Analytical Techniques 1: Practical solution

Exercise 1.1

a. The two values of 0 will have affected both the mean (pulled towards 0) and, to a greater extent, the SD (inflated). The effect on the SD will be greater because the deviations from the mean are squared. The two values of 0 would result in a negative skew (in fact -5.4) and a large kurtosis (in fact 35). The large negative skew is due to the two values lying a long way to the left of the main distribution and the large kurtosis is due to these values lying a long way from the centre of the distribution.

b. (i) Calculate the mean SBP for the 98 patients.

$$\frac{1}{100} \times \sum_{1}^{100} x_i = 132.2$$

$$\sum_{1}^{100} x_i = 132.2 \times 100 = 13220$$

As two removed values are both 0 therefore the sum of x_i will remain unchanged. Therefore, the new mean is:

$$\frac{1}{98} \times \sum_{1}^{98} x_i = \frac{13220}{98} = 134.9\ mmHg$$

b. (ii) Calculate the standard deviation of SBP for the 98 patients.

Need to use formula for $SD = \sqrt{\frac{1}{n-1}\left[\left(\sum x_i^2\right) - n\bar{x}^2\right]}$

We know that for the **100** patients: $\sqrt{\frac{1}{99}\left[\left(\sum x_i^2\right) - 100 \times 132.2^2\right]} = 20.7$

Rearranging this gives that $\sum x_i^2 = 1790104.51$

Now we know that the 2 patients who have been removed both had values of 0. So their contribution to the $\sum x_i^2 = 0$. So for the 98 patient the $\sum x_i^2$ will be unchanged.

Therefore, the new SD for 98 patients is:

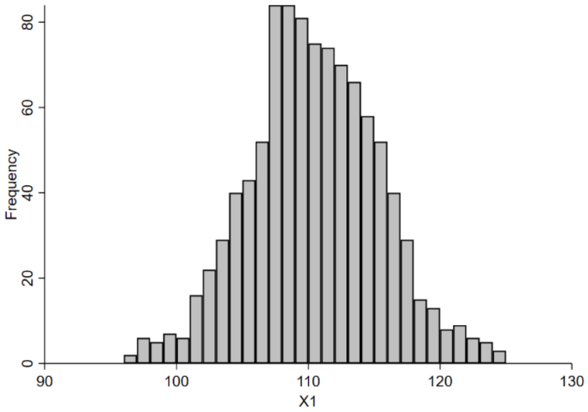
$$SD = \sqrt{\frac{1}{97}\left[1790104.51 - 98 \times 134.9^2\right]} = 8.3\ mmHg$$

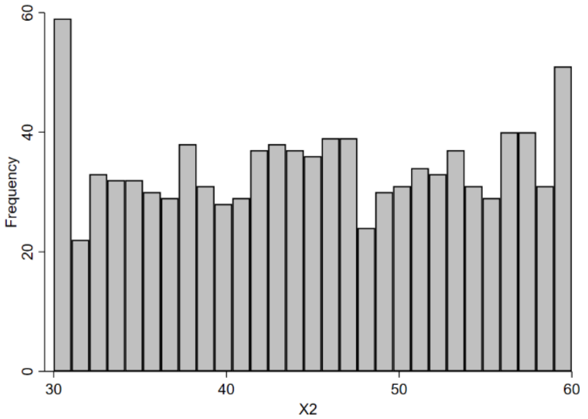
c) Since we are told that SBP is approximately normally distributed the mean and median values will be approximately the same, i.e. around 134.9mmHg. Therefore, we'd expect around 50% of patients to have SBP>135mmHg. We could use the standardised normal distribution to estimate the proportion of patients with SBP>145, e.g. Prob(SBP>145) = 1 – Φ(z) where z = (145-134.9)/8.3 = 1.205. So Prob(SBP>145)=1- Φ(1.205) = 0.114 or 11.4%.

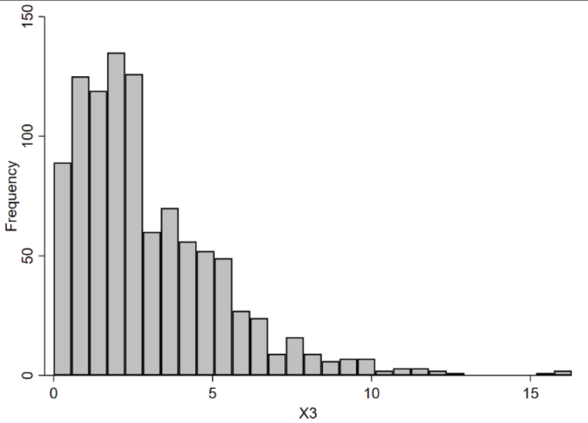
So

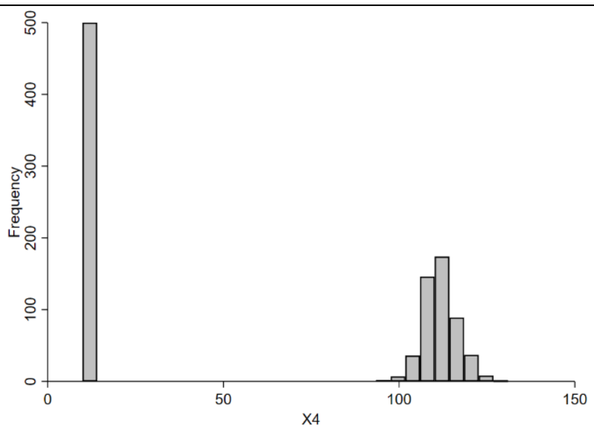
Analytical Techniques 1: Practical solution

Exercise 1.2

<table>
<tr><td>
</td><td>Mean = 110
Median = 110
SD = 5
Skew = 0.1
Kurtosis = 3

Distribution appears approximately normal so mean/SD would be fine for location/spread</td></tr>
<tr><td>
</td><td>Mean = 45
Median = 45
SD = 8.7
Skew = 0.0
Kurtosis = 1.8

Although distribution is symmetric it is clearly not normally distributed – probably best to report median/IQR</td></tr>
<tr><td>
</td><td>Mean = 3
Median = 2
SD = 2.4
Skew = 1.6
Kurtosis = 6.7

Distribution is positively skewed – report median/IQR for location/spread</td></tr>
<tr><td>
</td><td>Mean = 61
Median = 54.5
SD = 50
Skew = 0.0 (why?)
Kurtosis = 1.0 (why?)

Clear problem with this distribution – inappropriate to report any summary statistic for location/spread</td></tr>
</table>