6.2: Finding Outliers Using Quartiles

We can use quartiles to identify *outliers* or data points that are wildly discrepant with the rest of the data. For this application, we need another definition of data dispersion :

$$\text{Interquartile Range} = IQR = Q_3 - Q_1 \tag{6.2.1}$$

With the IQR any data value that satisfies:

(a) less than $Q_1 - (1.5 \times IQR)$
or
(b) greater than $Q_3 + (1.5 \times IQR)$

…is considered an outlier. This is one of many ways one can define an outlier. As we will discuss below, it is a robust way of identifying outliers.

Example 6.4 : Consider the data of Example 6.2. We found

$$Q_1 = 9 \qquad Q_2 = 14 \qquad Q_3 = 20 \tag{6.2.2}$$

so,

$$IQR = Q_3 - Q_1 = 20 - 9 = 11. \tag{6.2.3}$$

Following our rules for finding outliers, we compute:

(a) lower acceptable value limit

$$\begin{aligned} &= Q_1 - (1.5 \times IQR) \\ &= 9 - (1.5 \times 11) \\ &= 9 - 16.5 = -7.5 \end{aligned}$$

(b) upper acceptable value limit

$$\begin{aligned} &= Q_3 + (1.5 \times IQR) \\ &= 20 + (1.5 \times 11) \\ &= 20 + 16.5 = 36.5 \end{aligned}$$

and 50 > 36.5 so 50 is considered an outlier.

□