Descriptive Statistics:
Central Tendency and Variance

Variables/
Scales of
measurement

Bar graph
Pie Chart
Histogram
Polygon

Mode
Median
Mean

Highway MPG of mini-compact gas-powered cars (N=12)

Ordered Raw Scores: 22 22 23 24 24 24 25 27 28 28 30 31

Frequency Table

Score	Frequency	Cumulative Freq.
22	2	2
23	1	3
24	3	6
25	1	7
26	0	7
27	1	8
28	2	10
29	0	10
30	1	11
31	1	12 (N)
50		

Mode: Most frequently occurring score

What is the mode in this distribution? 24

What happens to the mode when we add an extreme score to the distribution? It stays the same

Mode

- Advantages
  - It does not change when we add an extreme score to the distribution
  - It is the only measure that can be used for nominal data, such as gender
- Disadvantages
  - More than one value may be “most” frequent. No a true mode.
  - The mode varies dramatically from sample to sample (random sampling variation)

Highway MPG of mini-compact gas-powered cars (N=12)

Ordered Raw Scores: 22 22 23 24 24 24 25 27 28 28 30 31

Frequency Table

Score	Frequency	Cumulative Freq.
22	2	2
23	1	3
24	3	6
25	1	7
26	0	7
27	1	8
28	2	10
29	0	10
30	1	11
31	1	12 (N)
40 ~~or~~ 50		

Median: The score about which 50% of the scores fall above and below; middle score, or average of 2 middle scores, in a distribution ordered from lowest to highest (or highest to lowest)

What is the median in this distribution? 24.5

What happens to the median when we add an extreme score to the distribution?
It changes to 25 (middle score in the distribution)

Median

- Advantages
  - Not very sensitive to extreme scores
  - Most stable measure when distributions are open-ended
- Disadvantages
  - May not represent an observed score
  - Not as stable as the mean across random samples
  - Less useful mathematically

Highway MPG of mini-compact gas-powered cars (N=12)

Ordered Raw Scores: 22 22 23 24 24 24 25 27 28 28 30 31

Frequency Table

Score	Frequency
22	2
23	1
24	3
25	1
26	0
27	1
28	2
29	0
30	1
31	1
40 or 50	

Mean: $\overline{X} = \frac{\sum X}{N}$

What is the mean in this distribution?

22+22+23+24+24+24+25+27+28+28+30+31=308

N=12

$\overline{X} = \frac{308}{12} =$ 25.67

What happens to the mean when we add an extreme score to the distribution?

$\overline{X} = \frac{358}{13} =$ 27.54

It changes to 27.54 (more sensitive than the median)

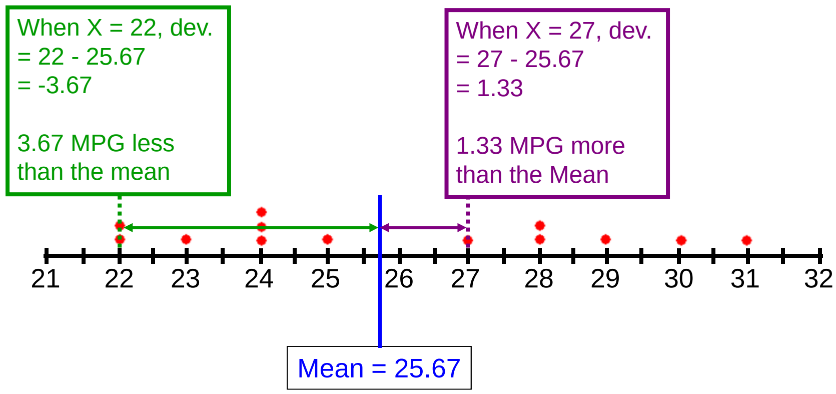
Mean

- The mean is the most frequently used measure of central tendency because:
  - It takes into account all scores in the distribution.
  - It may be used for other statistical computations in ways that the median and mode cannot.
  - It is generally stable from sample to sample.

- But it is NOT the best measure when:
  - The data are nominal or ordinal (use mode instead).
  - The data are skewed or have extreme scores; (use median instead).

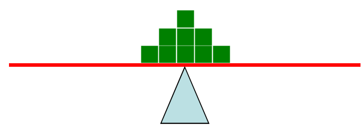
Scores naturally deviate from the mean

$X - \bar{X}$

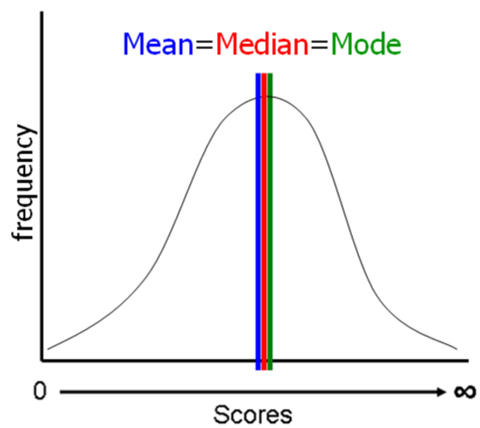


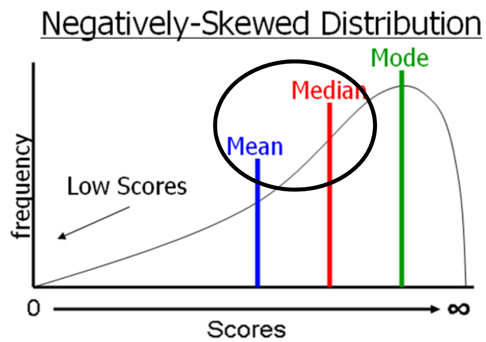
The mean acts like a fulcrum.

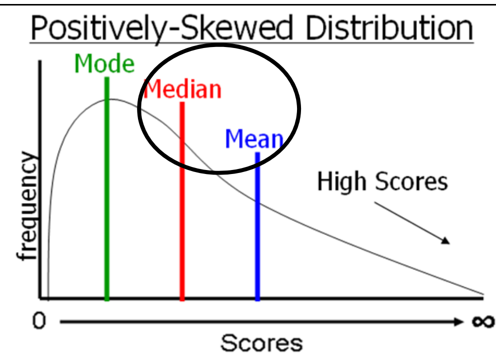
$$\sum d_i = \sum \left(x_i - \bar{X} \right) = 0$$



Symmetrical Distribution







2. Measures of central tendency depict one aspect of a distribution (*viz.,* location). That value, however, is relatively uninformative about the distribution of scores unless it is supplemented with a measure of *dispersion* (i.e., how variable or "spread out" the scores are). There are several ways to numerically represent dispersion, including:

   a) Range (highest score – lowest score)

      i. 4 6 8:

         1. Range: $R = 8 - 4 = 4$
         2. Mean: $\overline{X} = 4 + 6 + 8 = \frac{18}{3} = 6$

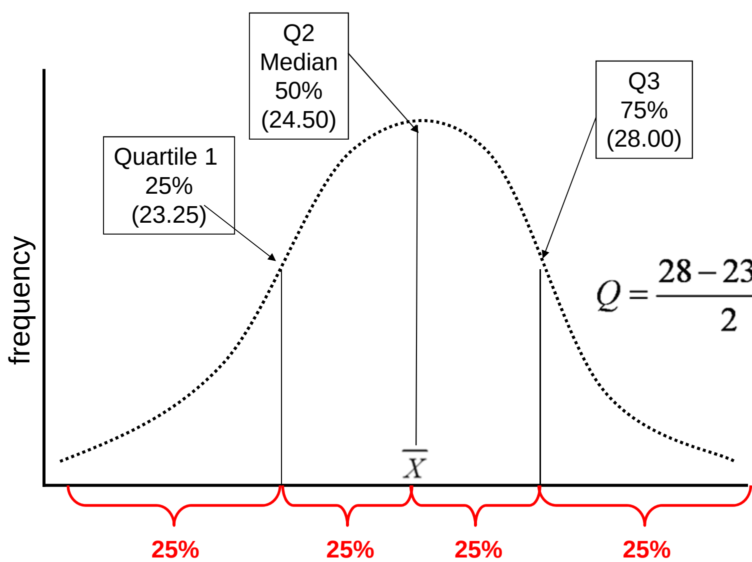
      ii. 2 6 10:

         1. Range: $R = 10 - 2 = 8$ ← **More Dispersion**
         2. Mean: $\overline{X} = 2 + 6 + 10 = \frac{18}{3} = 6$

Quartiles

- Q_1 - point below which 25% of the scores fall (above which 75% fall)
- Q_2 (Mdn) - point below and above which 50% of the scores fall
- Q_3 - point below which 75% of the scores fall (above which 25% fall)

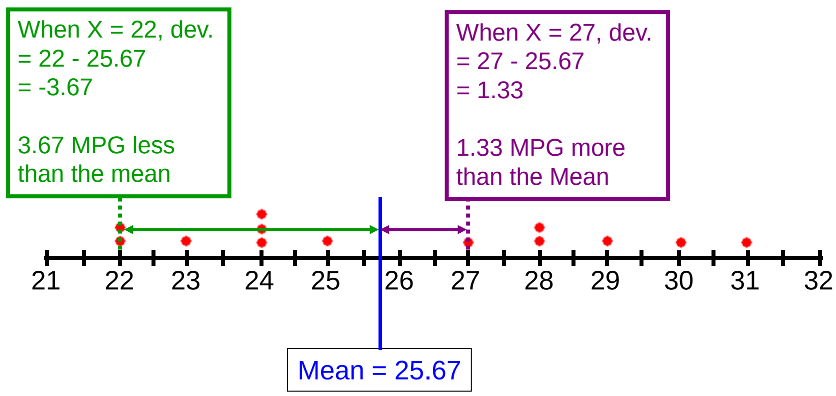
Semi-inter*quart*ile Range $Q = \frac{Q_3 - Q_1}{2}$



$$Q = \frac{28 - 23.25}{2} = 2.375$$

Scores naturally deviate from the mean.

On average, how much?
The variance and the standard deviation help answer this question.



$$\Sigma\left(X_i-\overline{X}\right)^2 = \text{What does all this mean?}$$

$$\Sigma\left(X_i-\overline{X}\right)^2 = \left(X_1-\overline{X}\right)^2+\left(X_2-\overline{X}\right)^2+\left(X_3-\overline{X}\right)^2+\ldots+\left(X_n-\overline{X}\right)^2$$

$$\Sigma\left(X_i-\overline{X}\right)^2 = SS = \text{Sum of Squares, or Sum of Squared values}$$

Example: Exam Scores

X	$\left(X-\overline{X}\right)$	$\left(X-\overline{X}\right)^2$
65	65 - 79.20 = -14.20	201.64
82	2.80	7.84
91	11.80	139.24
73	-6.20	38.44
85	5.80	33.64
$\Sigma X = 396$		$SS_x = 420.80$
$\overline{X} = 79.20$	Σ A.K.A. $sum = 420.80$	

Variance and Standard Deviation

If the sum of squares is $SS = 420.80$

The variance is $SD^2 = \dfrac{SS}{N-1} = \dfrac{420.80}{N-1} = \dfrac{420.80}{4} = 105.10$

And the standard deviation is

$$SD = \sqrt{\frac{SS}{N-1}} = \sqrt{\frac{420.80}{N-1}} = \sqrt{\frac{420.80}{4}} = \sqrt{105.10} = 10.25$$

Standard Deviation

- The standard deviation is the most frequently used measure of variability because (like the mean) it:
  - takes into account all scores in the distribution.
  - is highly stable from sample to sample.
  - is useful for further statistical computations.

How does variability affects the distribution?