

QUEEN MARY, UNIVERSITY OF LONDON

MTH 4106

Introduction to Statistics

Assignment 2

For handing in on 24–25 January 2012

You should attempt all of these questions, as they are designed to help you to learn and understand the material in the course.

The ‘Feedback’ question is the one for handing in. Write your name, student number and group number at the top of your answer before handing it in. Staple all the pages together. Either (i) hand it to your allocated tutor when (s)he asks for it during your allocated Minitab laboratory session on Tuesday 24 January 2012, or (ii) post it in the red post-box on the ground floor of the Maths building before 1400 on Wednesday 25 January 2012.

If you want help on any of the other questions, or want to check that you have done them correctly, you may ask any tutor during your laboratory session or ask me in any of my office hours.

1 Some foresters planted 16 *Angophora* trees and 16 *Casuarina* trees. One of the *Angophora* trees died. After 54 months, the heights of the remaining trees were measured in metres. The data were as follows.

Angophora: 8.8, 7.3, 8.6, 7.8, 6.8, 7.1, 6.8, 8.8, 5.9, 8.5, 4.1, 8.9, 9.1, 5.8, 8.4.

Casuarina: 4.8, 4.5, 4.9, 6.7, 7.0, 7.7, 7.0, 5.8, 5.5, 5.1, 4.1, 4.0, 5.4, 5.2, 5.5, 5.0.

[Source: *Experimental Design and Analysis for Tree Improvement*, second edition, by E. R. Williams, A. C. Matheson and C. E. Harwood, published by CSIRO Publishing in 2002.]

Make a stem-and-leaf plot for each set of data.

2 For the VW petrol data as originally given in mpg, we have

$$\sum_{i=1}^{20} x_i = 727, \quad \sum_{i=1}^{20} x_i^2 = 26591 \quad \text{and} \quad \sum_{i=1}^{20} (x_i - \bar{x})^3 = -340.6350.$$

Calculate the sample mean \bar{x} , the sample standard deviation s , and the coefficient of skewness.

3 (Feedback) Please do not use Minitab at Queen Mary on Tuesdays except in your allocated hour.

The following data set gives the height in centimetres of thirty female athletes. The data were collected at the Australian Institute of Sport, courtesy of Richard Telford and Ross Cunningham. Heights above the line correspond to basketball players, those below the line to rowers.

195.9	189.7	177.8	185.0	184.6	174.0
186.2	173.8	171.4	179.9	193.4	188.7
169.1					
177.9	177.5	179.6	181.3	179.7	185.2
177.3	179.3	175.3	174.0	183.3	184.7
180.2	180.7	176.0	156.0	179.7	

Use Minitab to write a short report including the following.

- Draw two boxplots, one for each sport.
- Draw two stem-and-leaf plots, one for each sport.
- Comment on the similarities and differences in the heights of the two groups of athletes which your plots have shown.
- Calculate the mean, median, minimum, maximum, lower and upper quartile, and standard deviation for each group of athletes. Do these statistics confirm your comments in part (c)?

Write your report in such a way that it is informative to the manager of the university sports centre. In particular, this means that diagrams should be clearly labelled, units of measurement should be shown, and comments should be set in the context of what the data are about.

4 Let $x_1, y_1, x_2, y_2, \dots, x_n, y_n$ be real numbers, where $n \geq 2$. Prove that

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{n \sum_{i=1}^n (x_i y_i) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n(n-1)}.$$

5 Let x_1, \dots, x_n be data, not all the same. For $i = 1, \dots, n$, put $y_i = ax_i + b$, where a and b are constants and $a \neq 0$.

Prove that the sample correlation coefficient between x and y is equal to

$$\begin{cases} +1 & \text{if } a > 0 \\ -1 & \text{if } a < 0. \end{cases}$$