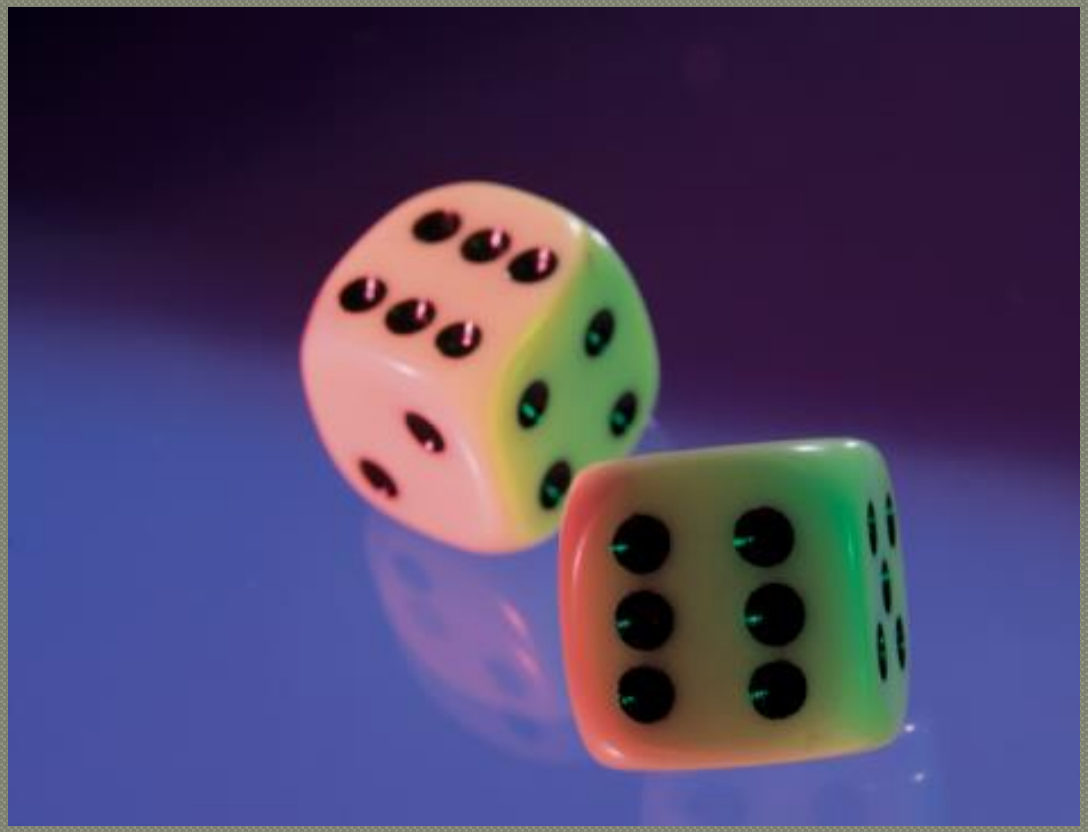
Statistics in Psychology



Why statistics?

- After an experiment or other research method, helps us….
  - 1. Organize/summarize data
  - 2. Describe the data
  - 3. Make inferences

Descriptive Statistics

- Helps us organize material
  - Measures of Central Tendency
    - Marks the “center” of data
    - (mean, median, mode)
    - Data can be “bimodal”
  - Measures of Variability
    - (range, standard deviation)

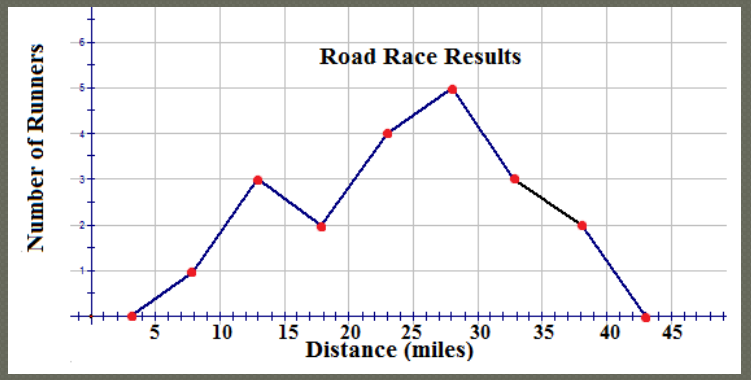
Frequency Distribution

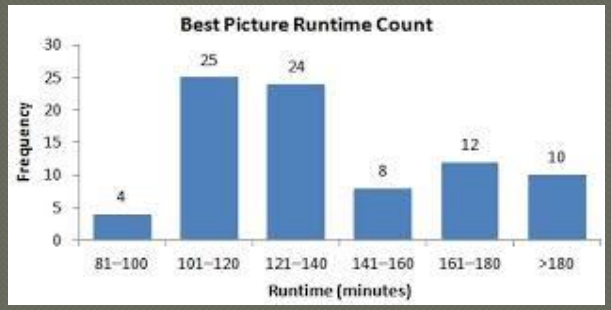
- Table or graph that shows how often different numbers/scores appear in a set of scores
  - Table, frequency polygons, histograph

Frequency Distribution

10/4/04

Score	Frequency	Cum Frequency	Percentile	Percentile Rank
40	1	20	90.91	97.50
35	1	19	79.55	92.50
34	3	18	77.27	82.50
31	1	15	70.45	72.50
29	2	14	65.91	65.00
28	1	12	63.64	57.50
27	2	11	61.36	50.00
26	2	9	59.09	40.00
24	3	7	54.55	27.50
23	1	4	52.27	17.50
22	2	3	50.00	10.00
17	1	1	38.64	2.50





Representing Data

LO 1.14 How Tables and Graphs Represent Patterns in Data

- Frequency distribution: a table or graph that shows how often different scores appear in a set of scores

Table 1.1

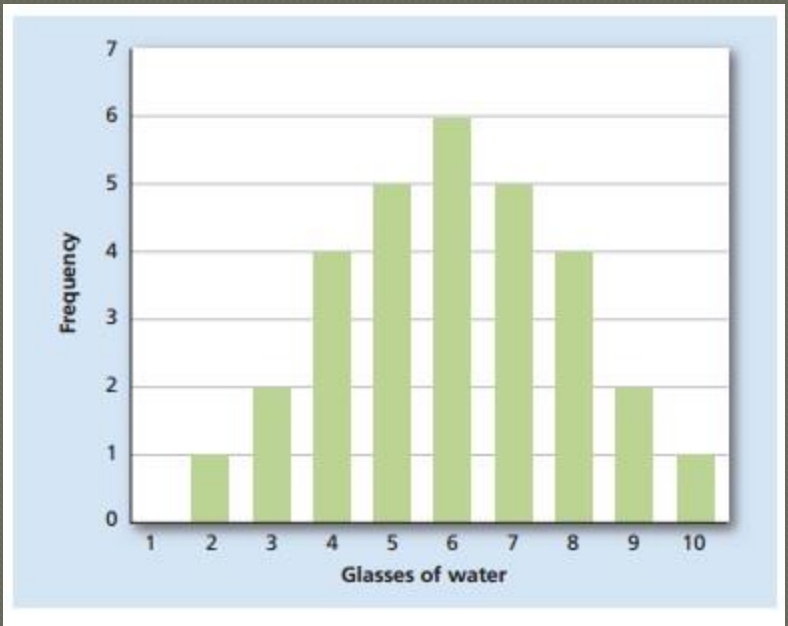
A Frequency Distribution

NUMBER OF GLASSES PER DAY	NUMBER OF PEOPLE OUT OF 30 (FREQUENCY)
1	0
2	1
3	2
4	4
5	5
6	6
7	5
8	4
9	2
10	1

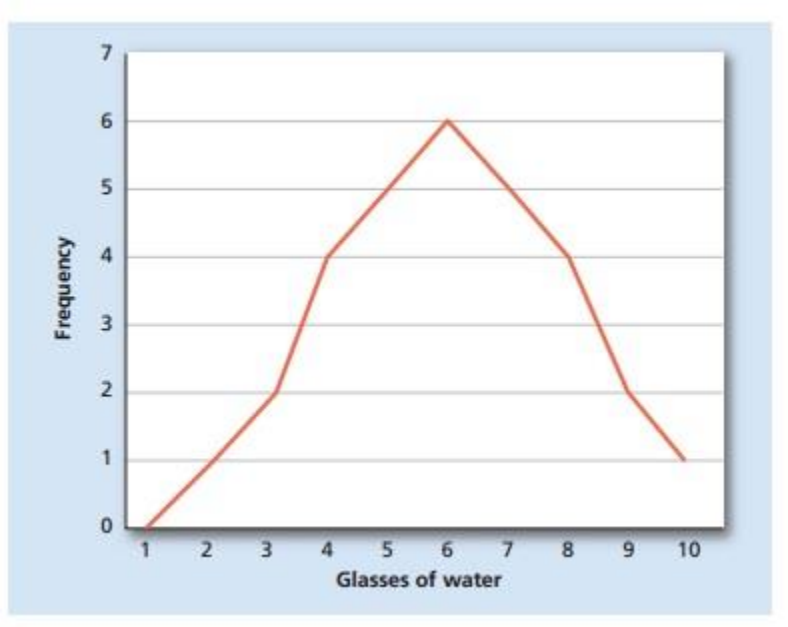
Representing Data

LO 1.14 How Tables and Graphs Represent Patterns in Data

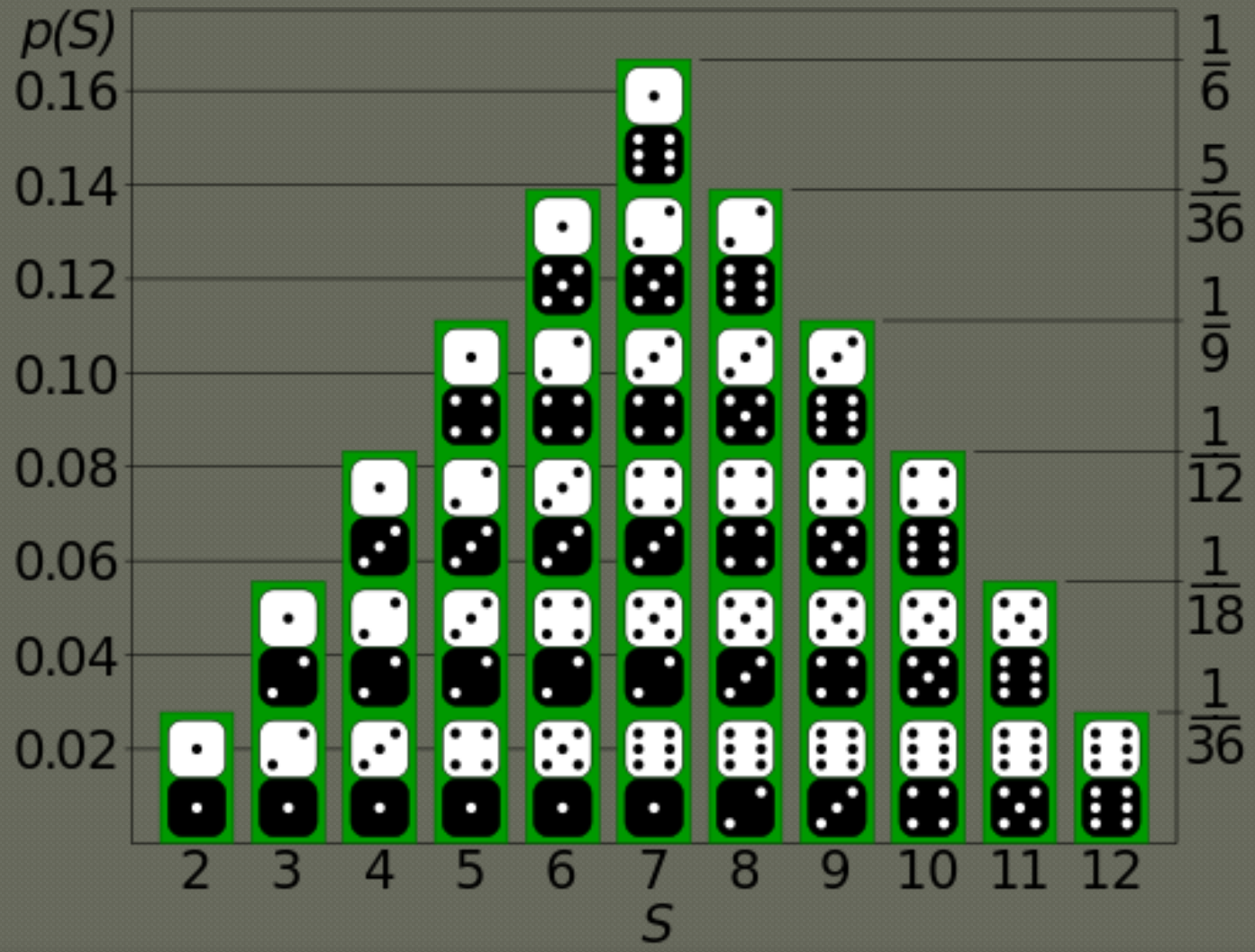
- Histogram (bar graph)



- Polygon (line graph)



Normal Distribution and Dice Activity



Representing Data

LO 1.14 How Tables and Graphs Represent Patterns in Data

- The normal curve (bell curve)
  - Very useful because it has very specific relationships to measures of central tendency and measures of variability
    - Standard deviation

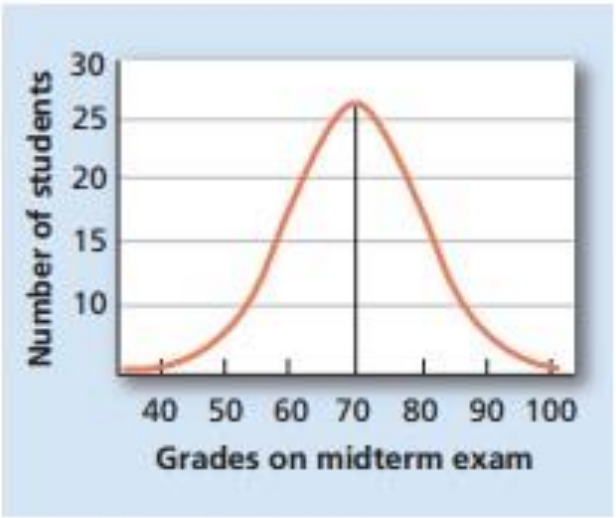


Figure 1.6 The Normal Curve

The normal curve, also known as the bell curve because of its unique shape, is often the way in which certain characteristics such as intelligence or weight are represented in the population. The highest point on the curve typically represents the average score in any distribution.

Measures of Variability

- http://education-portal.com/academy/lesson/measures-of-variability-range-variance-standard-deviation.html#lesson

Standard Deviation

- How far does each score (on average) deviate from the mean?
  - How far away is each score from the average?

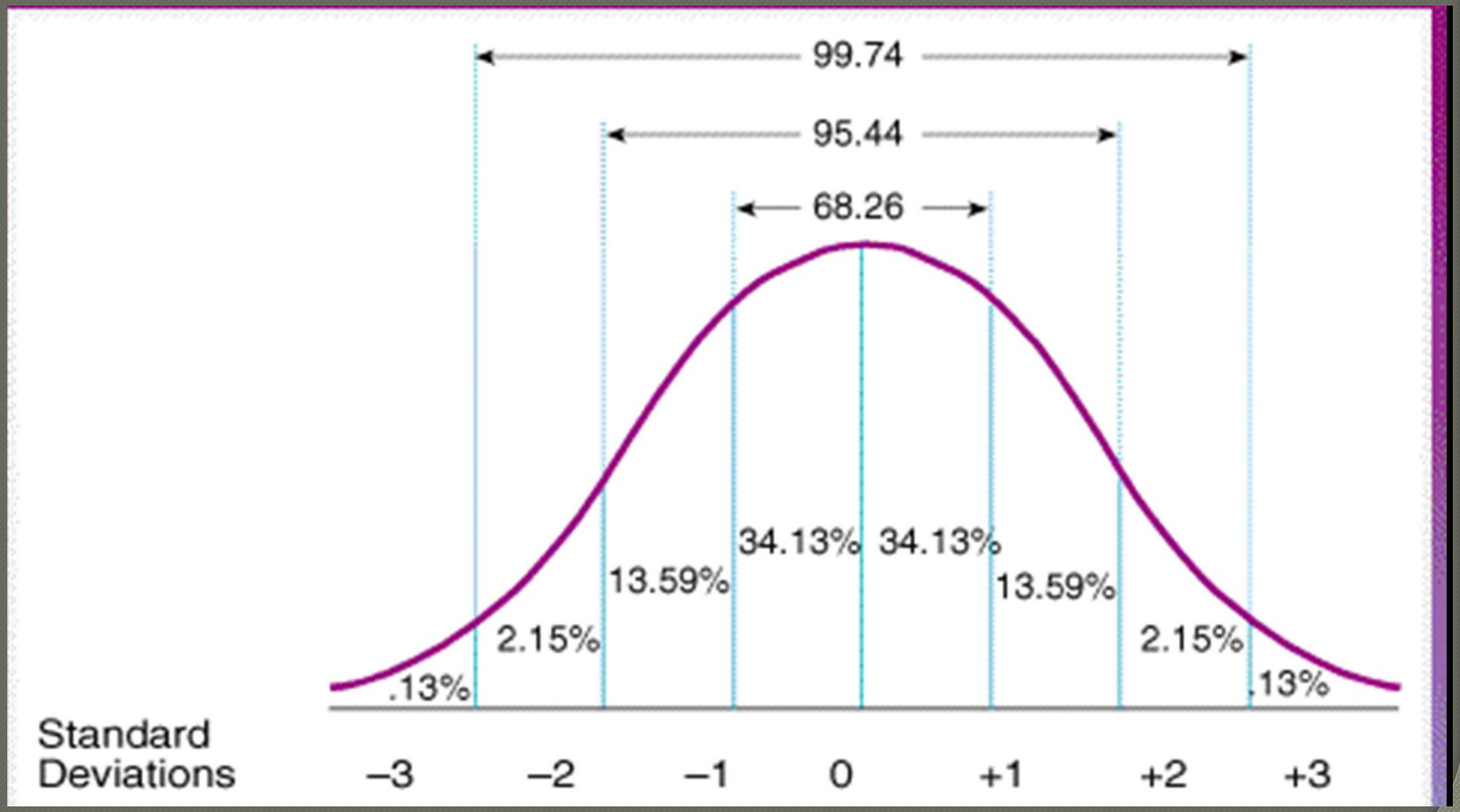
Set 1	**5**	**5**	**5**	**5**	**5**	**5**
Set 2	0	0	0	10	10	10

- Mean of each?
  - 5
- How far does each number deviate from the mean?
  - 1st set:
    - 0 – standard deviation = 0
  - 2nd set:
    - 5 – standard deviation = 5

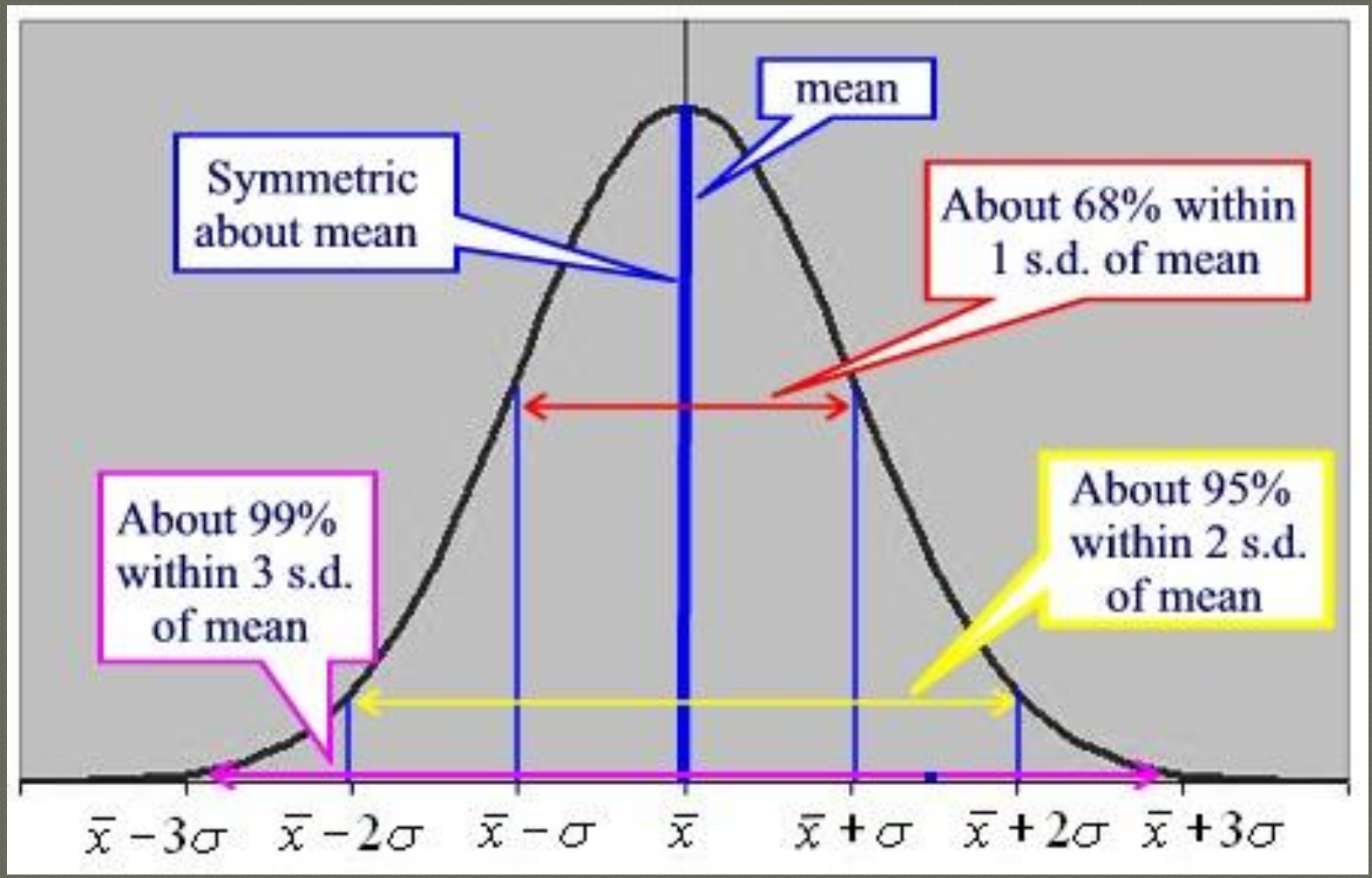
Standard Deviation (cont.)

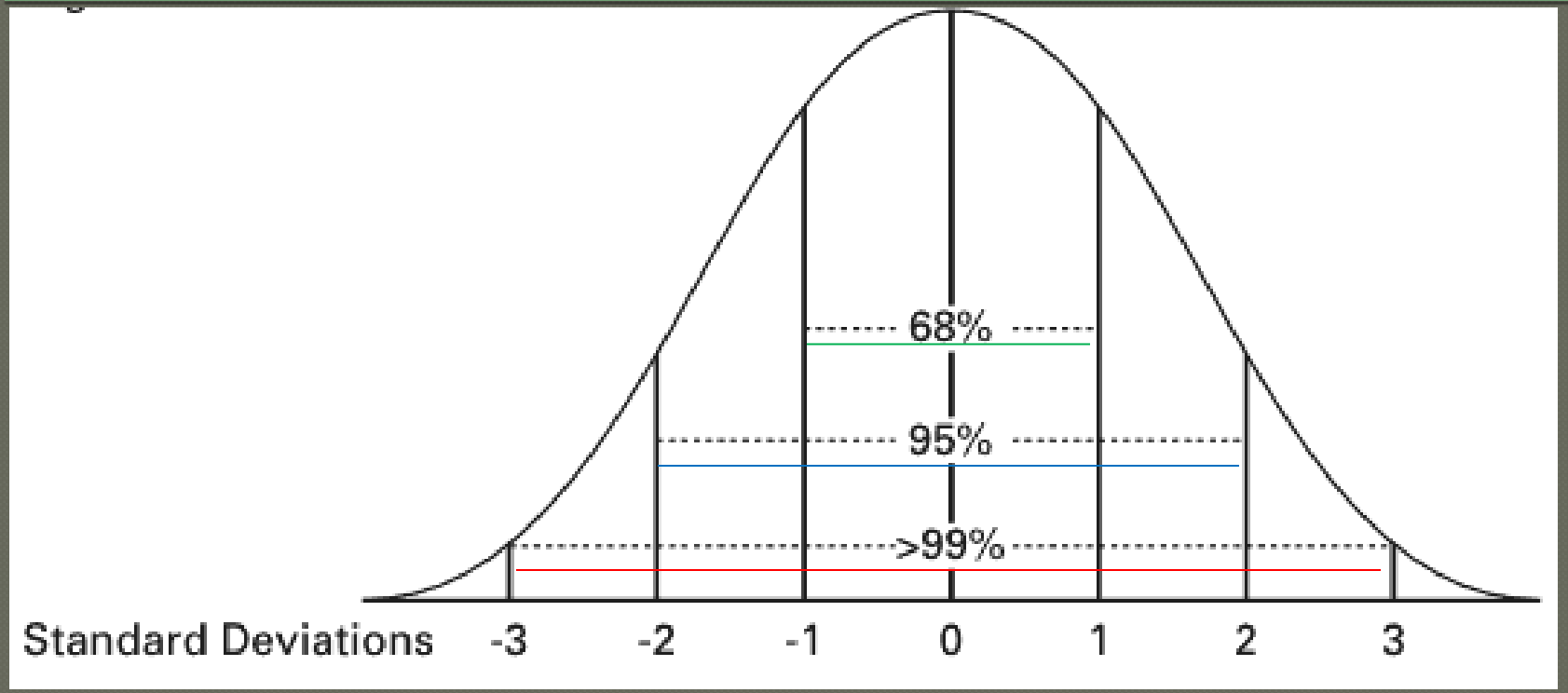
- Z score
  - How many standard deviations is the number/score away from the mean?
- If the mean (average) test score in this class was 70, and the average distance of scores from the mean was 10 (stnd.dev.), and your z score was -2, what was your test score?
  - +1?
  - +.5?
  - 0?
- Tompkins draw this

Normal Distribution and Standard Deviation

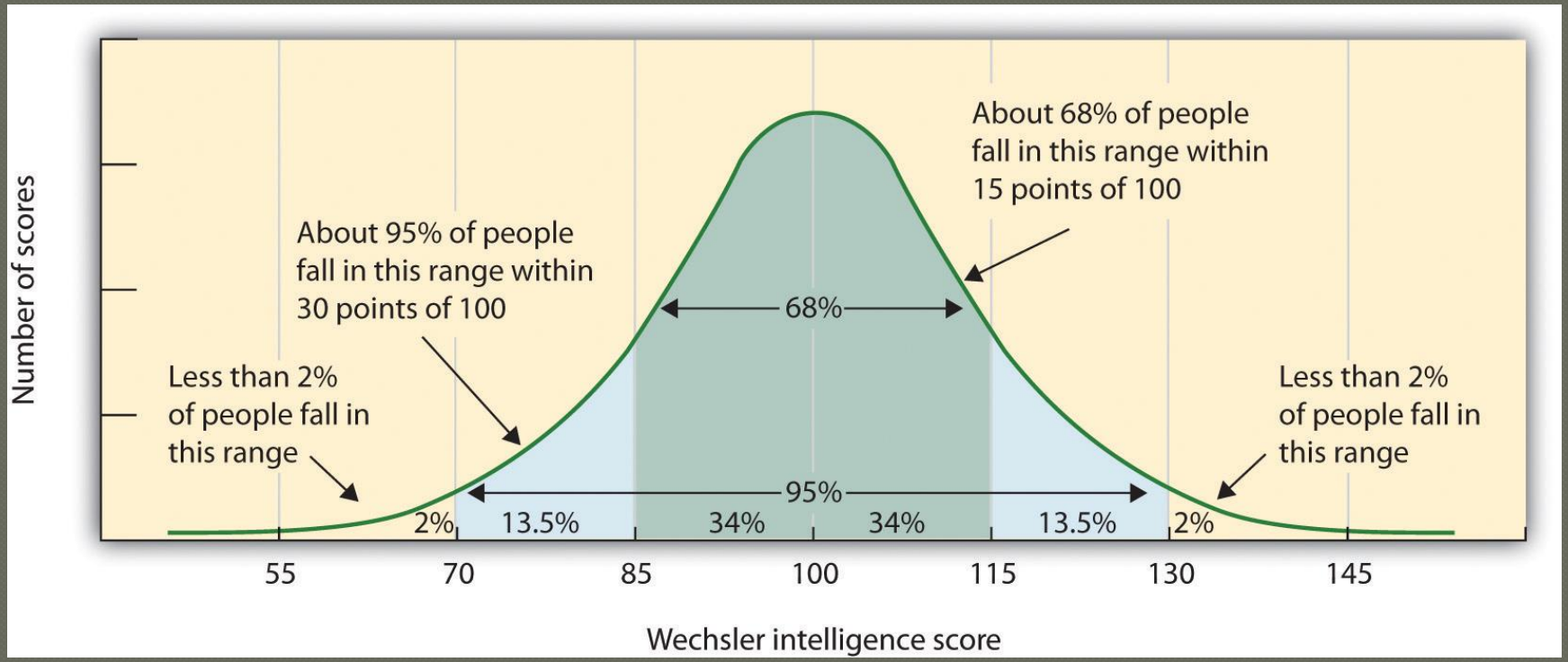


Note z scores and how they fit with percentiles




68%
95%
>99%
Standard Deviations
-3
-2
-1
0
1
2
3

Normal Distribution and IQ



Skewed Distributions

- When a distribution of data includes an extreme score or two (outliers), the distribution will be skewed

- Negatively skewed
  - When the skew is caused by a particularly low score (or group of scores)
- Positively skewed
  - When the skew is caused by a particularly high score (or group of scores)

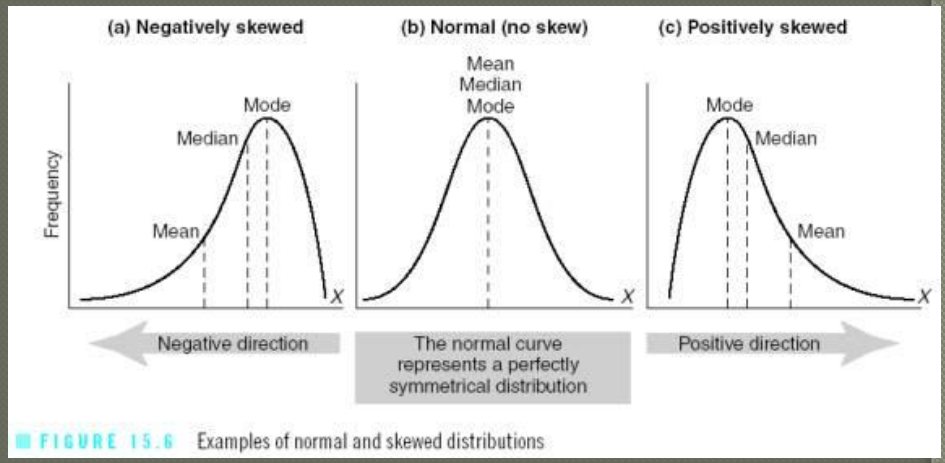


FIGURE 15.6 Examples of normal and skewed distributions

Representing Data

LO 1.14 How Tables and Graphs Represent Patterns in Data

- Skewed distribution
  - Positive or negative

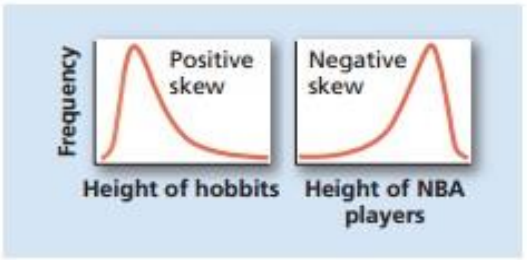


Figure 1.8 Skewed Distribution

These frequency polygons show how distributions can be skewed in two different directions. The graph on the left represents the frequency of heights among hobbits (the little people from the fantasy *The Lord of the Rings*) and is positively skewed because the long "tail" goes to the right, or positive direction. The graph on the right shows the frequency of heights among NBA basketball players and is negatively skewed—the tail points to the left.

- Bimodal distribution

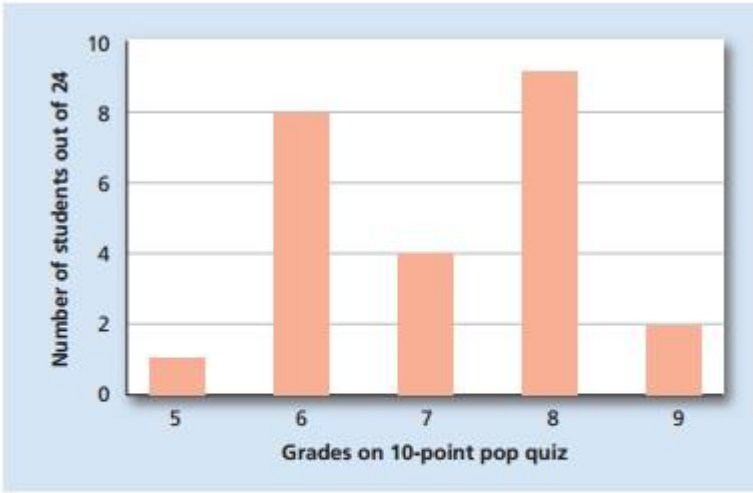
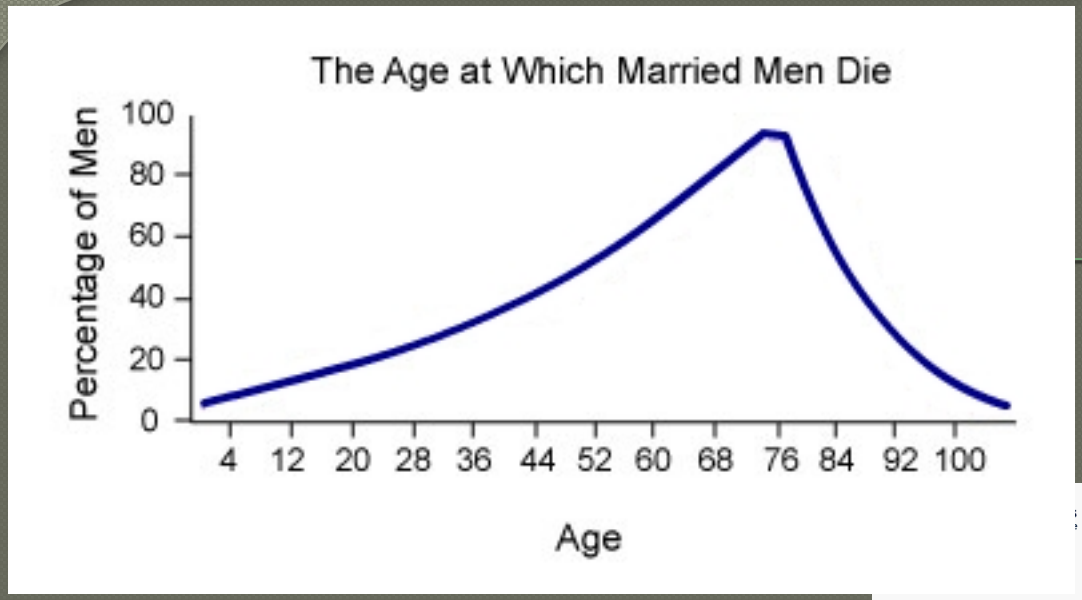
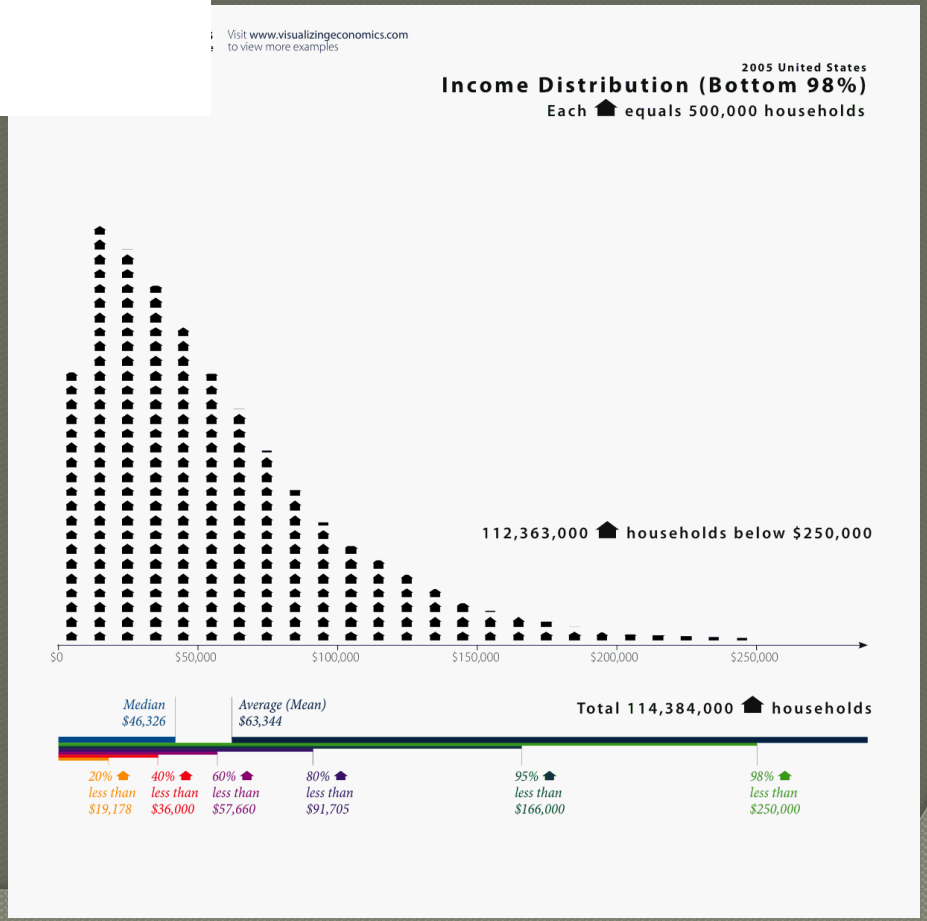


Figure 1.9 A Bimodal Distribution

When a distribution is bimodal, it means that there are two high points instead of just one. For example, in the pop-quiz scores represented on this graph there are two "most frequent" scores—6 and 8. This most likely represents two groups of students, with one group being less successful than the other.


The Age at Which Married Men Die
Percentage of Men
100
80
60
40
20
0
4 12 20 28 36 44 52 60 68 76 84 92 100
Age


Visit www.visualizingeconomics.com
to view more examples
2005 United States
Income Distribution (Bottom 98%)
Each ⌂ equals 500,000 households
112,363,000 ⌂ households below $250,000
$0
$50,000
$100,000
$150,000
$200,000
$250,000
Median
$46,326
Average (Mean)
$63,344
Total 114,384,000 ⌂ households
20% ⌂ less than $19,178
40% ⌂ less than $36,000
60% ⌂ less than $57,660
80% ⌂ less than $91,705
95% ⌂ less than $166,000
98% ⌂ less than $250,000

Statistics and Central Tendencies

LO 1.15 Statistics that Examine Central Tendencies

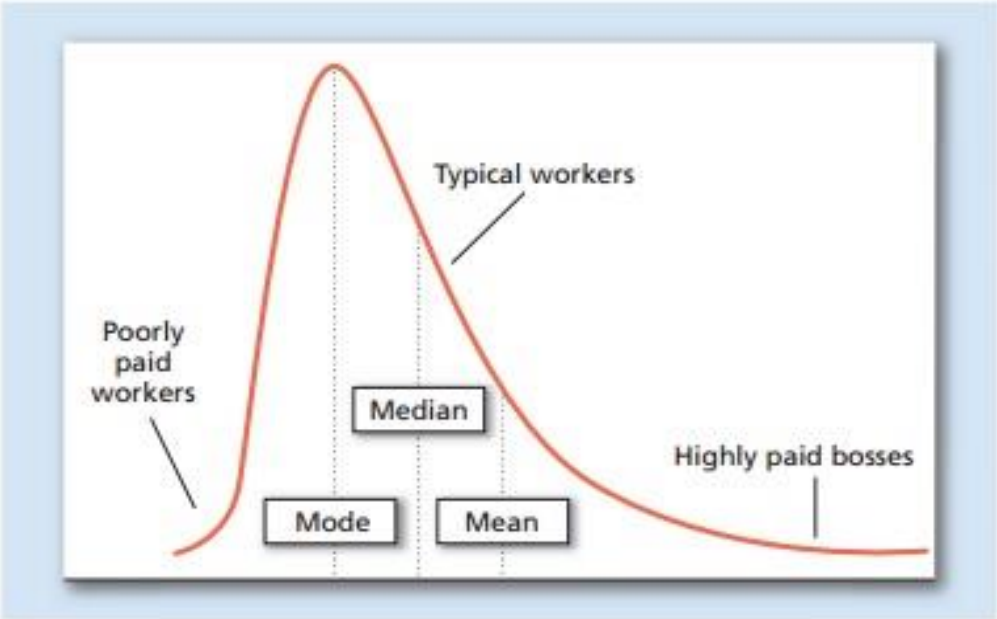


Figure 1.10 Positively Skewed Distribution

In a skewed distribution, the high scores on one end will cause the mean to be pulled toward the tail of the distribution, making it a poor measure of central tendency for this kind of distribution. For example, in this graph many workers make very little money (represented by the mode) while only a few workers make a lot of money (the tail). The mean in this case would be much higher than the mode because of those few high scores distorting the average. In this case, the median is a much better measure of central tendency because it tends to be unaffected by extremely high or extremely low scores such as those in this distribution.

Skewed Distribution

Inferential Statistics

- Don’t describe data…. These types of statistical data allow researchers to ***infer*** if the data can be applied to the larger population

Statistical Significance

- Can the results of this experiment be generalized to the broader public? Could it be that the results really only fit this group?
- Statistical significance uses math equations to determine the probability (p) that the results of the experiment happened by chance.
- We can say that something is “statistically significant” if $p<.05$
  - The probability that the results happened by chance is less than 5%

Measures of Variability

LO 1.16 Statistics that Examine Variations in Data

Figure 1.11 IQ Normal Curve

Scores on intelligence tests are typically represented by the normal curve. The dotted vertical lines each represent one standard deviation from the mean, which is always set at 100. For example, an IQ of 116 on the Stanford-Binet Fourth Edition (Stanford-Binet 4) represents one standard deviation above the mean, and the area under the curve indicates that 34.13 percent of the population falls between 100 and 116 on that test. The Stanford-Binet Fifth Edition was published in 2003 and it now has a mean of 100 and a standard deviation of 15 for composite scores.