MATU9D2 : PRACTICAL STATISTICS

Practical 7

Spring 2017

`

- Minitab :

  SECTION 1
  - : Transforming Data
  - : One Sample Chisquared Interval
  - : F test

  SECTION 2
  - : Scatter Plots
  - : Correlation

- Handout 2 of 2

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
Faculty of Natural Sciences
University of Stirling

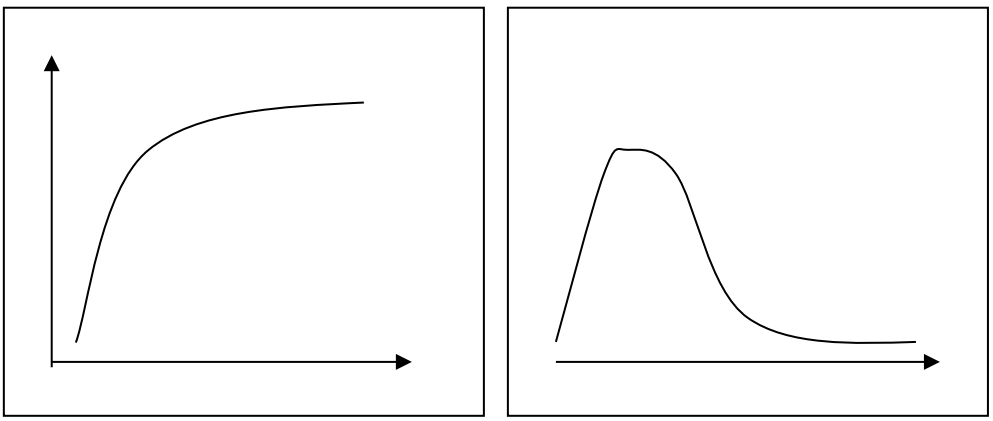
SECTION 1

1. Transforming Data

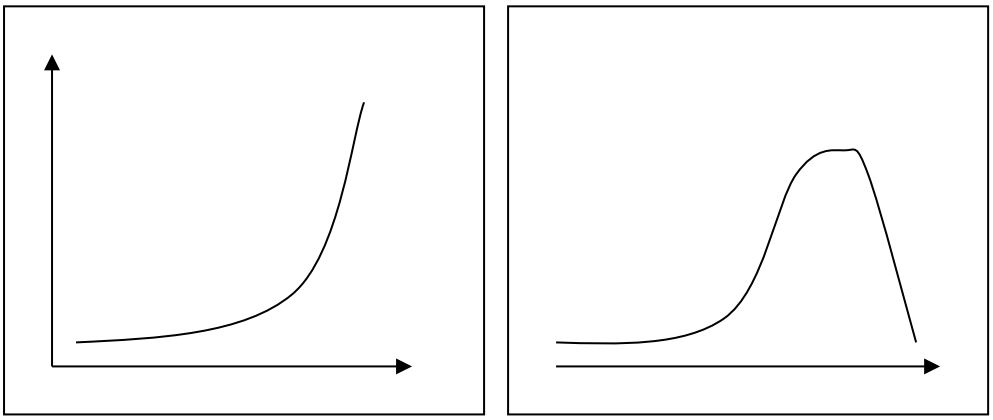
If the Normal Probability plot suggests we cannot assume Normality of our data then we should try 'transforming our data to Normality'.

What to do if the data is not normally distributed?
Try the functions suggested by the plots below then re-draw the Probability Plot for this new data – if Normal now, then analyse this new data.

1. If the Probability plot in Minitab is as below (left) it suggests that the distribution is positively skewed (check histogram- schematic below right) then try natural log, log 10 or square root.



2. If the Probability plot in Minitab is as below left it suggests that the distribution is negatively skewed (check histogram – schematic below right) then try exponential or square.



Menu Commands :

Calc -> Calculator
-> Store result in variable i.e. New Column eg.C2
-> Click on Expressions Block
-> Enter the transformation e.g. Loge(C1)
-> OK

2. One Sample Problem – Variance or Standard Deviation

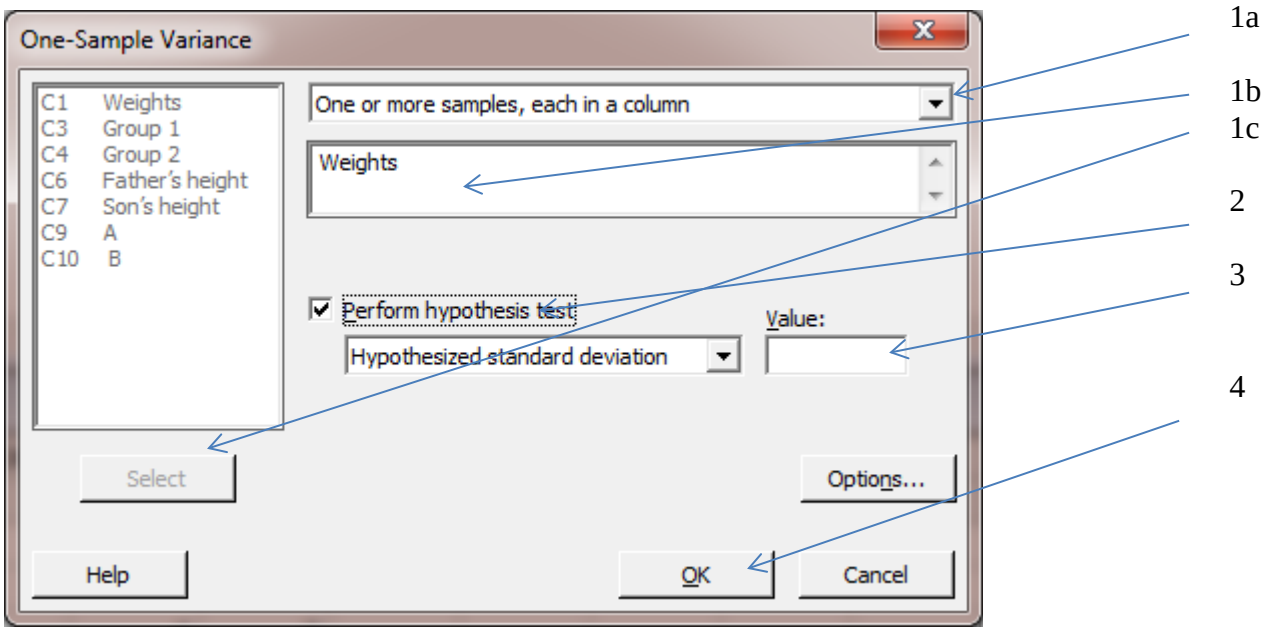
Situation : *The data is collected from one group of experimental units*
Question : *About the standard deviation or variance of responses.*
Condition : **Is the data Normally distributed?** **YES**

You can perform any of the following test (or equivalent confidence interval)
Null hypothesis H_0 : $\sigma = \sigma_o$ against
Alternative H_1 : $\sigma \neq \sigma_o$

Access the **Stat Menu** -> **Basic Statistics** -> **1Variance**

In the dialogue box,

1. Select the Columns you want to use as One or more samples, each in a column
   - 1a Click on 1a
   - 1b Choose the column from the list
   - 1c Click Select
2. Choose Perform Hypothesis Test
3. Enter the Hypothesized Standard Deviation :
4. If you do not want to change the Confidence Level or H_1 – Click OK



3. Two Sample Problems I – Independent Groups

In this case, after checking whether the data is normally distributed you should first check we can assume that the variances are equal.

F-test to homogeneity of variance.

Situation : *The data is collected from two **independent** groups of experiments / individuals.*

Question : *Are the variances of the responses equal?*

Condition : **Is the data Normally distributed? YES**

You will be testing the following hypotheses :

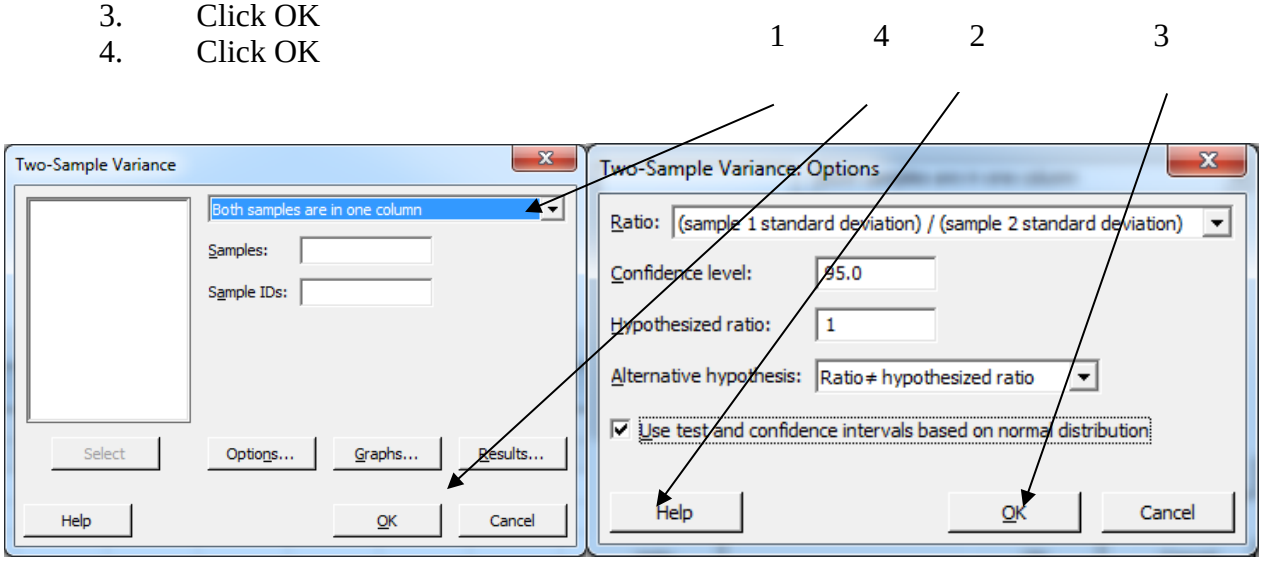
Null hypothesis H_0 : $\sigma_1^2 = \sigma_2^2$ against

Alternative H_1 : $\sigma_1^2 \neq \sigma_2^2$

Access the **Stat Menu -> Basic Statistics -> 2 Variances**

In the dialogue box,

1. Select the columns you want to plot as either Both Samples are in one column or Each Sample is in its own column depending on your data layout
2. Click on Options – second box appears – Choose Use test and CI based on normal distn
3. Click OK
4. Click OK



SECTION 2 : Relationships between Quantitative Variables

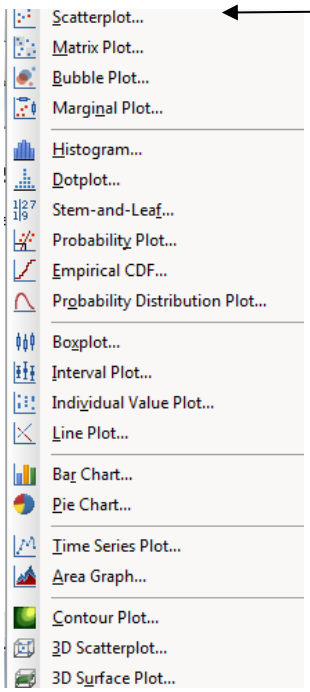
Correlation (r value) gives a measure of the linear relationship between two quantitative variables.

1 Plot the Data

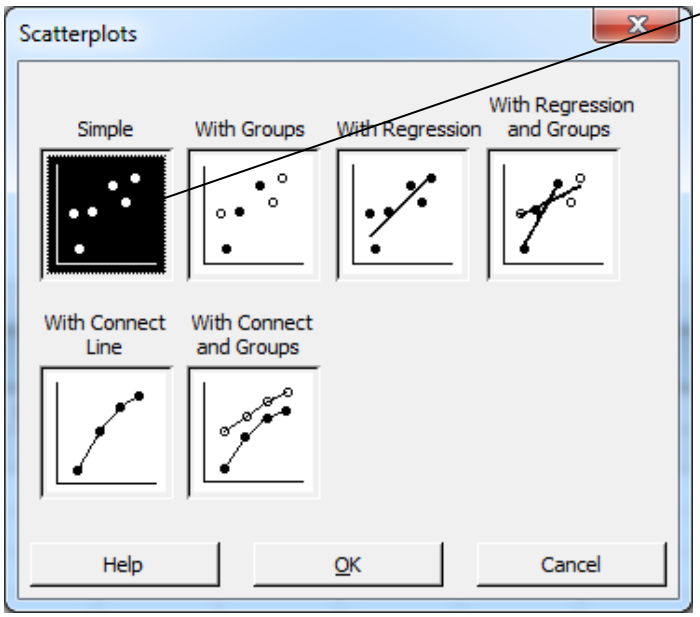
In all cases when looking at relationships - **PLOT THE DATA**

i.e. a bivariate or XY plot or Scatter plot.

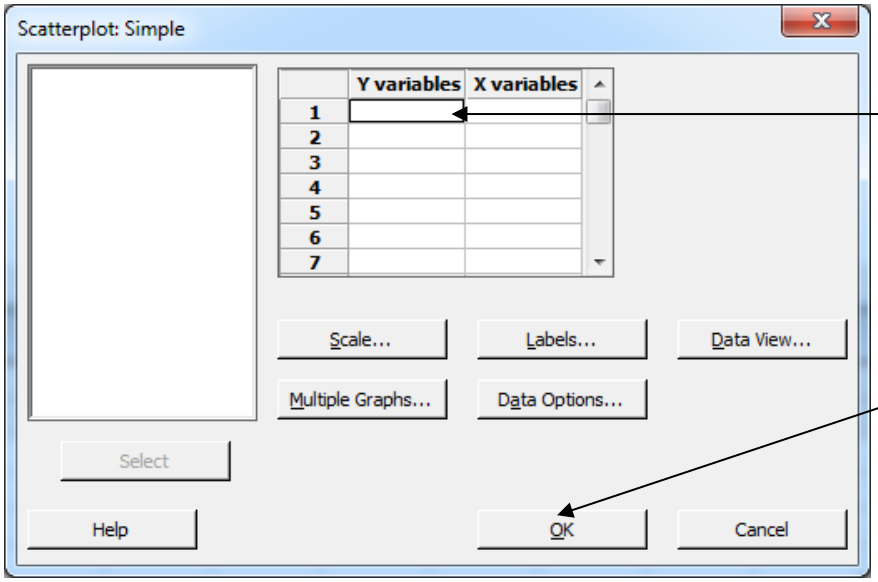
Access the Graph Menu then following choice is presented.



1. Choose Scatter plot.



2. Choose Simple



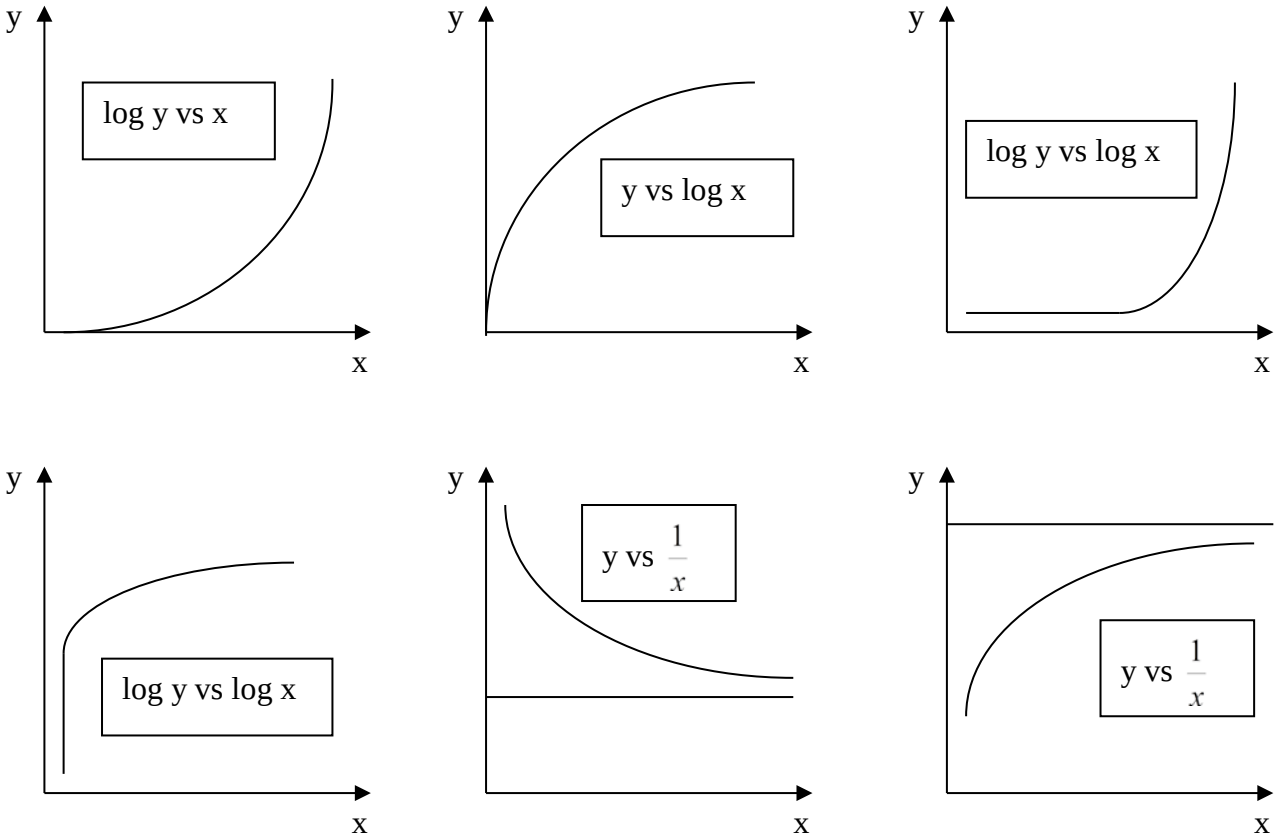
3. Select the Column with the y data and the x data.

4. Click OK

2. <u>Transformations</u>

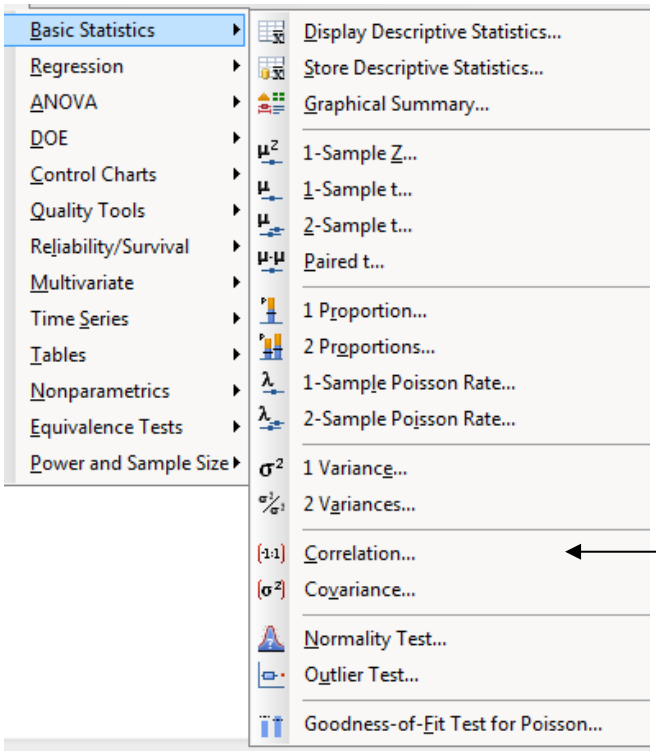
If the relationship is not linear - try to transform the data e.g. take the logarithm.

The transformation necessary will be identified from the shape of the graph.

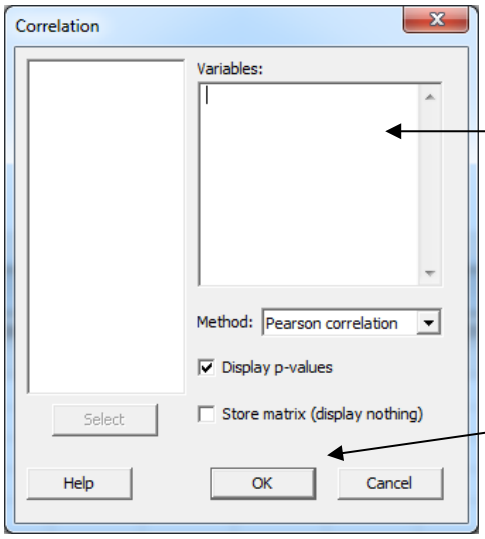


3. Calculating the Correlation Coefficient (r)

Access the Stat Menu then Basic Statistics the select Correlation (see below)



1. Choose Correlation



2. Select the columns containing the data as Variables

3. Click OK

Example Output

Correlations: Age, Grades

```
Pearson correlation of Age and Grades = 0.278
P-Value = 0.000
```

Result of test

$H_o : \rho = 0$ vs $H_1 : \rho \neq 0$

e.g. $p<0.05$ so correlation significantly different to zero at 5% level

Questions

1. 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 standard deviation different from 0.6kgs?

2. 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 variances? Which t test should we use to compare the means?

3. We are given the following data and required to answer the 3 questions below.

x	10	12	14	16	18	20	22	24	26	28
y	25	24	22	20	19	17	13	12	11	10

(a) Take the data given above and construct a scatter diagram.
(b) Find the correlation coefficient (r) for this data.

2. Construct a scatter diagram for the following data
(**Note : this is one set of data with 20 pairs).**

x	1	2	3	4	5	6	7	8	9	10
y	10	10	11	12	12	13	15	18	21	25
x	11	12	13	14	15	16	17	18	19	20
y	26	29	33	39	46	60	79	88	100	130

Find the correlation coefficient.

Now find the natural logarithm of y and re-calculate the two statistics. How would you interpret your results?