



COURSE REFERENCE

MiC Quality Online Courses

FREE ONLINE MODULE

“INTRODUCTION TO STATISTICS”

http://www.micquality.com/introductory_statistics/

www.micquality.com

STATS

PRIMER IN STATISTICS

PRIMER IN STATISTICS



This publication belongs to a series of **Course Reference** booklets that accompany our online courses. This booklet summarizes material covered in our **Primer in Statistics** online course.

For information about all our courses and **free Six Sigma resources** visit our Web site at www.micquality.com:

- [Free Online Module](#) "Introduction to Statistics"
- [Free Glossary](#) with **500+** Six Sigma and statistical terms
- [Free Online Excel Primer](#)
- [Reference tables, Six Sigma and Normal Distribution calculators](#)

Glen Netherwood, MiC Quality

CONTENTS

Useful Statistical Measures	1
Normal Distribution	2
Useful Values from the Normal Distribution	3
Six Sigma 'Sigma Values'	3
Histograms and Pareto Charts	4
Stem and Leaf Plots	5
Scatter Graphs	6
Correlation	6
Multi-Vari Charts	7
Box Plots	8
Grouped Data	9
Percentiles	9
Normal Probability Plots	10
Standard Error	11
Confidence Intervals	12
Hypothesis Tests	13
The p-Value	14
One & Two Sided Tests	15
Using z-Values	16
Type I & II Errors and Power	17
Sample Size	18
Other Hypothesis Tests	19
Sampling Methods	20
Measurement Scales	21

APPENDICES

Appendix 1 - Normal Distribution Tables

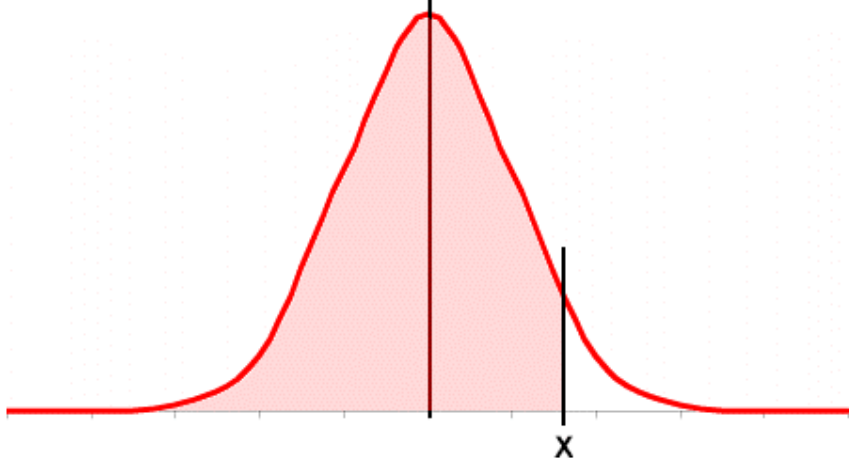
OUR STUDENTS SAY: Jennifer McClare, Engineer, Canada

"Very practical, lots of examples, easy to understand. Rather than just a review of math, the course was very applied with a number of very practical real-world examples. It showed me that I already knew enough to be making improvements in processes, but just didn't know how to apply it. The email support was very thorough and contained personal responses, not "canned" answers; individual attention was at least as good as in a classroom setting, if not better."

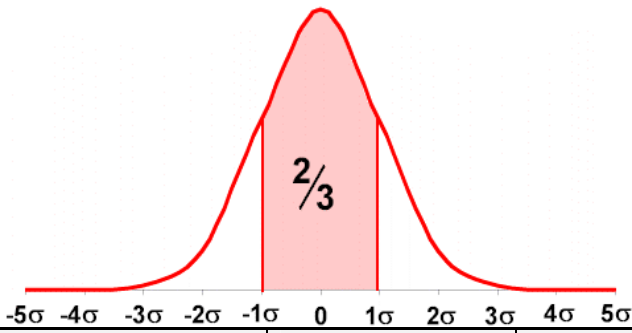
USEFUL STATISTICAL MEASURES

MEASURE	DESCRIPTION	FORMULA	EXCEL FUNCTION
Mean/ Average	The sum of the sample values divided by the sample size. The sample statistic \bar{x} is an unbiased estimate of the population mean μ .	$\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}$	=AVERAGE(data_range)
Median	The middle value after the sample values have been sorted into order by magnitude. If there are an even number of values in the sample, the average of the two middle values.		=MEDIAN(data_range)
Mode	The most common value in the sample.		=MODE(data_range)
Range	The difference between the largest and smallest values in the sample.		=MAX(data_range) - MIN(data_range)
Variance	An estimate of the variation or dispersion of the process from which the sample was drawn. The sample statistic 's ² ' is an unbiased estimator of the population parameter σ^2 .	$s^2 = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n - 1}$	=VAR(data_range)
Standard Deviation	The square root of the variance. Often preferred as a measure of process variation. The sample statistic 's' is an estimator of the population parameter ' σ '. This method of calculating the standard deviation is known as the Root Mean Square Error (RMSE) method.	$s = \sqrt{\frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}{n - 1}}$	=STDEV(data_range)

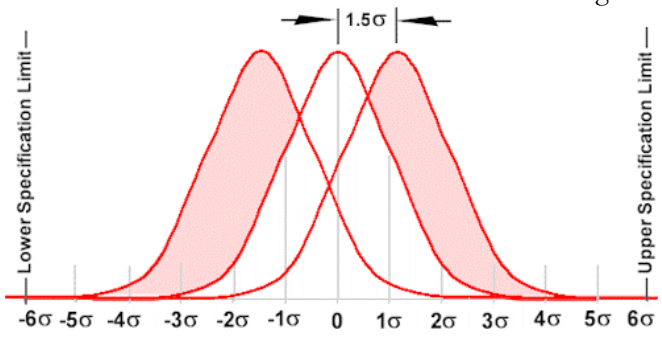
NORMAL DISTRIBUTION

<p>DESCRIPTION</p>	<p>The output of many types of processes can be represented by the normal distribution. The normal distribution is often used to estimate the proportion of process output that will lie within a specific range of values (for example, the proportion of the process output that will be within specification).</p> <p>The normal distribution shape results from the ‘common cause’ variation in a process. It can be used to reveal the presence of ‘special causes’.</p>
<p>DISTRIBUTION</p>	
<p>EXCEL FUNCTION</p>	<p>The proportion of the process output smaller than ‘X’ can be found in Excel using:</p> <p>=NORMDIST (X, Mean, Standard Deviation, TRUE)</p>
<p>TABLES</p>	<p>The proportion of the process output smaller than ‘X’ can be also found from Normal Distribution Tables (see Appendix 1).</p>
<p>Z-SCORE FORMULA</p>	<p>The z-score is calculated from:</p> $z = \frac{x - \mu}{\sigma}$ <p>‘μ’ process mean, estimated from the sample mean ‘\bar{x}’ ‘σ’ process standard deviation, estimated from the sample standard deviation ‘s’</p>
<p>EXAMPLE</p>	<p>A process produces bars with a mean length of 50mm and a standard deviation of 5, what proportion of the bars will be shorter than 64mm?</p> <p>Answer: The z-score is 2.80. From tables find that 0.9974 (99.74%) of the bars will be shorter than 64mm.</p>

USEFUL VALUES FROM THE NORMAL DISTRIBUTION

DISTRIBUTION	<p>About two thirds of the process output lie within one standard deviation either side of the process mean. Other values to remember:</p> 			
VALUES	Standard Deviations either side of the mean	Approximate amount	Exact amount	
	±1	two-thirds	68.27%	
	±2	95%	95.45%	
	±3	99.7%	99.73%	

SIX SIGMA 'SIGMA VALUES'

SIX SIGMA	<p>The 'six sigma' approach uses the 'sigma value' to measure the number of DPMO (Defects per Million Opportunities). The calculation assumes that the process mean is 1.5 standard deviations from the target.</p> 			
VALUES (ppm – parts per million)	Sigma Level	Yield (%)	Defective ppm	
	1	30.23	697700	
	2	69.13	308700	
	3	93.32	66810	
	4	99.3790	6210	
	5	99.97670	233	
	6	99.9999660	3.4	

HISTOGRAMS AND PARETO CHARTS

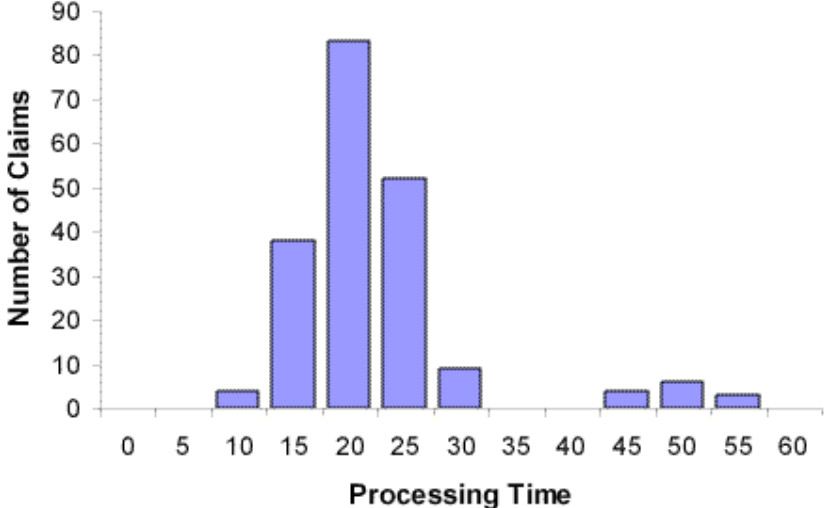
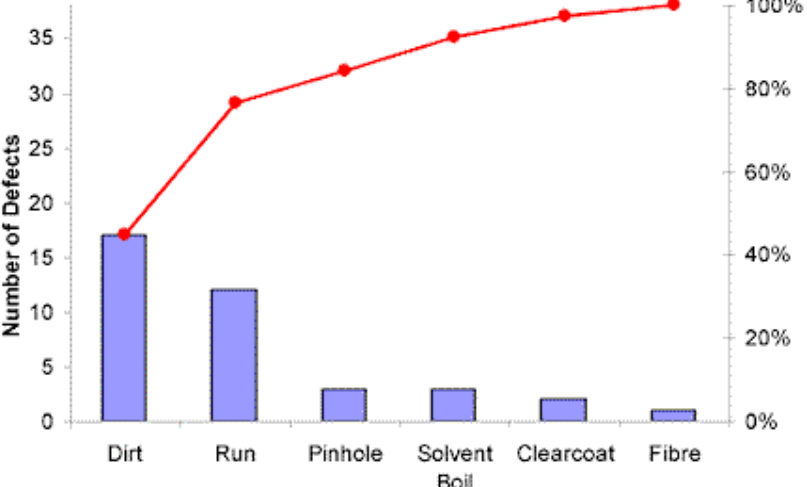
CHART TYPE	Frequency Histogram																		
DESCRIPTION	A useful way to represent the distribution of a set of data.																		
EXAMPLE	<p>The example shows processing time for insurance claims.</p>  <table border="1"> <caption>Data for Frequency Histogram Example</caption> <thead> <tr> <th>Processing Time Range</th> <th>Number of Claims</th> </tr> </thead> <tbody> <tr><td>10-15</td><td>5</td></tr> <tr><td>15-20</td><td>38</td></tr> <tr><td>20-25</td><td>85</td></tr> <tr><td>25-30</td><td>52</td></tr> <tr><td>30-35</td><td>10</td></tr> <tr><td>45-50</td><td>5</td></tr> <tr><td>50-55</td><td>7</td></tr> <tr><td>55-60</td><td>4</td></tr> </tbody> </table>	Processing Time Range	Number of Claims	10-15	5	15-20	38	20-25	85	25-30	52	30-35	10	45-50	5	50-55	7	55-60	4
Processing Time Range	Number of Claims																		
10-15	5																		
15-20	38																		
20-25	85																		
25-30	52																		
30-35	10																		
45-50	5																		
50-55	7																		
55-60	4																		

CHART TYPE	Pareto Chart																					
DESCRIPTION	Used to sort the 'significant few' from the 'trivial many'. Often combines a sorted histogram and a cumulative graph.																					
CHART EXAMPLE	<p>The example shows the frequency of occurrence of types of paint defects on an automotive part.:</p>  <table border="1"> <caption>Data for Pareto Chart Example</caption> <thead> <tr> <th>Defect Type</th> <th>Number of Defects</th> <th>Cumulative Percentage</th> </tr> </thead> <tbody> <tr><td>Dirt</td><td>17</td><td>42.3%</td></tr> <tr><td>Run</td><td>12</td><td>54.3%</td></tr> <tr><td>Pinhole</td><td>3</td><td>57.6%</td></tr> <tr><td>Solvent Boil</td><td>3</td><td>60.6%</td></tr> <tr><td>Clearcoat</td><td>2</td><td>62.3%</td></tr> <tr><td>Fibre</td><td>1</td><td>63.5%</td></tr> </tbody> </table>	Defect Type	Number of Defects	Cumulative Percentage	Dirt	17	42.3%	Run	12	54.3%	Pinhole	3	57.6%	Solvent Boil	3	60.6%	Clearcoat	2	62.3%	Fibre	1	63.5%
Defect Type	Number of Defects	Cumulative Percentage																				
Dirt	17	42.3%																				
Run	12	54.3%																				
Pinhole	3	57.6%																				
Solvent Boil	3	60.6%																				
Clearcoat	2	62.3%																				
Fibre	1	63.5%																				

STEM AND LEAF PLOTS

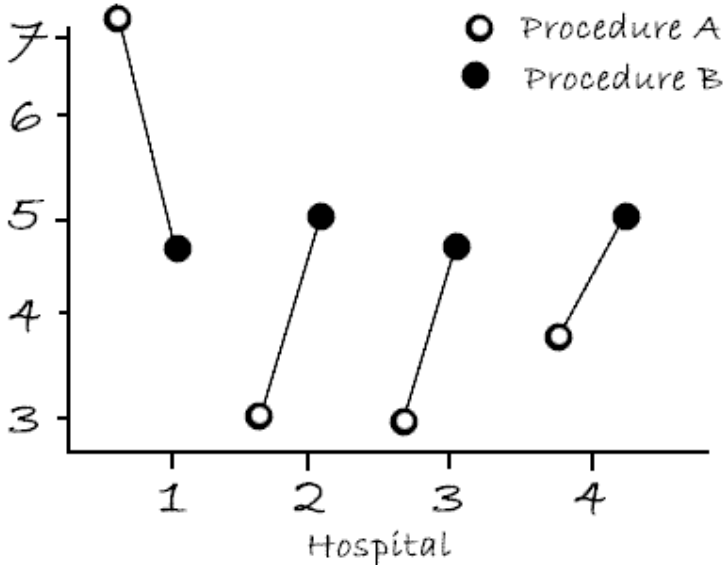
CHART TYPE	Stem and Leaf Plots																																																													
DESCRIPTION	Stem and Leaf Plots are similar to histograms, but retain some, or all, of the original data values.																																																													
EXAMPLE	<p>The following values:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tbody> <tr><td>38</td><td>10</td><td>60</td><td>90</td><td>88</td></tr> <tr><td>96</td><td>1</td><td>41</td><td>86</td><td>14</td></tr> <tr><td>25</td><td>5</td><td>3</td><td>16</td><td>22</td></tr> <tr><td>2</td><td>29</td><td>34</td><td>55</td><td>36</td></tr> <tr><td>37</td><td>36</td><td>91</td><td>47</td><td>43</td></tr> </tbody> </table> <p>Would form a stem and leaf plot:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Frequency</th> <th>Stem</th> <th>Leaves</th> </tr> </thead> <tbody> <tr><td>5</td><td>0</td><td>1235</td></tr> <tr><td>3</td><td>1</td><td>046</td></tr> <tr><td>3</td><td>2</td><td>259</td></tr> <tr><td>5</td><td>3</td><td>46678</td></tr> <tr><td>3</td><td>4</td><td>137</td></tr> <tr><td>1</td><td>5</td><td>5</td></tr> <tr><td>1</td><td>6</td><td>0</td></tr> <tr><td>0</td><td>7</td><td></td></tr> <tr><td>2</td><td>8</td><td>68</td></tr> <tr><td>3</td><td>9</td><td>016</td></tr> <tr><td>0</td><td>10</td><td></td></tr> </tbody> </table> <p>There are two values with a stem of '8', the values 86 and 88. These go into the row with stem '8'; the leaves are '6' and '8'.</p>	38	10	60	90	88	96	1	41	86	14	25	5	3	16	22	2	29	34	55	36	37	36	91	47	43	Frequency	Stem	Leaves	5	0	1235	3	1	046	3	2	259	5	3	46678	3	4	137	1	5	5	1	6	0	0	7		2	8	68	3	9	016	0	10	
38	10	60	90	88																																																										
96	1	41	86	14																																																										
25	5	3	16	22																																																										
2	29	34	55	36																																																										
37	36	91	47	43																																																										
Frequency	Stem	Leaves																																																												
5	0	1235																																																												
3	1	046																																																												
3	2	259																																																												
5	3	46678																																																												
3	4	137																																																												
1	5	5																																																												
1	6	0																																																												
0	7																																																													
2	8	68																																																												
3	9	016																																																												
0	10																																																													

SCATTER GRAPHS & CORRELATION

CHART TYPE	Scatter Graphs
DESCRIPTION	Scatter graph are used to explore associations between two variables. They are normally used when the data form natural pairs, and where there are many pairs. Standard X-Y graphs are normally used to explore a causal relationship between X and Y; typically X is varied systematically and the effect on Y measured.
EXAMPLE	
COMMENTS	Female Literacy and Birth Rate are associated, but there is not necessarily a causal relationship.

Correlation	
DESCRIPTION	Correlation is a measure of the strength of the relationship between the input and the output of a process. Correlation is measured by the 'Pearson Product Moment Correlation', known as 'R'. The value of 'R' varies from +1 to -1.
EXAMPLES	
COMMENTS	An R value of + 1 is perfect correlation. Values between -0.5 and + 0.5 show weak relationships.

MULTI-VARI CHARTS

CHART TYPE	Multi-Vari Charts																											
DESCRIPTION	<p>Multi-vari charts are a graphical method used to show the effect of more than one input variable on the output, including any interaction between the input variables.</p> <p>This is illustrated by an example. The table shows the Average Length of Stay (LOS) for two procedures, 'A' and 'B' at four different hospitals.</p> <table border="1" style="margin: 10px auto;"> <thead> <tr> <th>Hospital</th> <th>Procedure</th> <th>LOS</th> </tr> </thead> <tbody> <tr><td>1</td><td>A</td><td>7.25</td></tr> <tr><td>1</td><td>B</td><td>4.75</td></tr> <tr><td>2</td><td>A</td><td>3.00</td></tr> <tr><td>2</td><td>B</td><td>5.00</td></tr> <tr><td>3</td><td>A</td><td>3.00</td></tr> <tr><td>3</td><td>B</td><td>4.75</td></tr> <tr><td>4</td><td>A</td><td>3.75</td></tr> <tr><td>4</td><td>B</td><td>5.00</td></tr> </tbody> </table>	Hospital	Procedure	LOS	1	A	7.25	1	B	4.75	2	A	3.00	2	B	5.00	3	A	3.00	3	B	4.75	4	A	3.75	4	B	5.00
Hospital	Procedure	LOS																										
1	A	7.25																										
1	B	4.75																										
2	A	3.00																										
2	B	5.00																										
3	A	3.00																										
3	B	4.75																										
4	A	3.75																										
4	B	5.00																										
EXAMPLE																												
COMMENTS	<p>The chart clearly shows that Procedure A at Hospital 1 is the odd one out. Multi-vari charts can show the effects of more than two input variables. If there were three input variables, there would be a separate chart for each level of the third variable.</p>																											

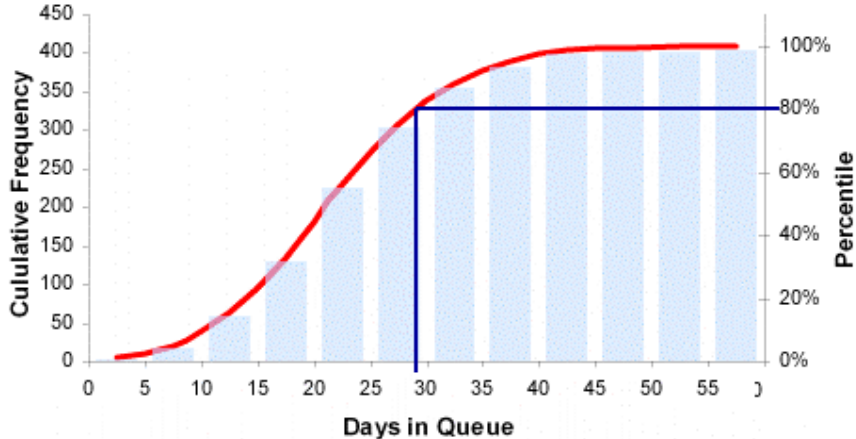
BOX PLOTS

<p>CHART TYPE</p>	<p>Box Plot</p>
<p>DESCRIPTION</p>	<p>A graphical method for representing a data set.</p> <div style="text-align: center;"> </div> <ul style="list-style-type: none"> • outliers are smaller than $Q1 - 1.5 \times (Q3 - Q1)$ or greater than $Q3 + 1.5 \times (Q3 - Q1)$ • whiskers are the largest and smallest data values that are not outliers
<p>EXAMPLE</p>	<p>The example shows a box plot used to compare the strength of five mixes of concrete:</p> <div style="text-align: center;"> </div>

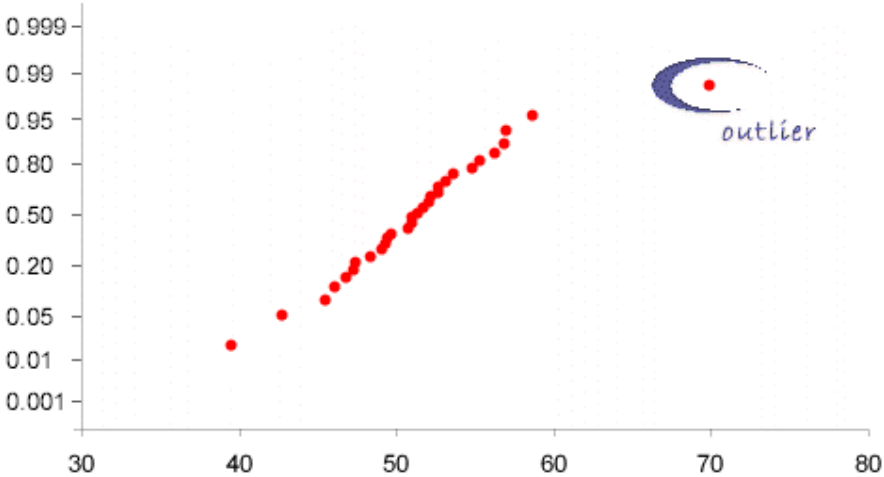
GROUPED DATA

MEASURE	DESCRIPTION	FORMULA
Mean of Grouped Data	Used where data are collected in ranges of values, rather than individual values (e.g., census data). The number of individual data values would typically be large.	$\bar{x} = \frac{\sum_{j=1}^k f_j M_j}{n}$ <p> f_j - frequency for group 'j' M_j - midpoint of the range (smallest + largest)/2 n - the total number of samples k - the number of groups </p>
Standard Deviation of Grouped Data		$s = \sqrt{\frac{\sum_{j=1}^k f_j (M_j - \bar{x})^2}{n - 1}}$

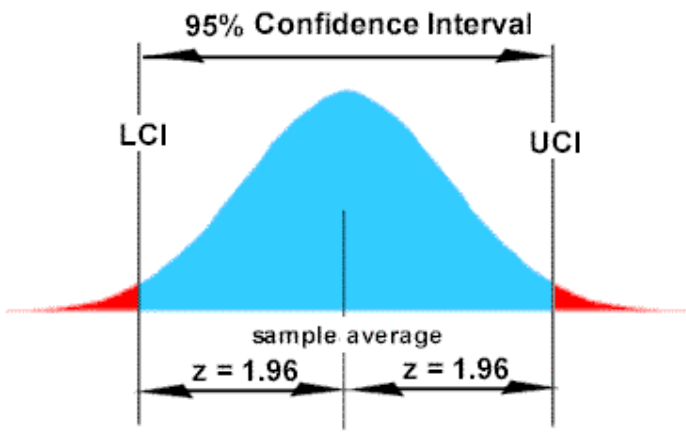
PERCENTILES

DESCRIPTION	Percentiles show the cumulative frequency and are often associated with grouped data.
EXAMPLE	<p>The chart shows how a frequency 'ogive' is created from grouped data (displayed as a cumulative histogram). It also shows, as an example, how the number of days corresponding to the 80th percentile is obtained.</p> 

NORMAL PROBABILITY PLOTS

CHART TYPE	Normal Probability Plot								
DESCRIPTION	A type of graph used to check if a sample conforms to a normal distribution, or to identify values that do not conform to a normal distribution (outliers).								
EXAMPLE	 <p>The 'y' scale of the normal probability plot is non-linear. It can be created using special graph paper or can be created in Excel using the function '=NORMSINV' as a transform as follows:</p> <table border="1" data-bbox="544 1465 1339 1581"> <thead> <tr> <th>Index</th> <th>F_i</th> <th>Y</th> <th>Ordered Data (x)</th> </tr> </thead> <tbody> <tr> <td>i</td> <td>$F_i = \frac{i-0.5}{n}$</td> <td>=NORMSINV(F_i)</td> <td>x_i</td> </tr> </tbody> </table> <p>Index values run from 1 through n, where n is the number of values in the sample.</p>	Index	F_i	Y	Ordered Data (x)	i	$F_i = \frac{i-0.5}{n}$	=NORMSINV(F_i)	x_i
Index	F_i	Y	Ordered Data (x)						
i	$F_i = \frac{i-0.5}{n}$	=NORMSINV(F_i)	x_i						

Standard Error	
DESCRIPTION	<p>A sample is used because the sample mean is a more accurate estimate of the population mean than a single value.</p> <p>The variation between sample means is less than the variation between individuals:</p>
STANDARD ERROR	<p>The standard deviation of the sample means is known as the ‘standard error’ and is calculated using the equation:</p> $\sigma_{\text{sample averages}} = \frac{\sigma_{\text{individuals}}}{\sqrt{n}}$
EXAMPLE	<p>The standard deviation of the characteristic is known to be 20. If a sample of four items is used then the standard error is:</p> $\sigma_{\text{sample averages}} = \frac{20}{\sqrt{4}} = 10$
CENTRAL LIMIT THEOREM	<p>The central limit theorem states that the distribution of sample averages will tend to conform to a normal distribution, no matter what the shape of the population distribution. The greater the sample size the greater this tendency.</p>

Confidence Intervals					
DESCRIPTION	The confidence interval is the span about the sample mean that contains the process mean to a defined probability.				
95% INTERVAL	<p>The 95% confidence interval is most often used as the basis for decision making:</p> 				
FORMULA	$\bar{x} - Z \times \sigma_{\text{sample averages}} \leq \mu \leq \bar{x} + Z \times \sigma_{\text{sample averages}}$ <p>This may also be written:</p> $\bar{x} - z \times \frac{\sigma_{\text{individuals}}}{\sqrt{n}} \leq \mu \leq \bar{x} + z \times \frac{\sigma_{\text{individuals}}}{\sqrt{n}}$				
EXAMPLE	<p>A characteristic is known to have a standard deviation of 20. A sample of four gives values:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td style="text-align: center;">512</td> <td style="text-align: center;">488</td> <td style="text-align: center;">507</td> <td style="text-align: center;">525</td> </tr> </table> <p>The mean equals 508. The 95% confidence limits are:</p> $508 - 1.96 \times 10 \leq \mu \leq 508 + 1.96 \times 10$ $488.4 \leq \mu \leq 527.6$	512	488	507	525
512	488	507	525		

Hypothesis Tests							
DESCRIPTION	Hypothesis tests are widely used as the basis for decision making. This primer looks at one specific type of test known as the z-test.						
HYPOTHESIS	<p>A hypothesis Test consists of two mutually exclusive statements:</p> <table border="1" style="margin-left: 40px;"> <tr> <td style="padding: 5px;">Null Hypothesis</td> <td style="padding: 5px;">H_0</td> <td style="padding: 5px;">a statement that makes a specific assumption about the parameter being studied</td> </tr> <tr> <td style="padding: 5px;">Alternative Hypothesis</td> <td style="padding: 5px;">H_a</td> <td style="padding: 5px;">a statement that contradicts the null hypothesis</td> </tr> </table> <p>The null hypothesis often reflects the status quo, or default assumption.</p>	Null Hypothesis	H_0	a statement that makes a specific assumption about the parameter being studied	Alternative Hypothesis	H_a	a statement that contradicts the null hypothesis
Null Hypothesis	H_0	a statement that makes a specific assumption about the parameter being studied					
Alternative Hypothesis	H_a	a statement that contradicts the null hypothesis					
EXAMPLE	<table border="1" style="margin-left: 40px;"> <tr> <td style="padding: 5px;">H_0</td> <td style="padding: 5px;">the mean strength equals 500</td> </tr> <tr> <td style="padding: 5px;">H_a</td> <td style="padding: 5px;">the mean strength does not equal 500</td> </tr> </table>	H_0	the mean strength equals 500	H_a	the mean strength does not equal 500		
H_0	the mean strength equals 500						
H_a	the mean strength does not equal 500						
p-VALUE	<p>The p-value is the probability of getting the observed data, or even more extreme results if the null hypothesis is true.</p> <p>NOTE: Do not fall into the common trap of believing that the p-value is the probability that the null hypothesis is true.</p>						
ALPHA VALUE	<p>This is known as the ‘level of significance’ and is the threshold probability for the p-value.</p> <p>If the p-value is less than the alpha value then:</p> <ul style="list-style-type: none"> • reject the null hypothesis and accept the alternative hypothesis <p>If the p-value is greater than the alpha value then:</p> <ul style="list-style-type: none"> • do not reject the null hypothesis 						
NOTE	<p>The statement ‘accept the null hypothesis’ is misleading and should be avoided because the null hypothesis is a ‘strawman’ proposition that cannot be proved, only disproved. If the p-value is greater than the alpha value it may just mean you don’t have enough data to prove that the null hypothesis is incorrect to the required level of significance.</p>						

The p-Value																
DESCRIPTION	The p-value is the probability of getting the results, or even more extreme results, if the null hypothesis is true.															
EXAMPLE	<p>The hypothesis is:</p> <table border="1" style="margin-left: 40px;"> <tr> <td style="padding: 5px;">H_0</td> <td style="padding: 5px;">the mean strength equals 500</td> </tr> <tr> <td style="padding: 5px;">H_a</td> <td style="padding: 5px;">the mean strength does not equal 500</td> </tr> </table> <p>The sample mean equals 508 and the standard error equals 10</p>	H_0	the mean strength equals 500	H_a	the mean strength does not equal 500											
H_0	the mean strength equals 500															
H_a	the mean strength does not equal 500															
FIGURE	<p style="text-align: center;"> Sample Mean (508) Hypothesized Mean (500) </p>															
INTERPRETATION	<p>The red area in the tails represents the p-value. It is the proportion of the curve that is further away from the hypothesized mean than the sample mean.</p> <p>NOTE: This is a two-sided test.</p>															
CALCULATING	<p>The p-value can be calculated using Excel. The areas in the tails are equal. It is easiest to calculate the area in the left hand tail and double it:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td colspan="5" style="text-align: center;">=NORMDIST(492,500,10,TRUE) *2</td> </tr> <tr> <td style="width: 20px;">B</td> <td style="width: 20px;">C</td> <td style="width: 20px;">D</td> <td style="width: 20px;">E</td> <td style="width: 20px;">F</td> </tr> <tr> <td></td> <td></td> <td style="text-align: center;">0.424</td> <td></td> <td></td> </tr> </table>	=NORMDIST(492,500,10,TRUE) *2					B	C	D	E	F			0.424		
=NORMDIST(492,500,10,TRUE) *2																
B	C	D	E	F												
		0.424														

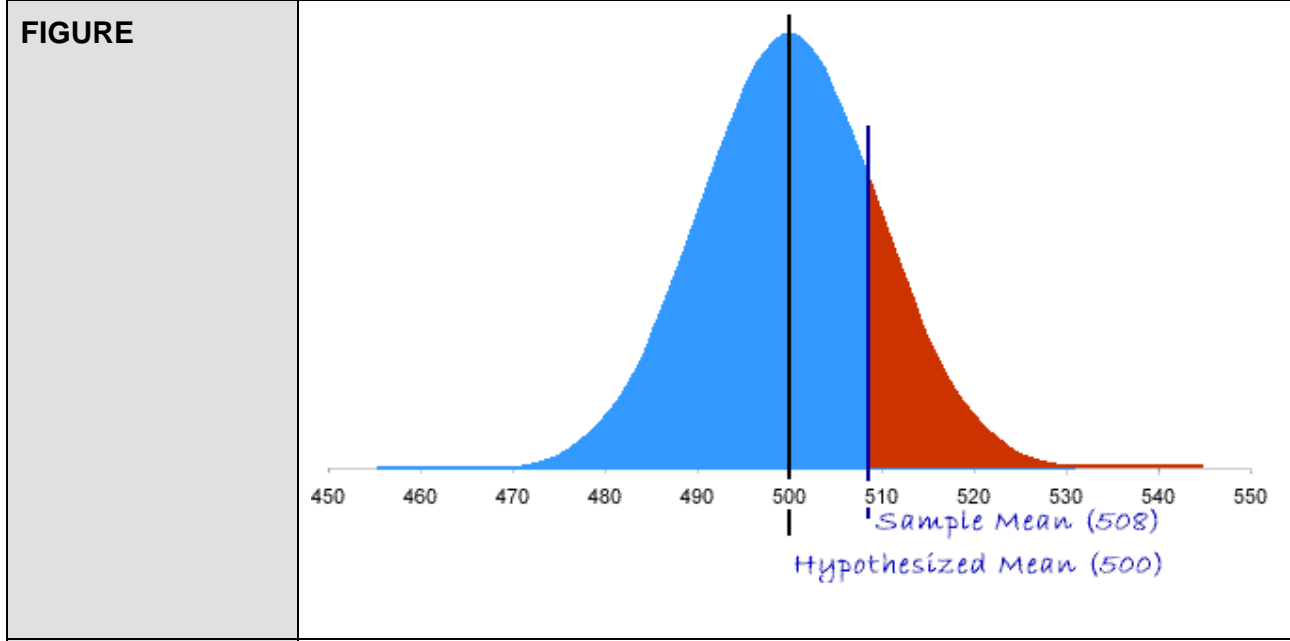
One & Two Sided Tests

DESCRIPTION Tests can be one sided or two sided. One sided tests can be either upper tail or lower tail. The example refers to an upper tail test.

EXAMPLE The hypothesis is:

H_0	the mean strength equals 500
H_a	the mean strength is greater than 500

The sample mean equals 508 and the standard error equals 10



INTERPRETATION The red area in the tails represents the p-value. It is the proportion of the curve that is greater than the sample mean.

CALCULATING The p-value can be calculated using Excel:

=1-NORMDIST(508,500,10,TRUE)			
B	C	D	E
		0.212	

Notice that the p-value for the upper tail test is half that for the two sided test given the same results.

Using z-Values										
DESCRIPTION	<p>Before computers were widely available it was necessary to calculate the value of z_0 and then look up the associated probability in normal distribution tables. Many other types of hypothesis test use a different distribution (eg. the t-distribution) in which the p-value depends on the sample size. It is not practical to have tables for every sample size.</p> <p>The workaround is to use tables that contain critical values of z_α, t_α etc. In this approach the p-value is never known and is not explicitly used.</p>									
TWO TAILED TEST										
EQUATION	$z_0 = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$									
TEST	<p>Do not reject the null hypothesis if:</p> $-z_{\alpha/2} < z_0 < z_{\alpha/2}$									
ONE TAILED TEST	<p>For a one tailed test all the 'alpha' will be in one tail and do not reject the null hypothesis if:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>$z_0 < z_\alpha$</td> <td>upper tail test</td> </tr> <tr> <td>$-z_0 < z_\alpha$</td> <td>lower tail test</td> </tr> </table>						$z_0 < z_\alpha$	upper tail test	$-z_0 < z_\alpha$	lower tail test
$z_0 < z_\alpha$	upper tail test									
$-z_0 < z_\alpha$	lower tail test									
z_α VALUES	α	0.10	0.05	0.025	0.02	0.01				
	z_α	1.28	1.65	1.96	2.05	2.33				

Type I & II Errors and Power			
DESCRIPTION	Hypothesis tests are subject to two types of error:		
	Type I Error	α	Rejecting the null hypothesis when it is actually correct
	Type II Error	β	Failing to reject the null hypothesis when it is incorrect
TYPE I ERROR	<p>The probability of a Type I Error is equal to the alpha value</p>		
TYPE II ERROR	<p>The probability of a Type II Error depends on the difference between the actual and hypothesized means</p>		
SUMMARY		Reality	
		H_0	H_a
Test Result	H_0	Conclusion Correct	Type II Error
	H_a	Type I Error	Conclusion Correct
POWER	The power is 1 minus the probability of a Type II Error: $\text{Power} = 1 - \beta$		

Sample Size	
DESCRIPTION	The design of a hypothesis test involves deciding on the difference to be detected and the power, then deciding on the necessary sample size.
STATISTICAL VS PRACTICAL SIGNIFICANCE	If the test is too sensitive it will detect differences that are statistically significant but negligible for practical purposes.
ILLUSTRATION	
	<p>The difference for a power '1 - β':</p> $\delta = (z_{\alpha/2} + z_{\beta}) \frac{\sigma}{\sqrt{n}}$ <p>The sample size to achieve a power '1 - β':</p> $n = \left(z_{\alpha/2} + z_{\beta} \right)^2 \frac{\sigma^2}{\delta^2}$ <p>δ departure from the hypothesized value $z_{\alpha/2}$ the z value corresponding to the level of significance z_{β} the z value corresponding to the Type II error probability n sample size σ standard deviation</p>
NOTE	Replace $z_{\alpha/2}$ with z_{α} for single sides tests, but use z_{β} for both double and single sided tests. Note that I have used absolute values for both 'z'
EXAMPLE	<p>If the standard deviation is 20 using alpha = 0.05 and power 0.9 to detect a departure of 10:</p> $n = (1.96 + 1.28)^2 \frac{20^2}{10^2} = 42$

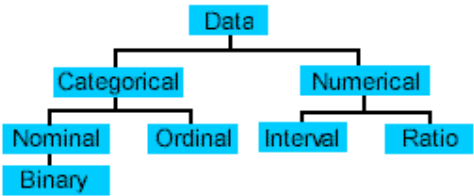
Other Hypothesis Test			
DESCRIPTION	The test used in this course is called the z-test. It is only one of a large range of hypothesis tests.		
MEANS	z-test	standard deviation known	H_0 : the means of two populations are the same
	t-test	standard deviation unknown	
	paired t	standard deviation unknown	as for the t-test but where the items in the two samples can be paired eg. before and after
	ANOVA	more than two populations involved	H_0 : the mean of at least one of the populations is different
PROPORTIONS	Binomial	exact result for small samples	H_0 : the proportion equals a standard
	Normal Approximation	larger samples $np > 10$ & $n(1-p) > 10$	H_0 : the proportion equals a standard or the proportions in two populations are the same
RATES	Poisson	exact result for small samples	H_0 : the rate equals a standard
	Normal Approximation	larger sample mean > 10	H_0 : the rate equals a standard or the rates in two populations are the same
VARIANCE	F-test		H_0 : the variances of the two populations are the same
GOODNESS OF FIT	Chi-Square test	the test uses the proportions in the sample and can be used as an alternative form of test for proportions	H_0 : the population conforms to a given distribution

Related Methods	
INTRODUCTION	Many methods used in statistics are not explicitly called hypothesis tests but implicitly use the concept.
REGRESSION ANALYSIS	Developing an equation to relate the 'response' to one or more variables. This is similar to drawing a graph except that the regression equation gives hard number with confidence intervals around the 'expected value'.
DESIGN OF EXPERIMENTS	This is similar to regression analysis but there are always multiple inputs. The experimental design uses a structure that minimizes the number of tests to be carried out but maximizes the amount of information obtained.
STATISTICAL PROCESS CONTROL	A method for the ongoing monitoring of processes. SPC aims to ignore period to period changes that are caused by normal random variation but identify changes that are significant in a timely fashion. SPC can be more effective than hypothesis testing because it identifies changes as they arise, and so problems can be fixed early. Including the effects of time based patterns also allows for more powerful analysis.

SAMPLING METHODS

Sampling Types	
SAMPLING	Inferential statistics is used to draw conclusions about a population based on a relatively small sample. The conclusions are only meaningful if the sample is representative of the population.
SIMPLE RANDOM SAMPLING	In a simple random sample, the sampling is carried out in a way that ensures that every member of the population has an equal chance of being selected. This can be done by numbering each item in the population, and then picking numbers from a hat, as in a lottery. An easier approach is to use a computer to generate random numbers.
STRATIFIED SAMPLING	This involves splitting the population into categories and then taking a random sample from each category. The size of the samples is proportional to the size of the category. Suppose a company wants to carry out a survey of employee satisfaction. A simple random sample would select employees at random. A stratified sample might select employees proportionally from each department, and level of management. In a small mixed gender group it may be appropriate to ensure that males and females are proportionally represented.
CONVENIENCE SAMPLING	A simple random sample is often impractical because some items are difficult to access, for example in products that are palletized. Sampling from the easy to access items may be acceptable if there is evidence that they are representative of the remainder of the batch.
SAMPLE HOMOGENEITY	In some types of test the sample is taken from an area that may not be homogeneous, for example a 500 gram sample of soil may represent 10 acres of land. In this case sub-samples are taken from over the area and thoroughly mixed to make sure it is representative of the plot.

MEASUREMENT SCALES

Measurement Scales	
DESCRIPTION	When you collect sample data you use a 'Measurement Scale'. The choice of scale affects the amount of information you will get from the data, and the mathematical operations that you can use with it. There are four types of measurement scale: nominal, ordinal, interval and ratio.
TAXONOMY	 <pre> graph TD Data[Data] --> Categorical[Categorical] Data --> Numerical[Numerical] Categorical --> Nominal[Nominal] Categorical --> Ordinal[Ordinal] Nominal --> Binary[Binary] Numerical --> Interval[Interval] Numerical --> Ratio[Ratio] </pre>
NOMINAL (ATTRIBUTE)	<p>This is the most basic measurement scale. Data are placed into categories that cannot be sorted into a logical order. An example would be marital status: single, married, divorced. Nominal data are also known as ‘attribute data’.</p> <p><i>arithmetic:</i> equal to <i>statistical:</i> mode</p>
ORDINAL	<p>Data are sorted into categories. The categories can be placed into a logical order, but the intervals between the categories are undefined. Often used for order of preference.</p> <p><i>arithmetic:</i> comparison (equal to, greater than, less than) <i>statistical:</i> median, mode</p>
INTERVAL	<p>Items are placed on a scale, with intervals that can be measured numerically. The numerical scale can be linear or non-linear (e.g. logarithmic). Zero does not mean the absence of the entity. Rating customer satisfaction on a scale of 1 to 10 would be an example of an interval scale. The Fahrenheit or Centigrade temperature scales are also interval scales because zero does not imply an absence of temperature.</p> <p><i>arithmetic:</i> comparison, addition, subtraction NOT multiplication, division. <i>statistical:</i> mean, median, variance</p>
RATIO	<p>Similar to an interval scale with the additional constraint that zero means the absence of the entity. Length or weights are measured on a ratio scale.</p> <p><i>arithmetic:</i> comparison addition, subtraction, multiplication, division <i>statistical:</i> mean, median, variance</p>

APPENDICES

APPENDIX 1

NORMAL DISTRIBUTION TABLES



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.40	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.30	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.20	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.10	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.00	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.90	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.80	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.70	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.60	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.50	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.40	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.30	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.20	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.10	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.50	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.40	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.20	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.00	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641



Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

MiC Quality Online Courses

SIX SIGMA PRIMER

This course provides an in-depth introduction to Six Sigma. It takes you through each stage of the DMAIC sequence using case studies to show you how, and when, to use the most important methods and tools.

TOPICS: Deployment, Six Sigma metrics, the DMAIC sequence, Lean methods, Design For Six Sigma (DFSS).

Recommended for everyone implementing Six Sigma or studying for a Black Belt.

STATISTICAL PROCESS CONTROL (SPC)

You will learn how to carry out process capability studies, use control charts effectively, and use the results to improve your processes.

TOPICS: Variation; process capability C_p , C_{pk} ; process performance P_p , P_{pk} ; X-Bar and R control charts; attribute control charts p , np , c , u .

A must for everyone involved in quality management, ISO9000, Six Sigma.

ADVANCED STATISTICAL PROCESS CONTROL

Covers 10 types of control charts for a variety of process situations, including short and high volume, as well as supporting process improvement and Six Sigma initiatives.

TOPICS: Given Standard, X-bar & s , Median, Demerits Per Unit (U), Individual and Moving Range (X_{mR}), Moving Average & EWMA, CuSum; Short Run SPC; PRE-control.

Recommended for everyone involved in quality management, ISO9000 and Six Sigma.

PRIMER IN STATISTICS

An introduction to statistics.

TOPICS: Mean, Median, Mode, Range, Variance, Standard Deviation, Normal Distribution, Testing for Normality, Grouped Data, Percentiles, Histograms, Pareto Charts, Box Plots, Stem & Leaf Plots, Multi-Vari Charts, Scatter Graphs, Correlation, Confidence Intervals, Hypothesis Testing, p-Values, Power, Calculating the Sample Size.

A must for everyone involved in Quality Management, Process Improvement and Data Analysis.

ADVANCED STATISTICS

Comprehensive coverage of the statistical methods for engineers and scientists. An excellent preparation for the American Society for Quality Six Sigma Black Belt (ASQ SSBB) exam.

TOPICS: Confidence Intervals; t-distribution; Hypothesis Testing; t-tests; Type I and II errors and Power; Chi-Square Distribution; Contingency Tables; Regression Analysis; Correlation; ANOVA; Probability; Binomial, Poisson & Hypergeometric Distributions

Recommended for scientists, engineers, Six Sigma Black Belts and Master Black Belts.

DESIGN OF EXPERIMENTS (DOE)

A practical guide for people who need to improve their processes using experimental design.

TOPICS: Full and Fractional Factorial Designs; Design Resolution; Hypothesis Testing; ANOVA; Analysis of Residuals; Screening Designs; Plackett-Burman Designs

A must for Six Sigma Black Belts and Master Black Belts; recommended for engineers.

ADVANCED DESIGN OF EXPERIMENTS

The course shows how to analyze the advanced methods of experimental design using Minitab.

TOPICS: Taguchi Signal to Noise Ratio and Taguchi designs, Response Surface Designs, "Hill climbing" approach for process optimum, Mixture Designs

A must for Six Sigma Black Belts; recommended for engineers and researchers.

MEASUREMENT SYSTEMS ANALYSIS (MSA/GAGE R&R)

SPC and DOE rely on the integrity of the measurement systems. This course provides a thorough treatment on how to evaluate and improve measurement systems.

TOPICS: Control Chart Methods; Repeatability & Reproducibility; Gage R & R Studies; Evaluating the Results; Using Minitab; ANOVA Methods; Capability, Bias, Linearity & Stability; Attribute Studies - long and short methods.

Recommended for quality managers, Six Sigma Black Belts and Master Black Belts.

"The fundamentals of statistics have been explained in a beautiful manner which makes them easy to understand." "I really appreciate your clear and practical explanations and the simulations! I've never really understood statistics very well. Your Primer was VERY helpful. I think I'm starting to understand this stuff!"

About MiC Quality

MiC Quality is a global provider of e-learning solutions. We provide online courses in the statistical methods used in quality assurance, process improvement, research and development, and **Six Sigma** programs. The courses are ideal for quality professionals, engineers, scientists, managers and supervisors who need to use statistics in their work. Our customers come from many industries including healthcare, manufacturing, biotechnology, electronics, IT, research and pharmaceuticals.

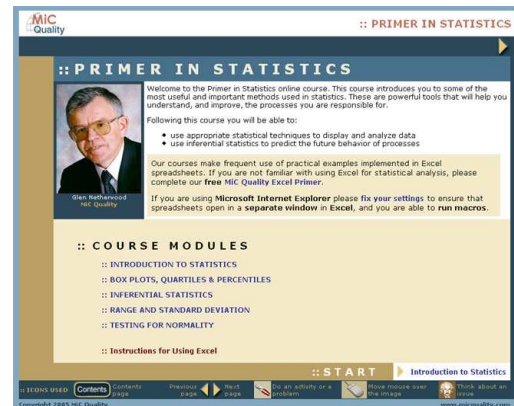
Benefits of MiC Quality online courses:

- :: **interactive** with exercises, simulations and case studies
- :: **extensive support** with individual coaching and feedback
- :: **comprehensive** with about 30 hours of in-depth study per course
- :: **flexible** self-paced learning available anywhere, anytime
- :: **effective** in developing practical knowledge and hands-on skills

MiC Quality online courses include:

- :: Six Sigma Primer
- :: Statistical Process Control (SPC)
- :: Advanced Statistical Process Control (SPC)
- :: Primer in Statistics
- :: Advanced Statistics
- :: Design of Experiments (DOE)
- :: Advanced Design of Experiments (DOE)
- :: Measurement Systems Analysis (MSA)/ Gage R&R

(see previous page for more information)



Solutions for Organizations and Individuals

Six Sigma, Process Improvement Programs

Use MiC Quality self-paced e-learning for the statistical component of Six Sigma methodology and reduce the need for expensive class-based courses.

Quality Management Systems

Provide training for group and individual staff members in statistical quality assurance methods to support **ISO9000**, **TS16949** and supplier accreditation requirements.

Professional Development for Associations

Form partnerships to provide members with professional development at discounted prices.

Career Development for Individuals

Support your professional career development, including the American Society for Quality qualifications of **Certified Quality Engineer (CQE)**, **Six Sigma Black Belt (SSBB)** and **Green Belt (SSGB)**.

Licenses and Partnerships

Site and group licenses are available. We welcome opportunities for new partnerships. Please **contact** us for more information.



Dragos Gabriel Marin

Analyst, Pratt & Whitney

"When I started the course my experience in statistics was a very traumatizing course at the university plus a number of unsuccessful

attempts of studying SPC from books. Now, at the end of the course, I can say that yes, **I understand the concepts, and I will apply them.** It is very well designed, the **e-mail support is excellent**, and it is **affordable.**"

- :: Try our **FREE** module
- :: Take our **FREE** Excel Primer
- :: **ENROLL** in our courses

Go online to:

www.micquality.com

Contact Information

MiC Quality

The Hill, 44 Sprys Lane,
PO Box 655, Hurstbridge
VIC 3099 Australia

Web: www.micquality.com

Freecall in the US: **1-888-240-5504**

Tel in Australia: **+61 3 9718 1112**

Fax in Australia: **+61 3 9718 1113**

Fax in the US: **(815) 846 8487**