**chapter 2. measures of central tendency**

- how to summarize and describe data? (descriptive statistics)
    - one goal: measure of central tendency
        = a single number that reports the centre of a distribution

- mean (average)
    - population mean: parameter, mu
    - sample mean: statistic, xbar
    - add up scores and divide by number of scores
    - write formulas using sigma notation

- median (50th percentile, $P_{50}$)
    - if odd number of data points, use middle data point
    - if even, use average of middle two data points
    - useful when data is skewed, or when there are outliers
    - to find middle data points
        if n is odd, n/2 rounded up
        if n is even, (n/2) and (n/2)+1

- mode (most common value)

- mean, median, and mode are all measures of central tendency
- choice between measures depends on purpose, e.g., mean income (tax collector) vs. median income (sociologist)

- skewed distributions
    - population or sample with many data points at one end of distribution and few at the other (opposite of: symmetric)
    - positively skewed (= right-skewed): tail longer at higher end
    - negatively skewed (= left-skewed): tail longer at lower end
    - for a skewed distribution the mean, median, and mode are usually different
        - mean is pulled up or down, in direction of skew

**chapter 3: measures of variability**

- mean, median, and mode give information about central tendency
    - but no information about how widely the scores are distributed

- measures of variability, or dispersion
    = differences between scores in a distribution

    - range = maximum - minimum
        - not robust; i.e., it is sensitive to outliers

    - interquartile range (IQR): 75th percentile - 25th percentile
        - percentile position: i = (P/100)n ("index method")
            - if integer, use average of data points in position i and i+1
            - if not integer, round up
        - first quartile = 25th percentile
        - second quartile = 50th percentile = median
        - third quartile = 75th percentile

    - variance = average squared deviation

        - formula for population; use sigma notation again
            - need to know the true population mean

        - formula for sample
            - different formulas for population and sample, because the sample mean is an estimate of the population mean, and so using the population formula on a sample gives a biased estimate of the population variance

This is as far as we made it in the lecture. We'll cover more material on measures of variability in the next lecture.