

Introductory Statistics Tutorial

Chapter 2 – Tools for Exploring Univariate Data

1. The weights of 9 stage II engineering students were recorded as part of a class experiment. The weights, in kilograms, of these 9 students were: 70, 75, 60, 102, 67, 85, 97, 60, 70.

(a) Draw a dot plot of the weights of the students.

(b) Comment on the main features in this sample.

2. At one stage in the process of producing silicon chips, a very thin layer of silicon oxide is deposited on a “wafer”. The wafer is then broken up into chips. Using the following data from *Technometrics* (1994), draw a stem-and-leaf plot of the thickness of silicon oxide in 30 such chips. The thickness has been measured in a special unit for very small distances called Angstrom units, Å.

840	900	930	940	950	960	970	980	990	990	1000	1000	1000	1010	1010
1030	1030	1030	1040	1040	1050	1050	1050	1050	1050	1070	1070	1100	1100	1120

(a) Complete the stem-and-leaf plot for these 30 thicknesses.

Units: 9 | 5 = 950Å

8	
8	
9	
9	
10	
10	
11	

(b) Calculate the five-number summary.

(c) Which of the following statements is **not** a feature of the data?

- (1) The interquartile range is 70Å.
- (2) The range is 270Å.
- (3) The mode is 1050Å.
- (4) The median is 1020Å.
- (5) Those observations with values 1100Å or more represent about 10% of the distribution of thicknesses.

3. (a) Draw a box plot for the following set of data:

18	19	21	21	23	23	23	27	29
29	30	31	35	41	49	55	78	

Five-number summary: (18, 22, 29, 38, 78)

- (b) Do you agree with the following statements? Discuss.

- (1) The distribution from which this sample is drawn is highly skewed.
- (2) The interquartile range is 21.
- (3) There are no observations greater than 78.
- (4) The observation 78 is an outside value for the box plot representing the above data.
- (5) The observation 18 is an outside value for the box plot representing the above data.

4. Which **one** of the following statements is **true**?

- (1) The mean is less affected by outliers than the median.
- (2) Outliers affect the standard deviation more than they affect the interquartile range.
- (3) The numbers of cars owned by a family is a continuous variable.
- (4) Box plots are good at distinguishing between unimodal and bimodal distributions.
- (5) When coding qualitative variables (i.e. using numbers to describe the outcomes) it is a good idea to work out the means and medians.

5. Do you agree with the following statements? Discuss.

- (1) It is a good idea to round off numbers when using them in a table for display purposes.
- (2) Dot plots should be used for samples with a small number of observations.
- (3) Box plots are not good for comparing centres of location and spreads of data.
- (4) Bar graphs cannot be used to display discrete data.

6. The five-number summary for a set of data is:

(10, 22, 37, 50, 60)

Which **one** of the following is **false**?

- (1) Each of the whiskers on the box plot of the data must be greater than 42 units in length.
- (2) It is not possible to determine the mean of the data from this five-number summary.
- (3) At least half of the observations are between 22 and 50 inclusive.
- (4) The interquartile range is 28.
- (5) None of the observations in the data set is an outside value on the box plot of the data.