Section 3.3 - Measures of Relative Standing

Measures of relative standing show location relative to other values in the data set.

The **z score** associated with a given data value is the number of standard deviations that value is from the mean.

$$z = \frac{x - \bar{x}}{s} \text{ (sample)}$$

$$z = \frac{x - \mu}{\sigma} \text{ (population)}$$

Notice that the *z* score is positive if the data value is greater than the mean and negative if the data value is less than the mean.

For reasons we will see later, most of the time we will round our *z* scores to the nearest hundredth.

Based on Chebyshev's Theorem and our observations about normal distributions, we have the following rules of thumb:

- Most ordinary data values lie have z scores between -2 and 2.

- Unusually small data values have z scores less than -2.

- Unusually big data values have z scores greater than 2.

Percentiles are measures of location that divide a data set into 100 groups with about 1% of the values in each group.

percentile of x = 100 * (# of values < x) / (total # of values)

(Round to the nearest whole number.)

To find the value associated with a given percentile:

- Arrange the data in ascending order.
- Solve the formula above for the # of values < x.
- Call your solution L.
- If L is a whole number, the value is the mean of the L-th data value and the (L+1)-th data value.
- If L is not a whole number, round it up to the nearest whole number. The value is the L-th data value.

Quartiles are measures of location that divide a set of data into four groups with about 25% of the values in each group.

The quartiles can be computed as the 25th, 50th, and 75th percentiles.

It is often easier to compute the quartiles as medians.

- The second quartile is the median.
- The first quartile is the median of the lower half (ignoring the actual data value at the location of the median).
- The third quartile is the median of the upper half (ignoring the actual data value at the location of the median).

There is not complete agreement among statisticians and authors on how to define and compute the quartiles.

The **interquartile range** (IQR) is the difference of the third and first quartiles.

Remember that about 50% of the data values lie between the first and third quartiles.

For a set of data, the **5-number summary** consists of the minimum value, the 1st quartile, the median, the 3rd quartile, and the maximum value.

A **boxplot** (box-and-whisker plot) is a type of graph showing the 5-number summary above a number line.

Constructing a boxplot...

Outliers are data values that lie more than 1.5 times the IQR above the 3rd quartile or below the 1st quartile. Compute the limits on the outliers as follows:

Any outliers are

less than Q_1 - 1.5 * IQR

or

greater than Q_3 + 1.5 * IQR.

In a **modified boxplot**, outliers are indicated with asterisks and whiskers are drawn only to the extreme values that are not outliers.