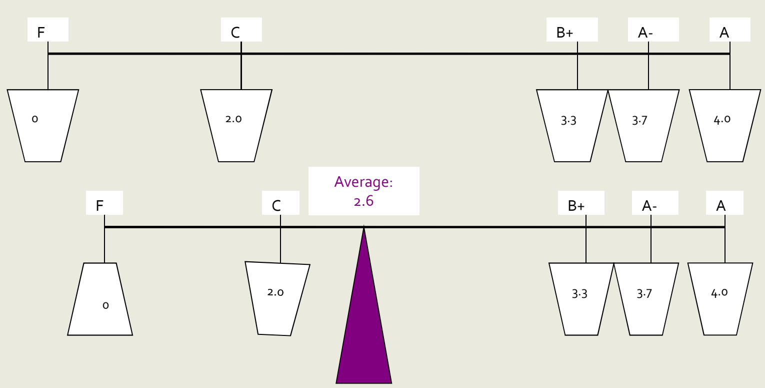
Online Lesson 1

Econ 29000
Kevin R Foster, CCNY
Spring 2011

Center and Spread of the Data

Average

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



Average

- Not necessarily representative: one person gets big bonus; other 99 get nothing, so average rises

Average

- The average is *often* a good way of understanding what happens to people within some group.
- But it is *not always* a good way.
- When projected forward, average is sometimes called “Expected Value”

Median & Mode

- **Median** is value of 50th percentile
- Value of observation in the middle (if odd number of observations) or average of two middle values (if even)
- Median bonus in example above is zero
- **Mode** is most common outcome

Spread around Center

- Want to know variation around the mean/median/mode
- Example: 2 hedge funds both average 10% returns
  – But one returns 9.5%, 10%, 10.5%
  – Other returns 0%, 10%, 20%
- Average deviation won't work; that's always zero (plus and minus cancel out)

Average Deviation = 0

The average of some N values, $X_1, X_2, \ldots X_N$, is given by $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$.

So what is the average deviation from the average, $\sum_{i=1}^{N}\left(X_i - \overline{X}\right)$?

We know that $\sum_{i=1}^{N}\left(X_i - \overline{X}\right) = \sum_{i=1}^{N} X_i - \sum_{i=1}^{N}\overline{X}$ and, since $\overline{X}$ is the same for every observation, $\sum_{i=1}^{N}\overline{X} = N\overline{X} = \sum_{i=1}^{N} X_i$, if we substitute back from the definition of $\overline{X}$. So $\sum_{i=1}^{N}\left(X_i - \overline{X}\right) = 0$. We can't re-use the average. So we want to find some useful, sensible function [or functions], $f(\cdot)$, such that $\sum_{i=1}^{N} f\left(X_i - \overline{X}\right) \neq 0$.

Standard Deviation

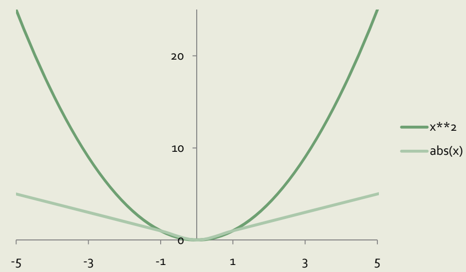
- Simple formula doesn't look simple

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

- Square the deviations so they're all positive (and don't cancel out)
- Then add up the squared deviations
- Take square root of sum

Other distances possible

- Any f( )
- For example often absolute value
- So mean absolute deviation $\frac{1}{N}\sum_{i=1}^{N}|X_i - \bar{X}|$



Distance measure and central measure

- Suppose just searched for a value, W, that minimized the distance, f(X_i – W)
- If f( ) were the absolute value, then set W = median
- If f( ) squares the distance, then set W = average

Coefficient of Variation

- Standard deviation divided by average
- Useful when there is no natural measure
- Such as Likert scale – e.g. "like on a scale of 1-10"
- In finance this is reciprocal of Sharpe ratio, the returns over safe divided by risk

Standardized Data

- Subtract mean
- Divide by standard deviation

$$z_i = \frac{x_i - \bar{x}}{s}$$