1.8 Summarizing Quantitative Data: Boxplots & Outliers (Part 1)

Besides serving as a measure of variability, the interquartile range (IQR) is used as a ruler for identifying outliers.

HOW TO IDENTIFY OUTLIERS:

Call an observation an outlier if it falls more than 1.5xIQR above the 3rd quartile or more than 1.5xIQR below the first quartile.

Low Outliers < Q1 - (1.5xIQR)

High Outliers > Q3 + (1.5xIQR)

It is important to identify outliers in a distribution for several reasons:

1. **They might be inaccurate data values.** Maybe someone recorded a value as 10.1 instead of 101. Perhaps a measuring device broke down. Or maybe someone gave a silly response, like a student in a class survey who claimed to study 30,000 minutes per night!

2. **They can indicate a remarkable occurrence.** For example, in a graph of career golf earnings, Tiger Woods is likely to be an outlier.

3. **They can heavily influence the values of some summary statistics**, like the mean, range, and standard deviation.

Example: Identify any outliers in the data.

a.) {23, 10, 13, 30, 26, 8, 25, 18}

Q3: 25.5 + 21 = 46.5
Q1: 11.5 − 21 = −9.5
IQR: 25.5 − 11.5 = 14 x1.5 = 21

NO OUTLIERS

b.) {35, 60, 20, 80, 95, 15, 40, 85, 75}

Q3: 82.5 + 82.5 = 170
Q1: 27.5 − 82.5 = −55
IQR: 82.5 − 27.5 = 55 x1.5 = 82.5

NO OUTLIERS

Example: Identify any outliers in the data.

c.) {88, 79, 94, 90, $\boxed{45}$, 71, 82, 88}

Q3: 89 + 21 = 110
Q1: 75 − 21 = 54
IQR: 89 − 75 = 14 x1.5 = 21

45 is an outlier

d.) {$\boxed{45}$, 18, 9, 25, 14, 7, 12, 9, 4}

Q3: 21.5 + 20.25 = 41.75
Q1: 8 − 20.25 = −12.25
IQR: 21.5 − 8 = 13.5 x1.5 = 20.25

45 is an outlier