

Class #3 - Monday February 8
Measures of Central Location (Mean & Median); Percentiles and Quartiles

Textbook readings:

- Ross, Sec 3.2: Sample Means & Sec 3.3: Sample Medians
- Phillips, Chapter 3

Key Concepts/Definitions & Spreadsheet Functions

- Sigma (Σ) notation for sums: for a sample of n data points whose values are x_1, x_2, \dots, x_n

$$\Sigma_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n$$

– spreadsheet function: =sum (data)

- **sample mean** (“x-bar”):

$$\bar{x} = \frac{\Sigma_{i=1}^n x_i}{n}$$

– spreadsheet function: =average (data)

- **sample median**: value m such that half of the data points in the sample are smaller than m (and hence half are larger than m)

– spreadsheet function: =median (data)

- **percentiles** (Ross, Sec 3.3.1, pp90-92) the p -th percentile is the value such that $p\%$ of the data points are smaller than that value

– Note that the median is the 50th percentile

– Percentiles are typically used when the number of data points is much larger than 100; often used in health care (e.g., height and weight) and education (e.g., SAT scores)

– spreadsheet function: =percentile (data, 0.95) would output the 95th percentile

- **quartiles** (Ross, Sec 3.3.1, p93): the quartiles break up the data set into quarters

– the 1st quartile (Q1) is the 25th percentile

– the 2nd quartile (Q2) is the 50th percentile, i.e., is the median

– the 3rd quartile (Q3) is the 75th percentile

- Note that the percentiles and quartiles are values in the range of the data set (i.e., are between the max and min)

- The min, 1st quartile, 2nd quartile/median, 3rd quartile, and max are used to create a graph called a **boxplot** (see Ross pp104-105 & WebWork ”HW2-Graphs” #10, 11, 13)