Location and Dispersion

**Notation**

Summation $\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$, often abbreviated $\sum x$  Subscripts can be arbitrary

Bracketed Subscripts: $x_{[1]}$ is lowest, $x_{[2]}$ is second lowest, $x_{[n]}$ is highest.

$x_{[i+\frac{1}{2}]}$ is midway between observation $i$ and observation $i+1$ sorted low to high

**Location**

Mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, median $x_{\left[\frac{n}{2} + \frac{1}{2}\right]}$, mode (most common value), midrange $\frac{x_{[1]} + x_{[n]}}{2}$

$x = \frac{\sum_{i=1}^{n} x_i}{n}$  $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$

**What are they and what are they good for?**

Stability of sample mean, median, mode, midrange.
(Stability not an issue when you have data for the whole population.)

Extremely advanced statistical tools are required to measure the probable error of an estimate of population median or mode based on sample data, so almost everyone estimates the population mean. We will too.

Xbar charts in statistical process control.
  Center line is Xbarbar
Dispersion

Range $x_{[n]} - x_{[1]}$, IQR $x_{[\frac{i}{4}+\frac{1}{2}]} - x_{[\frac{i}{4}+\frac{1}{2}]}$, MAD $\frac{\sum |x_i - \bar{x}|}{n}$, MSD $\frac{\sum (x_i - \bar{x})^2}{n-1}$, RMSD $s = \sqrt{s^2}$

(IQR = Interquartile range, MAD = mean absolute deviation of "average deviation," MSD = Variance or "Mean squared deviation from the mean," RMSD = Standard Deviation or Root mean squared deviation from the mean.)

What are they and what are they good for?

Ignore Section 4.7; we will "simplify" by using Excel

Outliers:
Tukey ("inner") fences p.58-59 ($x_{[\frac{i}{4}+\frac{1}{2}]} - 1.5 \times IQR$, $x_{[\frac{i}{4}+\frac{1}{2}]} + IQR$)

Three Sigma criterion ($\bar{x} - 3\sigma$, $\bar{x} + 3\sigma$)

Range charts and standard deviation charts in Statistical Process Control

Coefficient of Variation $\frac{s}{\bar{x}}$

Pearson Coefficient of Skewness vs. Excel =SKEW( function