ANSWERS CHAPTER 4

THINK IT OVER

think it over

TIO 4.1: Things like: checking for mistakes, making sure you are collecting the 'right' data, making sure your data is fit for purpose, making sure you are collecting the 'right' type of data (quantitative vs qualitative), making sure the data gives you the level of accuracy you require, etc.

TIO 4.2: The data point was an outlier and could have been a mistake or it could have been a valid datum. Never make assumptions concerning outliers; sometimes they can tell you more than the rest of the data. If you are not sure, investigate further.

TIO 4.3: It means it was the most popular amongst the students surveyed. The reason was to find out the most popular genre. It could be used to decide what sort of entertainment should be organised for the students. The sample comprised of sociology students so the results may not be applicable to, say, engineering students who would probably have different tastes.

TIO 4.4: Most programs will tell you the number of items entered; SPSS does. With Excel you can use the 'Count()' function to do this.

TIO 4.5: If you needed to look at individual responses. Grouping responses together may not be suitable and could possibly distort the analysis. Frequency means the number of data in a particular class.

TIO 4.6: If the pilot study revealed the class sizes were adequate, stick with them. If, on the other hand, some detail was lost, use more classes. All three questions are important.

TIO 4.7: It depends on the purpose of the study, how much detail is required, etc.

TIO 4.8: The word 'relative' implies a relationship between the class data and all the data. In other words, it gives you an overall 'picture' of what the data is telling you.

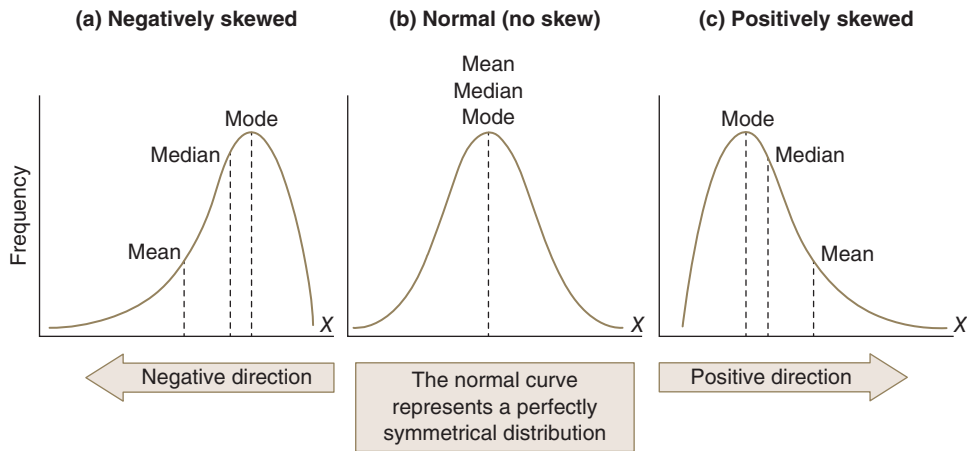TIO 4.9: A fixed width makes comparing data easier and misinterpretation less likely.

TIO 4.10: As with defining variables, labelling your graph is important. You may know what the graph shows, but will your audience, or will you remember in 12 months' time?

TIO 4.11: Bar graphs are not really appropriate for continuous data unless you have broken the data into classes. Line graphs are probably more appropriate since they demonstrate continuity.

TIO 4.12: Never leave cells empty! Someone else looking at your work would not be sure if it was deliberately left empty or just missed out. Programs like SPSS require you to code for missing values. If a value of, say, 999 was not within the data range, it could be used to indicate an empty cell.

TIO 4.13: No answer required.

TIO 4.14:

(a) Negatively skewed

Mode
Median
Mean

Negative direction

(b) Normal (no skew)

Mean
Median
Mode

The normal curve represents a perfectly symmetrical distribution

(c) Positively skewed

Mode
Median
Mean

Positive direction

Frequency

X

The first one would be symmetrical, the second positively skewed and the third negatively skewed.
It implies a horizontal line since the mode is a measure of the data that occurs the most.

TIO 4.15: It means that 80% of your peers scored higher than you in the exam, but you scored higher than 20% of your peers. If 80% of the class scored above 80%, than either the exam was very easy or you are in a class of exceptionally clever people!
Percent means out of 100. Therefore 1% means 1 out of 100. A percentile provides information on the data spread.

TIO 4.16: Quartiles are used when it is desirable to divide the data into four parts. For example, salaries, exam scores, etc.
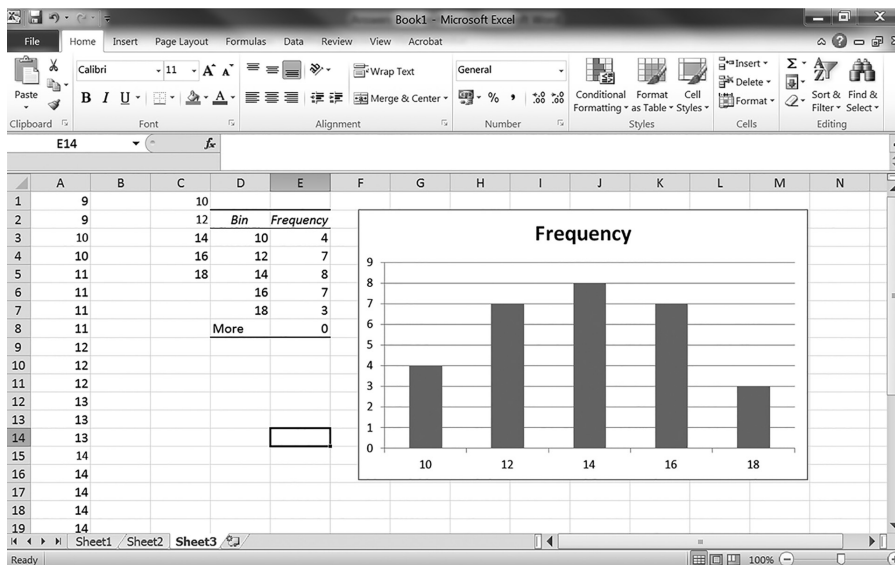
TIO 4.17: No answer required.

TIO 4.18: $\sigma^2 = \dfrac{\sum(x_i - \mu)^2}{N}$. I had forgotten to square the numerator.

TIO 4.19: You would need to find out if the results have been standardised. It is not enough to know they are normally distributed or to know the means and standard deviations. Standardising data enables like-for-like comparisons to be made.

EXERCISES

1. (a) A number of measures could be made: feelings of being prepared, amount of time spent revising, levels of feeling stressed, time scale from 12 hours to 1 hour.
   (b) On a scale from 0 to 10, or on a scale from 'very stressed' to 'not stressed at all'. People sometimes find it hard to quantify how they feel, so too many points on the scale can make it hard for people to distinguish between one level and the next.
   (c) It could be displayed as a histogram, bar chart or if you decide it could be classed as continuous data, as a line graph.
   (d) Conclusions would be difficult, since everyone has different abilities to cope with stress. You could say, something like, on average people who revise regularly, feel prepared (if your results show this) are less stressed compared with those who leave everything to the last minute.

2. (a) Crime statistics are regularly collected in the UK. They are not a true picture since the data only represents recorded crime, i.e. a crime that has been reported to the police service. The police service has to justify how it spends taxpayers' money, so 'positive' results are always beneficial.
   (b) Yes the conclusion is true but the 'severity' of the crimes reported in the article would make people think how taxpayers' money is being spent on what could be described as trivial crimes.
   (c) If the number of crimes solved was the only reported data, most people would think the police service were doing a great job and deserved all the funding they needed.

3. (a) The median because of the uneven spread of salaries.
   (b) The mean is 36,990 and the median 20,250. This indicates a positive skew.

4. It means that compared to the country as a whole, there are only 10% of towns with a higher crime rate. Not good, I would suggest you move.

5. Without standardising the data, you cannot make meaningful comparisons.

6. (a) The normal distribution for data set 1 would be 'flatter', i.e. the data is more spread out around the mean, whereas data set 3's dispersion is less, i.e. it is more closely grouped around the mean. This is what the interquartile range suggests (a measure of the middle 50%) and the mean deviation (datum – mean).
   (b) More of the business studies lecturers thought that £10 was reasonable.

7. (a) Group 2 by a long way compared to group 3.
   (b) Group 3 have £210 disposable income whereas group 2 have £180, therefore the difference is £30. Depending upon your interpretation of 'significant', there is not a huge difference.
   (c) The scale was omitted from the first chart, so interpretation depended solely on the visual difference between the columns, whereas chart 2 provided a reference via the vertical scale.

8. (a) It shows that yields have been rising over the past 6 years and the trend is for this to continue.
   (b) You could make a guess but that's all it would be.
   (c) This is even more tricky than (b) since anything could happen within the 4-year time scale. If conditions remained the same you could give an approximate prediction. But, as we know, conditions can change very quickly (for example, the world banking crisis of 2008).

9. (a), (b)

(c) Most of the 30 employees were between 11 and 16 minutes late.

10.

(a) Mean profit margin is 5.1%
(b) Calculating the individual product profits and total sales gives an overall profit margin of 4.23%. Calculating the mean of the % profit margins just gives you the mean of the profit margins. It does not tell you the overall % profit.
(c) By calculating the mean of the profit and dividing by the mean of the sales.