

03/02/2021

Descriptive Statistics for numerical data

Measures of Center (typical): mean (average), median, mode

Measures of Spread: standard deviation, variance, interquartile range (IQR), range

Measures of Position: percentiles, quartiles, deciles

Median is the middle of the data (50% of the data is below, and 50% is above)

Mode is the most common value in the data

Standard deviation: the average distance from the data to the mean

Interquartile range is the middle 50% of the data (distance between third quartile and first quartile)

Variance is the square of the standard deviation

Data Analysis Tool Pack

Mean – report one decimal place more than the original data

Standard deviation – typical to report two more decimal places than the original data

5-number summary : minimum, first quartile, median, third quartile, maximum

Histograms and Boxplots

Rule of thumb for histograms 5-20 bins depending on the amount data

Modal class – the class (bin/grouping) that has the most values in it

Shapes of distributions

Symmetric distributions generally have a central peak and trail off at equal rates on both sides

Skewed distributions have one tail longer than the other: right-skewed have longer right tails, and the left-skewed have longer left tails

Uniform distribution—all the bars are about the same height

Boxplots based on 5-number summary, Excel plots outliers

Mean is affected by outliers more than the median. So in our graphs we have large extreme value make the mean (average) larger than the median. The more skewed the graph is, the bigger the effect on the average (the larger the average will be in our graphs).

Weighted Averages

Calculate an average from a summary table

Things like grade calculations: 15% for homework, 45% for exams, etc.

We'll talk about scaling next time.