MATU9D2: PRACTICAL STATISTICS

Spring 2017

PRACTICAL SESSION 3

In this practical you will learn to:

1. Examine whether the data is Normally distributed
2. Perform a One Sample Z test and calculate the associated Confidence Interval.
4. Construct hypotheses for the above test.
5. Interpret the results of all of the above tests using both CI's and p values.

THERE ARE TWO SECTIONS IN THESE NOTES:

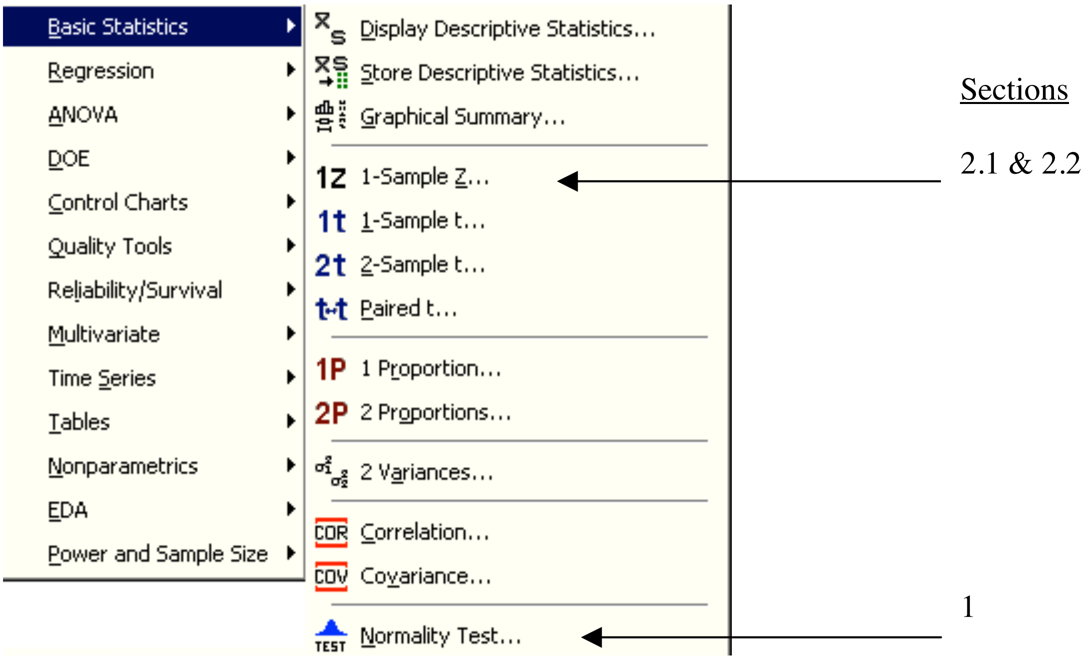
1. INSTRUCTIONS ON HOW TO PERFORM TASKS USING MINITAB.

2. A LIST OF EXERCISES TO DO USING THE ABOVE COMMANDS

C.A. Howie
Computing Science & Maths
School of Natural Sciences
University of Stirling

Introduction

In this Practical we will be using different options from the following menus:



> **Important**
>
> If the data can be assumed to be Normally distributed, we will perform parametric tests and construct parametric confidence intervals.
>
> Parametric tests assume that the data is normally distributed and use the mean as the measure of location and standard deviation (variance) as the measure of spread.
>
> All tests and confidence intervals in this Practical are parametric.

1. **Is the Data Normally Distributed?**

The formal statistical tests and confidence intervals that will use to 'objectively' answer real-life questions assume that the data comes from a Normal Distribution i.e. is Normally distributed.

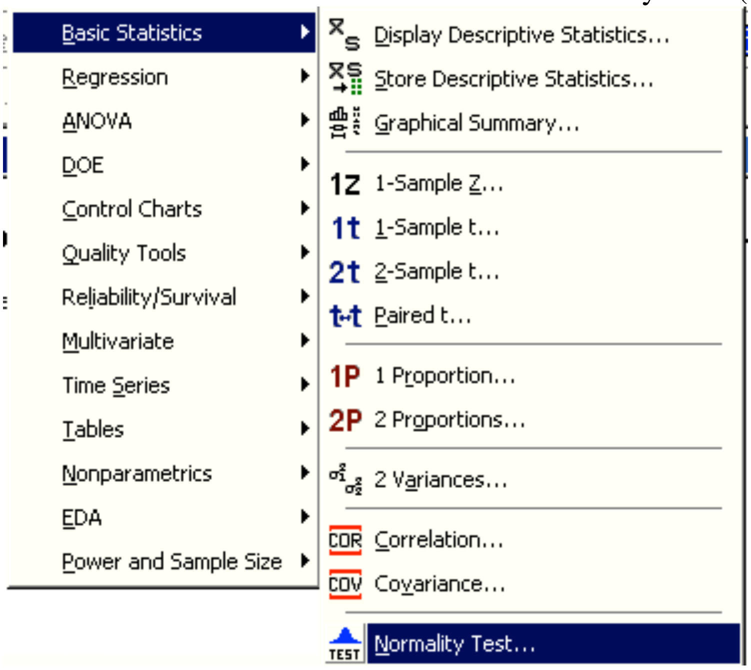
We could look at the histograms or stem and leaf plots then decide whether we think that they are bell-shaped. However, there are a couple of problems, firstly, different scales can make the plots look different and there are problems when we have small sets of data. (Also it is very subjective).

The method that we are going to use is less subjective and also uses each data point individually so it is very useful with small data sets.

The plot is called a 'Normal Probability Plot'.

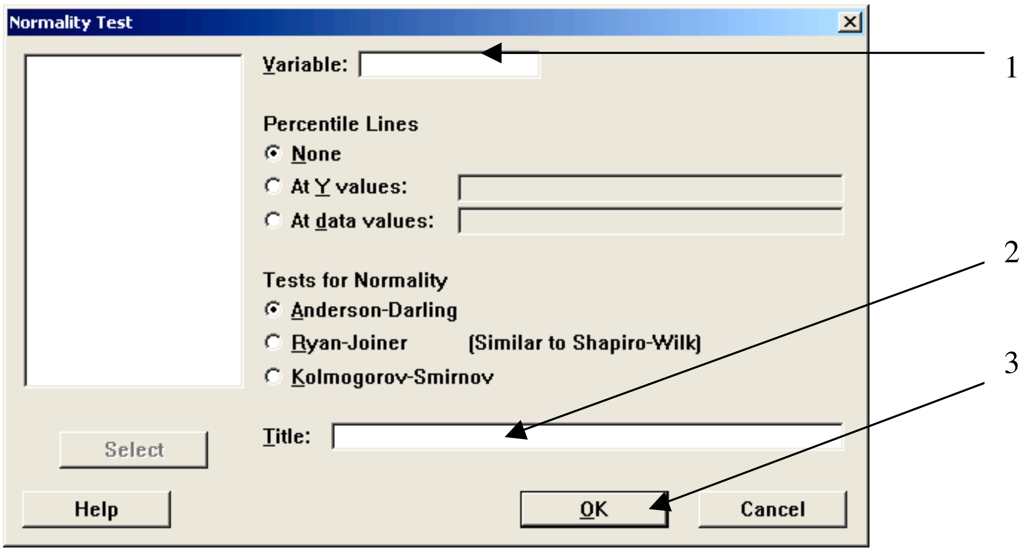
1.1 To create a Normal Probability Plot

Access Stat Menu -> Basic Statistics -> Normality Test (see over the page)

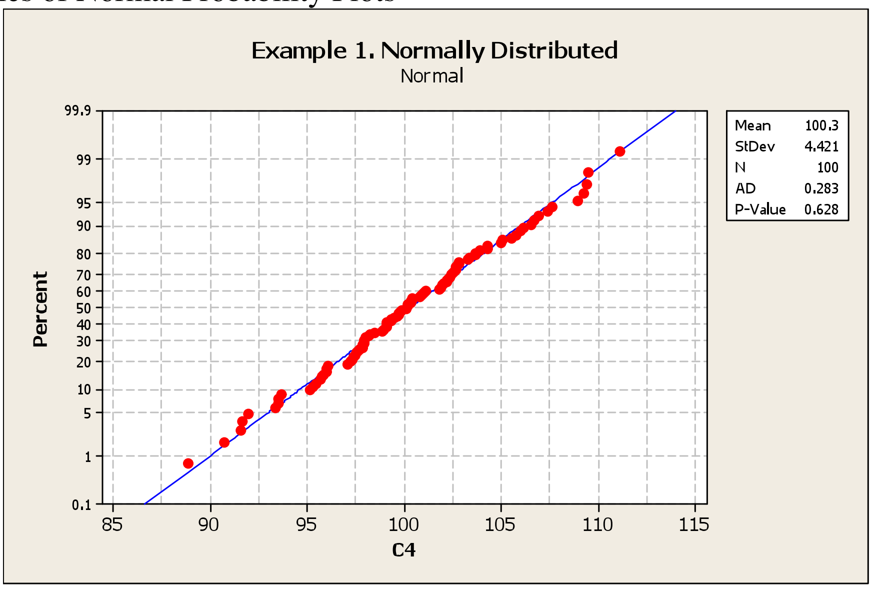


The following dialogue box appears:

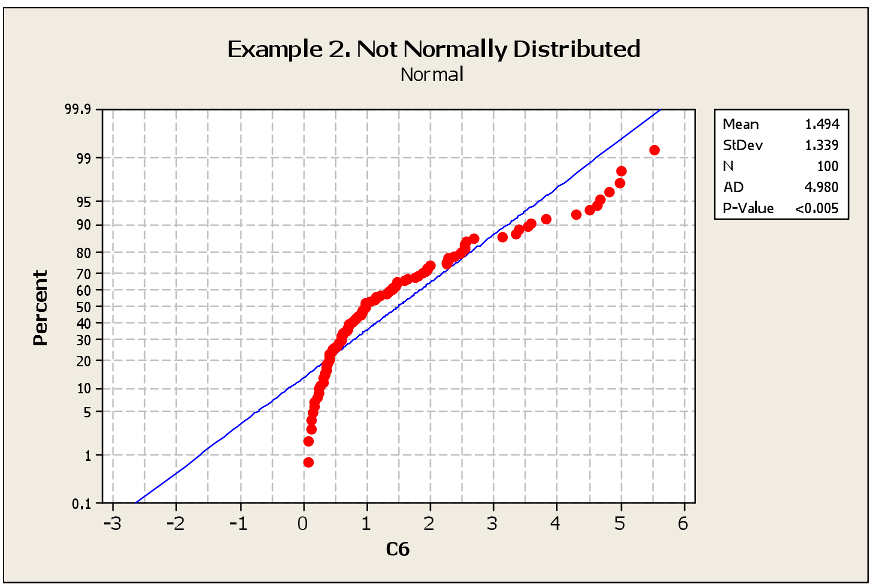
1. Select the column for which you want the plot as the Variable
2. Title the graph
3. Click OK



1.2 Examples of Normal Probability Plots



Data follows a straight line so can assume Normally distd.



Data follows a curved line so cannot assume Normally distd.

2. One Sample Problem

Situation : *The data is collected from* ***one group*** *of experiments /individuals.*
Question : *About the average response being some pre-determined value.*

2.1 <u>One Sample Z test</u>

Conditions: **Is the data Normally distributed?** **YES**
Is the variance known for the population? **YES**

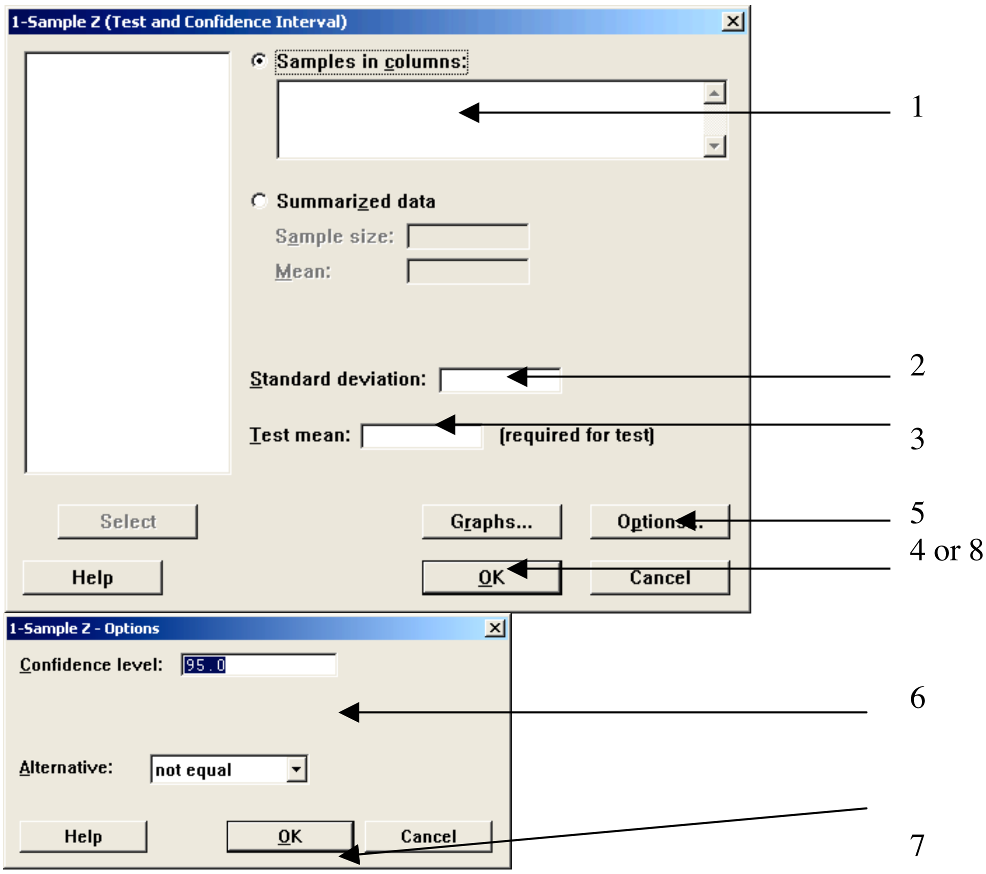
You can perform any of these 3 tests :

Null hypothesis H_0 : $\mu = \mu_o$ against
Alternative H_1 : $\mu \neq \mu_o$ or $\mu > \mu_o$ or $\mu < \mu_o$

Access the Stat Menu -> Basic Statistics -> 1-Sample Z

In the dialogue box,

1. Select the columns you want to plot as Samples in Columns
2. Enter the known standard deviation
3. Enter the claimed mean
4. If you do not want to change the Confidence Level or H1 – Click OK
5. If you do want to change the Confidence Level or H1 – Click Options - the second box appears
6. Change Confidence Level or Alternative
7. Click OK
8. Click OK



2.2 One Sample Z test Example Output

$H_o: \quad \mu = 95$

$H_1: \quad \mu \neq 95$

From the output below:

<u>Observed Test Statistic</u> Z = 6.60 ,$p<0.001$ and 95% CI for μ is (98.7, 101.9)

<u>Conclusion</u> : Sufficient evidence to reject Ho in favour of H1 at 5% (and 1%) level since $p<0.05$ (and $p<0.01$). Same conclusion using the 95% CI for μ i.e. since 95% CI for μ does not include 95, sufficient at 5% significance level to conclude that the population mean is significantly different to 95.

Hypotheses

Assumed sd

CI : states level

Observed Test Stat

One-Sample Z: Test 1

```
Test of mu = 95 vs not = 95
The assumed standard deviation = 4


Variable   N     Mean  StDev  SE Mean        95% CI            Z      P
Test 1    25  100.283  5.119    0.800  (98.715, 101.851)  6.60  0.000
```

p value if < 0.05 sig at 5% level
i.e. reject Ho in this case

EXERCISES

SECTION A

Retrieve the Minitab Project file containing the data and results that you saved during Workshop 2 or enter the data now!! (see below)

1. The following yields (kgs) were obtained from plots of a fixed size in a field of potatoes growing under the same fertiliser treatments.

28	21	14	17	24	19	22	21	16	26
20	24	21	19	17	15	18	22	23	20
26	18	24	20	19	23	22	18	20	22

   (i) Draw a Stem and Leaf plot.
   (ii) Calculate the Five Number Summary of the data
   (iii) Draw a Box and Whisker plot.

2. The rainfall was measured on nine days and the results (in mm) were :

   11.7 12.2 10.9 11.4 11.3 12.0 11.1 10.7 11.6

   Is the mean rainfall different from 12.1 mm (the average measured over the same period last year)? Draw appropriate plots and calculate numerical summaries to 'see' what you think.

3. The birth weights (kgs) of 36 babies born after normal pregnancies of 40 weeks were :

3.5	4.1	2.8	3.2	2.8	3.1	3.4	3.0	2.3
3.8	2.7	3.7	3.9	2.6	2.7	3.1	2.2	2.9
3.2	3.7	3.3	4.3	3.4	3.5	4.6	3.1	3.4
3.5	3.5	3.8	2.4	3.0	3.6	4.0	2.9	3.3

   Is the mean weight of babies born after a normal pregnancy 3.6kgs? Draw appropriate plots and calculate numerical summaries to 'see' what you think.

4. Drug levels (in ng/ml) in blood samples from two groups of subjects gave values of :

   Group 1 : 3.3 3.7 3.5 4.1 3.4 3.5 4.0
   3.8 3.2 3.7

   Group 2 : 3.2 3.6 3.1 3.4 3.0 3.4 2.8
   3.1 3.3 3.6

   Is there a difference in the means? Draw appropriate plots and calculate numerical summaries to 'see' what you think.

SECTION B

In each case from Section A, you should verify that the data is Normally distributed.

SECTION C

For the data below, state which test is appropriate and the null and alternative hypotheses.
Then use Minitab to perform the test and calculate the appropriate confidence interval.

1. The rainfall was measured on nine days and the results (in mm) were :

   11.7 12.2 10.9 11.4 11.3 12.0 11.1 10.7 11.6

   Is the mean rainfall significantly different from 12.1 mm (the average measured over the same period last year) ? The standard deviation is assumed known and is 0.45mm.

2. The birth weights (kgs) of 36 babies born after normal pregnancies of 40 weeks were :

3.5	4.1	2.8	3.2	2.8	3.1	3.4	3.0	2.3
3.8	2.7	3.7	3.9	2.6	2.7	3.1	2.2	2.9
3.2	3.7	3.3	4.3	3.4	3.5	4.6	3.1	3.4
3.5	3.5	3.8	2.4	3.0	3.6	4.0	2.9	3.3

   Is the mean weight of babies born after a normal pregnancy 3.6kgs?
   The standard deviation is assumed known and is 0.6kgs.