Module 2 – Solutions

2.1   (a)   (i) 34   (ii) 28.5   (iii) 26.4
           (b)   Mean is the lowest – the '4' depresses its value compared to the mode and the median.
           (c)   31, 25.5, 23.4

2.2   (a)   As long as the boss's true salary of €200,000 is still above the median, the median will be correct.
                      The mean will be too large, since the total of all the salaries will decrease by €2,000,000 – €200,000 = €1,800,000, once the mistake is corrected.
           (b)   The range will likely be too large. The boss's salary is probably the maximum, and a lower maximum would lead to a smaller range. The IQR will likely be unaffected, since the new maximum has no effect on the quartiles.
                      The standard deviation will be too large, because the €2,000,000 salary will have a large squared deviation from the mean.

2.3

2.4   (a)   The distribution of length of stays is skewed to the right, so the mean is larger than the median.
           (b)   The distribution of the length of hospital stays of female heart attack patients is skewed to the right, with stays ranging from 1 day to 36 days. The distribution is centred around 8 days, with the majority of the hospital stays lasting between 1 and 15 days. There are a relatively few hospitals stays longer than 27 days. Many patients have a stay of only one day, possibly because the patient died.

2.5   1.   As maths results increases physics results tend to also increase.
           2.   In general the more time spend exercising, tends to lead to a decrease in body mass.
           3.   There is no linear relationship between Maths results and the height of   students.
           4.   As the outside air temperature increases, heating bills tend to decrease.
           5.   As the daily hours of sunshine increases, the sale of suncream tends to get higher.
           6.   In general the older the car the less its value.

2.6   (a)   It is appropriate to calculate correlation. Both height of the drop and speed are quantitative variables, the scatterplot shows an association that is straight enough, and there are no outliers.
           (b)   There is a strong, positive, linear association between drop and speed; the greater the height of the initial drop, the higher the top speed.

2.7   The candidate might mean that there is an association between television watching and crime. The term correlation is reserved for describing linear associations between quantitative variables. We don't know what type of variables "television watching" and "crime" are, but they seem categorical. Even if the variables are quantitative (hours of tv watched per week, and number of crimes committed, for example), we aren't sure that the relationship is a linear. The politician also seems to be implying a cause-and-effect relationship between television watching and crime. Association of any kind does not imply causation.

2.8   There is no apparent association between the number of grams of fat and the number of milligrams of sodium in several brands of fast food burgers. The correlation is only r =0.199, which is close to zero, an indication of no association. One burger had a much lower fat content than the other burgers, at 19 grams of fat, with 920 milligrams of sodium. Without this (comparatively) low fat burger, the correlation would have been even lower.