

Descriptive Statistics

Measures of Central Tendency

Descriptive Statistics

Summarizing Data:

- Central Tendency (or Groups' "Middle Values")
 - Mean
 - Median
 - Mode

Mean

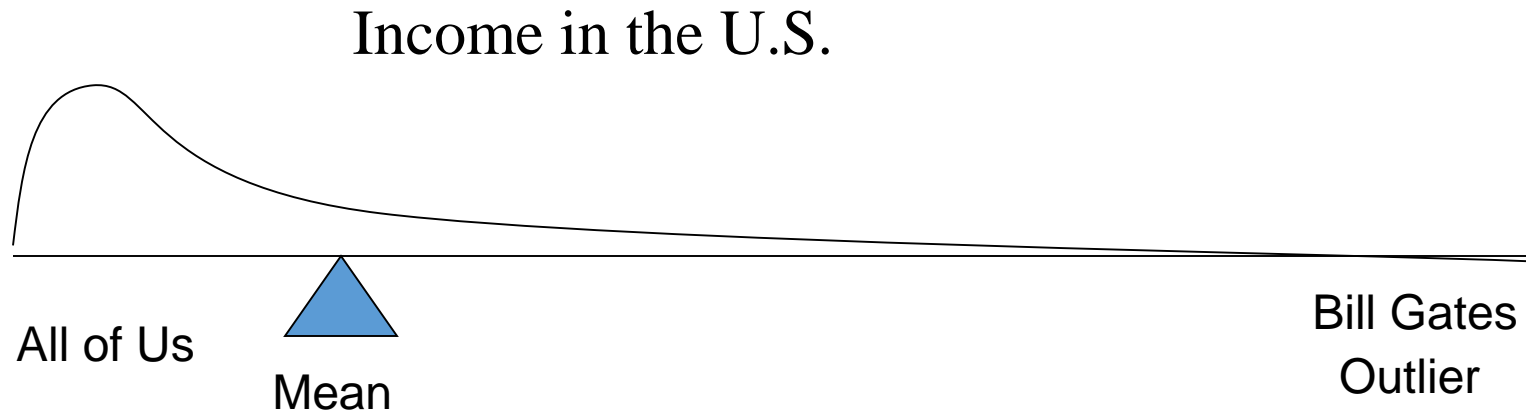
Most commonly called the “average.”

Add up the values for each case and divide by the total number of cases.

$$\frac{\sum_{i=1}^n x_i}{n} \equiv \mu \equiv \bar{X}$$

Mean

1. Means can be badly affected by outliers (data points with extreme values unlike the rest)
2. Outliers can make the mean a bad measure of central tendency or common experience



Median

The middle value when a variable's values are ranked in order; the point that divides a distribution into two equal halves.

When data are listed in order, the median is the point at which 50% of the cases are above and 50% below it.

The 50th percentile.

Median

Class A--IQs of 13 Students

89

93

97

98

102

106

109

110

115

119

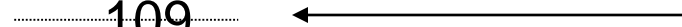
128

131

140

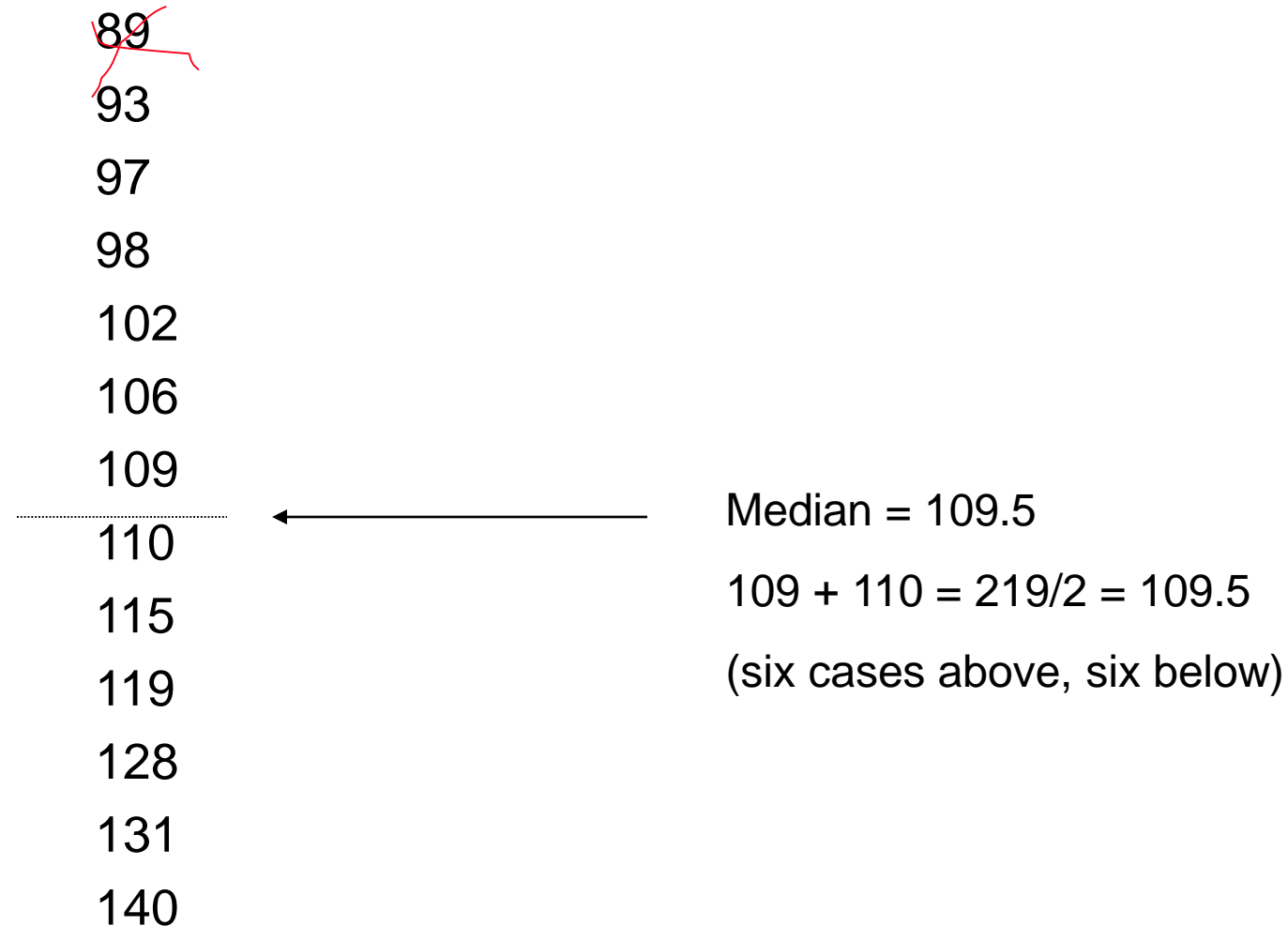
Median = 109

(six cases above, six below)



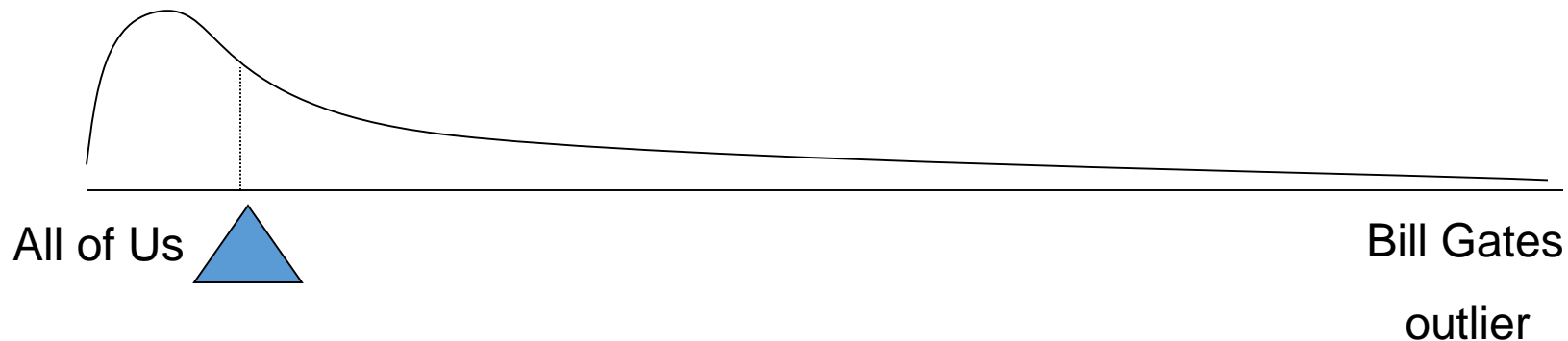
Median

If the first student were to drop out of Class A, there would be a new median:



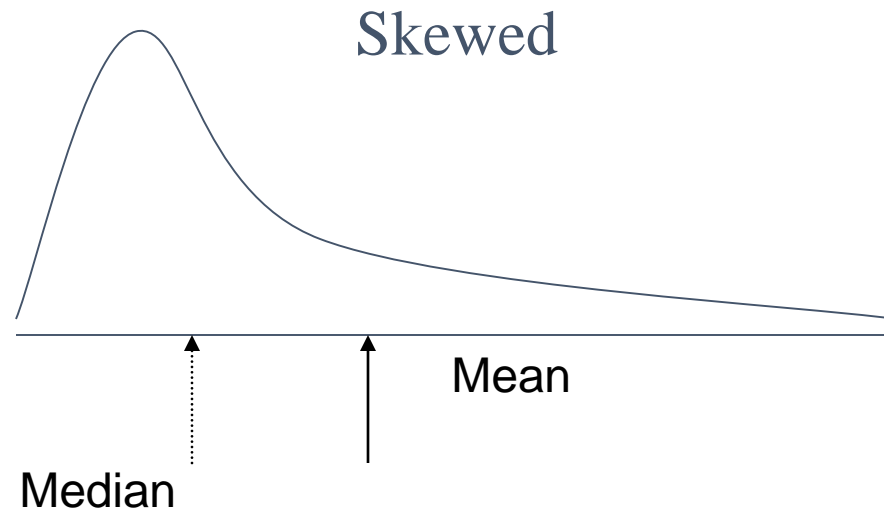
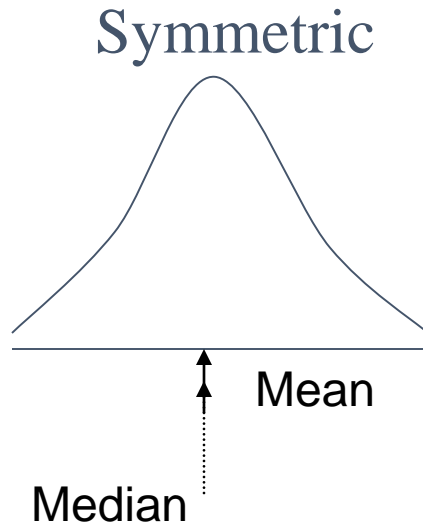
Median

1. The median is unaffected by outliers, making it a better measure of central tendency, better describing the “typical person” than the mean when data are skewed.



Median

2. If the recorded values for a variable form a symmetric distribution, the median and mean are identical.
3. In skewed data, the mean lies further toward the skew than the median.

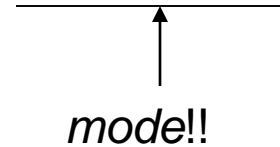


Mode

The most common data point is called the mode.

The combined IQ scores for Classes A & B:

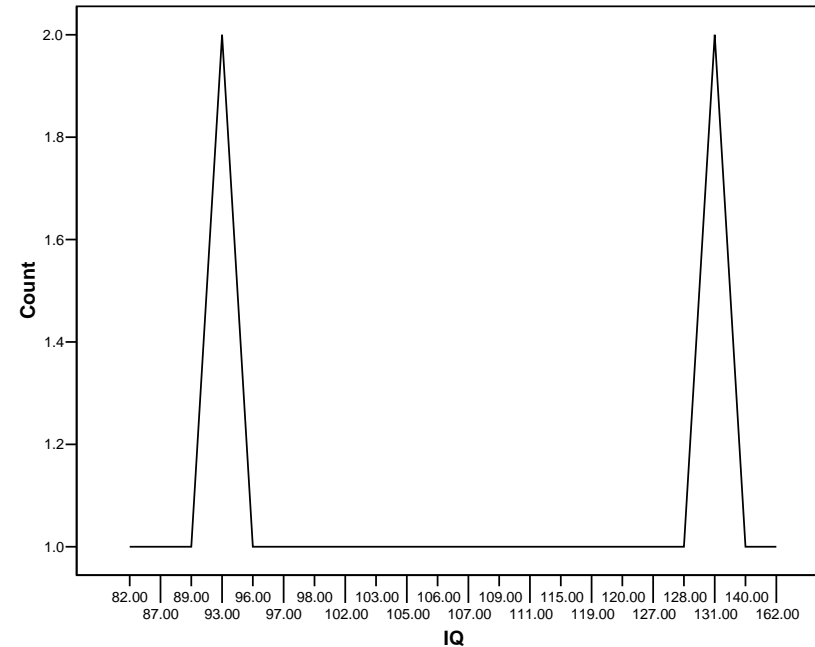
80 87 89 93 93 96 97 98 102 103 105 106 109 109 109 110 111 115 119 120
127 128 131 131 140 162



Mode

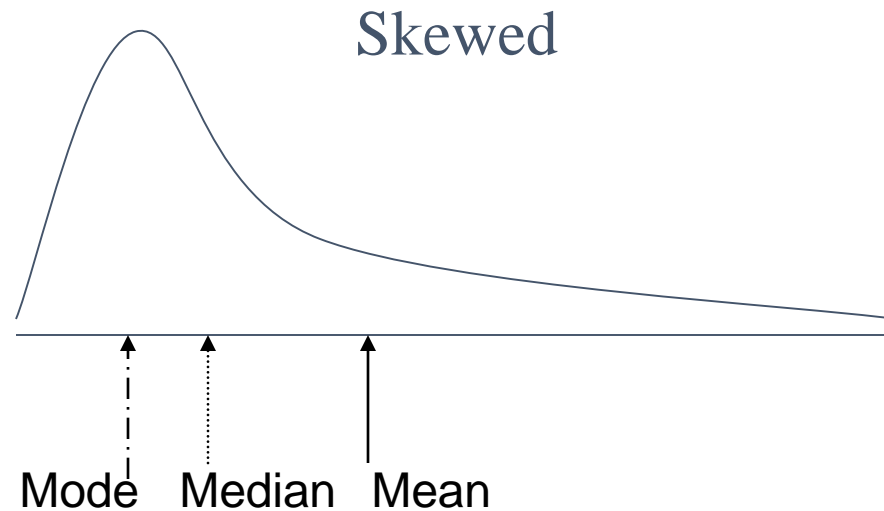
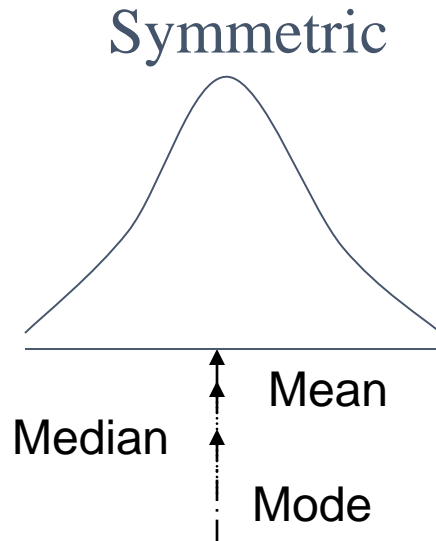
It may not be at the center of a distribution.

Data distribution on the right is “bimodal” (even statistics can be open-minded)



Mode

1. It may give you the most likely experience rather than the “typical” or “central” experience.
2. In symmetric distributions, the mean, median, and mode are the same.
3. In skewed data, the mean and median lie further toward the skew than the mode.



Summary: Measures of the Central Tendency

- Mean (algebraic measure) (sample vs. population):
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \mu = \frac{\sum x}{N}$$
 - Weighted arithmetic mean:
$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$
 - Trimmed mean: chopping extreme values
- Median: A holistic measure
 - Middle value if odd number of values, or average of the middle two values otherwise
 - Estimated by interpolation (for *grouped data*):
$$median = L_1 + \left(\frac{n/2 - (\sum f)l}{f_{median}} \right) c$$
- Mode
 - Value that occurs most frequently in the data
 - Unimodal, bimodal, trimodal
 - Empirical formula:
$$mean - mode = 3 \times (mean - median)$$