor.

Margin of Error	Sample Size 99%	Sample Size 95%	Sample Size 90%
1	2663	1537	1083
2	666	385	271
3	296	171	121
4	167	97	68
5	107	62	44
6	74	43	31
7	55	32	23
8	42	25	17
9	33	19	14
10	27	16	11
11	23	13	9
12	19	11	8
13	16	10	7
14	14	8	6
15	12	7	5
16	11	7	5
17	10	6	4
18	9	5	4
19	8	5	3
20	7	4	3

This application can be used effectively for determining initial sample sizes required for process studies or descriptive statistics of process information. It should not be used for hypothesis testing. Minitab is the recommended sample size calculator for hypothesis testing or experimental studies.

Sample Size Calculator Discrete

Open the Excel file named: BB_Sample_Size_Calculator.xls

FIREBRAND Immerse. Accelerate. Measure.

Estimated Sample Sizes for Discrete Data at 99%, 95% and 90% Confidence Levels

Enter Population defective rate (p)*	0.455	Margin of Error	Sample Size 99%	Sample Size 95%	Sample Size 90%
		0.01	16,507	9,527	6,711
		0.02	4,127	2,382	1,678
		0.03	1,835	1,059	746
		0.04	1,032	596	420
		0.05	661	382	269
		0.06	459	265	187
		0.07	337	195	137
		0.08	258	149	105
*p must be between 0 and 1, if unknown use .5		0.09	204	118	83
		0.1	166	96	68
		0.11	137	79	56
		0.12	115	67	47
		0.13	98	57	40

This application can be used effectively for determining initial sample sizes required for process studies or descriptive statistics of process information. It should not be used for hypothesis testing. Minitab is the recommended sample size calculator for hypothesis testing or experimental studies.

Sampling Exercise 2

Sampling for Estimation

Objective: To illustrate the use of the BB sample size calculator

- Each month we want to **monitor the errors made on cash deposits in each region**. We want to know where we are within ± 1%. We believe the error rate is 11.5%. How many samples should we take in each region, each month?

- After sampling for a few months, we have found that the error rate is only 8.1% within one of the branches. The typical number of deposits taken per month totals 9,000.
  Now how many samples should we take per month?

10 minutes

Nfinite = n / (1+ ((n – 1)/N))

Sample Size Concepts

Smaller sample sizes:

- Less Cost
- Quicker data collection
- Wider confidence intervals and/or more risk of missing the population parameter.

Larger sample sizes:

- Higher costs
- Longer time to get data
- Tighter confidence intervals and/or less risk of missing the population parameter.

Summary of key learning points

- Sample size for determining population mean or proportion depends on:
  - What level of risk you're willing to take.
  - What size difference you want to detect.
  - How much variation is in the population.
- Before collecting data, you should think about the sampling strategy and sample size requirements to ensure that you have an appropriate amount of data for drawing conclusions.
- Sample size determination should include practical considerations, like economics, as well.

Black Belt

Measurement System Analysis

Objectives of this module

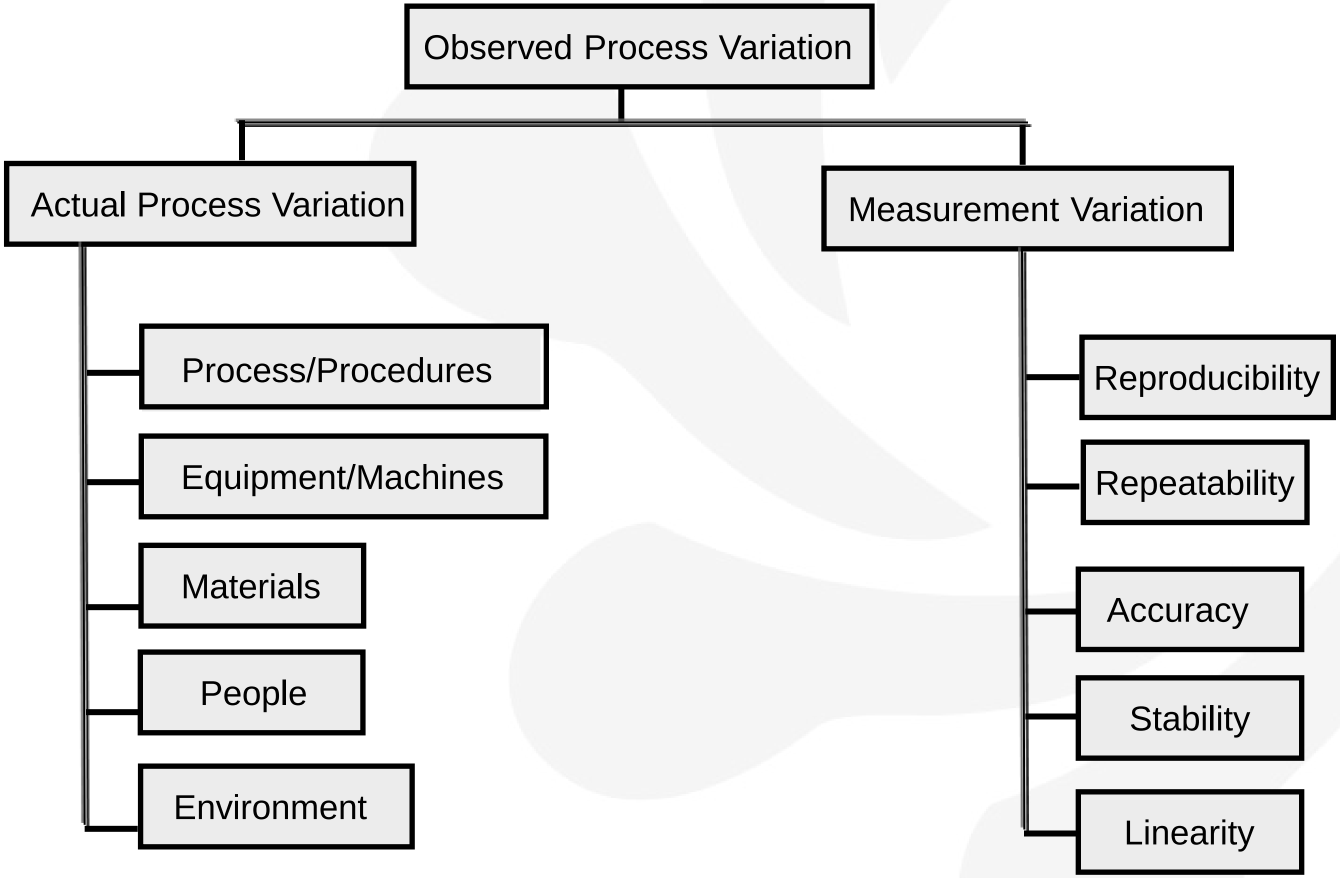
At the end of this module, you will be able to:

- Explain why we need to use MSA.
- Explain the concepts within MSA.
- Run and analyse a Gauge R&R for both continuous and discrete data.
- Be able to improve poor measurement systems.

Questions Lead

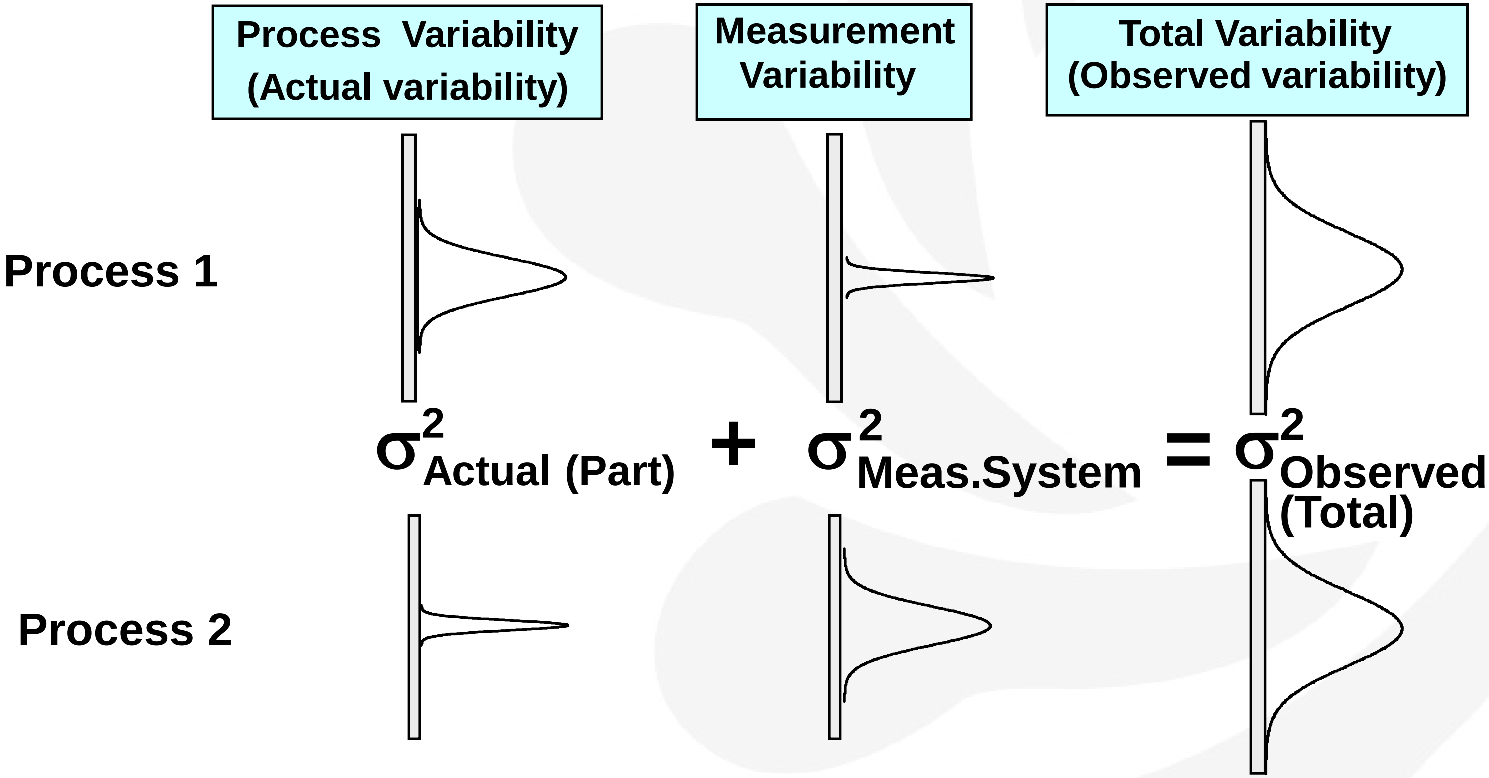
- Basic question: Can I trust my data?
  - When I get my data, can I rely on it to be accurate?
  - If I were to measure it twice would I get the same answer?
  - Does everyone measure it the same and get the same answer?
  - Will I still be able to trust my data later in the year?

Possible Sources of Observed Process Variation



To address actual process variability, the variation due to the measurement system must first be identified and separated from that of the process.

Observed Variation



- What are the strengths and weaknesses of the two processes?
- If customers were complaining about both of them, which one would you rather be assigned to?
- What would be one of your <u>first tasks</u> for each after completing the MSA?

Measurement Systems Analysis

- What is a measurement system? Think of the measurement system as a process that returns data. If you were to do a process map on it, including the inputs and outputs, that would be your measurement system. Common inputs are:
  - Items being measured.
  - People doing the measurements.
  - Operational definitions of the standards.
  - Work instructions for how to do the measurements.
  - Instruments used to take measurements.
  - The environment around the measurement process.
- What is an MSA? An MSA is a measurement of our measurement system to see if we can trust the data we are gathering.
- It is <u>not optional</u> for a Lean Six Sigma project! This applies even if the data is coming from an automated data collection system. We still need assurance that we can trust the data.
- Where applicable it will assess all five sources of variation: Repeatability, Reproducibility, Accuracy, Linearity and Stability.

Differences in Types of Measurement Systems

Automated Data Collection

- Need to validate that the information is correct by whatever method we can devise. Some possibilities include:
  - May need to do at least one manual calculation to check the computer.
  - May need to investigate the formulae and/or algorithms the computer uses to calculate the result.
  - May need to validate data inputs to the program. Can we trust those sources of data?

Data that is Measured (or Counted)

- Typically we use a traditional measurement systems analysis using a **Gauge R & R (Repeatability & Reproducibility) Study**.

Gauge R&R Concepts (1 of 4)

I. Measures of Accuracy

- *<u>Accuracy</u>* is a state where there is little difference between the **<u>observed average value</u>** of the measurements and some established reference value or standard (**<u>the real value</u>**).

II. Measures of Precision

- *<u>Precision</u>* means little variation exists between repeated measurements in relation to the current process variation.

III. Measure of both Accuracy & Precision

Gauge R&R Concepts (2 of 4)

I. Measures of Accuracy

- A. Bias: The difference between observed average measurement and a standard.
  - Example of bias: We are measuring the width of cheques and measure from the 1/16 inch mark instead of the end of the ruler. The bias will be a constant 1/16 inch.
- B. Linearity: Do we get the same accuracy in the measurement across all sections of the measurement scale? Poor linearity is seen when one end of the scale is less accurate than other parts of the scale.
  - Example of bad linearity: There are 10 potential errors we can make on applications. If there are 1 or 2 errors we can count them ± 1 error. But if there are 8 or more errors, we tend to miss the actual number by ± 3.

Gauge R&R Concepts (3 of 4)

II. Measures of Precision

- A. Repeatability: Variation when one person repeatedly measures (or counts) the same unit with the same measuring equipment or system.
- B. Reproducibility: Variation when two or more people measure (or count) the same unit with the same measuring equipment or system.
- C. Stability: Variation obtained when the **same person** measures the same unit with the same equipment or system over a period of time.
- D. Discrimination: The ability of the measurement system to divide measures into categories.
  - Example of poor discrimination: We want to reduce our month end closing time to one day from five, but the various processes are only measured in one day increments, most of them measuring one day to complete. We don't know where to attack because we can't discriminate between short and long processes.

Gauge R&R Concepts (4 of 4)

II. Measures of Precision, continued.

- E. Linearity: Previously discussed, but also affects the precision.

III. Measures of both Accuracy & Precision

- Correlation
  - Reproducibility meets bias.
  - When relative values of the items measured are the same, but one measurement system (or operator) is always higher.
  - Usually fixed by **recalibrating** the gauges or operational **definitions**.
- Example of bad correlation: A Glasgow unit counted the defects on a set of applications and found 25% errors. A Edinburgh unit did the same and found 4% errors. They traded sets and counted again. On the Edinburgh set, Glasgow counted about 25% errors and sure enough, the Edinburgh group counted about 4% errors on the Glasgow set.

Summary of Measurement Study Characteristics

- Discrimination- Can we know when we win? Do we know the difference between one thing and another when it happens?
- Accuracy- Is it what we think it is?
- Bias- Does someone or something have an effect on the number we see?
- Precision- Can we get the same number each time we try?
- Repeatability- Does the same result come from more tries?
- Reproducibility- Can some one else get the same number?
- Linearity- Is there a limitation in the system used to take measurements that creates inaccuracies?
- Stability- Is the measurement effected by a time related issue?

These questions apply whether we are doing a traditional GR&R study or whether we are looking at data from a computer!

Attribute Gauge

- What is an attribute gage?
  - A measurement system that is used to assign an attribute to an item by comparing it to a standard.
  - Many times the attribute is **'pass'** or **'fail'**, but does not have to be binary. Ordinal or nominal data work as well.
  - Can assess whether an item is good or bad, but does not indicate how good or how bad the item is.
- What does an Attribute Gauge R&R study do?
  - Can **assess your standard against a customer requirement**.
  - Determines if people use the same criteria to determine the attribute.
  - Quantifies the ability of people to consistently repeat their inspection/audit decisions.

Purpose of Attribute R&R

- To assess your inspection standards against your customers' requirements.
- To determine if inspectors across all shifts, all locations, etc. use the same criteria to determine "good" from "bad".
- To quantify the ability of appraisers to accurately repeat their inspection decisions.
- To identify how well these appraisers are conforming to a "known standard" which includes:
  - how often staff decide to **pass** truly **defective** product.
  - how often staff **fail** truly **acceptable** product.
- Discover areas where:
  - **training** is needed.
  - **procedures** are lacking.
  - **standards** are not defined.

Attribute GR&R—The Method

1. Select a minimum of 30 items from the process (more equals more confidence and is recommended).
   - 50% of the items in your study should have defects.
   - 50% of the items should be defect free.
   - If possible select border line (or marginal) good and bad samples.

2. Identify the employees for the study who may utilise the measurement system. Suggest 3 for the study.

3. Have each person, independently and in random order, assess these items and determine whether or not they pass or fail. (this step is repeated twice with a time delay between trials).

4. Enter the data into a Minitab data table and analyse the results.

5. Document the results. If necessary, implement appropriate actions to fix the inspection/audit/measurement process.

6. Re-run the study to verify the fix.

Exercise 1, Attribute Gauge R&R

- We want to **reduce handle time in our call centre** and **we believe one of the key Xs is whether the call passed a quality review**. But we are **not sure whether all call assessors would rate the calls in the same way**. If we gather data, but have incorrectly assessed the call, we will not be able to accurately interpret it so we decide to do a measurement study.
  - Thirty calls were recorded and a group of supervisors determined whether the call passed or failed.
  - Three operators were then asked to individually listen to the recorded calls and determine whether each call passed or failed.
  - Several weeks later, these same operators listened to the calls again in a random order and made the determination again.
  - The data are recorded in MSA_Ex_1_Attribute.MPJ.
- Analyse this measurement system **to see if we can utilise the operators' assessments** of the type of call **in our project**.

Exercise 1, Minitab Data Table

- Set up and run the study and enter the data into a Minitab data table as follows:

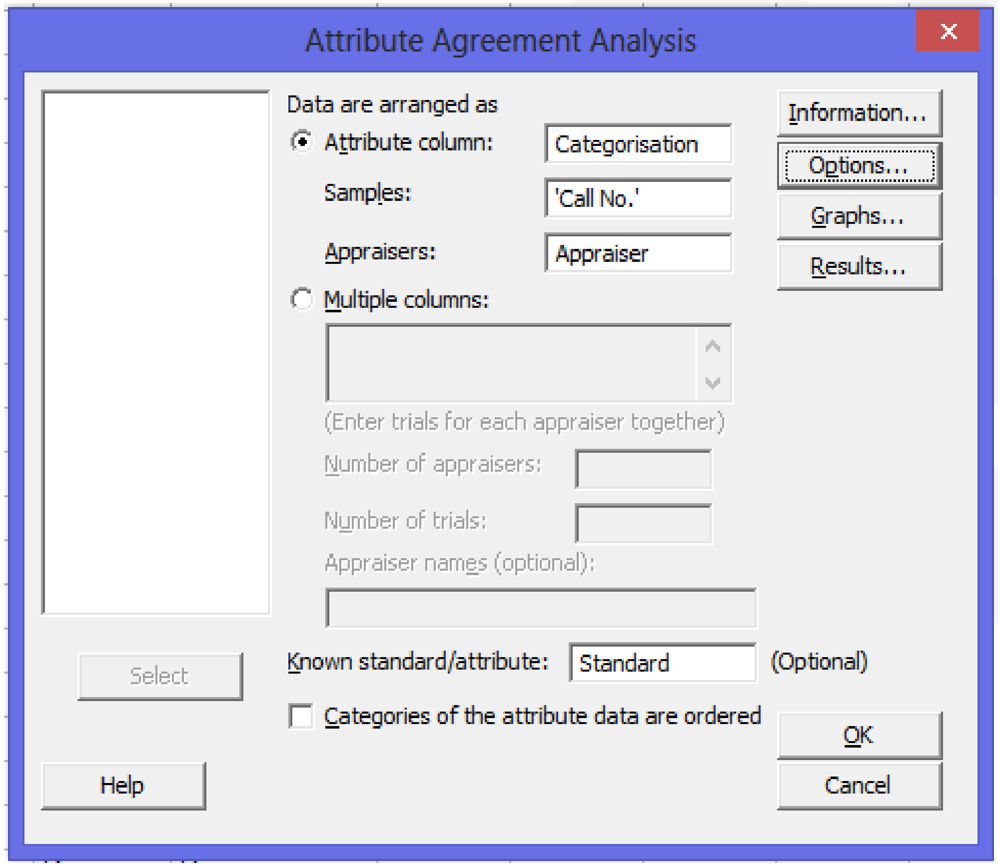
  - One column to identify the item being measured, in this example the call number.
  - One column to identify the standard or truth about the item.
  - One column for each appraiser with their evaluation of the item.

- Data Table: MSA_Ex_1_Attribute.MPJ

↓	C1	C2	C3-T	C4-T	C5-T
	Trial No.	Call No.	Appraiser	Standard	Categorisation
1	1	1	Alex	P	P
2	1	2	Alex	F	F
3	1	3	Alex	F	F
4	1	4	Alex	F	F
5	1	5	Alex	P	P
6	1	6	Alex	P	P
7	1	7	Alex	F	F
8	1	8	Alex	P	P
9	1	9	Alex	F	F
10	1	10	Alex	P	P
11	1	11	Alex	P	P
12	1	12	Alex	P	P
13	1	13	Alex	P	P
14	1	14	Alex	F	F
15	1	15	Alex	F	F
16	1	16	Alex	P	P
17	1	17	Alex	P	P
18	1	18	Alex	F	F
19	1	19	Alex	P	P
20	1	20	Alex	P	P
21	1	21	Alex	F	F
22	1	22	Alex	F	F
23	1	23	Alex	F	F

Exercise 1, Minitab Data Entry

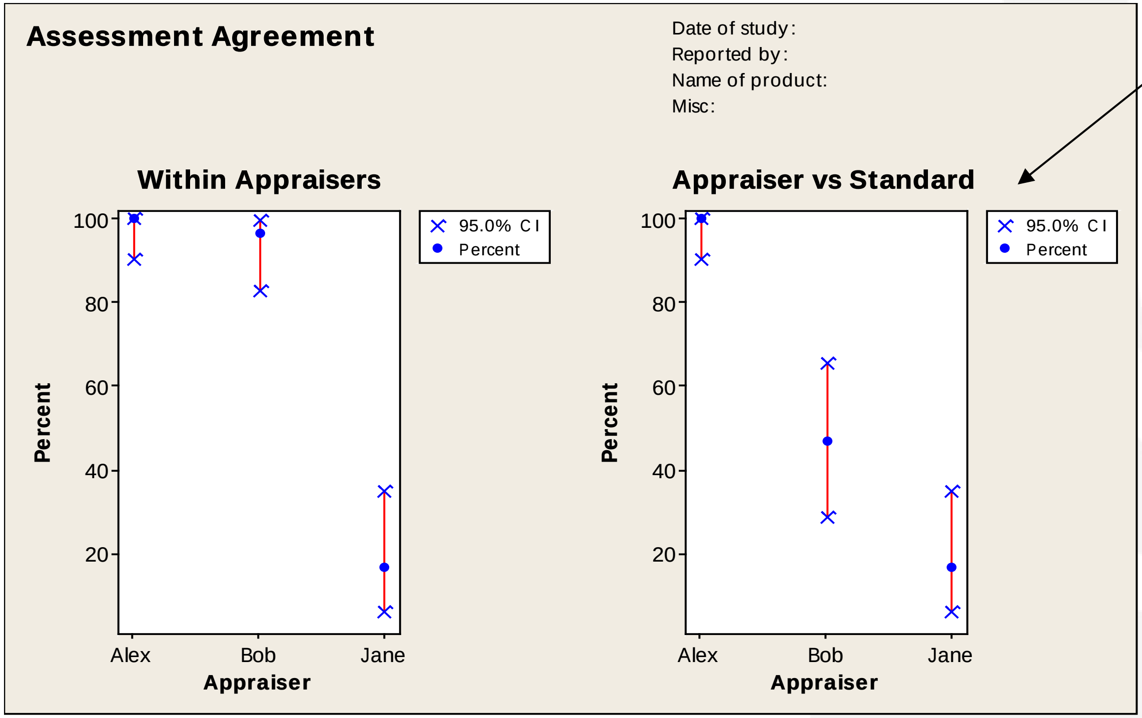
Stat > Quality Tools > Attribute Agreement Analysis



Our data is organised in a single column so using the top three variables.

To use the multiple columns option all the columns for one appraiser must be next to each other.

Exercise 1, Attribute GR&R Graphical Output



Appraiser vs Standard

Because there was an "expert" decision (a "known standard"), the graphical output shows that performance against that standard.

- Bob only agreed with the "expert" decision around on 46.7% of his assessed calls
- Jane agreed 16.7%
- The assessment by Alex matched the "expert" exactly.

The **Within Appraiser** results show the **repeatability** of the appraisers as a percentage score. The blue dot indicates the actual calculation, whilst the lines extending in either direction indicate the 95% confidence intervals for the result.

Confidence levels are used because a sample size of 20 or 30 (calls in this example) is not very large when dealing with attribute data.

Bob reached the same decision on 96.7% of his calls (Jane 16.7%, Alex 100%)

Exercise 1, Attribute GR&R Session Window Output

Overall study variation

This metric will indicate the overall indication of the measurement system.

For Attribute GR&R this means that we're looking for complete agreement across the appraisers, within appraiser and to the standard/expert assessment.

% Study Variation Criteria:

- **Good: <10%**
- **Marginal: 10% to 30%**
- **Bad: >30%**

Session Window Output

All Appraisers vs Standard

Assessment Agreement

# Inspected	# Matched	Percent	95% CI
30	3	10.00	(2.11, 26.53)

Matched: All appraisers' assessments agree with the known standard.

1 - (3/30) = 90% study variation

- **How good is the MSA? What do we need to fix?**

Exercise 1, Attribute GR&R Session Window Output

Within Appraisers

Assessment Agreement

Appraiser	# Inspected	# Matched	Percent	95% CI
Alex	30	30	100.00	(90.50, 100.00)
Bob	30	29	96.67	(82.78, 99.92)
Jane	30	5	16.67	(5.64, 34.72)

Matched: Appraiser agrees with him/herself across trials.

- Number Inspected: total number of items (calls) rated.
- Number Matched: number of calls where the operator determined the same rating (type of call) on both trials.
- Rater Score: Number Matched divided by Number Inspected.
- How repeatable are the operators?

Exercise 1, Attribute GR&R Session Window Output

Between Appraisers

Assessment Agreement

# Inspected	# Matched	Percent	95% CI
30	3	10.00	(2.11, 26.53)

Matched: All appraisers' assessments agree with each other.

- Between Appraisers % is the number items matched within raters (trial to trial) and matched between all other raters, divided by total number of items inspected
- All raters have to give the item (call) the same rating each time (repeatability) and the same rating as each other (reproducibility).
- Overall, how precise are the operators?

Note that the percentage agreement shown here is the same as the % vrs Standard as one of the appraisers (Alex) matched completely to the standard with all of his ratings

Exercise 1, Accuracy

Each Appraiser vs Standard

Assessment Agreement

Appraiser	# Inspected	# Matched	Percent	95% CI
Alex	30	30	100.00	(90.50, 100.00)
Bob	30	14	46.67	(28.34, 65.67)
Jane	30	5	16.67	(5.64, 34.72)

Matched: Appraiser's assessment across trials agrees with the known standard.

- Number of times the rater matched themselves **AND** the standard divided by the total number of trials.
- How accurate are the operators?
- What are the next steps?

Exercise 1, Diagnostics

- Included within the diagnostics are Kappa Statistics:
  - Kappa Statistics: _An index which compares the agreement against that which might be expected by chance. Kappa can be thought of as the **chance-corrected proportional agreement**.
  - The kappa coefficient equals +1 when there is complete agreement of the raters. When the observed agreement exceeds chance agreement, the kappa coefficient is positive, with its magnitude reflecting the strength of agreement as shown below.
    - 0.90 - 1.00 = Measurement system is excellent.
    - 0.70 - 0.89 = Measurement system is capable - continue improvements.
    - 0.50 - 0.69 = Measurement system is marginal - needs improvement.
    - 0.00 - 0.49 = Measurement system unacceptable.
  - Although unusual in practice, kappa is negative when the observed agreement is less than chance agreement. The minimum value of kappa is between -1.

Exercise 1, Diagnostics

Assessment Disagreement

Appraiser	# P / F	Percent	# F / P	Percent	# Mixed	Percent
Alex	0	0.00	0	0.00	0	0.00
Bob	8	61.54	7	41.18	1	3.33
Jane	0	0.00	0	0.00	25	83.33

```
# P / F:  Assessments across trials = P / standard = F.
# F / P:  Assessments across trials = F / standard = P.
# Mixed: Assessments across trials are not identical.
```

Fleiss' Kappa Statistics

Appraiser	Response	Kappa	SE Kappa	Z	P(vs > 0)
Alex	F	1.00000	0.129099	7.74597	0.0000
	P	1.00000	0.129099	7.74597	0.0000
Bob	F	-0.05727	0.129099	-0.44363	0.6713
	P	-0.05727	0.129099	-0.44363	0.6713
Jane	F	0.02356	0.129099	0.18252	0.4276
	P	0.02356	0.129099	0.18252	0.4276

What do the Kappa values above tell us?

Bob marked 8 call P (on both trials) when the standard was F. He also marked 7 calls F (on both trials) when the standard was P.

Accuracy issue

Of the 25 calls that Jane didn't get right ALL where due to inconsistency of marking between the first and second trials

Repeatability issue

Variable Gauge

- What is a variable gage?
  - A measurement system that is used to measure a continuous characteristic of an item or event.
  - Much of the time this will involve some sort of instrument to take the measurement.
  - Can assess not only whether an item is good or bad, but how good or how bad the item is.
- What does a Variable Gauge R&R study do?
  - Can assess your standard against a customer requirement and use a standard to assess the measurement device for linearity issues.
  - Measures the overall effectiveness of the people in using the measurement system, as well as the precision of the measurement system itself.
  - Should be the basis for evaluating all continuous measurement systems.

Purpose of Variable R&R

- To assess your measurement device/system against your customer's requirements.
- To determine if inspectors across all shifts, all locations, and so on, use the same device/system to accurately determine measurements.
- To quantify the ability of appraisers to accurately repeat their measurements.
- Discover areas where:
  - training is needed.
  - procedures are lacking.
  - standards are not defined.

Variable Gauge R&R—The Method

1. Select a minimum of 10 items from the process.
   - Items should cover the entire specification range.
2. Identify the staff members for the study who may utilise the measurement system. Suggest 3 people.
3. Have each person, independently and in random order, measure these items and determine the measurement.
4. Perform Step 3 a minimum of two times for each staff member (at least 2 trials).
5. Enter the data into a Minitab data table and analyse the results.
6. Document the results. If necessary, implement appropriate actions to fix the measurement process.
7. Re-run the study to verify the fix.

Exercise 2, Continuous Gauge R&R

- A LEAN Team needed to measure the time it took to complete various tasks.
  - **Ten** different workers were videotaped performing a task (each task is different for each worker); supervisors determined the standard time expected for each task based on historical data.
  - **Three** team members were then asked to view the tapes and time how long the task took.
  - Two days later, these same team members viewed the same tapes in a different order and timed the task again.
  - The data are recorded in MSA_Ex_2_Continuous.MPJ.
- Analyse this measurement system to see if we can trust the data for our LEAN project.

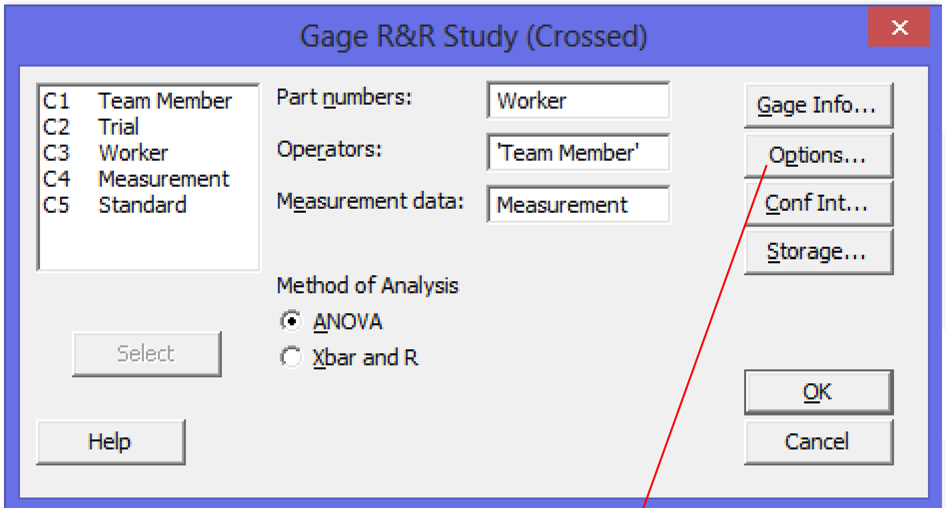
Exercise 2, Minitab Data Table

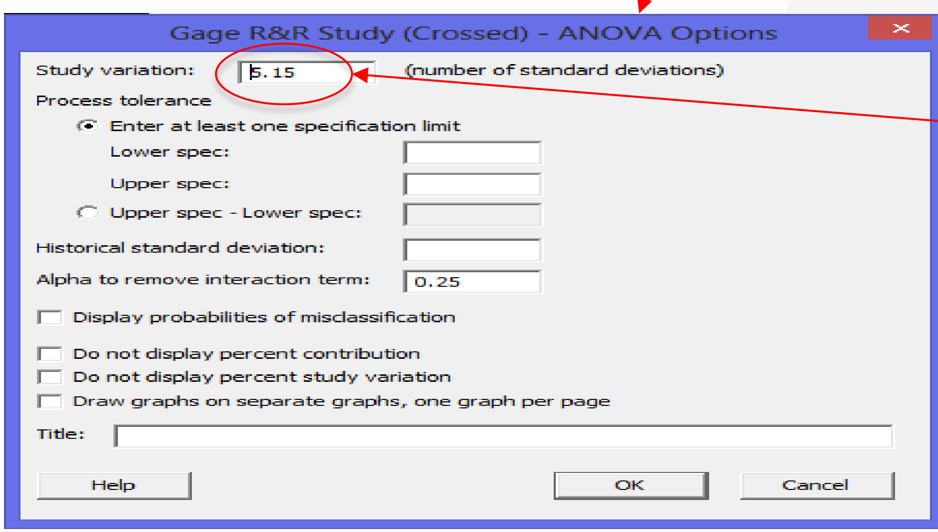
- Set up and run the study and enter the data into a Minitab data table as follows:
  - One column to identify appraisers.
  - One column to identify the trial.
  - One column to identify the item being measured, in this example the Worker.
  - One column for the measurement of the item.
- Data Table: MSA_Ex_2_Continous.MPJ

↓	C1-T	C2	C3	C4	C5
	Team Member	Trial	Worker	Measurement	Standard
1	Jack	1	1	0.887	0.90
2	Jack	2	1	0.885	0.90
3	Jack	1	2	0.815	0.81
4	Jack	2	2	0.828	0.81
5	Jack	1	3	0.885	0.89
6	Jack	2	3	0.888	0.89
7	Jack	1	4	0.910	0.90
8	Jack	2	4	0.918	0.88
9	Jack	1	5	0.919	0.88

Exercise 2, Minitab Data Entry

<u>Stat > Quality Tools > Gage Study > Gage R&R Study (crossed)</u>





"Crossed" indicates that every rater (Team Member) is going to measure every part (Worker). The parts are "crossed" to all raters.

In our example we have 3 "Operators" (Nick, Jack and Carole) assessing 10 "Part numbers" (the workers)

<u>Options</u>

Study variation can be left at 6 (which is the Minitab default) or set to "5.15" is which is used to give a 99.0% spread. This is a common industrial standard for GR&R studies

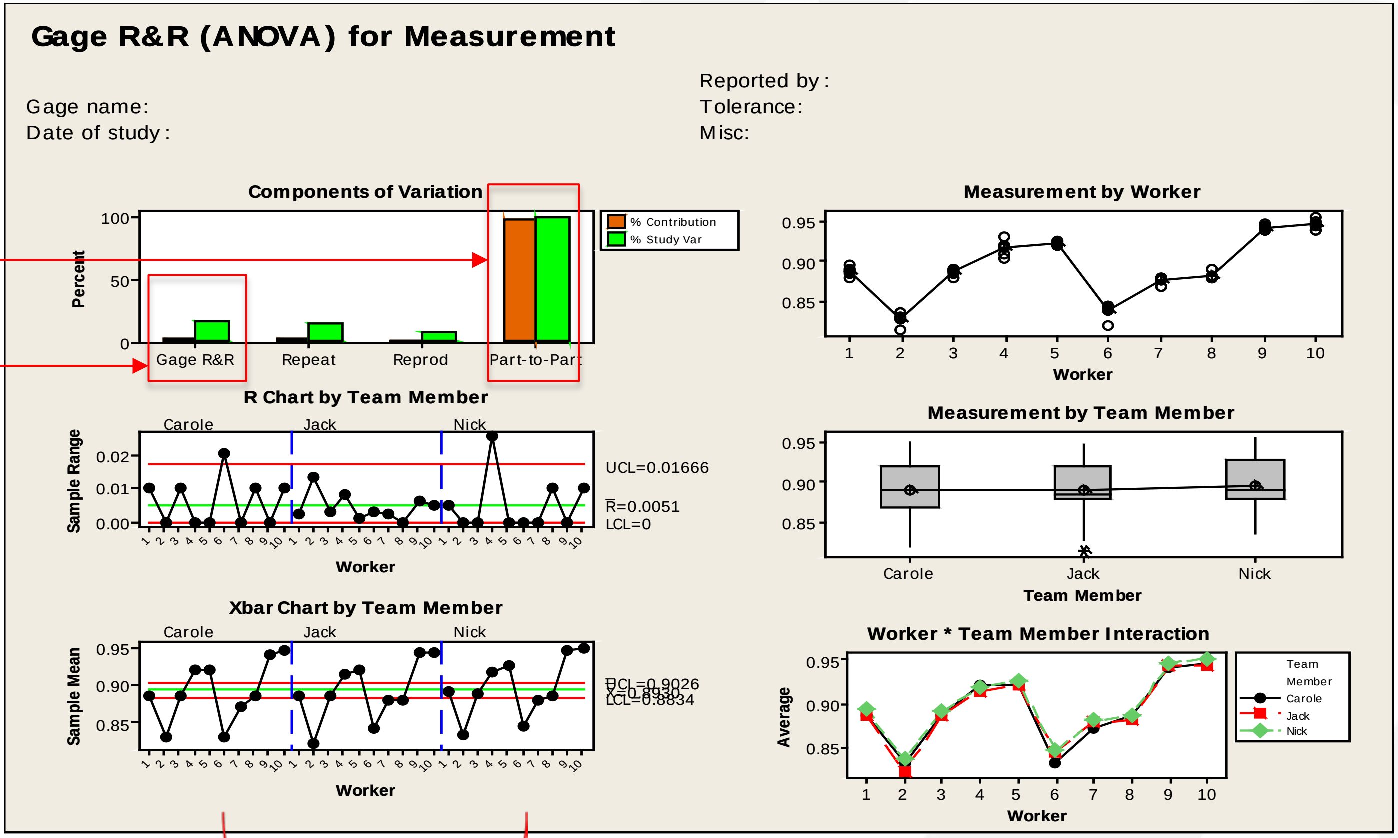
Exercise 2, Results – Graphical Output

The Gage R&R group of columns should be small in comparison to the part-to-part group of columns

The first (orange) column in the GR&R is the % contribution of variation.

The second (green) column shows the GR&R as a % of study variation

The middle two sets of columns represent the sub-components of Gage R&R separately (ie Repeatability and Reproducibility)

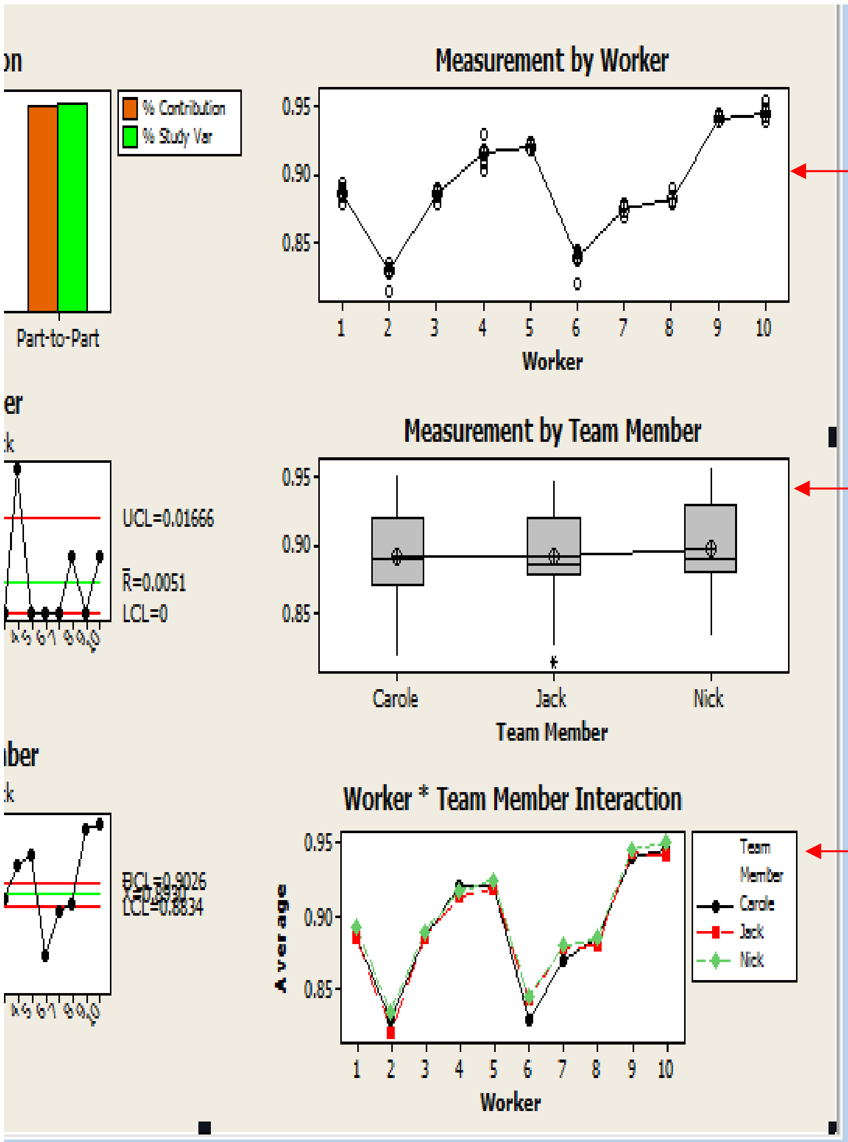


The two lower left charts are effectively control charts.

The middle chart is checking for unlikely results gathered. If any of the points are outside the red lines check for data gathering problems or with the assessment of a particular part.

The bottom chart is simply charting the mean of the measurements by appraiser and part.

Exercise 2, Results – Graphical Output (cont'd)



The top right chart shows all the results by part number (workers in this example). Each part number has six data points having been assessed by three workers twice.

The second chart helps show reproducibility by showing all the results by appraiser.

The boxplots show the spread of the data

The third chart is the same as the top chart but separates out the results by appraiser.

Exercise 2, Results – Session Window Output

Two-Way ANOVA Table Without Interaction

Source	DF	SS	MS	F	P
Worker	9	0.0843958	0.0093773	303.691	0.000
Team Member	2	0.0003670	0.0001835	5.943	0.005
Repeatability	48	0.0014821	0.0000309		
Total	59	0.0862450			

Gage R&R

Source	VarComp	%Contribution (of VarComp)
Total Gage R&R	0.0000385	2.41
Repeatability	0.0000309	1.93
Reproducibility	0.0000076	0.48
Team Member	0.0000076	0.48
Part-To-Part	0.0015577	97.59
Total Variation	0.0015962	100.00

Source	StdDev (SD)	Study Var (5.15 * SD)	%Study Var (%SV)
Total Gage R&R	0.0062056	0.031959	15.53
Repeatability	0.0055568	0.028617	13.91
Reproducibility	0.0027626	0.014227	6.91
Team Member	0.0027626	0.014227	6.91
Part-To-Part	0.0394682	0.203261	98.79
Total Variation	0.0399531	0.205758	100.00

Number of Distinct Categories = 8

The ANOVA is used to assess whether the part, assessor or interaction of part and assessor have a significant affect on the result.

The second set of data represents the GR&R in terms of its contribution towards variance – we want this small.

The third set of data provides the detail behind the top left graph in the graphical output

The number of distinct groups the measurement system is capable of distinguishing within the process variation present

The key numbers here are the **% Gauge R&R Variance Component**, the **% Gauge R&R** and the **number of distinct values**.

Exercise 2, Analysing the Results

- What is good and bad?

	% Variance Components	% Study Variation	Distinct Categories
Good	**<4%**	**<10%**	**>10**
Marginal	**4% to 9%**	**10% to 30%**	**4 to 9**
Bad	**>9%**	**>30%**	**<4**

- Whether you can use the measurement system depends somewhat on how broken your process is.
- What do we decide about our measurement system?

Fixing the Measurement System (Discrete or Continuous)

- Why?
  - If you don't have reliable data, you won't have a chance to do correct Analysis, make the right Improvements, or keep the process under Control.
- **How**?
  - **Find the cause of variability** in measurement system, **remove** defects, then **rerun** the MSA to verify.
- When?
  - Periodically (minimum of once per year) rerun a MSA to make sure the measurement system is stable and continues to generate reliable data. This **should be included in the Control Plan**.

How do you fix a Measurement System?

- If a dominant source of variation is **repeatability**, you need to *change the measurement system*.

- If you are using an **instrument**, and you find that the instrument is "state-of-the-art" and it is performing to its specifications, you *still have to change the measurement system.*

- If a dominant source of variation is **people** (**reproducibility**), you must address this via *training and definition of the standard operating procedure.* You should look for differences between people to give you some indication as to **whether it is a training, skill, or procedure problem**.

- If the process is repeatable and reproducible but **does not match the customer's interpretation**, must address *understanding the customers' requirements.*

- **This may be the first part of your project**!

Evaluation Questions to Diagnose Issues

- What are the **components** of the measurement system? Which might be weak?
  - Are **inspection or measurement procedures** written?
  - Is the detailed **process map for the measurement process** developed?
  - Are there **operational definitions for all characteristics**?
  - Are the inspectors/raters/operators **trained**? Is there a **certification** procedure for them?
- Where are the **sources** of variation?
  - Repeatability?
  - Reproducibility?
  - Accuracy?

Final Questions

- Have you picked the right measurement system? Is this measurement system associated with either **critical inputs or outputs**?
- Have we informed the right people of our results?
- Who owns this measurement system?
- Who owns trouble shooting?
- Does this process have a control plan in place and is MSA part of that control plan?
- What is the frequency of rerunning the MSA to check for stability?

Summary of key learning points

- Sources of variation come from both processes and measurement systems

- Ensure that you have a capable measurement system to allow you to focus on process change

- There are attribute and continuous measurement system analysis approaches which include comparison to between and within appraisers as well as to a known standard

- Should excessive levels of measurement system variation be found to exist, fix this first.

Recommended coach support points

- Setting up your Gage R&R study
- Interpretation of study results
- Identifying ways of improving the measurement system

FIREBRAND

Black Belt

Statistical Process Control

Objectives of this module

At the end of this module, you will be able to:

- Be able to use Statistical Process Control (SPC) in various phases of a Lean Six Sigma project.
- Understand process stability, types of variation and signals from control charts.
- Understand basic theory and assumptions to be able to interpret control charts.
- Be able to use rational sub-grouping.
- Be able to select and use the correct control charts based on the type of data and sampling strategy.

Questions Lead

Basic question: **Is the process stable?**

- What type of variation is in the process? Is it only common cause variation or is there special cause variation?
- Should I take action to change the process or leave the process alone?
- Did my process change force a change in the output?

Essence of SPC

“A process will be said to be predictable when, through the use of past experience, we can describe, at least within limits, how the process will behave in the future.”

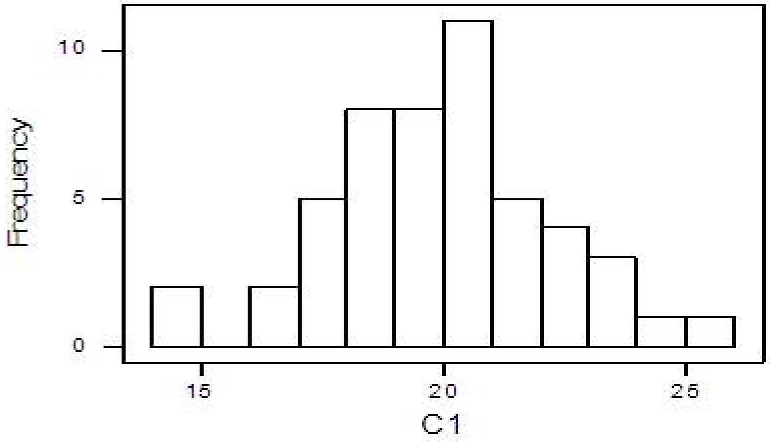
Walter Shewhart

as paraphrased by Donald J. Wheeler

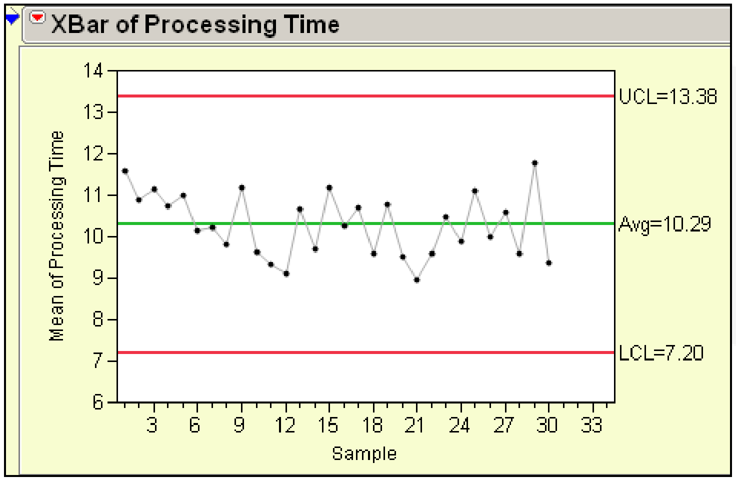
Distributions

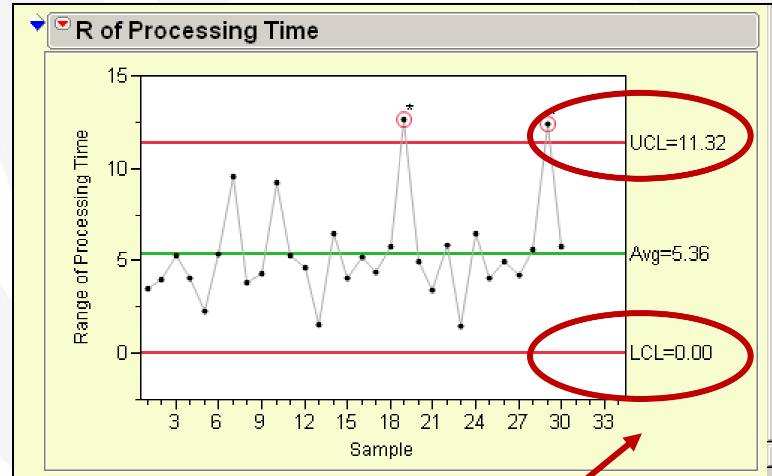
Distributions are the shape of frequency plots that show the observed fluctuations of the data.

*Therefore, statistical limits can be derived from these distributions to **predict the behavior** of the fluctuating pattern when there are **NO abnormal (special) causes**.*



Control Chart Components





time

Data points are plotted over time (shift, day, week, month, etc.).

- **Each data point represents either:**
  - **an individual observation such as budget variances.**
  - **a calculated value such as:**
    - **average processing time.**
    - **proportion of rejected transactions.**

- **Upper and lower control limits (circled) give the control chart the analytical power to determine if the process is stable.**

- **The centre line is the grand average of the observations.**

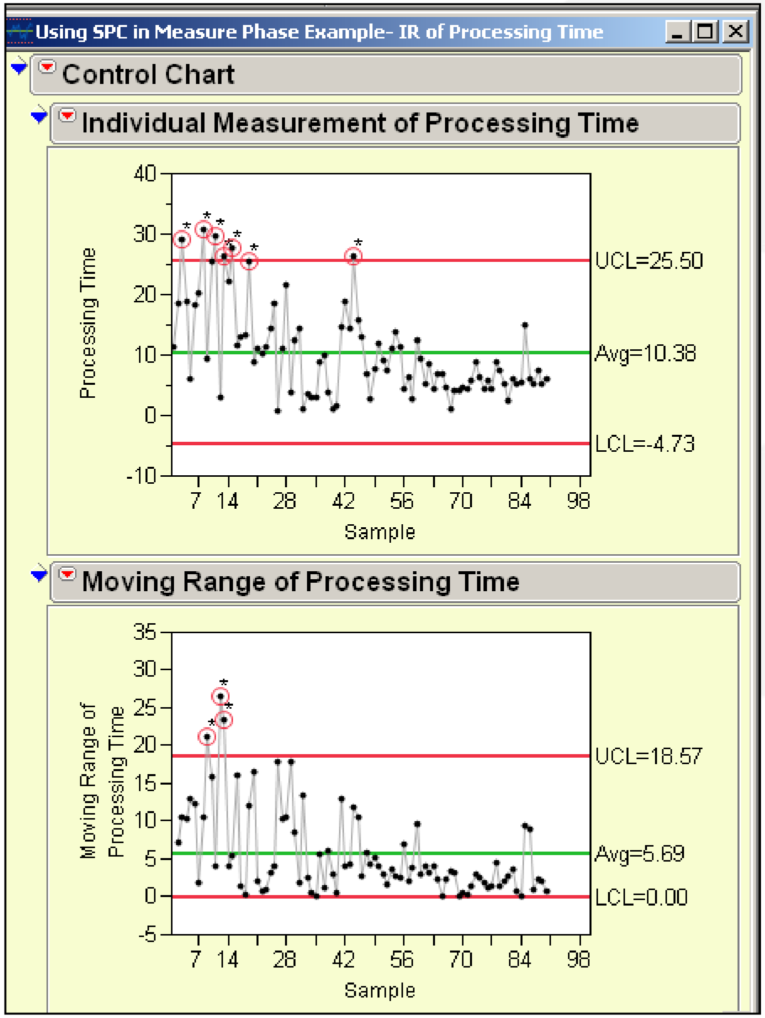
SPC

SPC is the basic tool for using statistical “signals” to monitor and/or improve process performance.

- As part of the Measure Phase, SPC is used to verify process stability and to identify and remove special causes of variation before we baseline the process.
- As part of the Analyse Phase, SPC can be used as an alternate type of hypothesis test.
- As part of the Improve Phase, SPC is used to verify process improvement and stability of our change.
- As part of Control Phase, SPC is used to monitor ongoing performance of the Y and to control the Xs.

Using SPC in Measure Phase

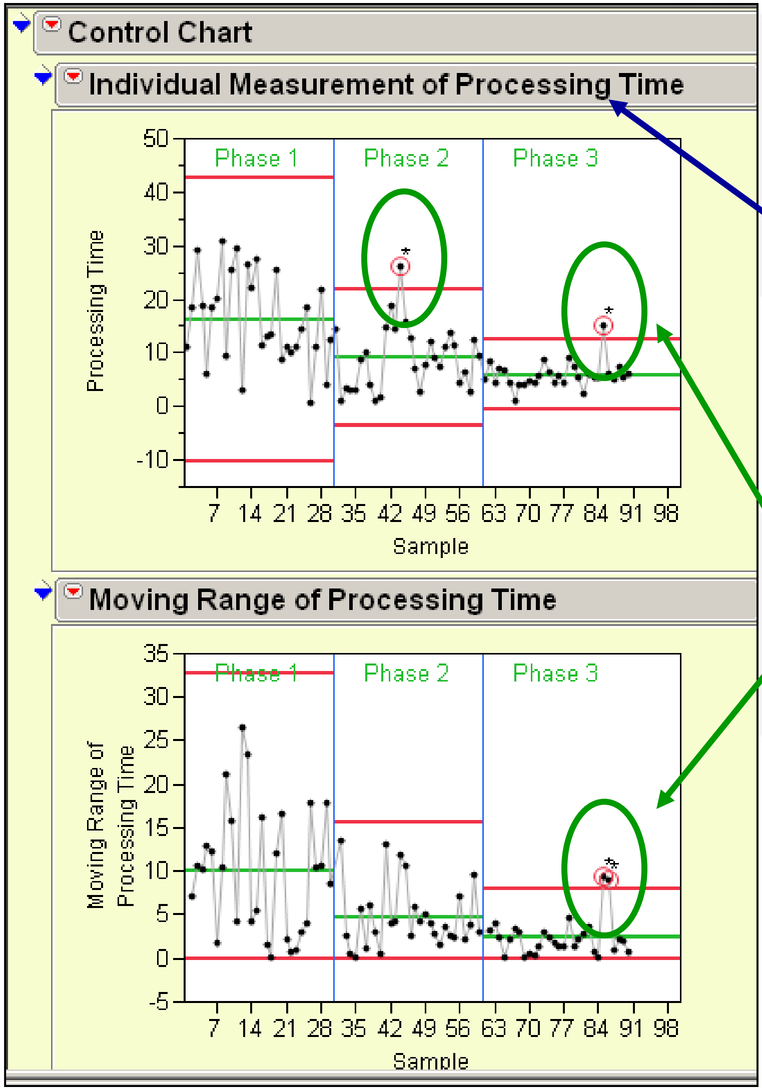
Evaluate process stability during the Measure Phase.



- If the current process is not stable, analysis might lead us to wrong conclusions.
- **Process capability conducted on an unstable process is not valid**.
- **Investigate** for causes of variation and **remove special causes** of variation.

Using SPC in Measure Phase

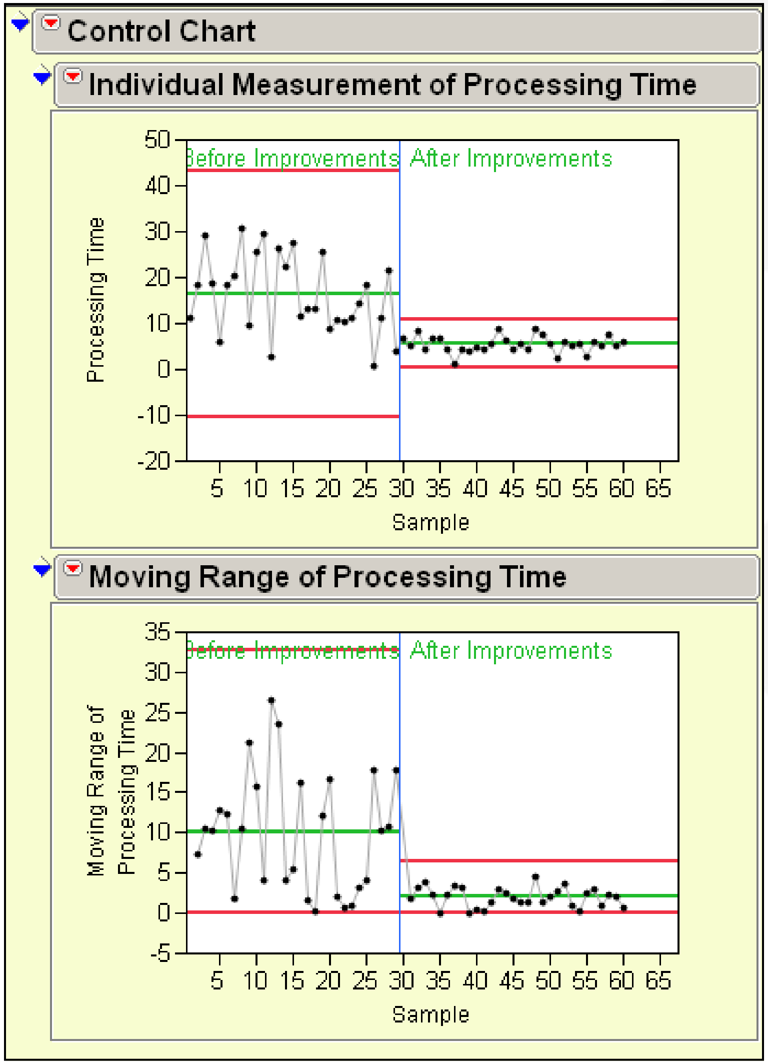
Results of the investigation:



- This process has changed twice during the data collection period.
- Only use the data since the last process change to calculate process capability.
- This process is out-of-control. Analysis revealed assignable variation.
- If the assignable condition can be removed from the process, the subgroup can be excluded when calculating process capability.

Using SPC in Improve Phase

During Improve Phase, use SPC to:



- Analyse process stability.
- Visualise process improvement.
- This chart shows us that our change has reduced both the variation and average of the process.
- Further the new process is showing statistical stability.

Using SPC in Control Phase

During the Control Phase, use SPC:

- To verify sustained stability of the improved process.
- As an integral part of a Control Plan.
  - The first step is to use statistical techniques to monitor process outputs (the Ys).
  - The true value of SPC is to use statistical techniques to control the critical inputs (Xs). Controlling inputs results in increased quality, increased productivity, and lower costs.

Types of Variation

- **Common Cause Variation:**
  - The natural variation of any process. The little ups and downs of the output of the process.
  - Exists in every process.
  - Caused by the combination of all the common cause variation of all the Xs.
  - We would like to minimise this variation, but generally speaking, this is the variation we have to live with.
- **Special Cause Variation:**
  - The unnatural, unexpected variation that can come into any process.
  - Caused by unique disturbances.
  - The goal is to find the cause of the variation and either **eliminate** it or, if it was variation in a good direction, **incorporate** it into the standard process.

Process Stability

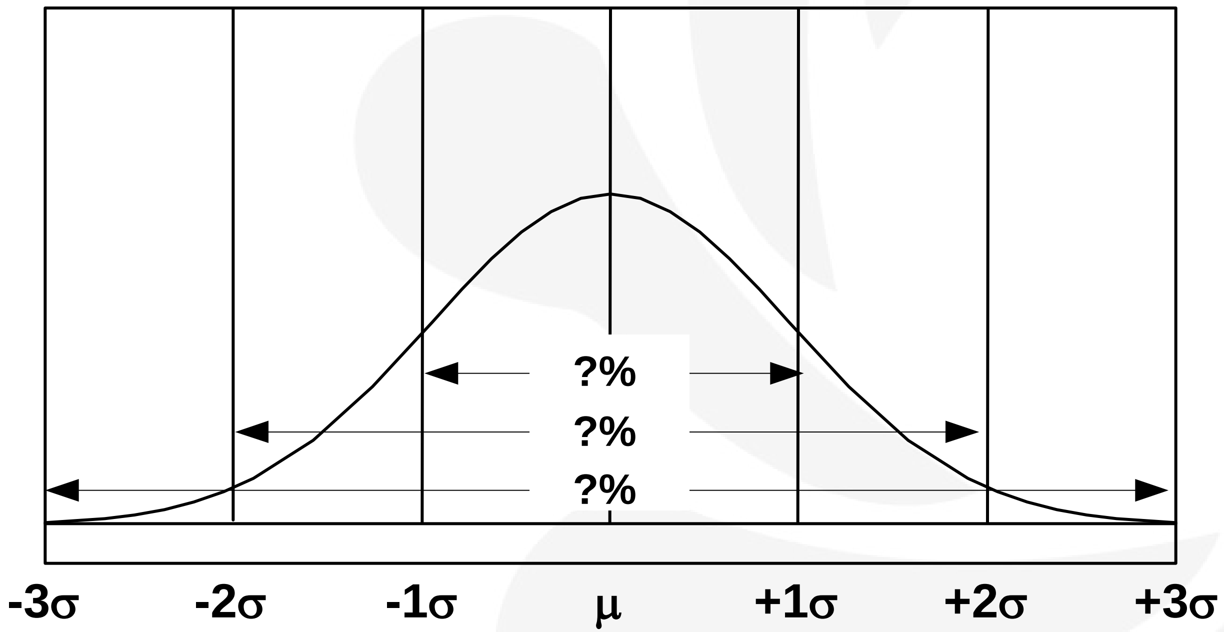
- Process stability is defined as the state where only common cause variation exists in the process.
- A Statistical Control Chart is used to determine whether a process is stable or unstable with respect to its variation.
- A Control Chart is a graphical representation of the process over time and the use of calculated “control limits” allows us to interpret whether the process variation is random (common cause) or non-random (special cause).
- Synonyms:
  - Stable
  - In control
  - Random variation
  - Predictable
  - Only common cause variation present.

Why is it important?

How you interpret variation . . .

True variation type...		Common Causes	Special Causes
	Common Causes	**Focus on systematic process change.**	**Mistake 1: Tampering. (increases variation)**
	Special Causes	**Mistake 2: Under-reacting. (missed prevention)**	**Investigate special causes for possible quick-fixes.**

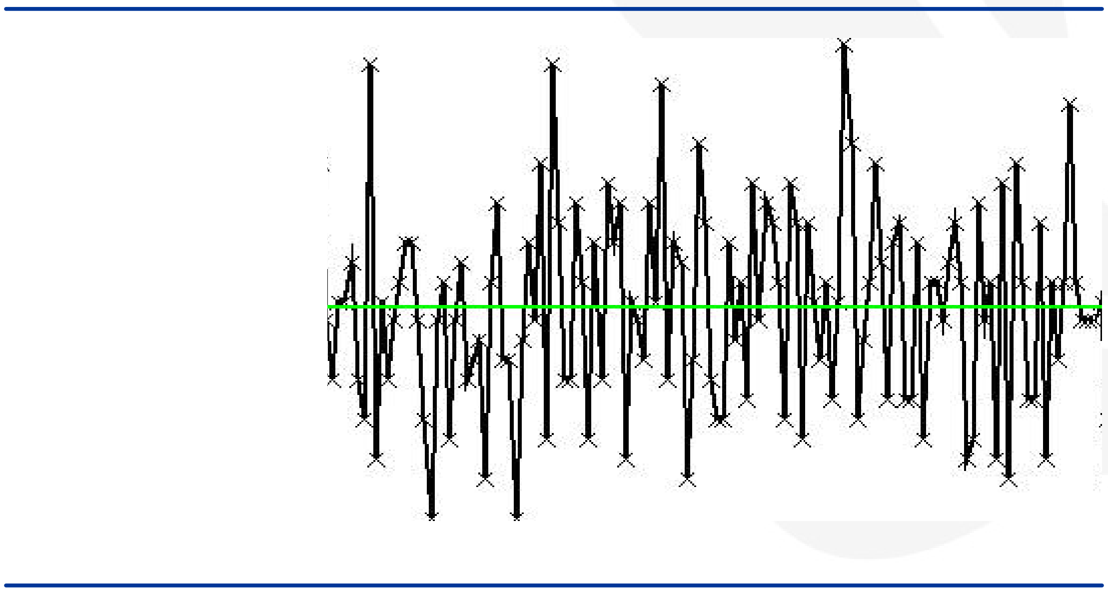
Control Chart Theory



For a normal distribution, what are the percentages of data that will fall between these standard deviation limits if no special causes are operating?

15

Control Chart Theory



- **If a data point falls outside the +/- 3 sigma limits, what is the probability that the reason it is there is because a special cause was operating?**

- **This is Control Chart theory in a nutshell.**

- **What would happen if your data were not normally distributed?**

Control Chart Theory

- **How many times in a row** could I flip a coin and get HEADS before you got **suspicious** that the coin was not fair?
- **How many times** could a process measure **above (or below) average** before you got **suspicious** that some special cause was operating?
- The last rule of control chart theory is that even if a process stays inside the control limits, if it behaves in a **non-random** fashion, the process is also out of control.

Control Limits vs. Spec Limits

Control Limits

- Defined based on process performance (+/- 3 estimated standard deviations from the mean).
- Help determine if your process is in control.
- Plotted on control charts.
- Change when there is a verified, significant change to your process.
- Represent the voice of the process.

Customer Spec Limits

- Defined based on feedback from the customer(s).
- Help determine if your process is producing defects.
- Plotted on histograms (never on control charts).
- Change when your customers say they do!
- Represent the voice of the customer.

Basic SPC

- Determine what will be controlled.
- Decide what data to collect.
  - Establish process measures
  - Determine sub-grouping strategy
- Determine which control chart to use.
- Collect and plot the data.
- Interpret the output.

Rational Subgrouping

- **Rational subgrouping** is the sampling strategy we use to sample our processes for SPC and control charts.

- "Rational" means that we have some rationale for how we collect and subgroup our data

- "Subgrouping means that when we take a sample, we collect a small group of items from the process, a subgroup.

- Three rationales:
  - How often do we sample?
  - How many do we sample?
  - Where in the process do we sample?

- By setting up the proper rational subgrouping strategy, I will be able to answer the questions I need to answer to control my process.

Rational Subgrouping

Subgrouping principles:

1. Never knowingly subgroup **unlike** things together. E.g. don't group securities trades with commodities trades.
2. Minimise the variation **within** each subgroup. E.g. take consecutive items to minimise any changes over time.
3. Maximise the **opportunity for variation between** the subgroups. E.g. if we want to know if the process changes from hour 1 to hour 2, take a sample at the start of each hour, not a random sample of the whole hour.
4. **Average (i.e. X-bar) across noise, not across signals**. E.g. again if we average readings through hour 1, we may be averaging out the signal that there was a change between hour 1 and 2.
5. If single values are collected, then use a subgroup size of 1, don't force a larger subgroup.
6. Establish **operational definitions** for the sampling procedure as well as the operational definitions for measurement procedure.

Exercise 1: Rational Subgrouping

Sampling

Objective: Develop sampling strategies

- Develop a sub-grouping strategy for processing time for loans. You are focusing on a single loan processing location with 7 teams. Each team has 8 processors.
  - How often would you gather samples?
  - How many samples would you take?
  - Where would you sample?

10 minutes

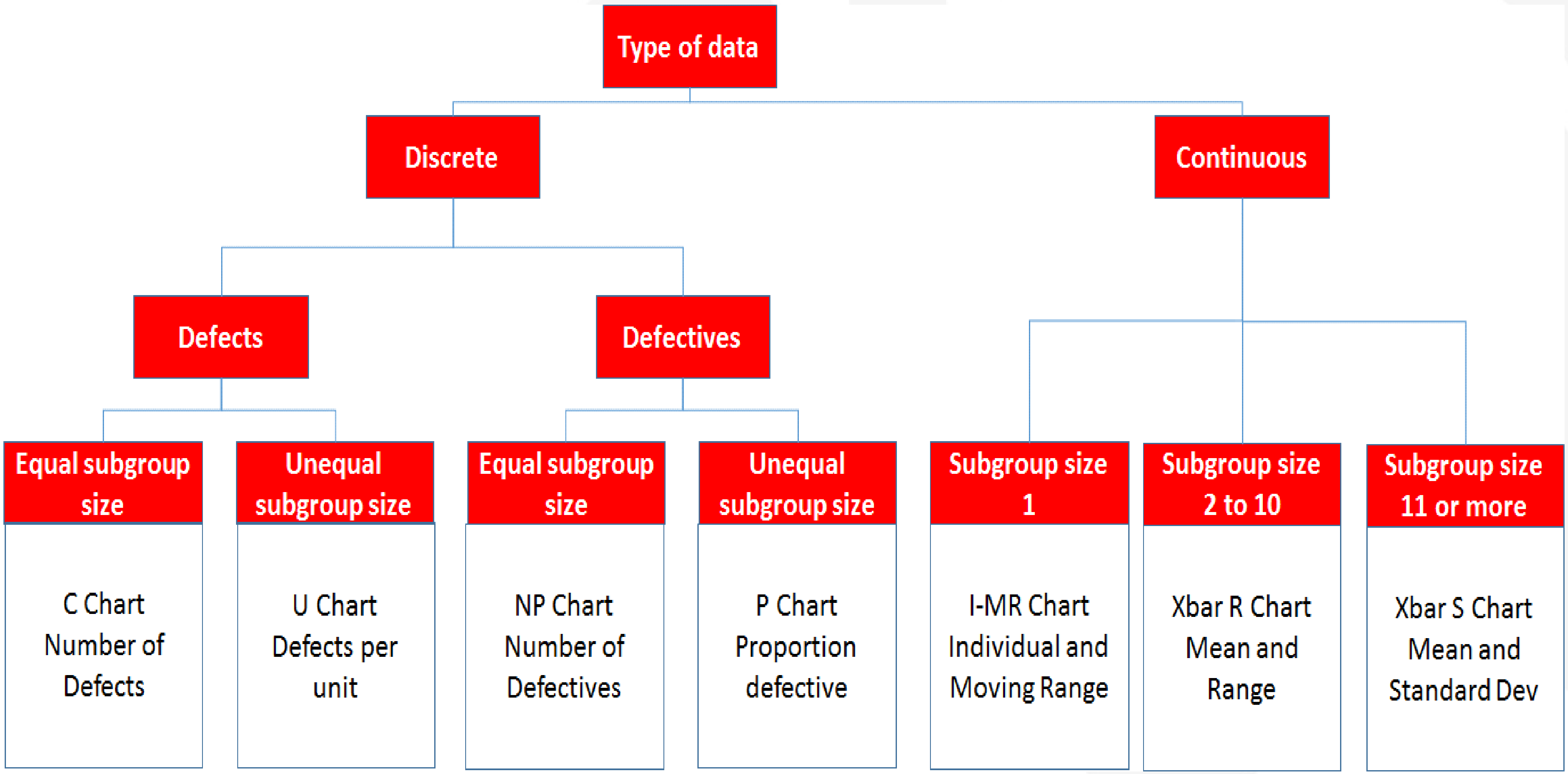
- Assume an investment house has 10 Securities Trading Sites. Our particular site has 35 traders. Each trader will do dozens of trades each day. Each trade takes a certain amount of **time** to complete and that time is recorded.
  - If the investment house wanted to monitor whether **the time it took to complete trades was stable**, what would be a good subgroup strategy?
  - If we wanted to monitor whether the completion time for our site was stable, what would be a good subgroup strategy?

Control Chart Types

- There are many types of control charts; however, the underlying principles of each are the same.

- The chart type selection depends on:
  - Data type: discrete vs. continuous?
  - If discrete data: defects or defectives?
  - Subgroup size: constant or variable?
  - If continuous data: 1 chart for **accuracy**, 1 chart for **precision**.

Control Chart Selection

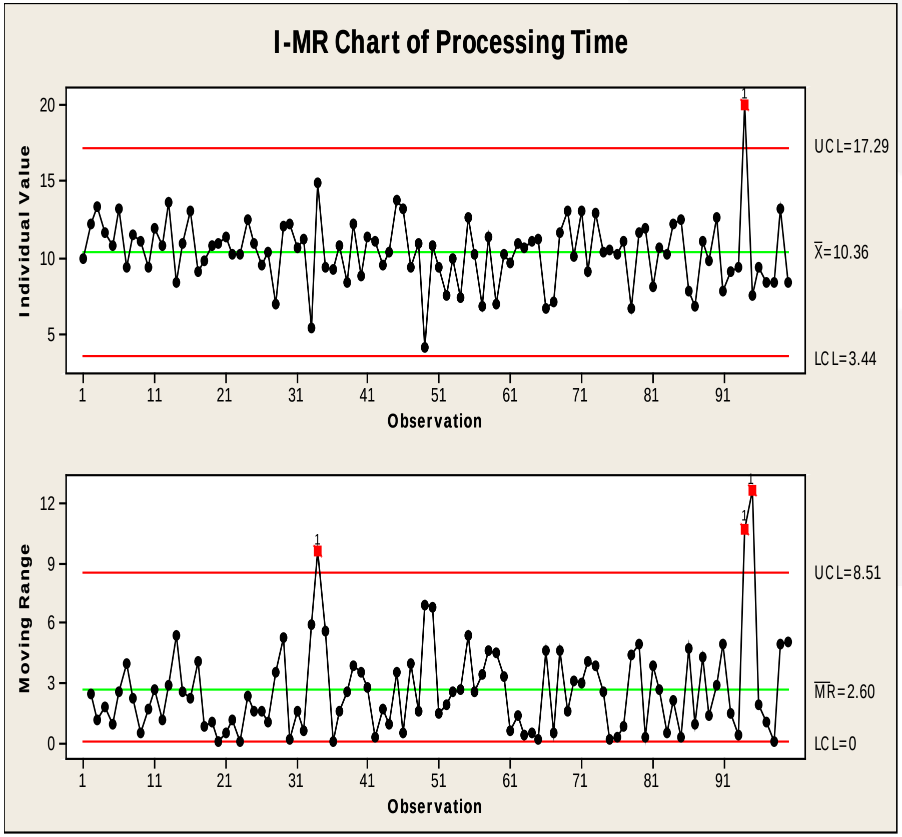


Subgroup Size and Number of Subgroups

- There is a BIG difference between **subgroup size** and the **number of subgroups**.
- Every time we go to our process and collect a "sample", we have one subgroup. When we collect this "sample", how many measurements do we take at one time? That is our subgroup size.
  - My rational subgroup is 5 wait times every hour. Every hour I measure the wait time of 5 customers. Every hour I get one subgroup and the subgroup size is 5. At the end of my 8 hour shift, I have 8 subgroups.
  - My rational subgroup is 100 loan applications every day. Every day I collect 100 loan applications to inspect to see if they are good or bad. Every day I get one sample and the subgroup size is 100. At the end of the week I have 5 subgroups.
  - My rational subgroup is every 30th expense report. Every day I inspect each 30th expense report and count the number of errors. At the end of every day I have one subgroup of variable size. At the end of the week I have 5 subgroups.

I-MR Chart

I-MR Chart (Individual and Moving Range Chart) is appropriate for continuous data and focuses on the variation between individual measures.



Use when:

- There is no basis for subgrouping.
- Each measurement represents one batch.
- Production rate is very slow such that measurements are widely spaced in time.

I-MR Chart

Sample Size = 1

Mean = $\overline{X}$

Control Limits = $\overline{X} \pm 3\left(\frac{\overline{MR}}{d_2}\right)$

$UCL_{MR} = D_4\overline{MR}$

$LCL_{MR} = D_3\overline{MR}$

Sample Size	d_2	D_3	D_4
2	1.128	-	3.267
3	1.693	-	2.574
4	2.059	-	2.282
5	2.326	-	2.114
6	2.534	-	2.004
7	2.704	0.076	1.924
8	2.847	0.136	1.864
9	2.97	0.184	1.816
10	3.078	0.223	1.777

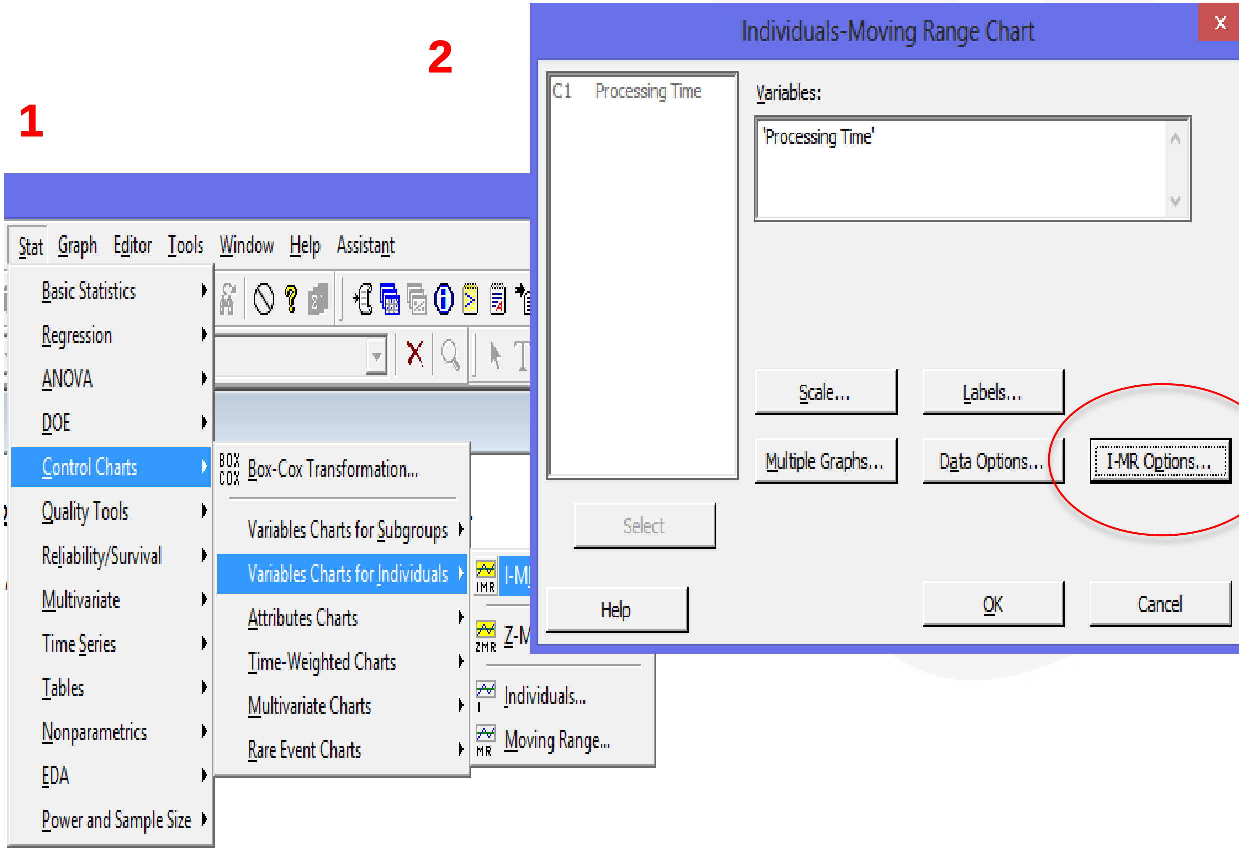
I-MR Charts in Minitab

Stat > Control Charts > Variable Charts for individuals > I-MR

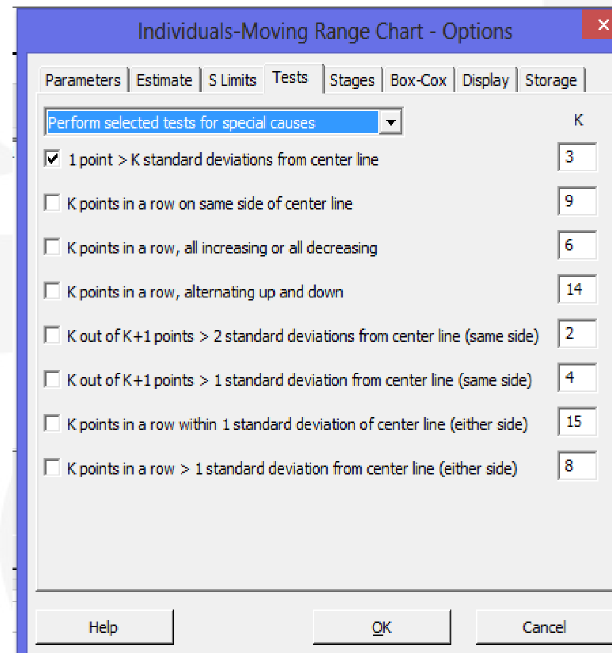
Data Table:
SPC_IMR_Chart_Example.MPJ

1

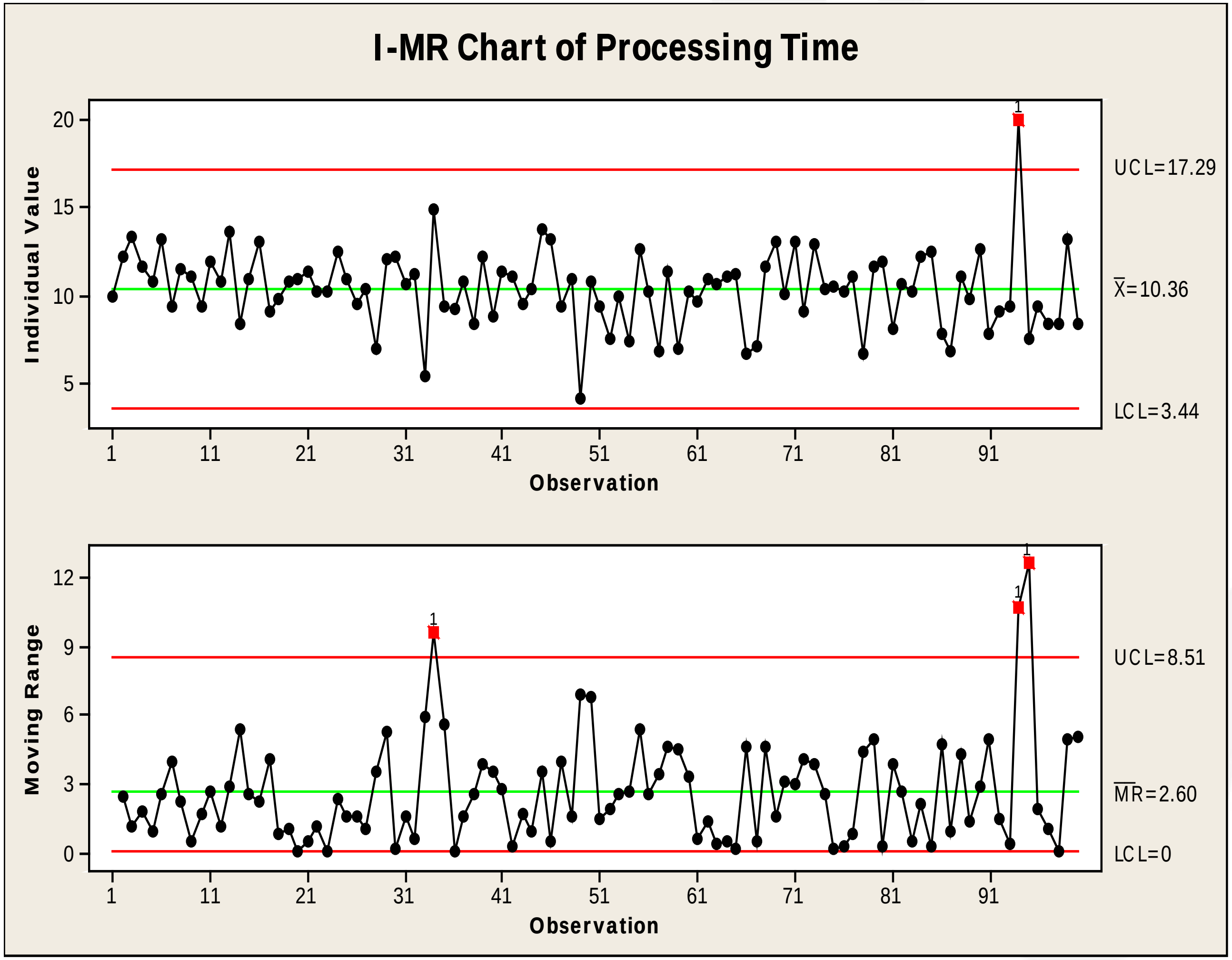
2



3

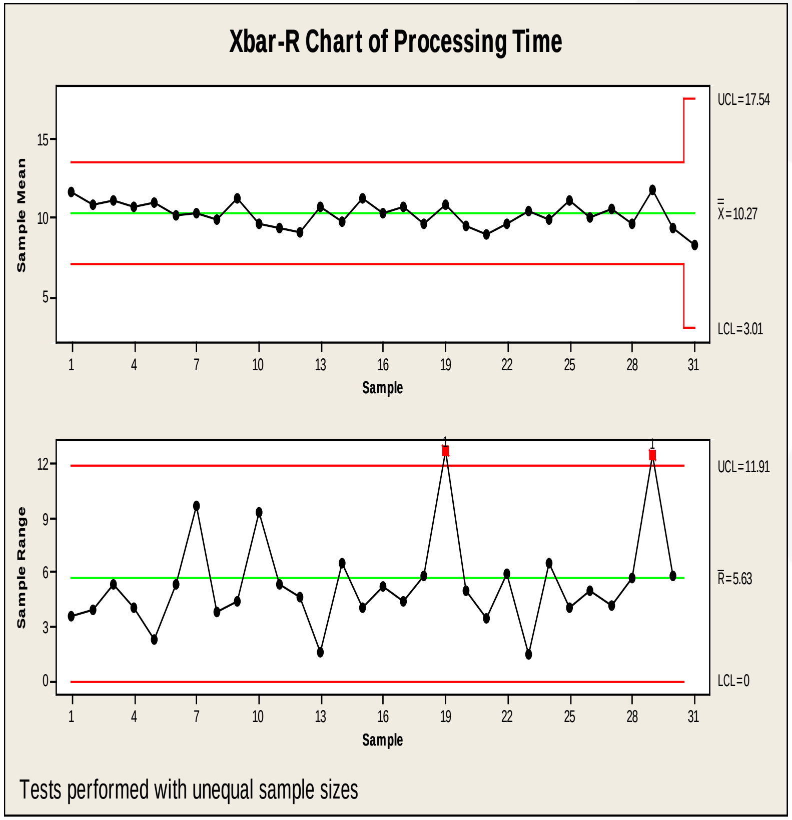


I-MR Charts in Minitab



XBar-R Chart

XBar-R chart is appropriate for continuous data when it is practical to collect frequent samples of subgroups.



R chart:

- Shows **changes** in the **"within"** subgroup variation.
- Asks "Is the variation in the measurements within subgroups stable?"

XBar chart:

- Shows **changes in the average value of the process.**
- Asks "Is the **variation between** the averages of the subgroups more than that predicted by the variation within the subgroups?“
- Very simple, yet sensitive chart for **tracking changes in the mean.**

XBar-R Chart

Mean $= \overline{\overline{X}}$

$CL_{\overline{X}} = \overline{\overline{X}} \pm A_2\overline{R}$

$UCL_R = D_4\overline{R}$
$LCL_R = D_3\overline{R}$

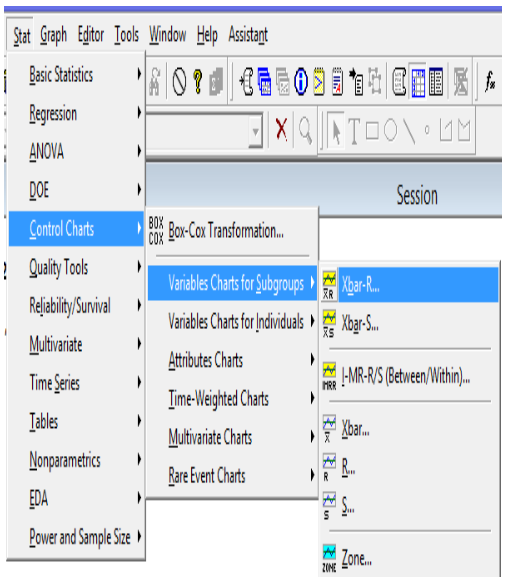
Sample Size	A_2	D_3	D_4
2	1.880	-	3.267
3	1.023	-	2.574
4	0.729	-	2.282
5	0.577	-	2.114
6	0.483	-	2.004
7	0.419	0.076	1.924
8	0.373	0.136	1.864
9	0.337	0.184	1.816
10	0.308	0.223	1.777

XBar-R Charts in Minitab

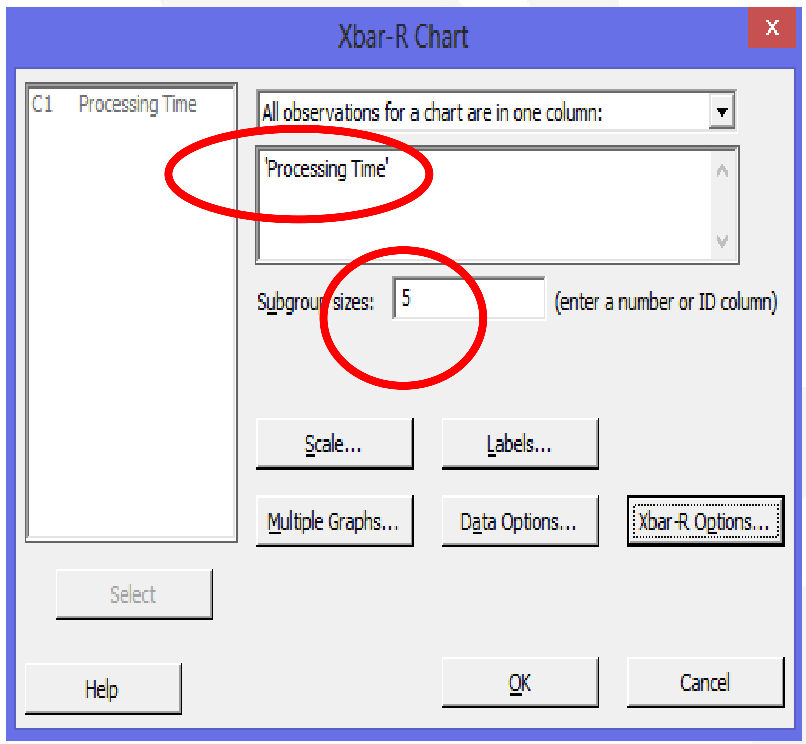
Stat > Control Charts > Variable Charts for Subgroups > Xbar-R

Data Table:
SPC_Xbar_R_Chart_Example.MPJ

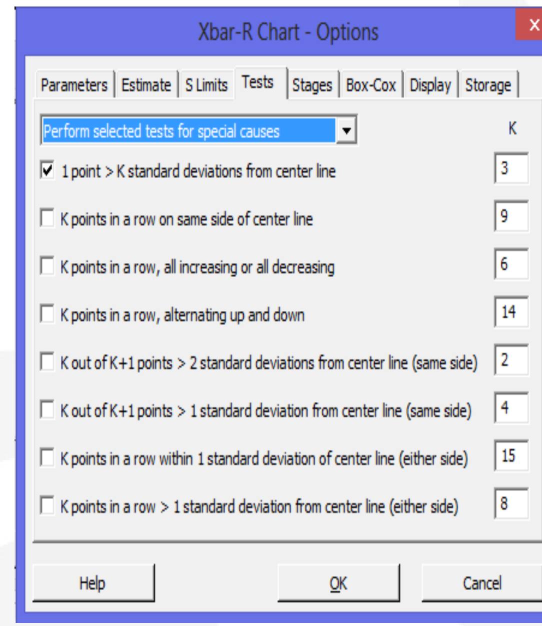
1



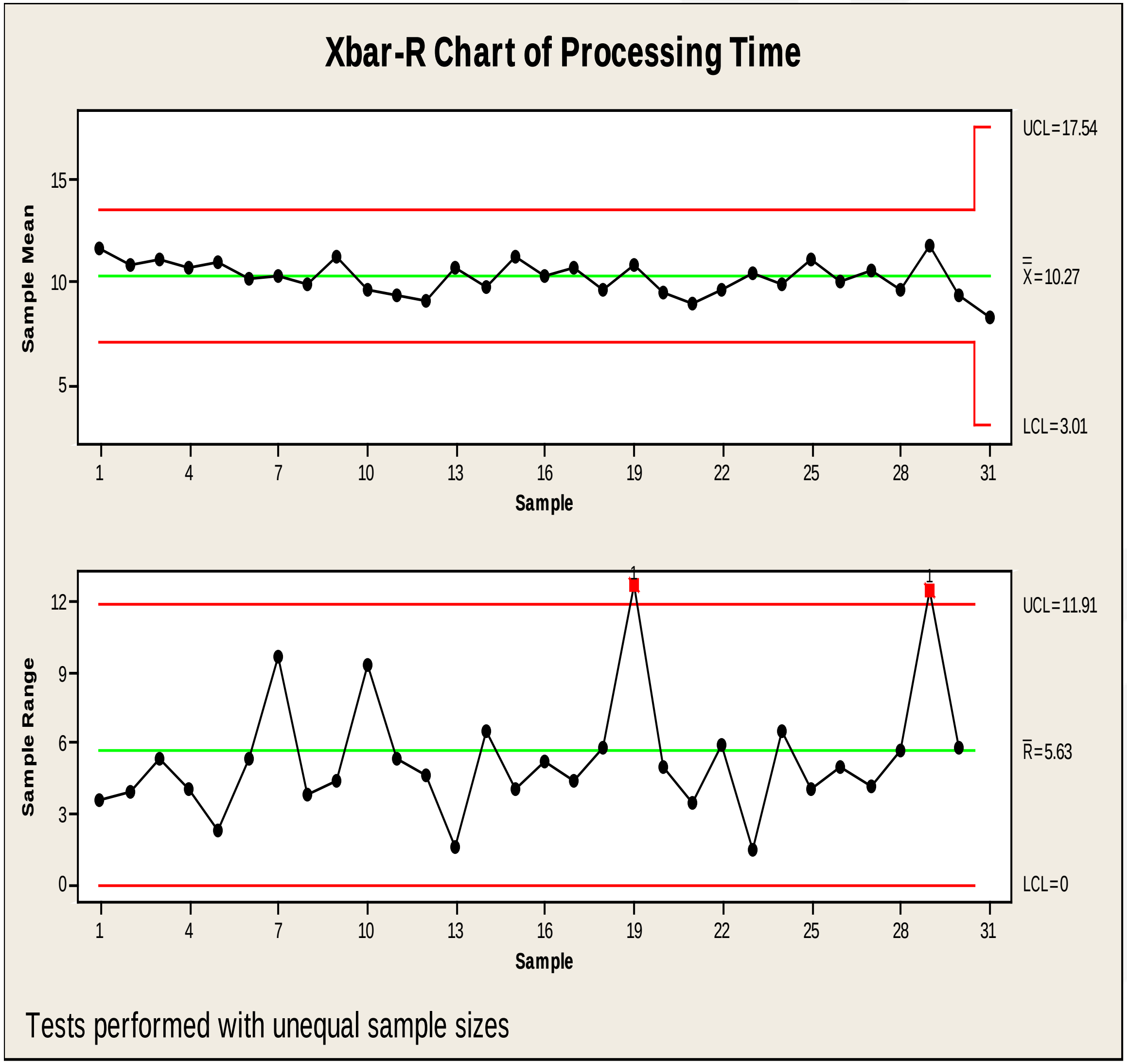
2



3

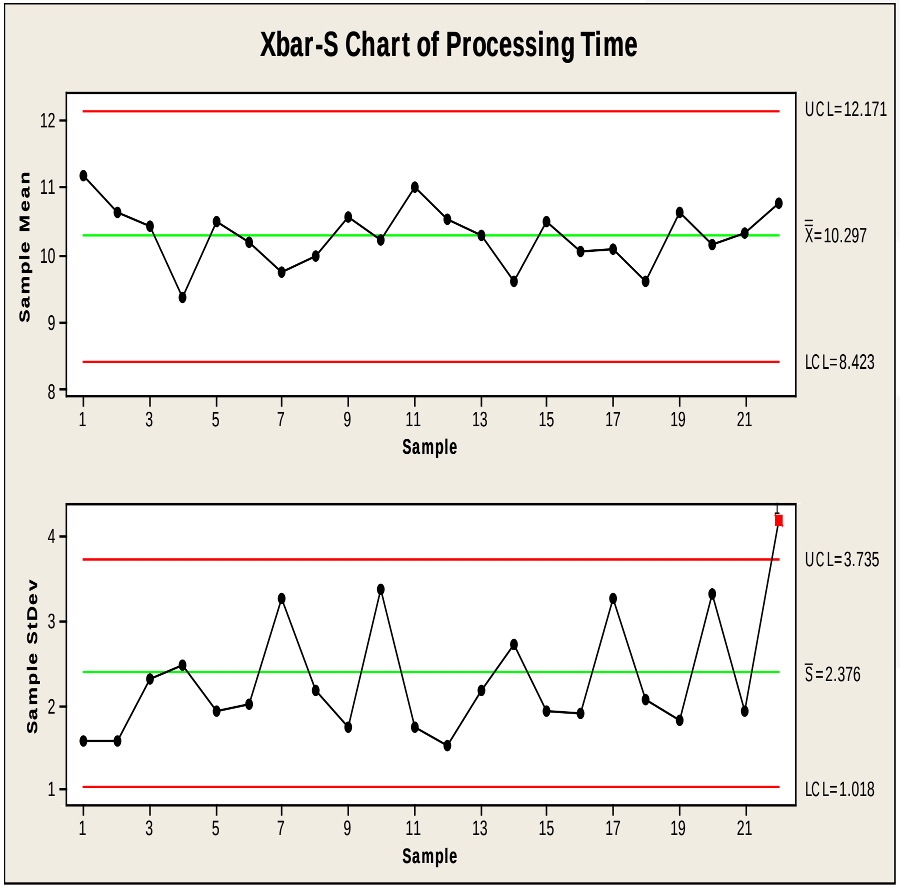


XBar-R Charts in Minitab



XBar-S Chart

XBar-S chart is appropriate for continuous data when the subgroup size is greater than 10.



S chart:

- Plots the subgroup standard deviations.
- Shows changes in the "within" subgroup variation.
- Asks "**Is the variation** in the measurements **within** subgroups stable?"
- Provides a **more precise estimate of process variation** than a R chart when the subgroup size is over 10.

XBar chart is the same as in the XBar and R charts.

XBar-S Chart

Mean $= \bar{\bar{X}}$

$CL_{\bar{X}} = \bar{X} \pm A_3\bar{s}$

$UCL_s = B_4\bar{s}$

$LCL_s = B_3\bar{s}$

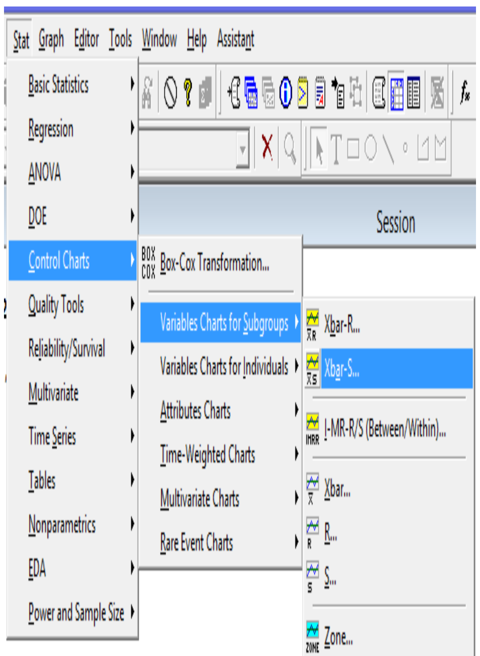
Sample Size	A_3	B_3	B_4
10	0.975	0.284	1.716
11	0.927	0.321	1.679
15	0.789	0.428	1.572
20	0.68	0.51	1.49
25	0.606	0.565	1.435

XBar-S Charts in Minitab

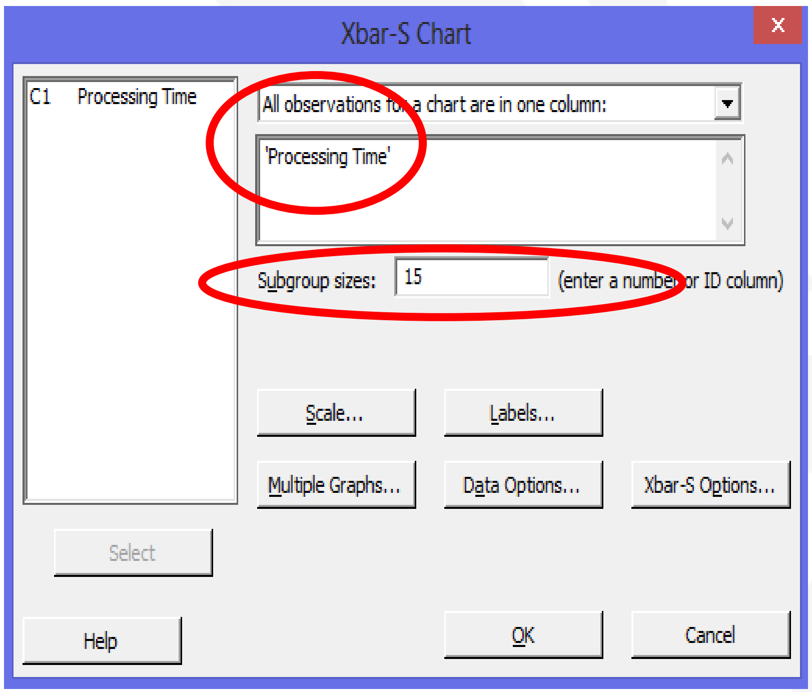
<u>Stat > Control Charts > Variable Charts for Subgroups > Xbar-S</u>

Data Table:
SPC_Xbar_S_Chart_Example.MPJ

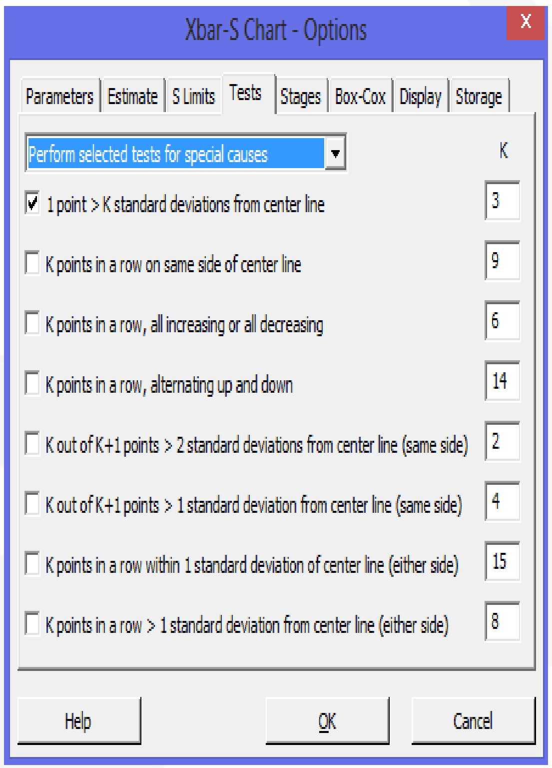
1



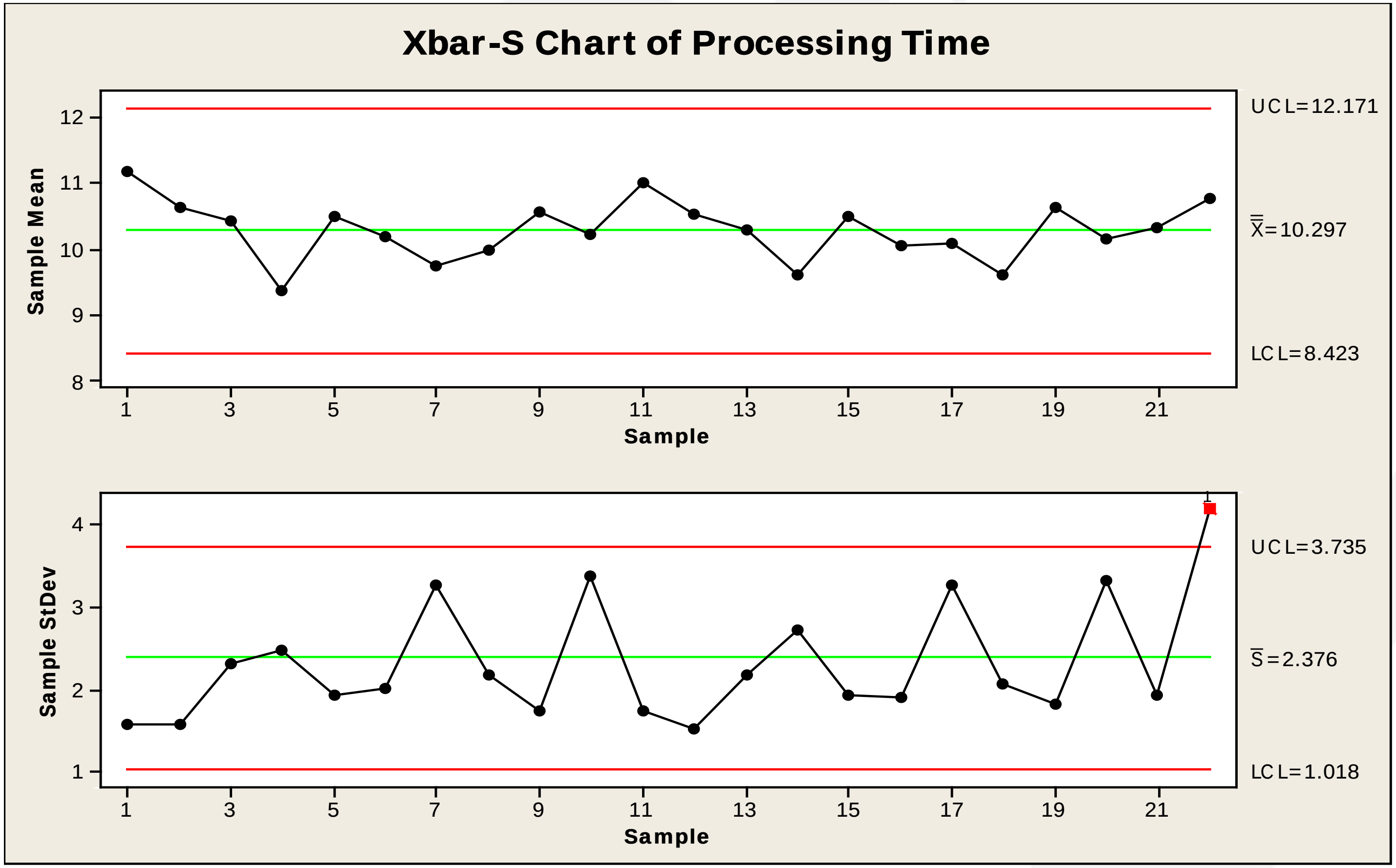
2



3



XBar-S Charts in Minitab



Control Charts for Discrete Data

Discrete data is based on counts or proportions of characteristics for individual items.

- For example, the number of Cash machines out of cash, the number of late shipments or the proportion of defective bills issued.
- Based on accept/non-accept decisions.
- May be applied in almost every operation where data is collected, i.e. continuous data can be evaluated as meeting the specification or not.

Defectives or Defects?

- Defective: The product or service, AS A WHOLE, may be unusable to the customer (form, fit, function). The product or service may be defective because of one or more defects. This is a binary decision; two categories, defective or not defective.

- Defect: The product or service while possessing flaws may still be usable to the customer (form, fit, function). Each flaw that does not meet customer specifications is a defect. Each item can have multiple defects and might still be usable. Or an item may have only one defect and because it is unusable, it would also be considered a defective. This is a count of one category, defects.

- Defective and Defect are generic terms: If we categorise as ‘male’ or ‘female’, we can’t say that one is defective. If we count the number of customers that arrive each hour, we can’t say that these are defects. Yet in both cases we could plot our data on control charts.

Defectives or Defects - Examples

- If we inspected 100 credit card application forms, what is the maximum number of defective forms could we have? What chart would we use?

- If we inspected 5 credit card application forms, what is the maximum number of errors we might see on them? Could all of those errors have been on one or two of the applications? What chart would we use?

- For selection of a control chart, are we interested in:
  - Counting how many units are defective? -or-
  - Counting how many defects there are in the measuring space?

Opportunity Space (1 of 2)

- When measuring items to determine if they are defective or not, the opportunity to find a defective is equal to the number of items. So the opportunity is the subgroup size.

- For defect control charts, there is a slightly different interpretation. Instead of looking at subgroup size, we look at **opportunity space**. Much of the time this opportunity is exactly the same as subgroup size.
    - For example, we are counting the number of errors on exactly the same loan application and exactly the same number of loan applications. The opportunity for errors in each subgroup is exactly the same. What chart would we use?
- But sometimes the opportunity is variable.
    - For example we are looking at exactly the same number of applications, but they are slightly different and contain a different number of fields. The opportunity for errors in each subgroup is now variable. What chart would we use?

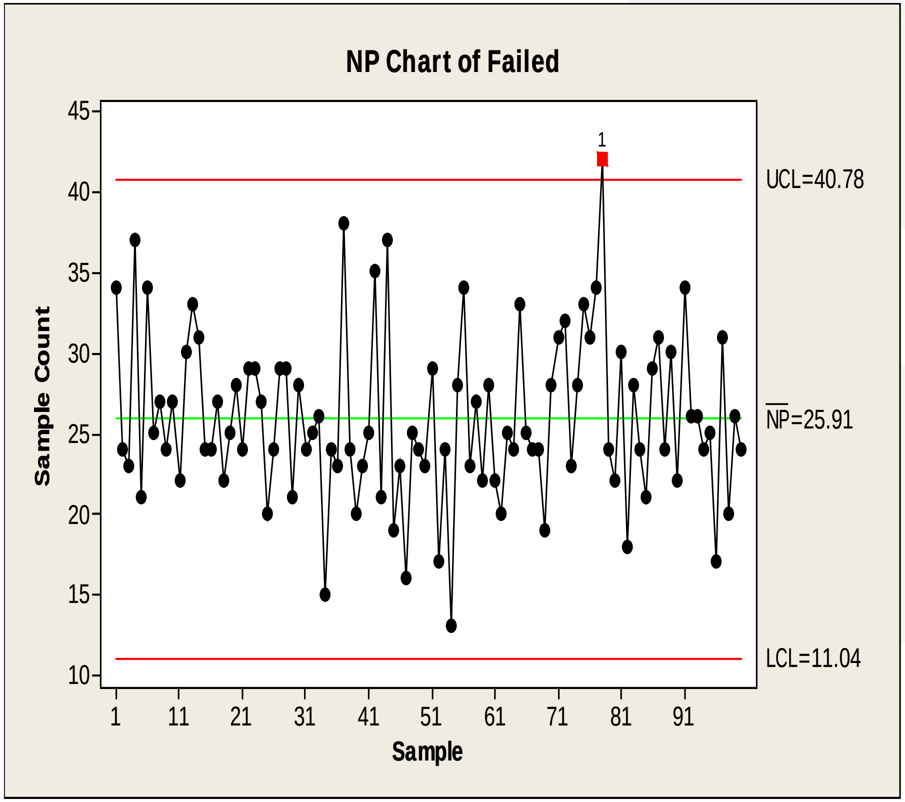
Opportunity Space (2 of 2)

- Another type of opportunity space is **time based subgrouping**. In this case **the area of opportunity I am inspecting is a period of time**.

  - For example I am counting the number of unplanned server outages every day. The subgroup size is one day. Are outages defective or defects? Since I cannot know the total number of possible outages that might occur, I cannot treat these as defectives, because to do so I would have to know the 100% level. I would have to treat these as defects in my constant (equal) area of opportunity, one day. Each day, I would then get one subgroup.
  - Another example: I am counting the number of FCIs (Failed Customer Interactions) every hour. There are two ways to do this. If I count the number of NOT failed interactions, I can do a P chart—defectives per total interactions every hour. This would be unequal subgroup size. But if I choose not to count the NOT failed interactions, I can again treat one hour as the constant opportunity and consider each FCI to be a defect.

NP Chart

Usage: Controlling number of defective units observed when the number of units inspected remains constant.



Assumptions:

- Each sample must have n distinct items.
- Each of the n distinct items must be classified as possessing or not possessing some attribute, typically a type of non-conformance to a specification.
- There is a constant failure probability from sample to sample.
- The items must be independent, that is they do not have non-conformances naturally found in clusters or groups.

NP Chart

Underlying distribution is Binomial.

Binomial Assumptions.

- **Two choices: success / failure, yes / no.**
- **There are n identical trials.**
- **Constant defect probability.**
- **The trials are independent – meaning that one trial does not influence the outcome of any other trial.**

Mean = $\overline{np}$

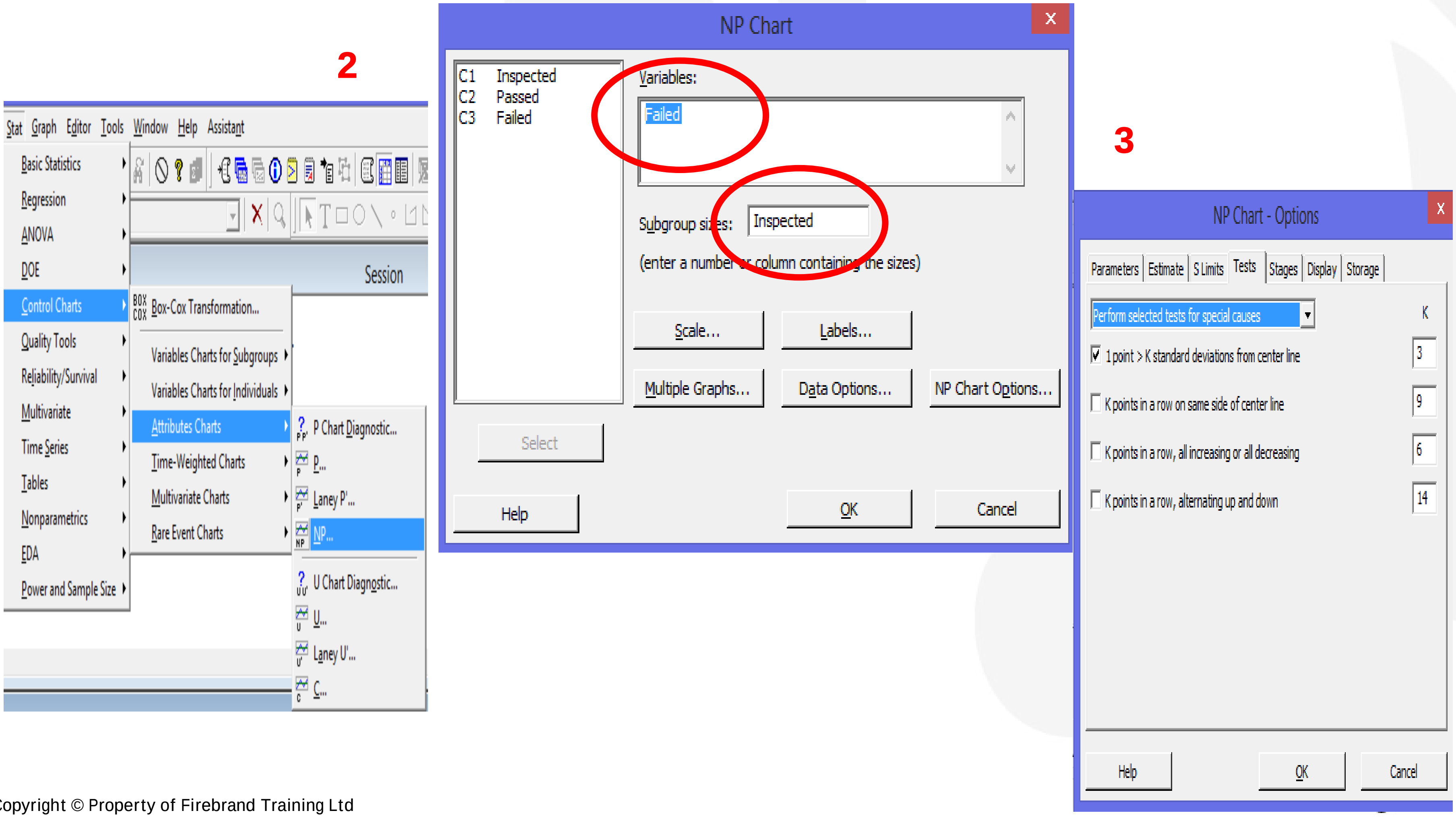
Control Limits = $n\bar{p} \pm 3\sqrt{n\bar{p}(1-\bar{p})}$

NP Charts in Minitab

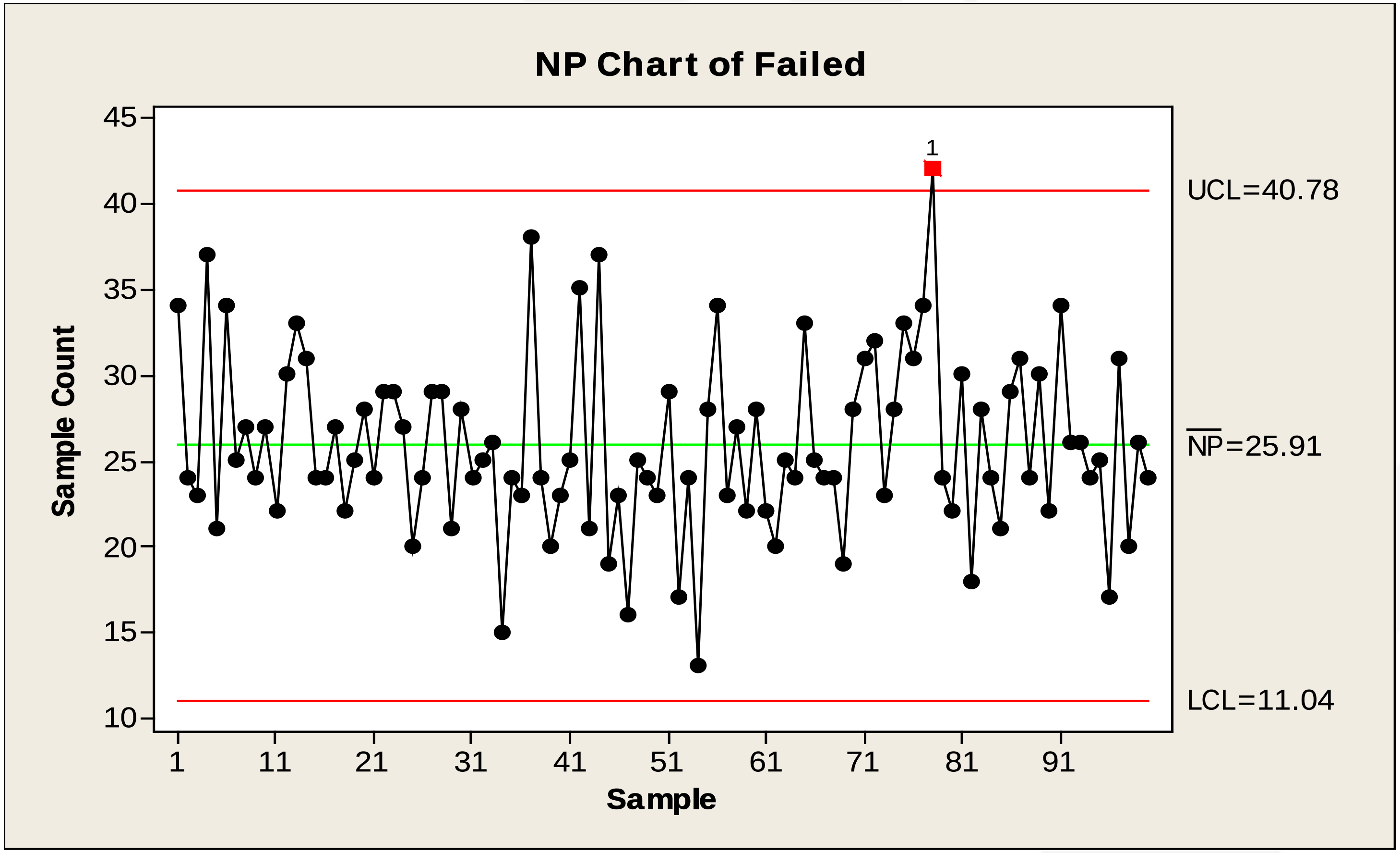
Stat > Control Charts > Attribute Charts > NP

Data Table:
SPC_NP_Chart_Example.MPJ

1 2 3

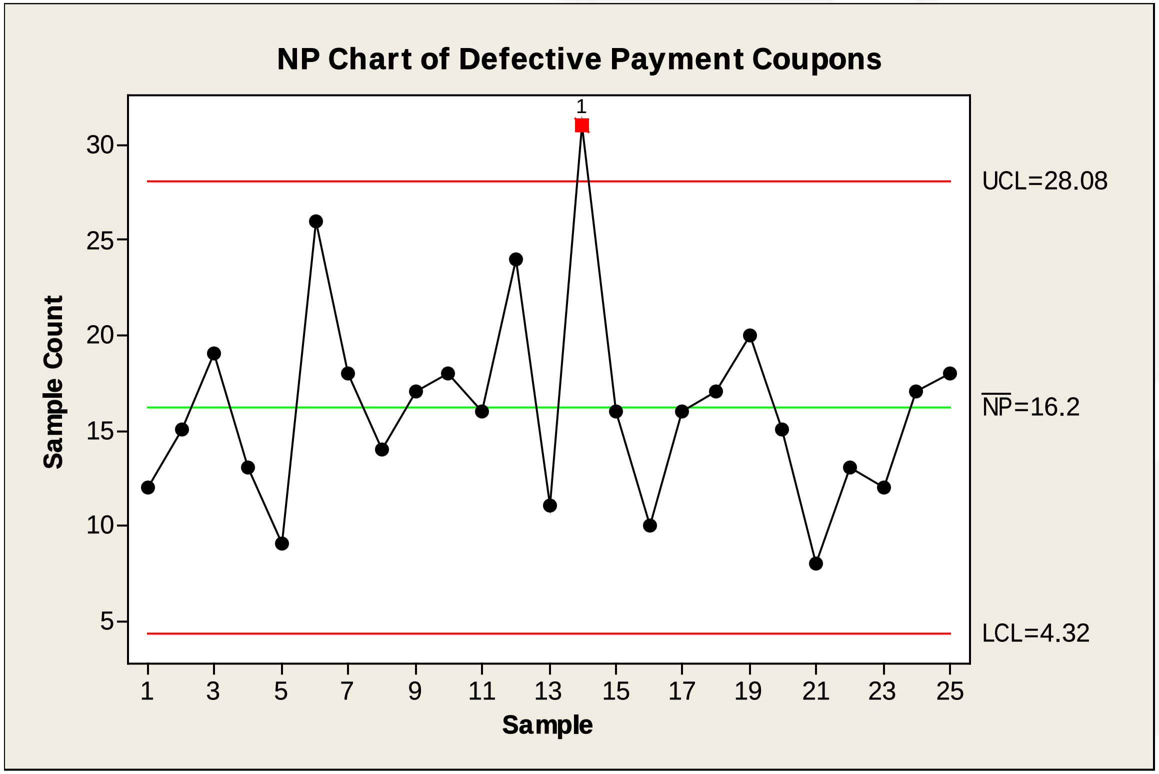


NP Charts in Minitab



NP Chart – Another Example

We have the results of sampling 500 overdue payments daily and determining how many are overdue because an inaccurate coupon was sent to the customer. What do we determine from our control chart? What number would be considered normal for this process? Data Table: SPC_NP_Chart_Coupon_Example.MPJ

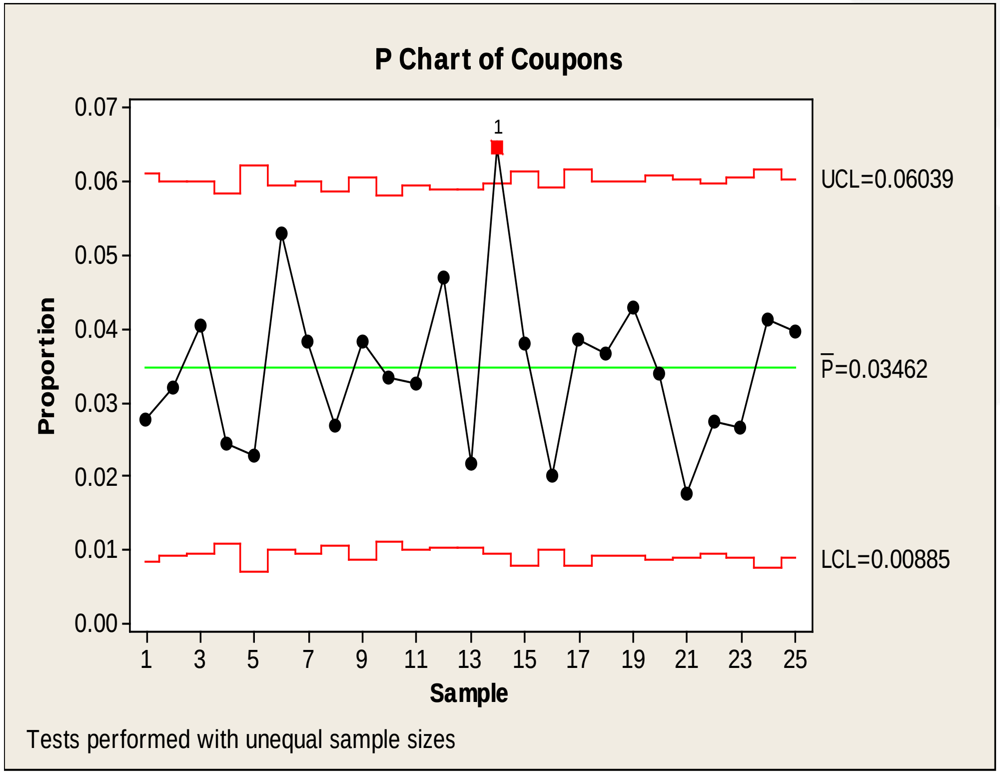


Data Set:
12
15
19
13
9
26
18
14
17
18
16
24
11
31
16
10
16
17
20
15
8
13
12
17
18

P Charts

The P chart is appropriate for discrete data classified into two categories such as good / bad or conforming / non-conforming.

The P chart is used with either a variable subgroup size (different size for each subgroup) or a constant subgroup size.



Assumptions:

- Constant defect rate (p).
- Two choices:
  e.g. success or failure.
- There are n identical trials.
- Trials are independent.

P Charts

Underlying distribution is Binomial.

Mean = $\dfrac{\text{Number of defectives}}{\text{Number of units inspected}} = \overline{p}$

Control Limits = where n_i = number of units inspected

The control limits will vary inversely with the square root of the subgroup size *n*, and therefore will vary from subgroup to subgroup.

$$\overline{p} \pm 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n_i}}$$

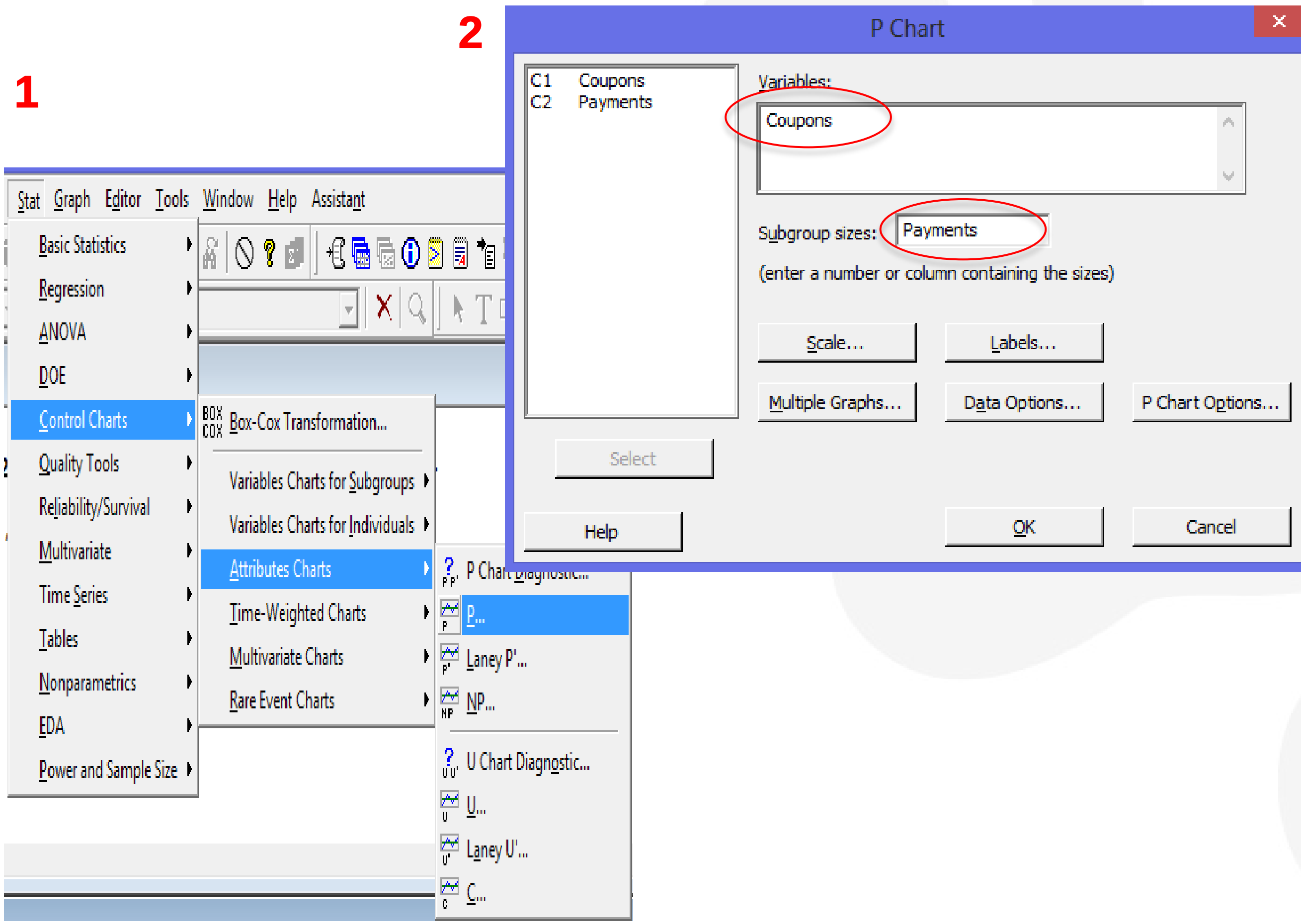
P Charts in Minitab

Stat > Control Charts > Attribute Charts > P…

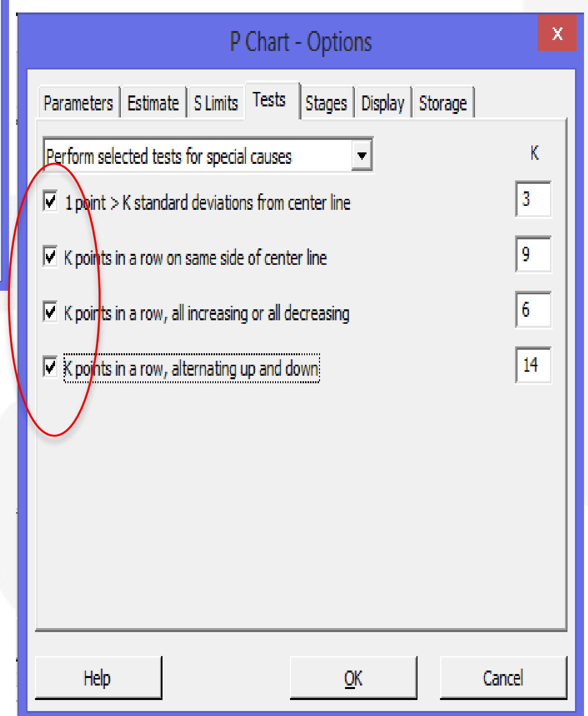
Data Table:
SPC_P_Chart_Example.MPJ

1

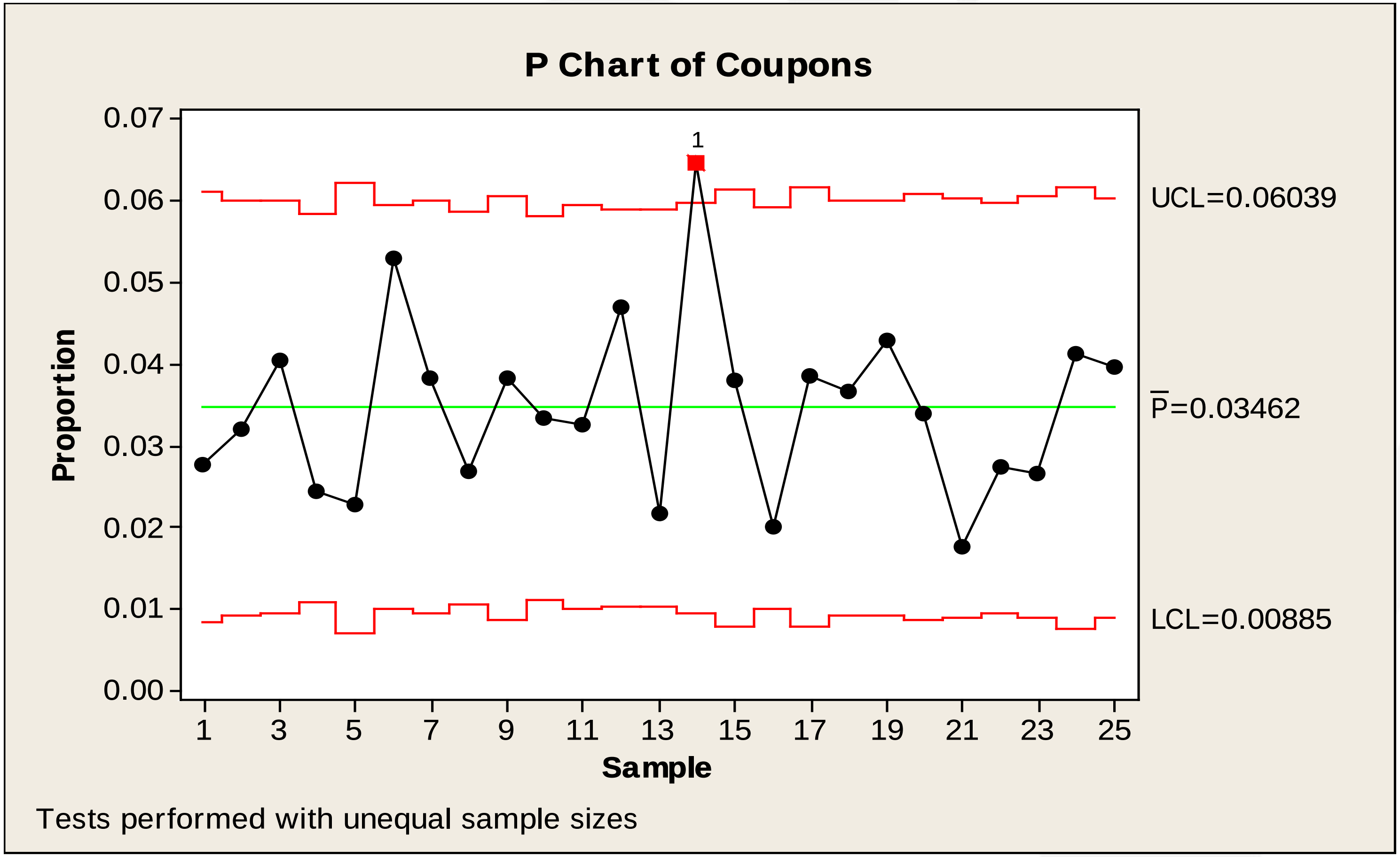
2



3

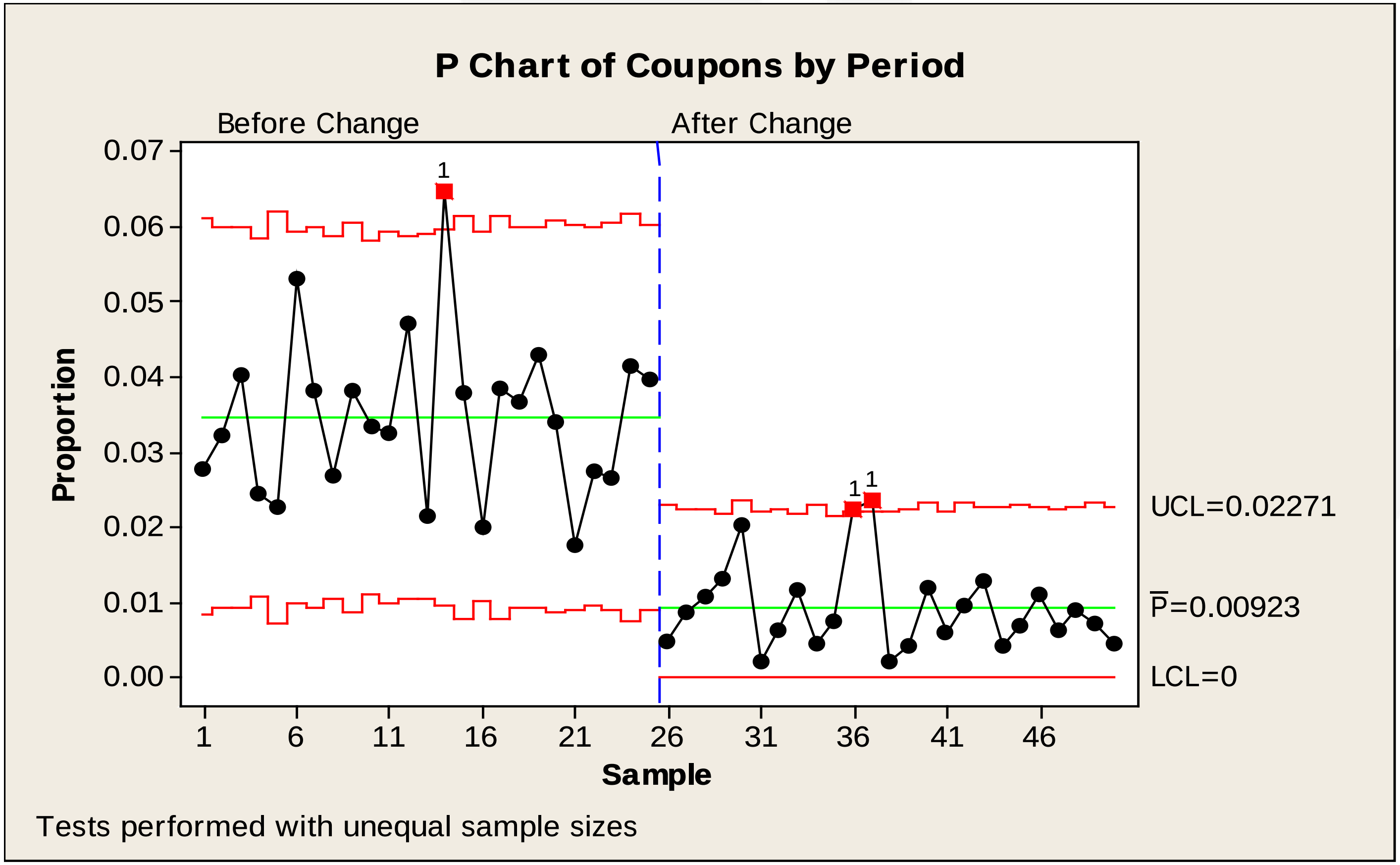


P Charts in Minitab



P Chart – Split Control Limits

Assume a change was made in the process to reduce the number of inaccurate coupons sent to customers. Data was collected for the periods before and after the change.

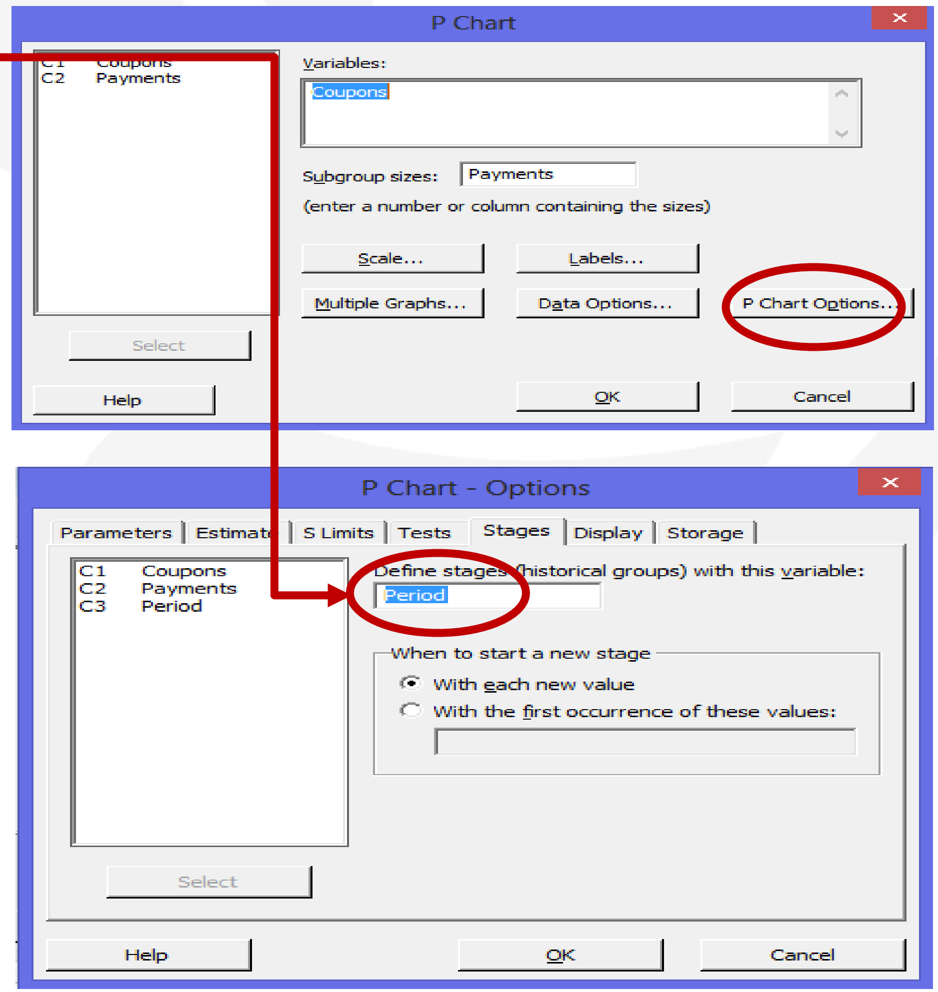


P Chart – Split Control Limits

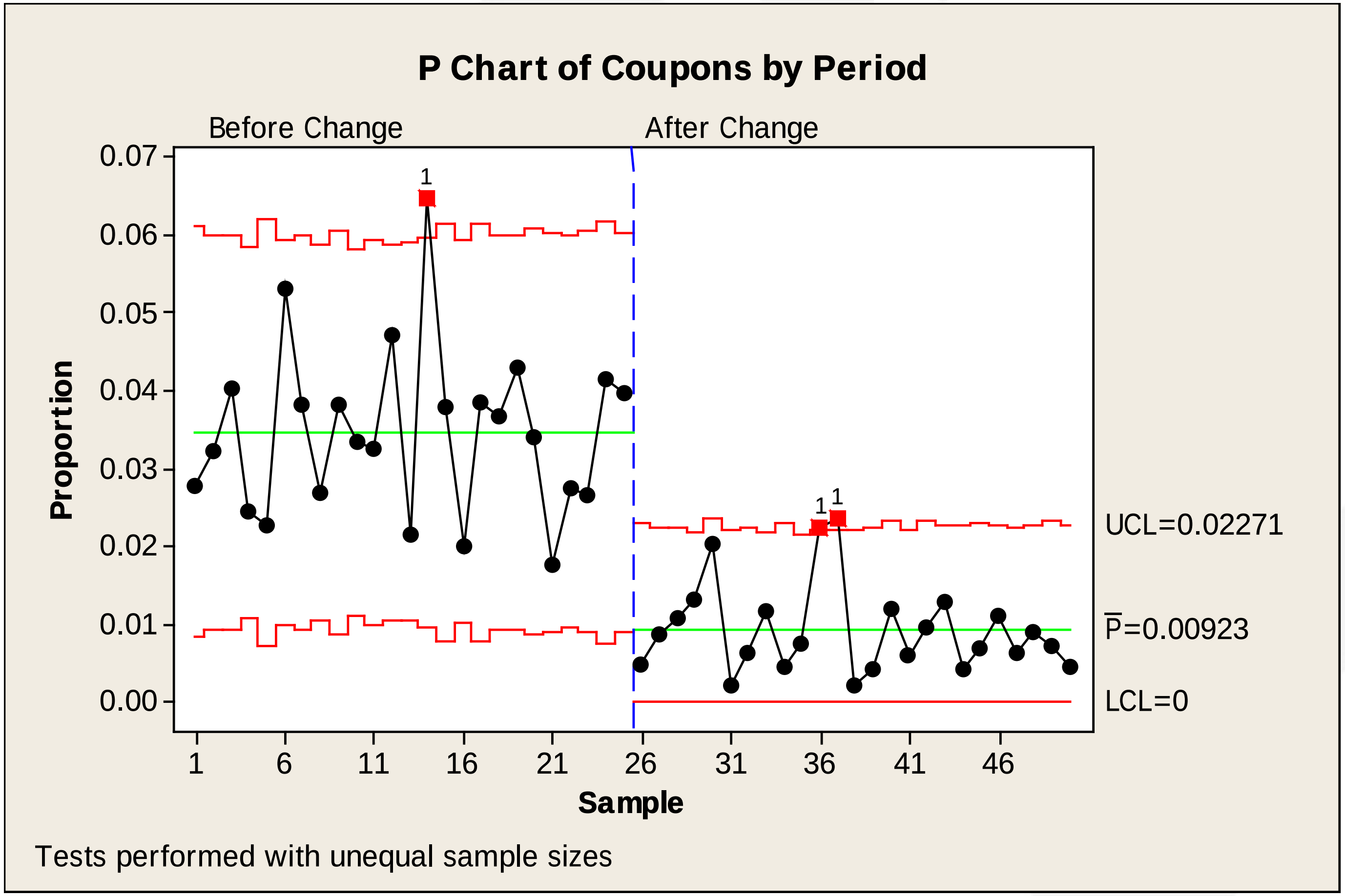
Data Table: SPC_P_Chart_Example_with_Limits_Recalculated.MPJ

Use the stages tab to split time periods

↓	C1	C2	C3-T
	Coupons	Payments	Period
13	11	509	Before Change
14	31	479	Before Change
15	16	421	Before Change
16	10	497	Before Change
17	16	416	Before Change
18	17	465	Before Change
19	20	465	Before Change
20	15	441	Before Change
21	8	454	Before Change
22	13	473	Before Change
23	12	451	Before Change
24	17	411	Before Change
25	18	453	Before Change
26	2	432	After Change
27	4	467	After Change
28	5	471	After Change
29	7	531	After Change
30	8	396	After Change
31	1	491	After Change
32	3	470	After Change
33	6	520	After Change



P Chart – Split Control Limits



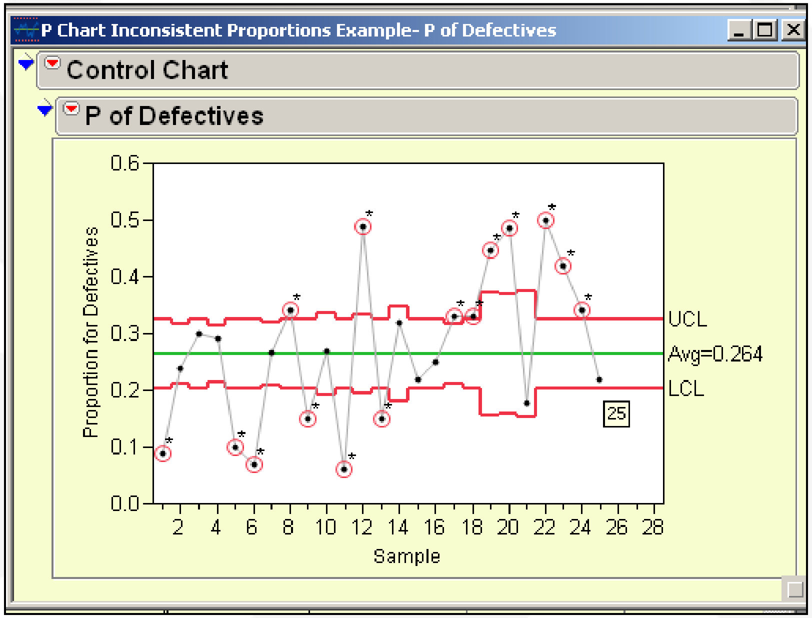
When *NOT* to Use NP and P Charts

1. When the percentages are based on <u>measurement data</u> rather than on count data, the NP and P charts are not appropriate.
   - Use instead either the IR chart or XBar-R chart.
   - Measurement data do not have a discrete area of opportunity for the denominator.
   - E.g.: Square footage is continuous. The average square footage of a house in MyTown is 2500 sq. ft. My house is 2000 sq. ft. My house is 80% of the MyTown average. It is a continuous percentage.

2. When <u>the ratio is not really a proportion</u>, do not use a P chart.
   - An IR chart will be more appropriate.
   - In order to use a P chart, the denominator of the ratio needs to define the area of opportunity for defectives.
   - E.g.: I calculate the "proportion" by dividing today's errors by the average daily production. The denominator is not really the area of opportunity so the ratio is not really the proportion of errors.

When *NOT* to Use NP and P Charts

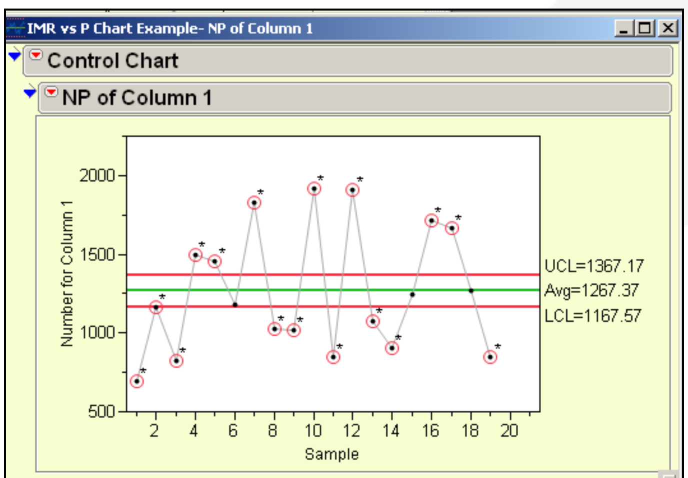
3. When the proportion of defectives is inconsistent.
   - The P chart and NP chart represent a stable proportion of defectives when the process is in control.
   - If the data include different processes, then different charts should be created for each unique process.
   - If the process has not yet achieved a stable state, it is too early to do a P or NP chart.



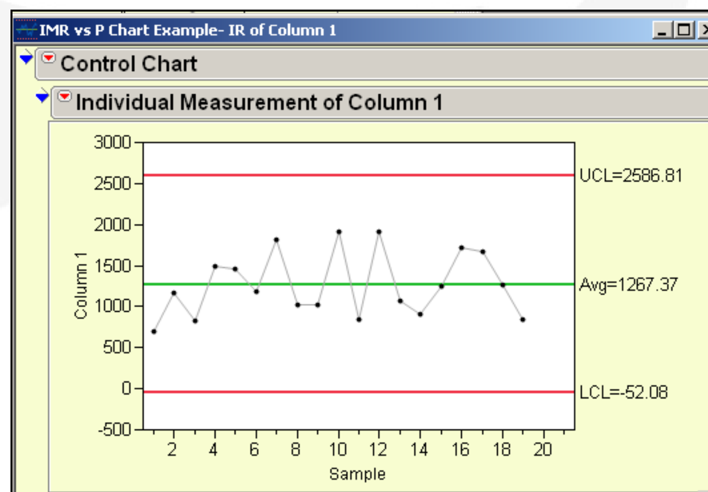
When *NOT* to Use NP and P Charts

4. When the subgroup size becomes very large or the defective proportion becomes very small.
   - The upper and lower control limits will be excessively close together which could create false ‘out-of-control’ alarms.
   - The general guideline is: if the number of samples in a subgroup (n) times the proportion of defect (p) is greater than 5 (np > 5) AND n(1-p) > 5, it is acceptable and many times more informative to use the IR chart.

	Column 1
1	698
2	1163
3	824
4	1493
5	1453
6	1182
7	1825
8	1028
9	1020
10	1917
11	848
12	1910
13	1074
14	909
15	1242
16	1713
17	1666
18	1270
19	845

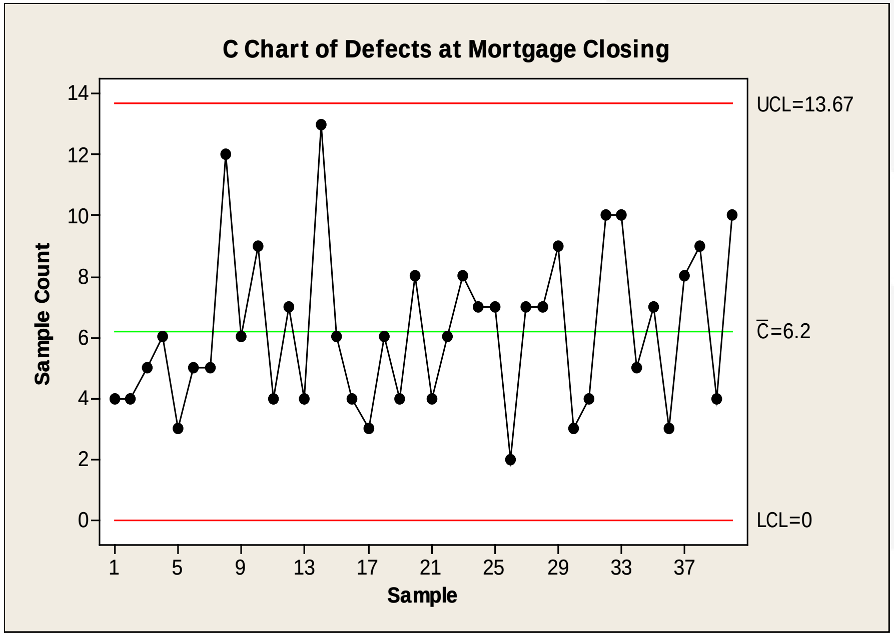


(Subgroup size = 10,000)



C Charts

Usage: Controlling the number of defects observed when the number of units inspected remains constant.



Assumptions:

- Each defect is a discrete event.
- Defects occur in a defined region or area.
- Defects are independent of each other.
- Defects are rare compared to what might be.

C Charts

Underlying distribution is Poisson.

Mean = $\frac{\text{Number of defects}}{\text{Number of samples}}$ = $\bar{c}$

Control Limits = $\bar{c} \pm 3\sqrt{\bar{c}}$

Standard Deviation = $\sqrt{\bar{c}}$

Note: *all* defects for each sample must be counted – not just the first defect found.

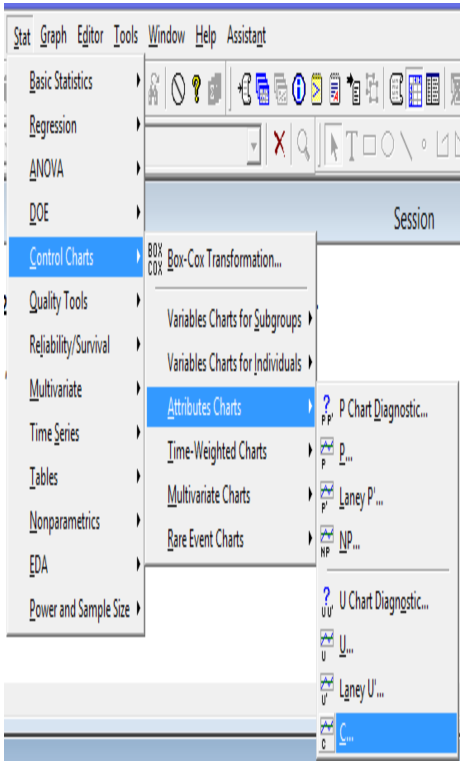
Since the standard deviation is equal to the square root of the mean, the c chart can be used to examine both mean shifts and changes in variation.

C Charts in Minitab

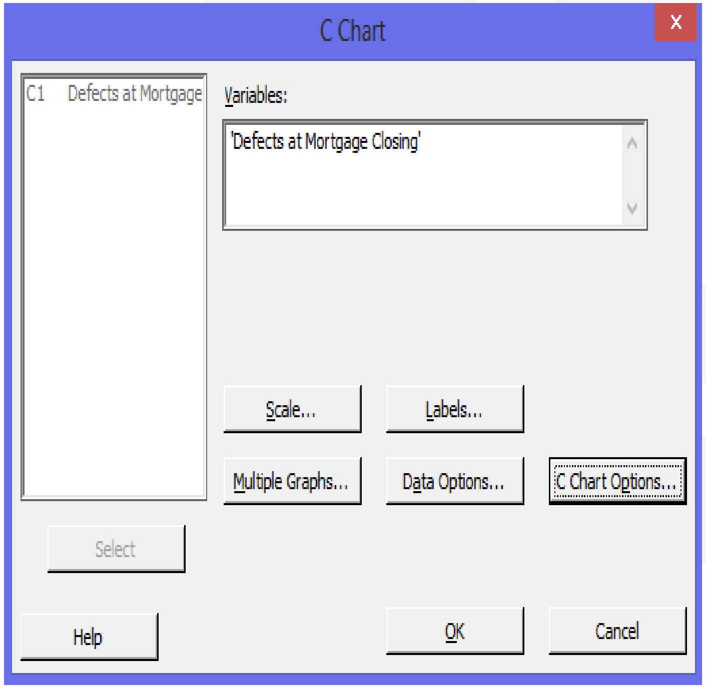
Stat > Control Charts > Attribute Charts > C…

Data Table:
SPC_C_Chart_Example.MPJ

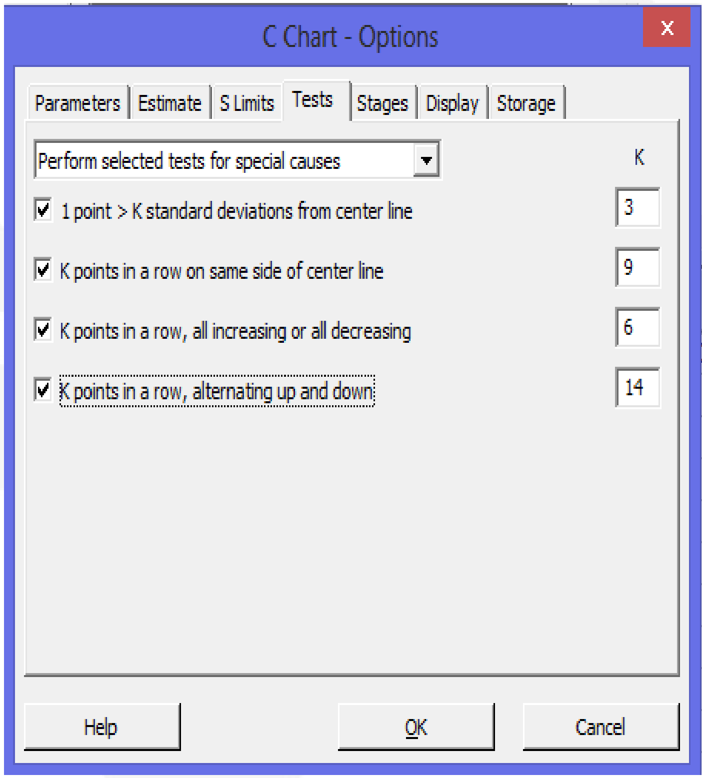
1



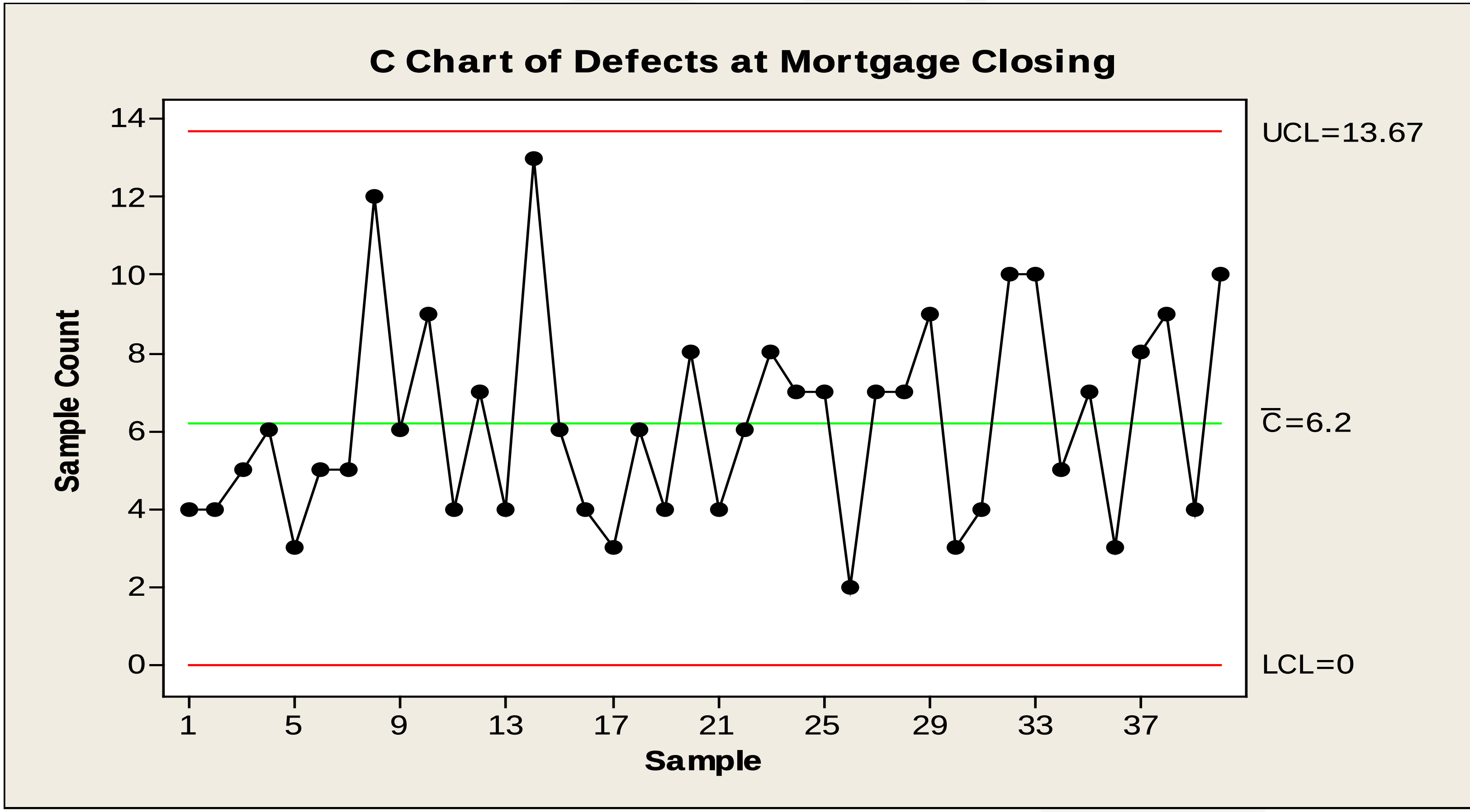
2



3

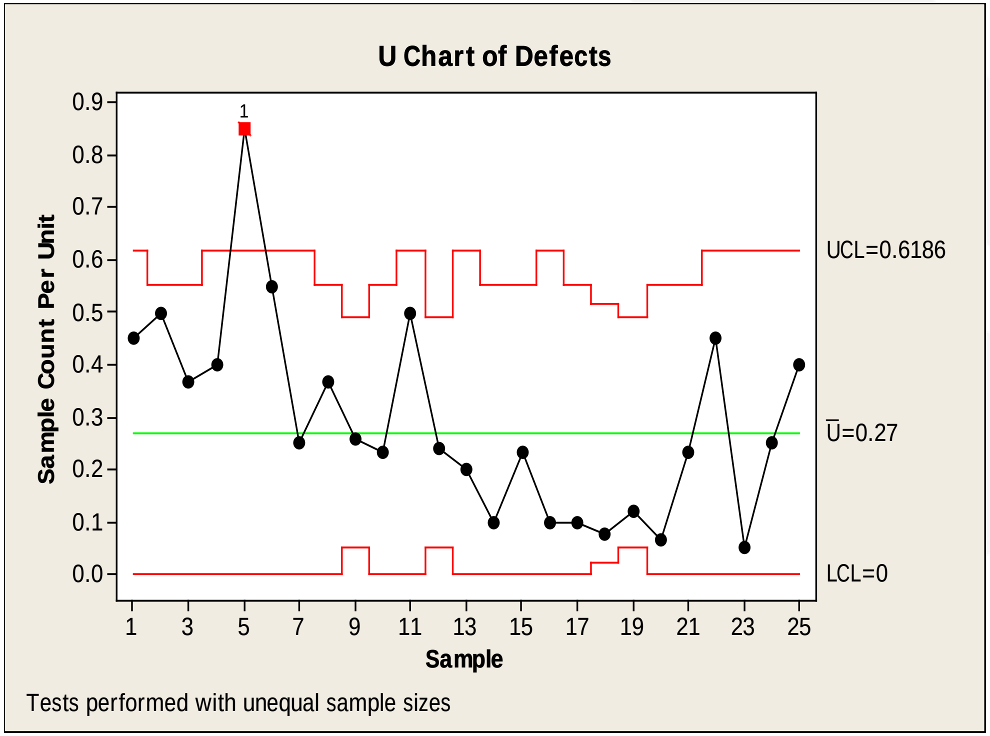


C Charts in Minitab



U Charts

Usage: Controlling the average number of defects per unit when the number of units inspected varies.



Assumptions:

- Each defect is a discrete event.
- Defects occur in a defined region or area.
- Defects are independent of each other.
- Defects are rare compared to what might be.

U Charts

Underlying distribution is Poisson.

Mean = $\dfrac{\text{Total Defects Found}}{\text{Number of units inspected}} = \bar{u}$

Control Limits =

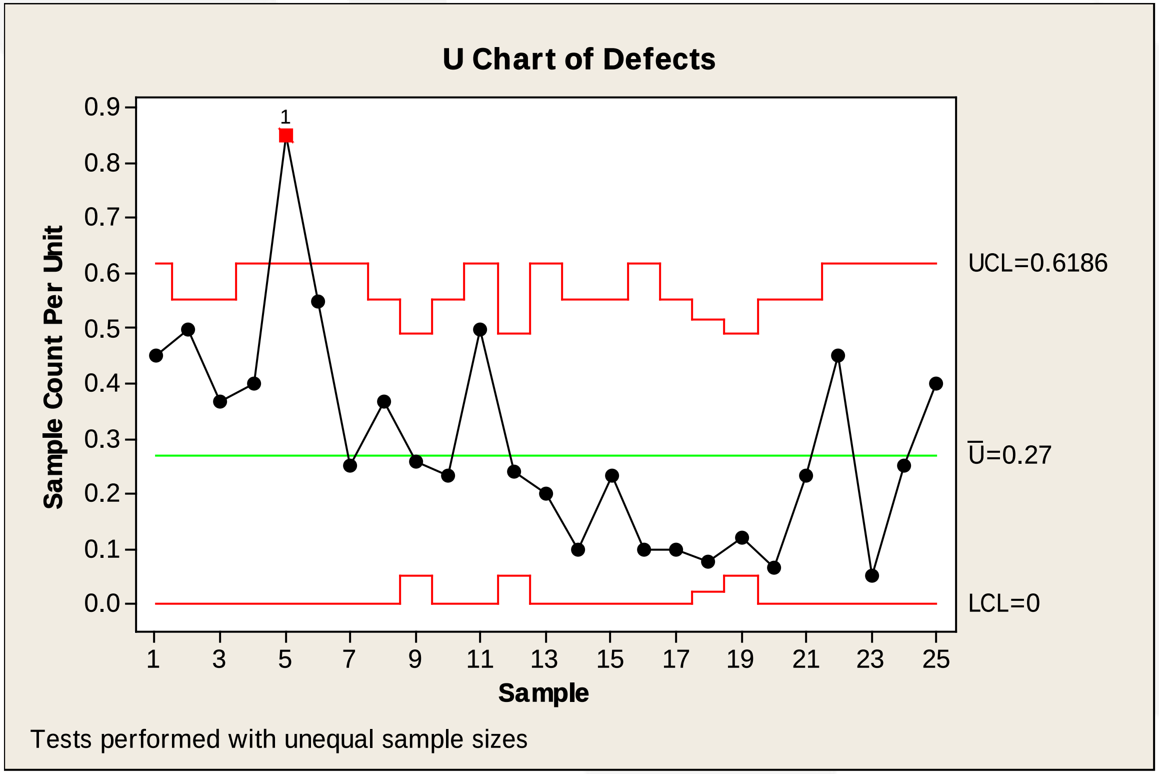
$$\bar{u} \pm 3\sqrt{\frac{\bar{u}}{n_i}}$$

where n_i = number of units inspected

U Charts in Minitab

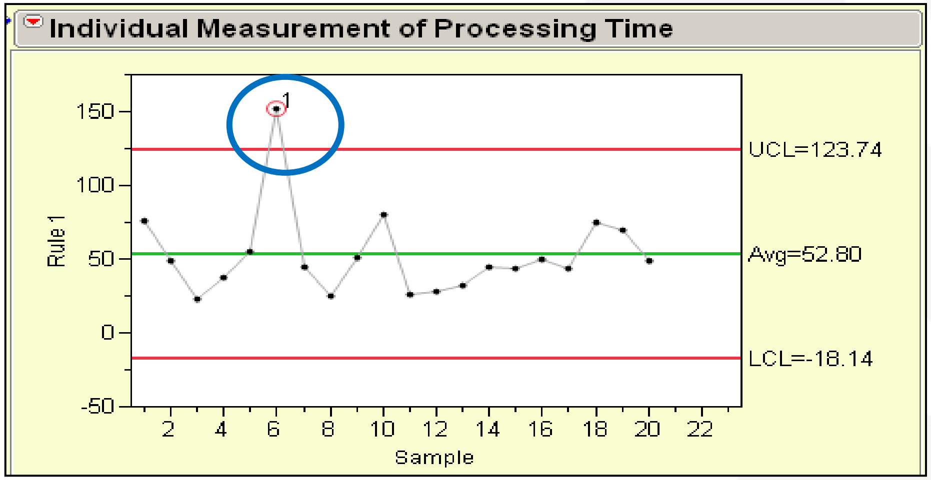
Defects	Claims Filed
9	20
15	30
11	30
8	20
17	20
11	20
5	20
11	30
13	50
7	30
10	20
12	50
4	20
3	30
7	30
2	20
3	30
3	40
6	50
2	30
7	30
9	20
1	20
5	20
8	20

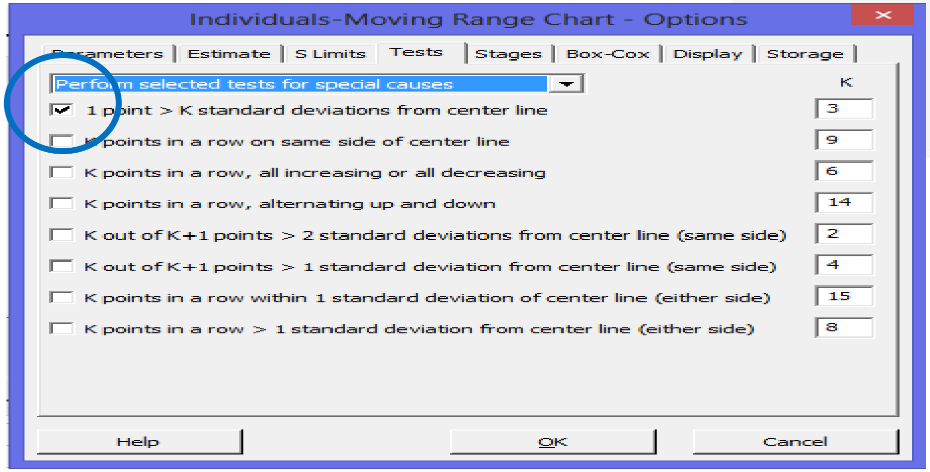
Data Table:
SPC_U_Chart_Example.MPJ



Western Electric Rules, Test 1

#1 One point more than 3 sigma from centre line (set as the only test default in Minitab).





Look for:

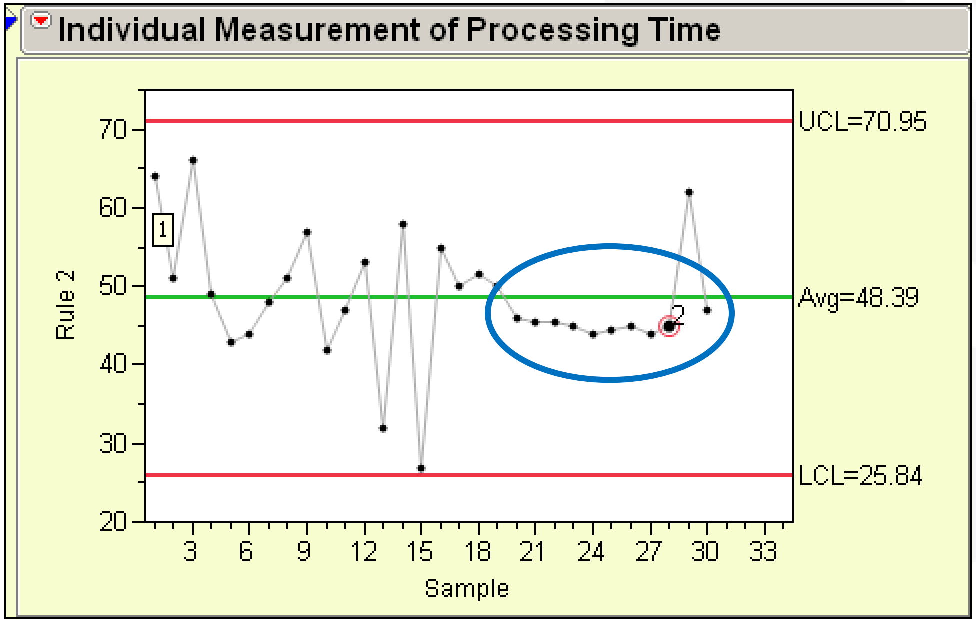
- Difference in process during this/these observations.
- Special causes operating at that specific time.

For all Tests, the first things to check are the obvious! If these are all good continue to investigate the process.

- Plotting errors (paper charts).
- Data entry errors.
- Measurement errors.
- Sampling errors.

Western Electric Rules, Test 2

#2 Nine points in a row on same side of centre line.



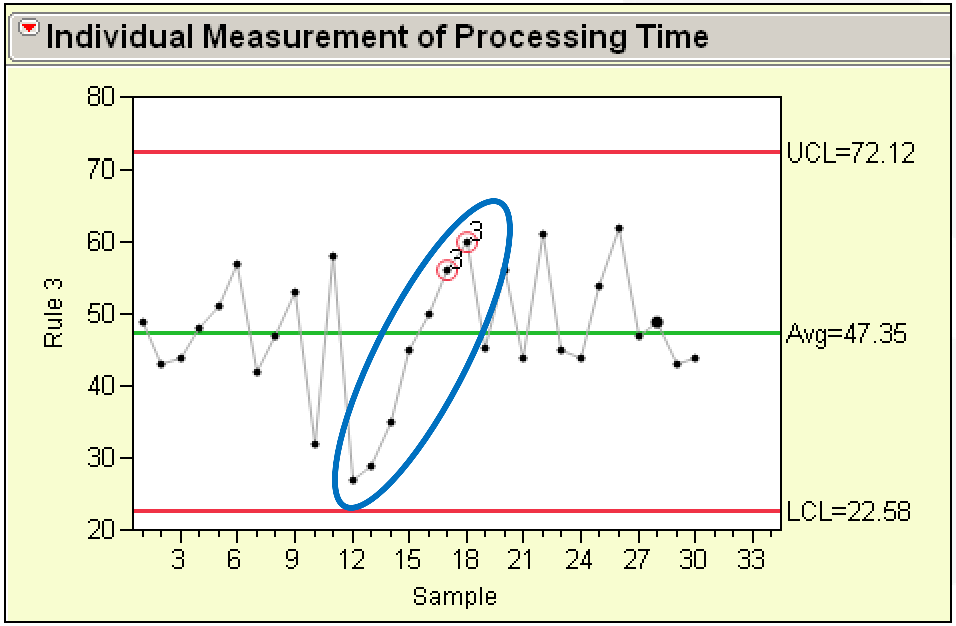
Look for:

- A special cause that would cause a small mean shift.
- New employees.
- Temporary process changes.

You need to turn this test on in Minitab so identify these special cause

Western Electric Rules, Test 3

#3 Six points in a row, all increasing or all decreasing.



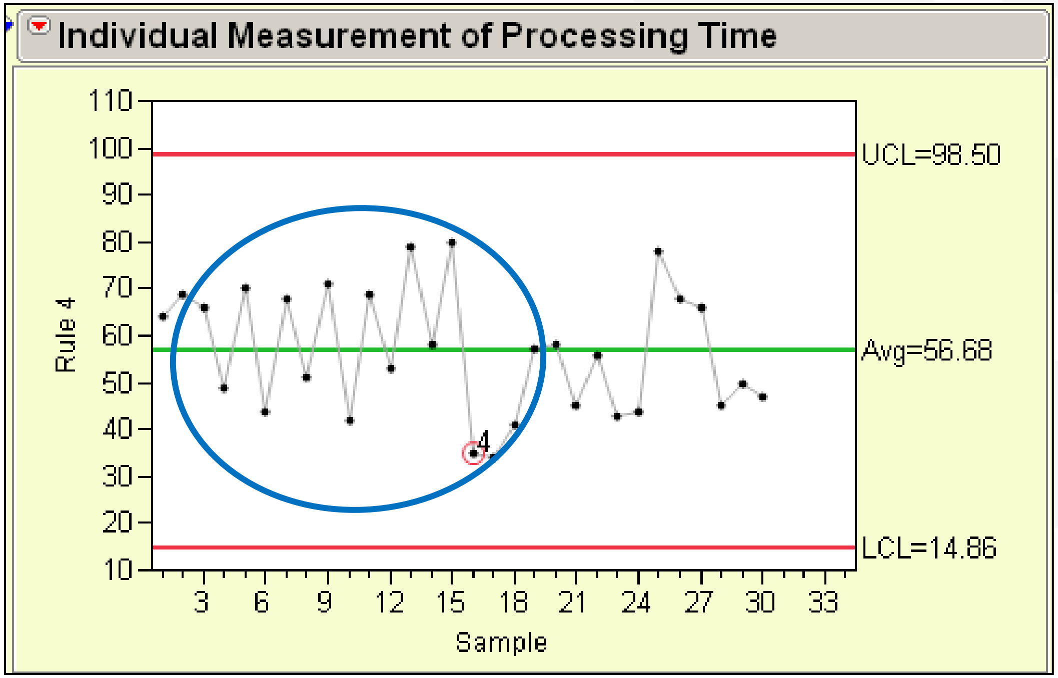
Look for:

- Gradual introduction of changes in the process.
- Better / changed supervision
- Increase / decrease in skill level.
- Introduction of SPC or improvements made in other areas that impact your process.
- Cumulative effects of some input.

You need to turn this test on in Minitab so identify these special cause

Western Electric Rules, Test 4

#4 Fourteen points in a row, alternating up and down.



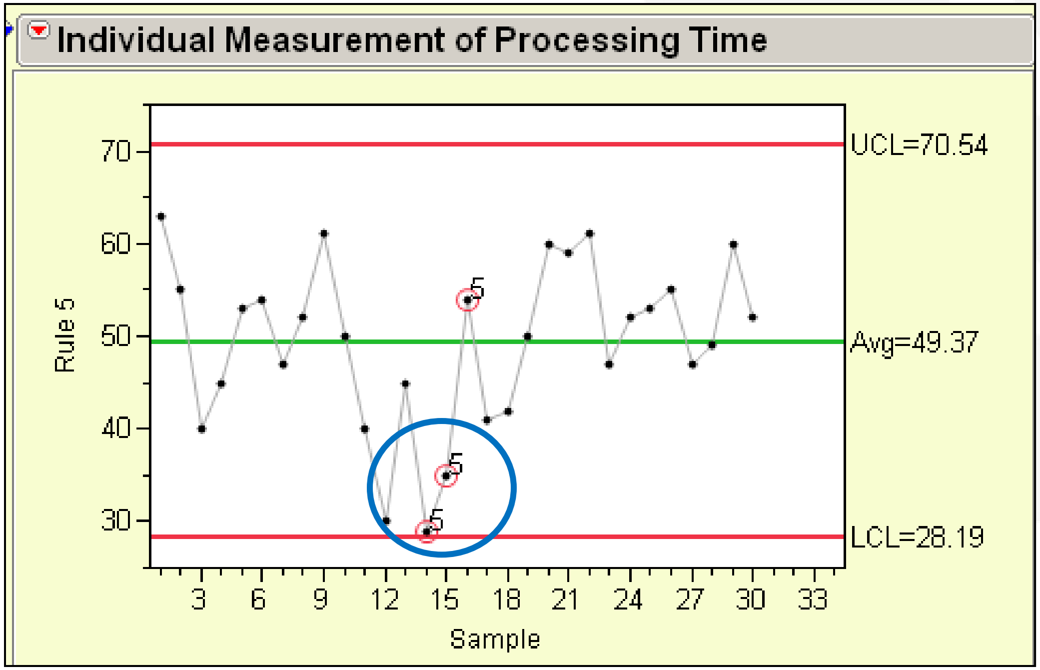
Look for:

- Different shifts or employees.
- Different batches.
- Different environmental conditions.

You need to turn this test on in Minitab so identify these special cause

Western Electric Rules, Test 5

#5 Two out of three consecutive points more than 2 sigma from centre line (same side).



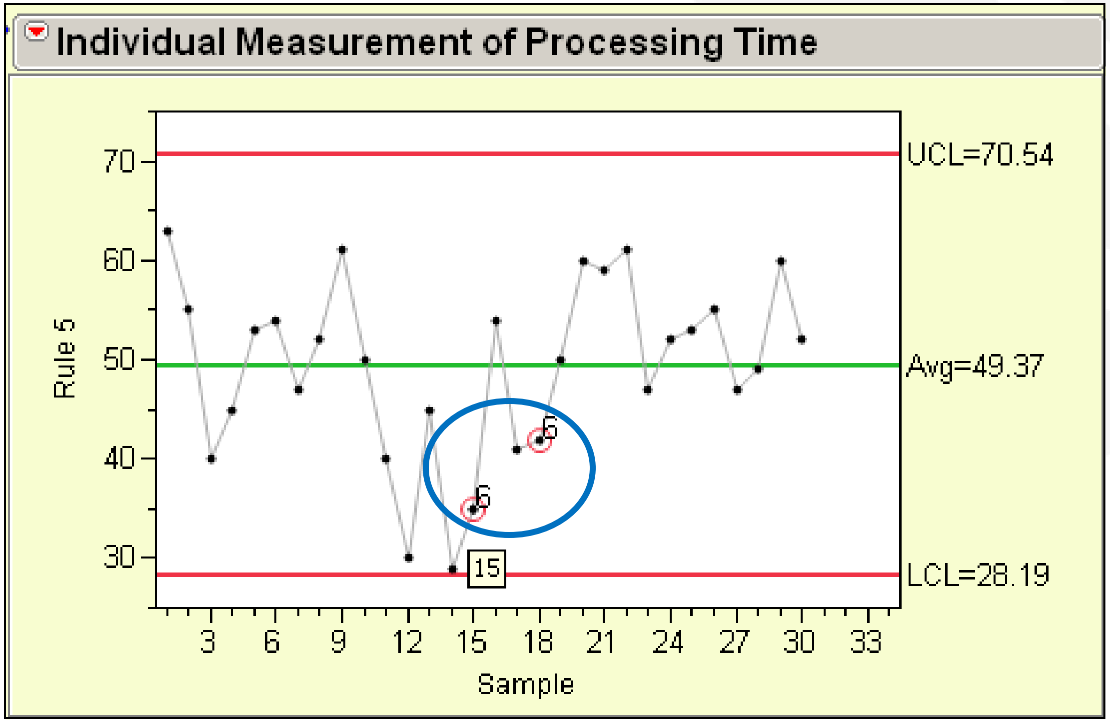
Look for:

- Special causes that would reduce the mean slightly.

You need to turn this test on in Minitab so identify these special cause

Western Electric Rules, Test 6

#6 Four out of five consecutive points more than 1 sigma from centre line (same side).



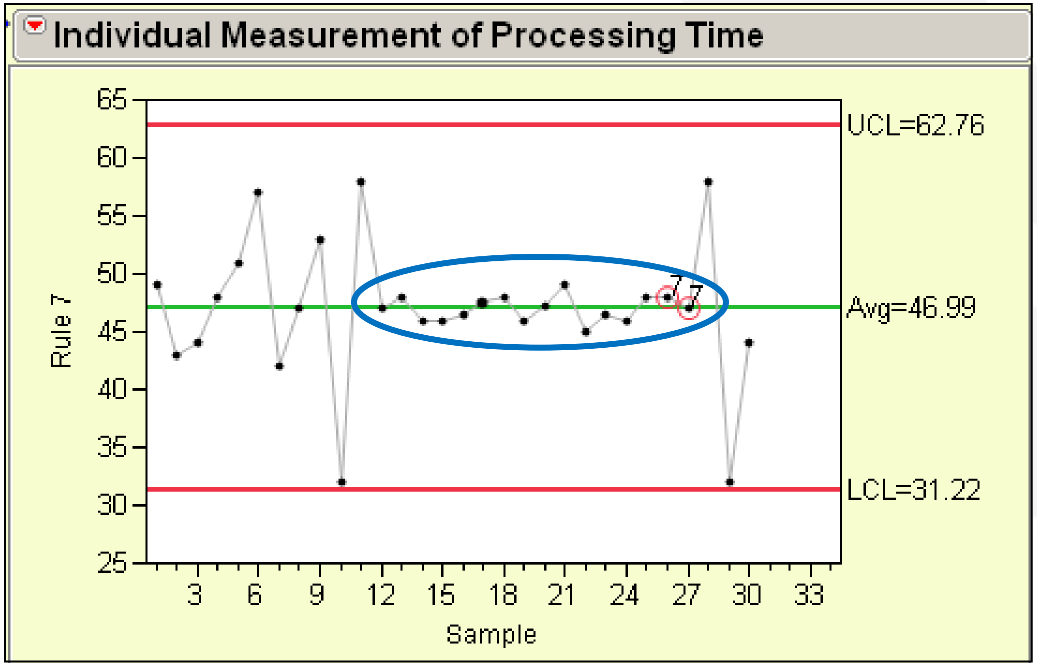
Look for:

- Special causes that would reduce the mean slightly.

You need to turn this test on in Minitab so identify these special cause

Western Electric Rules, Test 7

#7 Fifteen points in a row within 1 sigma of centre line (either side).



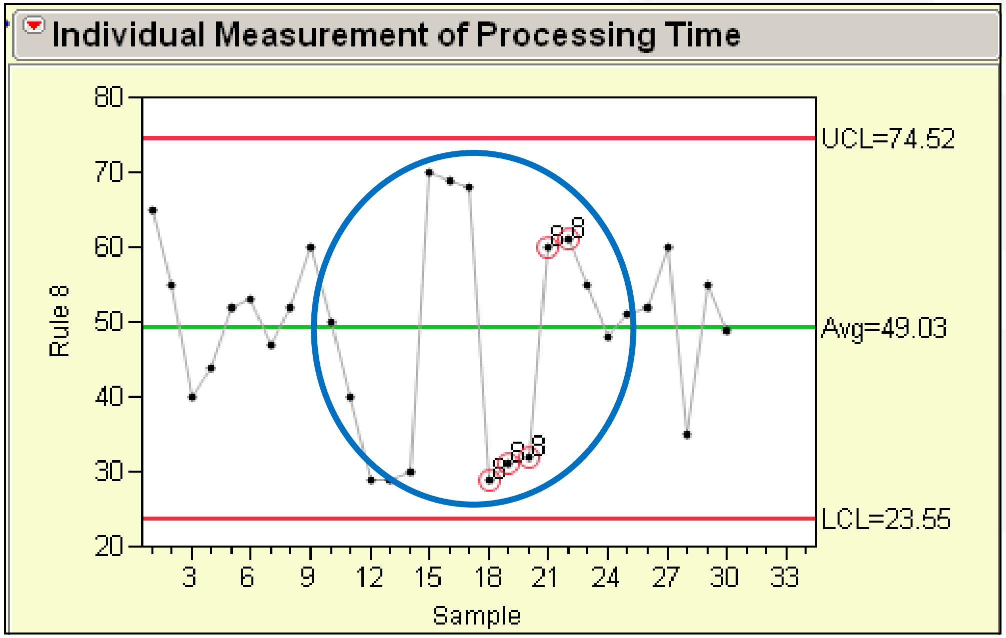
Look for:

- Special causes that would reduce the process variation.
- "Cherry picking" the samples.

You need to turn this test on in Minitab so identify these special cause

Western Electric Rules, Test 8

#8 Eight points in a row more than 1 sigma from centre line (either side).

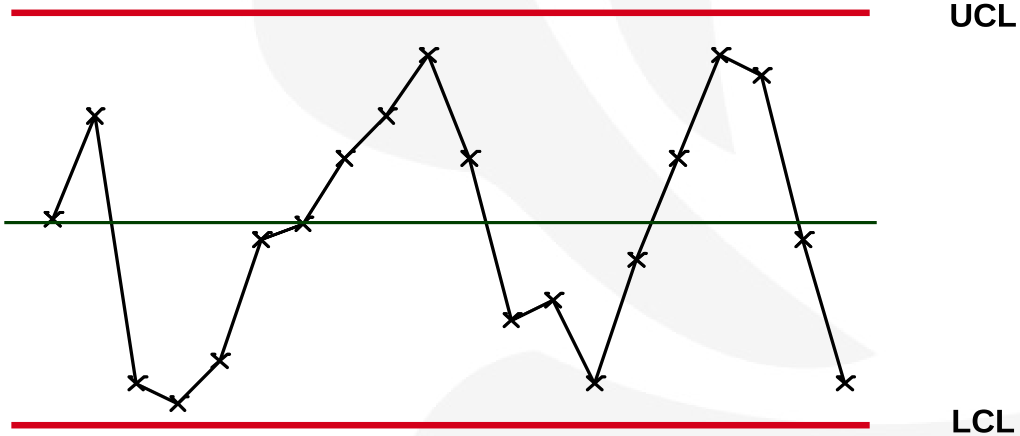


Look for:

- Special causes that drive the average in different directions.
- Different employees.
- Tampering with the process.

You need to turn this test on in Minitab so identify these special cause

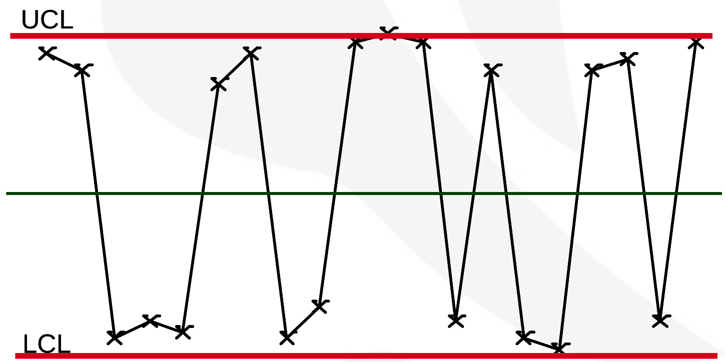
Cycling pattern example:



Possible causes:

- Seasonal or periodic influences.

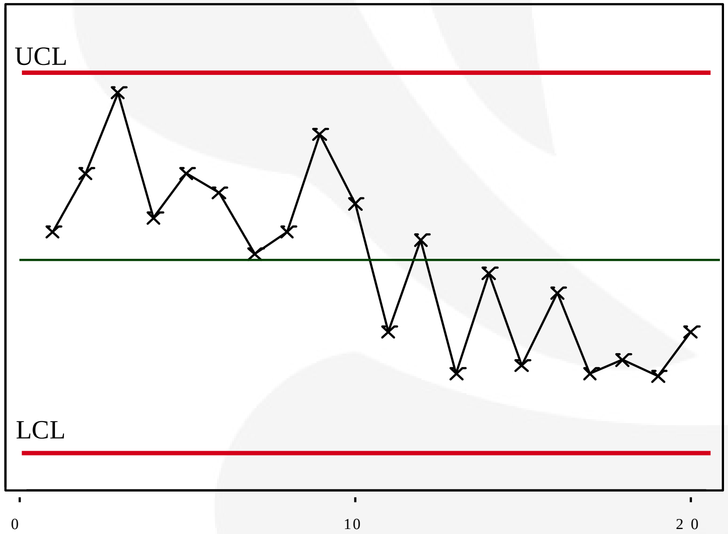
Mixture pattern example:



Possible causes:

- Two or more overlapping process output distributions.
- Tampering.

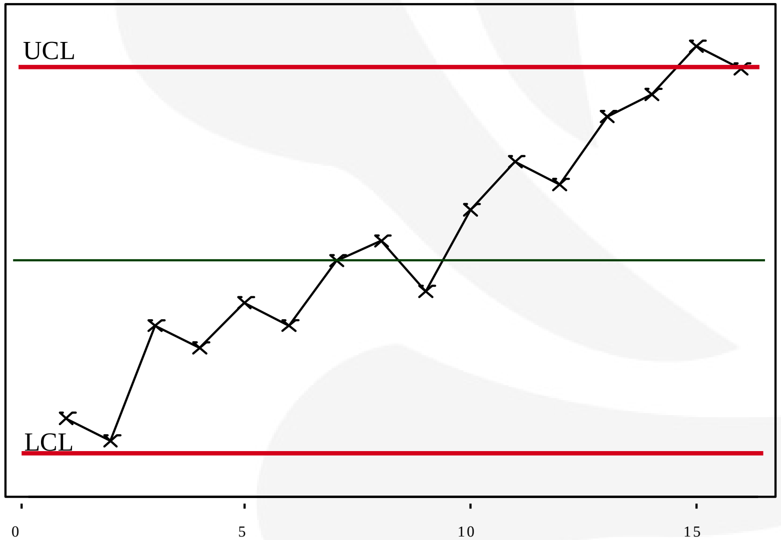
Shift in process example:



Possible causes:

- Introduction of new inputs.
- A change in inspection methods or standards.
- A process improvement.

Trend pattern example:



Possible causes:

- A cumulative effect over time, such as an increase in associate knowledge with more experience.
- A gradual effect due to **noise variables**, such as an increase in the cycle time to process loans when volume increases.

Exercise 2

Creating Control Chart

Objective: Create and interpret control charts

Create the appropriate control chart and interpret what it is telling you.

1. CRE samples the time to close a loan, sub-grouped by day. Is the process time to close a loan in control? Do you think the process is capable of achieving a 60 hour upper limit? Data in: SPC_Exercise_2_Problem_1.MPJ.
2. Every month, 2000 checking account statements are reviewed closely and every problem found is recorded. Based on analysis of the monthly data, is the process in control? Data in: SPC_Exercise_2_Problem_2.MPJ.
3. The process owner for New Account Openings wants to know if the process is predictable? Is the process capable? Data in: SPC_Exercise_2_Problem_3.MPJ.

30 minutes

Exercise 2

4. An online website response time is measured on a daily basis. Over the time span covered by the data several improvements have been implemented. On a single control chart, vary the control limit calculations based on improvements. Data in: SPC_Exercise_2_Problem_4.MPJ.

5. A call centre is monitoring AHT, the average handle time. Is their process in control? If not can you give some direction? Data in: SPC_Exercise_2_Problem_5.MPJ.

6. On a weekly basis, an IT team is monitoring links to the Intranet that do not point to a valid URL. As the Intranet changes this number can go up or down, but they want to maintain the process in control such that the number does not get out of hand. What would be the best chart? With the given data, would they have reacted to any points? Data in: SPC_Exercise_2_Problem_6.MPJ.

Summary of key learning points

- Control Charts monitor the stability of a process over time
- Control Charts help users distinguish between common cause and special cause variation
- Rational subgrouping is the sampling strategy to be used when control charting
- The type of control chart to be used in a given scenario is driven by the type of data, whether defects/defective units are being monitored and the subgroup size
- Control chart limits should only be drawn after a minimum of 20 data points are collected
- The Western Electric rules are criteria applied to the data collected which highlight special causes
- Other patterns in your collected data may be identified while not violating the Western Electric rules

Recommended coach support points

- There is significant within subgroup variation
- When you want to create a new phase (stage) in your control chart
- If you are unsure when to draw your control charts limits or declare your process stable
- There appears to be a pattern in the data you are collecting but no violation of the Western Electric rules

Black Belt

Hypothesis Testing I

Objectives of this module

At the end of this module, you will be able to:

- Review the basic usage and methods for Hypothesis Testing.
- Understand the statistical principles behind Hypothesis Testing.
- Be able to select, run and interpret Hypothesis Tests with confidence.

Questions Lead

- Basic question: Is there a difference in the output between these groups?
  - Does one procedure give us a shorter waiting time than the other?
  - Is one of my sales teams regularly selling more than the others?
  - I made a process change. Did our revenue increase?
  - Did the new machine reduce the variation in processing time?

Where We Normally Use HT

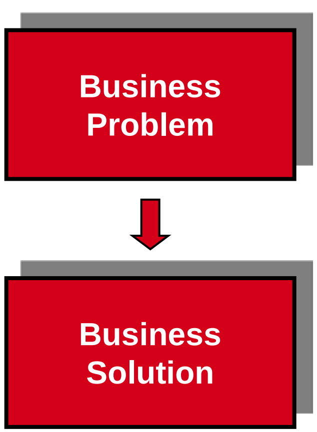
- In the Analyse Phase to verify the potential vital few inputs to confirm which ones are the true vital few inputs.
- In the Improve Phase to test potential solutions to validate and quantify the improvement.
- In the Control Phase to validate that the implementation is delivering the predicted results.
- Any time the question arises whether there is a statistically significant difference between the output of two or more groups.

Solving the Business Problem

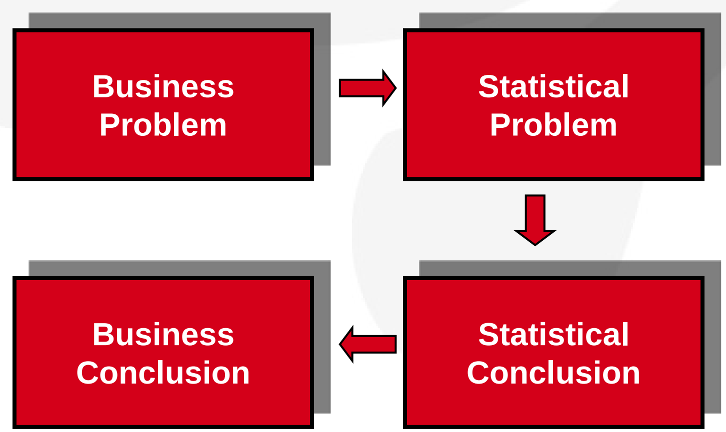
- Too often we look at a business problem, observe it for a while or take a look at some data and make our decision.

- The Six Sigma approach is to take the business problem and turn it into a statistical problem via the hypothesis testing method. Then we can solve the statistical problem to gain true knowledge and confidence about what is really happening. We apply that knowledge to solve the business issue.

The Typical Approach



The Six Sigma Approach



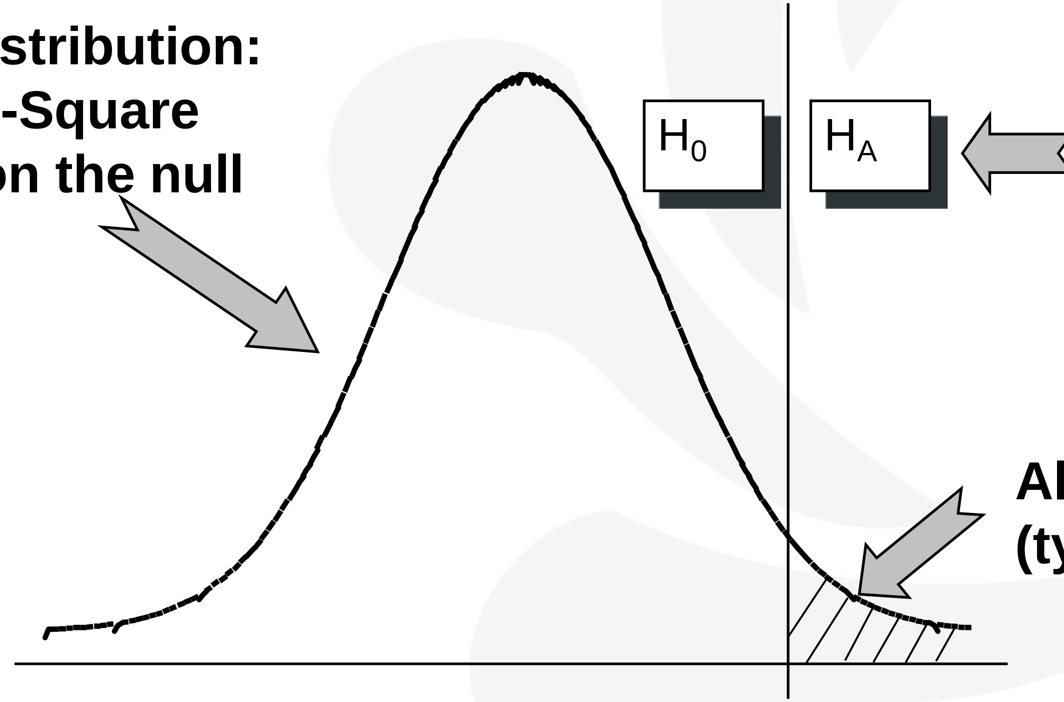
What is Hypothesis Testing?

A statistical hypothesis is a claim about whether there is a difference in output. We gather samples of the data and test them to see if the claim we made is true.

- Some examples:
  - There is no difference in the average A&E waiting time days by hospital
  - New-hire call takers follow procedures more consistently than experienced tellers.
  - Sales revenue in shopping centres in London is higher than those in the Birmingham.

How Hypothesis Testing Works

Sampling Distribution: Z, t, F or Chi-Square depending on the null hypothesis.



Options we must choose between depending on where the test statistic falls.

Alpha Risk Area (typically 5%)

Test Statistic: numerical value calculated by a statistical equation based on the null hypothesis and can fall anywhere along this line.

Decision Point, or Critical Statistic: numerical value that yields the alpha risk area.

Where the Numbers Come From

Decision Point: The decision point is based on what percentage of the area under the distribution matches the α risk level we are willing to accept. This point is also called the **critical statistic**.

This number is read from statistical tables that can be found in most statistical text books.

Test Statistic: The test statistic is calculated using a mathematical formula matched to the test you are performing. There are different formulas to test for differences betweens means, medians, variances and proportions. Examples will follow.

P-Value

The **p**-value is the actual **proportion of the area** under the distribution that is beyond the test statistic. Thus, if the α risk is 5% and the test statistic calculates to exactly the decision point at 5%, the p-value will be 0.05. If the p-value is 0.035, then 3.5% of the area under the distribution is beyond the test statistic. [Note: This applies for one sided tests. For two sided tests, it is the area beyond the test statistic multiplied by 2.]

Another way to describe the **p-value is the probability that the H_0 is true**. In other words the **p-value is the probability that the difference observed is due to mere random chance instead of a real difference in populations**.

Method for Hypothesis Testing

0. What practical question do you want to answer?
1. Define Null Hypothesis, H_0
2. Define Alternate Hypothesis, H_a
3. Choose Risk Level, α
4. Get Sample (Strategy, Size)
5. Run Test (See Hypothesis Testing Roadmap)
6. Make Decision
0. What is the answer to the practical question?

Hypothesis Statements

- Hypothesis statements come in pairs:
- The Null Hypothesis (H_0) claims that the **measurements are equal**, or **no change has taken place**. There is no difference in the outputs of the different groups.
- The Alternate Hypothesis (H_a) claims that the measurements are **not equal**, or that **some change** has taken place. There is a difference in the outputs of different groups.

Writing Hypothesis Statements

Hypothesis statements are recorded in a mathematical format. They include:

1. A notation of the population parameter being tested:

Population Parameter	Notation
Mean	**μ (mu)**
Variance	**σ^2 (sigma squared)**
Median	**η (eta)**
Proportion	**p**

2. A mathematical symbol for the claim to be tested:

Claim of Hypothesis	Symbol Used
No difference	**=**
Difference with 1 or 2 groups	**≠, <, or >**
Difference with more than 2 groups	**At least 1 group is different**

Examples of Hypothesis Statements

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

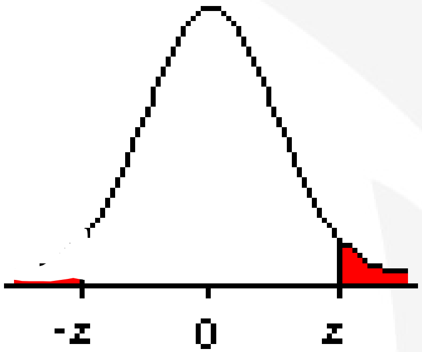
$$H_0 : \sigma^2_{Before} = \sigma^2_{After}$$
$$H_a : \sigma^2_{Before} > \sigma^2_{After}$$

$$H_0 : \eta_a = \eta_b = \eta_c$$
$$H_a : \text{At least one } \eta \text{ is different.}$$

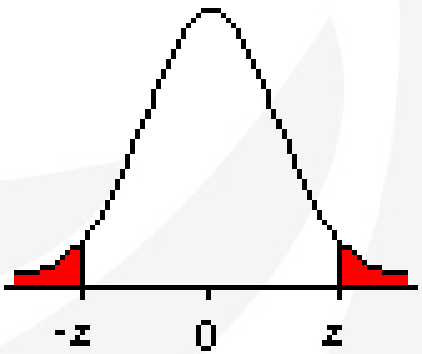
$$H_0 : p_1 = (\text{Goal of})\ 25\%$$
$$H_a : p_1 < 25\%$$

One-Sided versus Two-Sided Tests

A test that is concerned only with whether the difference is greater than <u>or</u> less than (but not both) is called a one-sided test. We can put all the α risk in one tail.



If the test involves inequality, then the test is called two sided. We have to split the α risk between both tails. In such cases, the Z_α value is substituted by $Z_{\alpha/2}$. There is no change in β or Z_β. This applies to the t distribution as well.

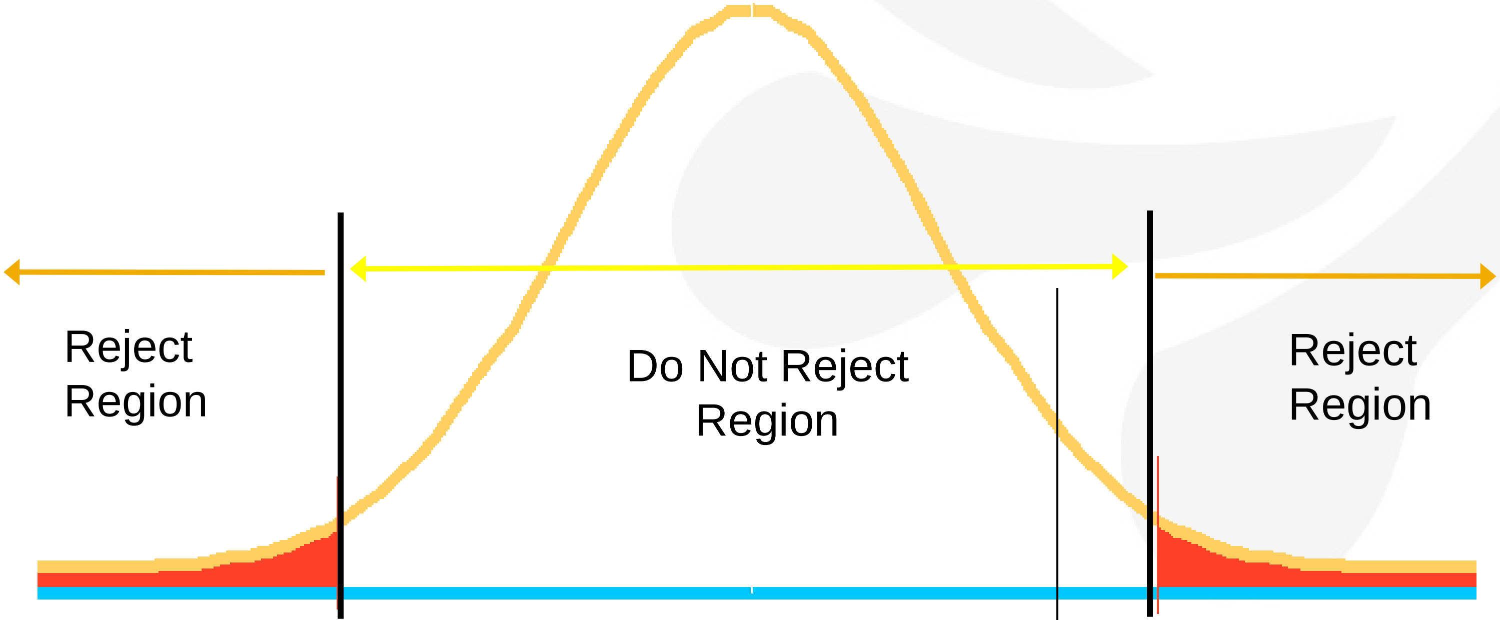


Two Sided (Two "Tailed") Test

Hypotheses (μ_t is the target mean)

H_0: $\mu = \mu_t$

H_a: $\mu \neq \mu_t$

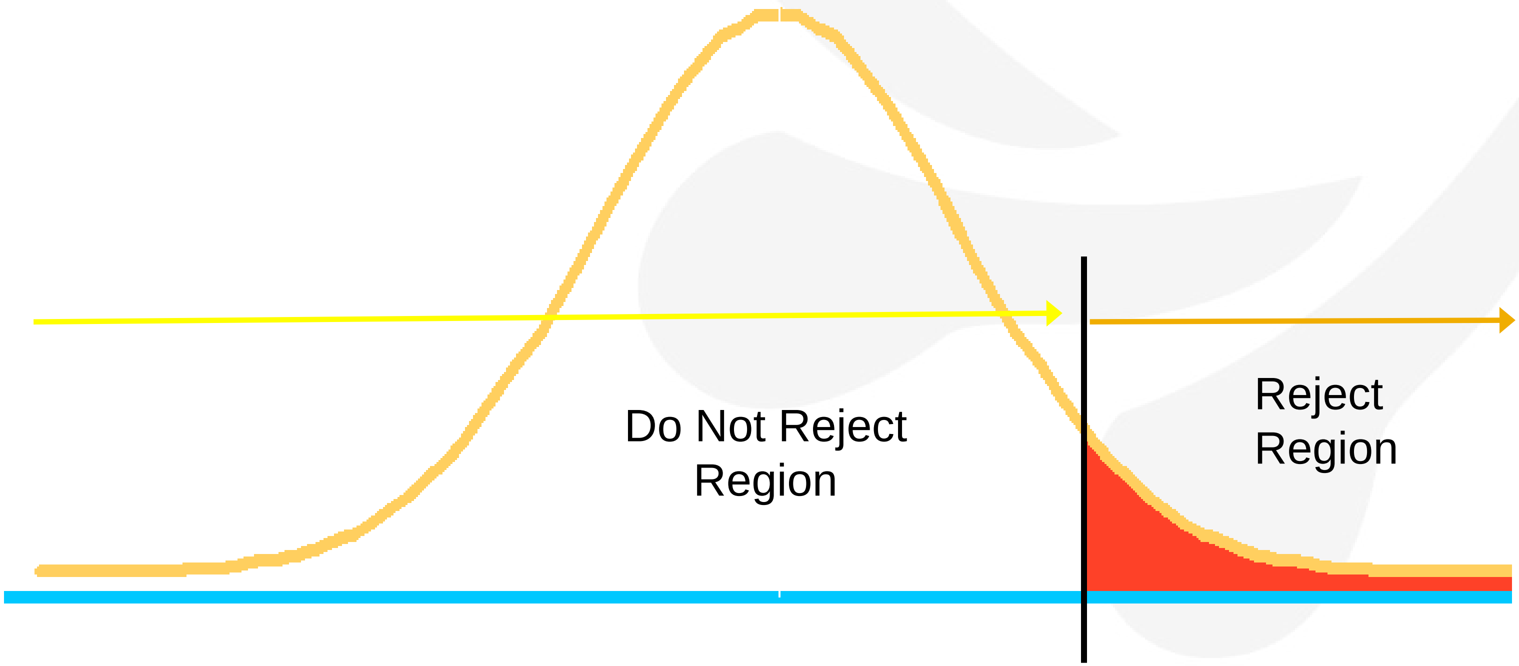


One Sided (One "Tailed") Test

Hypotheses (μ_t is the target mean)

$H_0: \mu \geq \mu_t$ *or* $H_0: \mu \leq \mu_t$

$H_a: \mu < \mu_t$ $H_a: \mu > \mu_t$



Exercise 1

Hypothesis statements

Objective: Revision of hypothesis statements

Frame hypothesis statements for the following scenario:

A particular call centre has an average call length of 4.7 minutes and a standard deviation of 2.2 minutes. A Green Belt project's goal is to drop the average to 3.0 minutes. How do we test whether or not they succeeded?

10 minutes

Furthermore, if the call length standard is 5.3 minutes, the Green Belt indicated that 21% of the calls prior to the project exceeded the standard but after the changes, only 18% exceeded the standard.

Method for Hypothesis Testing

0. **What practical question do you want to answer?**
1. **Define Null Hypothesis, H_0**
2. **Define Alternate Hypothesis, H_a**
3. **Choose Risk Level, α**
4. **Get Sample (Strategy, Size)**
5. **Run Test (See Hypothesis Testing Roadmap)**
6. **Make Decision**

0‘. **What is the answer to the practical question?**

Alpha Risk

Ways to understand alpha (α) risk:

- **The risk associated with finding that something is significant when in reality it is not.**
- **False positive.**
- **Another term for α is “producer’s risk”, because we may make unnecessary changes to our processes.**
- **An “overactive” α finds causes that do not exist.**
- **The risk of chasing an unimportant X.**
- **The risk of thinking you made an improvement when you really didn’t.**
- **α risk is commonly set at 5%, or α = 0.05.**

Beta Risk

Ways to understand beta (β) risk.

- **The risk associated with concluding that something is not significant when in reality it is.**
- **False negative.**
- **Another term for β is "consumer's risk", because the customer receives poor service or a defective product.**
- **A "lazy" β fails to find causes.**
- **The risk of missing an important X.**
- **The risk of thinking you didn't make an improvement when you really did.**
- **β risk is commonly set at 10%, or β = 0.10.**

How Alpha and Beta Relate

Truth is: \ Test says:	H_o	H_a
H_o	Correct Decision	Type I Error α
H_a	Type II Error β	Correct Decision

α Risk is the risk of making a Type I Error.

A fire alarm sounds, but there is <u>no</u> fire. Or we deny a loan to a credit worthy person.

β Risk is the risk of making a Type II Error.

The fire alarm is silent, but there <u>is</u> a fire. Or we approve a loan for a non-credit worthy person.

Only One Risk at a Time

Because we only make one decision on a hypothesis test, we can only make one of the two errors.

If: $H_0: \mu = \mu_0$ $H_A: \mu \neq \mu_0$ And: p-value $< \alpha$

Reject the null with risk of α ($1-\alpha$ confidence). Risk of false positive; Type I Error.

If: $H_0: \mu = \mu_0$ $H_A: \mu \neq \mu_0$ And: p-value $> \alpha$

Fail to reject the null with risk of β ($1-\beta$ power). Risk of false negative; Type II Error.

Trap: "Fail to reject the null with risk of α." If we fail to reject the null we cannot make a Type I error; there is no alpha risk, only beta.

We manage the α risk by choosing the risk level and comparing the p-value to it. We manage the β risk by increasing the sample size. We will discuss sample size shortly.

Method for Hypothesis Testing

0. What practical question do you want to answer?
1. Define Null Hypothesis, H_0
2. Define Alternate Hypothesis, H_a
3. Choose Risk Level, α
4. Get Sample (Strategy, Size)
5. Run Test (See Hypothesis Testing Roadmap)
6. Make Decision

0'. What is the answer to the practical question?

Sampling Terminology

n: The number of units making up the sample size. May be expressed differently depending on the situation. For a Design Of Experiment, n may be the number of experimental runs. In a two sample t-test, n could represent the number of observations for each group.

α: **(alpha)** Your chance of a false positive, which is the p-value at which you start calling things "statistically significant".

β: **(beta)** Your chance of a false negative.

δ: **(delta)** The size of the real effect you want to be sure to detect if in fact it is there. Often expressed as a multiple of σ.

σ: **(sigma)** The standard deviation of the noise variation when factors are held fixed.

Confidence: The confidence you have that you have not made a false positive statement. Confidence = **1-α**

Power: Your chance of detecting a real effect, i.e., declaring it to be statistically significant. You want this high. Power = **1-β**

General Sample Size Formulas

The sample size formula presented in the Sampling module is actually a simplified version of the general formula that is given below:

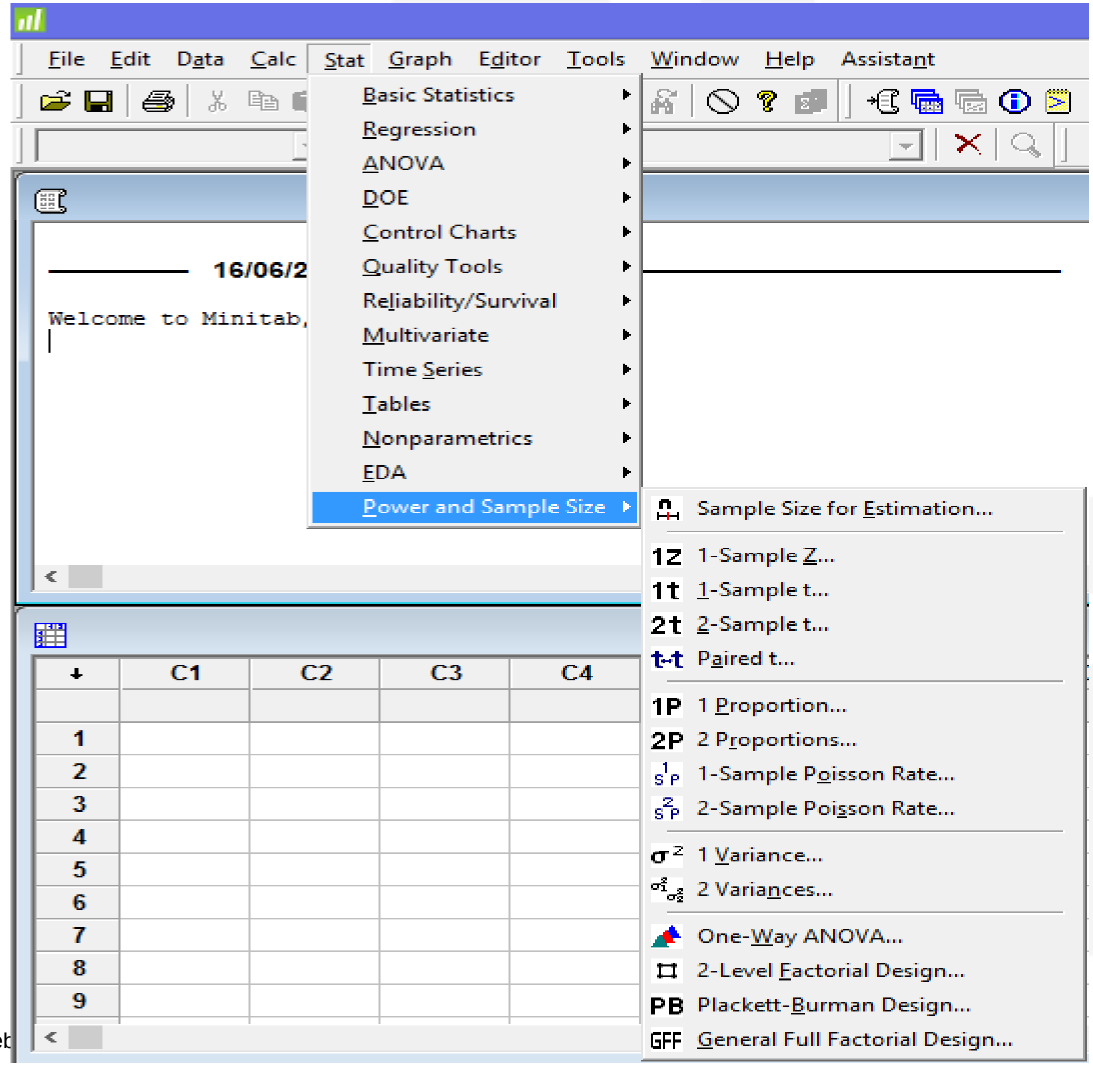
$$\mathbf{n} = \left[\frac{(Z_{\alpha/2} + Z_{\beta})s}{\delta}\right]^2 \text{ for continuous data}$$

$$\mathbf{n} = \left[\frac{(Z_{\alpha/2} + Z_{\beta})}{\delta}\right]^2 p(1-p) \text{ for discrete data}$$

Computers use the t Distribution instead of the Z Distribution for more accuracy. These are shown to demonstrate how the beta risks are included when choosing sample size.

Using Minitab for Sample Size

Stat > Power and Sample Size



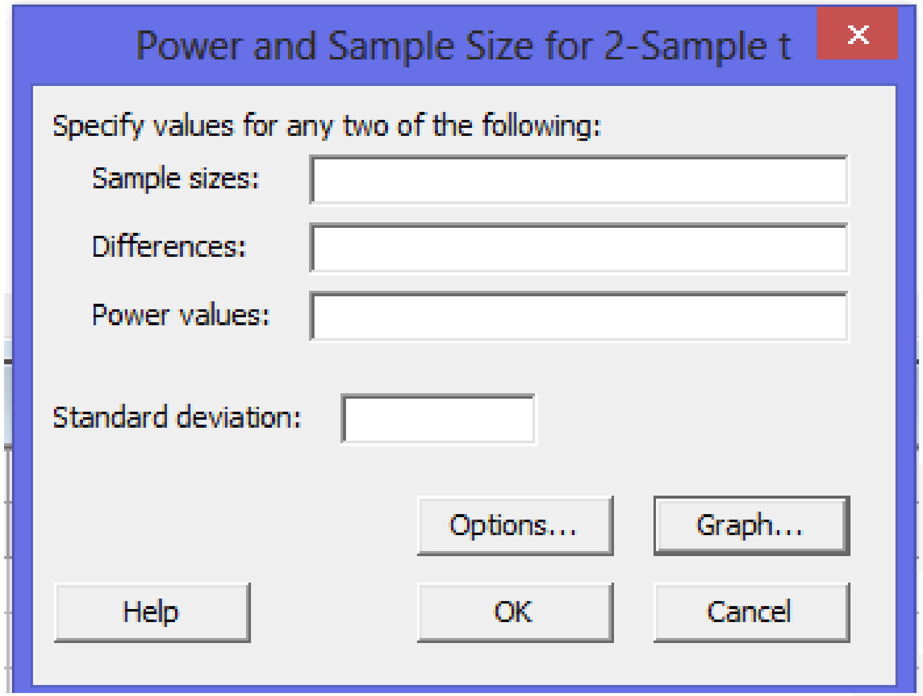
Exercise 2

We want to see if the average time to process an application is the same for two processing centres. Our best (planning) estimate for the average time is around 15 days with a standard deviation, σ = 2 days.

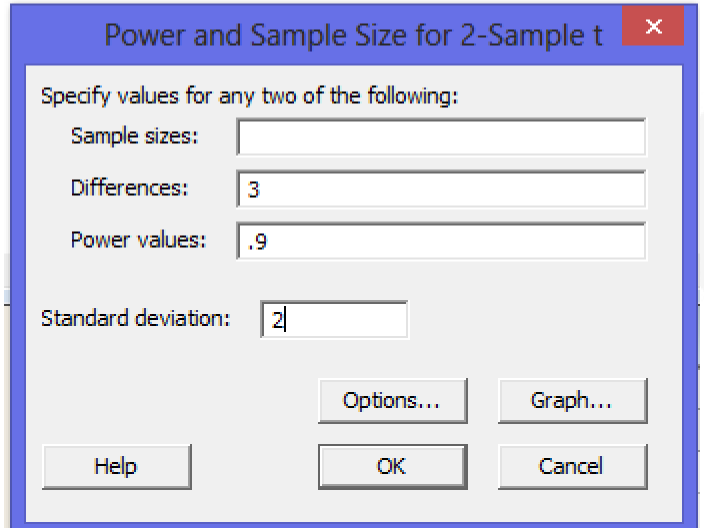
The sample size must be large enough to provide a 90% chance of detecting a difference (if it exists) in the average processing times, as small as 3 days (because a 3 day difference is of "practical" significance to us). Using an alpha risk of 0.05, what sample size would you recommend?

Exercise 2 – Choosing the Option

Stat > Power and Sample Size > Two Sample- t



Exercise 2 - Answer



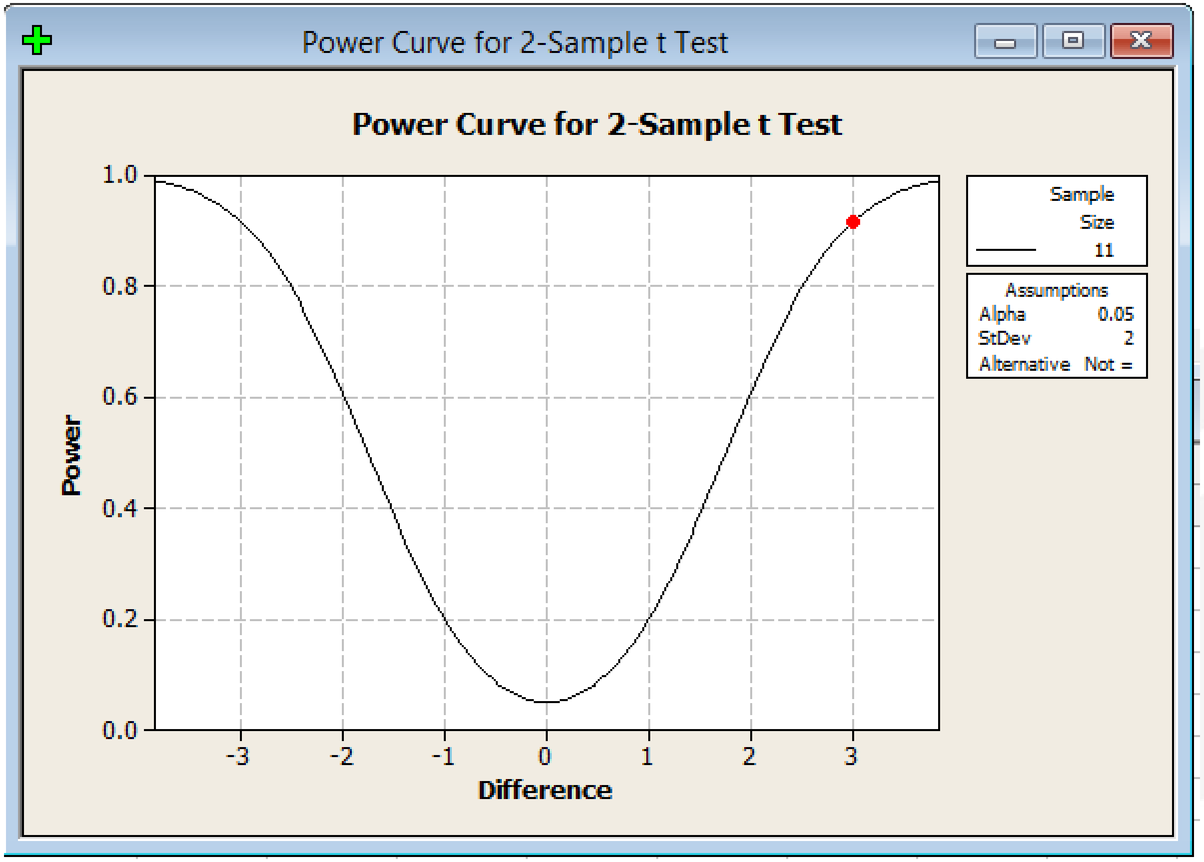
- **Enter two of Difference to detect, Sample Size or Power and Minitab will calculate the third.**
- **How many samples will we have to take from each Centre?**

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0.05  Assumed standard deviation = 2

            Sample  Target
Difference    Size   Power  Actual Power
         3      11     0.9      0.916899
The sample size is for each group.
```

Exercise 2 - Tradeoffs

- **By selecting "Display Power Graph" we can also see how our "difference to detect" works with our sample size of 13 and Power.**
- **What we can see from the graph below is that as our difference to detect falls, with a constant sample size our Power falls**



Exercise 3

Sampling Trade Offs

Objective: Understand the relationship between Power, Alpa risk, Precision and sample size

Fill in the relationships chart

Case		power		n
	Constant	Constant	↓	?
	Constant	?	Constant	↑
	Constant	↑	?	Constant
4	↓	Constant	Constant	?
5	↓	Constant	?	Constant
6	↓	?	Constant	Constant

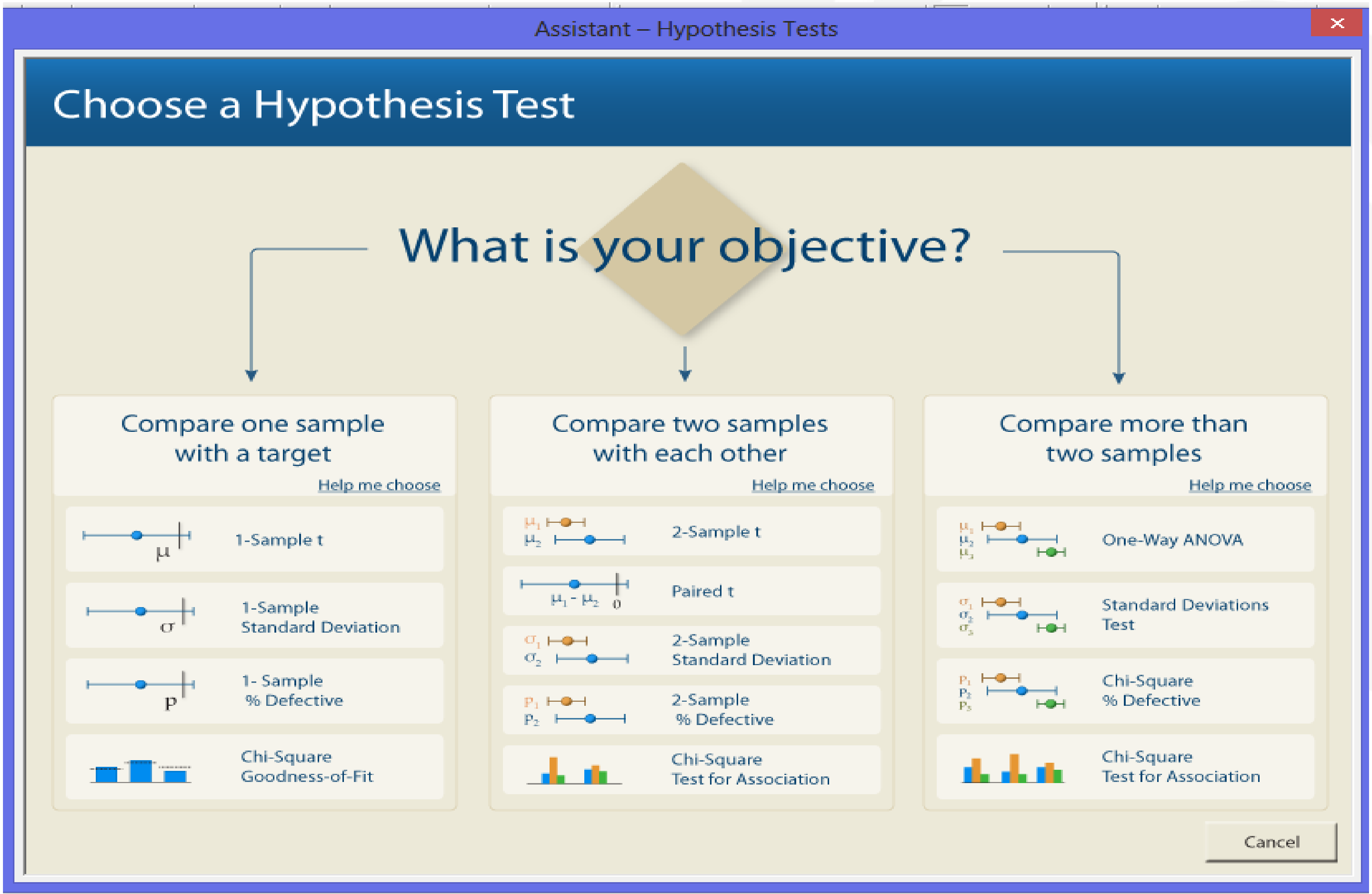
10 minutes

Method for Hypothesis Testing

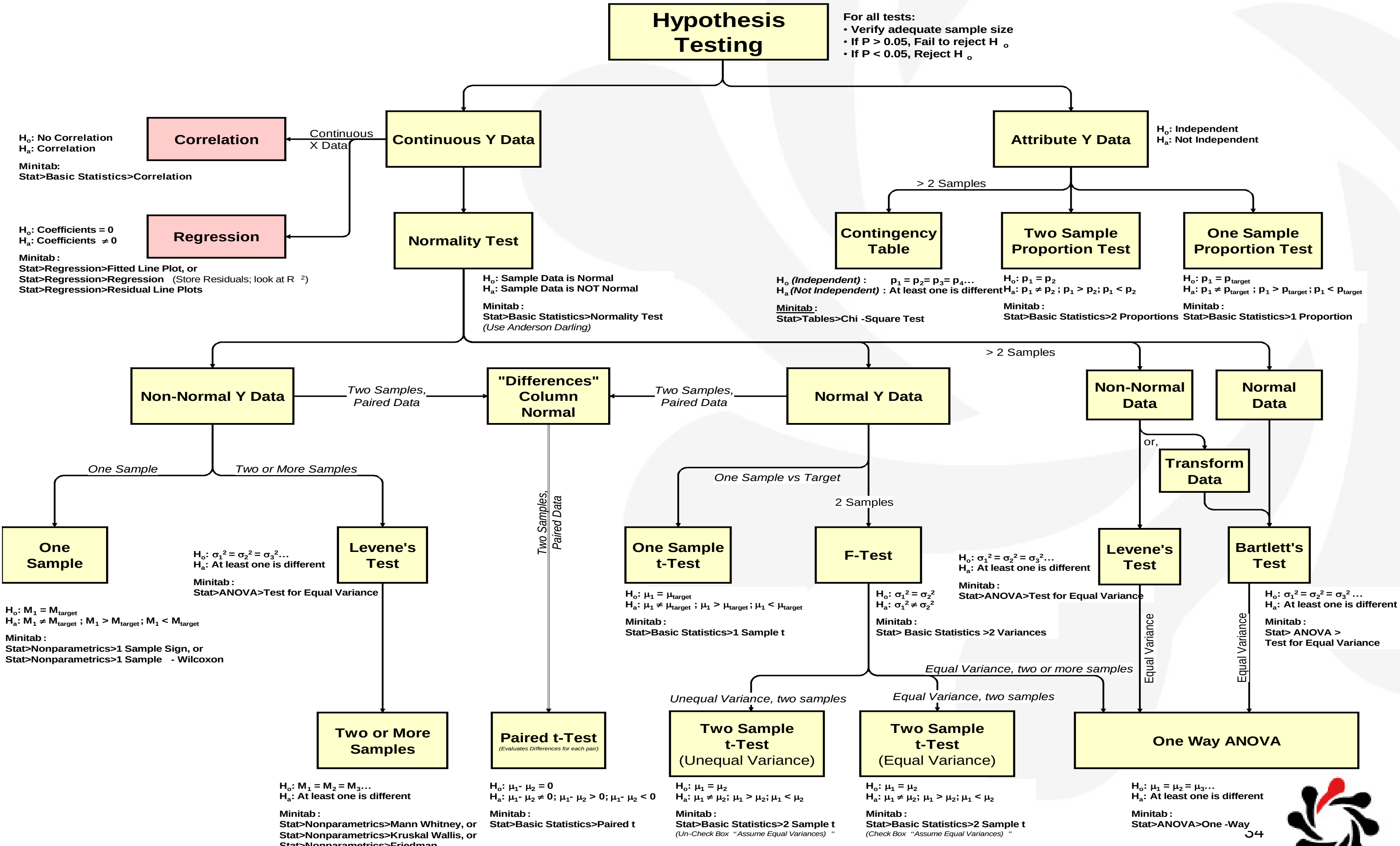
0. What practical question do you want to answer?
1. Define Null Hypothesis, H_0
2. Define Alternate Hypothesis, H_a
3. Choose Risk Level, α
4. Get Sample (Strategy, Size)
5. Run Test (See Hypothesis Testing Roadmap)
6. Make Decision

0'. What is the answer to the practical question?

Minitab Assistant



Hypothesis testing roadmap



Exercise 4, Sample Mean vs. Target

- At our year end sales meeting, we discussed ways to increase the revenue in our ten sub-regions and came up with a new sales strategy that we implemented in January. The 1st quarter data is in and we need to know if we are hitting our goal of £180,000 average across the ten.

- Actually we know we hit the goal for 1st Quarter since the average was £181,033. But our Master Black Belt advisor wants us to be able prove (to a 95% confidence) whether the difference is real and sustainable, or whether it might just as easily go below the goal for next quarter.

- If the results are that we cannot prove a sustainable difference, what is the power of the test to detect a £1,033 difference?

T-test assumptions

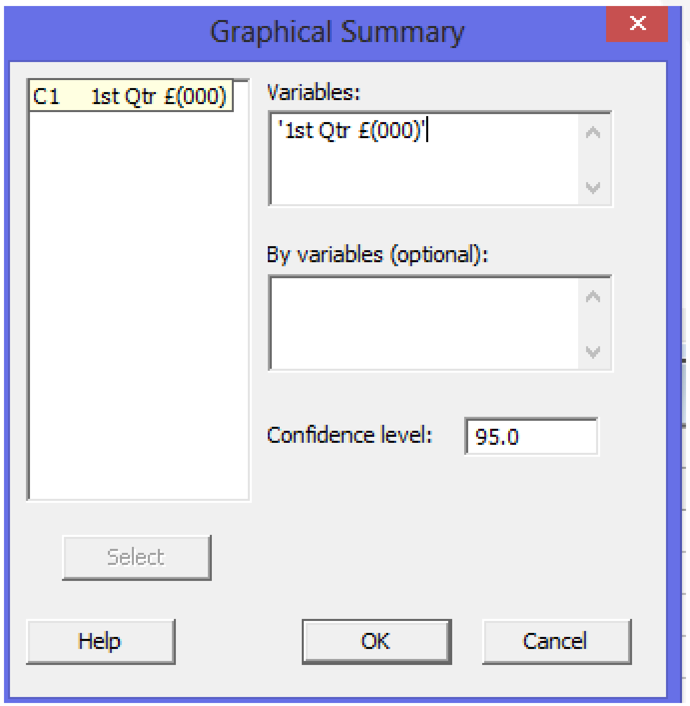
There are basic assumptions which must be met for a t-test to be robust

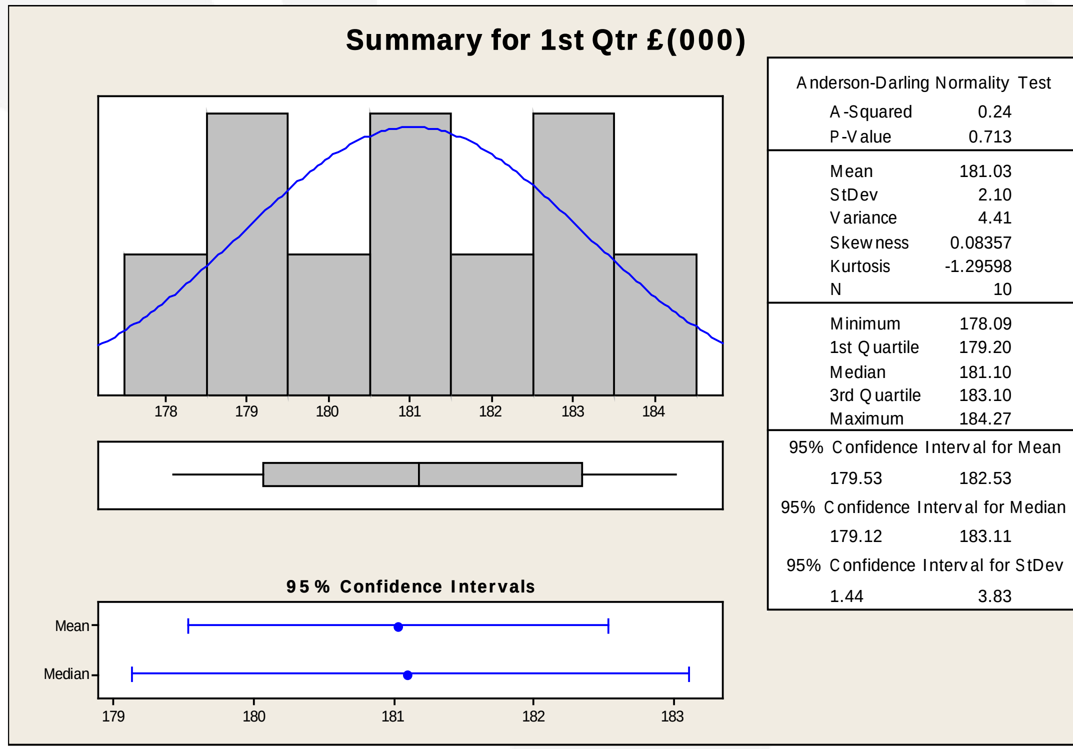
- Response is normally distributed (each level)
- Response is independent (each level)
- Equal variance (for one to one comparison only, each level)

Exercise 4, Normality Test

Stat > Basic Stats > Graphical Summary

Data Table:
Hyp1_One_Level_Ex4.MPJ

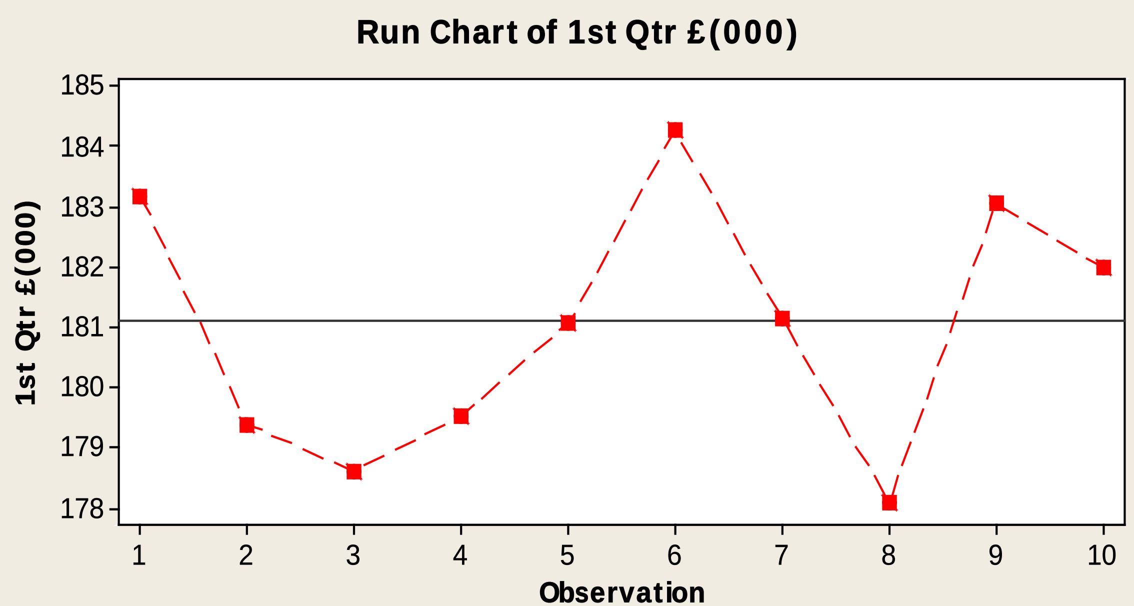




Can we assume normality?

Exercise 4, Independence

Stat > Quality tools > Run Chart



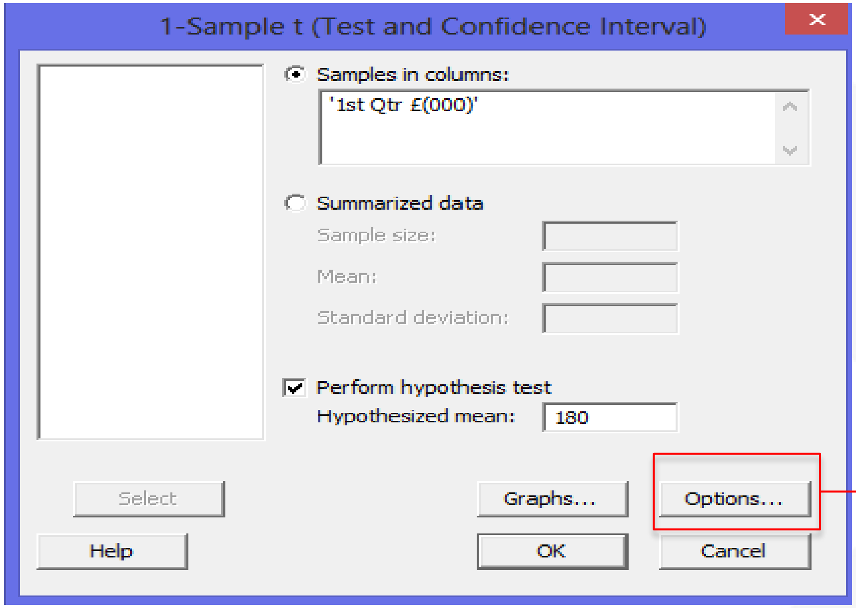
Number of runs about median:	5
Expected number of runs:	6.0
Longest run about median:	4
Approx P-Value for Clustering:	0.251
Approx P-Value for Mixtures:	0.749

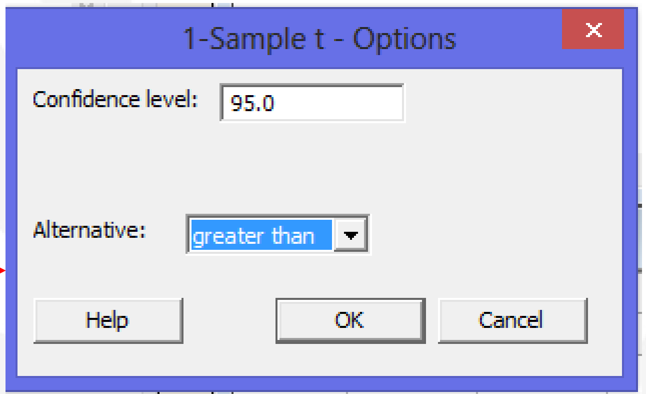
Number of runs up or down:	5
Expected number of runs:	6.3
Longest run up or down:	3
Approx P-Value for Trends:	0.135
Approx P-Value for Oscillation:	0.865

P-values created to highlight any clustering, mixtures, trends or oscillation.

Are we safe to continue?

Exercise 4, Testing the Mean





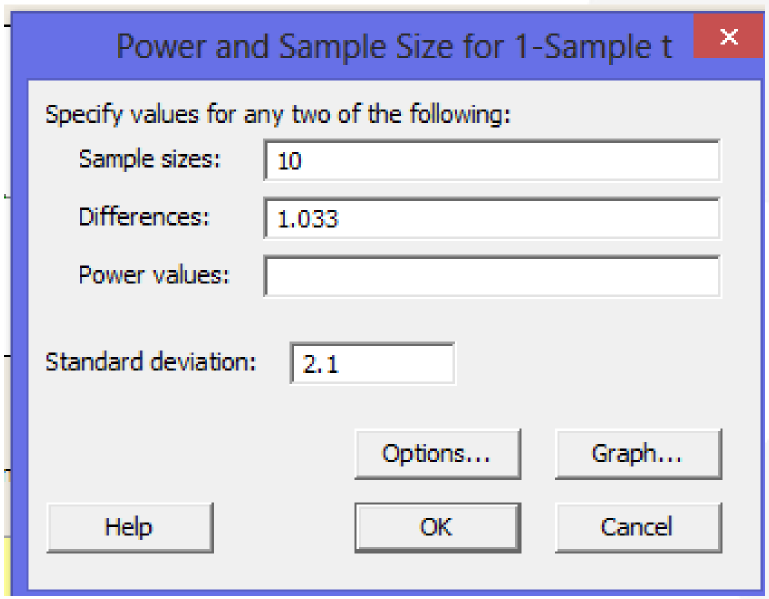
One-Sample T: 1st Qtr £(000)

Test of mu = 180 vs > 180

Variable	N	Mean	StDev	SE Mean	95% Lower Bound	T	P
1st Qtr £(000)	10	181.033	2.100	0.664	179.815	1.56	0.077

Exercise 4, Power of the Test

Ok, we couldn't prove we we're different to £180,000. What was the power of the test to prove a difference of £1,033?



Power and Sample Size

```
1-Sample t Test

Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05  Assumed standard deviation = 2.1


              Sample
Difference      Size     Power
     1.033        10  0.285426
```

With 10 samples and a standard deviation twice the size of the difference, we had less than a 30% chance of proving a difference.

What about Non-Normal Data?

<u>Parametric vs. Nonparametric:</u>

For non-normal data we use nonparametric methods. These procedures do not reference specific parameters.

Parameters are the values of Shape, Spread and Centering. For a normal distribution the shape is normal, the spread is s, the center is m. For every value of m and s, there is another normal distribution that is defined.

A nonparametric hypothesis test looks at the sample data and calculates a test statistic based on the medians without reference to the parameters. It determines whether that test statistic is inside or outside of the chosen risk level (i.e. beyond the decision point).

The reason we don’t use nonparametric tests exclusively is that they tend to have less power (to detect a difference) than the parametric tests.

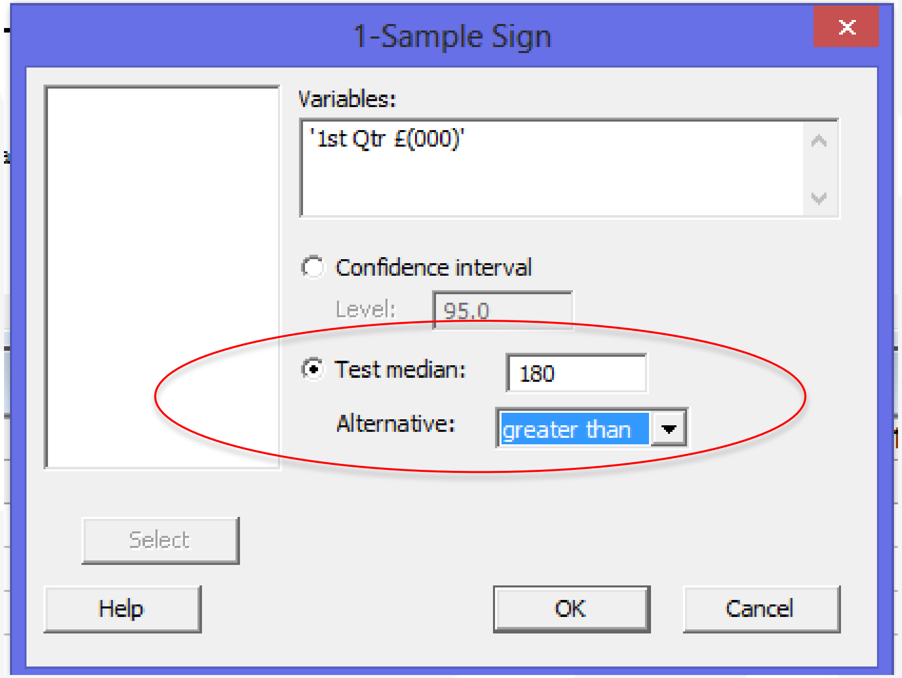
Exercise 4, Sample Median vs. Target

Sticking with our sales example: For our 1st quarter sales, we only have 10 sub-regions. That may be a little shaky to base normality on.

Would we get a different answer if we did not assume normality?

Exercise 4, Non-parametric Test

Stat> Non-parametrics >1-Sample Sign



Sign Test for Median: 1st Qtr £(000)

Sign test of median = 180.0 versus > 180.0

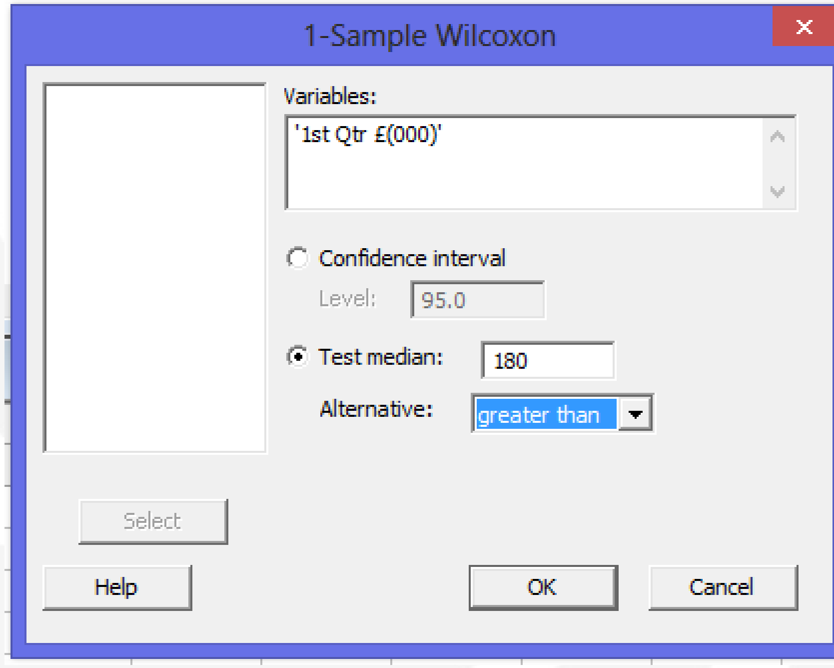
	N	Below	Equal	Above	P	Median
1st Qtr £(000)	10	4	0	6	0.3770	181.1

Same result as parametric equivalent

Exercise 4, Non-parametric Test

An alternative test …. 1 sample Wilcoxon

Stat> Non-parametrics >1-Sample Wilcoxon



Same result

Wilcoxon Signed Rank Test: 1st Qtr £(000)

Test of median = 180.0 versus median > 180.0

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
1st Qtr £(000)	10	10	41.0	0.093	181.1

1-Sample Sign Test

- **Used when evaluating whether a median is larger than a standard but not by how much**
- **Here's how it works:**
  - **The formula breaks the data into discrete buckets of 'below the median', 'equal to the median' and 'above the median.**
  - **It then does statistical tests on whether there is a difference in the size of the buckets.**

Wilcoxon Signed Rank Test

- **The test assumes that the population, while not necessarily normal, is symmetric.**
- **How it works:**
  - **Subtract each data point from the hypothesized median.**
  - **Rank the absolute value of those differences.**
  - **Sum up the ranks of the differences that were above the hypothesized median and separately sum the ranks below.**
  - **Subtract the sum of the negatives from the sum of the positives.**
  - **If the null is true, this difference of the sums of the ranks will be zero. Minitab does the statistics giving us a p-value to indicate whether this difference is significantly different than zero.**

More powerful than 1-sample sign as it does not break the data into binary categories but does not work for discrete data and assumes that the population, while not normal, is symmetric

Two Levels - Exercise 5

- We currently use a title company in St. Louis to do our title searches for Home Loans & Insurance (HL&I). But we have a competitor in Omaha who offers to do them at a 20% price reduction. We are always interested in reducing costs, but we don't want to sacrifice service. Service in this application can be measured by the average hours to return a title search. As long as Omaha is not over 2 hours longer than St. Louis, we will be fine.
- We set up a test to run for three hours. During that time we will route some titles to St. Louis and some to Omaha. Then we will measure how long it takes for them to come back.
- The test results are back! Is it worth our while to get more data? If so, how many samples might we need to make sure we don't miss a difference of 2 hours?

Reminder

We have some data assumptions to be checked before we can launch into a 2 sample t-test (from earlier)

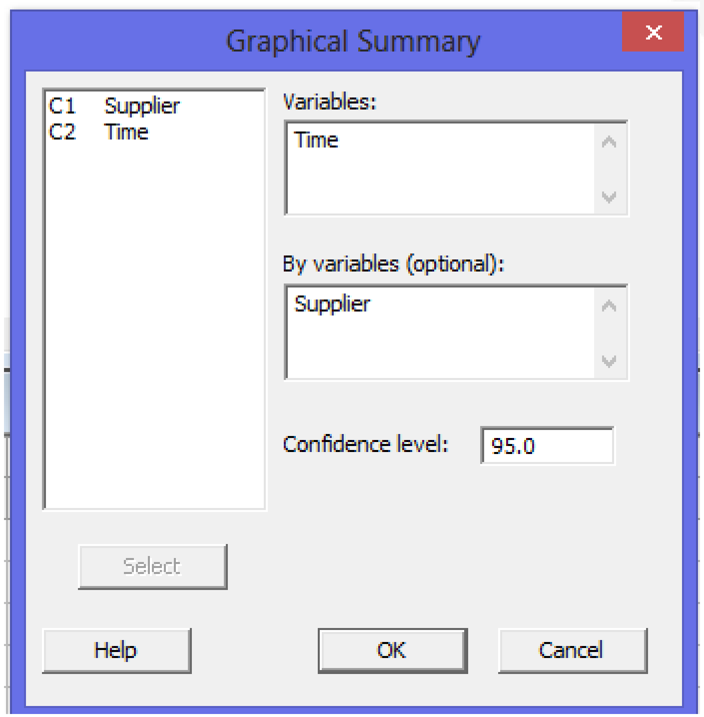
There are basic assumptions which must be met for a t-test to be robust

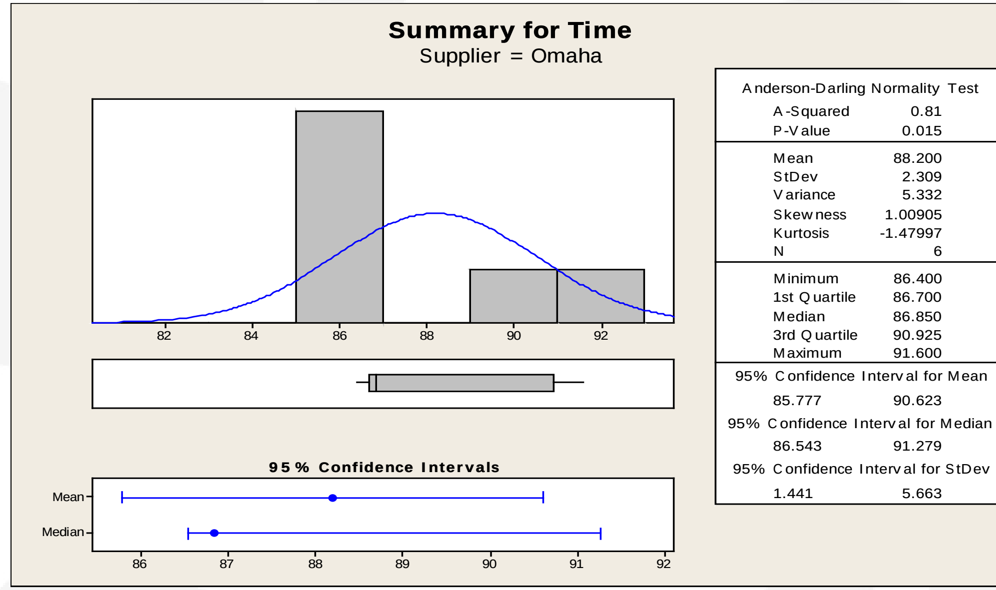
- Response is normally distributed (each level)
- Response is independent (each level)
- Equal variance (for one to one comparison only, each level)

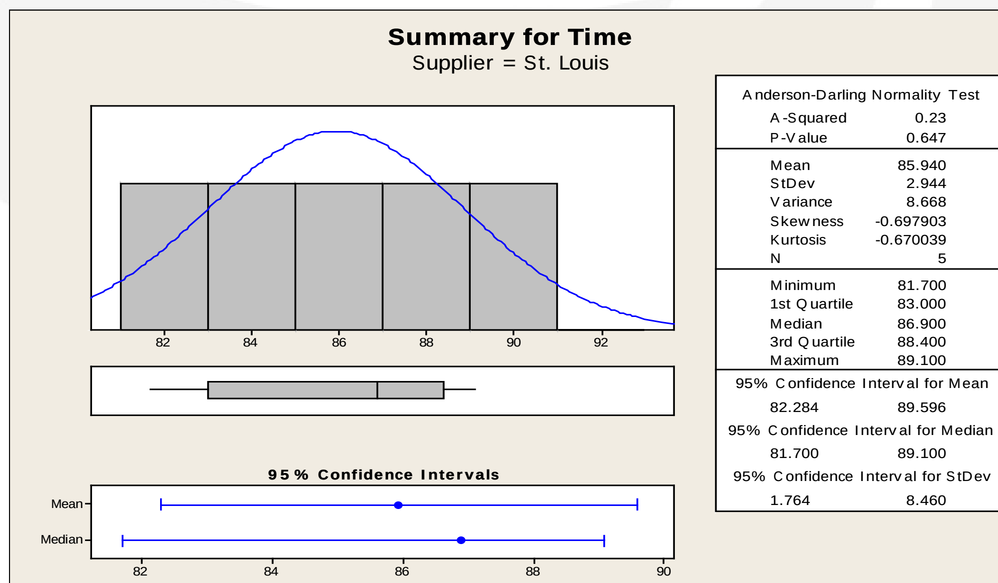
Exercise 5, Checking Normality

Data Table:
Hyp_I_Two_Levels_Ex5.MPJ

Stat > Basic Statistics > Graphical Summary

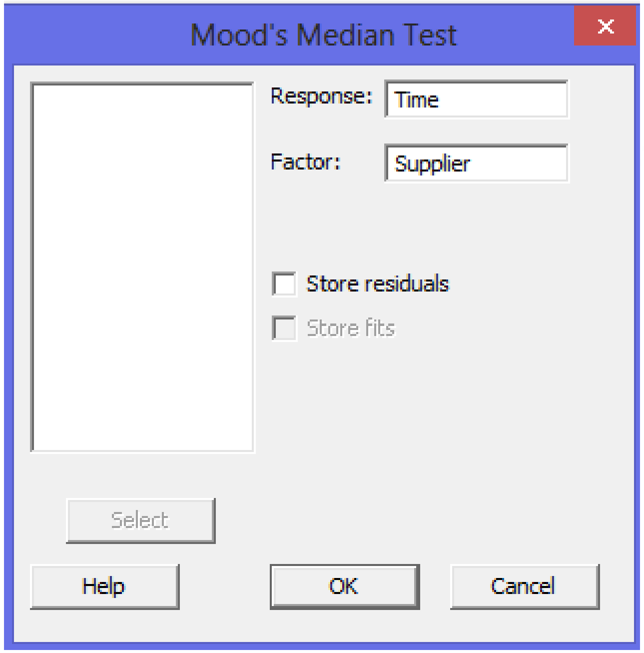






Can we assume normality?

Exercise 5, Run Moods Median test



Mood Median Test: Time versus Supplier

```
Mood median test for Time

Chi-Square = 0.05    DF = 1    P = 0.819
                                        Individual 95.0% CIs
Supplier   N<=  N>  Median  Q3-Q1  --------+---------+---------+--------
Omaha        4   2   86.85   4.23                (*-------------)
St. Louis    3   2   86.90   5.40  (---------------*------)
                                   --------+---------+---------+--------
                                          84.0      87.0      90.0

Overall median = 86.90
* NOTE * Levels with < 6 observations have confidence < 95.0%
```

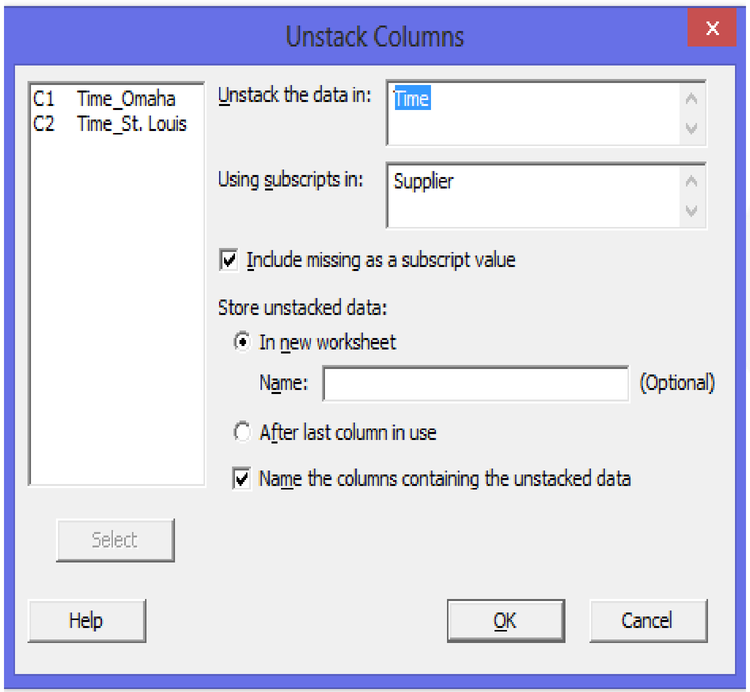
Use this as your p-value.
Fail to reject the null hypothesis

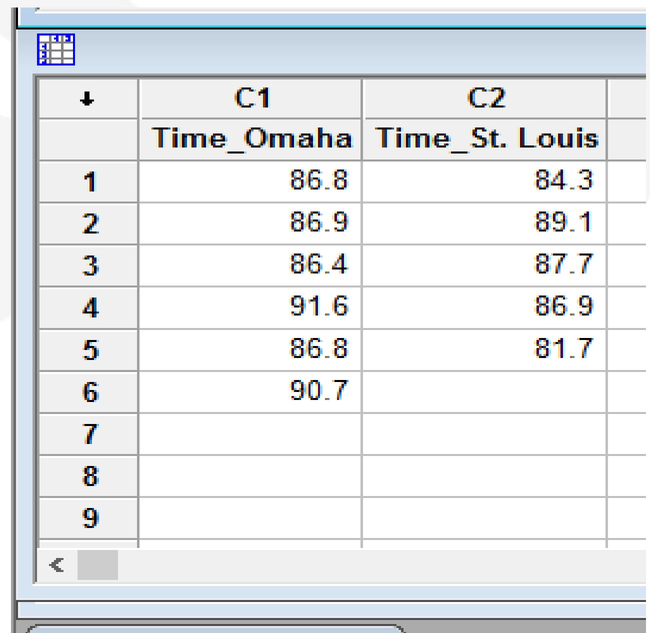
Exercise 5, Unstacking columns illustration

As we are comparing two samples of non-normal continuous data (so they have a similar shape), we can also use the Mann-Whitney test.

To allow us to perform our statistical test we'll need to unstack our collected data……

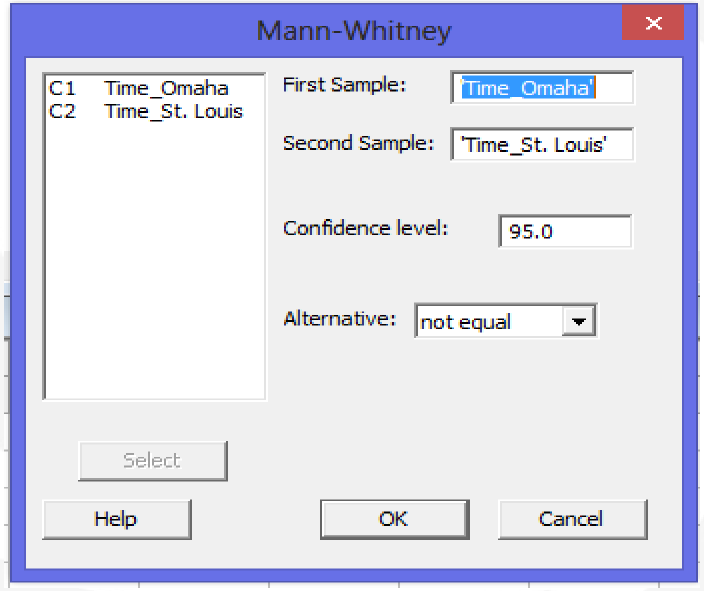
<u>Data > Unstack Columns</u>





Exercise 5, Run the test

Your data is now in appropriately formatted for the test



Mann-Whitney Test and CI: Time_Omaha, Time_St. Louis

```
                N  Median
Time_Omaha      6  86.850
Time_St. Louis  5  86.900


Point estimate for ETA1-ETA2 is 2.500
96.4 Percent CI for ETA1-ETA2 is (-2.199,6.401)
W = 39.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.5839
The test is significant at 0.5822 (adjusted for ties)
```

Use this as your p-value.
Fail to reject the null hypothesis

Two Medians

Mann-Whitney Test:

Assumes that the samples have similar shapes even though they don't have to be normal.

Works by combining the data points, sorting them numerically and assigning a rank (1st, 2nd, 3rd, etc) to each point. Then ranks for each level are summed. If the null hypothesis is true (there is not a difference in the population medians), then the sum of the ranks for each level will be the equal.

Moods Median Test:

Is very robust to the shape of the data and outliers. This is the "workhorse" or non-parametric testing.

Works by calculating an overall median of the data points. It then tests the proportion of points above and below that overall median at each level.

Next Steps?

- Since we had not been able to prove a difference between the two suppliers, are we ready to switch suppliers?
- What should be our next step?
- What sampling strategy should we use?

Exercise 6, Two Levels

- For the next test of St. Louis and Omaha, we want to send 50 titles to each and measure the time to return.
- We set up the next test to send 50 each to both suppliers. Once again we measure the time it takes for them to come back. If Omaha isn't more than 2 hours slower than St. Louis, we will pursue a switch.
- Test results are back! What do we decide?

T-test assumptions

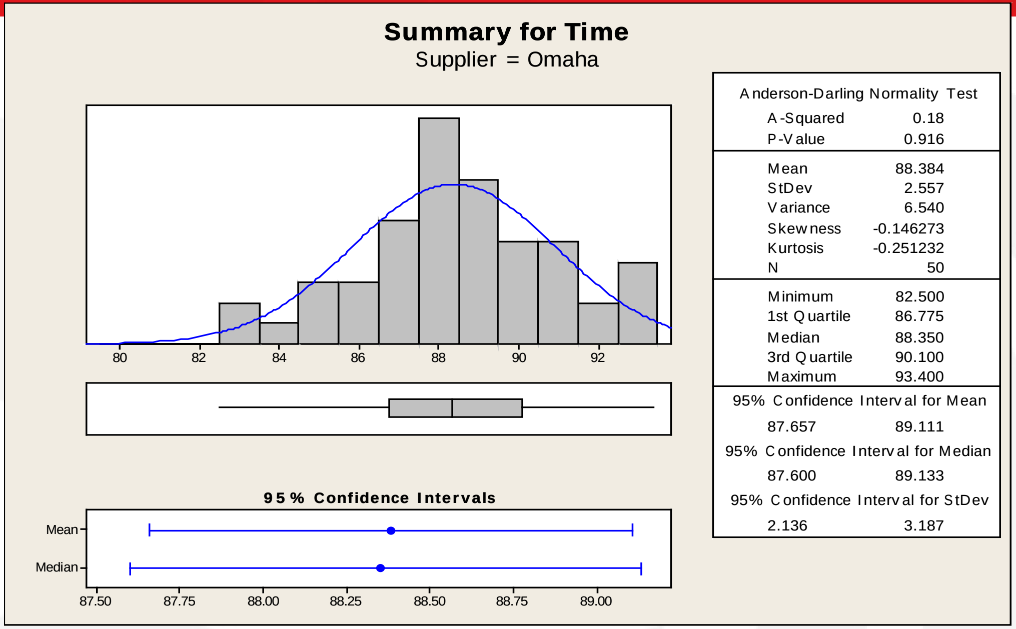
There are basic assumptions which must be met for a t-test to be robust

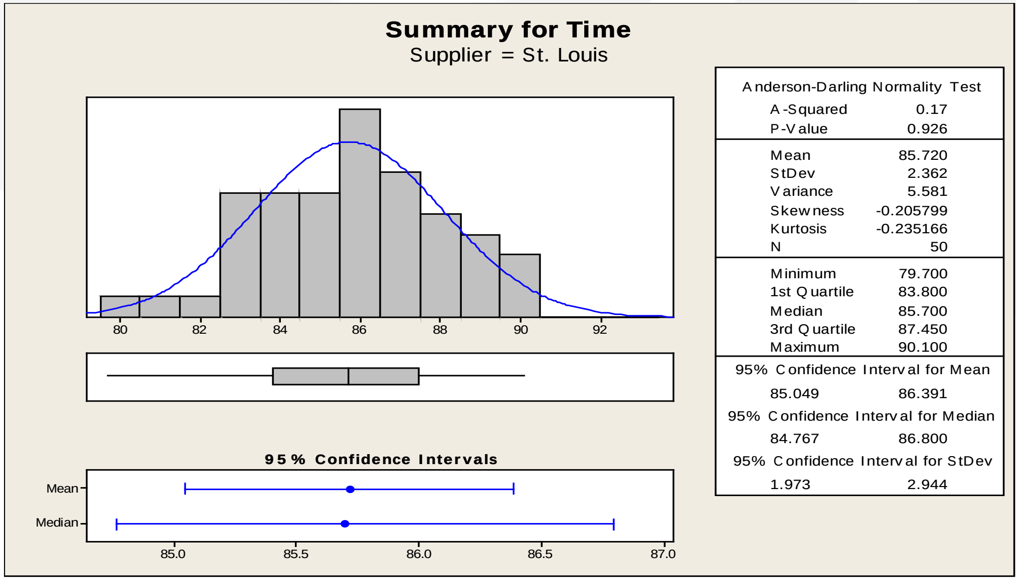
- Response is normally distributed (each level)
- Response is independent (each level)
- Equal variance (for one to one comparison only, each level)

Exercise 6, Checking Normality Assumption

Stat > Basic Statistics > Graphical Summary

Data Table:
Hyp1_Two_Levels_Ex6.MPJ

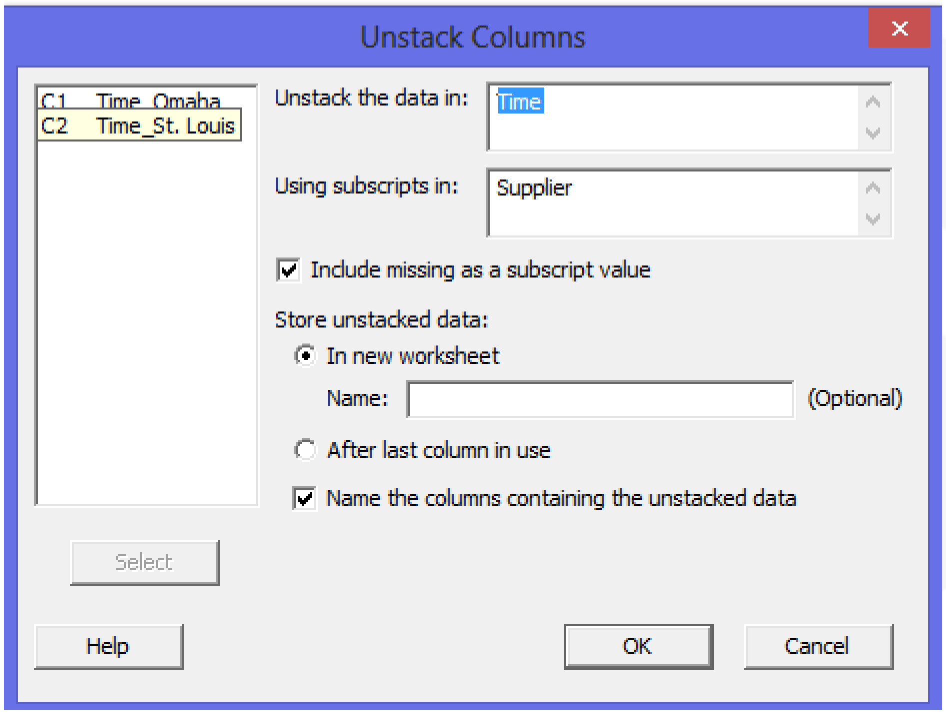


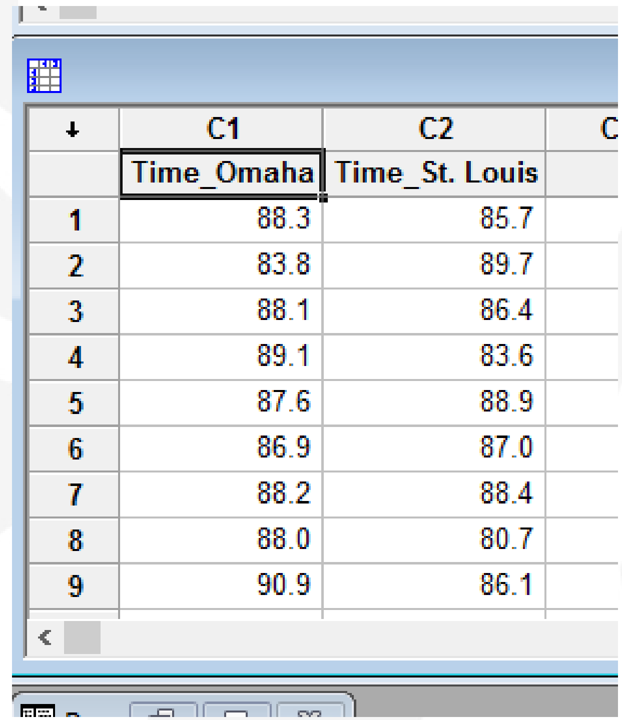


Can we assume normality?

Exercise 6, Checking Independence Assumption

Unstack columns to create a run chart for both cities

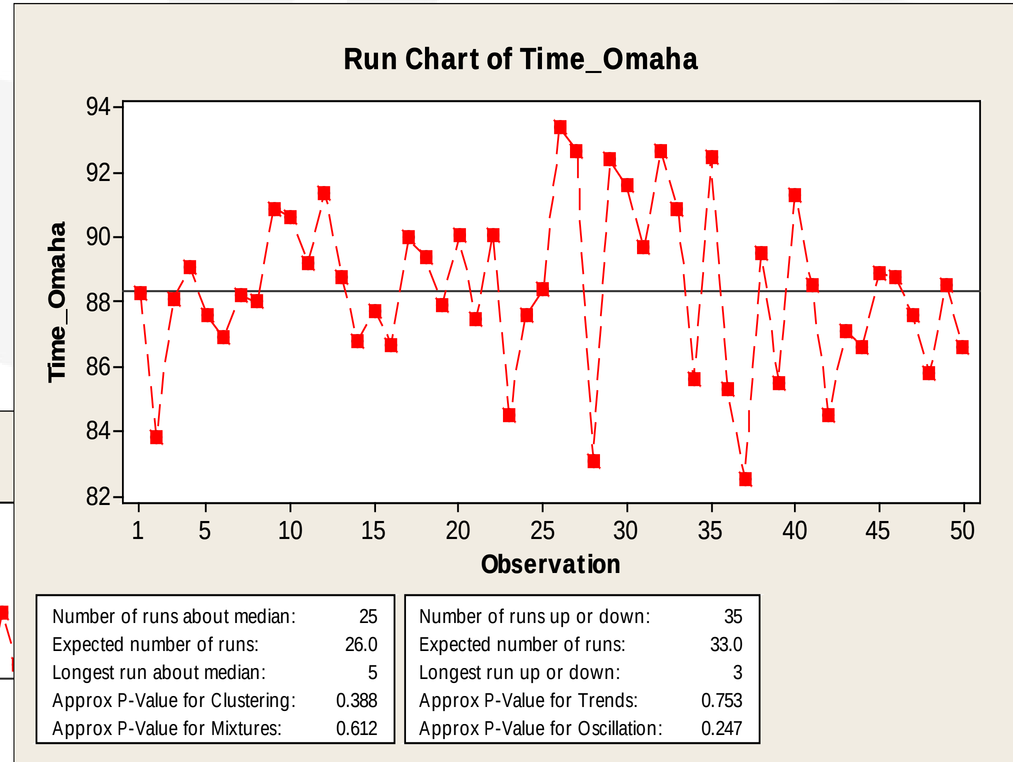


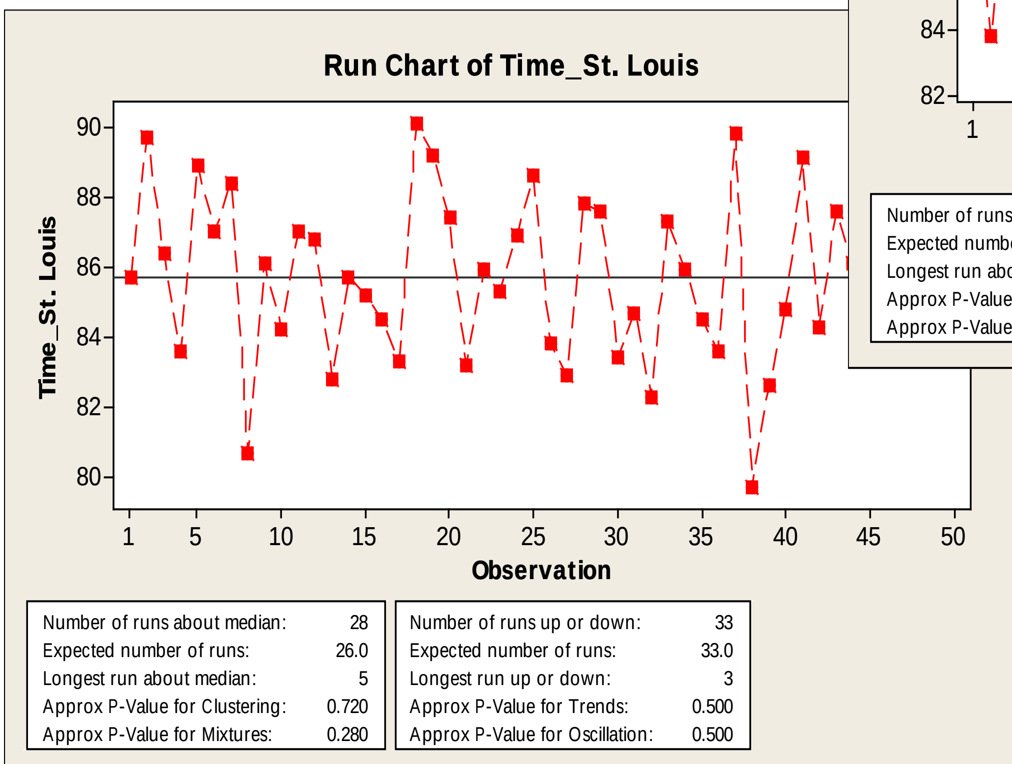


Exercise 6, Independence (cont'd)

Create a Run Chart for each city

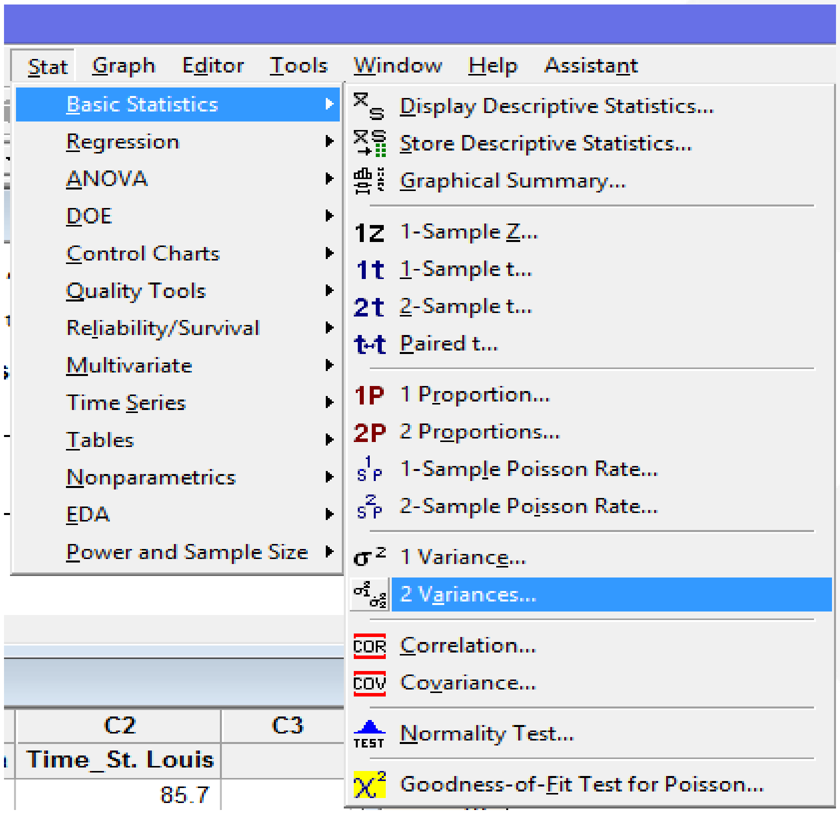
Stat > Quality Tools > Run Chart

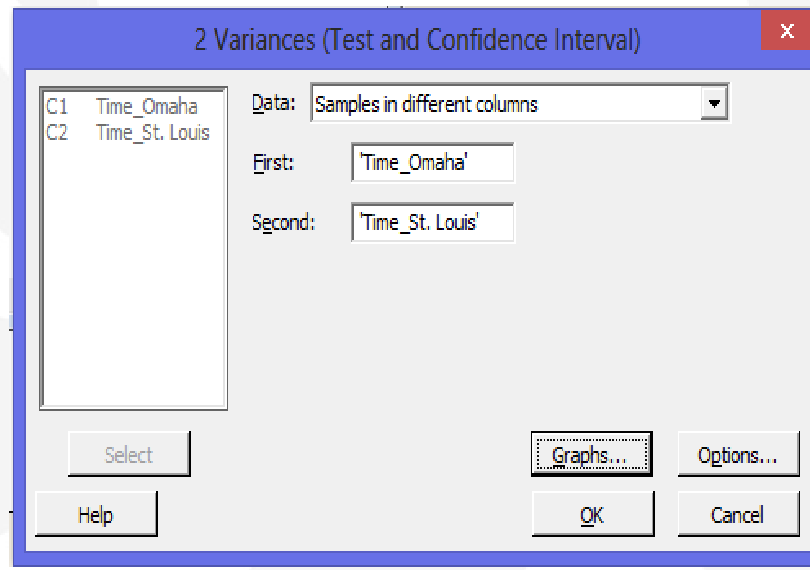




Safe to continue?

Exercise 6, Checking Equal Variance Assumption





Exercise 6, Checking Equal Variance Assumption (cont'd)

Test and CI for Two Variances: Time_Omaha, Time_St. Louis

Method

Null hypothesis	Sigma(Time_Omaha) / Sigma(Time_St. Louis) = 1
Alternative hypothesis	Sigma(Time_Omaha) / Sigma(Time_St. Louis) not = 1
Significance level	Alpha = 0.05

Statistics

Variable	N	StDev	Variance
Time_Omaha	50	2.557	6.540
Time_St. Louis	50	2.362	5.581

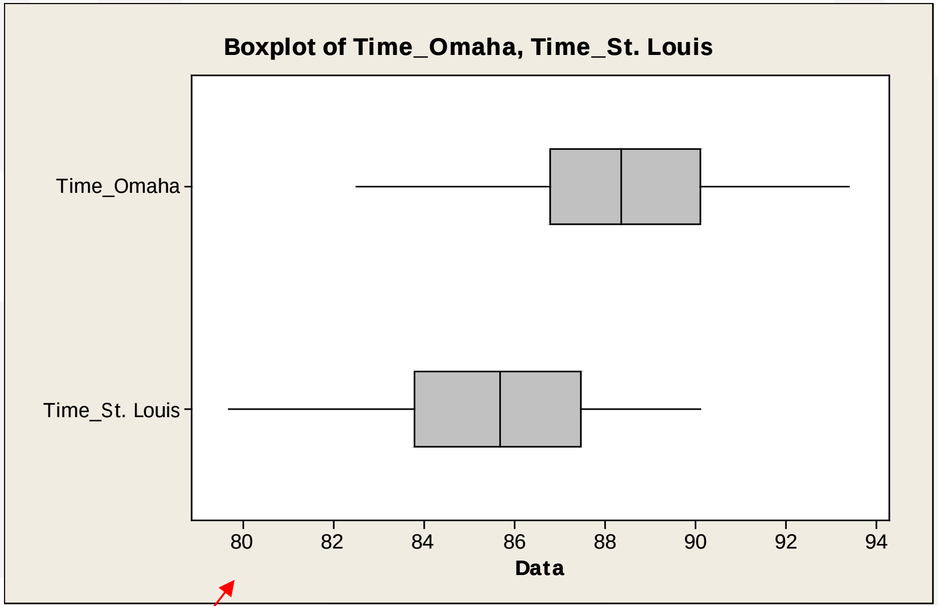
Ratio of standard deviations = 1.083
Ratio of variances = 1.172

95% Confidence Intervals

Distribution of Data	CI for StDev Ratio	CI for Variance Ratio
Normal	(0.815, 1.437)	(0.665, 2.065)
Continuous	(0.772, 1.438)	(0.596, 2.069)

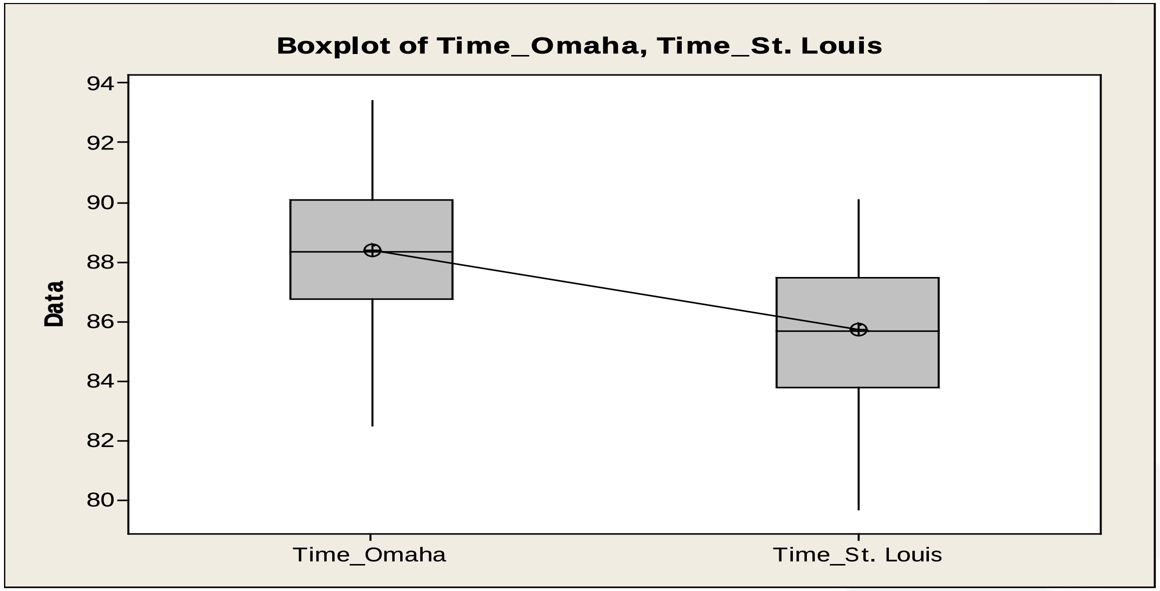
Tests

Method	DF1	DF2	Test Statistic	P-Value
F Test (normal)	49	49	1.17	0.581
Levene's Test (any continuous)	1	98	0.13	0.718



Graphically
Rule of thumb – check for a 3:1 difference in the spread

Exercise 6, Run the test



Graphically, does the time look different by city?

Two-Sample T-Test and CI: Time_Omaha, Time_St. Louis

```
Two-sample T for Time_Omaha vs Time_St. Louis

                 N   Mean  StDev  SE Mean
Time_Omaha      50  88.38   2.56     0.36
Time_St. Louis  50  85.72   2.36     0.33

Difference = mu (Time_Omaha) - mu (Time_St. Louis)
Estimate for difference:  2.664
95% CI for difference:  (1.687, 3.641)
T-Test of difference = 0 (vs not =): T-Value = 5.41  P-Value = 0.000  DF = 97
```

Confidence interval for the difference does not “straddle” zero

P-value <0.05, what do we conclude?

Exercise 6, Answer the business question

Always ensure you answer the business question.

Telling a senior stakeholder that because the p-value is low, the null must go won't go down too well!

Business question:	Is there a difference in processing times between St Louis and Omaha?
Business answer:	Yes, St Louis is faster. The estimate for how much faster is between 1.6 and 3.6 hours. Omaha could be more than two hours slower so we should not change supplier based on our results

How could we be more specific about the difference?

Two Means, Paired Data

Paired t Test:

- **Paired comparisons are used when a one-to-one correspondence exists between data in the populations (e.g. same customer before and after or same employee using two methods).**
- **It is particularly useful when there is more variation within one level of the factor, than there is between the levels (e.g. the before and after populations have a standard deviation of 20, but the average change to each pair is only 3).**
- **In paired comparisons we work with the difference between the paired observations. The test essentially performs the one-sample t-test on these differences to determine if the mean is significantly different than a specified delta (d), usually zero.**

Exercise 7, Two Means, Paired Data

Data Table: Hyp1_Paired_Data_Ex7.MPJ

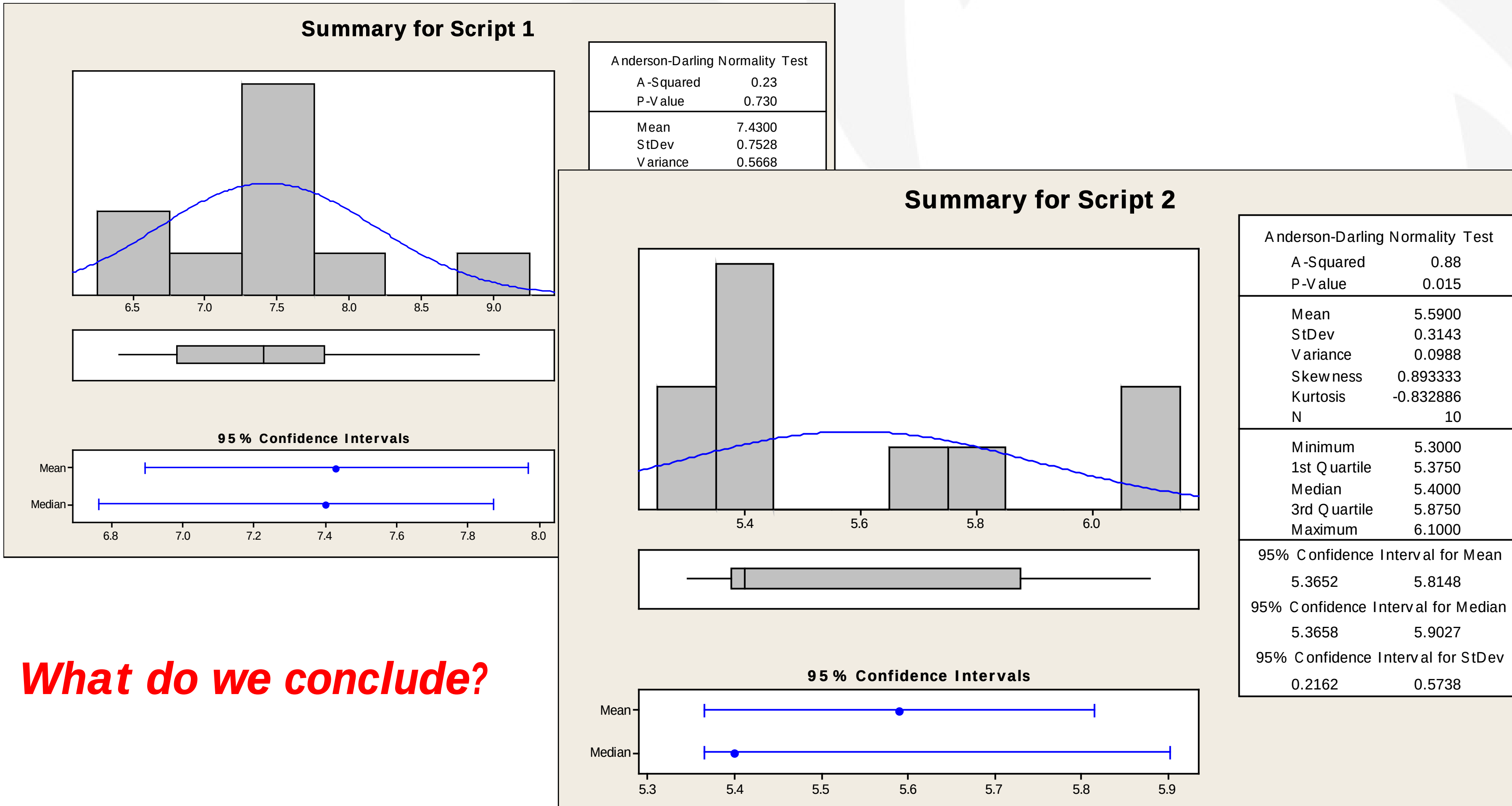
- We want to increase the number of referrals from our call centre for sales of other products. The project team has come up with two potential scripts that they think may help the new process. They would like to test these to see if one is better than the other.

- How might we set up a sampling plan? There is a lot of variation between staff, between calls and between days.

- We decide set up a test plan that we will run for 4 weeks. We are going to have 10 call takers try out the new scripts. Half the staff will start with script 1 and half with script 2. Every 7 days, the staff will switch to the other script. We collect data and record the average number of referrals per shift.

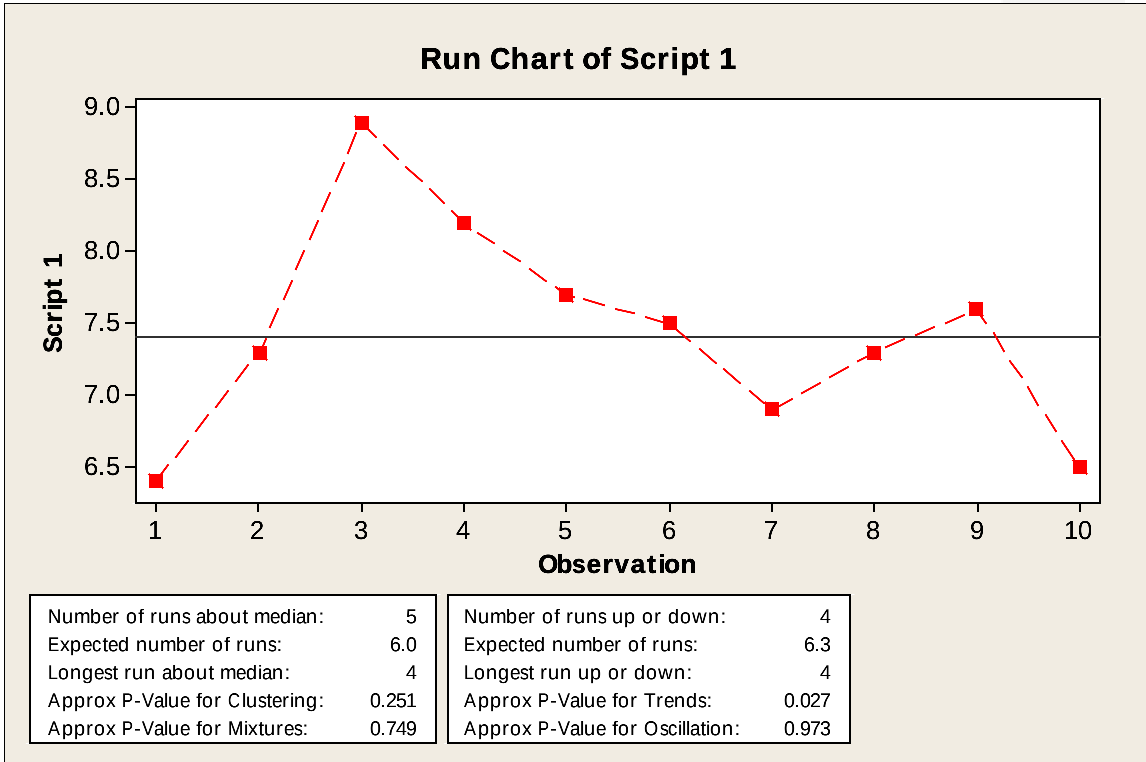
Assess Normality

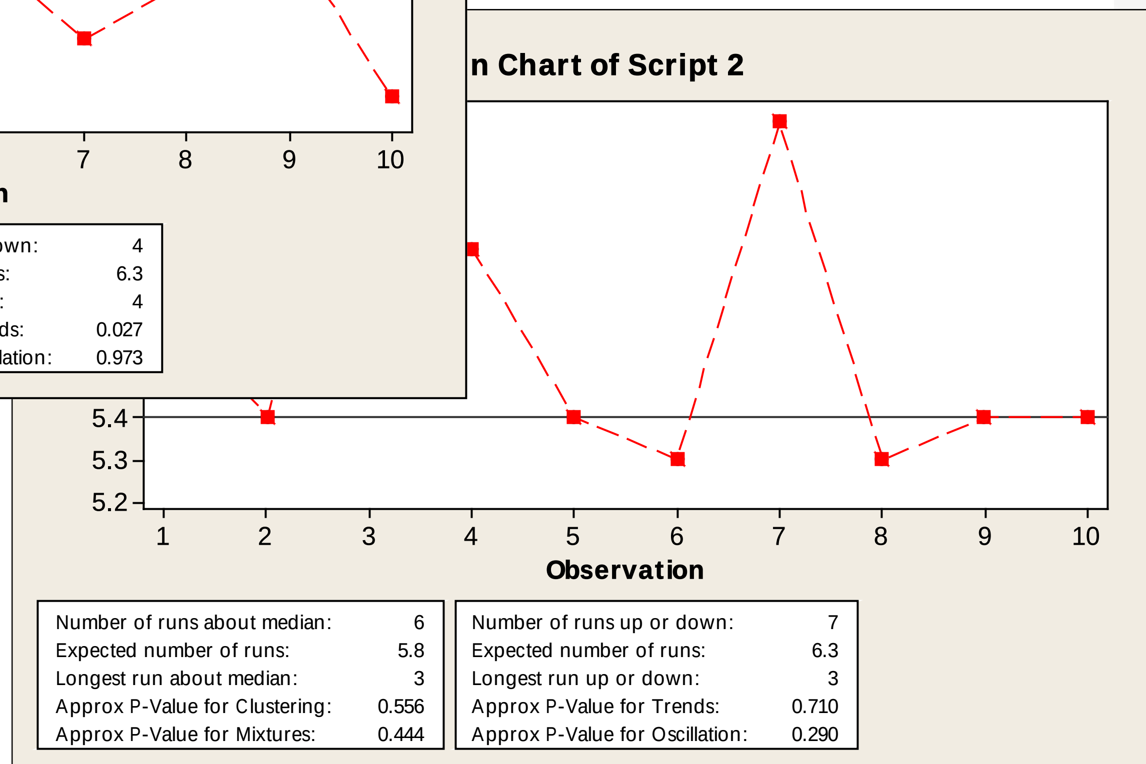
Stat > Basic Stats > Graphical Summary



What do we conclude?

Assess Independence



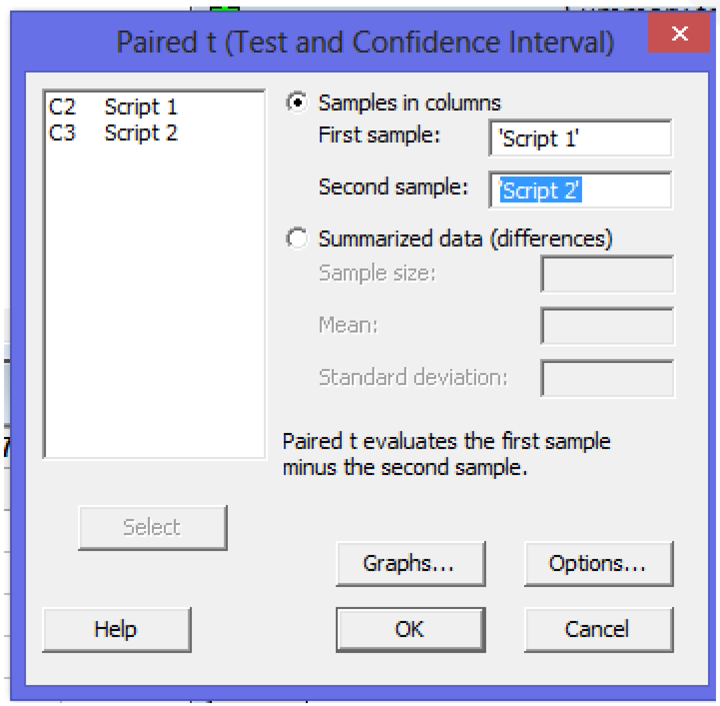


What do we conclude?

Exercise 7, Run the test

Our data is not normal but for illustrative purposes, lets run a paired t-test

<u>Stat > Basic Statistics > Paired t</u>



Paired T-Test and CI: Script 1, Script 2

Paired T for Script 1 - Script 2

	N	Mean	StDev	SE Mean
Script 1	10	7.430	0.753	0.238
Script 2	10	5.590	0.314	0.099
Difference	10	1.840	0.720	0.228

95% CI for mean difference: (1.325, 2.355)

T-Test of mean difference = 0 (vs not = 0): T-Value = 8.08 P-Value = 0.000

Two Medians: Paired data

Wilcoxon Signed Rank Test:

This is the non-parametric test corresponding to the paired t-test.

Works by ranking the data in each column and then subtracting the ranks to get a difference. Like the Paired t-Test, if the null hypothesis is true, each member of the pair will have the same rank and the difference will be zero.

There is no function in Minitab to perform the Wilcoxon Signed Rank Test on paired data.

A t statistic is manually calculated and compared to the critical t-statistic looked up in the t Table.

Summary of key learning points

- Ensure that the 6 step hypothesis testing approach is followed to move between business problem and business solution
- Alpha and Beta risks are key considerations when hypothesis testing
- Sampling for hypothesis testing is different from sampling for estimation
- Hypothesis tests are either parametric or non-parametric
- Parametric tests have some assumptions which must be satisfied for the results to be relied upon

Recommended coach support points

- When considering alpha risk, power and difference to detect
- Helping you choose the appropriate hypothesis test given the business question to be solved
- When preparing statistics on non-parametric paired data
- Validating your results before sharing them with the business

Black Belt

Coaching

Objectives

One of the roles of a Black Belt is to provide coaching to Green Belt candidates.

- **Objective:**
  - **Improve coaching skills through role playing and feedback.**
  - **Experience coaching Lean Six Sigma concepts.**

Qualities of a Great Coach

- Think back to some of the best coaches you have had—maybe a parent, a teacher, a sports coach . . .

- What made him/her a great coach?

Exercise

- Break into teams for the exercise.
- Choose a role for the first scenario.
  - Green Belt, Black Belt, Observers
- GB selects a question to be coached on, BB coaches (5-10 min.), Observers give feedback on coaching techniques (5-10 min.).
- Repeat until everyone has had an opportunity to be in each role.

Green Belt Role

- Select a question you would like coaching on. If possible, it should be a real question from your project. That way you have a context that the BB can work within. There are some suggestions on the next slide.

- Be earnest. Be yourself. Don't try to "act" like a difficult person.

Possible GB Questions

- Where might a control chart be used in my project?
- What is the best brainstorming tool to use?
- I need to get better VOC.
- Could you review the hypothesis testing roadmap and help me chose which test to use?
- Where should I use an MSA in my project?
- Where do I go with my project next?
- What can I do to find the root cause?
- How can I find more cost savings in my project?

Black Belt Role

- **Treat the scenario as if it were a real coaching session with a GB.**
- **Try to provide guidance and useful information to the GB.**
- **Try to help the GB move their project along.**

Observers Role

- **Silently observe the coaching session.**
- **Make notes if needed.**
- **At the end, provide feedback to the BB on their coaching (some suggestions on the next page).**

Feedback Suggestions

- Did the BB seek to understand the problem thoroughly before giving coaching?
- Did the BB make good use of questions? Leading questions?
- Did the BB tell the GB what to think or lead and challenge the GB to think?
- Did the BB speak to the issue or get off track?
- Was the issue addressed?
- How simple or complex did the BB make the coaching?
- How well did the BB do technically?
- Any other suggestions?

Black Belt

Analysis of Variance (ANOVA)

Objectives of this module

At the end of this module, you will be able to:

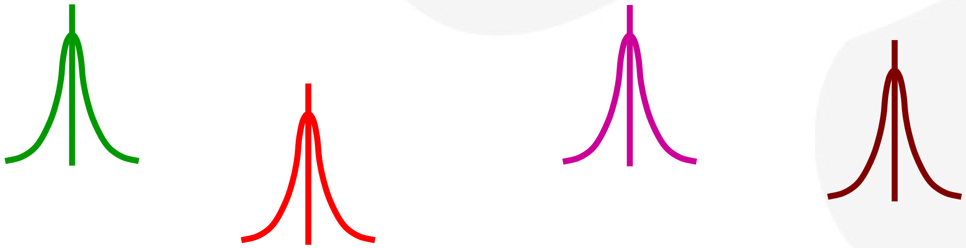
- Understand how Analysis of Variance (ANOVA) works to determine statistical significance of the relationship between a continuous Y (dependent variable), and single or multiple Xs (independent variables).
- Understand the concepts of within and between variation and signal to noise ratios.
- Be able to perform a one-way ANOVA to determine differences between means of different levels of one factor, e.g., Y = Sales £ per store; Factor = Shelf Location; Levels = Top, Middle, Bottom.
- Be able to perform non-parametric tests to determine differences between medians of different levels of a factor and know when to use these tests instead of a one-way ANOVA.
- Be able to perform a two-way ANOVA to identify the impact different levels of two factors and their interactions have on the Y

Questions Lead

- Basic question: Is there a difference in the output between these groups?
  - Manufacturing – An engineer is trying to determine whether there is any relationship between the machine on which the part was made and its polishing time
  - Transactional – A manager wants to understand how different attendance policies impact productivity
  - Marketing – A marketing executive want to know whether store location has an impact on sales

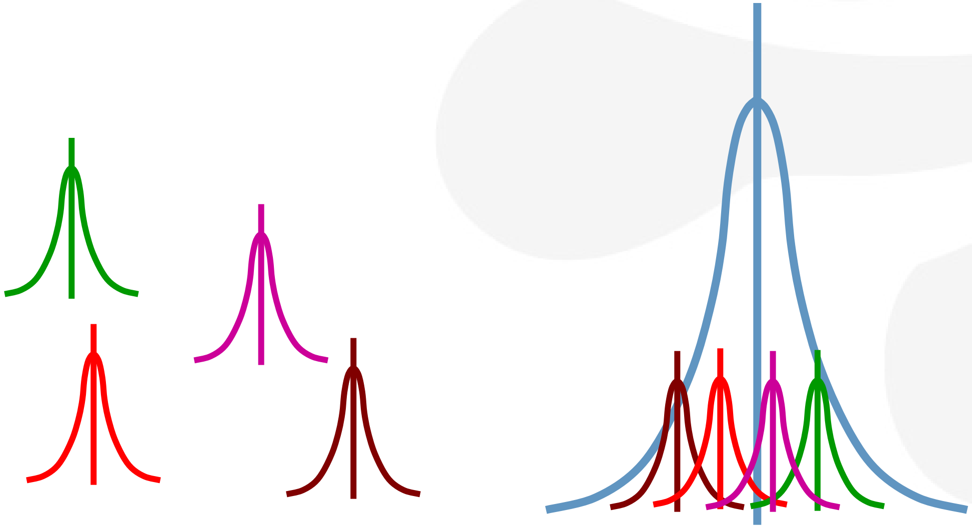
How does ANOVA work?

- How can an analysis of variance determine if there are differences between the means?
- Suppose we had a factor, X, with four levels and we sampled the process and measured the Y value several times at each of the four levels and the data looked like the four small distributions below.
- Our question is whether this factor is one of the vital few Xs. Is the average of the Y at each of the four levels the same or different?



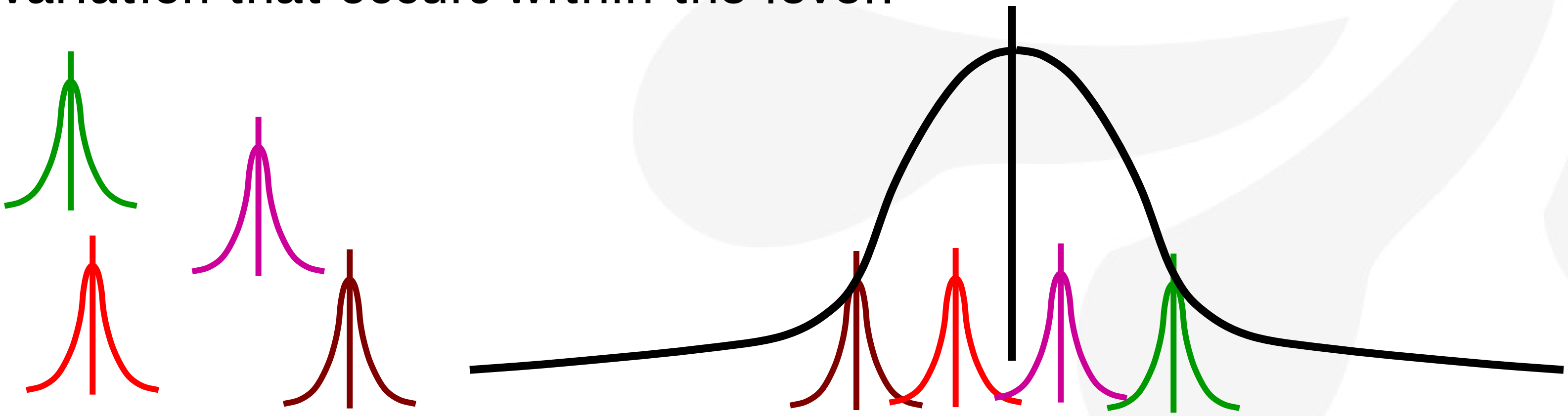
How does ANOVA work?

- Suppose we combined all of the data from the four levels into one group and drew another histogram of those data and it looked like the tall distribution below.
- Visualise all four of the small distributions fitting into the larger one.
- Is there any way the means of the four small ones can be very far from each other and the variation of the large distribution be so tight?

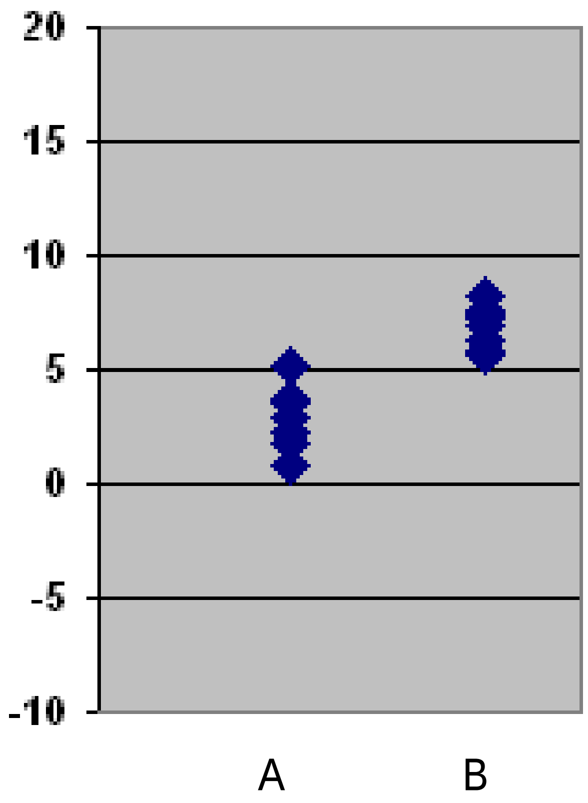


How does ANOVA work?

- Again visualise all four of the small distributions fitting into the larger one.
- Is there any way the means of the four small ones can be very close to each other and the variation of the large distribution be so wide?
- ANOVA determines if the variation between the averages of the levels is greater than could reasonably be expected from the variation that occurs within the level.

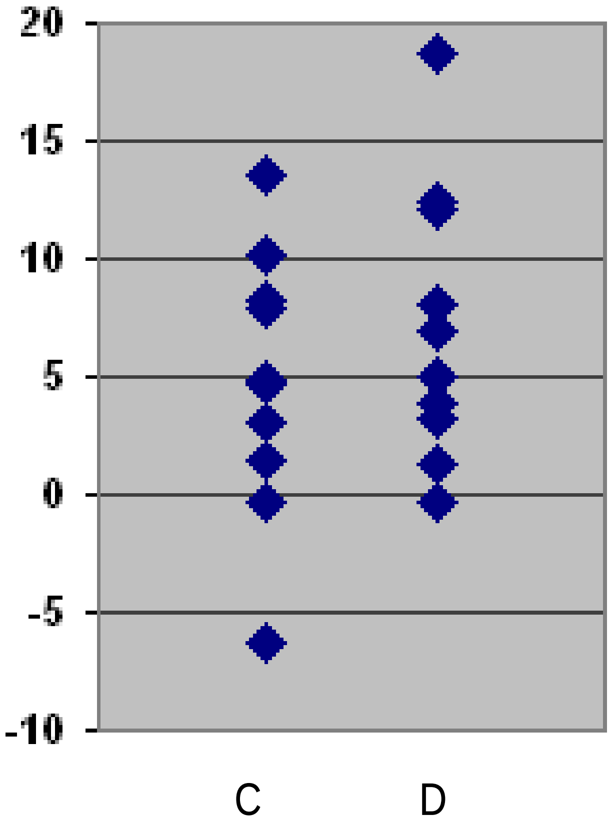


Within and Between Variation (1 of 3)



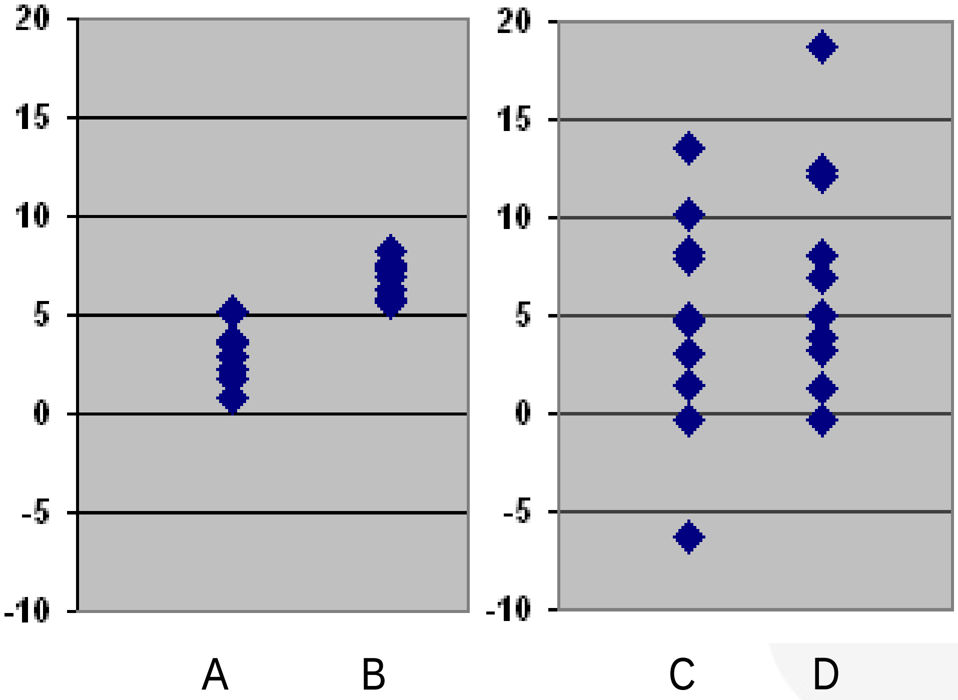
- Our process was sampled at A and B, n = 10.
- Are the means of A and B the same or different?
- If you said ‘different’, how confident are you?

Within and Between Variation (2 of 3)



- This process was sampled at C and D, n = 10.
- Are the means of C and D the same or different?
- If you said 'different', how confident are you?

Within and Between Variation (3 of 3)



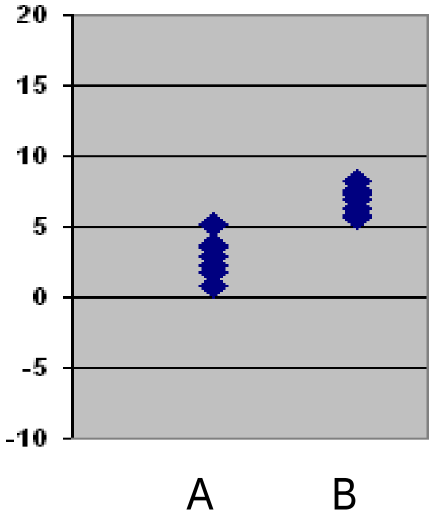
- Population Truth:

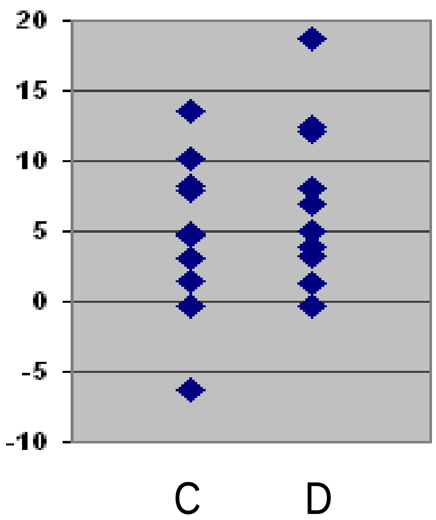
  $m_A = 3$ $m_B = 7$

  $m_C = 3$ $m_D = 7$

- The truth is that there is exactly the same difference between C and D as there is between A and B.
- Why were you confident that A and B were different, but not confident that C and D were different?

Signal and Noise





- More Population Truth:

  $m_A = 3 \quad s_A = 1 \qquad m_B = 7 \quad s_B = 1$

  $m_C = 3 \quad s_C = 7 \qquad m_D = 7 \quad s_D = 7$

- Xbar is the sometimes called the signal from the data, and s is sometimes called the noise in the data.
- Describe your confidence in the equality of the means using the words "signal" and "noise".

Terminology & Types of ANOVA

Terminology:

- Factor – An X. An input to a process. An independent variable in a statistical test or experiment.
- Level – A discrete value or setting for a factor.

Types of ANOVA:

- One-way ANOVA – A test for a single factor, X, that has more than two levels and a continuous Y response variable. Used to test for differences in the Y when measured at different levels of the factor. Are the means of the Y the same when the X goes from one level to another?
- Two-way ANOVA – A test for multiple factors, two or more Xs that have multiple levels and a continuous Y response variable. Used to see which Xs have a statistically significant effect on the Y.

One-Way ANOVA

A one-way ANOVA is a hypothesis test, and it follows the same procedure starting with the hypothesis statements.

H_0: $m_1 = m_2 = m_3 = m_4 \; . \; . \; . = m_k$

H_a: At least one m_k is different

Why wouldn't we just do a series of t-tests?

Exercise 1, One-way ANOVA

A manager at ABC Retail needs to determine if there are different starting salaries at four different shopping centres in Newtown after an assistant complained that her friends in the other shopping centres are all making more money with the same experience. Below are the salary data for the 24 assistants that we hired within the last 3 months.

- How many factors do we have?
- How many levels are there?
- Could there be other factors?

Centre 1	17.7	16.8	18.1	16.3				
Centre 2	18.1	19.9	21.7	18.6	19.0	19.5		
Centre 3	20.4	19.5	21.7	19.9	20.4	20.4		
Centre 4	15.0	17.7	16.8	17.2	18.1	18.6	18.1	16.3

(Salaries £000)

ANOVA Fundamentals (1 of 2)

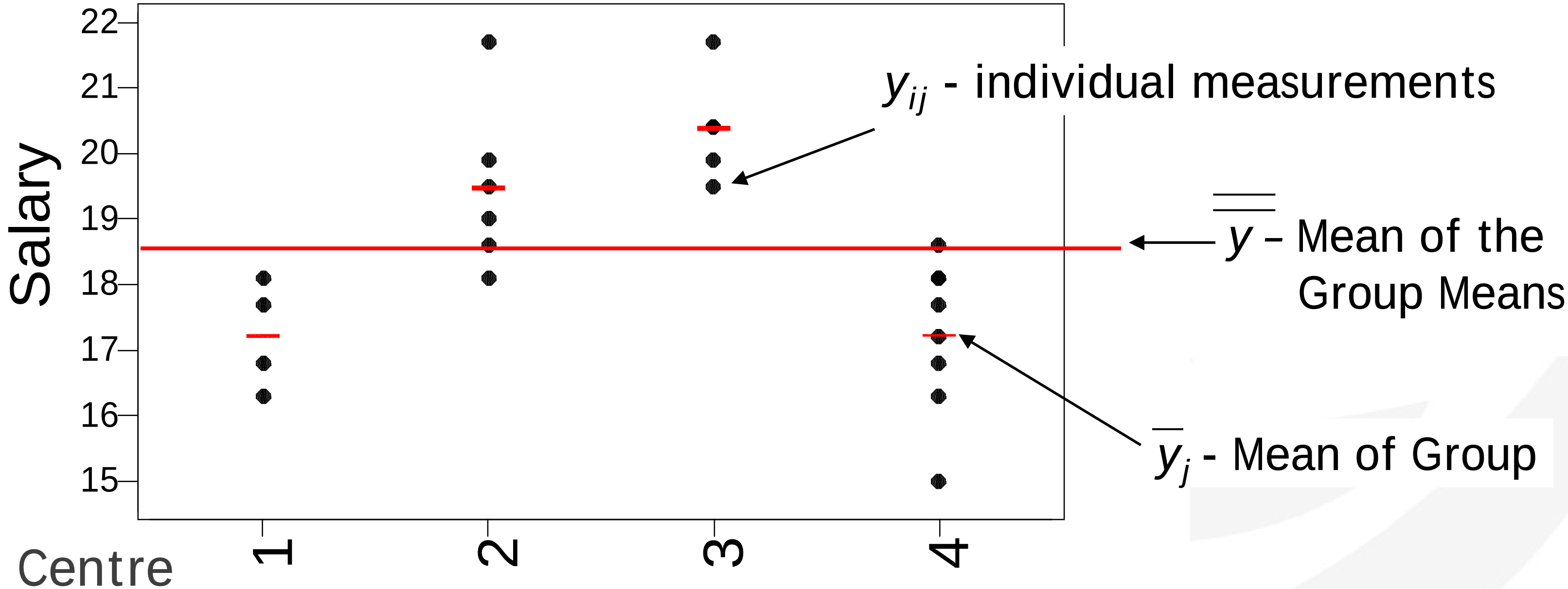
ANOVA looks at three sources of variation:

- Total = total variation in all observations
- Between = variation between levels of a factor
- Within = variation within each level (noise or statistical error)

Total = Between + Within

ANOVA Fundamentals (2 of 2)

Dotplots of salary by centre
(group means are indicated by red lines



Centre 1	17.7	16.8	18.1	16.3				
Centre 2	18.1	19.9	21.7	18.6	19.0	19.5		
Centre 3	20.4	19.5	21.7	19.9	20.4	20.4		
Centre 4	15.0	17.7	16.8	17.2	18.1	18.6	18.1	16.3

(Salaries in thousands of pounds)

Sum of Squares

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij}-\bar{\bar{y}})^2=\sum_{j=1}^{k}n_j(\bar{y}_j-\bar{\bar{y}})^2+\sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)^2$$

SS(Total) = SS(Between) + SS(Within)

SS (Total) = Individuals - Grand mean
SS (Between) = Group mean - Grand mean
SS (Within) = Individuals - Group mean

The Sums of Squares equation is the method for calculating the between and within variation.

The ANOVA Table

To determine whether we can reject the null hypothesis we must calculate a test statistic (F ratio) using the Analysis of Variance table.

SOURCE	DF	SS	MS	F	P
Factor (Between)	k-1	SS (Factor)	SS(Factor) / (k-1)	MS(Factor) / MS (Within)	
Error (Within)	[(n-1) – (k-1)]	SS (Error)	SS (Error) / [n – k]		
Total	n-1	SS (Total)			

Source? Source of what?
Why is the source “Between” called Factor?
Why is the source “Within” called Error?
What is another term for Error?
In practical terms what is the F ratio telling us?
What do you think large F ratios mean?

What are Degrees of Freedom?

degrees of freedom ≡ currency in statistics

We earn a degree of freedom for every data point we collect.

In our example there were 24 data points, we earned 24 degrees of freedom (DF).

We spend a degree of freedom for each parameter we estimate.

To calculate the Sum of Squares what parameters did you have to calculate? (Hint: See slides 17 and 18.)

In ANOVA, the degrees of freedom are as follows:

$$df_{total} = n - 1 = \# \text{ of observations} - 1$$

$$df_{factor} = k - 1 = \# \text{ of levels (or groups)} - 1$$

$$df_{within} = df_{total} - df_{factor}$$

One-way ANOVA: Salary versus Centre

Source	DF	SS	MS	F	P
Centre	3	46.81	15.60	13.79	0.000
Error	20	22.62	1.13		
Total	23	69.43			

Degrees of Freedom

Sums of Squares

Mean Sums of Squares = SS/DF

F = 13.79 = 15.60/1.13 = MS(Between)/MS(Within)

The $F_{critical}$ for this example is $F_{3,20,.05} = 3.10$

With a p-value of <.0001, what conclusions can you draw about H_0? Reject or Fail to Reject?

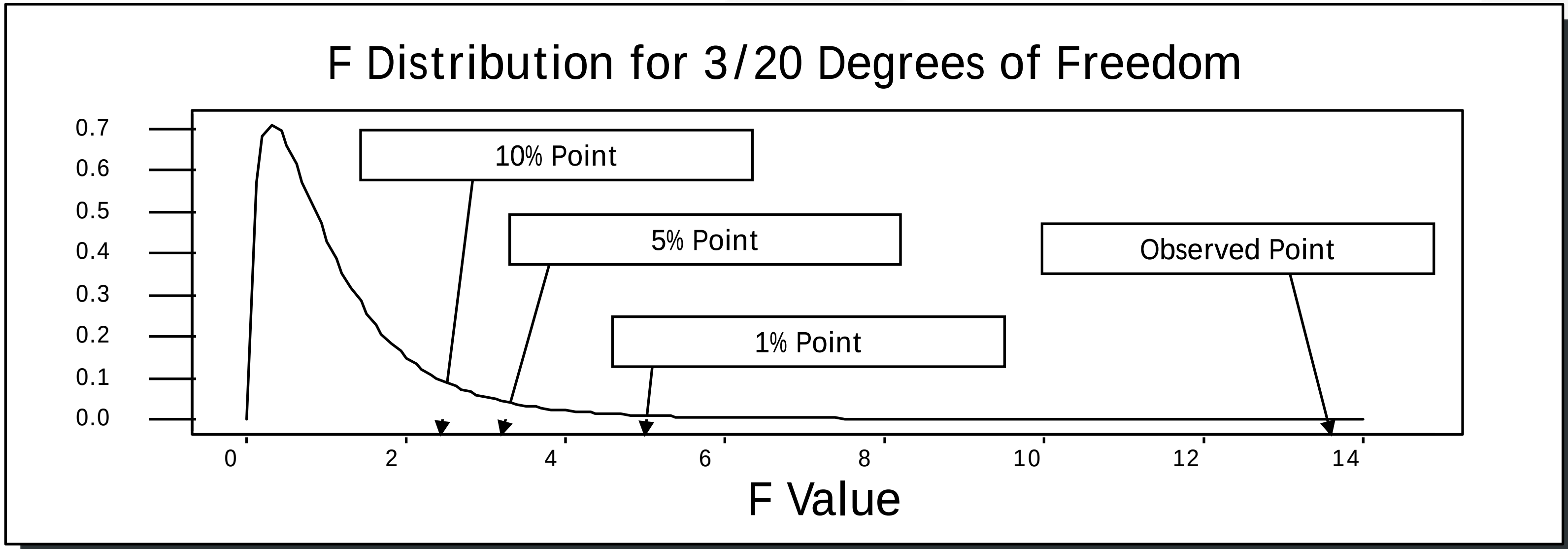
The F-Distribution

The picture below shows the shape of the F distribution for 3/20 degrees of freedom (there are multiple graphs based on Degrees of Freedom).

Our critical statistic (where 5% alpha sits) is at approximately 3.1.

The F ratio has been calculated as 13.79.

What does this say about the probability of the shopping centres salaries?



ANOVA Requirements (1 of 2)

ANOVA is a type of modeling and as such there are some inaccuracies that can creep into the analyses and yet the model can look good on the surface.

To discover problems, there are some requirements that have to be imposed on the data and verified. Some will be verified via the raw data and some will be verified via the residuals.

A residual is the difference between the value of an actual data point and value predicted by the model (which in the One-way ANOVA test is also the value of its group mean).

ANOVA Requirements (2 of 2)

- Raw Data Requirements
  - Normality. This is the least worrisome requirement, but if there are gross non-normalities in any of the levels, we should use non-parametric testing instead of the one-way ANOVA.
  - Equal Variances. The variances of each level of the factor have to be (statistically) equal. If not, we cannot run a one-way ANOVA. We will use non-parametric testing instead.
- Residual Requirements
  - Normality. Residuals have to be normally distributed.
  - Mean of Zero. The residuals have to average zero. This is not a problem given modern software and it is easy to check.
  - Independent. The value of one data point does not influence the value of the next one. Plotting the residuals in time order would give us an indication of problems here.
  - Equal Variance across the levels of the factor.

Steps to Run an ANOVA

Step 0: Begin the six step hypothesis testing procedure. Once you determine that the ANOVA is the test you wish to run, continue with Step A.

Step A: Take a practical look at the data.

Step B: Take a graphical look at the data.

Step C: Verify the raw data requirements.

Step D: Run the ANOVA test in Minitab.

Step E: Interpret the results of the test.

Step F: Verify the residual requirements.

Step G: Apply the answer to the business question!

Exercise 1

Step 0: The business question is whether there is a true difference in the salaries based on which centre an assistant is working at.

Step 1: H_0: $m_1 = m_2 = m_3 = m_4$

Step 2: H_a: At least one m is different.

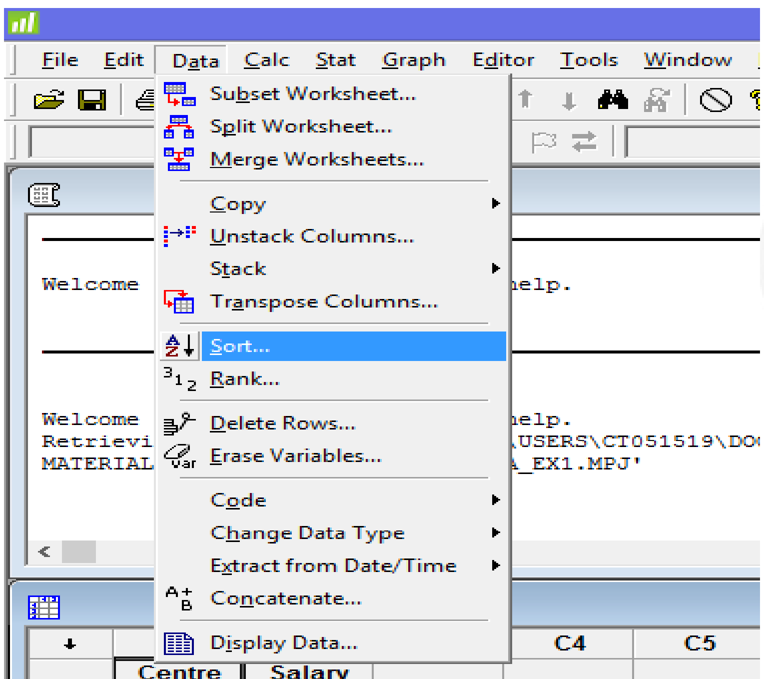
Step 3: a = 0.05

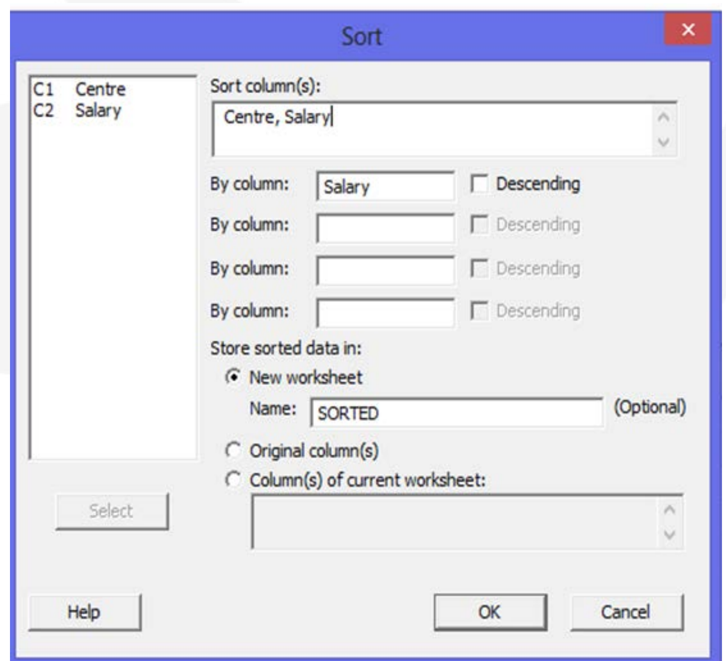
Step 4: Data: Open ANOVA_Ex1.MPJ

Step 5: ANOVA Analysis including a practical and graphical look and including the verification of the requirements.

Step A, Practical Look at the Data

- There are a lot of ways to take a practical look at the data, but one of the most common ones is to sort the data and see if anything jumps out.
- Any clearly improbable numbers may indicate a measurement system or data entry problem.
- Look for any patterns that would give an indication of significance.





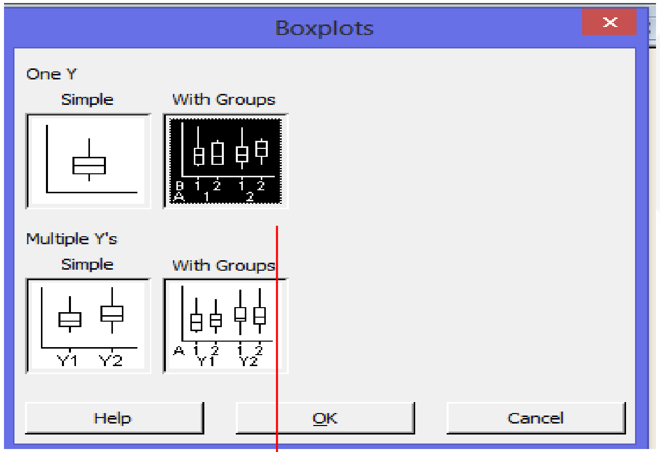
Step A, Practical Look at the Data

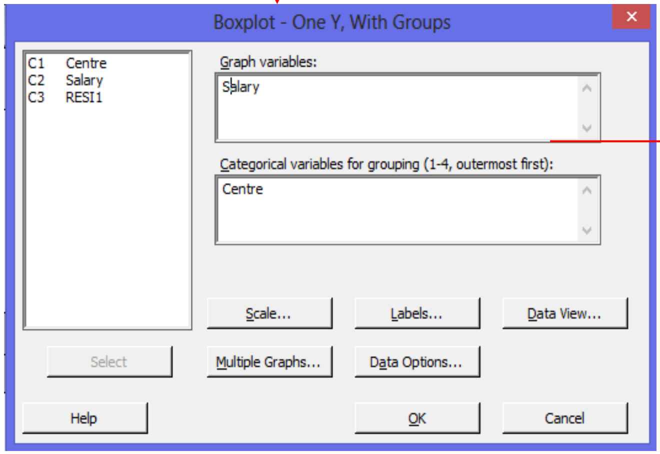
	C1	C2	C3	C4
	Centre	Salary		
1	4	15.0		
2	1	16.3		
3	4	16.3		
4	1	16.6		
5	4	16.8		
6	4	17.2		
7	1	17.7		
8	4	17.7		
9	1	18.1		
10	2	18.1		
11	4	18.1		
12	4	18.1		
13	2	18.6		
14	4	18.6		
15	2	19.0		
16	2	19.5		
17	3	19.5		
18	2	19.9		
19	3	19.9		
20	3	20.4		
21	3	20.4		
22	3	20.4		
23	2	21.7		
24	3	21.7		
25				

- **What do you see?**
- **With a small sample size, how confident are you that it's a true difference?**

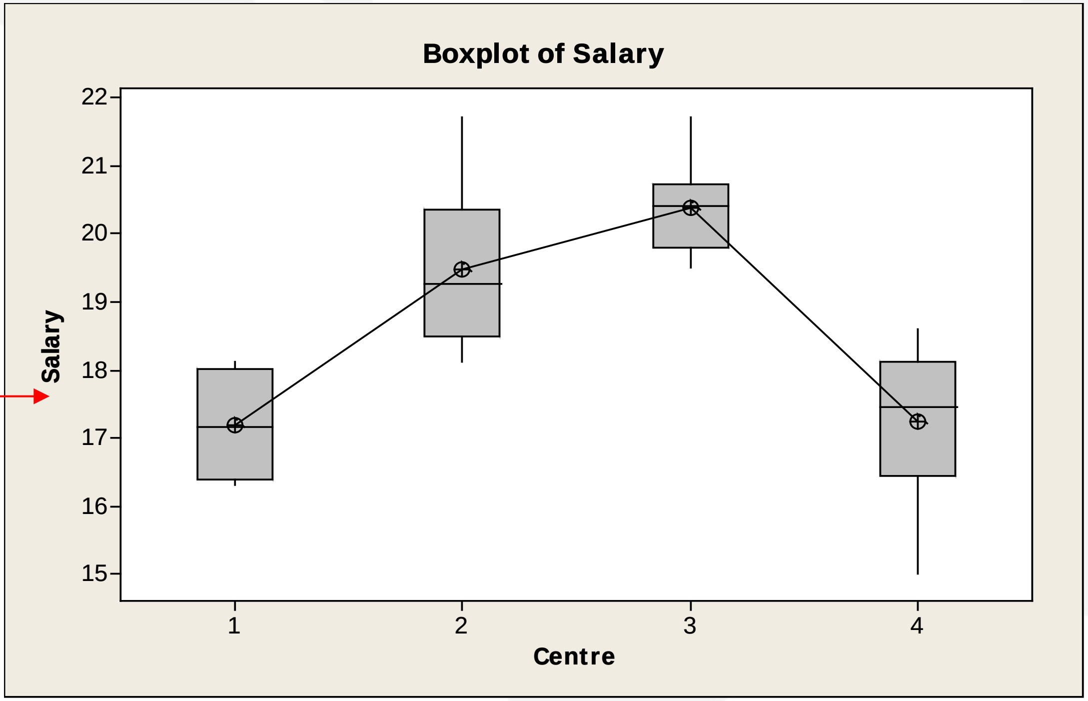
Step B, Graphical Look at the Data

Graph > Boxplot ...





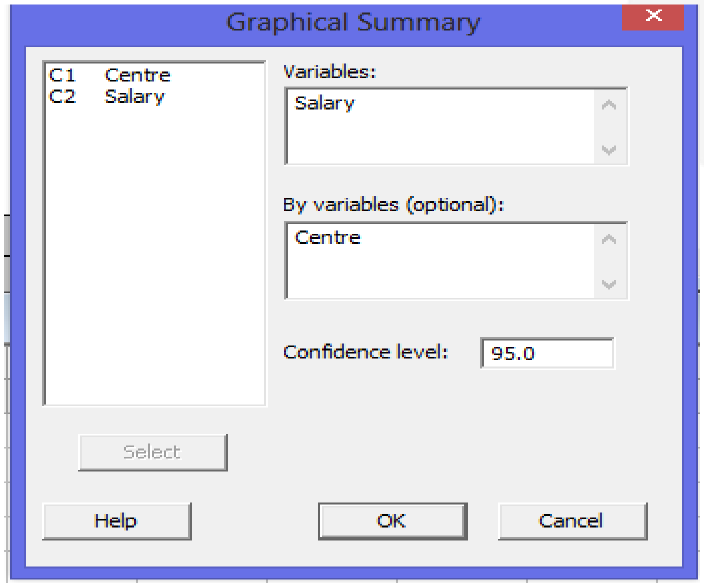
What does the box plot suggest?

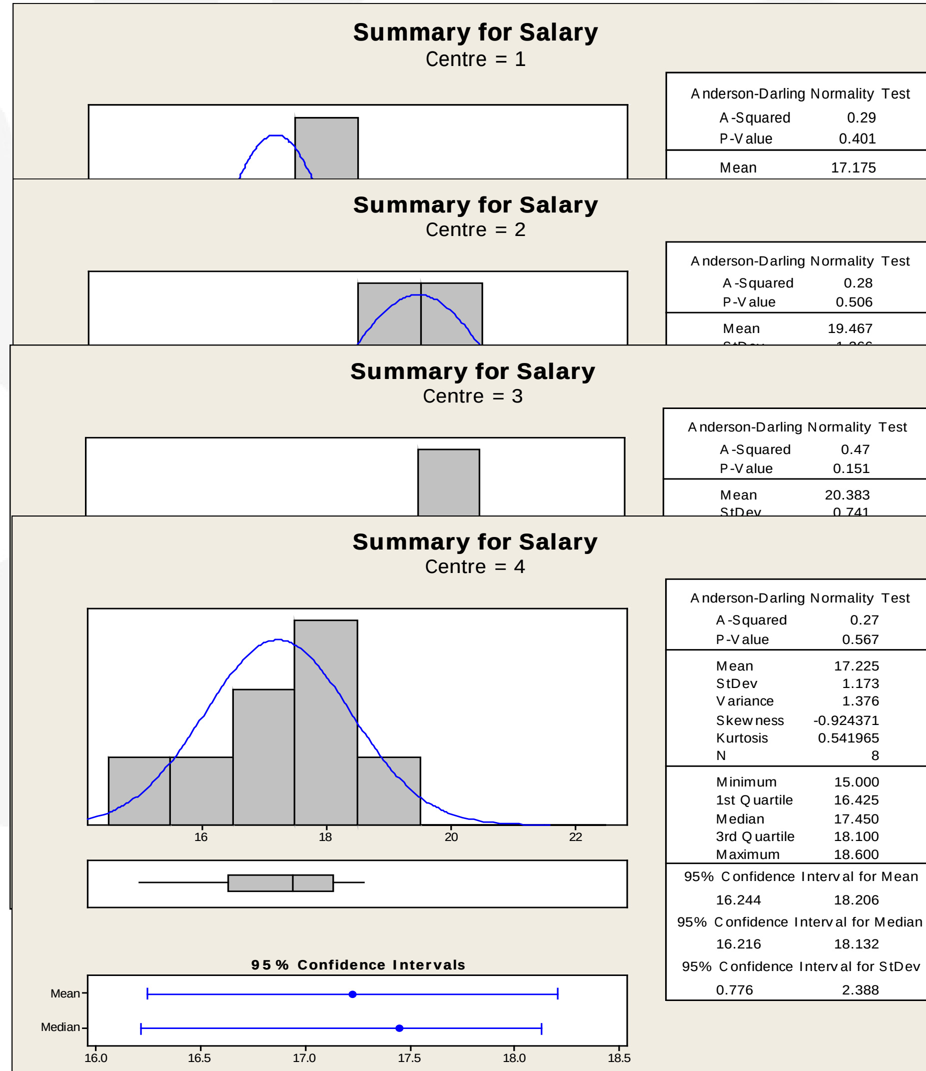


Step C, Test the Raw Data - Normality

Normality Test:

<u>Stat > Basic Statistics > Graphical Summary</u>

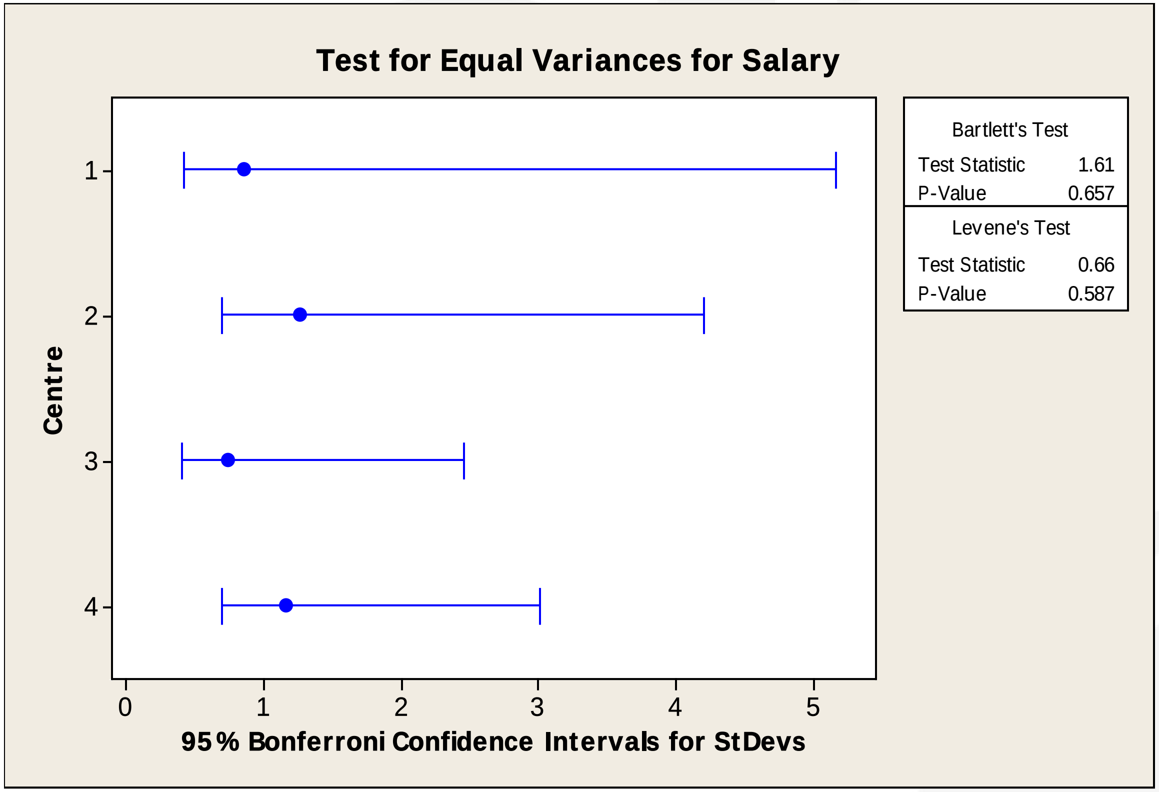




- **P-values for the Anderson Darling Test for the four centres are 0.401, 0.506, 0.151, and 0.567.**
- **What do we conclude?**

Step C, Raw Data – Equal Variances Test (at each factor level)

Stat> ANOVA > Test for Equal Variances



- **Use Levene's test if even one level is non-normal; otherwise use the Bartlett Test.**
- **What do we conclude?**

Steps D/E, ANOVA Test

<u>Stat > ANOVA > One-way</u>

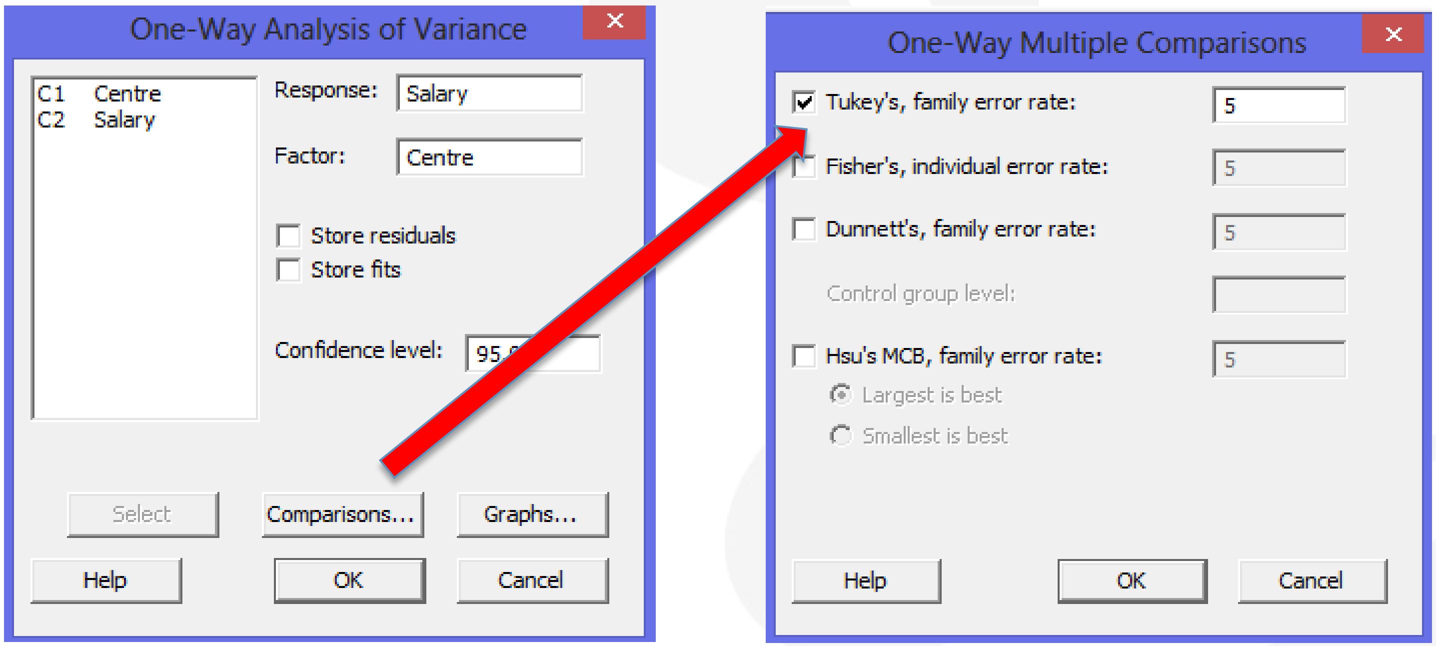
One-way ANOVA: Salary versus Centre

Source	DF	SS	MS	F	P
Centre	3	46.81	15.60	13.79	0.000
Error	20	22.62	1.13		
Total	23	69.43			

- **Recognise this?**
- **What do we conclude?**
- **What is Step 6 in hypothesis testing?**
- **What question does this not answer?**

Steps D/E, Pairwise Comparisons

Stat > ANOVA > One-way > Comparisons >Tukey's



Steps D/E, Pairwise Comparisons

```
Grouping Information Using Tukey Method

Centre  N    Mean  Grouping
3       6  20.383  A
2       6  19.467  A
4       8  17.225    B
1       4  17.175    B

Means that do not share a letter are significantly different.


Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of Centre

Individual confidence level = 98.89%


Centre = 1 subtracted from:

Centre   Lower  Centre  Upper  ---------+---------+---------+---------+
2        0.369   2.292  4.214                    (-------*-------)
3        1.286   3.208  5.131                        (-------*-------)
4       -1.774   0.050  1.874            (------*------)
                               ---------+---------+---------+---------+
                                     -2.5       0.0       2.5       5.0


Centre = 2 subtracted from:

Centre   Lower  Centre   Upper  ---------+---------+---------+---------+
3       -0.803   0.917   2.636                    (------*------)
4       -3.850  -2.242  -0.633        (-----*-----)
                                ---------+---------+---------+---------+
                                      -2.5       0.0       2.5       5.0


Centre = 3 subtracted from:

Centre   Lower  Centre   Upper  ---------+---------+---------+---------+
4       -4.767  -3.158  -1.550  (-----*------)
                                ---------+---------+---------+---------+
                                      -2.5       0.0       2.5       5.0
```

Two ways to make conclusions:

1. **A's are not significantly different from other A's, but are different from B's.**

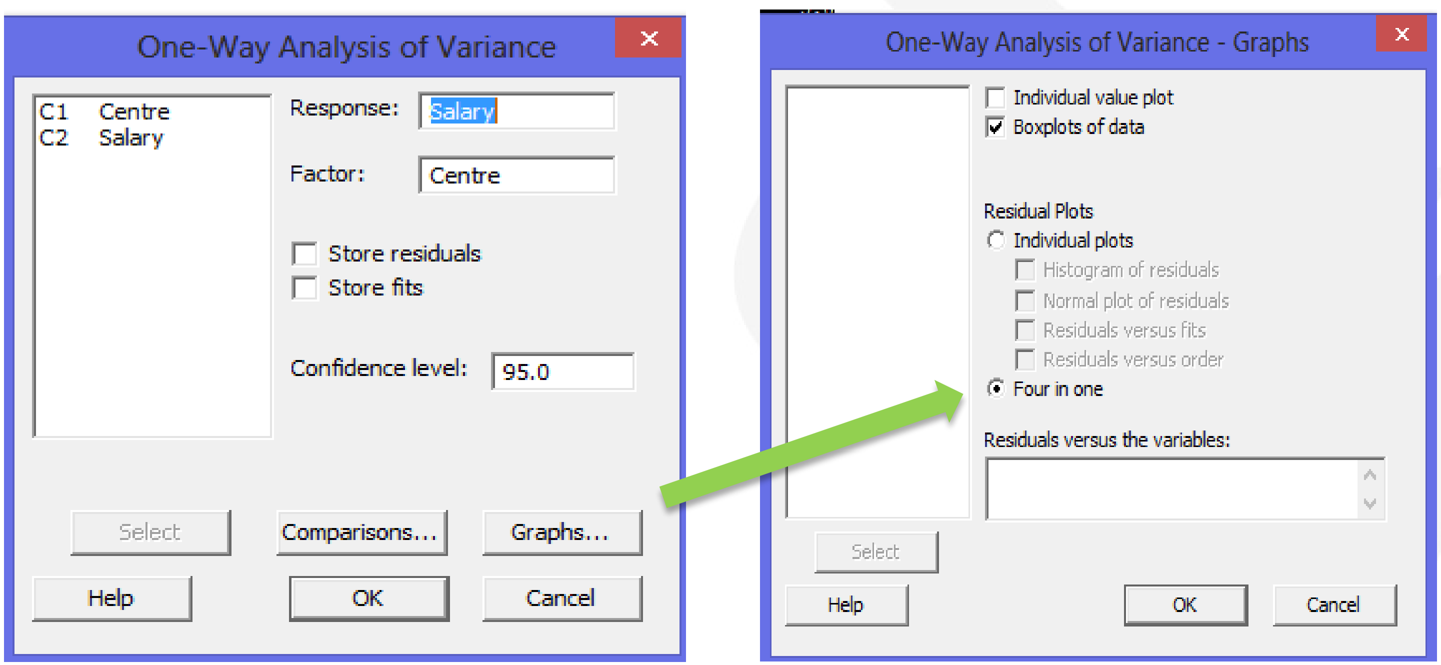
2. **Confidence interval for each pair. Is a zero difference within the confidence interval?**

Step F, Validate Residual Requirements

- After running the ANOVA we have to test the residual requirements: Normally Distributed, Mean of Zero, Independent, Equal Variance across levels of the factor.

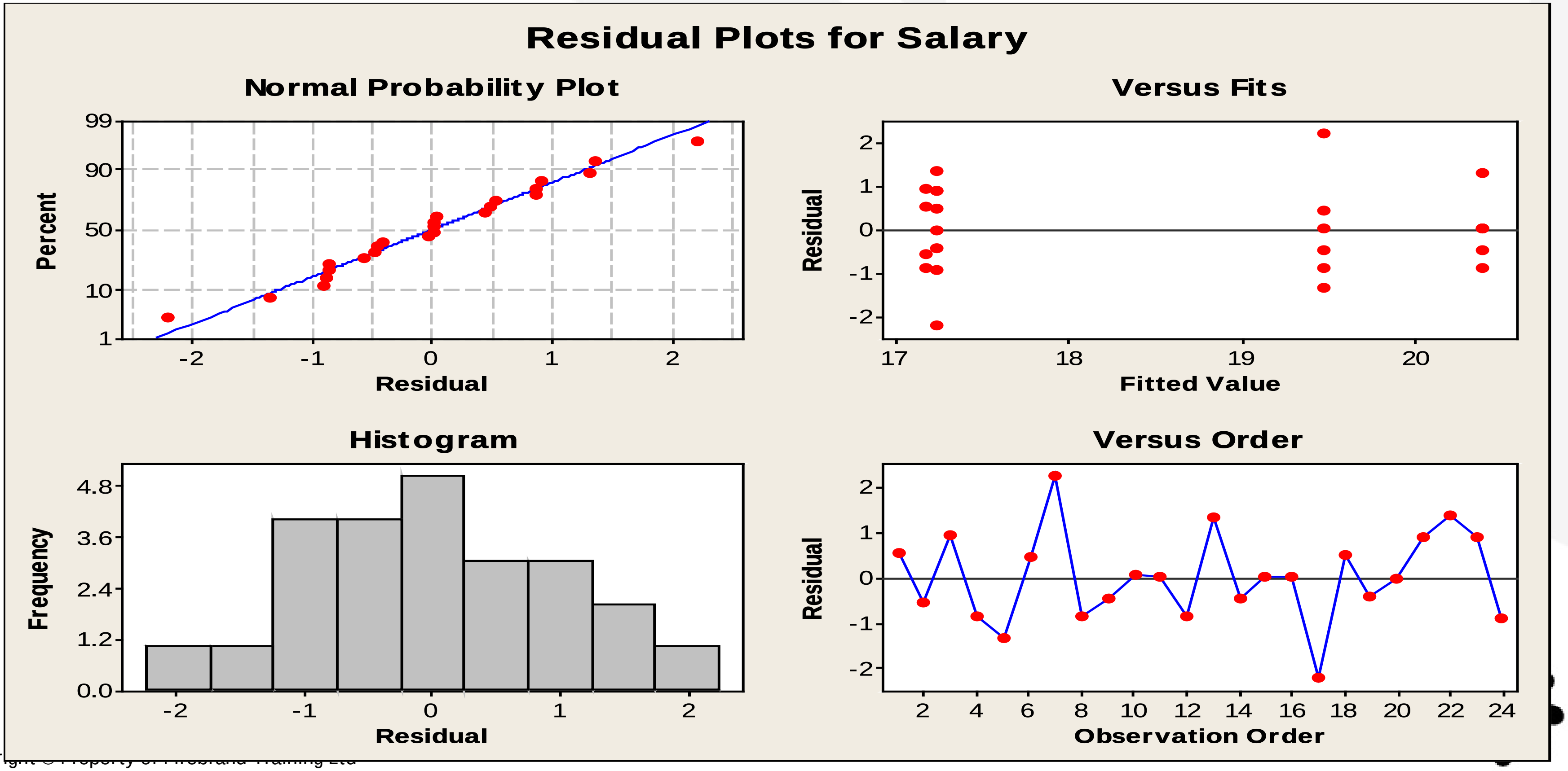
Stat > ANOVA > One-way

Click on graphs and select Four in one



Step F, Validate Residual Requirements

- Four in one
  - Normally distributed
  - Equally distributed around zero
  - Constant variance
  - Independent



Step F, Answer the business question

Step 0: The business question is whether there is a true difference in the salaries based on which centre an assistant is working at.

Based on our analysis we can conclude that salaries at centres 2 & 3 are different from 1 & 4.

Exercise 2

One-way ANOVA

Objective: To illustrate the use of ANOVA in solving business problems

- A marketing team wants to evaluate new VRU sequences for activating credit cards. A test was done to compare two new VRU sequences for entering information. If we can reduce the time even by a few seconds, we can save a LOT of money.
- We set up a test to test our current sequence against two potential improvements. We gather the data. After you analyse the data, what is your recommendation?

20 minutes

The data is in:
ANOVA_Ex_2.MTP

Non-Normal Data

- **The ANOVA test is robust to non-normal data, but if one or more of the levels are distributed very differently from the normal distribution the ANOVA test could give us misleading results. In that case we should use non-parametric tests to test for differences.**

- **Non-parametric in ANOVA is the same as discussed in Hypothesis Testing; we don’t assume a shape or parameters about the population distributions and for testing the central tendency we hypothesise about the medians.**

- **Additionally with non-parametric tests, we don’t have any requirements on the residuals.**

Non-parametric Exercise 3

- **We have ten investment sales offices in each of four similarly sized cities. We are piloting a new sales plan and have selected these four cities as the target. We want to know if the plan works equally well in all four cities.**
- **At the end of the first day of the pilot, we get the sales results from all of the offices. What do the early results suggest?**
- **The data are in ANOVA_Ex3.MTP and are recorded in $M.**

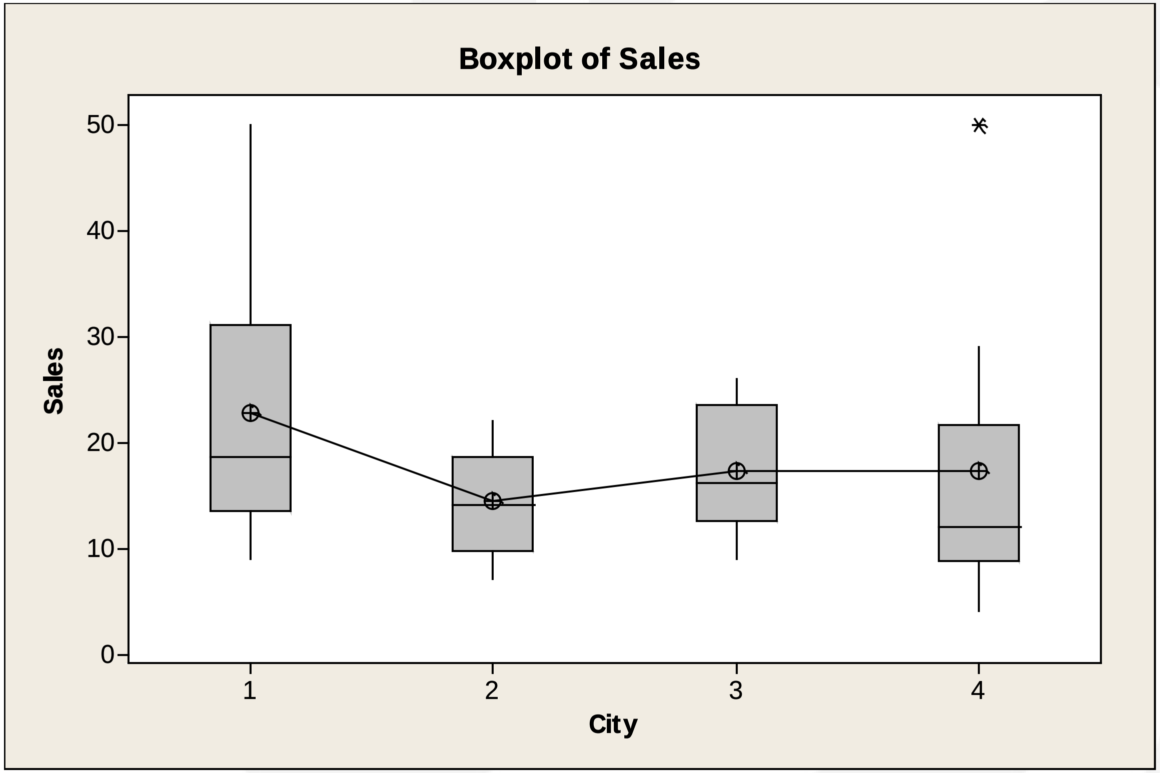
Steps to Run a Nonparametric Test

- **The steps for a non-parametric test start off just like the one-way ANOVA, but in Step C, if we see gross non-normalities, we will switch to non-parametric testing.**
- **Step 0: Begin the six step hypothesis testing procedure. Once you determine that the ANOVA is the test you wish to run, continue with Step A.**
- **Step A: Take a practical look at the data.**
- **Step B: Take a graphical look at the data.**
- **Step C: Verify the raw data requirements. If the data from any level show gross non-normalities continue with Step D.**
- **Step D: Run the nonparametric test in Minitab.**
- **Step E: Interpret the results of the test.**
- **Step F: Apply the answer to the business question!**

Steps A/B, Practical and Graphical

File Edit Data Calc Stat G

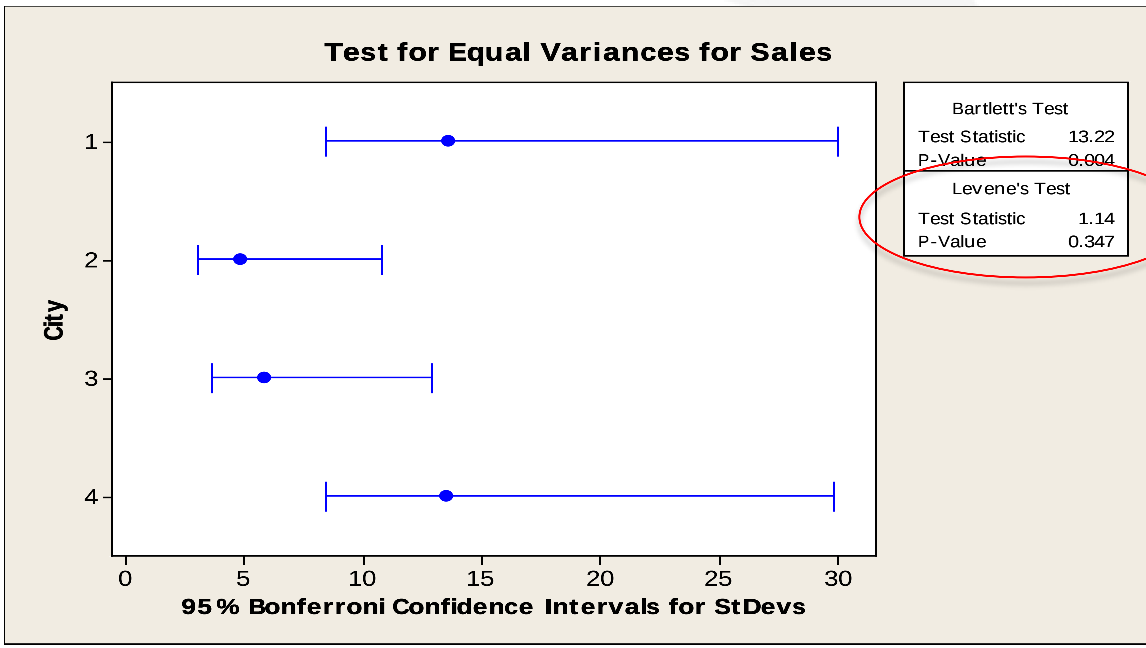
↓	C1	C2
	City	Sales
1	4	4
2	2	7
3	4	8
4	1	9
5	2	9
6	3	9
7	4	9
8	2	10
9	3	11
10	4	11
11	4	11
12	1	12
13	2	13
14	3	13
15	4	13



What do you learn or suspect from the practical and graphical look at the data?

Step C, Test the Raw Data Requirements

- **Normality Test**
  - **P-values for the Anderson Darling Test for normality the four cities are 0.032, 0.912, 0.607, and 0.024**
  - **What do we conclude?**
  - **Might we have predicted this from the box plots?**
- **Test for Equal Variance Test**



Levene's Test

P-value = 0.347

What do we conclude about the equal variances test?
What is our next step?

Three Non-parametric Tests

<u>Mann-Whitney</u> - used for comparing two groups

<u>1-sample sign</u> - used for comparing paired data

<u>Kruskal-Wallis</u> - used to look for differences between groups

All three of these test the hypothesis that the population <u>medians</u> are not equal:

Step 1: H_0: $\eta_1 = \eta_2 = \eta_3 = \eta_4$
Step 2: H_a: At least one η is different.

Which test will we want to use to evaluate day one of our pilot?

Steps D/E/F, Wilcoxon Test

Stat>Non-parametrics>Kruskal-Wallis

Kruskal-Wallis Test: Sales versus City

```
Kruskal-Wallis Test on Sales

City         N   Median   Ave Rank       Z
1           10    18.50       25.1    1.44
2           10    14.00       17.4   -0.97
3           10    16.00       22.2    0.53
4           10    12.00       17.3   -1.00
Overall     40                20.5

H = 3.21  DF = 3  P = 0.360
```

- **What decision do we make?**
- **How do we answer the business question?**

2-way ANOVA

- **2-way ANOVA is used when we have multiple discrete factors each of which have more than one level and we want to know how different levels of two or more factors AND their interaction have on the Y.**
- **2-way ANOVA uses the same approach as 1-way, applying the variation within and variation between to determine which Xs are significant.**
- **2-way ANOVA will also provide us with a predictive formula for the Y - think Y = f(X).**
- **Since this again is a predictive model, we will have similar requirements on the data as with the One-way ANOVA.**

2-WAY ANOVA Requirements

- **<u>Raw Data Requirements</u>—There are no Raw data requirements for 2-way ANOVA. Due to the complexity of having more than one factor, the model can only be evaluated using Residual Analysis.**
- **<u>Residual Requirements</u>—These are the same as for the One-way ANOVA.**
  - **<u>Normality</u>. Residuals have to be normally distributed.**
  - **<u>Mean of Zero</u>. The average residual is zero. This is not a problem given modern software and it is easy to check.**
  - **<u>Independent</u>. The value of one data point does not influence the value of the next one. Plotting the residuals in time order would give us an indication of problems here.**
  - **<u>Equal Variance</u> of the residuals across the full range of predicted Y values.**

2-way ANOVA —Exercise 4

- **Let's revisit the problem of salaries at Shopping Centres with four different Centers.**
- **Do you think that "Shopping Centre" is the only X?**
- **Although there are many Xs, for this exercise we are only going to expand our list by one, "Recruiter."**
- **The data are in ANOVA_Ex4.MTP**

Steps to Run a 2-way ANOVA

Step 0: Make sure you understand the business question. Verify that your Y is continuous data and your Xs are discrete.

Step A: Take a practical look at the data.

Step B: Take a graphical look at the data.

Step C: Specify the model and run the test in Minitab.

Step D: Interpret the results of the test.

Step E: Reduce the model (if needed)

Step F: Verify the residual requirements.

Step G: Apply the answer to the business question!

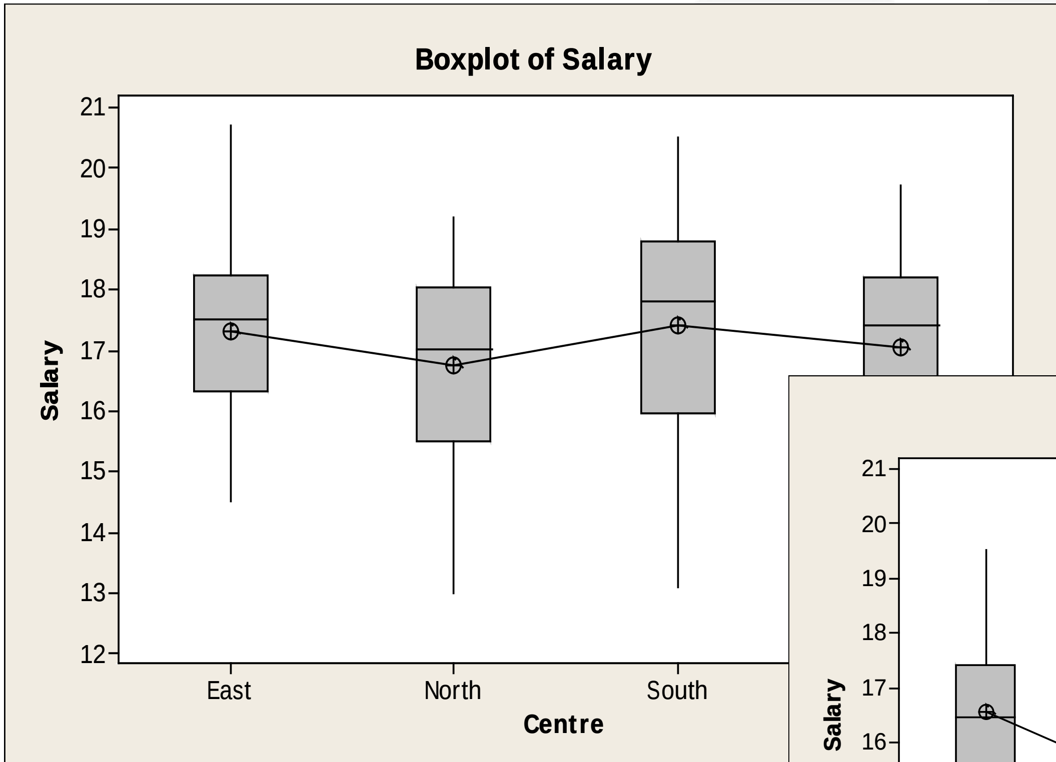
Step A, Practical Look at the Data

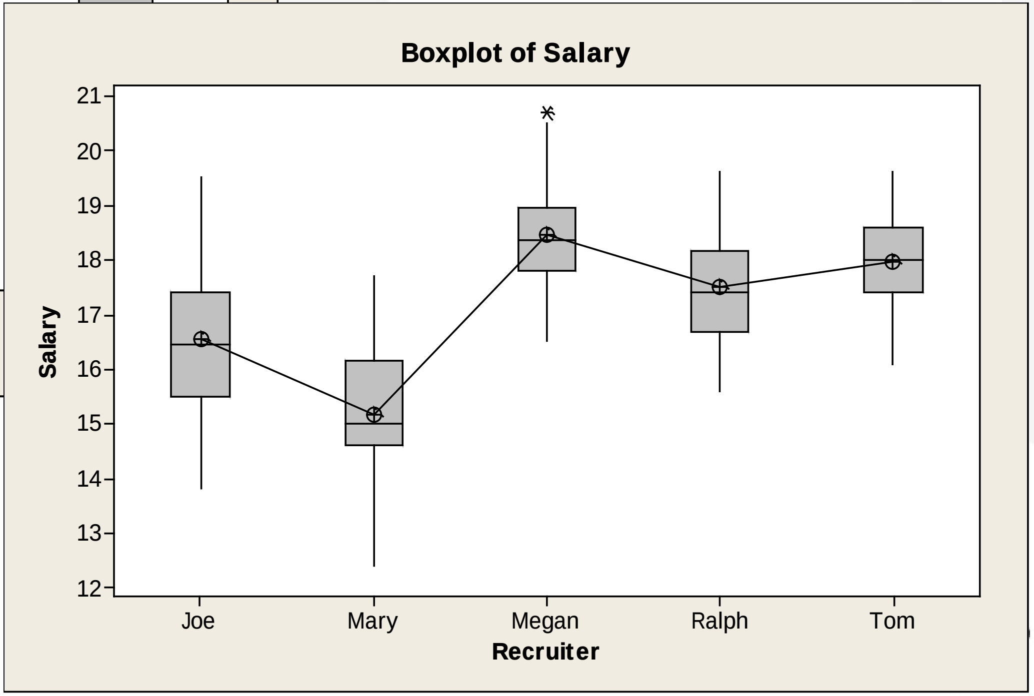
↓	C1-T	C2-T	C3
	Centre	Recruiter	Salary
1	West	Mary	12.4
2	West	Mary	12.7
3	North	Mary	13.0
4	South	Mary	13.1
5	North	Joe	13.8
6	North	Mary	13.8
7	South	Mary	13.9
8	West	Mary	14.1
9	East	Mary	14.5
10	North	Joe	14.7
11	North	Mary	14.7
12	South	Mary	14.7
13	East	Mary	14.7
14	West	Mary	14.7
15	South	Joe	14.8
16	North	Mary	14.8
17	West	Mary	14.8
18	East	Mary	14.9
19	East	Mary	14.9
20	South	Mary	15.0

What do you see?

Step B, Graphical Look at the Data

Stat > ANOVA > One-way



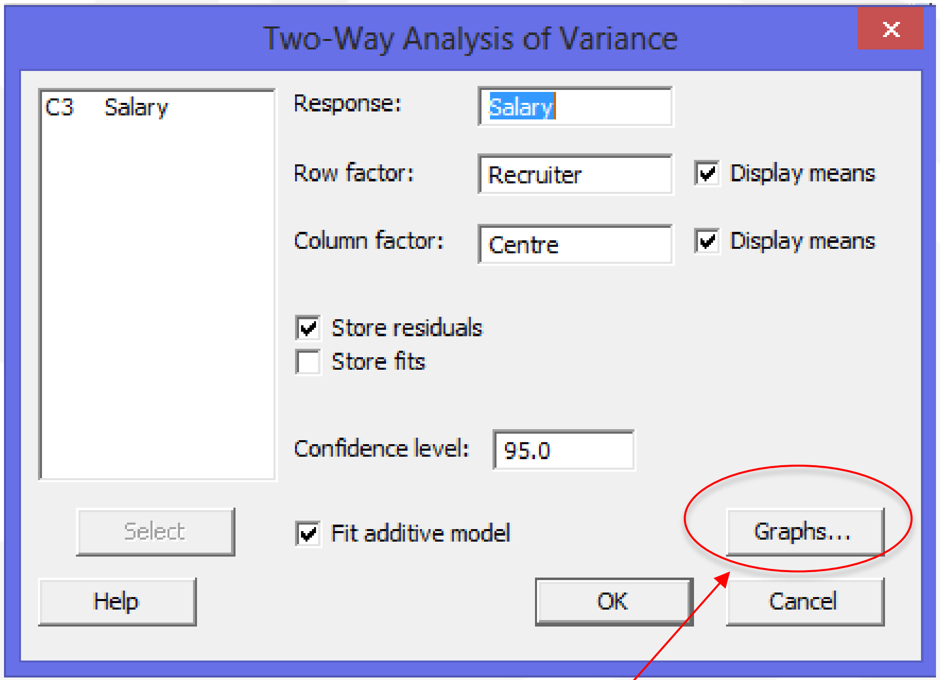


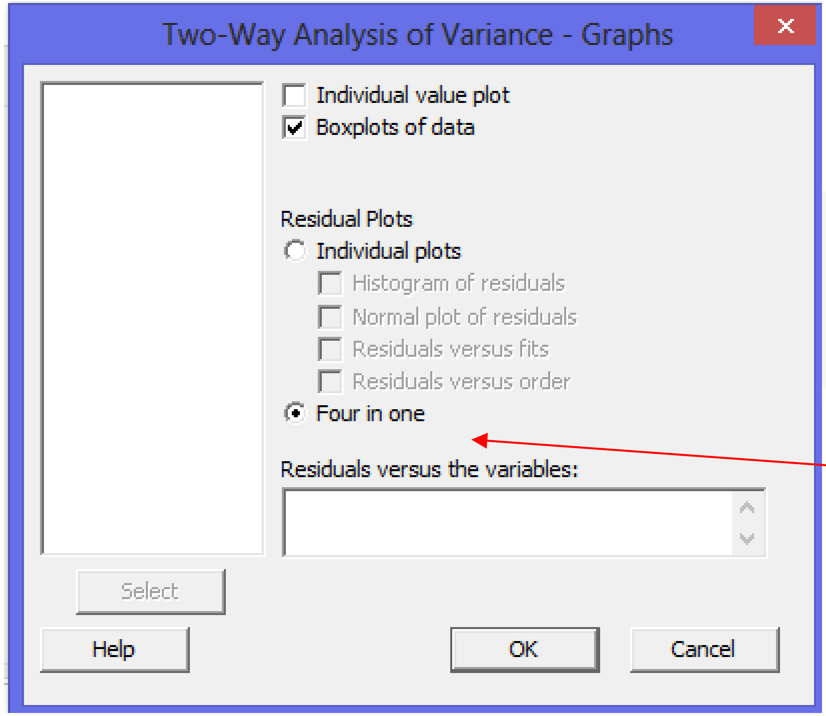
- **What do the box plots suggest?**
- **What do the comparison circles indicate?**

Step C, Specify the Model

<u>Stat > ANOVA > Two-Way...</u>

Select ***Display Means and Store residuals***





Select ***Four in one***

ANOVA Table for the Model as a whole

H_0: The model is not significant (none of the Xs are significant).

H_a: The model is significant (at least one of the Xs is significant).

(In this analysis the means we are hypothesizing about are the means of the factors, not the levels of the factors.)

Two-way ANOVA: Salary versus Recruiter, Centre

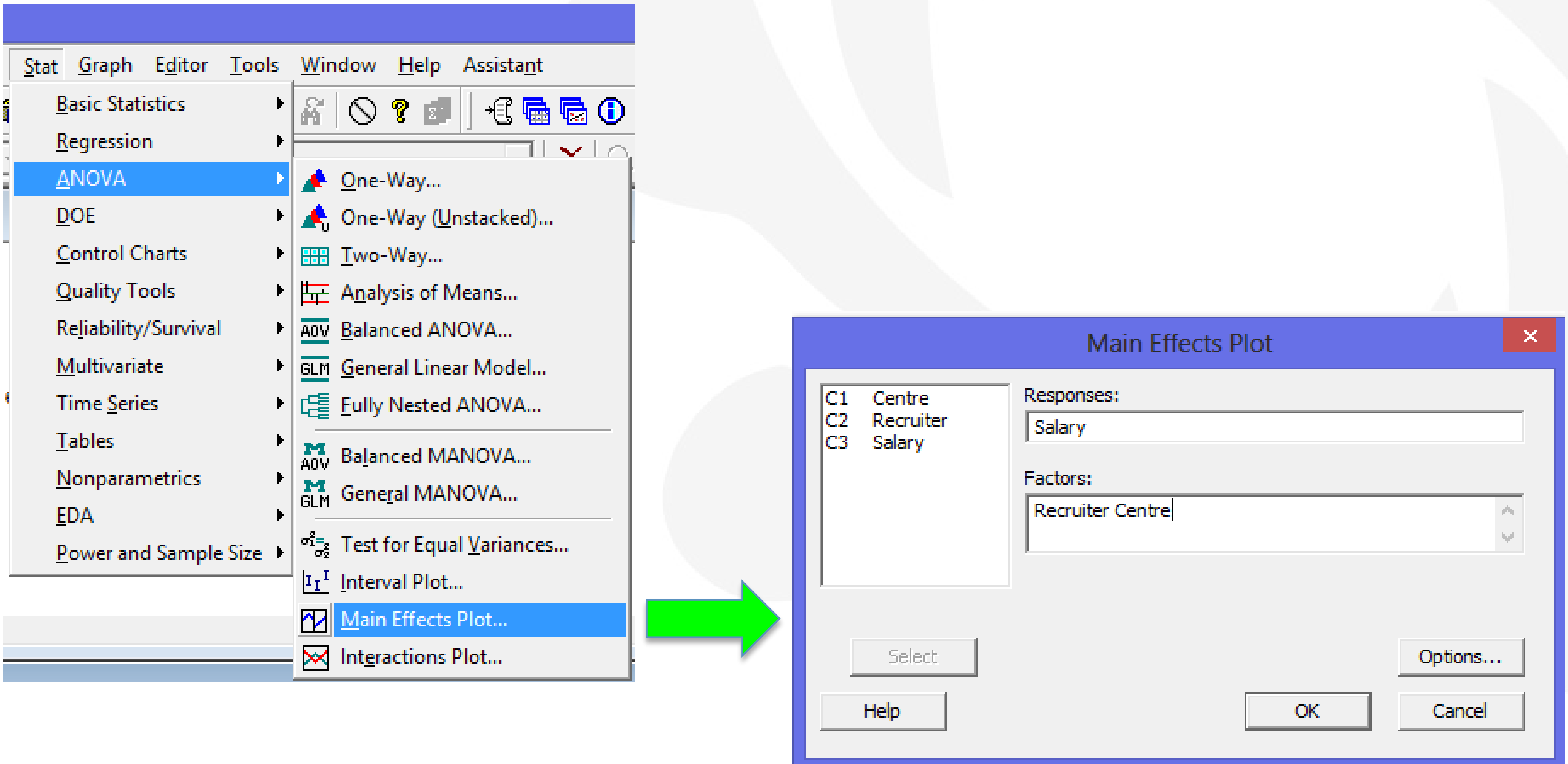
Source	DF	SS	MS	F	P
Recruiter	4	195.072	48.7681	40.75	0.000
Centre	3	8.724	2.9081	2.43	0.069
Interaction	12	11.685	0.9737	0.81	0.636
Error	120	143.609	1.1967		
Total	139	359.090			

S = 1.094   R-Sq = 60.01%   R-Sq(adj) = 53.68%

What decision do we make?

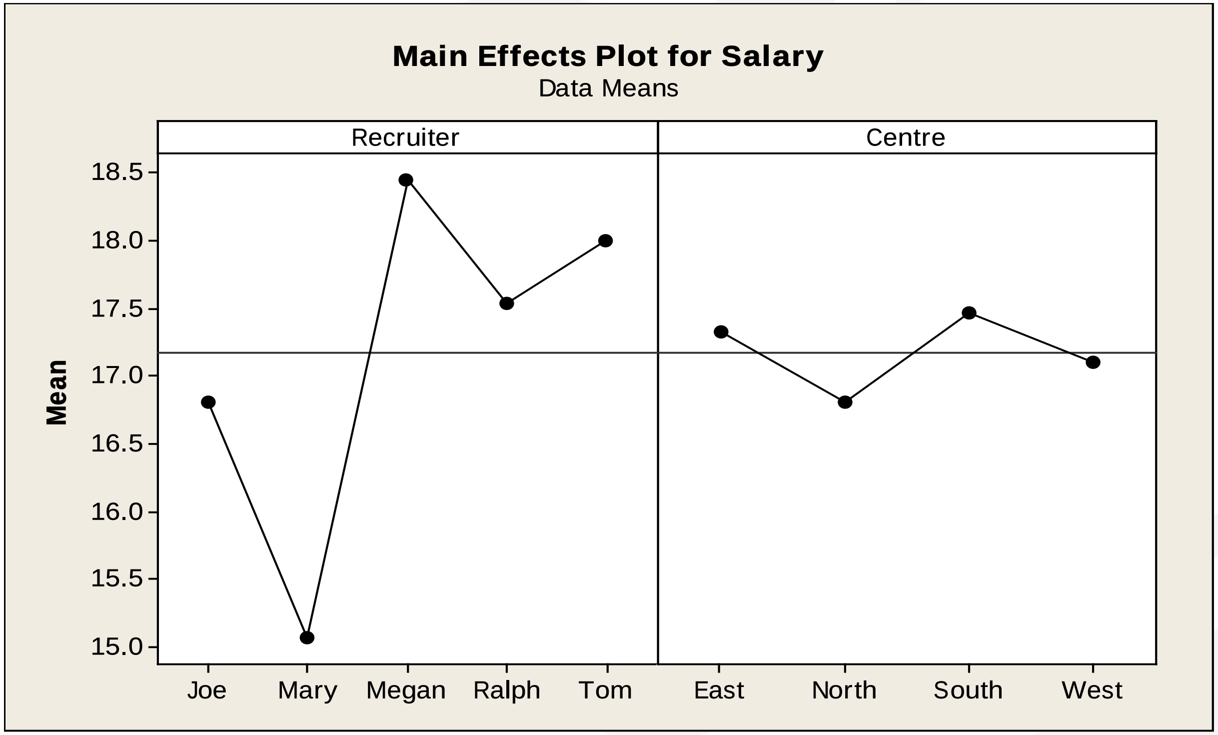
Step D, Results – Main effects Plot

Lets look at the main effects plot



Step D, Main Effects Plots

Model Specification Plots



Graphically, which factor has the largest impact?

Step F, Residual Analysis

- **After running the two-way ANOVA we have to test the residual requirements the same way as with the One-way ANOVA.**
- **By selecting the tick box as we created our model, a new column has been created on our worksheet**

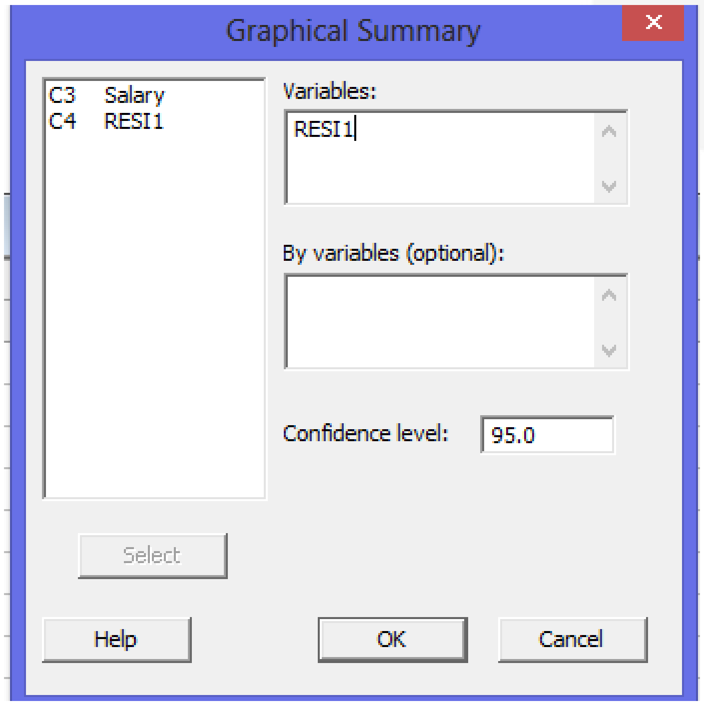
↓	C1-T	C2-T	C3	C4
	Centre	Recruiter	Salary	RESI1
1	East	Ralph	15.9	-1.78071
2	East	Ralph	19.6	1.91929
3	East	Ralph	17.2	-0.48071
4	East	Ralph	17.6	-0.08071
5	East	Ralph	17.5	-0.18071
6	East	Ralph	17.2	-0.48071
7	East	Ralph	16.6	-1.08071
8	East	Joe	17.7	0.74429
9	East	Megan	18.9	0.30500
10	East	Megan	17.9	-0.69500
11	East	Megan	18.7	0.10500
12	East	Megan	18.3	-0.29500
13	East	Megan	17.9	-0.69500
14	East	Megan	18.2	-0.39500

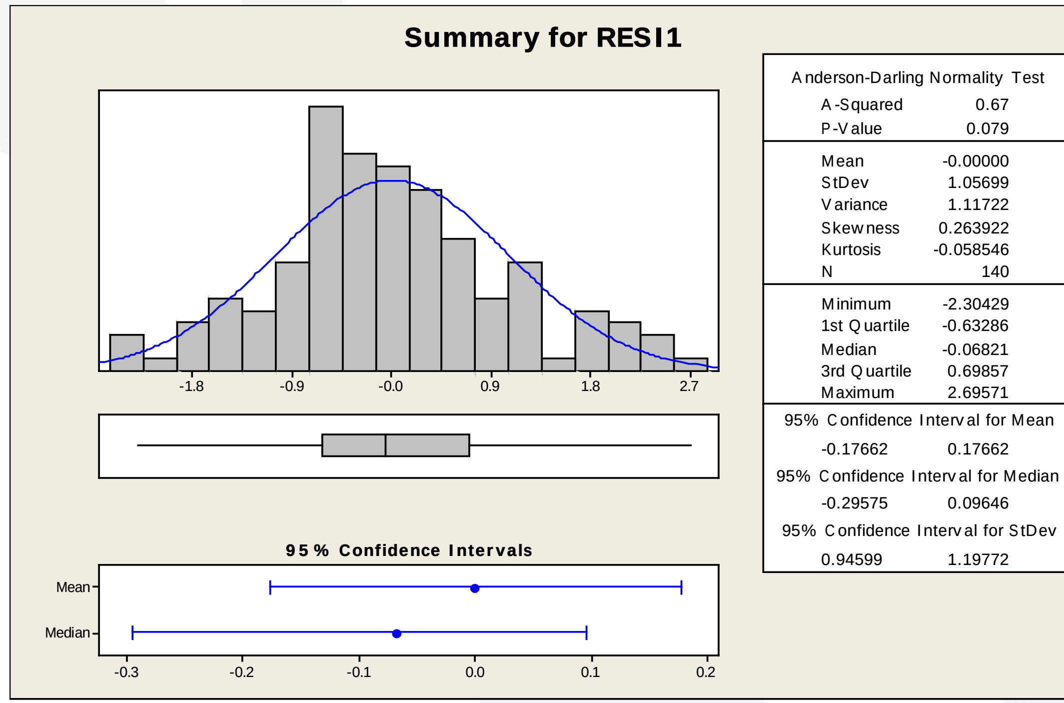
Residuals

Normally Distributed, Mean of Zero

Test for Normality and Mean of Zero

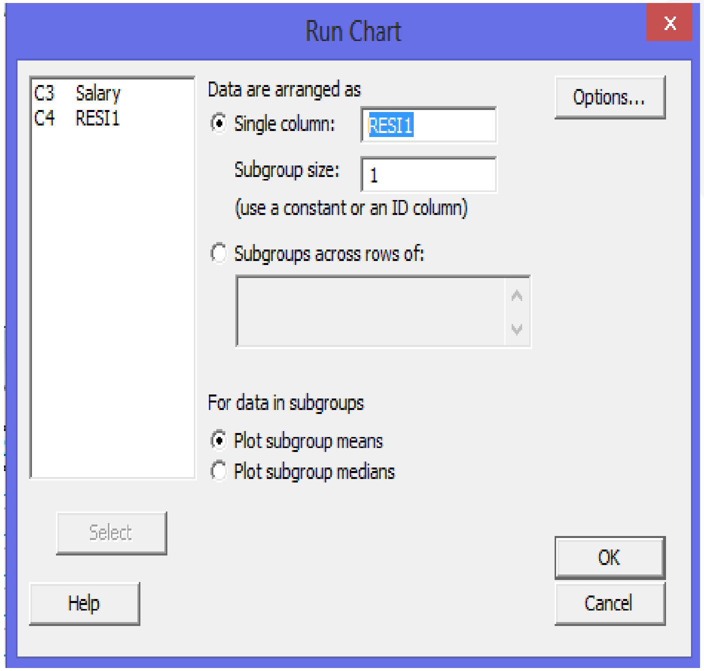
<u>Stat > Basic Stat > Graphical Summary</u>

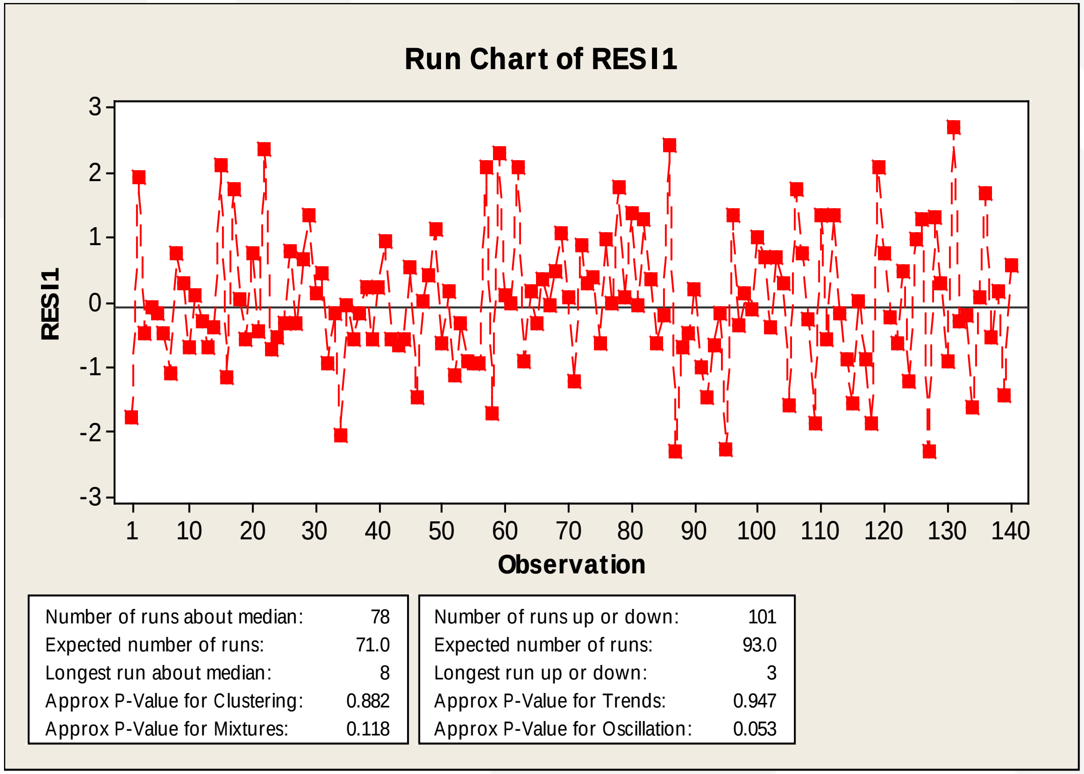




- **P-value for the Anderson-Darling Normality Test is 0.079**
- **Mean is 0.00000**
- **What do we conclude?**

Stat > Quality Tools > Run Chart



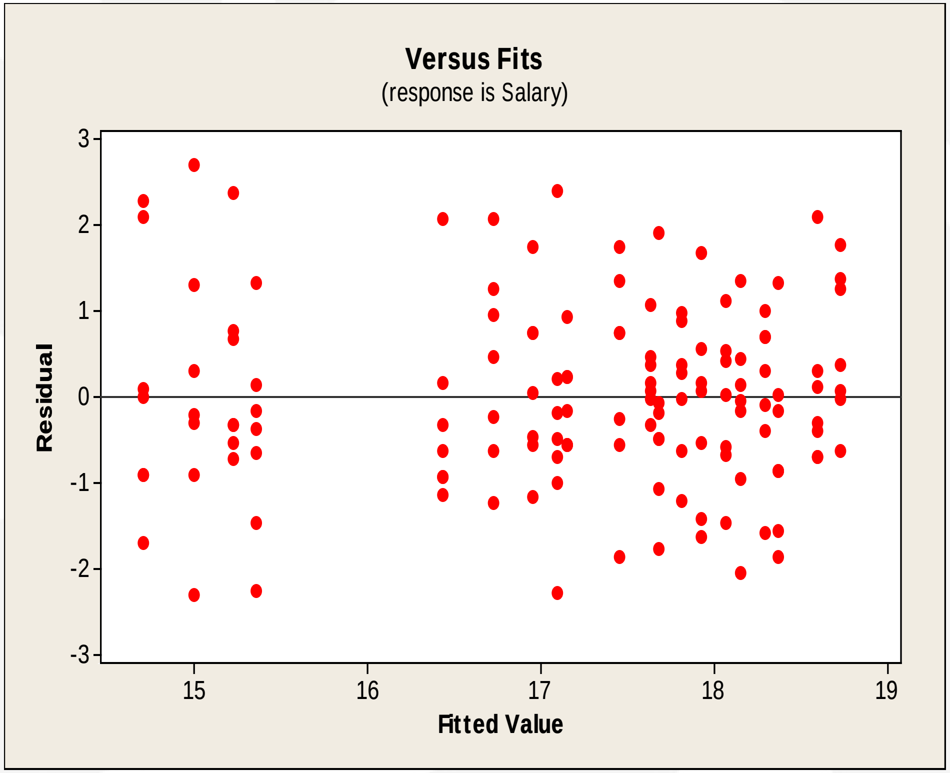


Always select subgroup size of 1 for individual measurements

P-values created to highlight any non-random patterns in the data collected

Equal Variance

- **On this graph we are looking to see that the vertical scatter of the residuals (the Y on this graph) is relatively equal across all the predicted X values and that there are no patterns.**
- **How would you interpret this graph?**



Step G, Answer the business question

- Most importantly…answer the business question.

As a reminder, we wanted to determine whether recruiter and/or shopping centre were significant factors in salary level.

Our answer….

Who recruits the assistant is a major driver in determining the salary level

Shopping centre or the interaction of shopping centre and recruiter are not.

Summary of key learning points

- ANOVA uses the Sums of Squares method to derive signal to noise ratios
- Use of the ANOVA platform requires there to be some hygiene tests to be successfully completed on the data
- Non-parametric tools can be used when the data fails these tests
- Two-way ANOVA can be used to assess multiple discrete factors and their interaction on a continuous Y

Recommended coach support points

- If you have an unbalanced design error
- If the p-value is low and you only have one significant factor.
- When specifiying factors that as crossed or nested.

FIREBRAND

Black Belt

Hypothesis Testing II

Objectives of this module

At the end of this module, you will be able to:

- To review the basics of Hypothesis Testing.
- To identify the principles behind Hypothesis Testing.
- To develop confidence in selecting and running Hypothesis Tests.

Questions Lead

- **Basic question:** **Is there a difference in the output between these groups?**

  - Does one service offering give us fewer unhappy customers than the other?
  - Is one of my sales teams regularly closing more deals than the others?
  - I made a process change. Are we making fewer errors now than we were before?
  - Does the medical plan chosen depend on which region of the country the employee lives in?

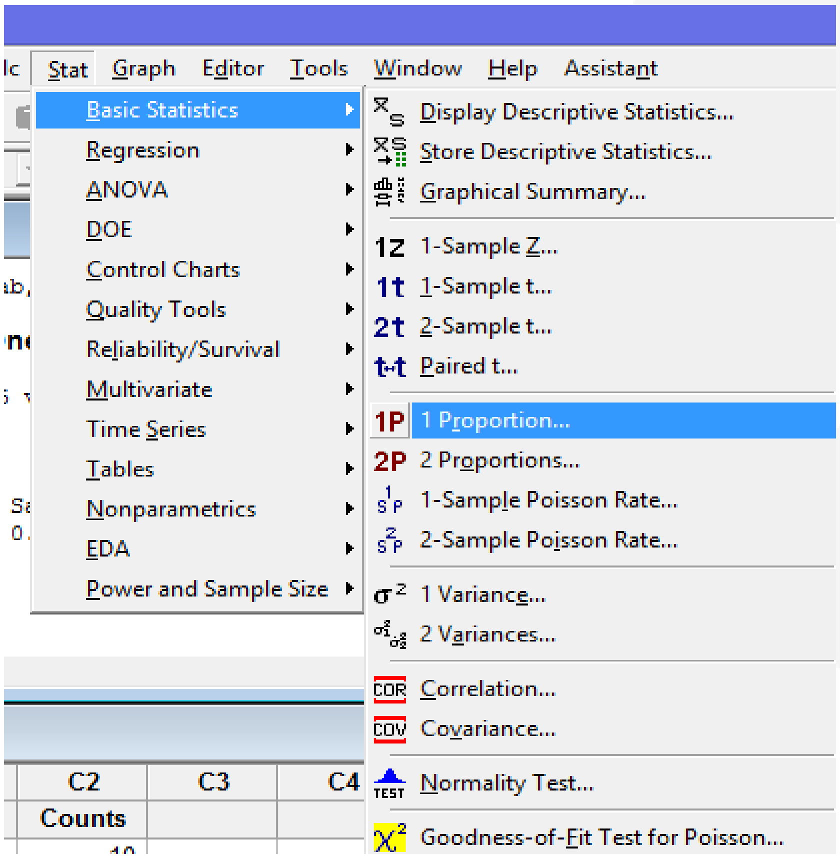
Exercise 1, Proportion vs. Target

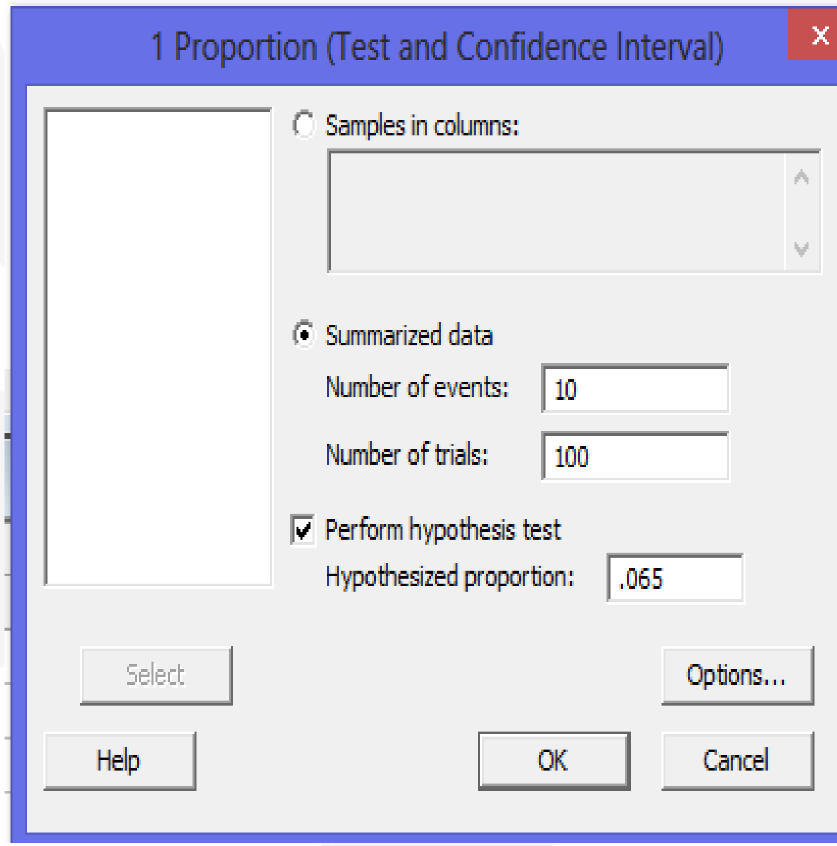
- The Marketing Manager is trying to develop a new e-mail advertisement campaign. He designs an ad and sends out 100 e-mails and finds that 10 responses are sent back. He is all excited because he believes he has developed an ad that beats the national average for responses to an e-mail advertisement which he read was 6.5%.

- Do these data confirm his belief or burst his bubble?

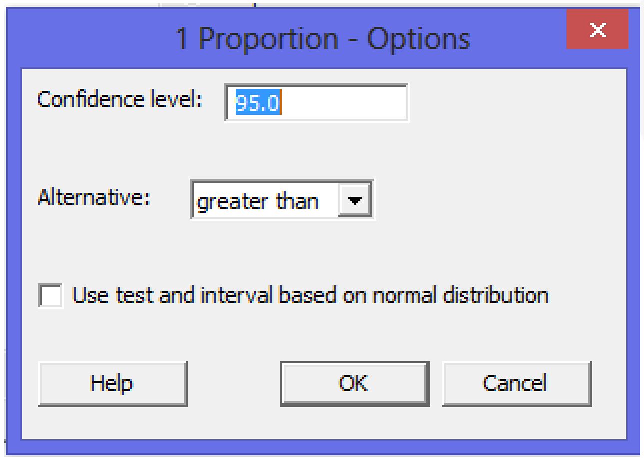
Exercise 1, 1-Proportion test

What’s the hypothesis statement?





Exercise 1, 1-Proportion test



Test and CI for One Proportion

```
Test of p = 0.065 vs p > 0.065

                                 95% Lower    Exact
Sample    X    N  Sample p          Bound  P-Value
1        10  100  0.100000       0.055263    0.115
```

What do we decide?

Exercise 2, Proportions vs. Targets

- An E-Commerce department is creating a new sales tool for On-Line that they will launch next quarter. They have based their profit plan on the proportion of sales from this tool for each of five products as shown below:
  - Product A: 10%
  - Product B: 10%
  - Product C: 20%
  - Product D: 20%
  - Product E: 40%

15 minutes

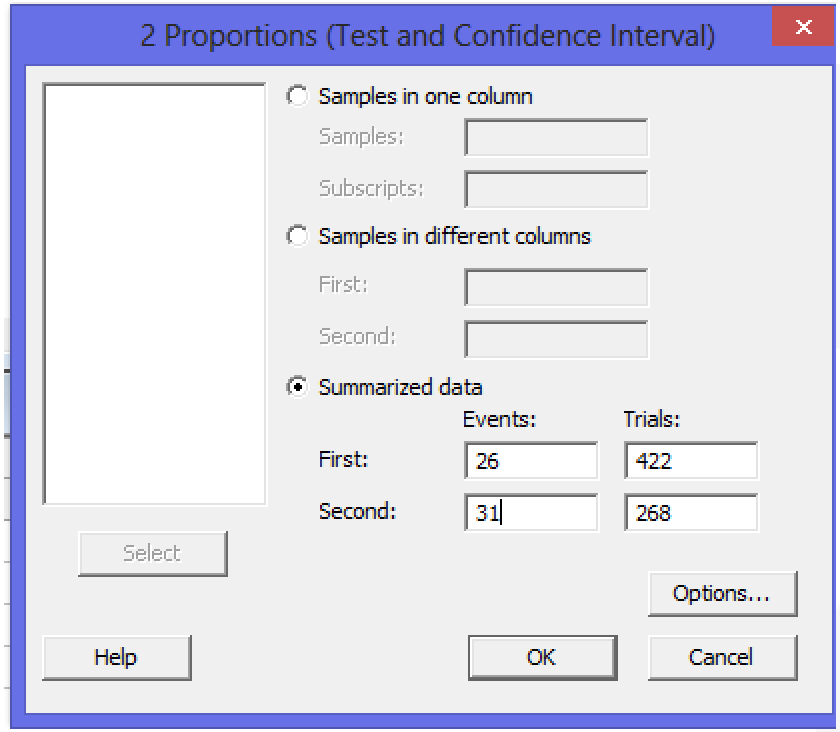
- They set up a small test run that will sample customers randomly as they go to on-line. When this sample of customers goes to the web site, they will be offered these products with the proposed sales pitch.

- After the test the following numbers of products were sold:
  - A: 31
  - B: 41
  - C: 45
  - D: 61
  - E: 82
- Are you comfortable launching this new tool with this profit plan?

Exercise 3, 2-Proportions

- **A bank has made a concerted effort to increase the number of loans made to minority owned small businesses in its area. Sampling the data for last year, it was noted that out of 422 loans to small businesses, 26 had minority ownership. The first half of this year, they had made 268 loans of which 31 had minority ownership.**
- **Have their efforts to increase loans to minority owned small businesses been successful? It appears that they have been successful, but can we statistically prove it with these sample sizes?**

Exercise 3, 2-Proportions



Test and CI for Two Proportions

```
Sample   X    N  Sample p
1       26  422  0.061611
2       31  268  0.115672

Difference = p (1) - p (2)
Estimate for difference:  -0.0540603
95% CI for difference:  (-0.0986979, -0.00942259)
Test for difference = 0 (vs not = 0):  Z = -2.37  P-Value = 0.018

Fisher's exact test: P-Value = 0.015
```

Key numbers circled, what is the message we take back to the business?

Exercise 4, Proportions vs. Proportions

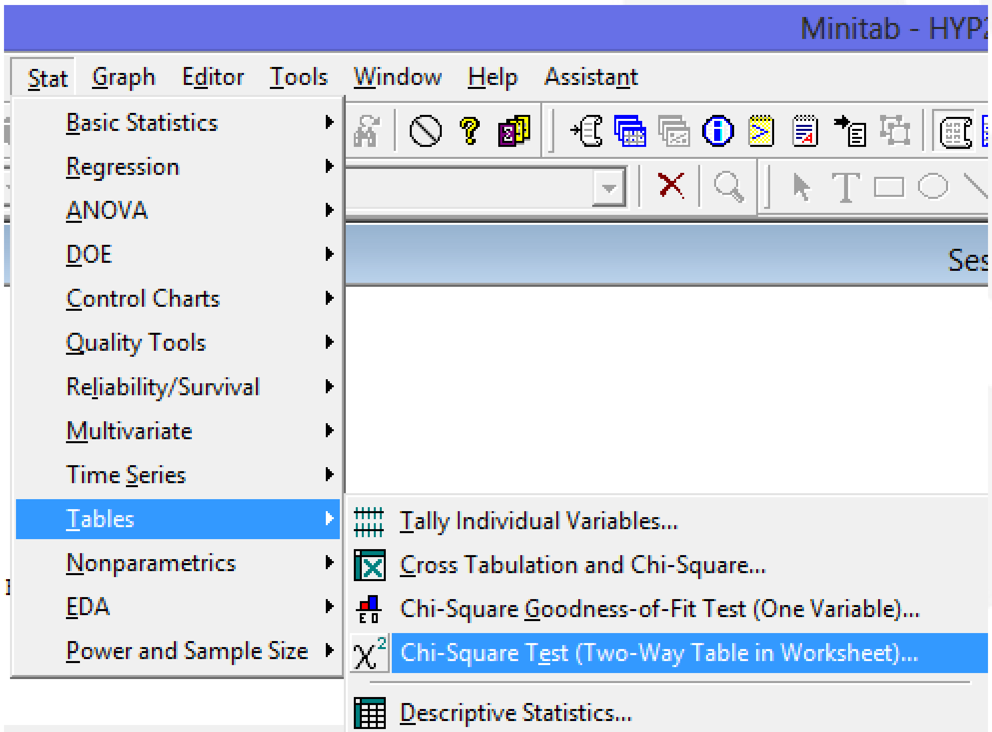
We want to know if there is any truth to "more errors are made on Monday and Friday than other days of the week." At different store locations, we sample a number of transactions randomly throughout the week and get these results:

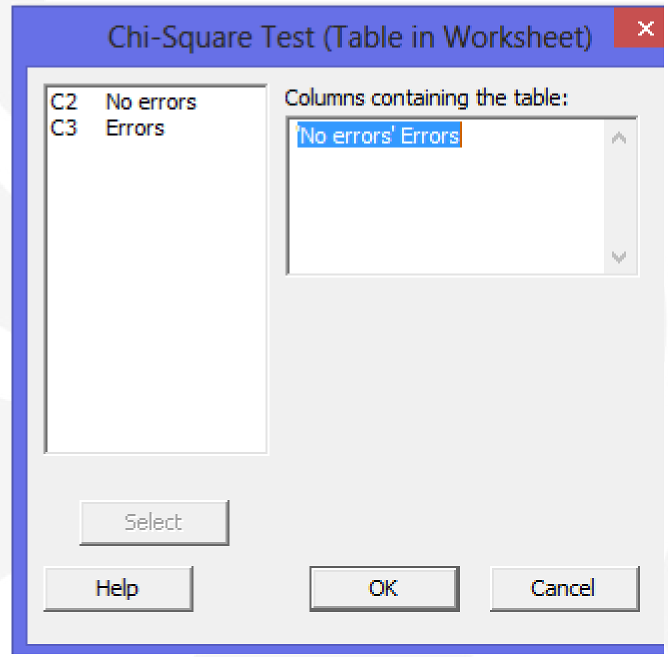
Day of Week	No Errors	Errors
Monday	170	14
Tuesday	215	28
Wednesday	192	17
Thursday	177	12
Friday	168	32

Exercise 4, Minitab

Data Table: Hyp2_Discr_Many_Levels_Ex4

Stat > Tables > Chi-Square test





Exercise 4, Analysis Results

Chi-Square Test: No errors, Errors

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

        No errors   Errors   Total
    1         170       14     184
           165.51    18.49
            0.122    1.090

    2         215       28     243
           218.58    24.42
            0.059    0.525

    3         192       17     209
           188.00    21.00
            0.085    0.763

    4         177       12     189
           170.01    18.99
            0.288    2.574

    5         168       32     200
           179.90    20.10
            0.787    7.049

Total         922      103    1025

Chi-Sq = 13.342, DF = 4, P-Value = 0.010
```

- **What decision do we make?**
- **Look for the highest Cell Chi^2 number to see which cell is driving the difference.**
- **Which day is different?**
- **Good or bad different?**

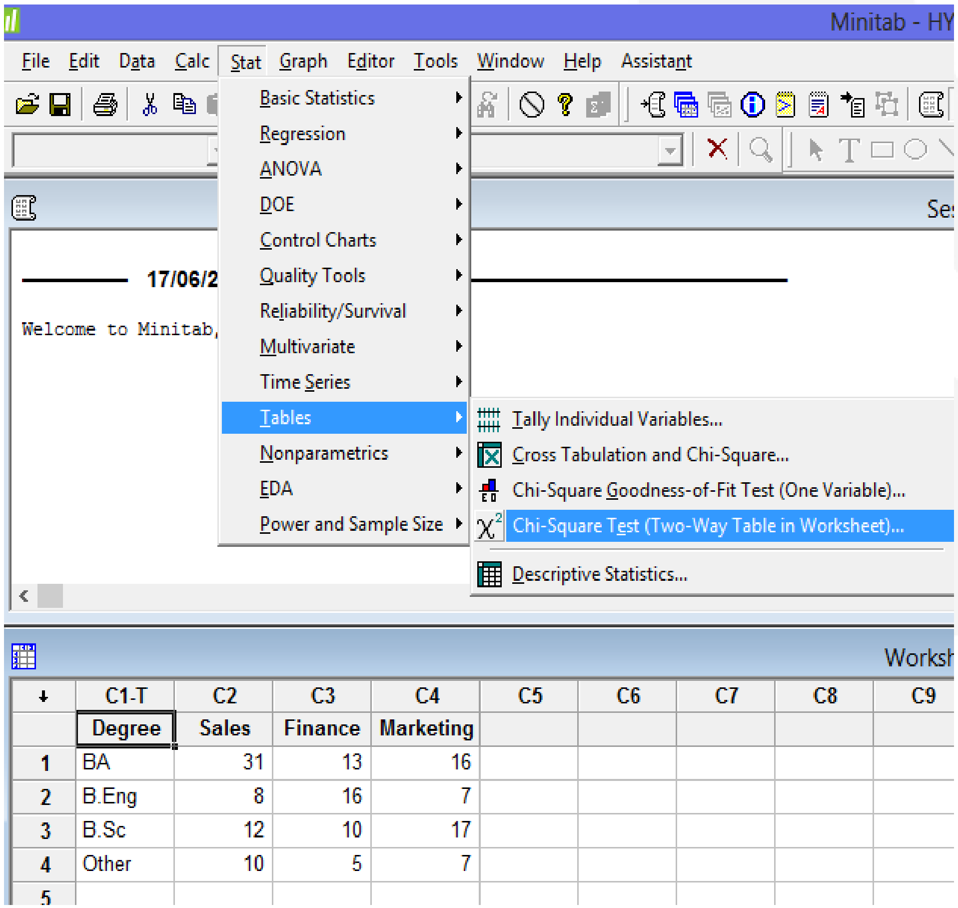
Exercise 5, Two Factors

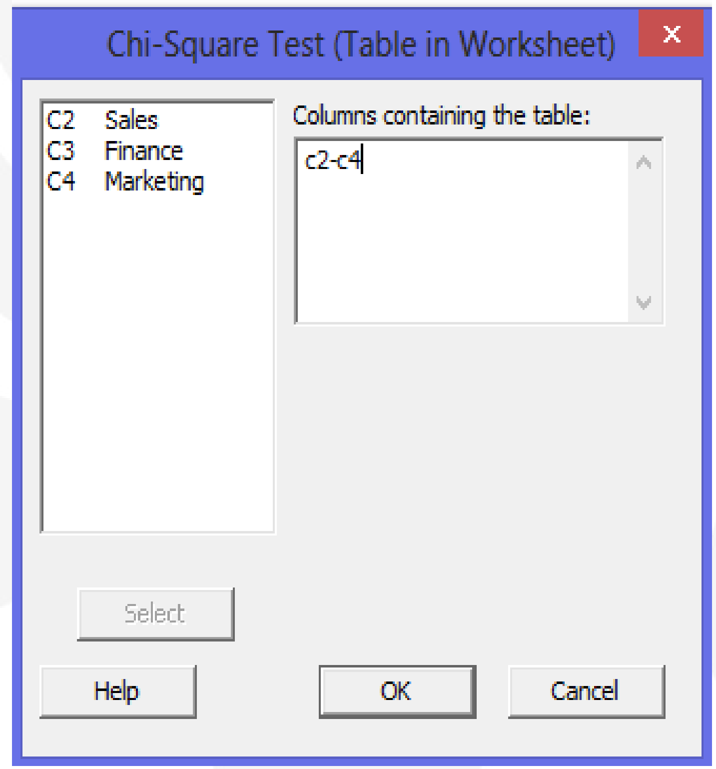
- **A recruitment function wants to understand whether or not the candidates degree is a significant factor in the type of job they have**
- **The following data has been collected and pre-loaded into Minitab**
- **This data can be found in the file: Hyp2_Disc_2_Factors_Ex5**

Degree	Sales	Finance	Marketing
B.A.	31	13	16
B.Eng	8	16	7
B.Sc	12	10	17
Other	10	5	7

Exercise 5, Minitab

Stat > Tables > Chi-Square test





Exercise 5, Session Output

Chi-Square Test: Sales, Finance, Marketing

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected
counts
        Sales  Finance  Marketing  Total
    1      31       13         16     60
        24.08    17.37      18.55
        1.989    1.099      0.351

    2       8       16          7     31
        12.44     8.97       9.59
        1.585    5.502      0.697

    3      12       10         17     39
        15.65    11.29      12.06
        0.852    0.147      2.024

    4      10        5          7     22
         8.83     6.37       6.80
        0.155    0.294      0.006

Total      61       44         47    152

Chi-Sq = 14.702, DF = 6, P-Value = 0.023
```

- **Use the p-value to make the decision.**
- **What decision do we make?**
- **Look for the highest Cell Chi^2 number to see which cell is driving the difference.**
- **Which cell(s) is/are different?**

Chi-Sq statistic 14.702

P-value = 0.023

How does Chi-Square work?

- **In order to determine whether there is a relationship between the two discrete variables being analysed, the chi-square test:**
  - Uses actual observations to compute expected frequencies for each cell
  - Uses the difference between the observed and the expected to calculate a chi-square statistic for each cell
  - Sums the chi-square statistic for all the cells
  - Determines a p-value for the chi-square statistic calculated
  - We determine that there is no significant difference between the observed and expected frequencies if the p-value>0.05

The expected frequency for each cell

- The expected frequency for each cell is calculated by dividing the row total of the actual frequencies by the grand total and multiplying by the column total

$$\text{Expected frequency} = \frac{\text{Row Total}}{\text{Grand Total}} \times \text{Column Total}$$

$$\text{Expected frequency} = \frac{60}{152} \text{ X } 61 = 24.08$$

Chi-Square Test: Sales, Finance, Marketing

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected
counts
           Sales   Finance  Marketing   Total
     1        31        13         16      60
           24.08     17.37      18.55
           1.989     1.099      0.351

     2         8        16          7      31
           12.44      8.97       9.59
           1.585     5.502      0.697

     3        12        10         17      39
           15.65     11.29      12.06
           0.852     0.147      2.024

     4        10         5          7      22
            8.83      6.37       6.80
           0.155     0.294      0.006

Total         61        44         47     152

Chi-Sq = 14.702, DF = 6, P-Value = 0.023
```

The Chi-Square statistic for each cell

- The difference between the cells in the table is measured by the chi-square statistic
- The Chi-Square statistic for each cell is calculated by squaring the difference between the expected and the observed

$$\text{Cell Chi-Square Statistic} = \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

$$\text{Chi-Square Statistic} = \frac{(31-24.08)^2}{24.08} = 1.989$$

Chi-Square Test: Sales, Finance, Marketing

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected
counts
          Sales  Finance  Marketing  Total
    1        31       13         16     60
          24.08    17.37      18.55
          1.989    1.099      0.351

    2         8       16          7     31
          12.44     8.97       9.59
          1.585    5.502      0.697

    3        12       10         17     39
          15.65    11.29      12.06
          0.852    0.147      2.024

    4        10        5          7     22
           8.83     6.37       6.80
          0.155    0.294      0.006

Total        61       44         47    152

Chi-Sq = 14.702, DF = 6, P-Value = 0.023
```

Overall Chi-square = (Cell 1) + (Cell 2)+…..(Cell n)

Summary of key learning points

- **Proportion tests are used when we have discrete independent and dependent variables**

- **Use proportion tests to test differences for comparison of discrete responses to:**

  - **a target proportion**
  - **against another one or many other proportions**

- **A Chi-square statistic is calculated for each cell with large values indicating significance**

- **The sum of all Chi-square statistic yields the overall statistic used to determine whether the factor was significant**

Black Belt

Surveys

Objectives of this module

At the end of this module, you will be able to:

- Be able to apply the basic concepts of survey construction.
- Understand what methods are available to validate a new survey.
- Know how to pre-test a survey
- Know various methods to analyse survey data.
- Understand the concept of **non-response bias**, how that can affect our conclusions and how to test for it.
- Be able to apply confidence intervals to survey results.

Questions Lead

- **Basic question: What are our customers saying?**
  - **Did we get good responses from our customers? Are the results valid?**
  - **Are we asking the right questions?**
  - **How sure are we that we truly understand what they are saying?**

Questionnaires Seek Two Types of Data

- **<u>Facts</u>: About the person, product, and use experience:**
  - **Name, age, income, tenure, etc.**
  - **Experience with product or process**
  - **Description of problems**

- **<u>Opinions</u>: About the offering (ours or competitors):**
  - **Importance of features/functions/benefits**
  - **Satisfaction with our offering**
  - **Satisfaction with competitors' offerings**
  - **Reaction to new concepts**

Basic Principles of Questionnaire Design

1. Before beginning:
   - Define what you want to accomplish with the questionnaire
   - Prioritise those items
2. Take time to consider the respondents. They control:
   - What you can ask
   - The words you can use
   - The methods you can use
   - The concepts you research
3. Don’t impose:
   - Your values
   - Your language
   - Your perceptions
4. Ask questions that the respondents:
   - Want to answer truthfully
   - Have knowledge or opinions about

Survey Design Steps

1. **Define the objectives.**
2. **Select the data collection method.**
3. **Decide the type of questions to be asked.**
4. **Draft the questions.**
5. **Organise the questions into an overall design.**
6. **Validate the survey.**
7. **Pre-test the survey.**
8. **Implement the survey.**

Step 1, Define the Objectives

- **Study objectives**
- **Question objectives**
  - **How many questions?**
  - **What is the point of <u>each</u> question?**

Step 2, Select Method

Type of Survey/ Characteristic	Mail	Phone	Phone* Automated Call Back	In-Person	Group Sessions Written	Electronic
Data Collection Costs	Low	Moderate	Moderate	High	Moderate	Low
Time Required To Collect	High	Low	Medium	High	Medium	Low
Response Rate	Low	Moderate	Low	High	High	Moderate to Low
Interviewer Bias	None	Moderate	None	High	Low	None
Acceptable Length Of Survey	Medium	Short (Maximum 15 Minutes)	Medium	Medium (Maximum 45 Minutes)	Long	Short
Ability To Obtain Open-Ended Responses	Low	Low	Low	High	High	Medium
Perceived Anonymity	High	Low	Moderate	Low	Moderate	None

Step 2, Select Method (continued)

- **Selecting the method is important because of all the characteristics on the previous slide, but also because it dictates the types of questions that can be asked.**
  - **Questions for the respondent to read can be very different from questions that an interviewer can ask.**
  - **Some methods allow for illustrations, some don't.**
  - **Some methods allow for follow up questions, some don't.**

Step 3, Decide the Type of Questions to be Asked

- **Closed-ended questions:**
  - **Yes, no**
  - **Rankings or ratings**
  - **Specific data**
- **Open-ended questions:**
  - **Opinions**
  - **Describe something**

Rating Scales (1 of 2)

- **Length of the scale:**
  - **The longer the scale, the more discriminating.**
  - **But the longer the scale, the less likely people are to use it after about 8 to 10 questions.**
  - **If you are going to have more than 10 questions use a shorter scale (5 to 7 points). Pre-testing will help you decide.**
  - **Some research suggests that the human mind can at most juggle 7 categories.**
- **There is no evidence that reliability of the scale:**
  - **Is different whether the scale points were labeled with words or numbers.**
  - **Is different whether the scale points were all labeled or just the end points.**
  - **Is different with a neutral point in the scale vs. not.**

Rating Scales (2 of 2)

- **Much of the power of a scale comes from having applied it in several studies to learn how respondents use it.**
  - **Even if there are some inaccuracies such as non-response bias (discussed later), using the same scale over time can be an accurate means of tracking progress.**
  - **Using the same scale can also give us a means for determining stability over time.**
  - **Tip: Check to see if the respondents are actually using all the values on the scale. If not, we should ask if we need that many values.**

Step 4, Draft the Questions

1. **Ask specifically for the information you want; avoid ambiguity.**
   - **The question, "What do you think of our benefits program?" could mean:**
     - **"Have you had any problems with our benefits program?"**
     - **"Do you think we have the right benefits?"**
     - **"How does our program rate versus other world class companies?"**

Step 4, Draft the Questions (continued)

2. Ask about one concept or dimension at a time.

- **Your laptop was repaired by your IT department. You then received a questionnaire with this question:**

Competence:
Did you find the tech knowledgeable and able to assist you?
○ Excellent ○ Good ○ Satisfactory ○ Need Improvement ○ Unacceptable

- **Your response: “He was very able to assist me. I have no way of knowing how knowledgeable he is. Didn't you know how knowledgeable he was before you hired him?”**

Step 4, Draft the Questions (continued)

3. **Clearly and consistently communicate to respondents the type of suitable answers.**
    - **In 2000, this question from a company that makes baby formula was asked of new mothers (less than 5 weeks):**
    - **"How long was your baby in the hospital?"**

Although most gave the answer in "days", a large percentage gave the answer in "_______?______".

Step 4, Draft the Questions (continued)

4. **Only ask questions in which respondents are willing and able to answer accurately.**
   - **You receive a questionnaire from the Leader of your Business Unit asking: "How likely are you to successfully complete your Black Belt project on time and on budget?"**
   - **What is your answer?**
   - **What are other questions that respondents might not want to answer truthfully?**

Step 4, Draft the Questions (continued)

5. **Make sure questions are asked consistently by all interviewers and understood in the same way by all respondents.**

- **A company surveying its sales force asked "What percent of your time is spent selling products that are not covered by your bonus plan?"**

- **If asked how to calculate this percentage, what would you say?**

Step 4, Draft the Questions (continued)

6. Review each question and think how the result will be used.

- **Ask colleagues or stakeholders:**
  - **"If we had this result … would it help us understand the issue better?" "How would it help us to make a better decision?"**
  - **Sometimes a manager will describe how he/she wants to use the results and it will be in a manner which is inconsistent with how you plan to design the study.**

- **Write an analysis plan for the questions that assures:**
  - **Believability**
  - **Accuracy**
  - **Accessibility**

Step 4, Final Thoughts & Examples (1 of 2)

- **If you ask a question, people will answer it—no matter how stupid the question.**
- **People will answer the question they hear, not what you thought you asked.**

Real question asked at trial:

Question: Doctor, how many autopsies have you performed on dead people?

Answer: All my autopsies have been performed on dead people.

Gary, all your responses must be oral.

Question: OK?

Answer: Oral

Question: How old are you?

Answer: Oral

Step 5, Organise the Survey

- **Do the actual design of the survey.**
  - **Write the introduction and instructions.**
  - **Format it to make it easy to follow and understand.**
  - **If it is being transferred from one method to another, make sure it is adapted to the new survey methodology.**
  - **Sequence the questions for maximum accuracy and completion rates.**

Step 5, Organise the Survey (continued)

- **Sequencing Rule: Funnel from general to specific.**
  1. **"Describe the worst experience you have had with the company?."**
  2. **"What is your overall satisfaction with the company?"**
- **Sequencing Rule: To increase the completion rate, the first question(s) should be:**
  - **Relevant**
  - **Interesting**
  - **Easy to answer**
  - **Non-threatening**
  - **Closed or partly closed with few categories**

Step 6, Validate the Survey

- Factor analysis and regression can be used to analyse surveys that ask multiple questions on a topic or "construct". A construct is an idea – usually a complex idea that is hard to examine through just one question or component.

- One construct a company may survey regularly is **customer satisfaction**. That construct is made of up **several components** – access, service, rates, etc. Customers could be asked about these separate components to get a better idea about the complex construct of customer satisfaction:
  - How satisfied are you with the access to your account with the company?
  - How satisfied are you with the service you receive?
  - How satisfied are you with the price?
  - How satisfied are you with the company overall?

- If your survey does not measure responses to questions that make up a "construct" or "constructs", but contains more straightforward data collection (how old are you, how many cars do you own, did you vote in the last election, etc), these tools will be less helpful.

Step 6, Validate the Survey (continued)

- **Ensuring the survey is valid (through analysis) leads to confidence and credibility in the results.**
- **Procedure to validate your survey:**
  1. **Determine if the questions "hang together" as ideas, topics, or constructs**
  2. **Determine if you have misaligned or redundant questions**
  3. **Determine if you're missing any significant ideas in the construct (Linear Regression)**

Step 7, Pre-test the Survey

1. Read the questions aloud.

- **Reading aloud often produces a different tempo than reading to oneself.**
- **Check to determine if verbal word emphasis is needed to make the questionnaire work. Any difference in these?**
  - **<u>WHY</u> did you buy that product?**
  - **Why did <u>YOU</u> buy that product?**
  - **Why did you buy <u>THAT</u> product?**

Step 7, Pre-test the Survey (continued)

2. **Have a colleague read the questions aloud.**
    - **Ask the colleague to tell you, in their own words, what they think the questions asked.**
3. **Ask real subjects to take the survey.**
    - **Observe respondents if possible.**
        - **Do they pause at certain points as if puzzled?**
        - **Do they ask for clarification?**
        - **How long does it take?**
        - **Do they speed up in the middle of a long set of attribute ratings?**

4. **Debrief the trial respondent about the questionnaire.**
   - **Walk through the questions with them.**
     - Ask their opinions & interpretation. What did they understand the question to mean?.
     - Did they feel comfortable that they could really answer the question truthfully?
     - Very important!—What did they want to tell you that the questionnaire didn't ask or didn't permit?
5. **Re-write and pre-test again.**

Step 8, Implement the Survey

- **If you have completed the first six steps well, this one is easy. Just begin the survey.**

- **If the questions are asked by people, get early feedback from the questioners about any problems. Also look at the early results to make sure you are getting the type of answer you expected and be ready to make revisions if the respondents are not understanding a question.**

Survey Analysis

- **Survey results are used to make statements about a population. We have to exercise caution when we analyse the results and state our conclusions.**
- **Two very common causes of survey misinterpretation are:**
  - **Bias due to low response rate (non-response bias).**
  - **Lack of statistical validity due to low sample size.**
- **This section discusses proper analysis techniques to avoid pitfalls due to these two issues.**

Who chooses to respond?

- **The sampling strategy for surveys is very often Random Sampling. Every member of the population has an equal chance of getting a survey. It is a great strategy!**
- **HOWEVER, the subgroup within that sample that chooses to respond is NOT random (the respondents are self-selected).**
- **People who respond tend to have stronger motivation to do so than people who do not respond. E.g. very bad or very good experience, desire to "vent", very high delight, etc.**
- **Therefore the survey results from the responding group are often not representative of the overall population.**

Response rate *R*

- **In most cases, only a fraction of the people who are offered a survey will respond.**
- **This fraction of people who choose to respond is the response rate, *R.***
- **If 1000 people are offered a survey and 150 people respond (complete the survey), the response rate is:**

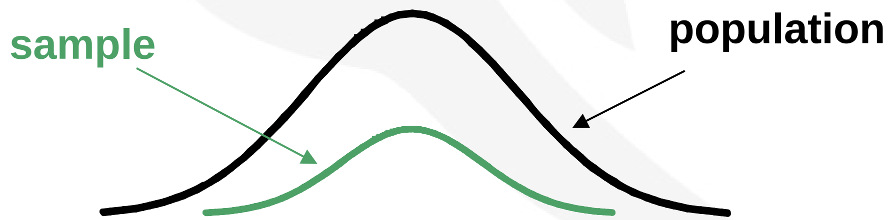
$$R = \frac{150}{1000} = 0.15 = 15\%$$

- **If our sample size calculator called for 200 samples and we expected a 10% response rate, how many surveys would we need to send out?**

$$n_{\text{offered}} = \frac{n_{\text{needed}}}{R}$$

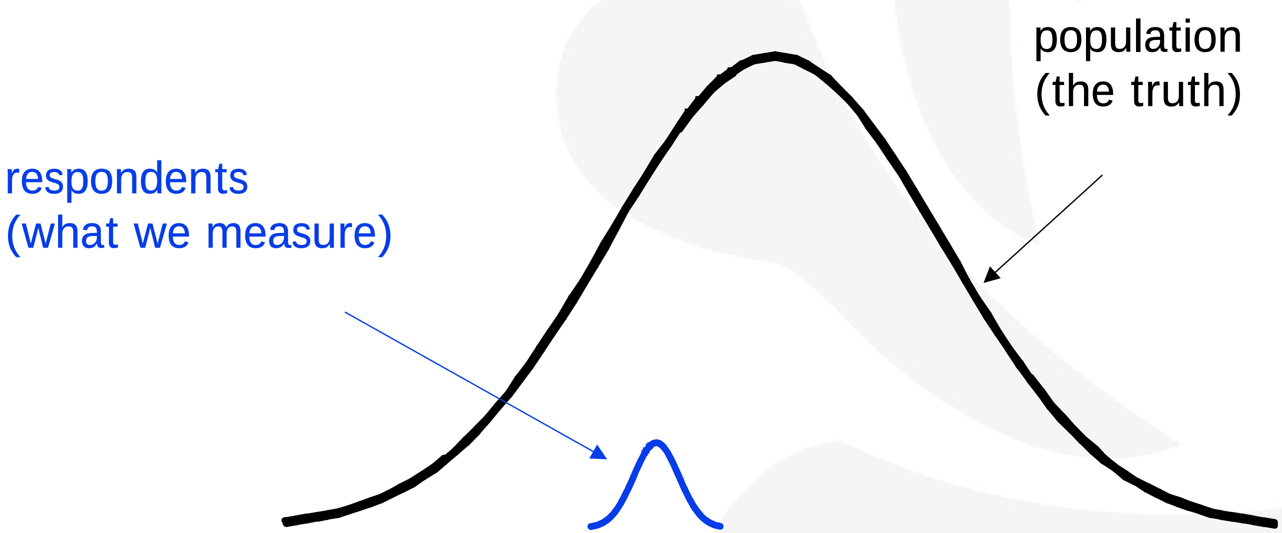
Validity of the Results

General Law of Inferential Statistics: Inferential statistics are valid if and only if an unbiased, representative sample is taken from the greater population.



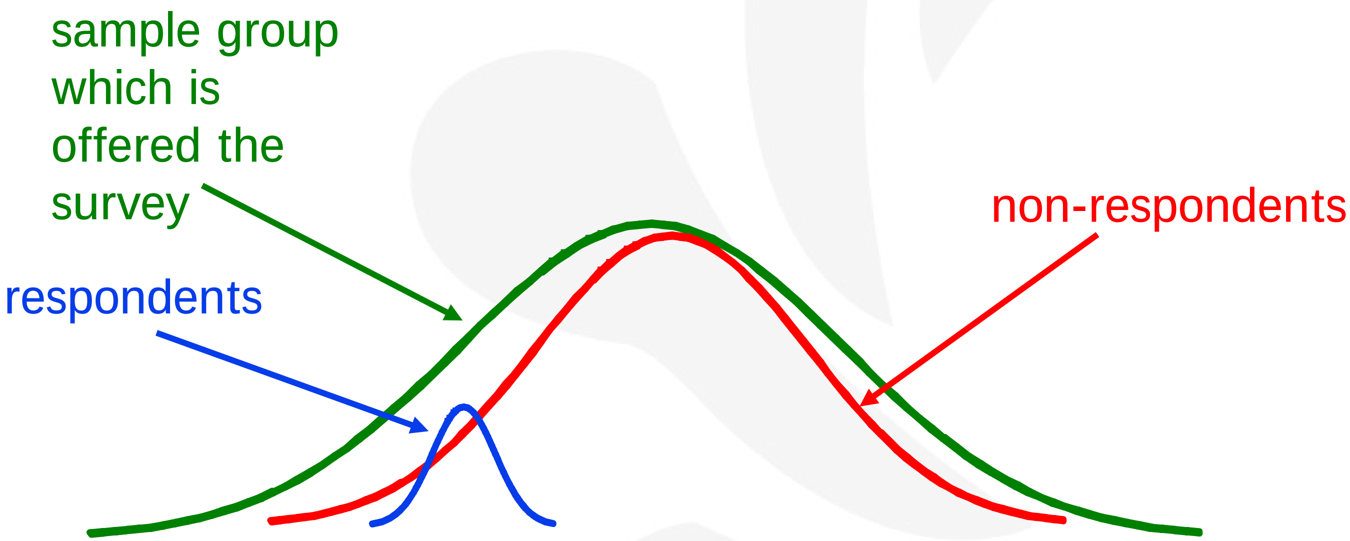
- **Because the respondents of a survey are typically NOT a representative sample of the population, sampling statistics may NOT be valid.**
- **The lower the response rate *R*, the greater the potential bias in the results.**
- ***"Any survey with a response rate lower than 85% must undergo a non-response bias analysis before any of the data can be used." US National Centre for Education Statistics, Standard 4-4***

Do our survey results reflect reality?



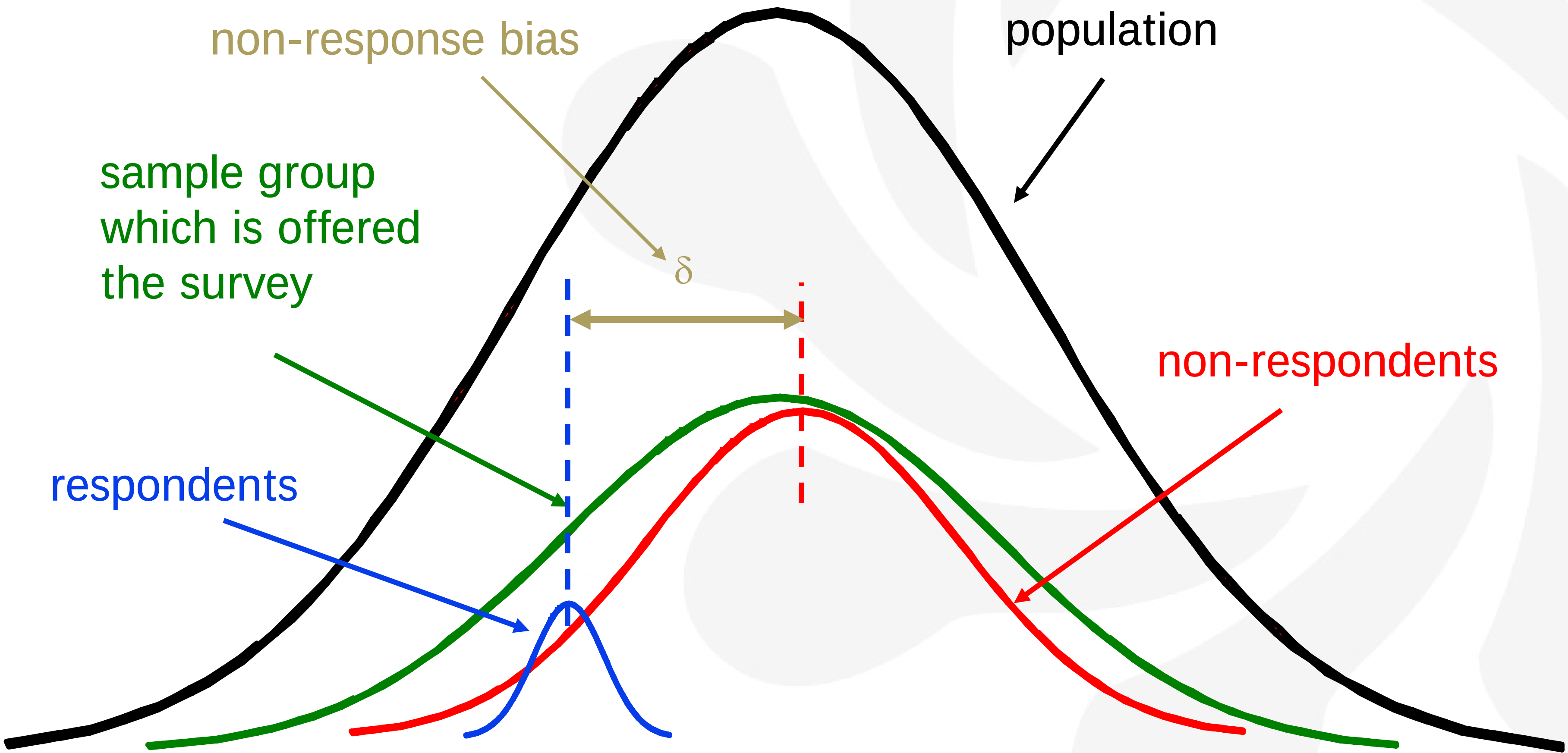
- **Note that initially we only have data from the group of respondents.**
- **If the response rate is low (R < 85%) we CANNOT assume that the respondents give us a representative sample of the population.**

Two Distinct Groups in Our Sample



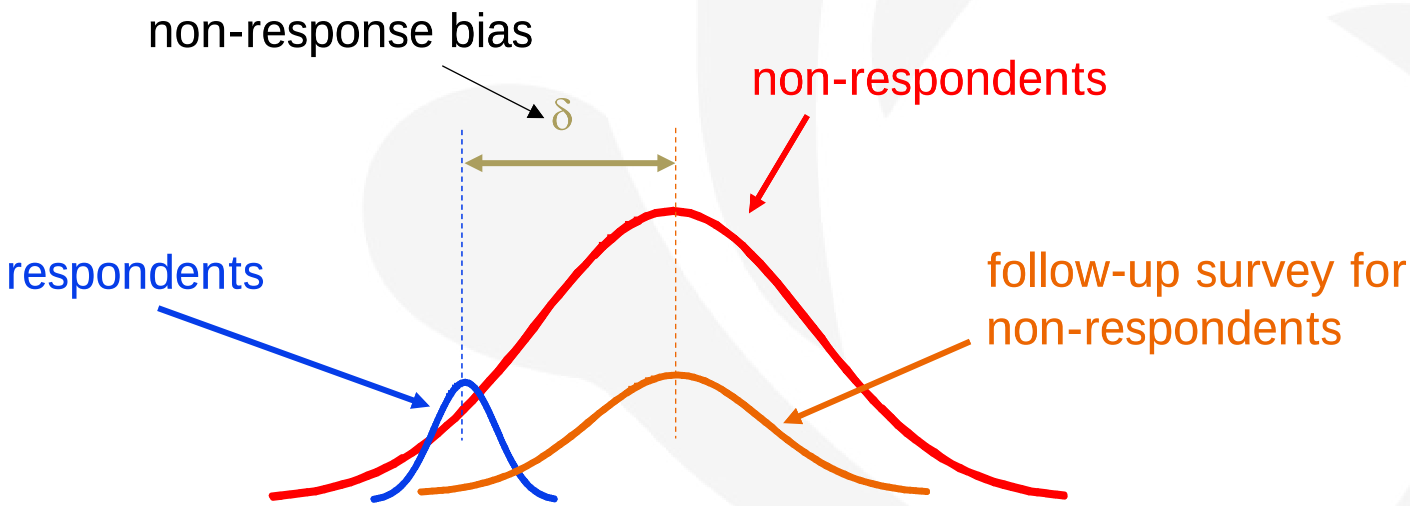
- **The sample of people who are offered the survey is divided into two groups: those who responded and those who did not.**
- **These two groups may be very different.**
- **We only have information on the responding group.**
- **But to tie our results to the population, we must probe the "silent majority" of the non-respondents to get the non-response bias.**

The Non-Response Bias δ



***The non-response bias** d is the average difference between the respondent group and the non-respondent group* **.**

Measuring the Non-response Bias



<u>How we do it:</u>

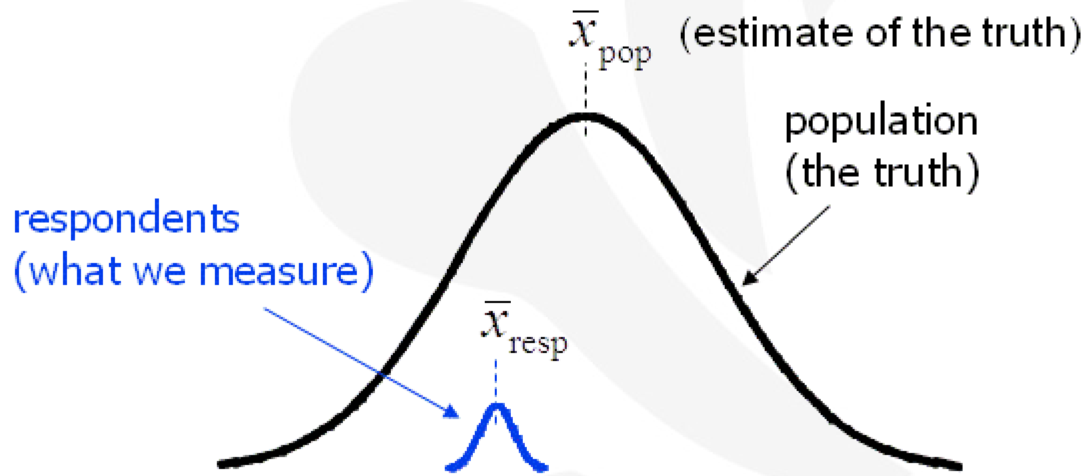
- **Administer a follow-up survey to a random sample of the non-responding group which asks the same questions as the original survey.**
- **This group should be incentivised to respond--the response rate of this follow-up survey must be at least 70%.**
- **The non-response bias <u>for each question</u> on the survey must be measured by subtracting the mean score of the responding group from the mean score of the non-responding group:**

$$\delta = \bar{x}_{non} - \bar{x}_{resp}$$

Note About Incentives

- Using incentives is a standard way of increasing response rates for surveys.
- Studies have shown that proper use of incentives should not introduce a bias in the results.
- Be sure to completely decouple the incentive from the response.
- In other words, make sure it is clear that the incentive will be given independent of how high or low the survey scores are.

Correcting for the Non-response Bias



Once we know the non-response bias δ for each question, we can calculate the estimate of the true population mean score ($\bar{x}_{pop}$) from the respondents mean score ($\bar{x}_{resp}$). For each survey question:

$$\bar{x}_{pop} = \bar{x}_{resp} + (1 - R)\delta$$

Non-response Bias Example (Ave. Scores)

A telephone survey was conducted by a banking centre asking the following three questions (to be answered on a scale from 1 to 10):

1. **How satisfied are you with the hours of operation of our banking centre?**
2. **How satisfied are you with the quality of service from our banking centre?**
3. **How satisfied are you with wait times at our banking centre?**

500 customers were randomly selected and were called. 45 of those called agreed to take the survey. The average scores were as follows:

initial survey results:

Question	Ave. Score
1	8.5
2	7.4
3	9.0

Because the response rate was low (R = 45/500 = 9%) the results are potentially biased. Therefore a non-response bias analysis is required…

Non-response Bias Example (Ave. Scores)

- **Out of the 455 people who did NOT respond to the initial survey, 40 people were randomly chosen and called again.**
- **This time, they were offered an incentive for their participation.**
- **36 out of the 40 people agreed to take the survey (follow-up survey response rate > 70%).**
- **The follow-up results are shown below:**

follow-up survey results (non-response group):

Question	Ave. Score
1	9.2
2	8.5
3	8.3

We can now calculate the non-response bias and therefore estimate the true population average scores for each question…

Non-response Bias Example (Ave. Scores)

We first calculate the non-response bias for each question:

$$\delta = \bar{x}_{non} - \bar{x}_{resp}$$

We can then calculate the estimate of the true population score:

$$\bar{x}_{pop} = \bar{x}_{resp} + (1 - R)\delta$$

where *R* is the initial survey response rate of 9%.

The results for this example are:

Survey Question	Initial Survey Ave. Score $\bar{x}_{resp}$	Non-respondents Ave. survey Score $\bar{x}_{non}$	Non-response Bias δ	Population Ave. Score $\bar{x}_{pop}$
1	8.5	9.2	+0.7	9.1
2	7.4	8.5	+1.1	8.4
3	9.0	8.3	- 0.7	8.4

Non-response Bias Analysis Guidelines

- **The non-response bias correction must be performed separately for each survey question—the non-response bias will vary from question to question.**
- **If the process or scenario surrounding the survey changes, then the non-response bias will likely change as well.**
- **Therefore, if the process has been improved or changed, then a new non-response bias analysis must be performed for subsequent surveys even if the survey questions have not changed.**

Discrete Data Proportion Results

- **What if the survey results are expressed in two categories, e.g. delighted vs. not delighted?**
- **In that case, the non-response bias is expressed in terms of the proportion delighted for each survey question rather than average score.**
- **If *p* represents the proportion of delighted customers then the non-response bias for each question is:**

$$\delta = p_{non} - p_{resp}$$

where p_{resp} is the proportion delighted from the responding group (initial survey results) and p_{non} is the proportion delighted from the non-responding group (follow-up survey results).

Estimate of Population Proportion

Our estimate of the true population proportion delighted is then:

$$\hat{p}_{pop} = p_{resp} + (1 - R)\delta$$

Note that this formula is completely analogous to the previous formula for average score:

$$\bar{x}_{pop} = \bar{x}_{resp} + (1 - R)\delta$$

Non-response Bias Example (Proportions)

Let's look at an example of a survey given by a call centre in which we are measuring the proportion of customers giving top-two box scores (9 or 10 on a 10-point scale) for the following questions:

1. **What is your satisfaction with the call center answer time?**
2. **What is your satisfaction with the associate to whom you spoke?**
3. **What is your overall satisfaction with the company?**

320 customers who called the centre within the last week were mailed surveys having the above 3 questions. 48 people mailed back the surveys. The top-two box (TTB) percentages were:

initial survey results:

Question	TTB Percent
1	84%
2	51%
3	38%

Again our response rate was low (*R* = 48/320 = 15%) and therefore we need to conduct a non-response bias analysis…

Non-response Bias Example (Proportions)

- **Out of the 272 people who did NOT respond to the initial survey, 50 people were randomly chosen and given the survey again.**
- **As before, they were incentivised so that 45 out of the 50 people responded (we need the follow-up survey response rate > 70%).**
- **The follow-up results are shown below:**

follow-up survey results (non-response group):

Question	TTB Percent
1	76%
2	68%
3	53%

Again, we calculate the non-response bias to get the estimate of the true population average scores for each question…

Non-response Bias Example (Proportions)

Non-response bias:

$$\delta = p_{non} - p_{resp}$$

Estimate of the true population TTB percent:

$$\hat{p}_{pop} = p_{resp} + (1 - R)\delta$$

where *R* = 15%

The results for this study are then:

Survey Question	Initial TTB Percent p_{resp}	Non-respondents TTB Percent p_{non}	Non-response Bias δ	Population TTB Percent $\hat{p}_{pop}$
1	84%	76%	- 8%	77%
2	51%	68%	+17%	65%
3	38%	53%	+15%	51%

What's the Alternative?

- **When our response rates are low, we must spend additional time and resources conducting a non-response bias analysis.**
- **Is there a better alternative?**

Yes! Try to get the response rate of your initial survey to be greater than 85%.

Then you don't need a non-response bias analysis.

Uncertainty of the Survey Results

- **Surveys for many purposes:**
  - **evaluating performances of individuals**
  - **evaluating performances of fulfillment centres, different branches of stores etc**
  - **evaluating desirability of products or services**
- **When we use surveys for evaluation purposes, we should always use our sample results to draw conclusions about the populations (all customers, all transactions, total performance of an employee).**
- **Therefore, we need to take the sampling uncertainty into account; maybe it was just an unlucky sample.**
- **This is done by calculating the confidence intervals for our results.**
- **If we fail to display the confidence intervals of our numbers, then our results can be misleading.**

Class Activity: Confidence intervals with surveys

Objective: Calculating CIs when assessing survey results

- **Confidence intervals for surveys are no different from confidence intervals for any kind of measurement data. Review the materials as you need to and calculate confidence intervals for these two situations.**

A survey was given to 20 employees (chosen randomly) asking 3 questions (to be answered on a scale from 1 to 10). Below are the results of the survey. The first table contains the averages and standard deviation treating the data as if it were continuous. The second table shows the percentages of Top Two Box answers received. Calculate the confidence intervals for these six responses.

Question	Ave. Score	St. Dev.
1	7.8	1.0
2	6.7	1.8
3	5.6	2.2

Question	TTB Percent
1	50%
2	35%
3	30%

Summary of key learning points

- People who choose to respond to surveys often have systematically different views than people who do not respond.
- Thus a non-response bias analysis should be performed for surveys where there is significant risk.
- **All survey results must include confidence intervals to depict the uncertainty due to the finite sample size.**
- Without proper analysis, survey results can cause us to make **WRONG** decisions:
  - Reward or punish people based on statistical noise.
  - Overreact to statistically insignificant changes in customer satisfaction scores.
  - Make invalid performance comparisons.
  - Fail to detect real process improvements due to negative response biases.

We must ensure there is appropriate analysis into survey misinterpretations

Black Belt

Multiple Linear Regression

Objectives of this module

At the end of this module, you will be able to:

- Complete a simple Linear Regression Analysis.
- Understand the Sum of Squares calculation.
- Complete a Multiple Linear Regression Analysis.
- Test the validity of model assumptions through **residual analysis**.

Questions Lead

- Basic question: Is there a relationship between my continuous inputs and my continuous output?
  - Do these Xs drive a change in my Y? All of them or just some of them? **Which** X is most important?
  - As these Xs increase, what happens to my Y?
  - How much does my Y change as the individual Xs increase?
  - Can we **predict** the Y assuming the X values? Can we **set** the X values to get the Y we want?

Application Examples

- Coffee shop – does **tenure** of staff affect the **wait time** of customers?

- Profitability – does the amount of **time** sales representatives spend with clients positively affect **revenue**?

- Call Centre Response Time – does the **number of calls per day** affect the **response time**? How about the **number of operators** in the call centre? Is the answer the same if we hold the other constant?

Regression is a Type of Modeling

“ All models are wrong, some are useful ! ” —
George Box

“The hallmark of good science is that it uses models and theories but never believes them.” —
Martin Wilk

Types of Models

- **Simple Linear Regression**
  - **One predictor variable (X)**
  - **Example: $Y = b_0 + b_1X + e$**

What does "e" refer to?

- **Multiple Linear Regression**
  - **More than one predictor variable (X_1, X_2, …, X_k)**
  - **Example: $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + e$**

Regression Requirements

Regression is a type of modeling and as such there are some inaccuracies that can creep into the analyses and yet the model can look good on the surface.

To discover problems, there are some **requirements** that have to be imposed on the data and verified.

Some will be verified via the **raw data** and some will be verified via the **residuals**. A residual is the difference between the value of an actual data point and value predicted by the model.

Regression Requirements (continued)

Raw Data Requirements

- Data Type. Regression is an analysis for a continuous Y and continuous Xs. However it works well for count data (discrete) if there is sufficient discrimination; i.e. there are enough different values of the counts. There are better tools (e.g. ANOVA) for categorical Xs.
- Linearity. The Y varies linearly with a change in the Xs. A violation of this will be seen in the Equal Variance test of the residuals.

Residual Requirements

- Normality. Residuals have to be normally distributed.
- Mean of Zero. The residuals should average zero. This is not a problem given modern software and it is easy to check.
- Independent. The Y value of one data point does not influence the value of the next one. Plotting the residuals in time order would give us an indication of problems here.
- Equal Variance across the range of the Y.

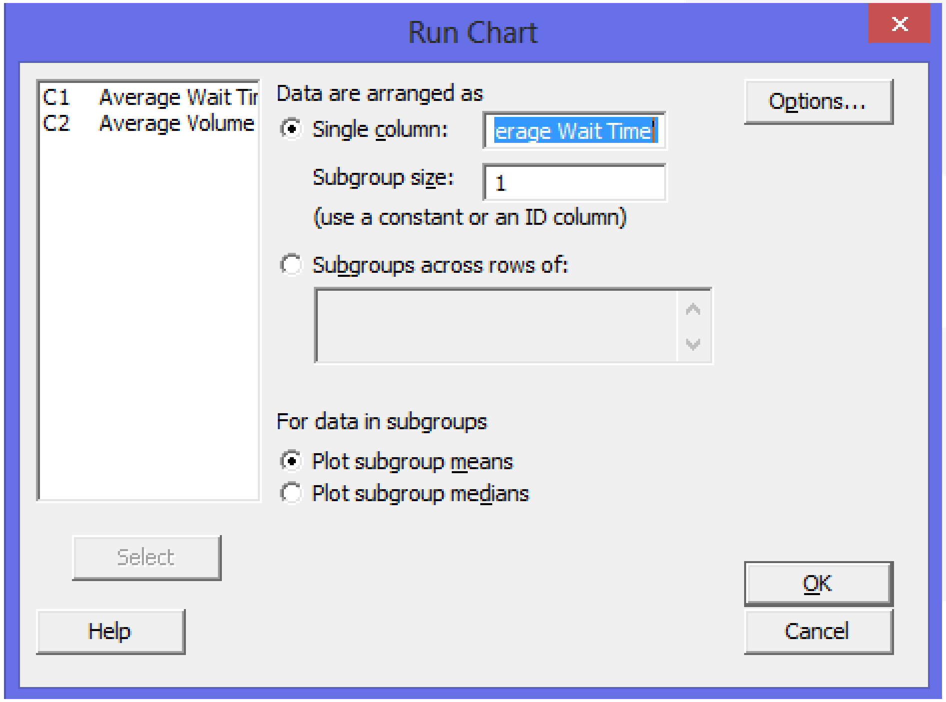
Exercise 1: Simple Linear Regression

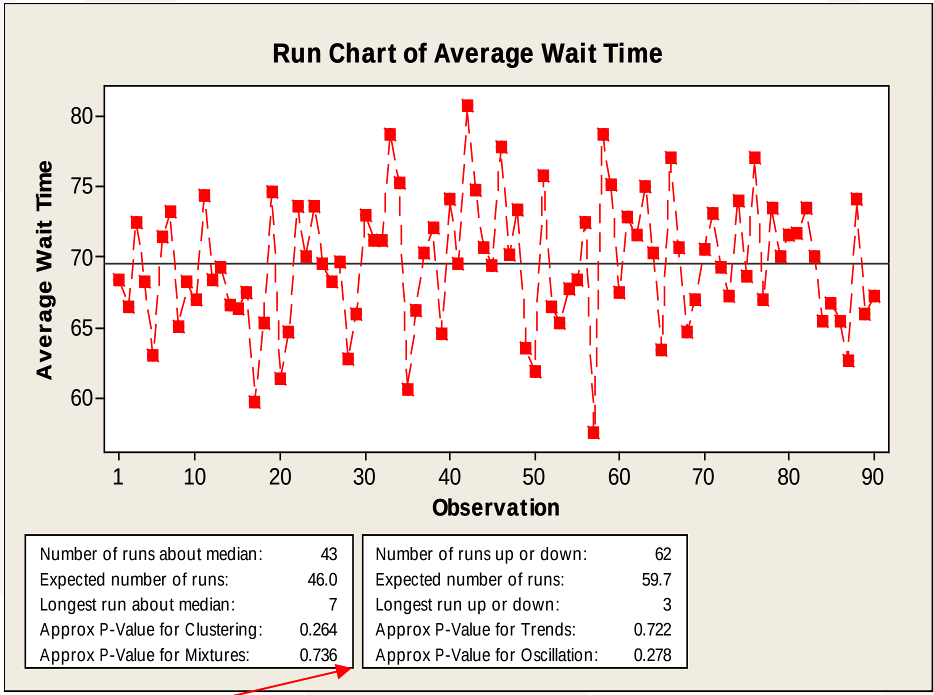
- **We need to reduce the waiting time for calls coming into our call centre.**
- **We think one factor driving the wait time might be the call volume. Data Table: Regr_Ex_1_Simple.MPJ**
  - **Business Question: Do wait times go up as the call volume increases?**
  - **Data: Wait time is certainly continuous. Call volumes are discrete counts, but they vary between four and six thousand per hour. Is this enough discrimination to continue with a regression analysis?**

Exercise 1: Simple Linear Regression

- Is the response independent?

Stat > Quality Tools > Run Chart



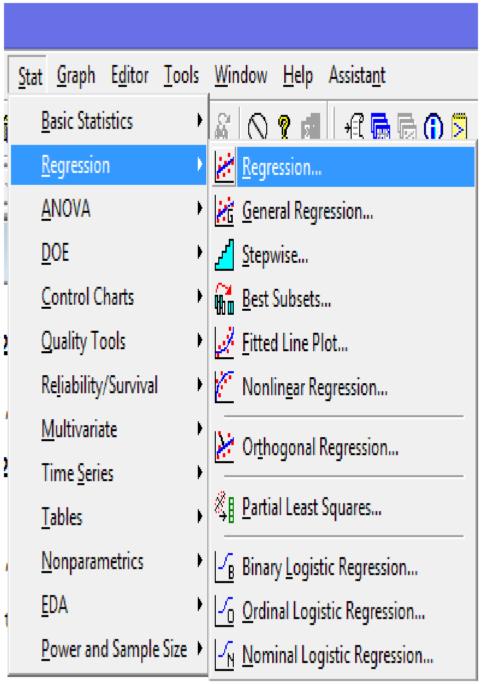


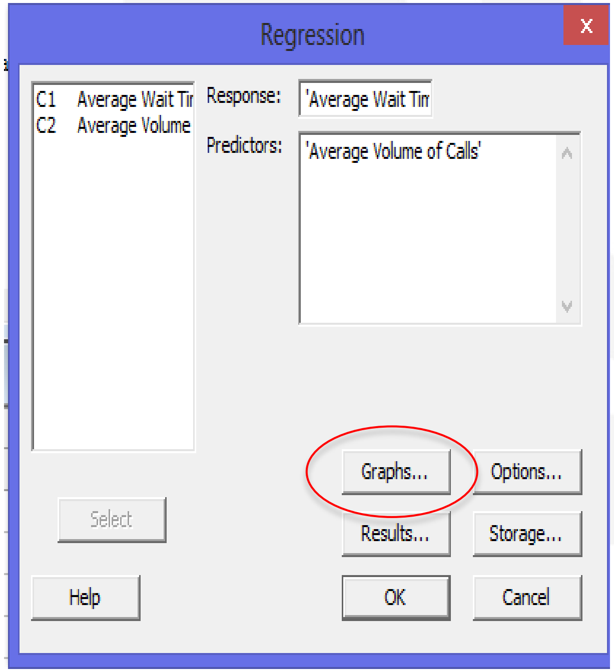
Ho: response = independent
Ha: response is not independent
Fail to reject the null hypothesis (response is independent)

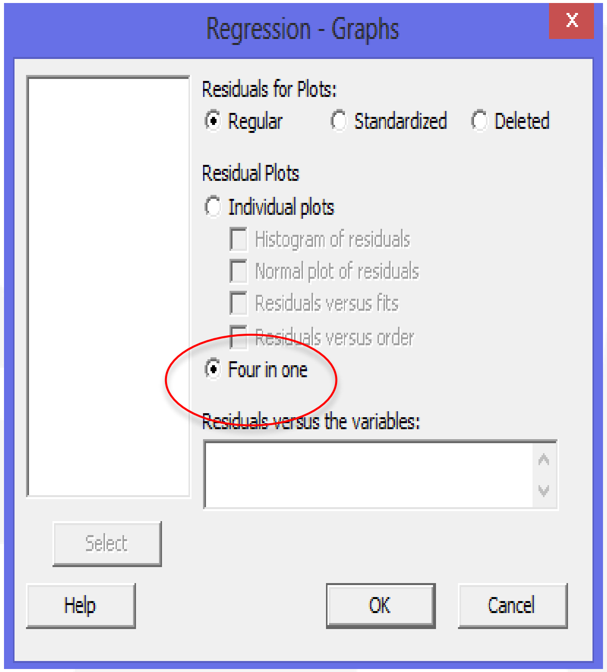
Exercise 1: Simple Linear Regression

Create model and gather information on residuals

Stat > Regression…





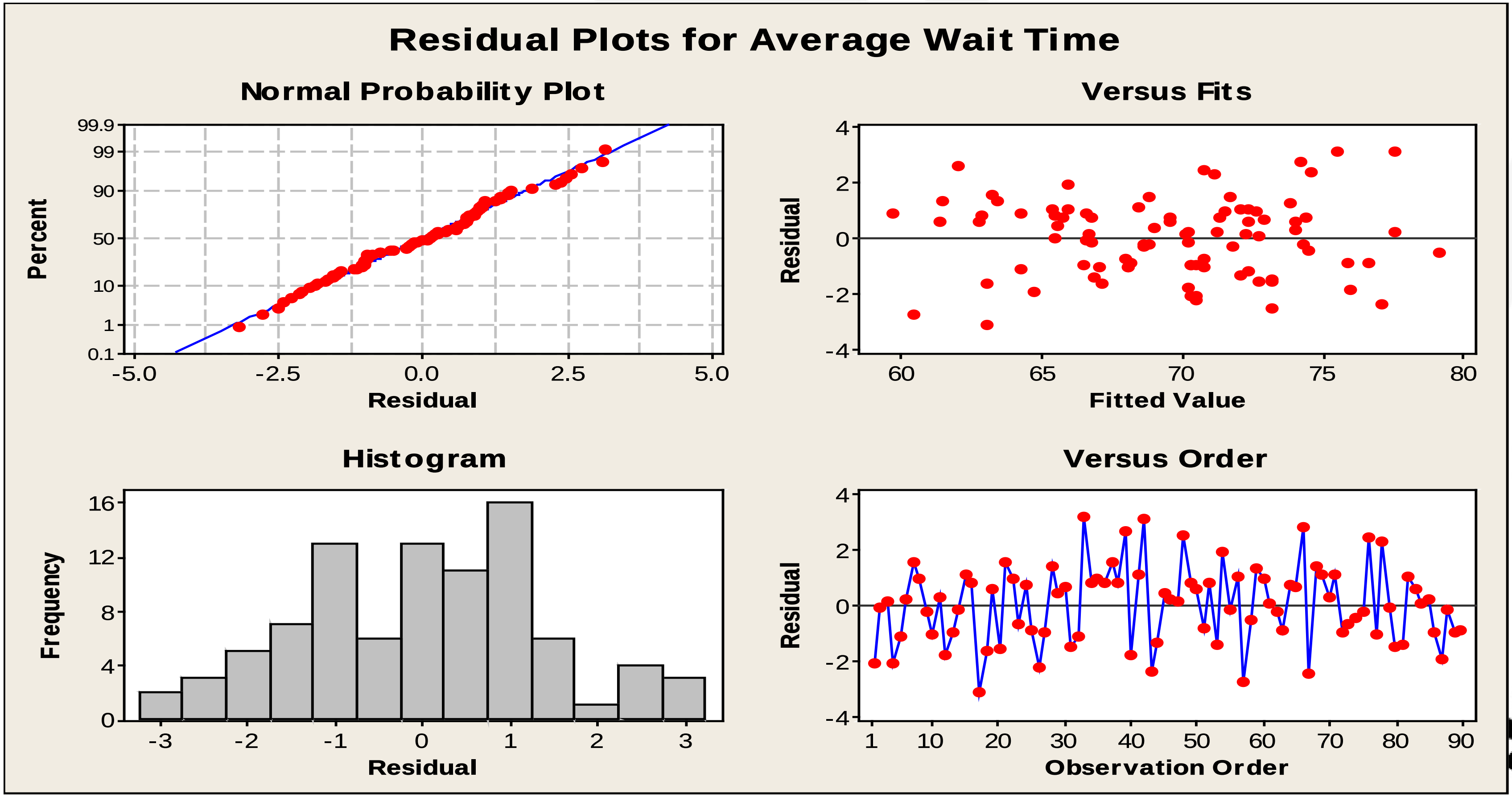


Exercise 1: Residuals analysis

- Do our residuals meet our requirements?

Normally distributed
Equally distributed around zero

Constant Variance
Independent



Exercise 1: Analysis – Is the model significant?

Regression Analysis: Average Wait Time versus Average Volume of Calls

```
The regression equation is
Average Wait Time = - 37.2 + 0.0214 Average Volume of Calls

Predictor                    Coef     SE Coef       T      P
Constant                  -37.186       3.626  -10.25  0.000
Average Volume of Calls 0.0213753   0.0007258   29.45  0.000

S = 1.39314   R-Sq = 90.8%   R-Sq(adj) = 90.7%

Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  1683.4  1683.4  867.34  0.000
Residual Error  88   170.8     1.9
Total           89  1854.2
```

Prediction equation

Significant factor

Coefficient of Determination
(see next slide)

Our model is significant

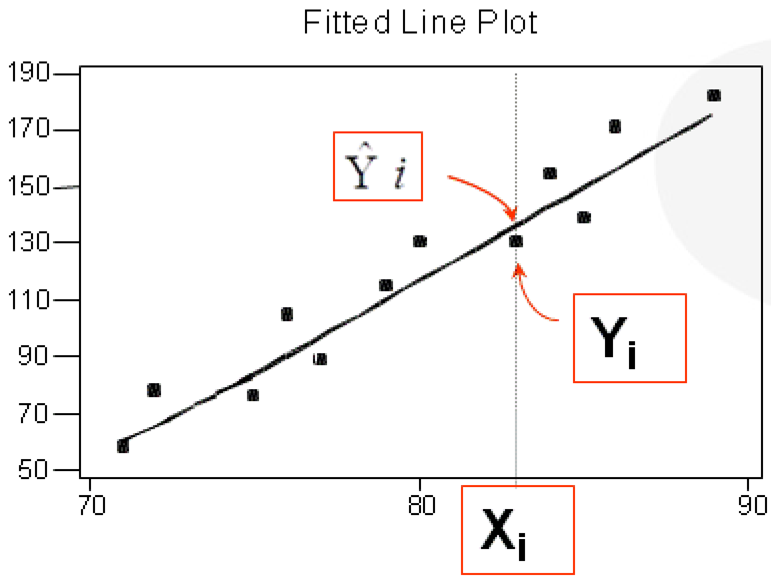
Coefficient of Determination, R-sq/R^2

- In simple linear regression, the measure R-sq (R-squared), R^2 is known as the “coefficient of determination”

- R^2 is a measure of the proportion of the variability in the output that is explained by the input

- The values of R^2 and R^2adj should generally be greater than 75% and should be within 5% of each other

Method of Least Squares



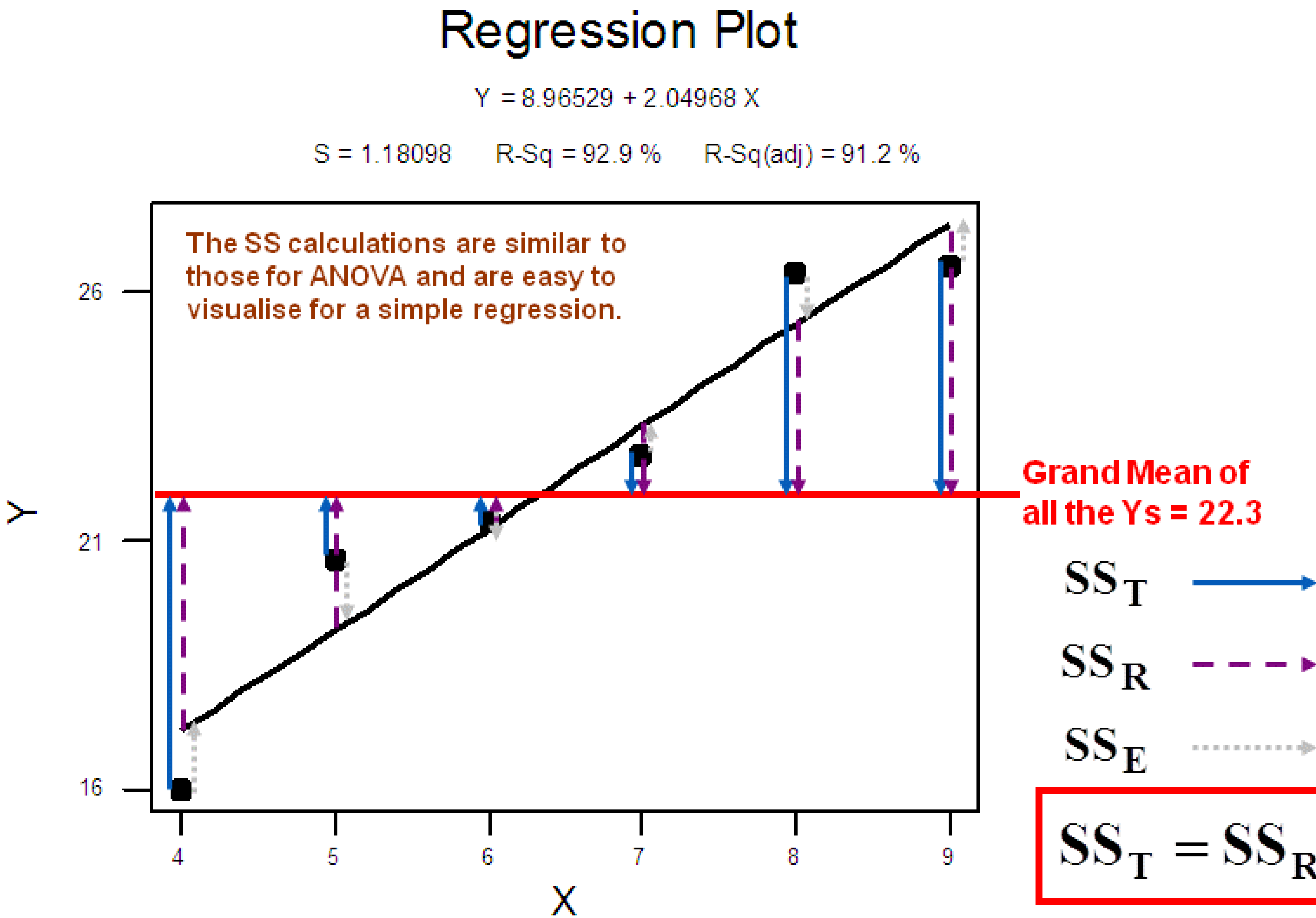
Y_i = Y value of the nth data point.

X_i = X value of the nth data point.

$\hat{Y}_i$ = Y value of the line at X value of X_i.

($Y - \hat{Y}$) is called a "residual" or error of prediction.
Least Squares Method minimises the Sum of Squares of error.

Visualising Sum of Squares



$$SS_T = SS_R + SS_E$$

R^2 Calculation

$$\mathbf{SS_T = SS_R + SS_E}$$

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

Where…

SS_R = Sums of Squares for the Regression

SS_E = Sums of Squares for the Error

SS_T = Sums of Squares for the Total

R^2 (coefficient of determination) measures the amount of variability of the response variable (Y) accounted for by the predictor variable (X); the proportion of total variation explained by the regression.

Multiple Linear Regression Equation

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n$$

Where…

X = independent variables (predictors)

Y = dependent variable (response)

b_0 = Y intercept (Where the line crosses Y axis, at X = 0, for simple linear regression. For multiple linear regression, the value of Y when all Xs = 0.)

b_x = coefficient for that variable (slope of curve dependent on that variable)

The b_x coefficients represent the contribution of the respective variables to the prediction of the response.

Partial Correlation

- Partial correlation forms the basis of multiple regression analysis. It is the concept that enables us to determine whether X_1 is correlated with Y, after controlling all other independent variables.
- For example, did you know that hair length is correlated to height? The taller a person is, the shorter the hair. No, really, this is true!
- But might there be another X that influences this outcome?
- Do you think that the partial correlation of height would be significant to hair length after controlling (taking out the effect of) the other variable?
- This is one reason why it is important to try to identify all the Xs and if possible analyse them together.

R^2 Versus Adjusted R^2

- Due to the math behind it, R^2 values almost always increase (but never decrease) when a predictor (X) is added to the model, regardless of whether the predictor is statistically significant.

- The adjusted R^2 value does not simply increase with the addition of more predictors. Each time a factor is added to a model, Adjusted R^2 will decrease UNLESS the factor significantly explains variation in the Y.

- So although we can use the R^2 value in Simple Linear Regression, we need to use the Adjusted R^2 value for Multiple Linear Regression.

Adjusted R^2 Value

$$R^2_{adj} = 1 - \frac{SS_E/(n-k-1)}{SS_T/(n-1)} = 1 - \left(\frac{n-1}{n-k-1}\right)(1-R^2)$$

Where…

***n* = sample size**

***k* = number of predictor (independent) variables used in regression model**

SS_E = Error (or Residual) sum of squares

SS_T = Total sums of squares

Exercise 2: Multiple Linear Regression

- **Big East Bank has 100 call centres. Data was collected from these 100 call centres on the time it takes to answer a customer call once it starts ringing. Your manager suspects that at least one of four factors affects the 'Answer Time'.**
- **These factors are: average volume of calls, average number of staff on duty, the average time spent on answering the question and/or the number of trunk lines available to send calls to other centres if the 'Answer Time' gets over a certain value.**
- **Your manager has asked you, as a Black Belt, to determine the relationship between these factors and answer time.**
- **The available data are in Regr_Ex_2_Multiple.MPJ.**

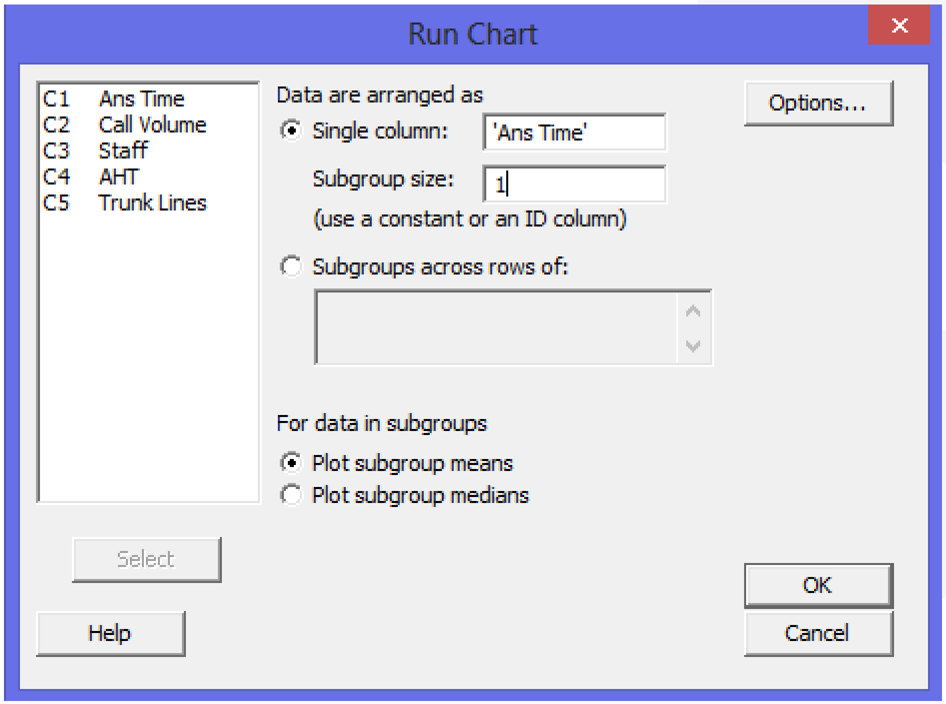
Exercise 2: Multiple Linear Regression

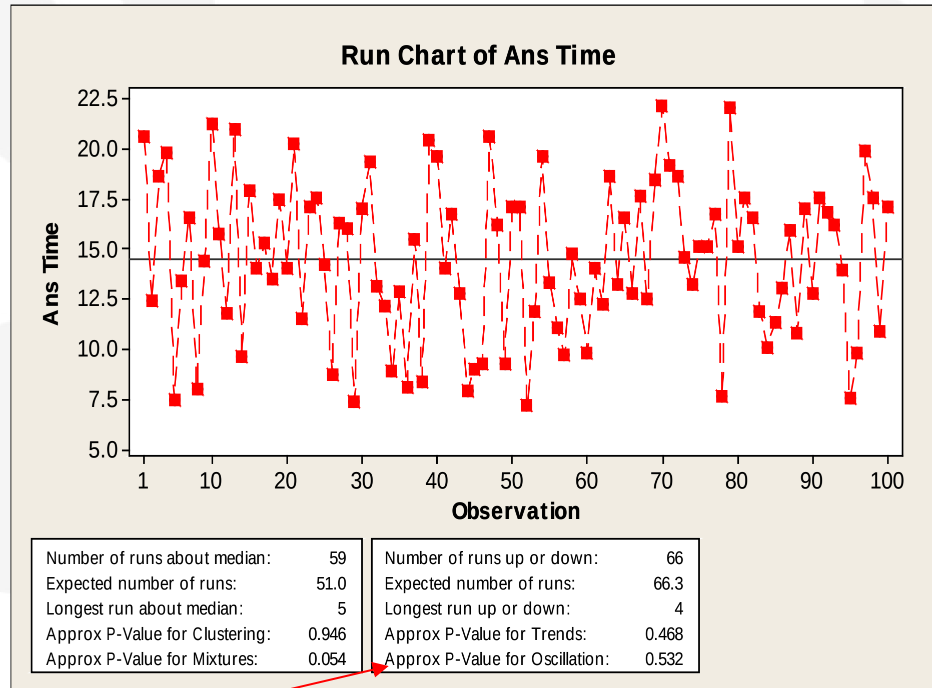
- **Business Question**: Do these inputs affect the output and by how much?
- **Data**: Y and Xs and type of data:
  - **Y: Time to Answer—continuous.**
  - **X_1: Volume of Calls—discrete but plenty of discrimination.**
  - **X_2: Number of staff—discrete but sufficient discrimination.**
  - **X_3: Average Handle Time—continuous.**
  - **X_4: Number of Trunk Lines—discrete with borderline discrimination. Let's assume, for now, it is sufficient and keep an eye on it.**

Exercise 2: Multiple Linear Regression

- Is the response independent?

<u>Stat > Quality Tools > Run Chart</u>





Ho: response = independent
Ha: response is not independent
Fail to reject the null hypothesis (response is independent)

Multicollinearity

- **Multicollinearity occurs when one (or more) of the Xs in a model is correlated to one or more of the other Xs. This can give those Xs more (or less) mathematical influence in the model than they deserve.**

- **Additionally, if we seek to model a process, we would like to do so with the fewest Xs needed. If two are correlated with each other, only one will be needed in the model.**

- **Variance Inflation Factor (VIF) is a measure of one factor acting simultaneously with another and can be found in the regression table**

- **Examples may include:**

  - **The cost of raw materials in £ and $ both being included as Xs in your model**

  - **Number of rooms and square meters of the house both appearing**

Multicollinearity (cont'd)

- **Guidelines for VIFs:**

  =1.0: absolutely no collinearity

  <5.0: little collinearity

  >10.0: major collinearity

- **From 5 to 10: gray area – worth checking out**

- **Take action for large (>10) VIF values. How much greater than 10 is less important than the fact that it is greater than 10.**

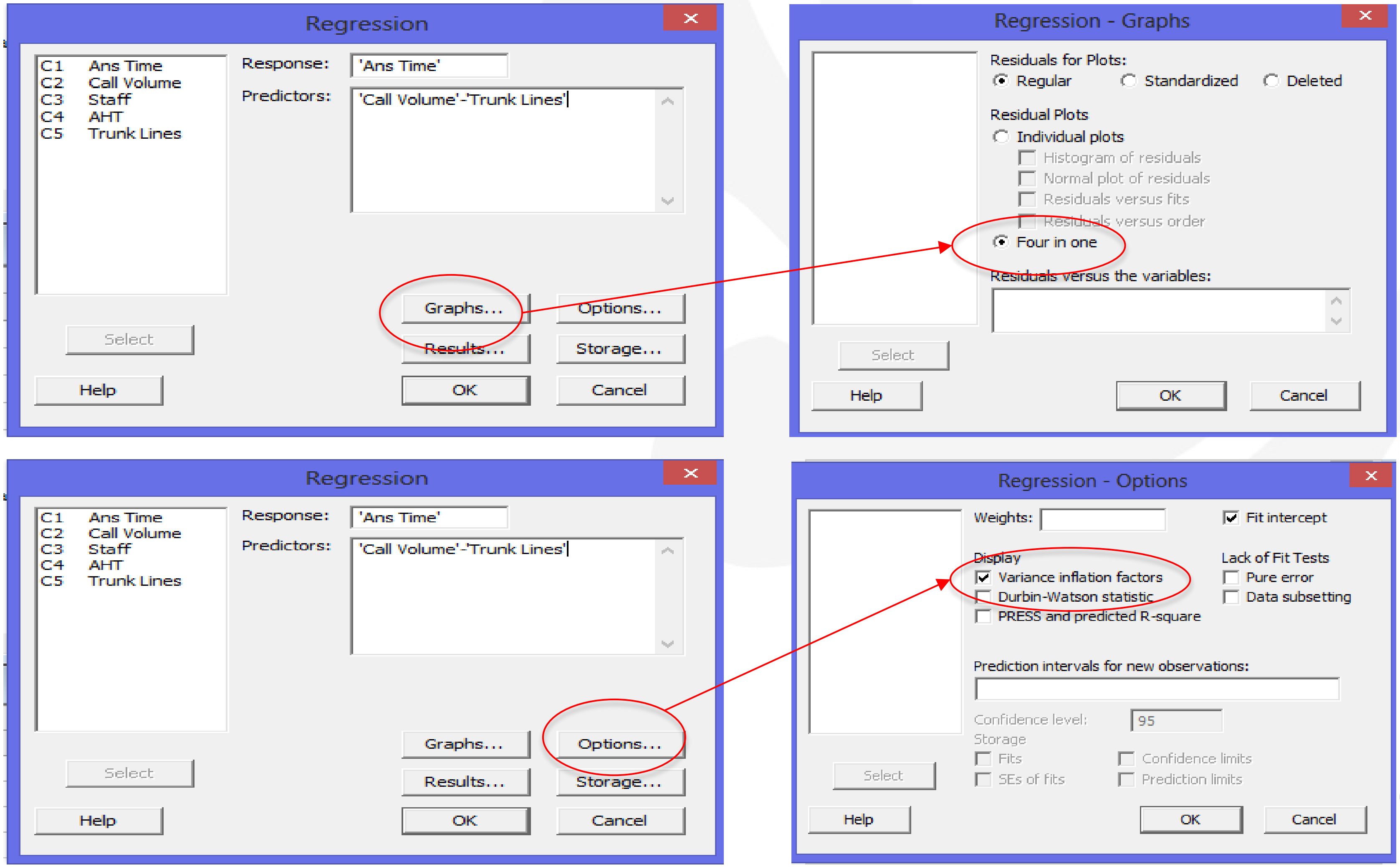
- **What decision do we make on collinearity? Do we need to eliminate one or more Xs?**

So our action is to eliminate one of any pair of Xs that are collinear.

Exercise 2: Create the model

Create model and gathering information on residuals and VIFs

Stat > Regression…

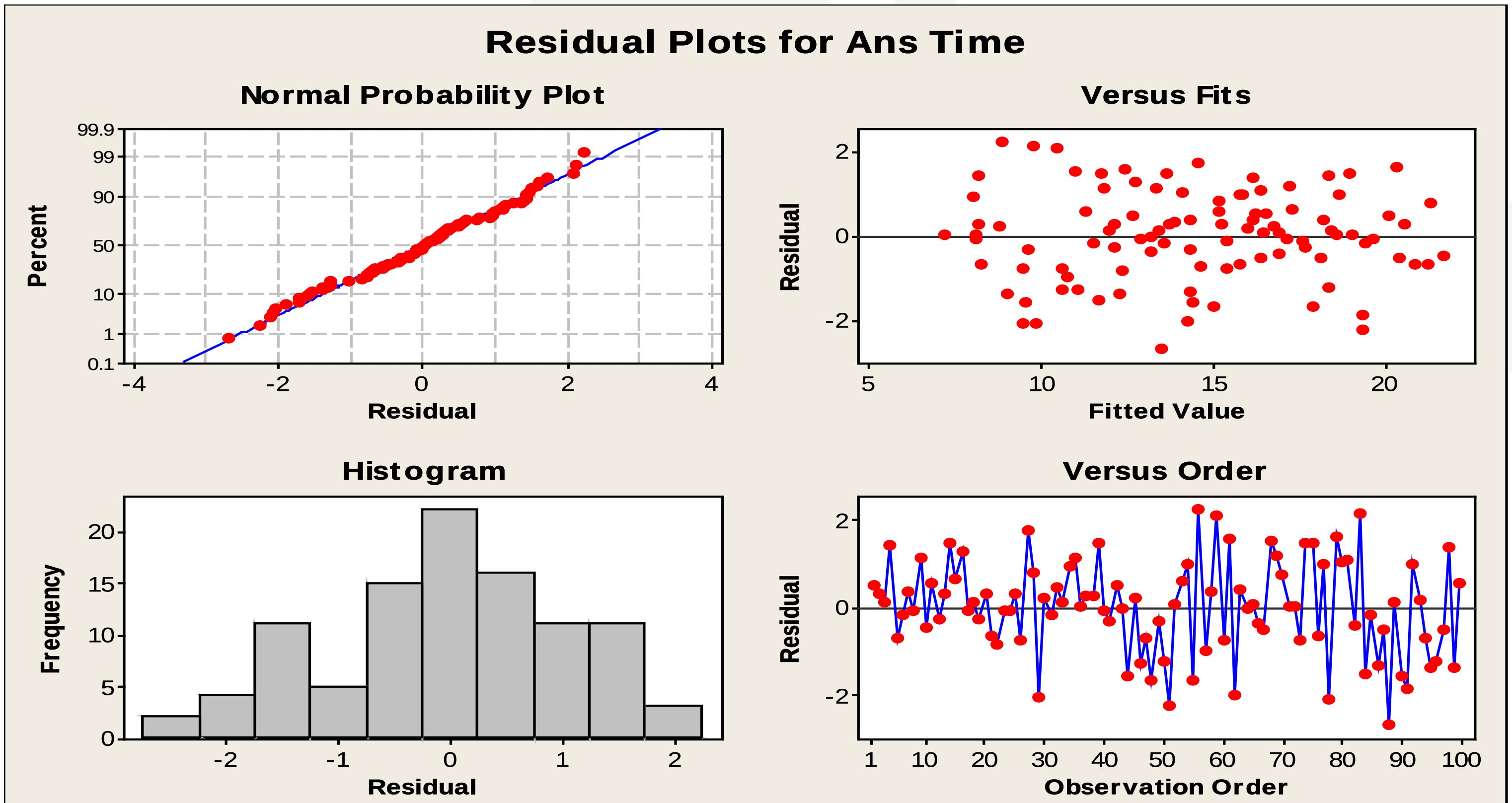


Exercise 2: Residuals analysis

- Are residuals good?

Normally distributed
Equally distributed around zero

Constant Variance
Independent



Exercise 2: Analysis – Is the model significant?

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	1393.31	348.33	293.31	0.000
Residual Error	95	112.82	1.19		
Total	99	1506.13			

- **Based on the p-value, what do we determine?**

What is our next step?

F Ratio - Behind the Scenes

Minitab calculates the F statistic for the regression model and uses the appropriate F table to determine the p-value. This is the same process used in ANOVA; it is the same signal to noise ratio.

$$\boxed{F_0 = \frac{SS_R/k}{SS_E/(n-k-1)} = \frac{MS_R}{MS_E}}$$

If F_0 exceeds $F_{\alpha,\, k,\, n-k-1}$, then we reject H_0. The F_{crit} is looked up in the appropriate F table.

Where...

SS_R = Regression Sums of Squares

k = number of predictor variables

SS_E = Error (or residual) Sums of Squares

n = number of samples

MS_R = Regression mean square

MS_E = Error (or residual) mean square

Exercise 2: Assess Variance Inflation Factor

Regression Analysis: Ans Time versus Call Volume, Staff, ...

```
The regression equation is
Ans Time = - 26.1 + 0.00900 Call Volume + 0.0297 Staff + 0.00028 AHT
           - 0.985 Trunk Lines
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	-26.128	5.369	-4.87	0.000	
Call Volume	0.009002	0.001231	7.31	0.000	6.153
Staff	0.02967	0.02900	1.02	0.309	5.990
AHT	0.000280	0.004097	0.07	0.946	1.032
Trunk Lines	-0.98543	0.04180	-23.57	0.000	1.056

S = 1.08975   R-Sq = 92.5%   R-Sq(adj) = 92.2%

Variance Inflation Factor caused by multicollinearity.

Exercise 2: Address Multicollinearity

Regression Analysis: Ans Time versus Call Volume, Staff, ...

```
The regression equation is
Ans Time = - 26.1 + 0.00900 Call Volume + 0.0297 Staff + 0.00028 AHT
           - 0.985 Trunk Lines
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	-26.128	5.369	-4.87	0.000	
Call Volume	0.009002	0.001231	7.31	0.000	6.153
Staff	0.02967	0.02900	1.02	0.309	5.990
AHT	0.000280	0.004097	0.07	0.946	1.032
Trunk Lines	-0.98543	0.04180	-23.57	0.000	1.056

S = 1.08975 R-Sq = 92.5% R-Sq(adj) = 92.2%

Which pairs are correlated?

Consider eliminating one of a collinear pair of variables based on the answers to the following questions in this order:

1. Is one of them dependent on the other one?
2. Which has the highest p-value?
3. Which has the highest VIF?

What decision do we make here?

Exercise 2: Address Multicollinearity

Re-run regression (without Staff)

Regression Analysis: Ans Time versus Call Volume, AHT, Trunk Lines

```
The regression equation is
Ans Time = - 30.9 + 0.0101 Call Volume + 0.00061 AHT - 0.980
Trunk Lines
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	-30.875	2.703	-11.42	0.000	
Call Volume	0.0101496	0.0005072	20.01	0.000	1.044
AHT	0.000611	0.004086	0.15	0.881	1.025
Trunk Lines	-0.98043	0.04153	-23.61	0.000	1.042

S = 1.09002   R-Sq = 92.4%   R-Sq(adj) = 92.2%

Did this resolve our multicollinearity issues?

What is our next step?

Exercise 2: Reduce Model

Reduce Model - Eliminate Insignificant Variables* (without AHT)

Regression Analysis: Ans Time versus Call Volume, Trunk Lines

The regression equation is
Ans Time = - 30.7 + 0.0101 Call Volume - 0.980 Trunk Lines

Predictor	Coef	SE Coef	T	P	VIF
Constant	-30.746	2.549	-12.06	0.000	
Call Volume	0.0101415	0.0005017	20.21	0.000	1.032
Trunk Lines	-0.97984	0.04113	-23.82	0.000	1.032

S = 1.08451   R-Sq = 92.4%   R-Sq(adj) = 92.3%

Did this resolve our p-value issues?

*Based on p-value: if p-value is > 0.05, eliminate variables one at a time beginning with variable with largest p-value, re-run model, and repeat until all variables are significant

Exercise 2: Answer Business Question

Regression Analysis: Ans Time versus Call Volume, Trunk Lines

The regression equation is
Ans Time = - 30.7 + 0.0101 Call Volume - 0.980 Trunk Lines

Predictor	Coef	SE Coef	T	P	VIF
Constant	-30.746	2.549	-12.06	0.000	
Call Volume	0.0101415	0.0005017	20.21	0.000	1.032
Trunk Lines	-0.97984	0.04113	-23.82	0.000	1.032

S = 1.08451   R-Sq = 92.4%   R-Sq(adj) = 92.3%

There is a relationship between some of the factors identified

Answering time can be estimated using the above regression equation

Exercise 3: Simple vs. Multiple

Simple v Multiple Regression Analysis

Objective: Compare approaches

- One of our team is trying to evaluate whether there is any correlation between the price of homes and the amount of parking space and land on the property. His data is in Regr_Ex_3_Simple_vs_Multiple.MPJ.
- Create the regression model for each of the two Xs individually. Which is significant? What is the RSquare for each?
- What would you conclude?
- Create the regression model with the two Xs together. Which is significant? What is the RSquare Adj?
- What would you conclude now?
- What is your take-away from this exercise?

10 minutes

Exercise 4: Residuals

Residual Analysis

Objective: Validating the model

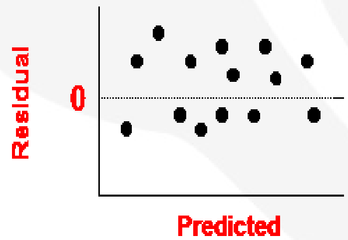
- Open Regr_Ex_4_Residuals.MPJ.
- X1 is random numbers from –10 to +10.
- Y1 is equal to X1 squared.
- Do a regression analysis with residuals.
- What do you see on the Residual by Predicted Value Plot?

- Do the same analysis on X2 vs. Y2.
- What might you conclude by looking at the Regression Plot, p-values, and RSquare? What do you see on the Residual by Predicted Plot?
- What is your take-away from this exercise?
- Is the Y2-X2 regression worthless?

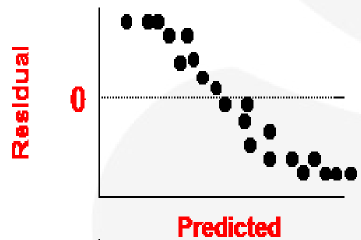
10 minutes

Residual Analysis

Objective: Interpretation of residuals

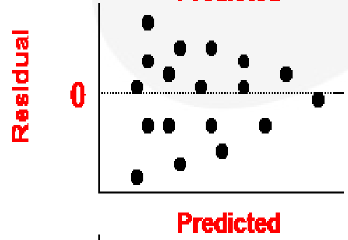


Interpretation?: ____________________

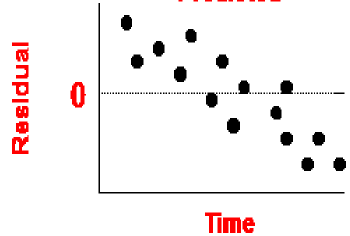


Interpretation?: ____________________

5 minutes



Interpretation?: ____________________



Interpretation?: ____________________

Exercise 6

Case Study

Objective: Practise use of tools end to end

- The sales manager at the Excellent Insurance Company collected data on all of his 100 sales people to see what might be driving the SVA, Shareholder Value Add.
- The Xs he collected data on included Industry Experience, Tenure (at Excellent), Client Time, Energy Level (assigned by the Sales Manager), (new customer) Contact Count, and Coaching Time (by designated coaches).
- Determine:
  - Which, if any, of the Xs significantly impact the SVA?
  - Is the model reliable?
  - What is the predictive equation?
- If you have any particular questions for the Sales Manager.
- The data is in Regr_Ex_6.MPJ.

15 minutes

Summary of key learning points

- **Use linear regression to model one or more continuous Xs with a continuous Y**
- **Linear regression has assumptions on independence of the response and on residuals**
- **The coefficient of determination explains the variability in the Y explained by the factors in the model**
- **Mulicollinearity of factors must be identified and removed from models when**
- **A regression equation is created which can be used to predict the responses**

Recommended coach support points

- If you suspect that there is a model but the relationship between the continuous variables is not linear
- You are unsure how to interpret the residual analysis created by your model

Black Belt

Design of Experiment (DoE)

Objectives of this module

At the end of this module, you will be able to:

- Choose the best type of experiment.
- Plan a designed experiment.
- Set up and run a designed experiment.
- Analyse and understand the results of a designed experiment.
- Understand the risks associated with using screening (fractional factorial) designs.

Questions Lead

- **Basic question: How can I efficiently test my process?**
  - **How can I identify the vital few Xs?**
  - **How can I prove causality?**
  - **How can I determine the best settings for the Xs?**
  - **How can I stress test my new design to see how it performs as the Xs vary?**
  - **I suspect that there is an interaction between two of my Xs. Is there a way to test for that?**

What is a DOE?

- Definition:
  - An organised way in which one changes one or more input variables (Xs) to see if any one of them, or any combination of them, affects the output (Y) in a significant way.
- Uses:
  - To screen out unimportant Xs from the important ones.
  - To prove causation by tying changes in Y directly to changes in the Xs. Seeks to minimise the effect of noise on the output, Y.
  - To discover interactions among the Xs, where one X has a multiplying effect on another X.
  - To develop a prediction equation that will allow us to manage our process at the optimal level.

Types of DOE Models

- Characterisation DOE (Full Factorial): Used to evaluate the main factors and interactions and if needed can provide a prediction equation. This strategy allows the team to understand the true functional relationship between the significant factors and the process response.
- Screening DOE (Fractional Factorial): Used to identify the main factors (Xs), and *generally ignores the interaction effects*. This strategy allows the team to reduce that list of potential X factors down to the most important ones. Typically used in the Analyse Phase.
- Confirming DOE: Used to ensure that the prediction equation matches reality.

DOE Terminology

- Do you remember Y = f (x)? In designed experiments Y = f (Factor A, Factor B…)
- Factors: Xs
- Levels (a.k.a. Settings): The selected values of the factors that we are going to test.
- Coded Values: A way of expressing your experimental settings using the following:
  - -1 represents the lowest or current level.
  - +1 represents the highest level or new level.
  - Where there is no low or high, it is purely arbitrary.
- Example:

Factor	Factor Name	Level	Coded Value
A	Method of Processing	• One piece • Batch	-1 +1
B	Team Size	• 20 • 40	-1 +1

DOE Terminology (cont.)

- **Standard Order:** A standard way to manually set up the run combinations of an experiment and not miss any.
- **Randomised Run:** A randomisation of the standard order that is used when actually running the experiment. It helps to minimise the effects of uncontrolled factors (i.e.. noise) variables.
- **Treatment:** Any one of many combinations of factors and levels; a.k.a. Treatment Combination.
- **Response:** The output variable, Y, that is measured during a run of the experiment.

2 Factors

1 Response

Run	A	B	Y
1	-1	-1	3.0
2	1	-1	8.0
3	-1	1	5.0
4	1	1	12.0

2 Levels

4 Treatments

Standard Order

The design matrix for factorial design with three factors and two levels is shown here in standard order.

Standard Order	X_1	X_2	X_3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

Note that each column is unique and sums to 0. Also for each level (+ or -) of an X, the levels of the other columns have equal numbers of +'s and –'s and therefore sum to zero. We say this design is orthogonal and balanced.

Effects Defined

- **Main Effects:** The impact of the individual factors on the response (Y).
- **Interaction:** The impact of a factor on the response (Y) may depend on the level of some other factor.

Factors Interaction

Run	A	B	A*B	Y
1	-1	-1	1	3.0
2	1	-1	-1	8.0
3	-1	1	-1	5.0
4	1	1	1	12.0

Main Effects

DOEs are analysed by looking at the effects that the significant factors have on the Y response.

Calculating the Main Effect for factor B:

Run	A	B	A*B	Y
1	-1	-1	1	3.0
2	1	-1	-1	8.0
3	-1	1	-1	5.0
4	1	1	1	12.0

Average of the +1 outputs – Average of the -1 outputs

Main Effect for B = [(5.0 + 12.0) / 2] – [(3.0 + 8.0) / 2]

= 8.5 – 5.5

= 3.0

When B changes from -1 to +1, the average response goes up by 3.0.

Interaction Effects

Calculating the Interaction Effect for factor combination A*B:

Run	A	B	A*B	Y
1	-1	-1	1	3.0
2	1	-1	-1	8.0
3	-1	1	-1	5.0
4	1	1	1	12.0

Average of the +1 outputs –
Average of the -1 outputs

Interaction A* B = [(3.0 + 12.0) / 2] – [(8.0 + 5.0) / 2]

= 7.5 – 6.5

= 1.0

When the interaction term changes from -1 to +1, the average response goes up by 1.0.

Full Factorial Terminology

- Full Factorial experiments are those in which all the combinations of the factors are represented in the design matrix. The example below is a two level, two factor design.
- A common type of experiment is called a "two to the k", 2^k.
- 2^k factorials refer to **k factors** each with 2 levels.
- A 2^2 factorial is a 2(factors) x 2(levels) factorial. This design can be done in 2*2 or 4 runs.
- A 2^3 factorial has 3 Xs, each with two levels. This would take 8 runs.
- A 3^4 factorial has 4 Xs, each with three levels. This DOE would have 3*3*3*3 = 81 runs

Run	A	B	A*B	Y
1	-1	-1	1	3.0
2	1	-1	-1	8.0
3	-1	1	-1	5.0
4	1	1	1	12.0

DOE Steps

- **Step 1: Define the experimental objective.**
  - What is the goal of the experiment?
  - What are the possible influential factors?
  - Do you need to reduce variation, shift the mean or both?
  - How much change is required for practical importance?
  - Are there multiple responses? Which are more important?
- **Step 2: Select the response variable, Y.**
  - Is the response quantitative or qualitative?
  - Is goal to maximise, minimise or hit a target?
  - What is current baseline?
  - Is response variable currently in statistical control?
  - Is measurement variation understood/adequate?
- **Step 3: Choose the Xs to test.**
  - Which inputs are the most likely drivers of the response?
  - What Lean Six Sigma tools have led you to believe this?
  - What other inputs, Xs, are there? How will you handle those during the experiment? What about noise variables?

DOE Steps

- **Step 4: Think about the inference space and choose the levels for the Xs.**
  - Where are the Xs set at now?
  - What levels do you want to know about when you finish?
  - Are there settings that you know create defects?
  - What levels are impractical?
  - Have you utilised the best judgment of the SMEs and Process Owners?
- **Step 5: Select an Experimental Design.**
  - Which type design matches your objective?
  - Which type matches your budget? Can you afford to get all of the information?
- **Step 6: Run the experiment.**
  - Change the levels of the factors and measure the Y output of the process for each treatment combination.

DOE Steps

- **Step 7: Analyse the experiment.**
  - Take a practical look at the data.
  - Fit the model.
  - Reduce the model by taking out insignificant Xs.
  - Analyse the residuals.
  - Look at the prediction profiler and/or display the prediction equation.
- **Step 8: Confirm the results.**
  - Based on the risk, decide what you need to do to confirm the results.

Exercise 1, 2^3 Full Factorial

- A Supply Chain executive has determined there is too much variation in processing times to complete incoming material requests. A suggestion is made to identify potential sources of variation so best practices can be identified and implemented.
- We begin our project passing the Define and Measure tollgates easily. Three of the Xs come to the top of our list as potential vital few Xs: Location, Group and Employee.
- As we go into the Analyse phase our team spends some time arguing about which one or ones is/are the vital few Xs, but we cannot come to agreement. We think about testing data, but find that very little exists. Finally we decide that the best approach is to run an experiment and test it for ourselves.

Exercise 1: Objective

- What is the goal of the experiment?
  - To identify the vital few Xs.
- What are the possible influential factors?
  - From the project so far we have identified Location, Group, and Employee.
- Do you need to reduce variation, shift the mean or both?
  - Reduce the variation. Which Xs are driving variation?
- How much change is required for practical importance?
  - We would like to reduce the variation by 50%.
- Are there multiple responses? Which are more important?
  - The other response we could look at is the mean time to process the requests, but that would be secondary.

Exercise 1 The Y

- Is the response quantitative or qualitative?
  - Quantitative (a.k.a. continuous): hours to process. [Note: Running a DOE with a discrete response is an advanced topic and not covered in this module.]
  - Measurement: Variance.
- Is goal to maximise, minimise or hit a target?
  - Minimise.
- What is current baseline?
  - Baseline from the Measure Phase is a standard deviation of 7.9 hours which is a variance of 62.41 hours.
- Is response variable currently in statistical control?
  - Again from the Measure Phase, we saw a stable process.
- Is measurement variation understood/adequate?
  - We passed the MSA in the Measure Phase.

Exercise 1 The Xs

- Which inputs are the most likely drivers of the Y?
  - Location, Group, Employee
- What Lean Six Sigma tools have led you to believe this?
  - In the Measure Phase we used the Process Map, C&E Diagram and C&E Matrix.
- What other inputs, Xs, are there? How will you handle those during the experiment? What about noise variables?
  - Randomise: Randomise our runs.
  - Control: Control the type of request. The complicated ones will be routed to Employees who are not part of the test.
  - Measure: Measure the volume of requests during the test.
  - Large Sample: Take a large sample to even out the variation between individual requests and use several employees in each treatment to average out the difference between employees.

Exercise 1 Factor Levels

- Where are the Xs set at now?
  - Location: We have both office based and home based employees.
  - Group: We do some in-house and some are out sourced.
  - Employees: We discussed the relative importance of experience.
- What levels do you want to know about when you finish?
  - Location and Group at all levels. We would like to know if first year employees are more variable than those with over one year tenure.
- Are there settings that you know create defects?
  - None that we know of.
- What levels are impractical?
  - employee with less than 6 months are still learning. So we will define "new" employees as those between 6 and 12 months.
- Have you utilised the best judgment of the SMEs and Process Owners?
  - Yes.

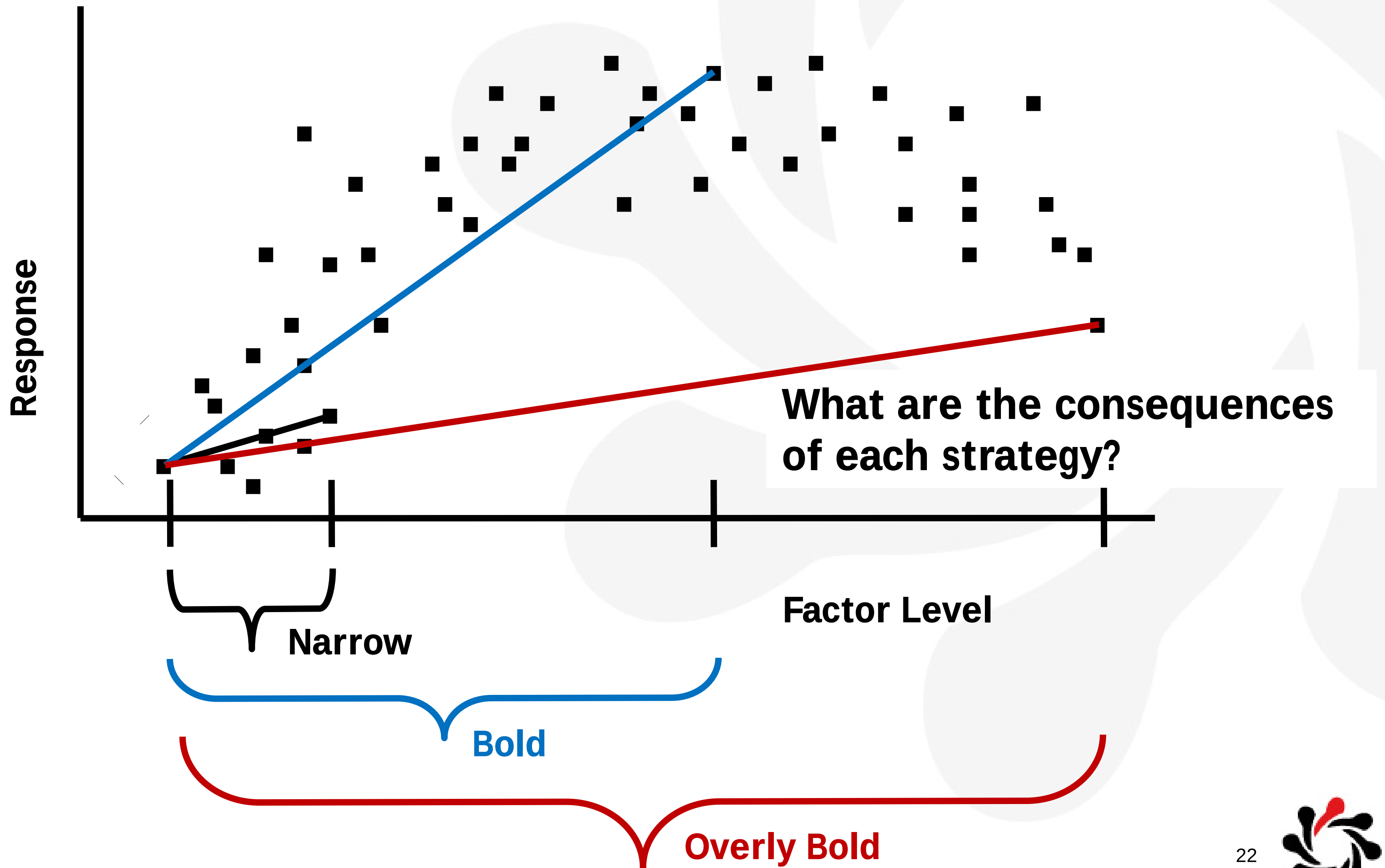
Inference Space

- Inference Space – ability to generalise about conclusions based on results of your experiment. Conclusions may be broad or narrow based on experimental strategy.

Narrow	**Broad**
1 Location	**All locations**
1 day	**Many months**
1 operator	**All operators**
1 system	**Multiple systems**

- What happens if we choose too narrowly?
- What happens if we choose too broadly?

Level Selection



Exercise 1: The Design

- Which type design matches your objective?
  - Screening: Which Xs are important? Is there a causal relationship?
  - Characterisation: How much do the Xs affect the Y? Are the Xs interacting with each other?
  - Optimisation: What is the absolute peak performance I can get from this process?
  - Confirming: Due to the risk, do I need to confirm the results?
- Which type matches your budget? Can you afford to get all of the information?
  - DOEs can be expensive to run. Screening, Characterisation, and Optimisation increase costs in that order, but also increase the information you get in the same order.
- Our basic objective in this experiment is Screening which many times will dictate a fractional factorial design. But with only three factors at two levels, a full factorial would be eight runs. Our budget is sufficient for eight runs.

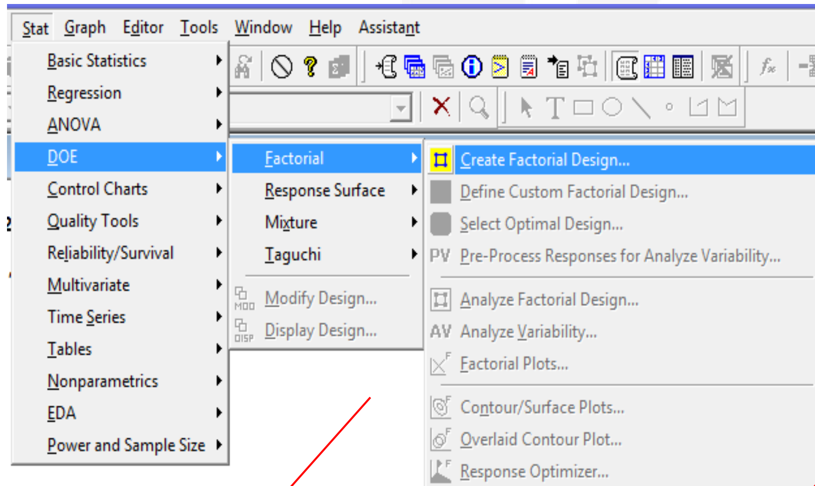
Exercise 1: The Design Process

We will be using Minitab for creating and analysing DOEs, starting with this example. The process steps are below:

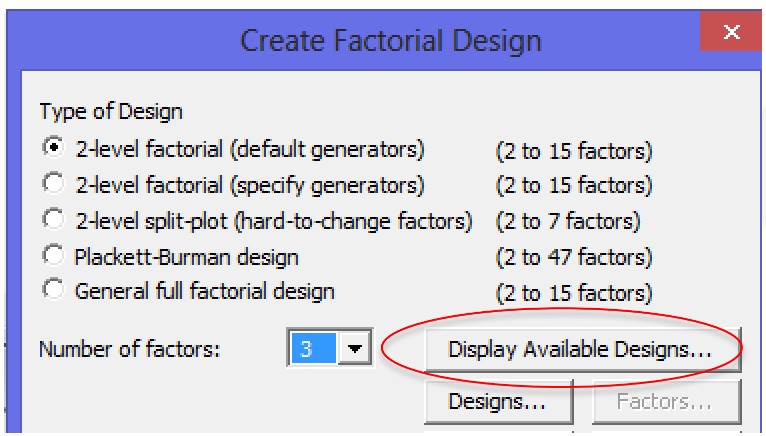
Enter response & factors → **Describe the DOE model** → **Select the number of runs** → **Check the design diagnostics** → **Specify output options** → **Make the design table**

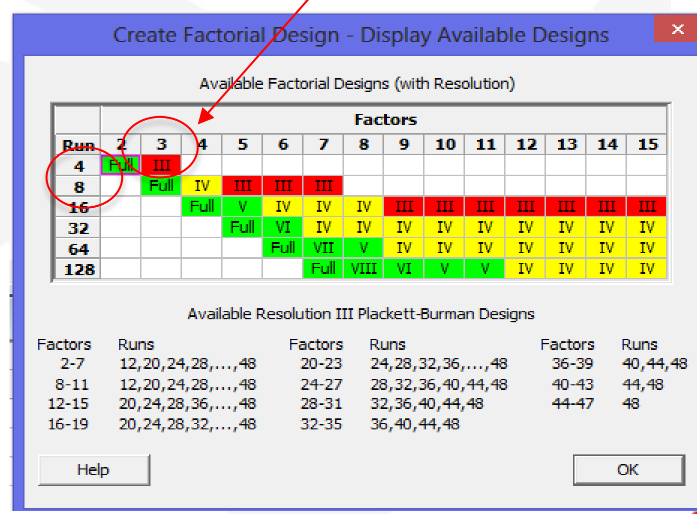
Exercise 1 Use Minitab to design the experiment

Stat > DOE > Factorial > Create Factorial Design



A full factorial design with 3 factors would be 8 runs

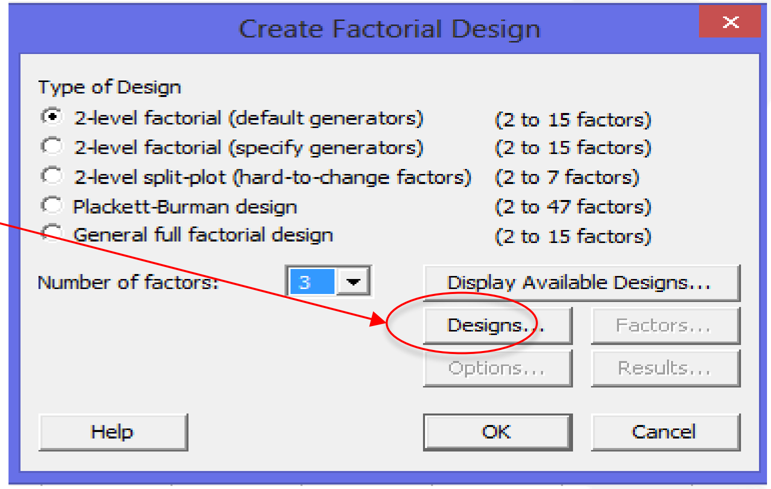




Select 3 factors (ie location, group, employees) then select display available designs

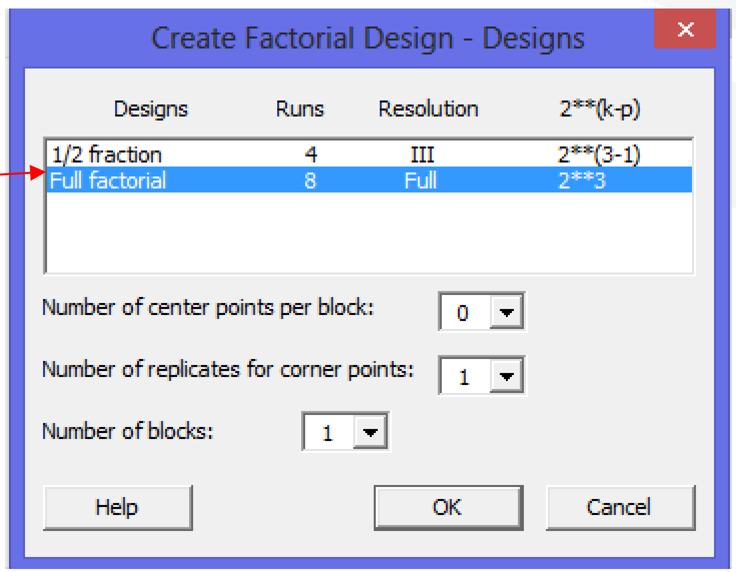
Exercise 1 Use Minitab to design the experiment

Select designs



Select Full Factorial

(we'll cover fractional factorials later)



Other design considerations

- Centre points
- Replicates
- Blocking

Repetition

- Repeats are simply setting up a treatment and **taking several measurements** before going on to the next run.
  - Repeats do not add additional runs to the experiment.
  - Necessary when your response measurement requires multiple values for calculation such as variance or mean.
  - Necessary when one measurement is not representative of the actual process. E.g., AHT in a call centre. There is so much variation between calls, that we will have to have the average of a large number of calls to see if a particular treatment combination affects the response.

- Degrees of Freedom
  - The extra data points do not increase the degrees of freedom because only one measurement (such as the average or standard deviation) of the treatment is used in the analysis.

Replication

- Replications are **duplicating** each treatment a number of times randomly inside the experiment.
  - Replicates do increase the number of experimental runs.
  - E.g., we set up [Office-In-New] and get data. Then we go on to another treatment. Later in the experiment we set up [Office-In-New] again and get another data point. Each treatment is run the same number of times depending on the number of replicates.
  - This differs from repetition in that the treatment is set up again, **giving the noise factors a chance to vary**.

- Degrees of Freedom
  - The extra data points here **add** extra degrees of freedom.
  - The extra degrees of freedom allow us to **detect smaller changes in the response** with the same confidence; the experiment has **more Power** to detect differences.

Blocking

- Blocking is **running different levels of a factor together instead of randomly** throughout the experiment.
  - For a key input, it might be one that is very difficult or expensive to change.

  - E.g., one of the factors may be running a large ad campaign. Due to the expense, we can only run it once and after it is run once it would have carryover effect to any runs afterward. So we run the first half of the DOE without the ad campaign, then run the ad campaign and the second half of the design.

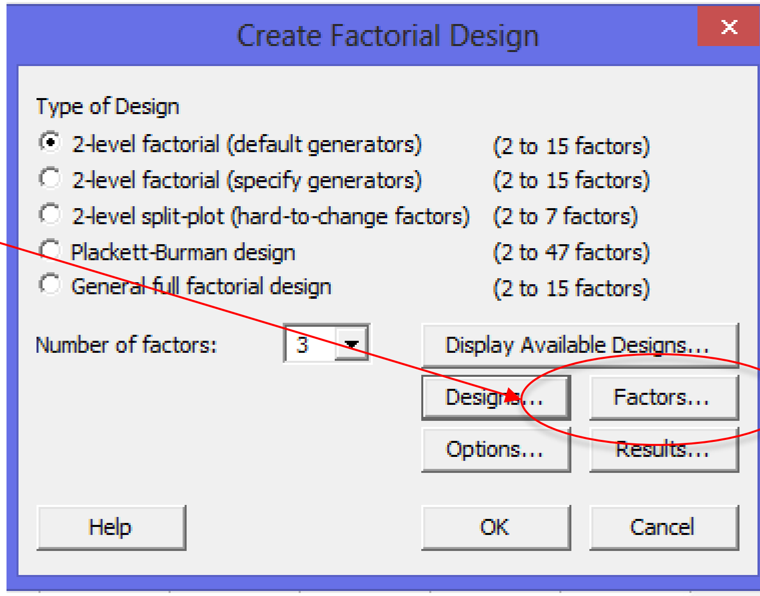
  - For a noise input, it might be that there is more variation caused by noise than by the factors we are testing. This happens many times when an experiment takes several days or weeks. So we can block the experiment into time periods when the noise variation is smaller. Minitab will then take that into account in the analysis.

Centre Points

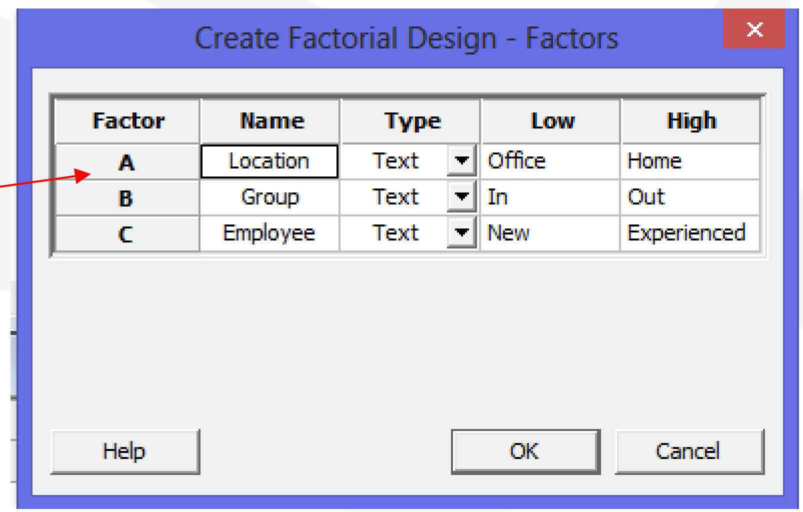
- When Designed Experiments are analysed, the assumption is made that the response to a change in a factor is **linear**, i.e. a straight line.
- **Centre Points check this assumption**. If you suspect that you may not have a linear response, you should get some coaching from your MBB.
  - A Centre Point is another run of the DOE where the factors are set exactly half way between the –1 and +1 levels. **It is coded as "0".**
  - Most of the time when the factors are discrete, this becomes impossible. There are some advance ways to check the centre point with discrete and continuous factors.

Exercise 1 Use Minitab to design the experiment

Select Factors



Type in the settings and labels for factors



Exercise 1 Use Minitab to design the experiment

Session Window Output

Full Factorial Design

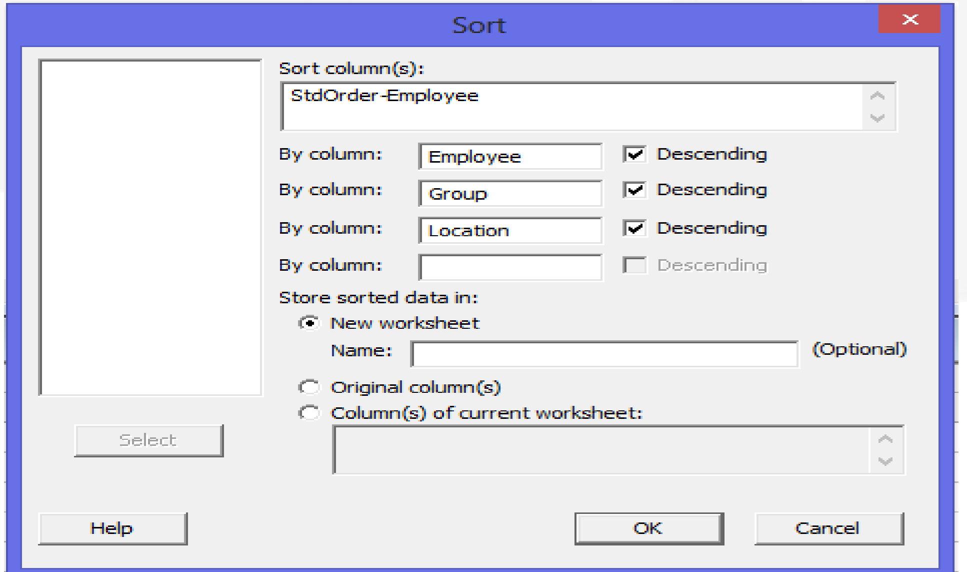
```
Factors:    3    Base Design:           3, 8
Runs:       8    Replicates:               1
Blocks:     1    Centre pts (total):       0

All terms are free from aliasing.
```

Note your worksheet order may be different as Minitab has randomised the standard order

(**FOR CLASSROOM PURPOSES ONLY!**)
Choose the Run Order as "Sort Right to Left" and click descending.
This will keep us together as a class.

Worksheet Output

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T
	StdOrder	RunOrder	CenterPt	Blocks	Location	Group	Employee
1	8	1	1	1	Home	Out	Experienced
2	4	2	1	1	Home	Out	New
3	6	3	1	1	Home	In	Experienced
4	2	4	1	1	Home	In	New
5	7	5	1	1	Office	Out	Experienced
6	1	6	1	1	Office	In	New
7	5	7	1	1	Office	In	Experienced
8	3	8	1	1	Office	Out	New
9							



Exercise 1 Design Table

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T	C8
	StdOrder	RunOrder	CenterPt	Blocks	Location	Group	Employee	
1	8	1	1	1	Home	Out	Experienced	
2	7	5	1	1	Office	Out	Experienced	
3	6	3	1	1	Home	In	Experienced	
4	5	7	1	1	Office	In	Experienced	
5	4	2	1	1	Home	Out	New	
6	3	8	1	1	Office	Out	New	
7	2	4	1	1	Home	In	New	
8	1	6	1	1	Office	In	New	
9								

Now we are ready to run the actual experiment and record the results here.
Would we run it in this order?
Describe how this experiment might be run in the business.

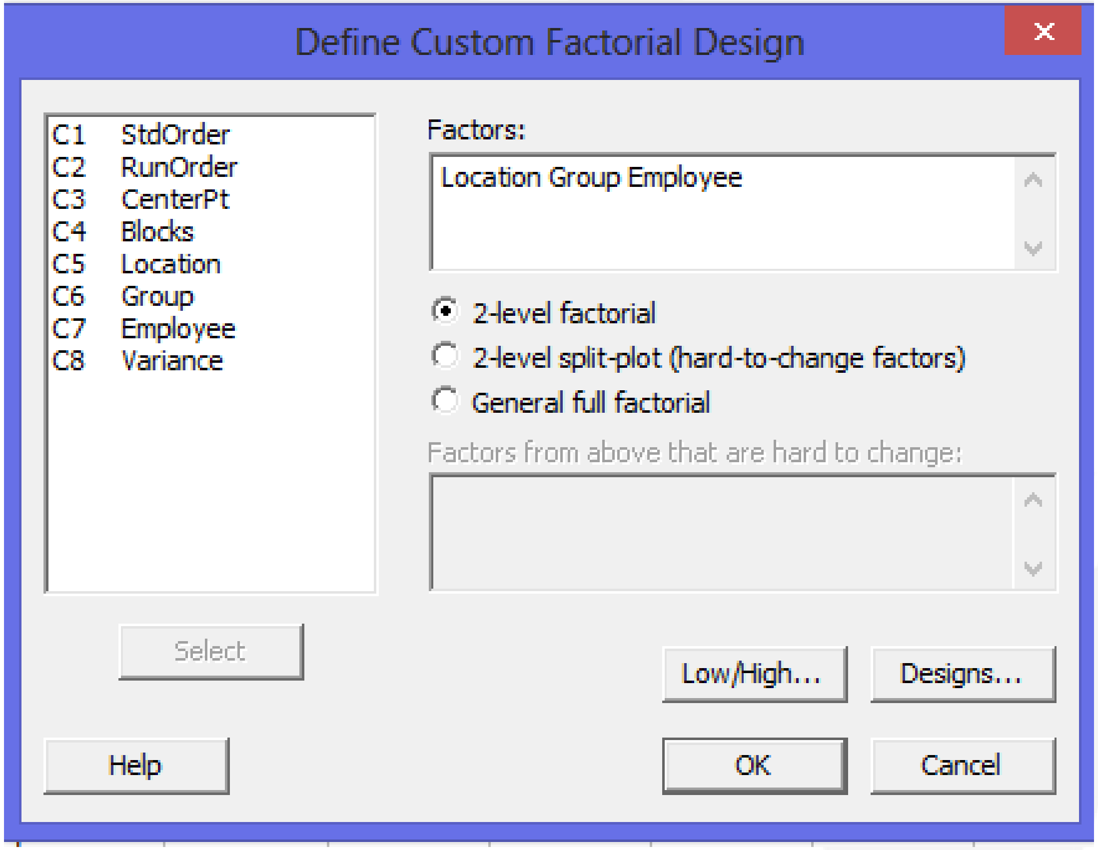
Exercise 1 Gather results

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T	C8
	StdOrder	RunOrder	CenterPt	Blocks	Location	Group	Employee	Variance
1	8	1	1	1	Home	Out	Experienced	22.24
2	7	5	1	1	Office	Out	Experienced	11.00
3	6	3	1	1	Home	In	Experienced	18.36
4	5	7	1	1	Office	In	Experienced	14.20
5	4	2	1	1	Home	Out	New	69.72
6	3	8	1	1	Office	Out	New	42.64
7	2	4	1	1	Home	In	New	58.20
8	1	6	1	1	Office	In	New	41.88
9								

After the experiment is run, enter the data into the Minitab data table.

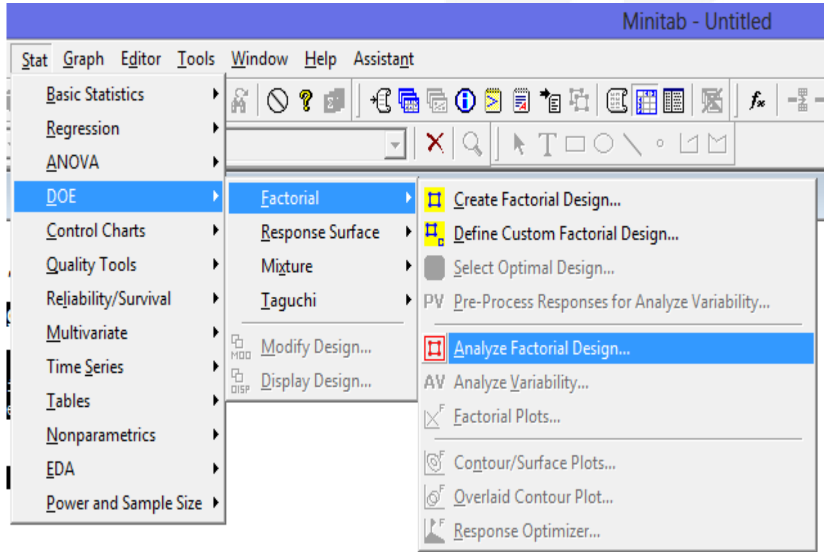
Exercise 1 Define Factorial Design

Stat > DOE > Factorial > Define Factorial design

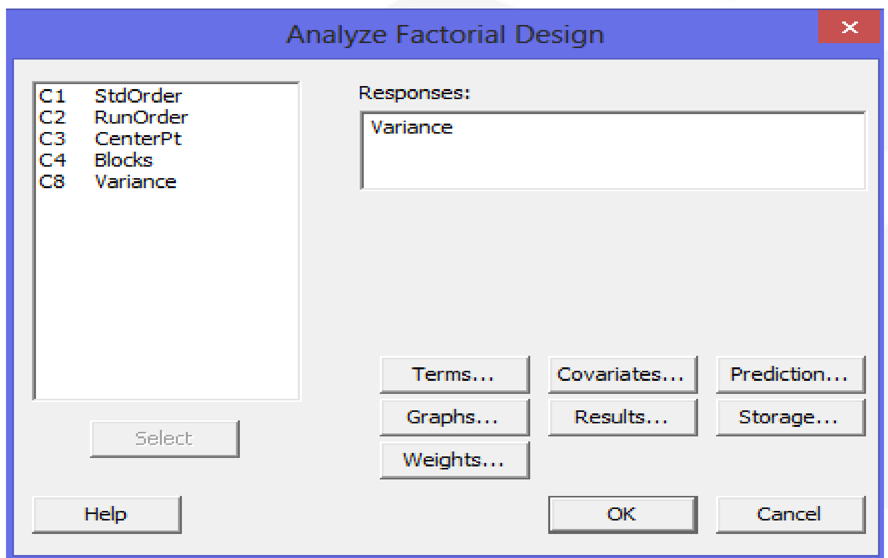


Exercise 1 Analyse results

Stat > DOE > Factorial > Analyse Factorial design

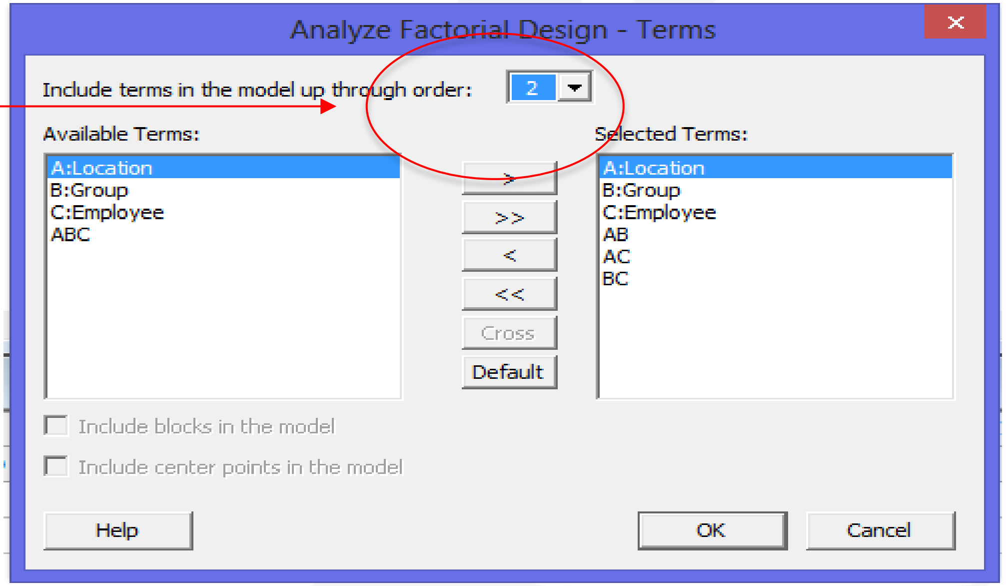


In Response add “Variance”



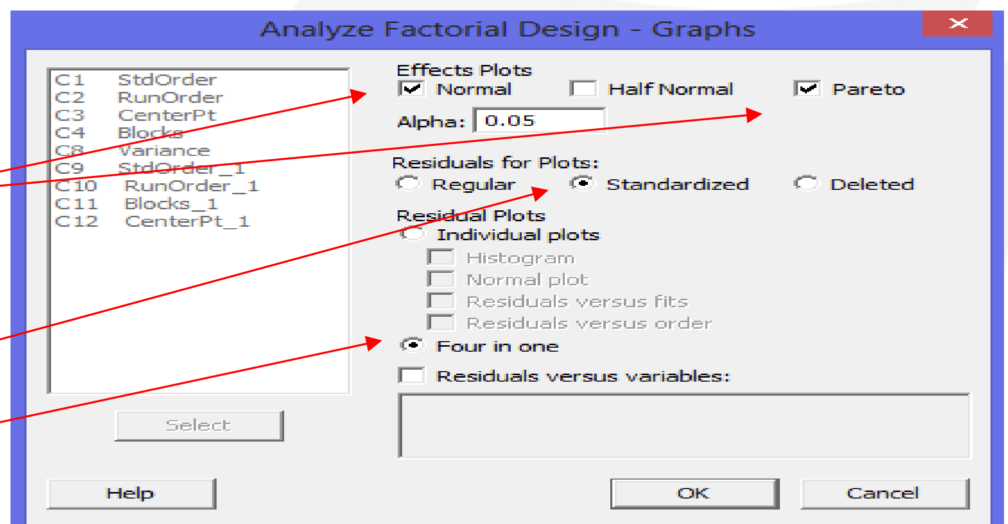
Exercise 1 Analyse results

Select “Terms” and include up to 2 factor interactions



We’re only going to look at two way interactions due to our small sample size (and limited degrees of freedom)

Select “Graphs and include Effects plots: Normal Pareto with Alpha set to 0.05



Residuals for plots: Standardised

Residuals plots: Four in one

Exercise 1 Evaluation

Ho: No factors/interactions have a significant effect on the response

Ha: At least one factor/interaction has an effect on the response

Session window output:

Analysis of Variance for Variance (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	3141.09	3141.09	1047.03	618.52	0.030
Location	1	432.18	432.18	432.18	255.30	0.040
Group	1	21.00	21.00	21.00	12.40	0.176
Employee	1	2687.91	2687.91	2687.91	1587.85	0.016
2-Way Interactions	3	154.60	154.60	51.53	30.44	0.132
Location*Group	1	39.78	39.78	39.78	23.50	0.130
Location*Employee	1	98.00	98.00	98.00	57.89	0.083
Group*Employee	1	16.82	16.82	16.82	9.94	0.196
Residual Error	1	1.69	1.69	1.69		
Total	7	3297.38				

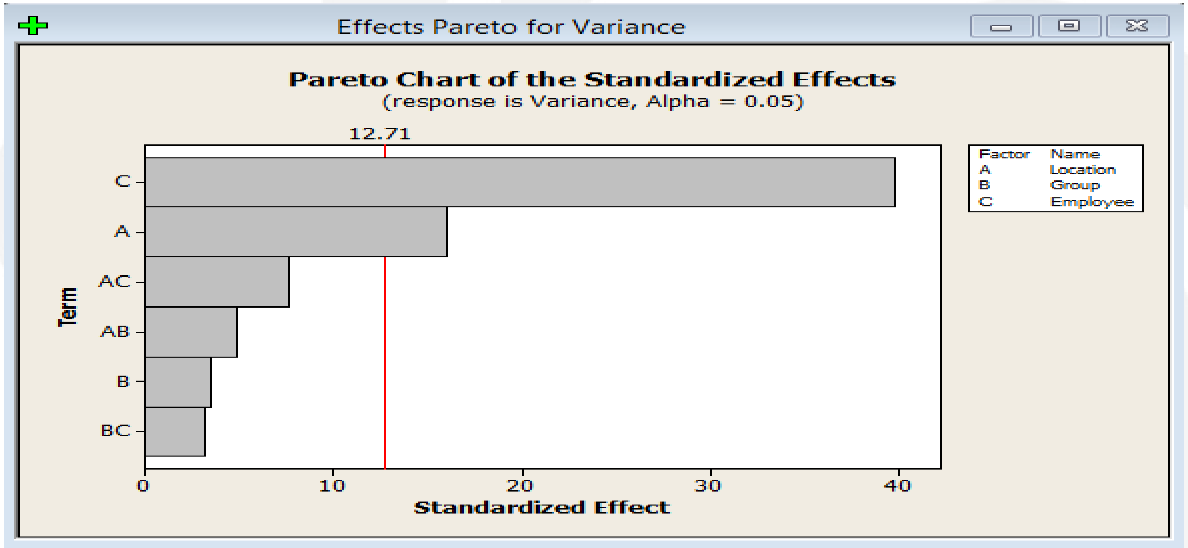
P-values for main effects < 0.05 so some factors (Location and Employee) significant

The interaction of any of the factors does not have a significant effect

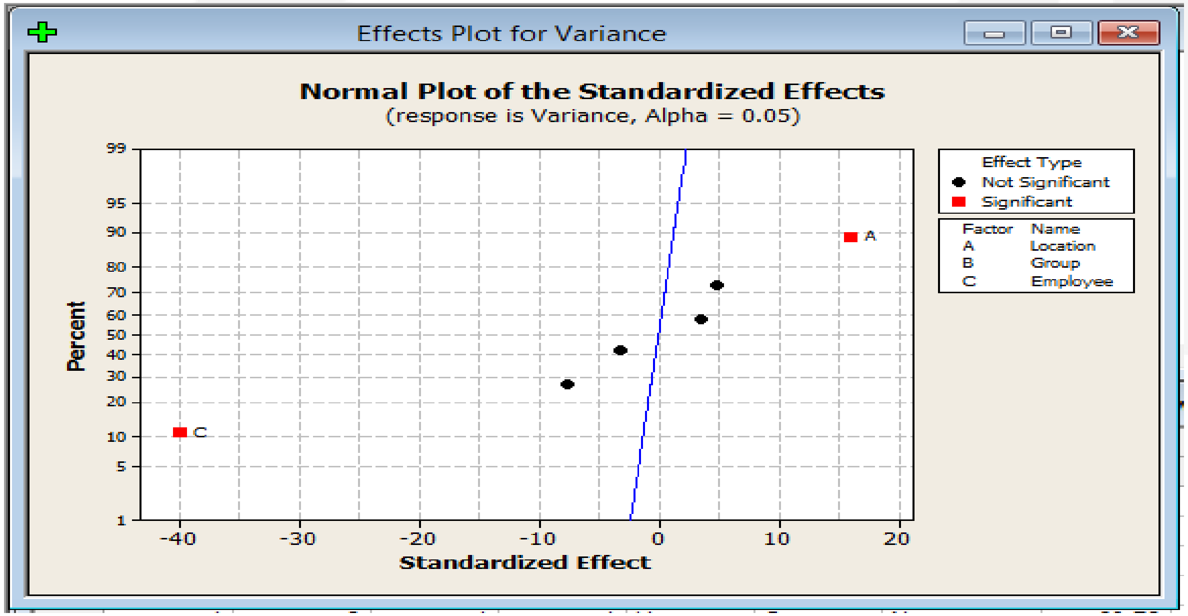
Exercise 1 Evaluation, Graphically

Factors A & C are beyond the red line indicating they are significant

The red line represents Alpha as a standardised value



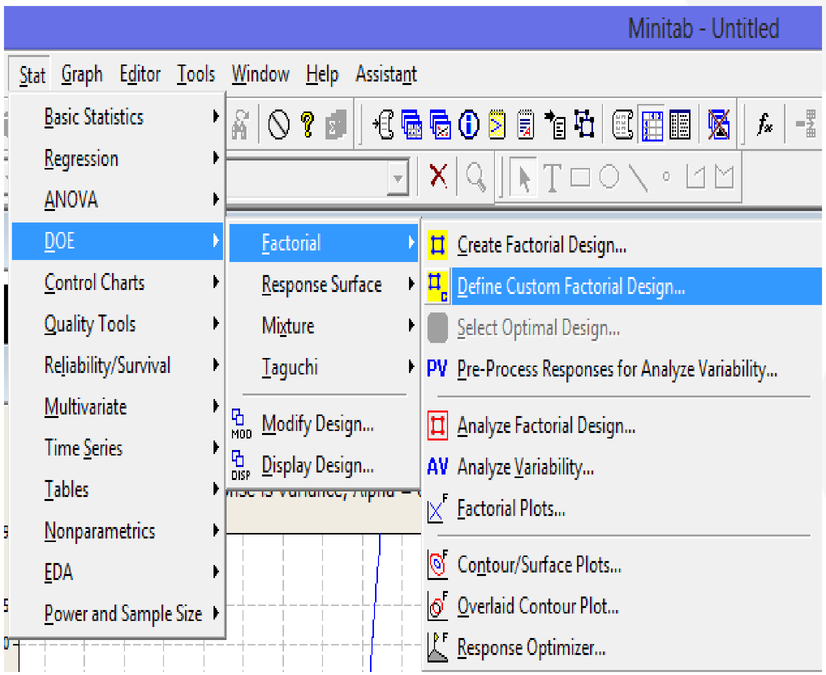
Another way to show it, significant factors highlighted in red

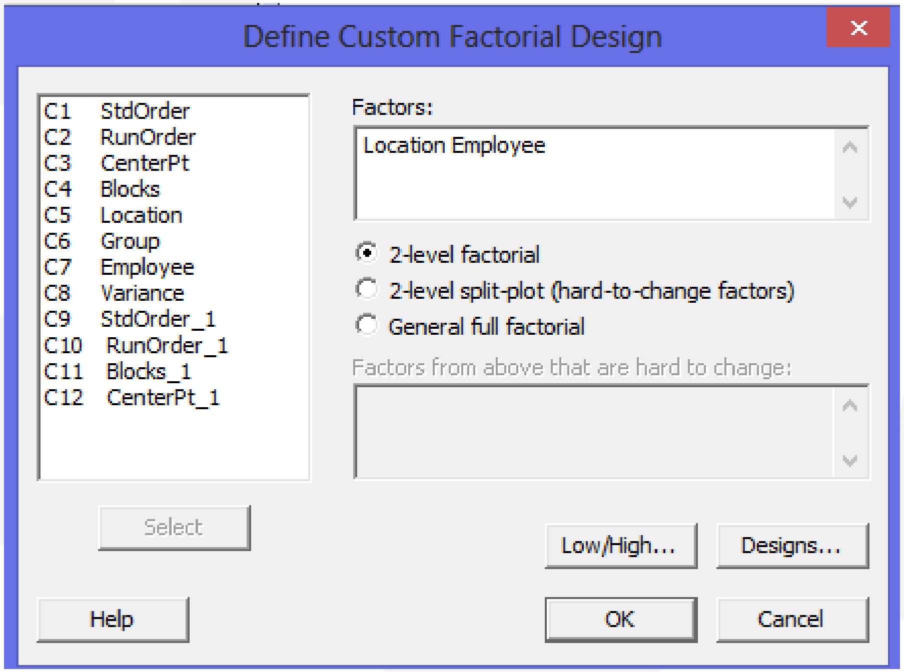


Exercise 1: Reduced Model

- Given we have now screened factors and their interactions, we can reduce our model to improve its accuracy

Redefine the factorial design

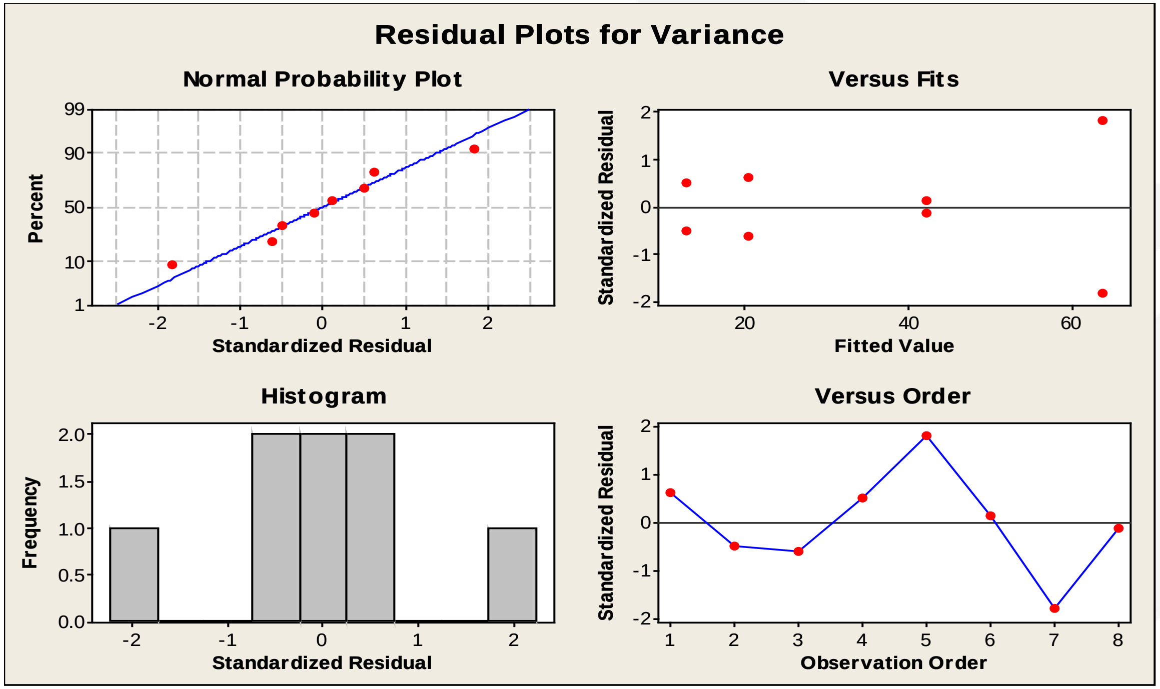




Now that we have a reduced model, what is our next step?

Exercise 1: Residual Analysis

- Residual Plots
  - Are residuals normal with a mean of zero?
  - Do residuals appear to be independent?
  - Do the variances of the residuals appear to be equal across the range of predicted values?
  - What is the overall conclusion of the Residual Analysis?

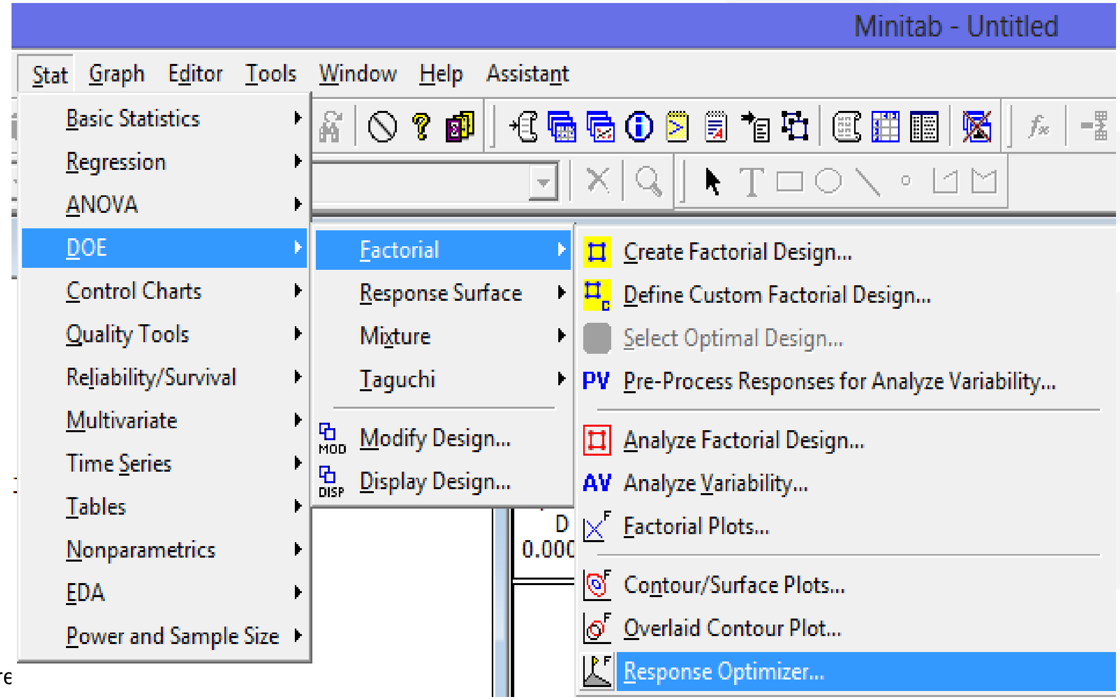


Exercise 1:Optimizer

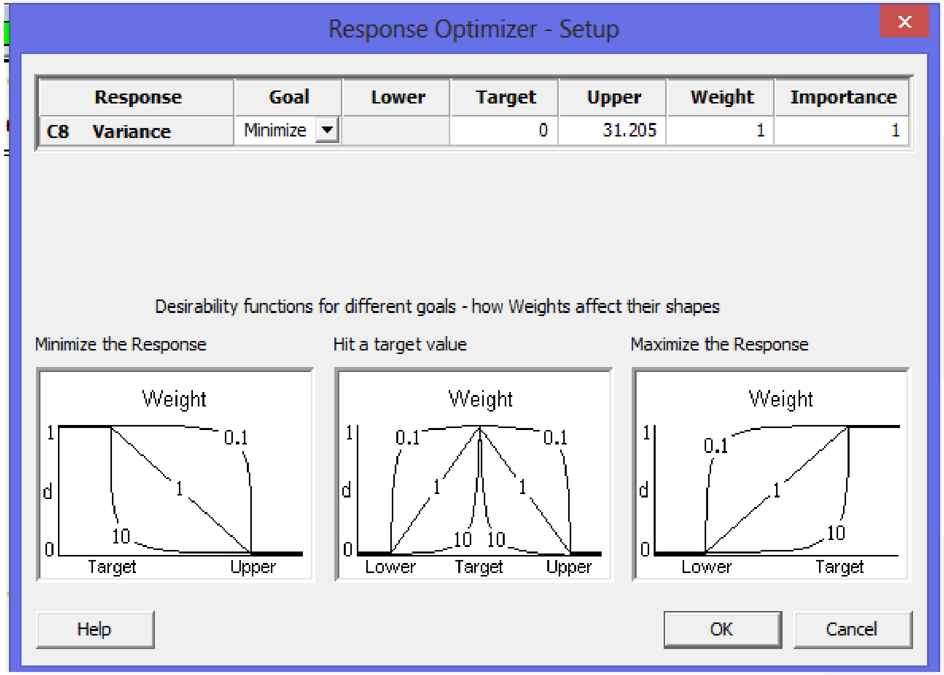
- Now that we have reduced the model to significant factors (and interactions) we can use the Minitab Response Optimizer tool to model our settings to achieve the best result

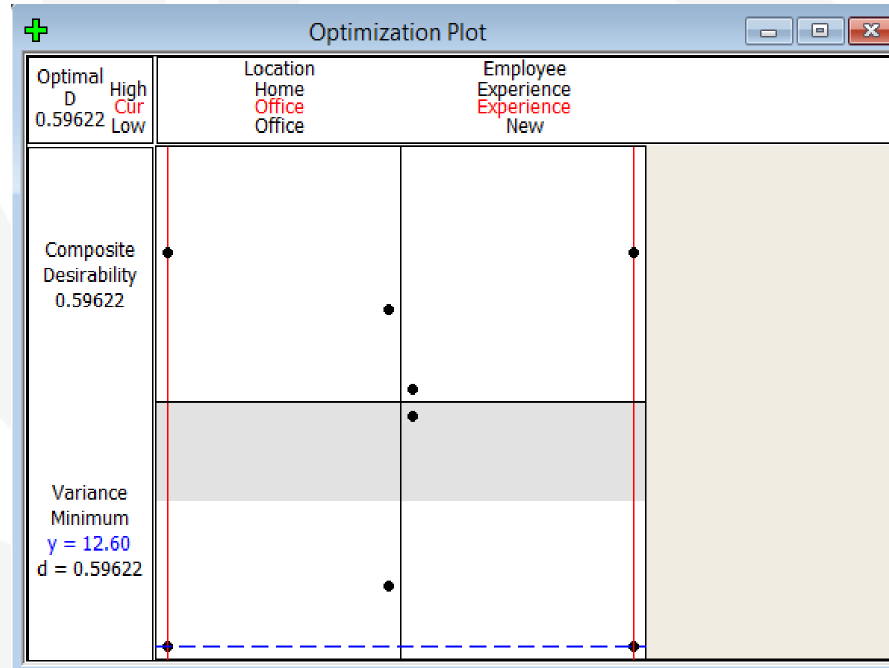
- Remember our objective was to minimise variation

Stat > DOE > Factorial > Response Optimizer



Exercise 1:Optimizer





- Move the red lines on the optimizer tool to achieve the lowest variation.
- Setting factors to levels:
  - Location = Office
  - Employee = Experience

Creates a minimum variation value of 12.6

Exercise 1: Confirmation

- If we were to run our experiment next month, would we get the same results?
- Can this model be applied in similar conditions? Will it be an easy roll-out? Do you need to test other locations, times, associates, or regions?
- Run another DOE or additional tests to confirm the results.
  - Optional depending on risk.
  - More important when the initial experiment was a fractional factorial DOE.

DOE Exercise 2

45 minutes

Present back your results!

We want to increase the acceptance rate of our new credit card mailings. The project has been ongoing and has reached the point where a DOE was run to test four factors. For this exercise, assume that steps 1 through 6 were done correctly. Open the data set DOE_Full_Factorial_Ex2.MPJ and analyse the data that were collected.

When we send out the real offer, where should we set the factors and what acceptance should we expect?

Factor 1 was whether the mailing was sent in white envelope (-1) or pastel (+1).

Factor 2 was whether it was mailed by bulk mail (-1) or first class (+1).

Factor 3 was the number of enclosures in the envelope 1 (-1) or 3 (+1).

Factor 4 was the time in days allowed for the customer to respond 5 (-1) or 30 (+1).

The response was the number that accepted the offer out of a 10,000 piece mailing.

Fractional Factorial DOE

Now let's explore the Fractional Factorial DOE to see the advantages and trade-offs.

Why Fractional Factorials?

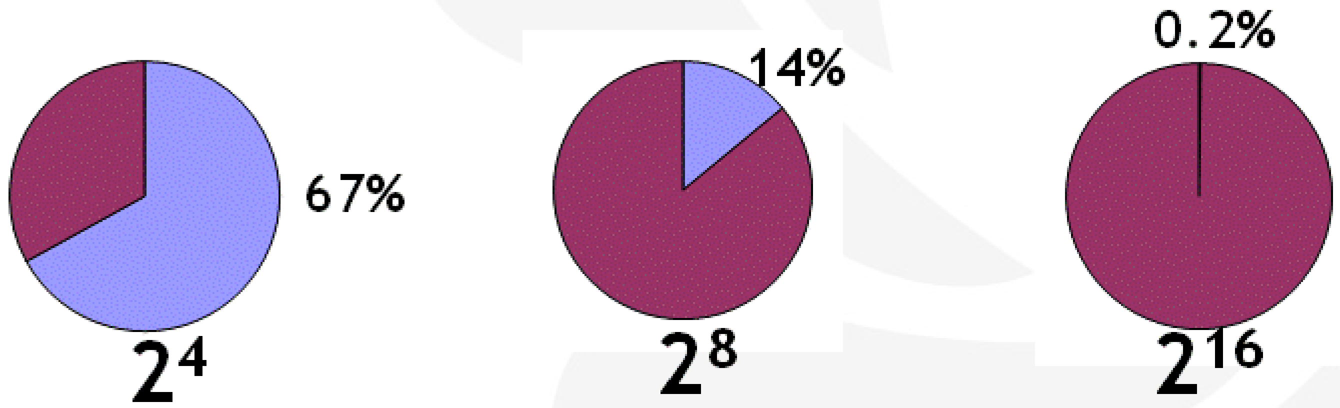
Key content

- You are a manager and one of your Black Belts wants to run a full factorial DOE to investigate which of 15 potential factors are key to improving your process. How would you respond?
  - Assuming you could conduct one treatment each minute, it would take the Black Belt 22 days and 18 hours to complete the full factorial DOE.
  - Even if we restrict the DOE to two way interactions, it would take 128 runs.
- Do you have the resources for this number of runs?

A Fractional Factorial DOE provides a practical alternative to the full factorial experiment.

How much do we want to know?

- What did we say about 3 way and higher interactions?
- So running experiments to characterise these higher order interactions in almost all cases is a waste.



The percent of total effects that are main factors or 2 way interactions (where most of our needed information lies).

When to consider Fractionals?

- Long list of potential vital few Xs.
- Belief that only a small proportion of the Xs have a substantial effect on the response.
- Limited budget for experimenting.
- Limited time that we experiment on a process.
- Fractional Factorials are typically performed early in a project as a means of identifying the vital few Xs while expending limited resources.

Exercise 3: Fractional Factorial

Returning to the Direct Mail example in Exercise 2, we want to increase the acceptance rate of a new credit card direct mail campaign. The project has been ongoing and has reached the point where a DOE was run to test four factors. What would it have looked like if we had run a fractional factorial?

Factor 1 (Envelope Colour) Does the colour of the envelope effect the response to the card offer?

Factor 2 (Mail Type) Is the offer response effected by how the offer was sent?

Factor 3 (# of Enclosures) Does the number of enclosures effect the responses?

Factor 4 (Time to Respond) Does the time allowed for the customer to respond in days have an effect on the acceptance offer?

The direct mail team will measure the number of responses per 10,000 piece mailing as the Y response for the DOE.

Full or Fractional Design?

- Due to budget constraints, your management would like to hold the experiment down to just 8 treatment combinations. Could you still run the DOE and get enough information about the four factors?
- The good news is a fractional factorial design will fit the run criteria. The full factorial design requires sixteen runs, the fractional can be accomplished with eight.
- The bad news is you must choose the correct combinations of runs. Which ones should you choose?

Choice 1: Which rows to delete?

If we set it up as a full factorial with 4 factors, we get 16 runs. Which rows should we delete to get down to our 8 runs?

	A	B	C	D
1	-1	-1	-1	-1
2	+1	-1	-1	-1
3	-1	+1	-1	-1
4	+1	+1	-1	-1
5	-1	-1	+1	-1
6	+1	-1	+1	-1
7	-1	+1	+1	-1
8	+1	+1	+1	-1
9	-1	-1	-1	+1
10	+1	-1	-1	+1
11	-1	+1	-1	+1
12	+1	+1	-1	+1
13	-1	-1	+1	+1
14	+1	-1	+1	+1
15	-1	+1	+1	+1
16	+1	+1	+1	+1

Choice 2: How do we set D?

If we set it up as a full factorial with 3 factors, we get our 8 runs. But how do we set the D factor for each of the runs?

	A	B	C	D
1	-1	-1	-1	?
2	+1	-1	-1	?
3	-1	+1	-1	?
4	+1	+1	-1	?
5	-1	-1	+1	?
6	+1	-1	+1	?
7	-1	+1	+1	?
8	+1	+1	+1	?

How to choose:

Using a 2^3 design, we can achieve the required 8 runs. How do we determine the settings for the fourth factor?

	A	B	C	AB	AC	BC	~~ABC~~ D
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Use the coded values of the 3-way interaction to set the levels of D.

Now we can examine 4 factors in only 8 runs.

Note that mathematically:

D = A*B*C

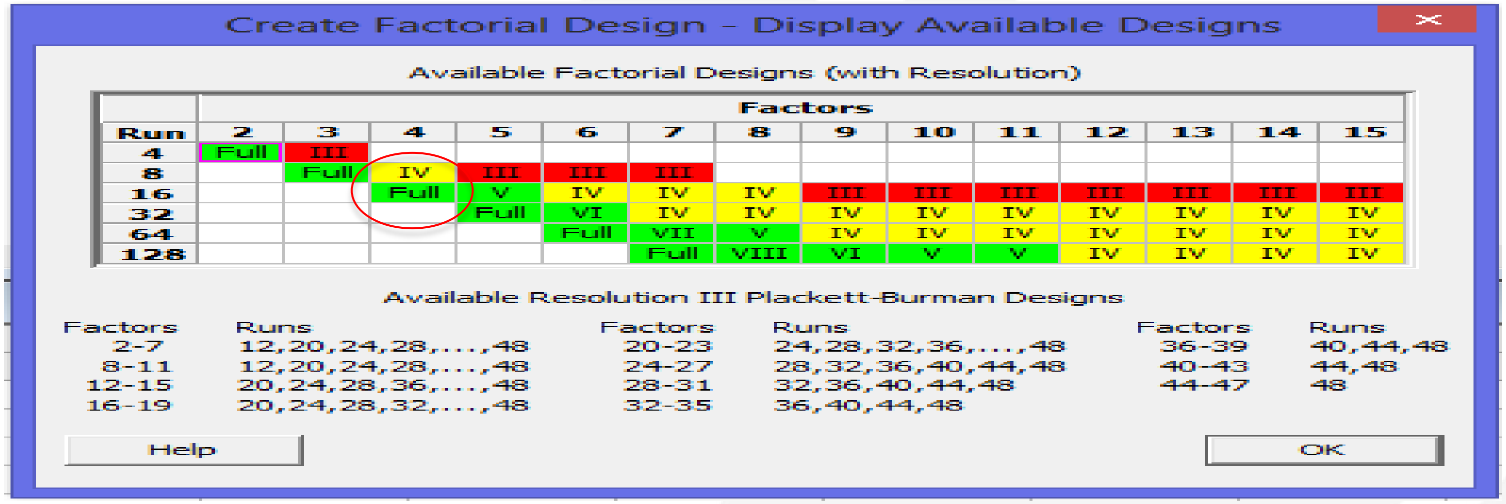
Since that column now has two names, we say that it has an alias or that D is aliased with ABC.

This is also known as confounding

Let's review the Minitab design generator.

Confounding or Aliasing

To create a four factor fractional factorial design, Minitab gives us the following design options:


Create Factorial Design – Display Available Designs

Available Factorial Designs (with Resolution)

Run	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	Full	III												
8		Full	IV	III	III	III								
16			Full	V	IV	IV	IV	III	III	III	III	III	III	III
32				Full	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV
64					Full	VII	V	IV	IV	IV	IV	IV	IV	IV
128						Full	VIII	VI	V	V	IV	IV	IV	IV

Available Resolution III Plackett-Burman Designs

Factors	Runs	Factors	Runs	Factors	Runs
2-7	12,20,24,28,...,48	20-23	24,28,32,36,...,48	36-39	40,44,48
8-11	12,20,24,28,...,48	24-27	28,32,36,40,44,48	40-43	44,48
12-15	20,24,28,36,...,48	28-31	32,36,40,44,48	44-47	48
16-19	20,24,28,32,...,48	32-35	36,40,44,48		

If we were to choose a 4 factor fractional factorial we have a resolution IV available. This would mean that:

1. Main effects would be confounded/aliased with three way interactions and
2. Two way interactions would be confounded/aliased with other two way interactions

An easy way to remember this is that they relate to all the possible ways to add to IV

Trade Offs

When we created the fractional design, factor D replaced the ABC interaction, because we assumed the risk of the ABC three-way interaction being significant to be very low.

What happens if the ABC interaction actually has a significant effect on the response?

Answer: You cannot tell which one is significant because they are not independent.

Further, the actual effect you see is really the sum of the effects of everything aliased together. So you don't really see D, but you see the sum of the effect of D and ABC. We make the assumption that the effect of ABC is very small.

Moreover, you can never make an estimate of the interaction(s) in the Identity (in this case ABCD) because A*B*C multiplied by D will always be +1.

Back to Exercise 3

We want to increase the acceptance rate of our new credit card mailings. The project has been ongoing and has reached the point where a DOE is to be run to test four factors. But we want to do it in 8 runs.

Factor 1 is whether the mailing is sent in white envelope or pastel.

Factor 2 is whether it is mailed by bulk mail or first class.

Factor 3 is the number of enclosures in the envelope.

Factor 4 is the time in days allowed for the customer to respond.

The response is the number that accept the offer out of a 10,000 piece mailing.

Exercise 3: Objective

- What is the goal of the experiment?
  - To determine the best way to send out the credit card offer.
- What are the possible influential factors?
  - From the project so far we have identified envelope, mail, enclosures and time.
- Do you need to reduce variation, shift the mean or both?
  - Raise the average response.
- How much change is required for practical importance?
  - We hope to increase acceptances by 20%.
- Are there multiple responses? Which are more important?
  - Only one response.

Exercise 3 The Y

- Is the response quantitative or qualitative?
  - Qualitative (a.k.a. discrete): number of offers accepted. Even though this is discrete, the number can vary through many hundreds, so it will analyse as if it were continuous.
  - Measurement: number accepted per 10,000 sent.
- Is goal to maximise, minimise or hit a target?
  - Maximise.
- What is current baseline?
  - Baseline from the Measure Phase is just over 500 per 10,000.
- Is response variable currently in statistical control?
  - Again from the Measure Phase, we saw a stable process.
- Is measurement variation understood/adequate?
  - We passed the MSA in the Measure Phase. This was basically counting the returns, so it was easy.

Exercise 3: The Xs

- Which inputs are the most likely drivers of the Y?
  - Envelope, mail, enclosures and time.
- What Six Sigma tools have led you to believe this?
  - In the Measure Phase we used the Process Map, C&E Diagram and C&E Matrix.
- What other inputs, Xs, are there? How will you handle those during the experiment? What about noise variables?
  - We don't think there are many other Xs. At least none that will affect our treatments differently.
  - Randomise: Since all treatment combinations will be sent out at once, there is no time component to randomise around. We will randomise the selection from the mailing list.
  - Large Sample: By sending out 10,000 of each treatment, we should be able to minimise sampling bias.

Exercise 3: The Levels

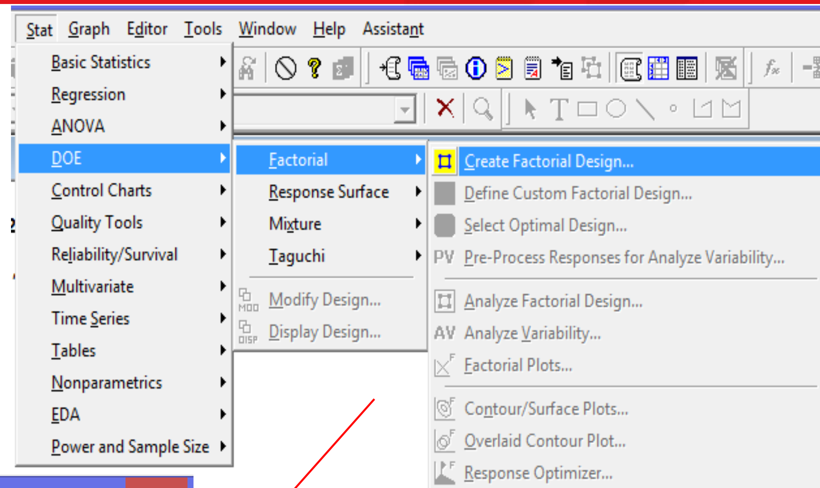
- Where are the Xs set at now?
  - Envelope: We generally use white.
  - Mail: We have used bulk and 1st class.
  - Enclosures: We normally use one.
  - Time: We have gone from 5 days to no limit.
- What levels do you want to know about when you finish?
  - Envelope: white and pastel.
  - Mail: bulk and 1st class.
  - Enclosures: 1 to 3.
  - Time: 5 to 30 days.
- Are there settings that you know create defects?
  - None that we know of.
- What levels are impractical?
  - None that we know of.
- Have you utilised the best judgment of the SMEs and Process Owners?
  - Yes.

Exercise 3: The Design

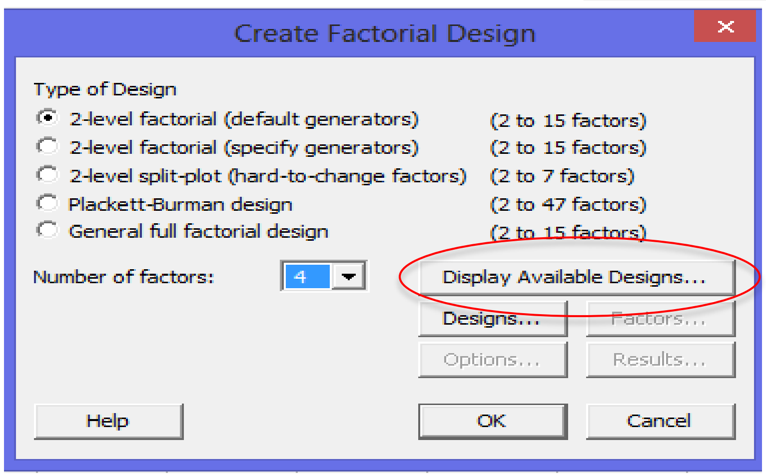
- Which type design matches your objective?
  - Our goal is characterisation.
- Which type matches your budget? Can you afford to get all of the information?
  - Our budget dictates that we try to do it with 8 runs; 80,000 piece mailing.
- We will have to use a fractional factorial.

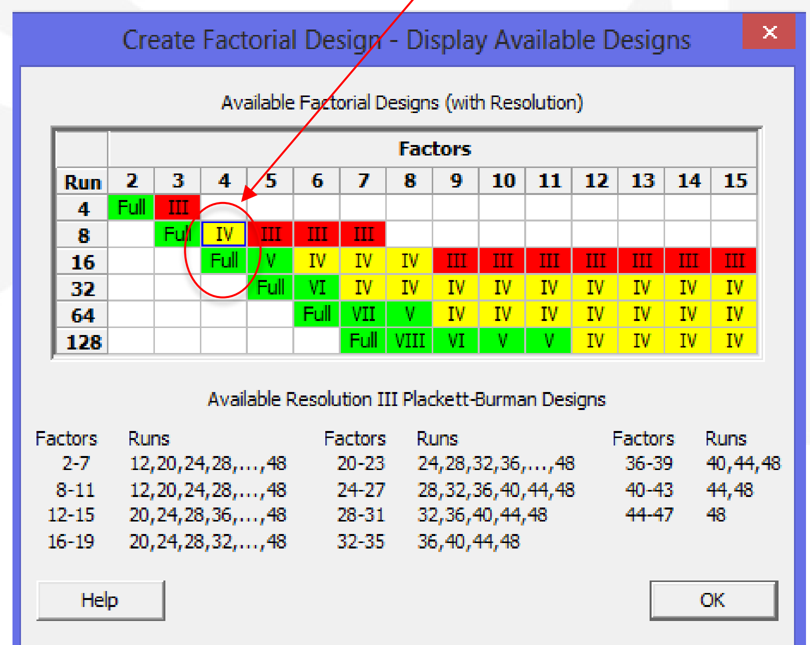
Exercise 3: Use Minitab to design the experiment

Stat > DOE > Factorial > Create Factorial Design



A full factorial design with 4 factors would be 16 runs but a fractional factorial only requires 8





Select 4 factors (ie envelope, mail, enclosures, time to respond) then select display available designs

Resolution

Resolution defines the degree of confounding (aliasing) in a fractional design

- The “cost” of running fractional factorials is that effects and interactions will be confounded
- We use a single Roman numeral, called resolution, to describe the degree of confounding
- Resolution tells us the combinations of interactions of the effects we expect to see confounded

Resolution

Resolution III designs

- Main effects are confounded with 2-way interactions
- They are not confounded with other main effects

Resolution IV designs

- Main effects are confounded with 3-way interactions
- 2-way interactions are confounded with other 2-way interactions

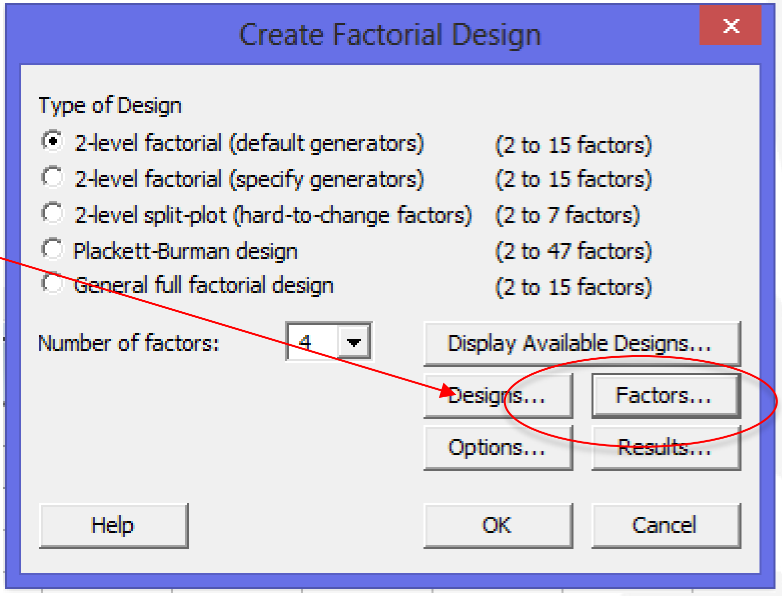
Resolution V designs

- Main effects are confounded with 4-way interactions
- 2-way interactions are confounded with other 3 way interactions

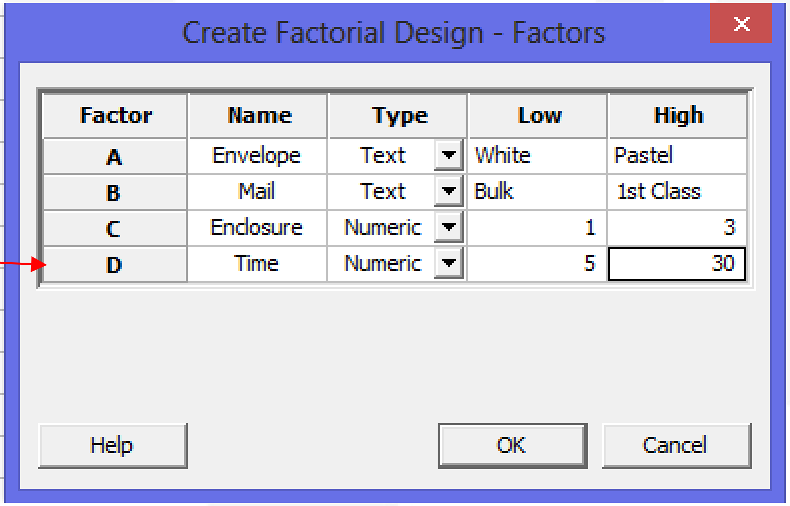
What resolution are we working to?

Exercise 3 Use Minitab to design the experiment

Select Factors



Type in the settings and labels for factors



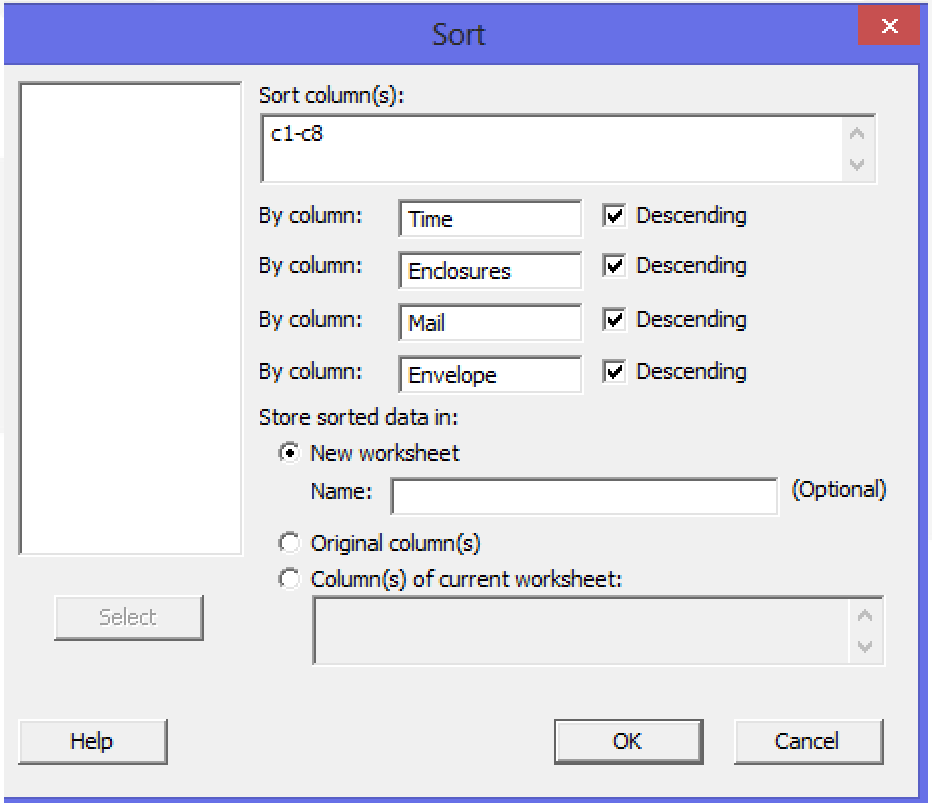
Factor	Name	Type	Low	High
A	Envelope	Text	White	Pastel
B	Mail	Text	Bulk	1st Class
C	Enclosure	Numeric	1	3
D	Time	Numeric	5	30

Exercise 3 Use Minitab to design the experiment

Session Window Output

Fractional Factorial Design

```
Factors:    4    Base Design:          4, 8    Resolution:   IV
Runs:       8    Replicates:              1    Fraction:    1/2
Blocks:     1    Center pts (total):      0


Design Generators: D = ABC


Alias Structure

I + ABCD

A + BCD
B + ACD
C + ABD
D + ABC
AB + CD
AC + BD
AD + BC
```



(**FOR CLASSROOM PURPOSES ONLY!**)
Choose the Run Order as "Sort Right to Left" and click descending.
This will keep us together as a class.

Exercise 3 Design Table

Worksheet 3 *

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T	C8-T	C9
	StdOrder	RunOrder	CenterPt	Blocks	Envelope	Mail	Enclosures	Time	
1	8	4	1	1	Pastel	1st Class	3	30	
2	5	8	1	1	White	Bulk	3	30	
3	3	1	1	1	White	1st Class	1	30	
4	2	5	1	1	Pastel	Bulk	1	30	
5	7	6	1	1	White	1st Class	3	5	
6	6	2	1	1	Pastel	Bulk	3	5	
7	4	3	1	1	Pastel	1st Class	1	5	
8	1	7	1	1	White	Bulk	1	5	
9									

Now we are ready to run the actual experiment and record the results here.
Would we run it in this order?
Describe how this experiment might be run in the business.

Exercise 3 Gather results

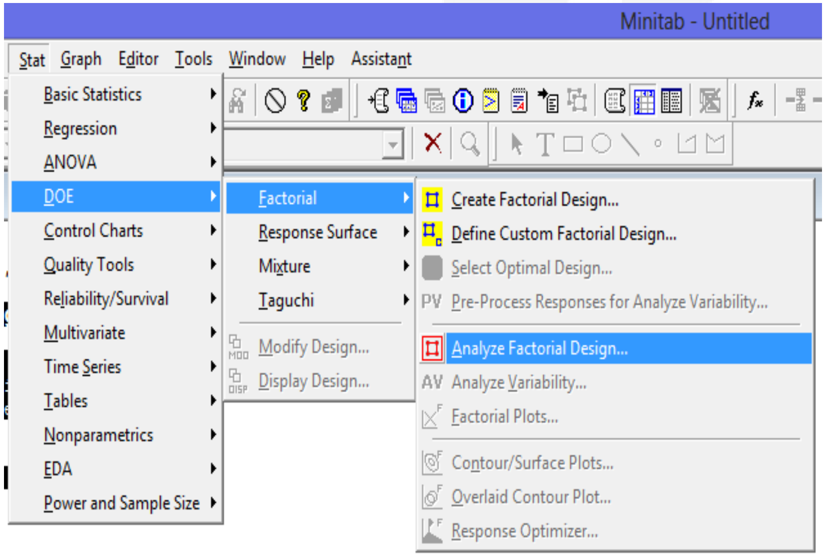
Worksheet 3 ***

↓	C1	C2	C3	C4	C5-T	C6-T	C7-T	C8-T	C9
	StdOrder	RunOrder	CenterPt	Blocks	Envelope	Mail	Enclosures	Time	Acceptance
1	8	4	1	1	Pastel	1st Class	3	30	482
2	5	8	1	1	White	Bulk	3	30	345
3	3	1	1	1	White	1st Class	1	30	612
4	2	5	1	1	Pastel	Bulk	1	30	476
5	7	6	1	1	White	1st Class	3	5	725
6	6	2	1	1	Pastel	Bulk	3	5	546
7	4	3	1	1	Pastel	1st Class	1	5	599
8	1	7	1	1	White	Bulk	1	5	456
9									

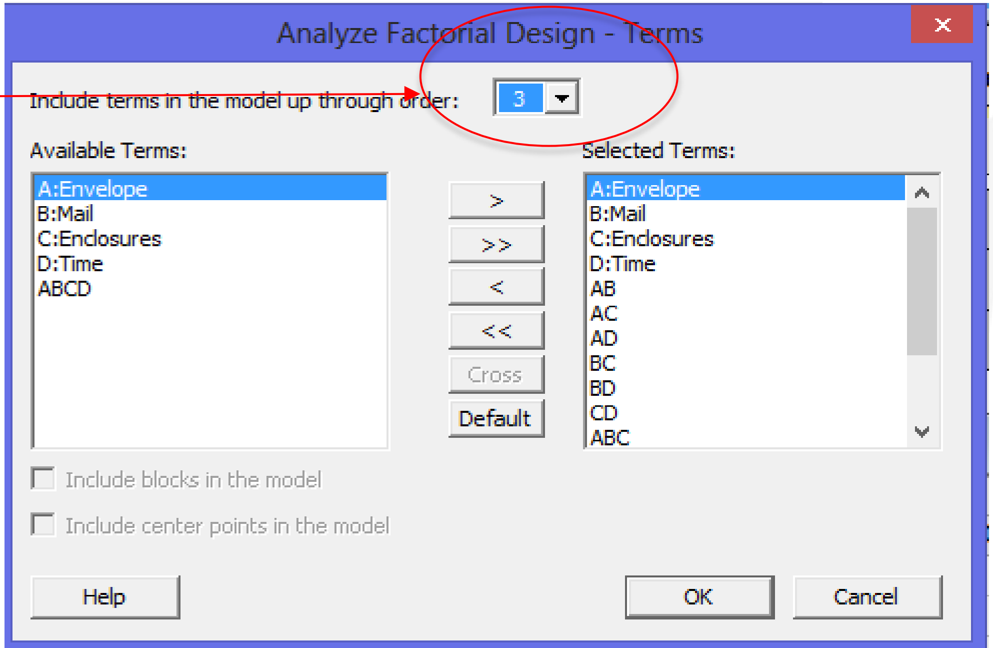
After the experiment is run, enter the data into the Minitab data table.

Exercise 3 Analyse results

Stat > DOE > Factorial > Analyse Factorial design

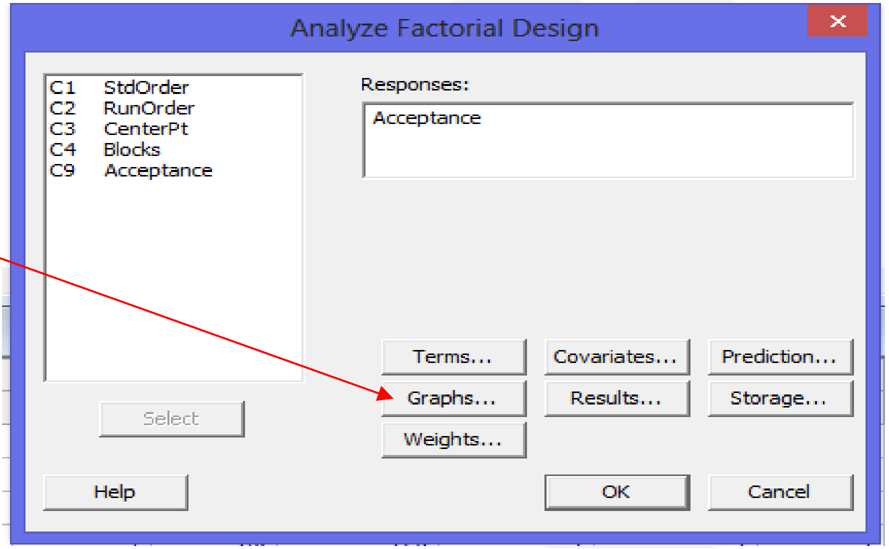


Select “Terms” and include up to 3 factor interactions for a resolution 4 factorial



Exercise 3 Analyse results

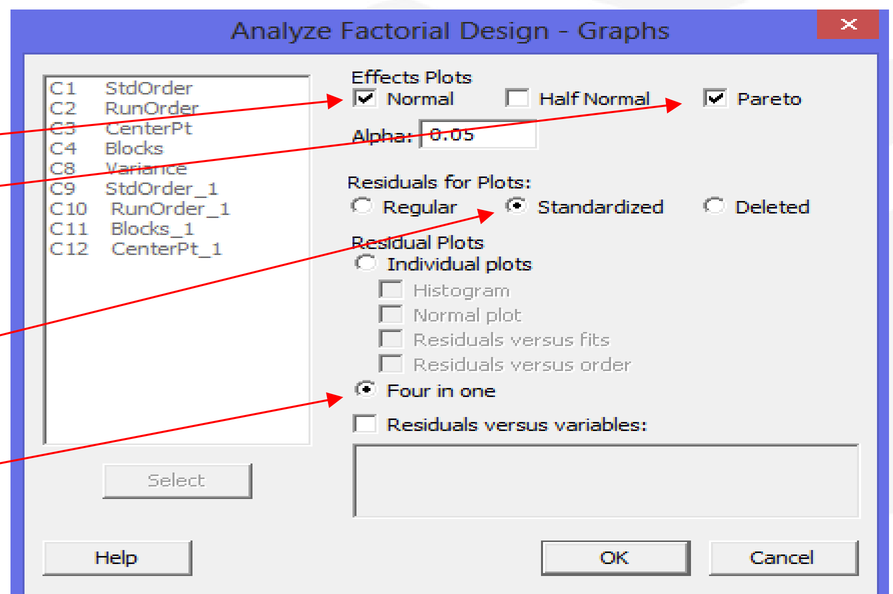
Select Acceptance as the response and select Graphs



Effects plots: Normal Pareto with Alpha set to 0.05

Residuals for plots: Standardised

Residuals plots: Four in one



Exercise 3 Evaluation

Session window output:

Analysis of Variance for Acceptance (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	4	65774.5	65774.5	16443.6	*	*
Envelope	1	153.1	153.1	153.1	*	*
Mail	1	44253.1	44253.1	44253.1	*	*
Enclosures	1	253.1	253.1	253.1	*	*
Time	1	21115.1	21115.1	21115.1	*	*
2-Way Interactions	3	28912.4	28912.4	9637.5	*	*
Envelope*Mail	1	28441.1	28441.1	28441.1	*	*
Envelope*Enclosures	1	300.1	300.1	300.1	*	*
Envelope*Time	1	171.1	171.1	171.1	*	*
Residual Error	0	*	*	*		
Total	7	94686.9				

Note there are no residual plots or p-values

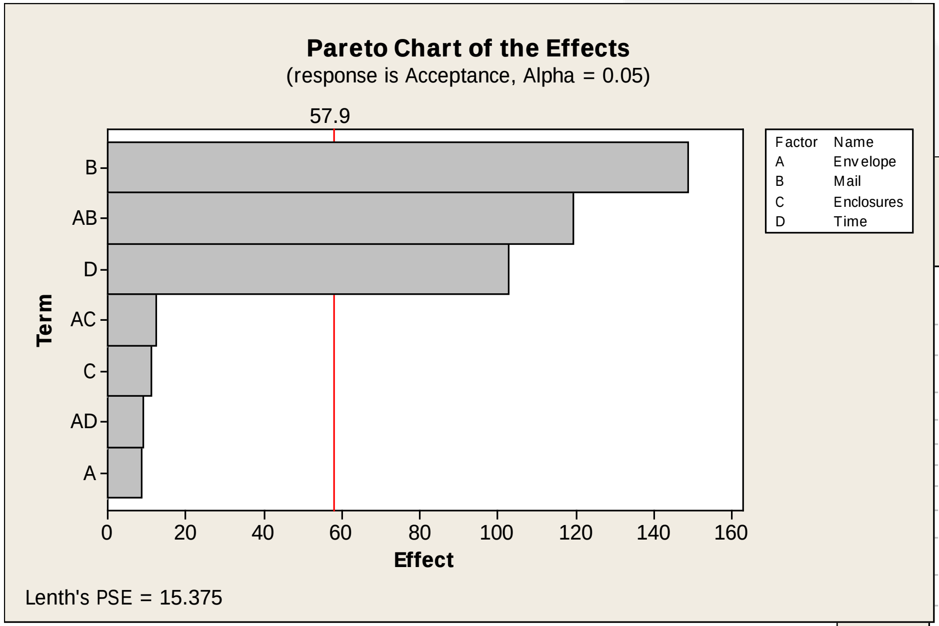
Reason – There is not enough degrees of freedom to give an estimate of residual error in the model. We need to give up some estimates in order to recover some degrees of freedom. We need to reduce the model.

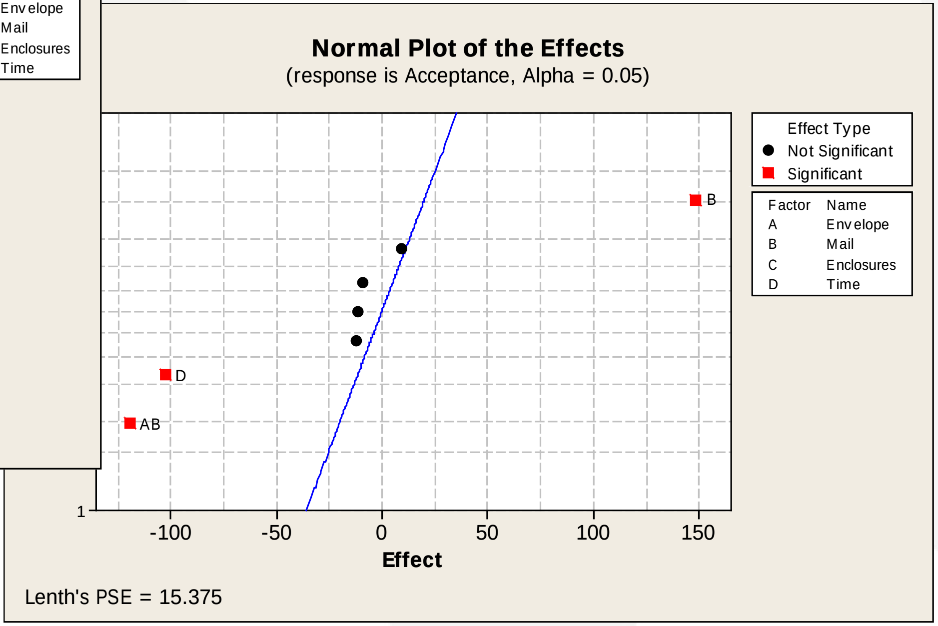
DOE> Factorial > Analyze Factorial Design >Terms

Exercise 3 Evaluation

Which factors should we take out?

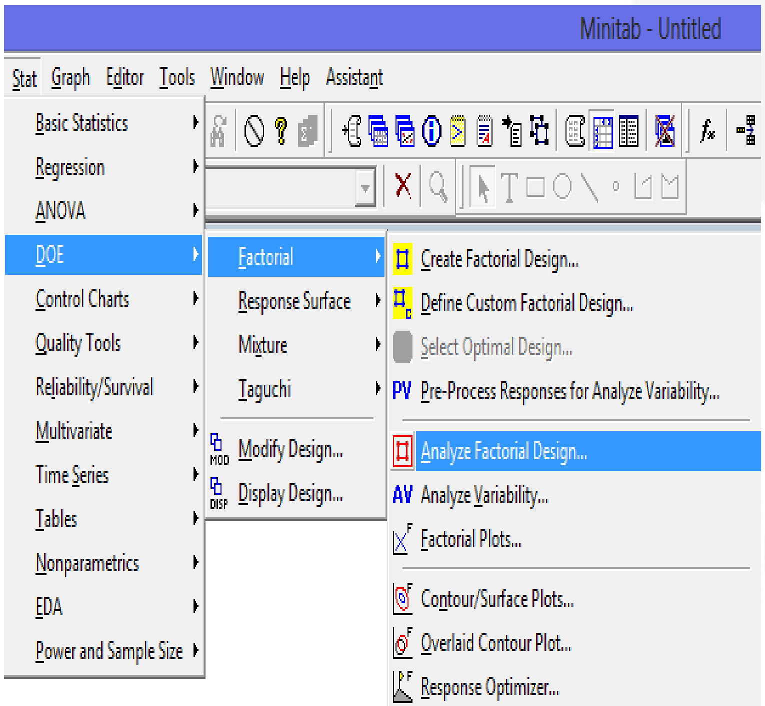
Our plots of effects gave us some clues:

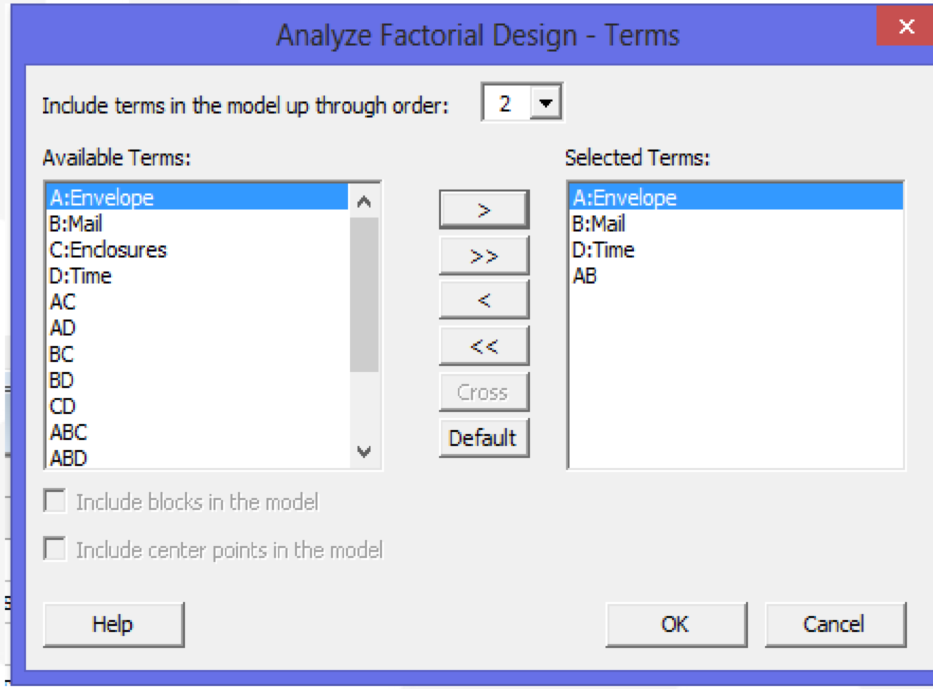




Our graphical analysis is showing us that Mail (B), Time (D) and the interaction of Envelope and Mail (AB) are significant

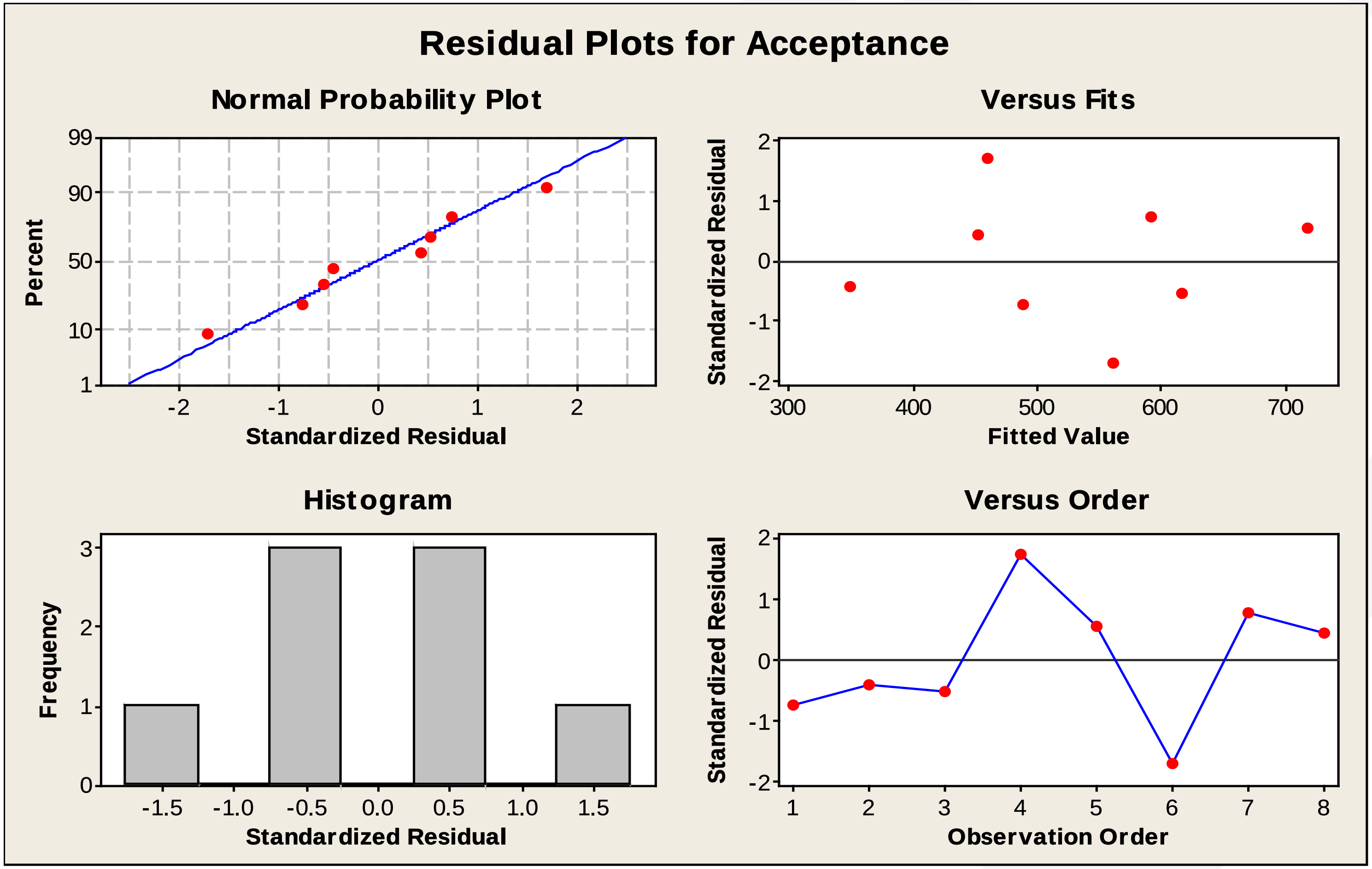
Exercise 3 Reduced Model





Although envelope (A) is not significant, you must keep it in the model because of its interaction with mail (B)

Exercise 3 Residuals check….



All good?

Exercise 3 We have p-values!!!

P-values for the regression co-efficients and Anova

Factorial Fit: Acceptance versus Envelope, Mail, Time

Estimated Effects and Coefficients for Acceptance (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		530.13	5.494	96.49	0.000
Envelope	-8.75	-4.38	5.494	-0.80	0.484
Mail	148.75	74.38	5.494	13.54	0.001
Time	-102.75	-51.38	5.494	-9.35	0.003
Envelope*Mail	-119.25	-59.62	5.494	-10.85	0.002

S = 15.5389 PRESS = 5151.11
R-Sq = 99.23% R-Sq(pred) = 94.56% R-Sq(adj) = 98.21%

Analysis of Variance for Acceptance (coded units)

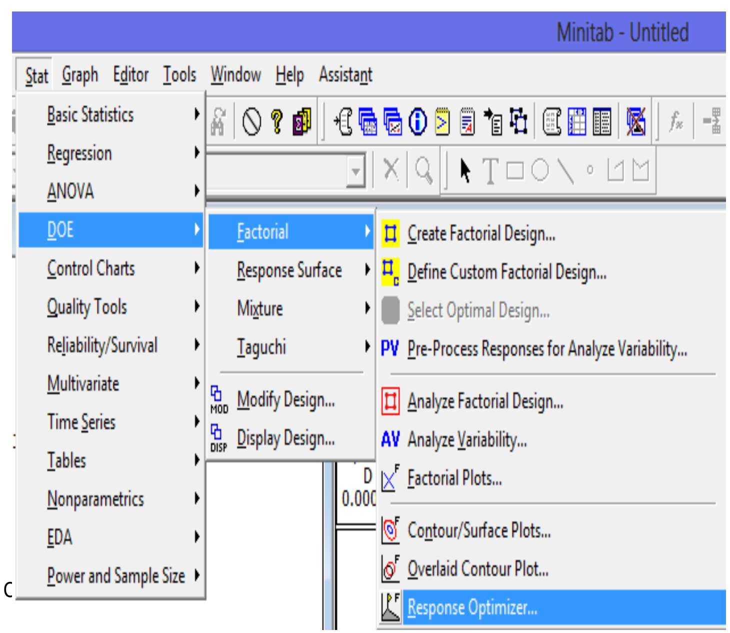
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	65521.4	65521.4	21840.5	90.45	0.002
Envelope	1	153.1	153.1	153.1	0.63	0.484
Mail	1	44253.1	44253.1	44253.1	183.27	0.001
Time	1	21115.1	21115.1	21115.1	87.45	0.003
2-Way Interactions	1	28441.1	28441.1	28441.1	117.79	0.002
Envelope*Mail	1	28441.1	28441.1	28441.1	117.79	0.002
Residual Error	3	724.4	724.4	241.5		
Total	7	94686.9				

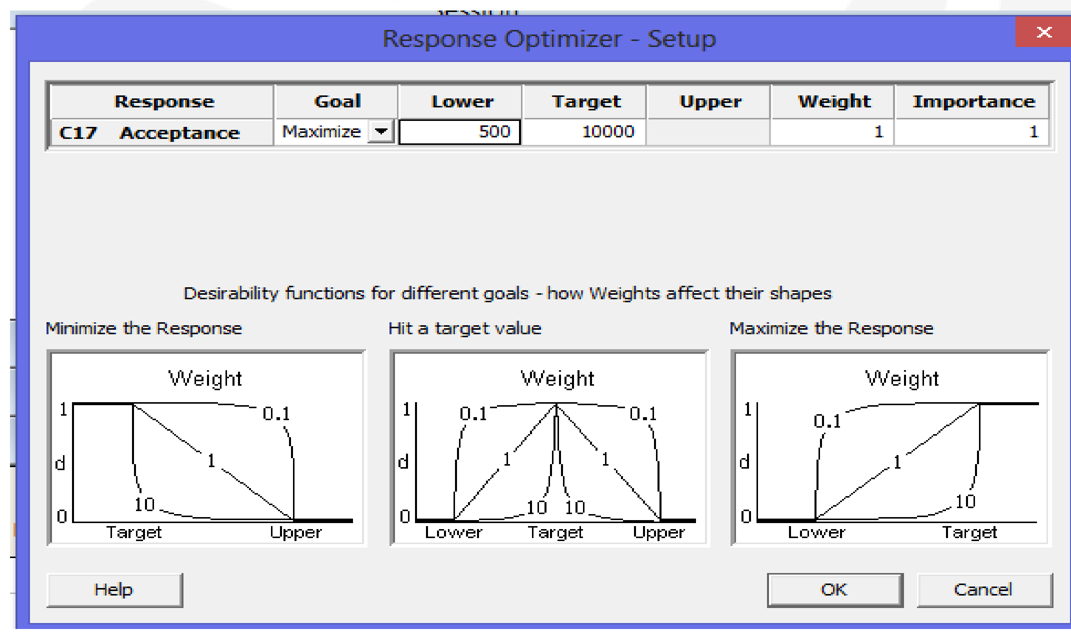
Exercise 3:Optimizer

- Now that we have reduced the model to significant factors (and interactions) we can use the Minitab Response Optimizer tool to model our settings to achieve the best result

- Remember our objective was to maximise response rate, Our baseline performance was 500 responses from 10,000 and we wanted to improve this by 20% to 600 responses

Stat > DOE > Factorial > Response Optimizer

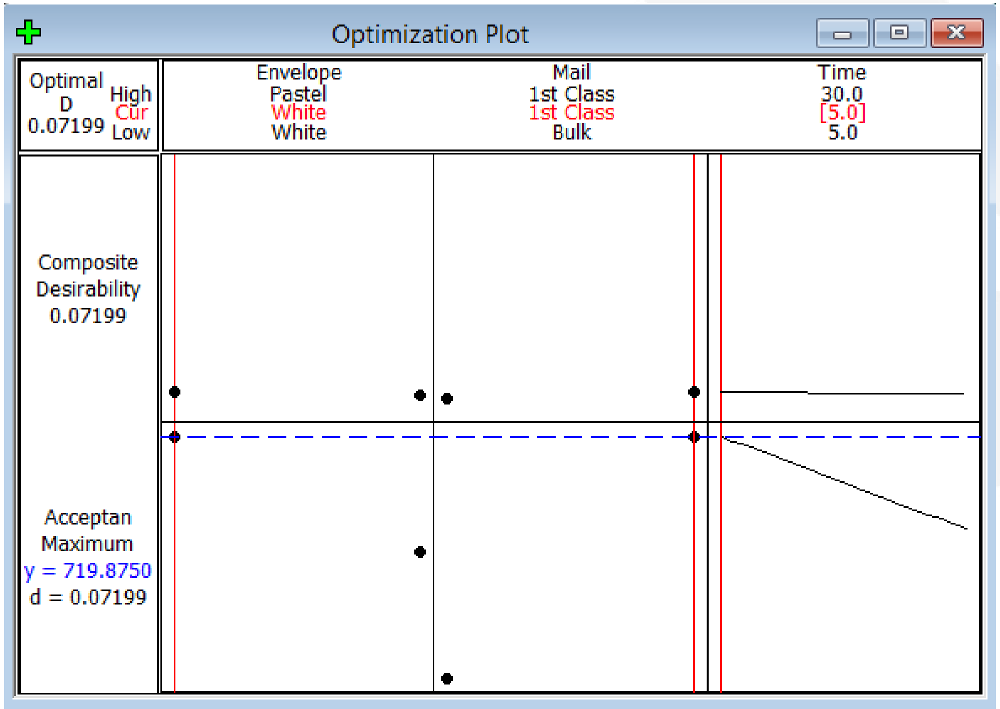




Exercise 3:Results!!!

- Based on setting our factors at:

Envelope:	White
Mail:	1st Class
Time:	5 days



Our estimated response rate is 720, an increase of 44% against a target improvement of 20%!

If we set the factor settings as shown above will we always get a response of 720? If not, why not?

Exercise 5

A bank is interested in increasing its customer satisfaction and know that one of its issues is the time customers wait in lines in Banking Centres. They have started a project to reduce wait time and have narrowed the list of potential root causes/potential improvements to 5, but are having trouble proving which ones are really the drivers.

Factor 1 is the scheduling of clerks. A Black Belt has been working on optimal scheduling model that he thinks will help. Scheduling: Normal (-1) Model (+1)

Factor 2 is phone calls that come into the Banking Centre and take clerks away from customers. Phones: Take calls in Banking Centre (-1) Route calls to call centres (+1)

Factor 3 is a potential new set of procedures for the clerks to follow when servicing the customers. Procedures: Old (-1) New(+1)

Factor 4 is a new idea to have a Floor Manager available to direct customers to the correct line and use a "mobile" clerk to take care of minor issues. Lobby Mgt: No (-1) Yes (+1)

Factor 5 is a simple idea to have signs set up to help customers find the correct line and have the correct information available. Signs: No (-1) Yes (+1)

Exercise 5

Measurement will be customer wait time from the time they enter the queue until they are being serviced by the clerk. Good representative data collection will be difficult because there is a lot of variation between customers, days, banking centres, etc. So we are planning to do a fairly large test to see the effect of our factors in several different settings. To do this we propose to test:

Two markets: Miami and Washington, D.C.

Eight randomly chosen Banking Centres in each market.

We will measure for two weeks.

We will measure nine different times during each day.

The Response that we will analyse in the DOE will be the average wait time of all the measurements for each treatment combination.

Because of the complexity of the experiment and the expense, we will do the experiment in 8 runs, a ¼ fraction.

Open the data set DOE_Frac_Factorial_Ex5.MPJ and analyse the results of this real experiment.

Summary of key learning points

Use Design of Experiment to:

- **Screen out unimportant Xs from the important ones.**
- **Prove causation by tying changes in Y directly to changes in the Xs. Seeks to minimise the effect of noise on the output, Y.**
- **Discover interactions among the Xs, where one X has a multiplying effect on another X.**
- **Develop a prediction equation that will allow us to manage our process at the optimal level.**

Recommended coach support points

- **Running a DOE with a discrete response**
- **If you suspect that you may not have a linear response**

Black Belt

Change Adoption

Objectives of this module

At the end of this module, you will be able to:

- Define Change Adoption (CA)
- Understand importance of CA for all changes or projects
- Provide high level definition of the CA model
- Overview each element of a change adoption model

Questions Lead

- **How do we ensure the changes put in place will be sustained?**
  - **Who do we need to communicate with?**
  - **What are the change messages?**
  - **How do we communicate the need for change?**
  - **When/how do we address change adoption issues?**

What is Change Adoption?

Change Adoption Is about leading the people side of change!

When the right ideas have the support of the right people… the success of the change is greater!

Knowing what, how and when to communicate the right messages at the right time is key.

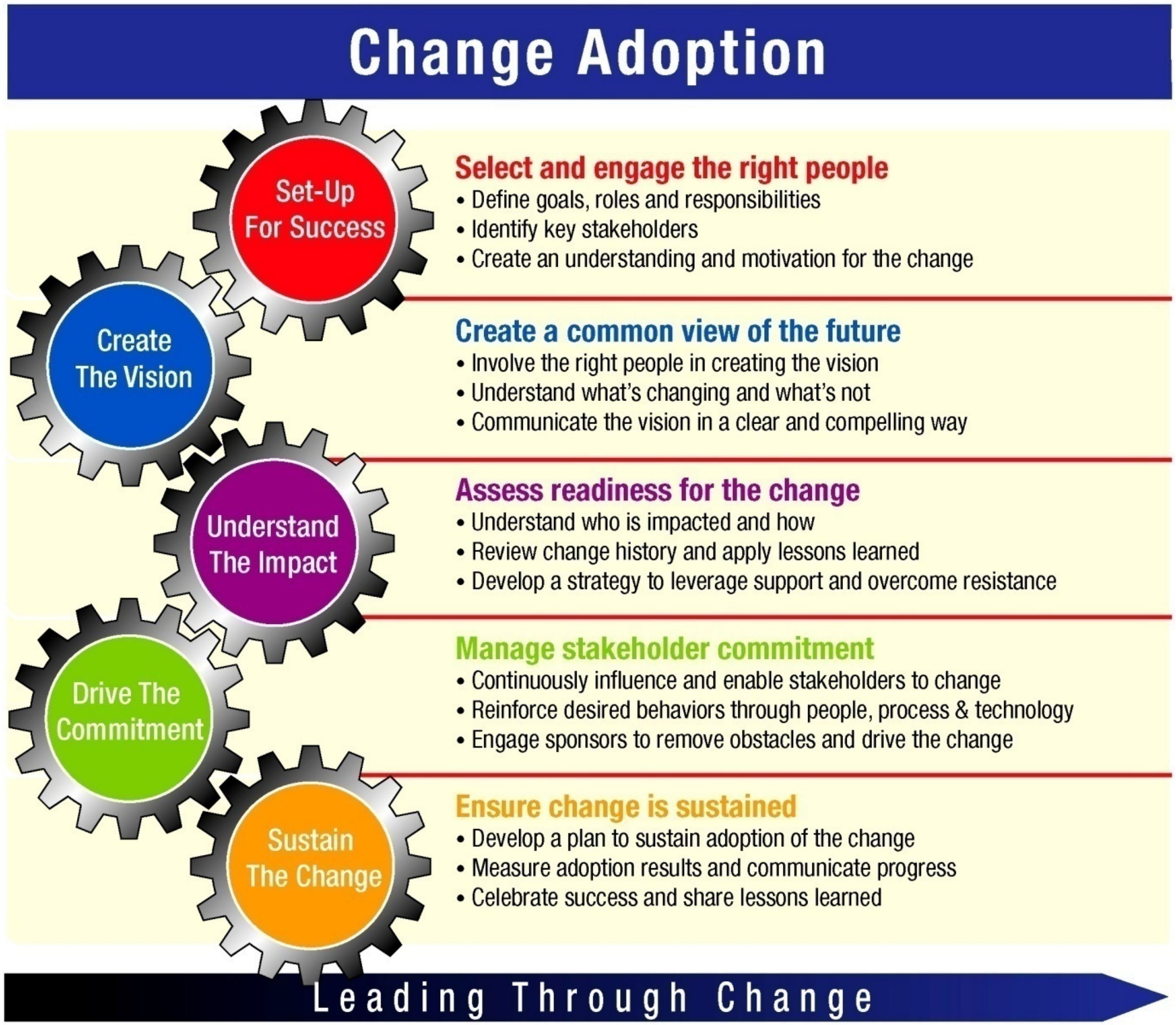
The Facts

- **Market research by McKinsey & Co. shows that *15% of all change projects completely fail***
- **Remaining change projects achieve *only 40-60% of anticipated results and only 33% of projects are deemed successful***
- **There is a *direct correlation between the level of project success and the factor called "change adoption"***
- **In a study by Prosci, research shows that 48% of project *results* are linked to project management activities and *52% are linked to change adoption activities***

Change Adoption Components

SET UP FOR SUCCESS
CREATE THE VISION
UNDERSTAND THE IMPACT
DRIVE THE COMMITMENT
SUSTAIN THE CHANGE

Change Adoption Model



Change adoption is a disciplined approach that views change as a journey versus an event.

The Change Adoption model is flexible and adaptable to projects of all sizes and at all levels.

The Change Adoption model elements and activities are interdependent and support the journey toward change adoption.

Component & Tools Overview

- The elements of the model are **interdependent and support the journey** toward change adoption
- The model outlines sequential work, but the dynamics of your particular change may require you to **shift back and forth, circle back, or work simultaneously between different elements**
- **Even the smallest changes** need leadership and the support of people to be effective

Select and engage the right people

- Define goals, roles and responsibilities
- Identify key stakeholders
- Create an understanding and motivation for the change

Benefits:

- Go slow to go fast
- Drives buy-in of the team
- Builds momentum and a sense of urgency
- Improves effectiveness and the ability to make the right decisions
- A “smart start” typically results in long-term success

Available Tools

- Stakeholder Management Grid
- Change Target Pie Chart
- Stakeholder Analysis

Identify Key Stakeholders

- GRPI (Goals, Roles, Process, Interpersonal relationships) Agreement
- Project Scope Statement
- In & Out of Frame
- RACIE Assessment (Responsible, Accountable, Consulted, Informed, Enabler)
- GRPI Assessment

Define Goals, Roles and Responsibilities

Create an Understanding and Motivation for the Change

- In & Out of Frame
- Risk & Reward Matrix
- Change Motivation Statement
- SWOT (Strengths, Weaknesses, Opportunities, Threats) Analysis

Example - Change Motivation Statement

Team Member	Team Member	Team Statement
"Improve customer scores and our performance as a company"	*"Prevent loss of customers and realise more profit"*	*"Why are we changing?"* *"Improve customer experience and realise measurable results"*

- **Alignment tool useful in answering the question, "Why are we changing?"**
- **Results in the "burning platform" message**
- **Creates a shared understanding of the change and motivation for the change**
- **Useful when forming project goals, when getting "push-back" or as a communications message**
- **Involves all team members**

Create the Vision

Create a common view of the future

- **Involve the right people in creating the vision**
- **Understand what's changing and what's not**
- **Communicate the vision in a clear and compelling way**

Benefits:

- **Everyone knows where they are headed**
- **Inspires and motivates people to change**
- **Guides a change team's decisions and actions**
- **Creates a sense of alliance**

Available Tools

- **Shared Vision Statement**

Involve the right people in creating the vision

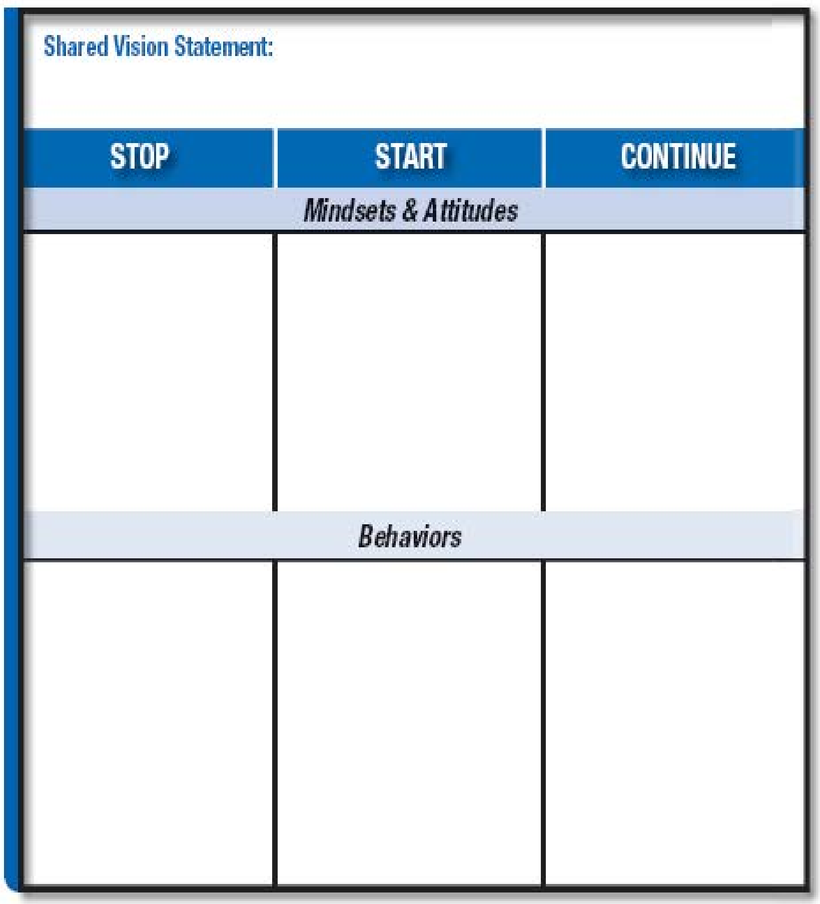
- **Vision Critical Success Factors (CSF's)**
- **Stop/Start/Continue**
- **What's Changing/What's Not**

Understand what's changing and what's not

- **Elevator Speech**

Communicate the vision in a clear and compelling way:

Example – Stop/Start/Continue



- Helps teams make a Shared Vision Statement behaviorally understood
- Identifies desired mindsets/ attitudes (ways of thinking, core values, emotions) and behaviors (actions) that need to stop, start or continue when the change is implemented
- Use when visioning with stakeholders or when there is uncertainty
- Creates a common view of the future

Understand the Impact

Assess Readiness for the Change

- **Understand who is impacted and how**
- **Review change history and apply lessons learned**
- **Develop a strategy to leverage support and overcome resistance**

Benefits:

- **Allows leaders to identify barriers to and drivers for successful change adoption**
- **Understand who is supportive of the change and who might be resisting it**
- **For those resisting the change begin to identify what their concerns are**
- **Create a strategy to influence the resistors to be more supportive**

Available Tools

- Stakeholder Communication Matrix
- Stakeholder Analysis
- Change Target Pie Chart

Understand who is impacted and how

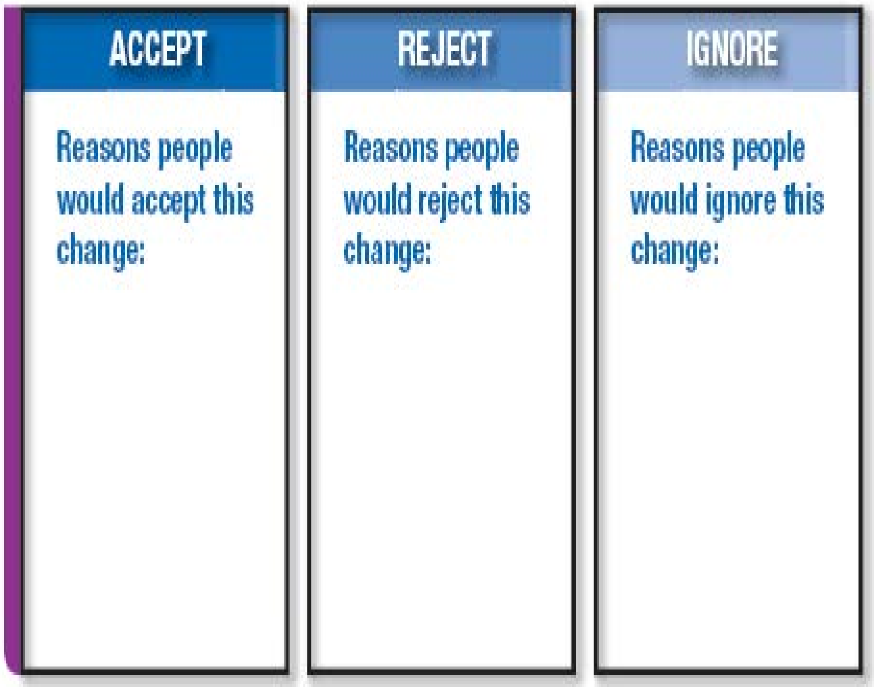
- Change Readiness Assessment

Review change history and apply lessons learned

- Accept-Reject-Ignore
- Sources of Motivation
- Overcoming Resistance
- Diagnosis of Resistance
- People, Processes, Technology Checklist
- Stakeholder Communication Matrix

Develop a strategy to leverage support and overcome resistance

Example – Accept-Reject-Ignore



- **Will help teams anticipate how stakeholders will react to the change announcement**
- **Targeted approach to communications, influencing, and stakeholder coaching**
- **Aids in understanding amount and type of resistance**
- **Allows for development of a strategy to leverage support and overcome resistance**

Drive the Commitment

Manage Stakeholder Commitment

- **Continuously influence and enable stakeholders to change**
- **Reinforce desired behaviors through people, processes, and technology**
- **Engage sponsors to remove obstacles and drive the change**

Benefits:

"Influencing stakeholders is <u>critical to the success</u> of every change initiative in every organisation I have ever worked in. By engaging the right people in the right way, you will make a significant, positive impact on how well the change is adopted."

- An Experienced Change Practitioner

Available Tools

- Stakeholder Analysis
- Influencing Strategy
- Communication Plan

Continuously influence and enable stakeholders to change

- Helping/Hindering

Reinforce desired behaviors through people, processes, and technology

- State of Change Agenda
- Change Sponsor Contracting Process
- Change Coaching Process

Engage sponsors to remove obstacles and drive the change

Example – Helping/Hindering

Organizational Factors	Helping	Hindering	Actions
Hiring & Staffing	Description	Description	Next Steps
Organizational Structure		Description	Next Steps
Resource Allocation	Description		Next Steps
Performance Management		Description	Next Steps
Reward & Recognition		Description	Next Steps
Goal Setting	Description		Next Steps
Learning & Development			Not Applicable
Process	Description		Next Steps
Technology		Description	Next Steps
Legal/Compliance			Not Applicable
Other:			

+5 +4 +3 +2 +1 -1 -2 -3 -4 -5

- Identifies organisation factors that help or hinder a change's potential for adoption
- Quantifies strength of the impact and leads to action plans
- Aids in planning for implementation or mitigate risks due to change resistance
- Supports change through alignment of people, process, and technology

Sustain the Change

Ensure change is sustained

- **Develop a plan to sustain adoption of the change**
- **Measure adoption results and communicate progress**
- **Celebrate success and share lessons learned**

Benefits:

- **Ensures achieve productivity planned because of the change**
- **Realise expected financial and non-financial benefits**
- **Protects the organisations investment as well as the team’s**

Sustaining the Change is What This is ALL ABOUT!

Available Tools

- Change Adoption & Sustainment Plan
- Change Adoption Metrics
- Change Adoption Leadership Assessment
- Change Adoption Sponsor Assessment

Develop a Plan to Sustain Adoption of the Change

- Change Adoption Metrics
- GRPI Assessment
- Change Readiness Assessment
- State of Change Agenda
- Change Adoption Leadership Assessment
- Change Adoption Sponsor Assessment

Measure Adoption Results and Communicate Progress

Celebrate Success and Share Lessons Learned

- Communication Plan
- State of Change Agenda

Example - Adoption and Sustainment Plan

Change Adoption & Sustainment Plan			
Elements of Plan	**Project Outcome #1**	**Project Outcome #2**	**Project Outcome #3**
Sustainment Strategies: Best practices for reinforcing and integrating this outcome into the fabric of the culture; making it "the new way"			
Measurement: Indicators that this outcome is being adopted & sustained by associates			
Monitoring: Strategies to ensure this outcome is being monitored and measured on an ongoing basis			
Support (Owner): The group or job role (with the resources and capability) to own and support this outcome long-term			

- **Helps a team identify the strategies, metrics, and support needed to ensure the change is long-lasting**
- **Used at beginning of the project to define a sustainment strategy; throughout the change process to ensure progress is made; and when communicating with sponsors and stakeholders**

What do Change Leaders Need to Do?

1. Engage
2. Create motivation for the change
3. Care about the impact to people
4. Provide private support
5. Provide public support
6. Leadership
7. Leverage their network
8. Practice 360-degree coaching
9. Monitor progress and resistance
10. Reinforce change

Appendix - CA Tollgate Questions (i.e., Report Out Elements) 1 of 2

- Set-up for Success and Sustainment
  - *Definition: Ensuring that everyone is on the same page going into the project and agree on the approach to sustain the change*
    - Discuss what the project is about, why it's important, what success looks like and the role stakeholders will play in driving success.
    - What activities have contributed to the development of the Change Adoption and Sustainment Plan?
- Assessing the Impact of the Change
  - *Definition: Assessing and understanding the impact of the change on our stakeholders*
    - Has the project team engaged critical support partners to assist in identification of impacts (Learning/Communication partners, Risk, Audit, GHR, Information Protection, etc)?
    - Have you and the project team assessed who is impacted by the change?
    - How are we engaging stakeholders to help them prepare for these impacts?
    - Are you actively mitigating any negative impacts? If so, what steps have been taken and what actions are planned?
- Management Routines
  - *Definition: Mgmt routines within the project team to ensure discussion concerning stakeholders and ensuring engagement at every step in the process.*
    - What routines do you have in place to ensure you remain in touch with your stakeholders ? Are you aware of their potential areas of resistance, level of support, ability to execute on the change?
    - Do we know what the points of resistance are and do we have a plan to move stakeholders from **Acknowledgement** to **Acceptance**?
    - What actions have been taken to overcome known resistance?

Appendix - CA Tollgate Questions (i.e., Report Out Elements), 2 of 2

- Stakeholder/Sponsor (or Change Target) Engagement
  - *Definition: Ensuring there are ongoing conversations, communications, education, interviews, etc with our stakeholders*
    - How are you actively engaging your sponsors to remove obstacles and drive the change?
    - What routines are in place or what leader assessments have been completed to ensure adequate sponsorship is sustained throughout the project?
    - How are your stakeholders being engaged at each level? (executive, leadership, staff levels)
    - Do you have a plan to move your stakeholders from **Acknowledgement** to **Acceptance**?
- Stakeholder Readiness
  - *Definition: Are stakeholders ready to accept and execute the change?*
    - Stakeholder Readiness Dashboard/Assessment
    - Do impacted associates have the knowledge, skill and ability to embrace the change and execute on it?
    - What has been done to engage those staff members/stakeholders impacted by the change?
    - Has anticipated resistance been addressed? In what way?

Summary of key learning points

- Change adoption is often thought of as a “soft” skill but is really anything but given it’s a principal driver behind why projects fail

- Using a model such as that described within this module at every step of your change effort will support you in making change and support the sustainment of your solutions

- There are many tools and techniques available to you, only some of which have been highlighted in this module, look them up!

- ***Don’t ignore the people side of change***

Black Belt

Logistic Regression

Objectives of this module

At the end of this module, you will be able to:

- Know when to use Logistic Regression.
- Understand basic concepts of Logistic Regression.
- Perform and analyse a Binary Logistic Regression.

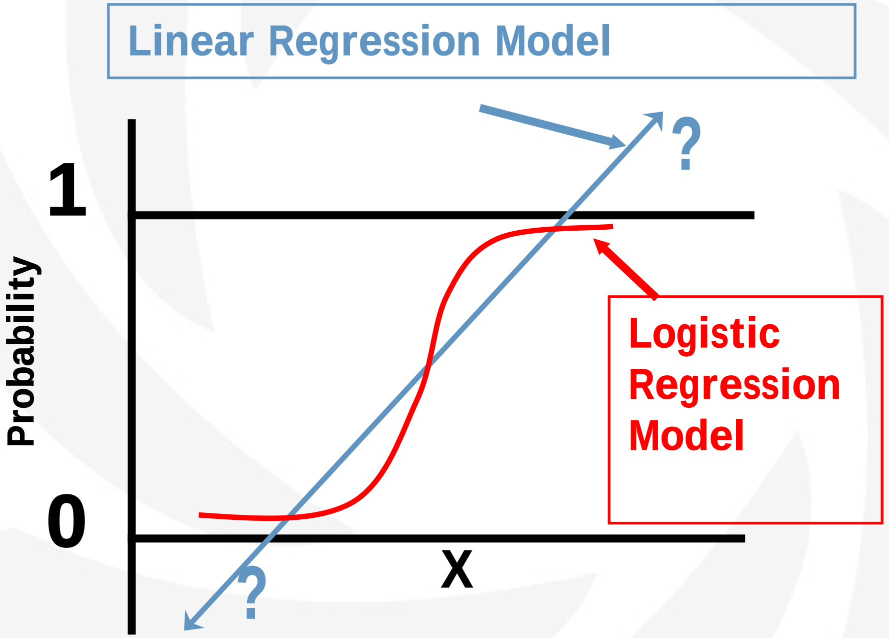
Questions Lead

- Basic question: Is there a relationship between my continuous or discrete inputs and my **discrete** output?

  - Do these Xs affect the probability of seeing my preferred Y output? All of them or just some of them?
  - As these Xs change, what happens to the probabilities?
  - How much does the probability of getting my preferred Y output change as the individual Xs change?

Why Use Logistic Regression?

- In Logistic Regression, we calculate the **probabilities** of an event occurring. These will be between 0 and 1.
- In Linear Regression, we calculate the actual predicted values. These can be greater than 1 and less than 0, but the probability of an event cannot be greater than 1 or less than 0.



Also remember that two of the assumptions of Linear Regression are that the residuals are normally distributed and are relatively constant across the values of X (homoscedastic). This WILL NOT happen when we have a discrete Y. The distribution of the probabilities would not yield residuals that are even approximately normal or homoscedastic.

How Logistic Regression Works

- It turns out that Logistic Regression, in a sense, modifies the data and then uses Linear Regression to analyse it.
- So if Linear Regression theoretically can go from -∞ to +∞, we need to modify our data to do the same. [Good news: the software does this in the background!]
- The first step is to **calculate the odds of an event happening** based on the probability.
- *Odds* - Probability that the event occurs divided by probability that the event does not occur or:

$$\text{Odds} = \frac{\text{Probabilit y}}{1 - \text{Probabilit y}}$$

Understanding Odds

- So how do odds help us?
- Examples:
  - The probability of success .50, Odds = $\frac{.5}{1-.5} = \frac{.5}{.5} = 1$
  - The probability of success .25, Odds = $\frac{.25}{1-.25} = \frac{.25}{.75} = .333$
  - The probability of success .99, Odds = $\frac{.99}{1-.99} = \frac{.99}{.01} = 99$
  - For very low probability, say .0001, the Odds approach zero. $\frac{.0001}{1-.0001} = \frac{.0001}{.9999} = .0001$
  - For very high probability, say .9999, the Odds approach +∞. $\frac{.9999}{1-.9999} = \frac{.9999}{.0001} = 9999$

Continuing the Transformation

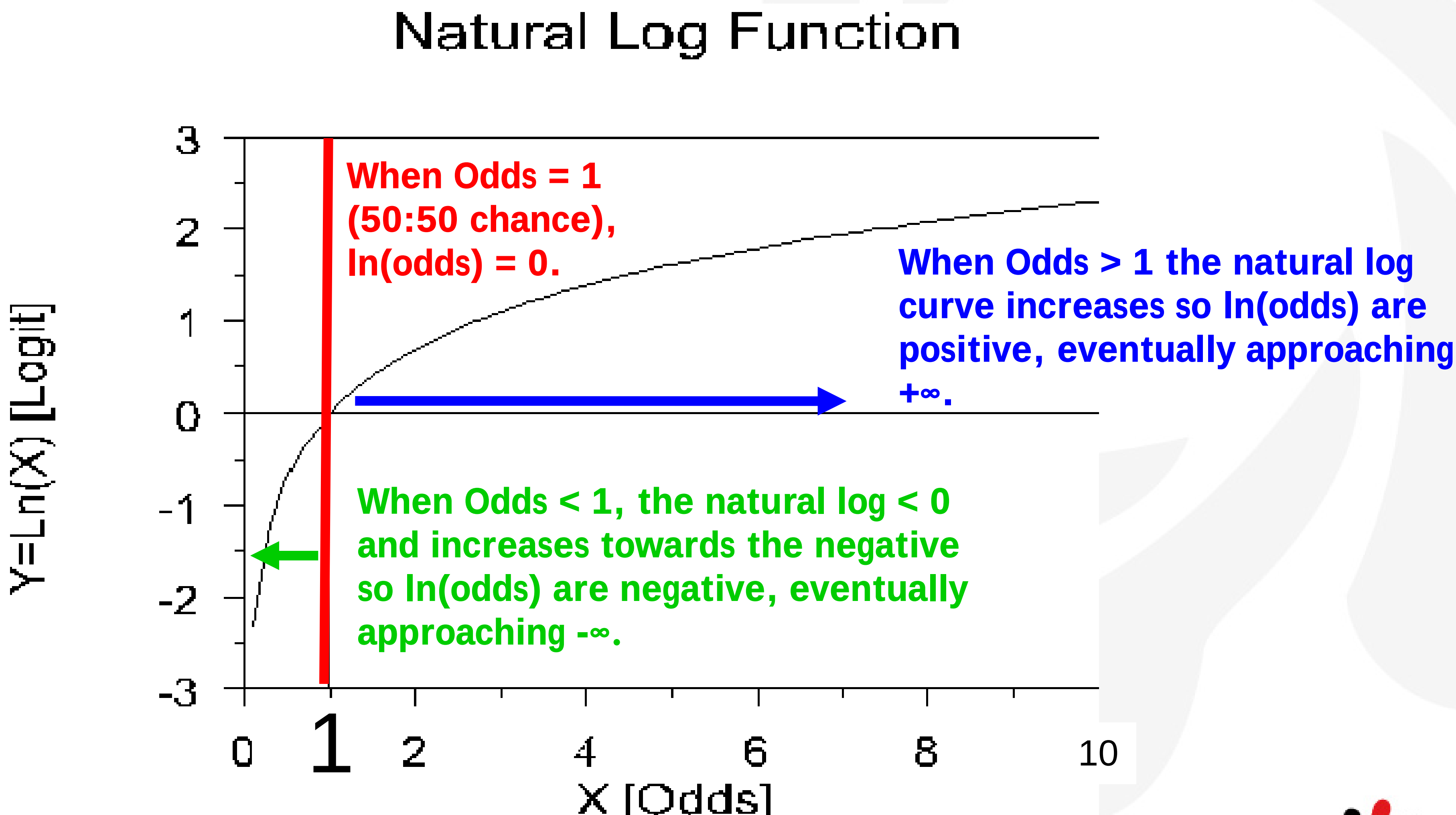
- If Odds takes us from [0 to 1] to [0 to +∞], the next step is to take the logarithm of the odds to get us to [-∞ to +∞].

- A logarithm is the number to which another number must be raised to get the answer.
  - Example: $10^3 = 1000$. The logarithm (base 10) of 1000 is **3**.

- In logistic regression, the base number that is used is the mathematical constant ‘e’ which is 2.7182.

- A graph of this is on the next slide.

Graph of the Transformation



What This Means

- Now we have a function of the probability that can go from [-∞ to +∞] called the **Logit function**.
- Minitab will now do a linear regression, solving for ln(odds) giving us an equation in the form: $\ln(\text{odds}) = b_0+b_1x_1+b_2x_2 \; . \; . \; . \; b_nx_n$.
- By raising e to the power calculated by this equation, we get the actual odds.
- But the business doesn't understand odds let alone ln(odds). We have to take this back to probabilities. Fortunately the math is fairly easy to go backwards from ln(odds) to get the probability. (And don't forget, we live in the computer age!)
- The only other concept we need is the concept of **Odds Ratios**. This is the means by which the mathematics evaluates how much of a change is caused by a difference in the level of a discrete X or by a one unit change in a continuous X. If an X can take the values of A or B, the odds ratio is the odds of A causing the event divided by the odds of B causing the event. For a continuous X say dollars, it is the odds of the event occurring at w$ divided by the odds of the event occurring at (w+1)$.
- The equations are on the next page.

Background Equations

$$\text{Odds} = \frac{\text{Probability}}{1-\text{Probability}} \therefore \text{Probability} = \frac{\text{Odds}}{1+\text{Odds}}$$

$$e^{\ln(\text{Odds})} = \text{Odds} \therefore \text{Probability} = \frac{e^{\ln(\text{Odds})}}{1+e^{\ln(\text{Odds})}}$$

$$\ln(Odds) = b_0 + b_1x_1 + b_2x_2 \ldots b_nx_n$$

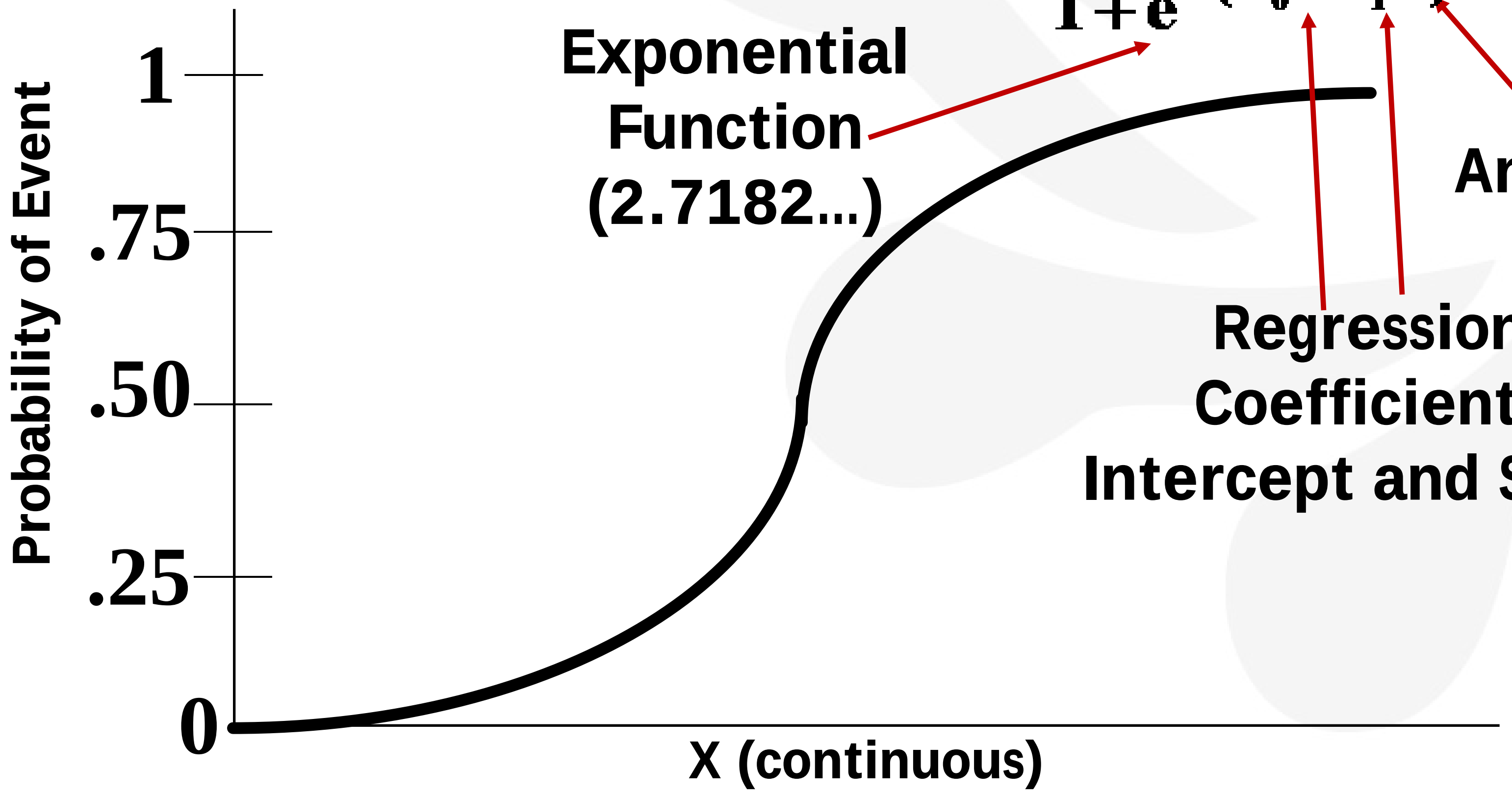
$$\text{Probability} = \frac{e^{b_0+b_1x_1+b_2x_2\ldots b_nx_n}}{1+e^{b_0+b_1x_1+b_2x_2\ldots b_nx_n}} = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2\ldots b_nx_n)}}$$

$$\text{Odds Ratio} = \frac{\text{Odds of Event at } X_a}{\text{Odds of Event at } X_b}$$

Logistic Curve (one continuous X)

By using the Logit function the binary scale is transformed into a continuous probability scale:

$$\text{Probability of Event} = \frac{1}{1+e^{-(B_0+B_1X)}}$$



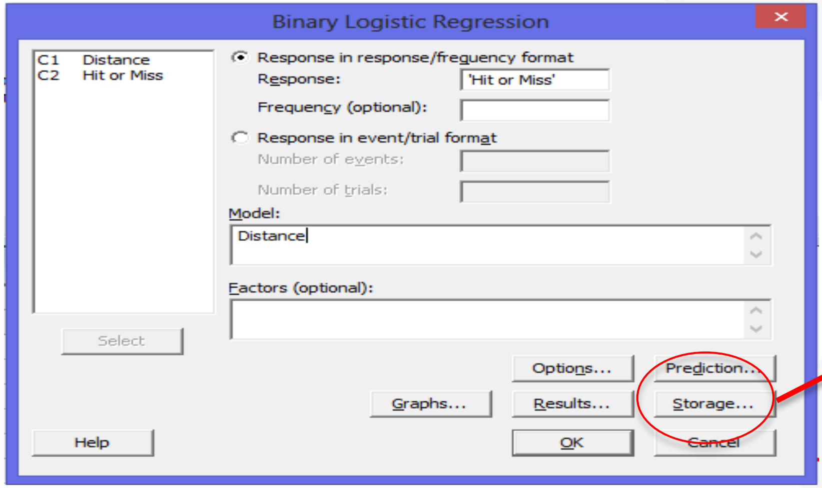
Exercise 1

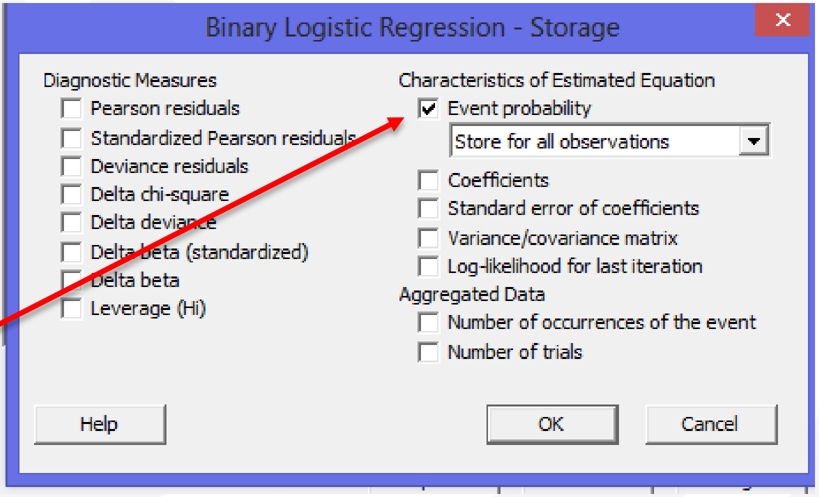
- **Break into small teams, take some shots at the basket and collect the following the data: Y = H or M (Hit or Miss) and Distance of the shot to the nearest ½ foot (e.g. 6.5).**
- **Randomise on people and distances. Use MANY different distances.**
- **Take about 30 shots.**
- **Create a data table in Minitab and record your data.**

↓	C1	C2-T
	Distance	Hit or Miss
1	1.5	H
2	10.0	M
3	8.0	H
4	13.0	M
5	2.0	H
6	6.0	H
7	6.0	H
8	1.0	H
9	13.5	M

Exercise 1, Minitab Analysis Inputs

Stat > Regression > Binary Logistic Regression





Exercise 1, Session Window Output

Binary Logistic Regression: Hit or Miss versus Distance

Link Function: Logit

Summary of the output of my test
(yours may be slightly different)

Response Information

Variable	Value	Count	
Hit or Miss	M	12	(Event)
	H	18	
	Total	30	

The Logistic Regression table
The p-value for the factor "Distance" is less than 0.05 indicating that it has a significant effect on whether we hit or miss.

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-11.0643	5.15574	-2.15	0.032			
Distance	1.34861	0.642374	2.10	0.036	3.85	1.09	13.57

Log-Likelihood = -4.990
Test that all slopes are zero: G = 30.400, DF = 1, P-Value = 0.000

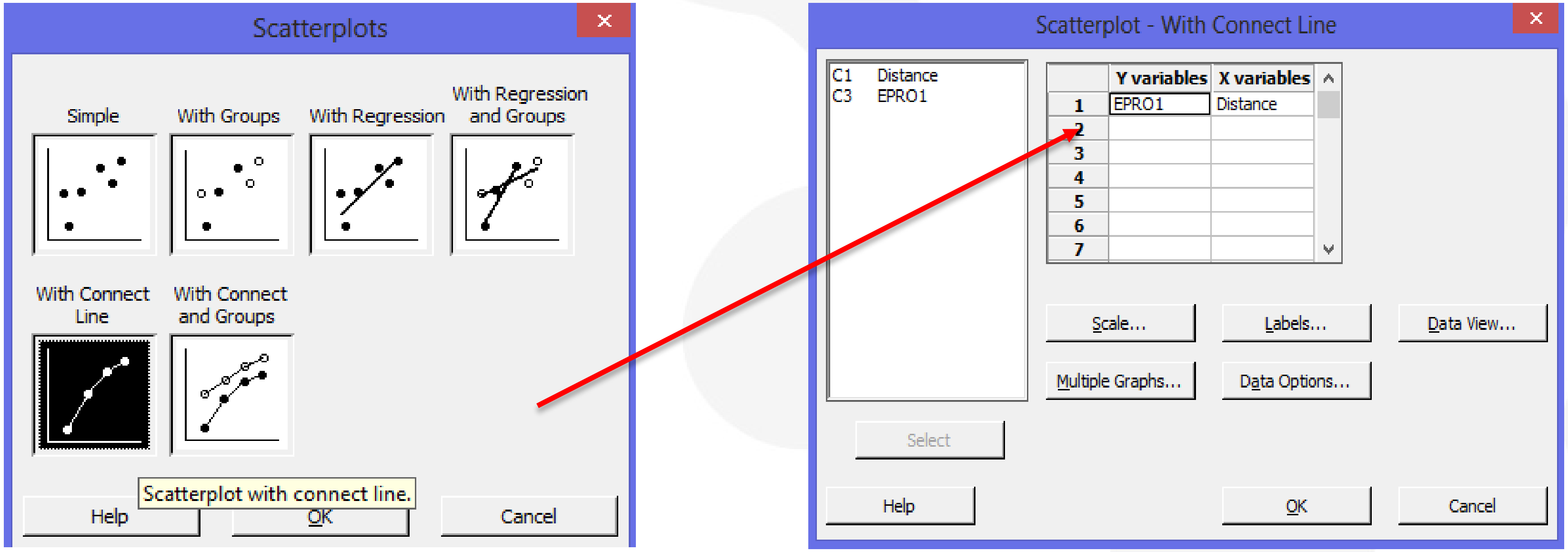
Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	3.92313	18	1.000
Deviance	4.43572	18	1.000
Hosmer-Lemeshow	0.96911	7	0.995

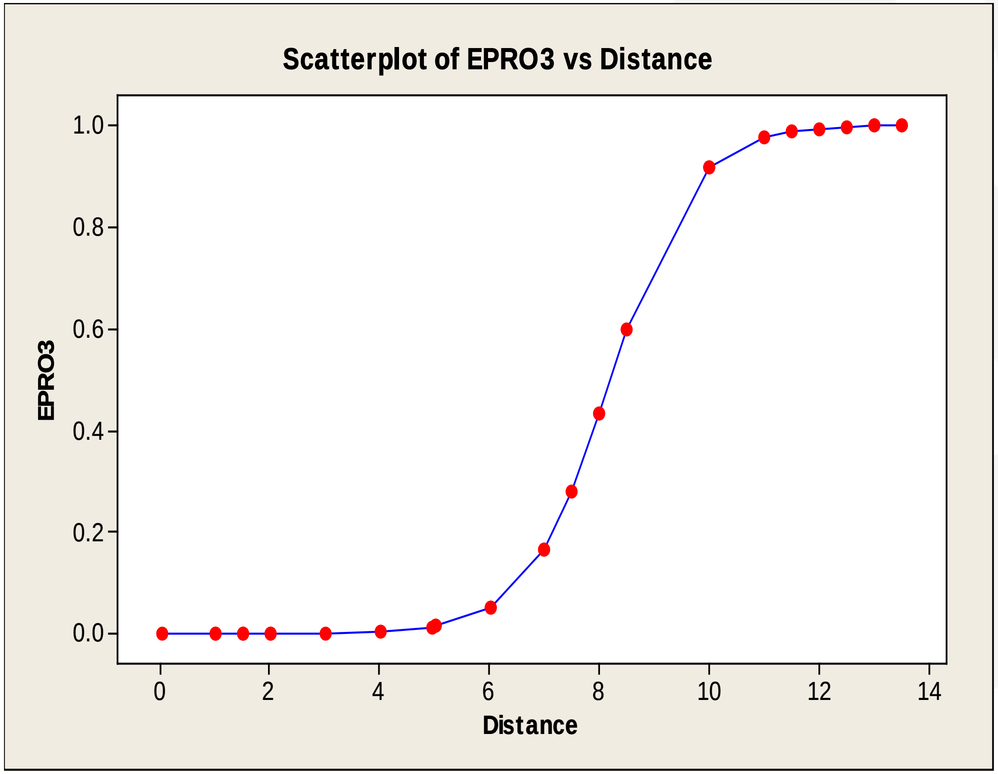
Exercise 1, Graphing the probability

Having stored the probabilities, you can now graph these to show how the impact of our continuous factor (distance) has on probability

<u>Graph > Scatterplot...</u>



Exercise 1, Graphical Output



Graphing the stored EPRO (Event probability) we now have a graph illustrating the impact our continuous factor has on the probability

Technical

Detailed Example Using Minitab

Exercise 2, Multiple Xs, Binary Y

Space Shuttle Challenger

- A Presidential Commission was interested in knowing if the Challenger Disaster could have been predicted given data available at the time of the incident in 1986. The Commission obtained data on **booster rocket o-ring** damage that existed PRIOR to the launch of Challenger. For each of the previous 23 shuttle flights, the o-rings had been inspected post flight.
- Questions to answer:
  - Do the data indicate temperature as an important factor affecting o-ring damage?
  - Is one booster side more susceptible to damage than the other?
  - If temperature is a factor, by extrapolation, what would be the probability of damage for all 3 o-rings at 31 deg F? (There are 3 o-rings in each booster rocket and all three have to fail for the rocket to fail.)

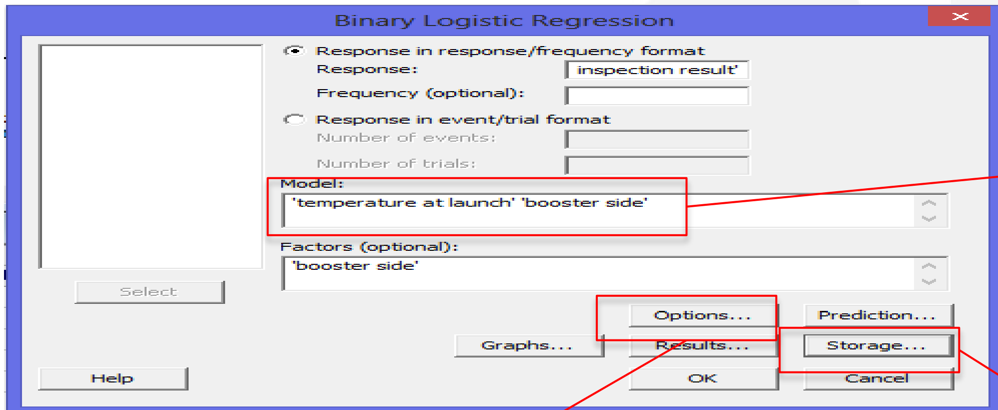
Exercise 2, Pre-work

- **What is the business question?**
  - **Is temperature a prediction of o-ring damage?**
  - **Is one booster more susceptible to damage than the other?**
  - **What is the probability of failure at 31º F?**
- **What are the X's? What type of data are the Y and X's?**
  - **X1: temperature at launch, continuous**
  - **X2: booster side, discrete**
  - **Y: o-ring inspection result, discrete**
- **Correct tool is binary logistic regression (for 3 or more outputs use nominal)**

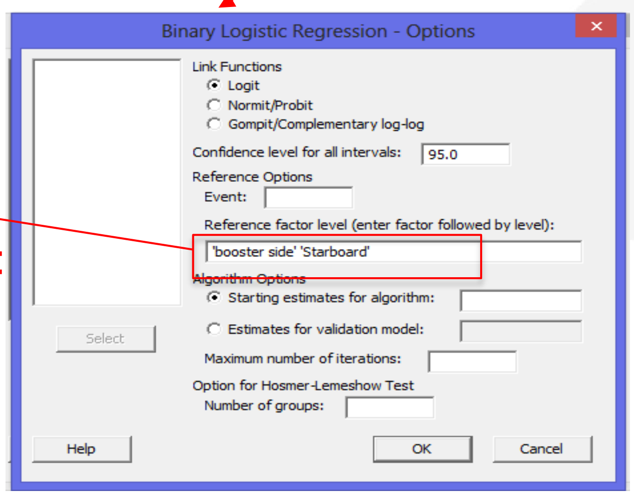
Data Table:
Logist_Regr_Ex_2_Space_Shuttle.MPJ

Exercise 2, Specify the model

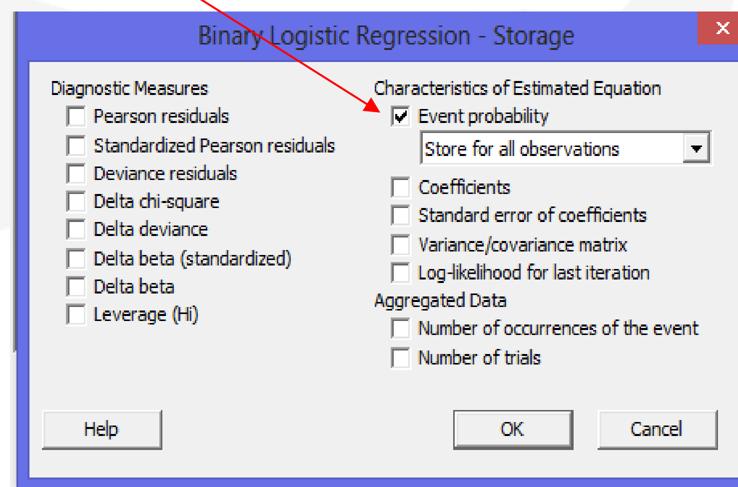
Stat > Regression > Binary Logistic Regression



Remember to add any discrete factors into this section too!



Specify the discrete reference factor level as: 'booster side' 'Starboard'



Exercise 2, Session window output

Is the regression model significant?

```
Log-Likelihood = -29.484
Test that all slopes are zero: G = 7.707, DF = 2,
P-Value = 0.021
```

- **What is the null and alternate hypothesis?**
- **H_0:**
- **H_a:**
- **What decision do we make?**

Exercise 2, Session window output

Do the data fit the model?

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	16.0694	29	0.975
Deviance	16.8002	29	0.965
Hosmer-Lemeshow	6.4024	7	0.494

- **What is the null and alternate hypothesis?**
- **H_0:**
- **H_a:**
- **What decision do we make?**

Which of the predictors are significant?

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-5.60157	3.05823	-1.83	0.067			
temperature at launch	0.118185	0.0466258	2.53	0.011	1.13	1.03	1.23
booster side							
Port	0.784925	0.751161	1.04	0.296	2.19	0.50	9.56

- **What is the null and alternate hypothesis?**
- **H_0:**
- **H_a:**
- **What decision do we make?**

Exercise 2, Session window output

Reduce the Model: Rerun without 'booster side' as an X.

Does all still look good?

Response Information

Variable	Value	Count	
o-ring inspection result	good	130	(Event)
	damaged	9	
	Total	139	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	Upper
Constant	-5.18042	2.99088	-1.73	0.083			
temperature at launch	0.116823	0.0460293	2.54	0.011	1.12	1.03	1.23

Log-Likelihood = -30.061

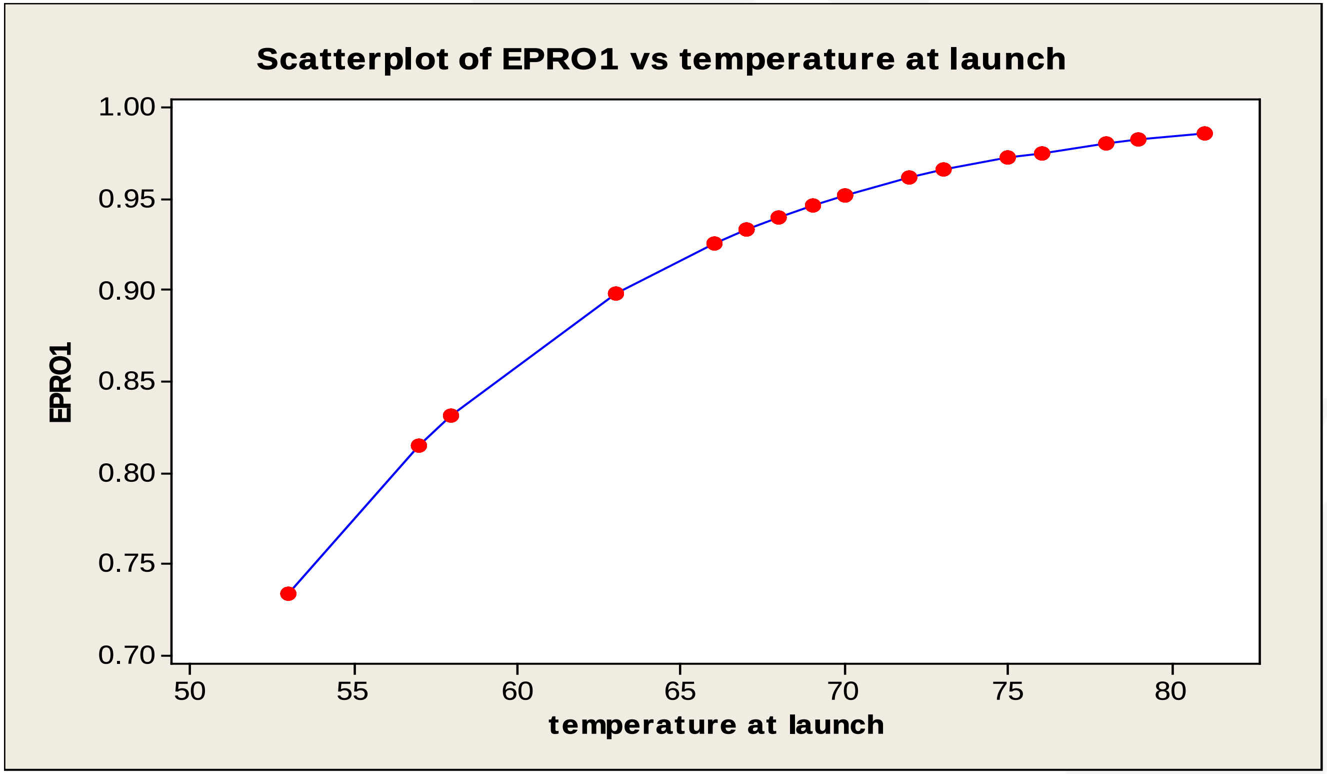
Test that all slopes are zero: G = 6.554, DF = 1, P-Value = 0.010

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	8.87647	14	0.839
Deviance	9.60189	14	0.791
Hosmer-Lemeshow	3.39879	5	0.639

Review the Scatterplot

We can see that the temperature at launch site has a clear impact on the probability of the event (a good o ring result)



Can we use this information to model what the result will be at 31degrees?

Exercise 2, Session window output

Look at our model to predict the result:

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-5.18042	2.99088	-1.73	0.083			
temperature at launch	0.116823	0.0460293	2.54	0.011	1.12	1.03	1.23

Remember that: $$\text{Probability of Event} = \frac{1}{1+e^{-(B_0+B_1X)}}$$

Using the predictive model above, we can predict the probability of damage to an o-ring:

$$EPRO(31) = \frac{1}{1 + e^{-(-5.18+(0.11682 \times 31))}} = \frac{1}{1 + 2.7182^{-(1.55858)}} = 0.826$$

At 31 degrees, the probability of a damage to all three o-rings and a catastrophic result is:

0.826 X 0.826 X 0.826 = 0.56 OR 56%!!!

Sample Size Considerations

- **An adequate sample size is needed to fit any regression model.**
- **Because of the nature of the response, discrete, there is not any proven sample size calculations for logistic regression and the ones embedded in some software are special. Sample size is determined by the count of each outcome in the data set, not the total number of observations.**
- **A practitioner's rule of thumb is the model cannot have more predictors than the lowest count for any outcome, divided by 10:**

$$\text{predictors} \leq \frac{\min(n_1, n_0)}{10}$$

- **In the Space Shuttle Exercise, we observed 9 damaged and 130 good. Therefore we would have to be just short of one predictor:**

$$\frac{\min(9{,}130)}{10} \leq \frac{9}{10} \leq .9 \text{ or } 1$$

Exercise 3, Credit Card Marketing

- A marketing department is planning a campaign to get card holders to upgrade their card which will provide certain benefits, but also increase the annual fee. Rather than mail the offer to the entire customer base, they would like to target the mailing to those most likely to upgrade.

- We have a sample of the data from last year's campaign which includes whether the customer upgraded, what their annual charges were (given in M$ annually), and whether our records indicated that they had an additional credit card.

- Analyse the data and determine what criteria should be used to target the mailing?

Data Table: Logist_Regr_Ex_3.MPJ

Exercise 3, Prework

Q: What is the business question?

A: Which criteria should be used to increase sales rate?

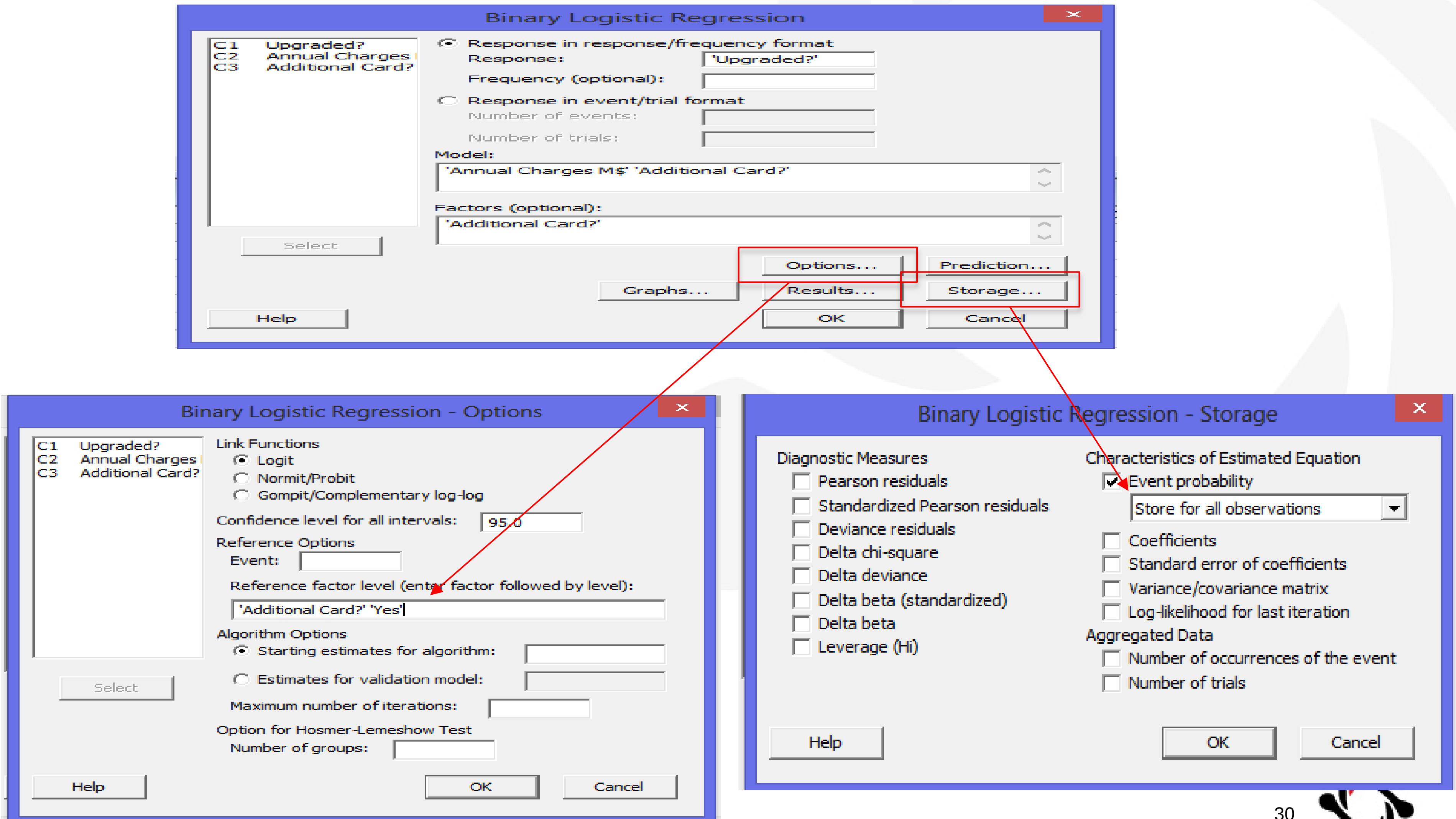
What are the potential Xs? What type of data are the Y and the Xs?

X1: Annual Charges $m, continuous
X2: Additional Card, discrete
Y: Whether the customer upgraded, discrete

Correct tool is Binary Logistic Regression

Exercise 3, Specify the model

Stat > Regression > Binary Logistic Regression



Exercise 3, Session window output

Is the regression model significant?

Do the data fit the model?

Which of the predictors are significant?

- **What does this tell us?**
- **What is our next step?**

Response Information

Variable	Value	Count	
Upgraded?	Yes	26	(Event)
	No	34	
	Total	60	

Logistic Regression Table

					Odds	95% CI	
Predictor	Coef	SE Coef	Z	P	Ratio	Lower	Upper
Constant	-4.28105	1.81869	-2.35	0.019			
Annual Charges M$	0.353471	0.118688	2.98	0.003	1.42	1.13	1.80
Additional Card?							
No	-2.92599	0.856194	-3.42	0.001	0.05	0.01	0.29

Log-Likelihood = -20.081
Test that all slopes are zero: G = 41.946, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	34.5485	49	0.941
Deviance	34.6168	49	0.940
Hosmer-Lemeshow	12.2585	8	0.140

Exercise 3, Session window output

As well as using the predictive equation, we can also look at Odds Ratios for the predictors.

Logistic Regression Table

					Odds	95% CI	
Predictor	Coef	SE Coef	Z	P	Ratio	Lower	Upper
Constant	-4.28105	1.81869	-2.35	0.019			
Annual Charges M$	0.353471	0.118688	2.98	0.003	1.42	1.13	1.80
Additional Card?							
No	-2.92599	0.856194	-3.42	0.001	0.05	0.01	0.29

- **As Annual Charges go up by $1,000, how much more likely is the customer to upgrade?**
- **How much less likely is a customer with an additional card to upgrade than a customer without an additional card?**

Summary of key learning points

- **Logistic regression is used to model and predict the outcome of a discrete/categorical output with one or more continuous inputs**

- **Binary Logistic Regression should be used when you have two potential outputs eg, Pass/Fail, Good/Bad**

Recommended coach support points

- Understanding odds ratios
- Developing a predictive equation
- Checking your sample size