AP Statistics

2019-01-16 1.3 Describing Quantitative Data with Numbers

Notes taken by: **Noah Overcash**

The mean (average, x with bar on top) = (sum of all)/n

- ∑ means "add them all up", so we can write:
  - See "2019-01-16 1.3 Describing Quantitative Data with Numbers - Formula 1.png"
- μ refers to the mean of a population
- The mean is weak as it is sensitive to extreme observations (outliers)
  - Therefore, it is not a **resistant measure** of the center.

To find the median (M):

- Sort items
- M = center, averaging if needed
- The median, unlike the mean, is a resistant measure.
- Symmetric distributions have similar means and medians
- Skewed distributions have the mean farther out into the tail

Interquartile range (IQR)

- Each quartile and the IQR are resistant as they are not affected by a few outliers
- IQR = Q[3] - Q[1]

Outliers are:

- More than 1.5•IQR above Q[3] or
- Less than 1.5•IQR below Q[1]

When outliers are found they should be explained.

- If they are due to bad measurements and/or data entry errors, they can be removed. Otherwise, statistical analysis should be done in a manner which would ignore the outlier.

The **five-number summary** shows a distribution, center, and spread:

- [minimum, Q[1], median, Q[3], maximum]

- Drawing a box-and-whiskers plot
  - Draw a central box from Q[1] to Q[3]
  - Draw a line denoting the median in the box
  - Extend lines from the end to the minimum and maximum (disregarding outliers)
  - Mark outliners with asterisks (*)
- Standard deviation (stddev)
  - S[x] = sample standard deviation
  - σ[x] = population standard deviation
  - n = sample size
  - v (variance) = (std dev)^2
  - Measures the average distance of objects from their mean
  - Because the standard deviation is based on the mean, it is not resistant. A few outliers can significantly increase the stddev.
  - The {sample/population}'s {measurement} vary about {stddev} from the mean {measurement}.
  - Median and IQR are better for describing skewed distributions.