

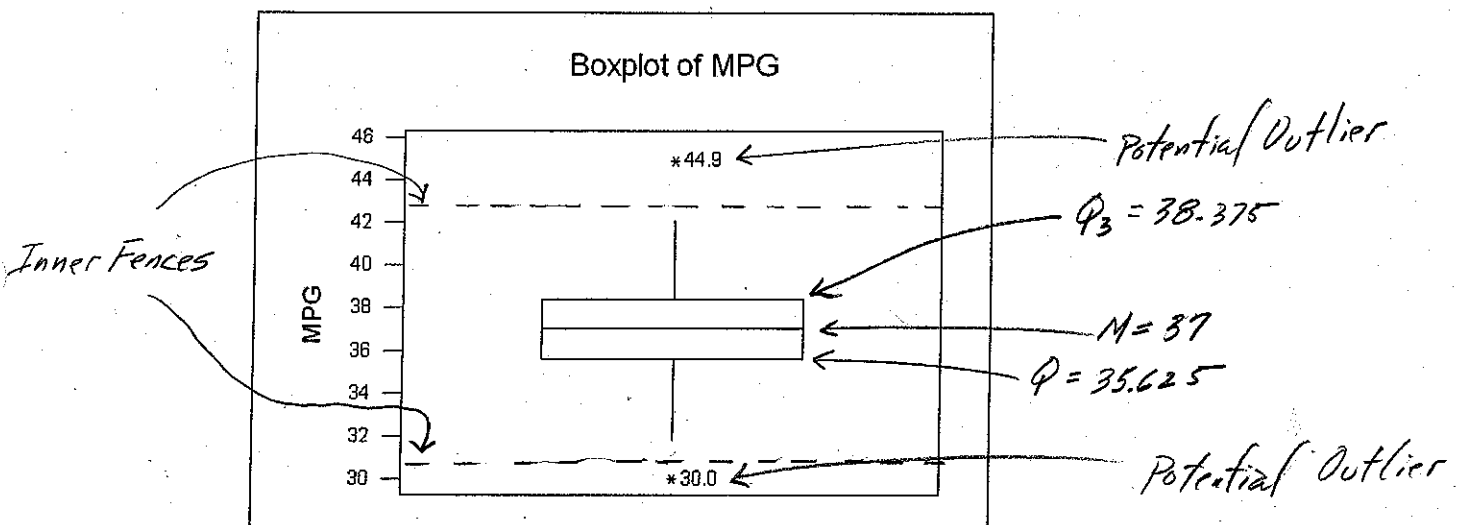
Detecting Outliers

An observation that is unusually large or small relative to the data values we want to describe is called an *outlier*.

One method for determining outliers is based on the quartiles of a data set. *Quartiles* are values that partition the data set into four groups, each containing 25% of the measurements. The *lower quartile* is designated by Q_1 or Q_L ; the *middle quartile* is the median; the *upper quartile* is designated by Q_3 or Q_U .

Box plots like the one below are based on the *interquartile range (IQR)* – the distance between the lower and upper quartiles. $IQR = Q_U - Q_L = Q_3 - Q_1$

The middle 50% of the of the MPG ratings, those in the interquartile range, fall inside the box. The vertical lines emanating from the box are called *whiskers*. Values less than $Q_L - 1.5(IQR)$ or greater than $Q_U + 1.5(IQR)$ are identified as *potential outliers* because they are extreme values that represent relatively rare. Those potential outliers are identified by asterisks (*) in Minitab. Values less than $Q_L - 3(IQR)$ or greater than $Q_U + 3(IQR)$ are called *outliers* and are also identified by asterisks (*) in Minitab.



Let's revisit Minitab's calculation of descriptive statistics for the MPG distribution.

Variable	N	Mean	Median	StDev	Minimum	Maximum	Q1	Q3
MPG	100	36.994	37.000	2.418	30.000	44.900	35.625	38.375

Use the above descriptive statistics to determine each of the following values:

$$\text{Interquartile Range (IQR)} = 38.375 - 35.625 = 2.75$$

$$Q_L - 1.5(IQR) = 31.5$$

$$Q_U + 1.5(IQR) = 42.5$$

$$Q_L - 3(IQR) = 27.375$$

$$Q_U + 3(IQR) = 46.625$$