Math 122
Introduction to Statistics
Measures of Center

MEAN

Mean/Average: The mean or average of a set of quantitative data is the sum of all of the values divided by the number of values.

Notation:

- Σ (Greek Sigma) denotes a SUM
- x - variable to denote individual data values
- Σx - sum of all values of x
- n - number of data values in a SAMPLE
- N - number of data values in a POPULATION
- $\bar{x}$ - the mean or average of a SAMPLE
- μ - the mean or average of a POPULATION

Sample Mean: $\bar{x} = \dfrac{\sum x}{n}$

Population Mean: $\mu = \dfrac{\sum x}{N}$

Example: The mean of the values 3, 4, 2, 3, 5, 6, 7, 5, 4, 5 is

$$\frac{3+4+2+3+5+6+7+5+4+5}{10} = \frac{44}{10} = 4.4$$

Reliability: The mean is reliable in the sense that different samples from the same population tend to have similar means which are close to the population mean.

Resistence: The mean is not resistent in the sense that it is affected by outliers.

Quantitative vs. Categorical: The mean is only useful for quantitative data.

MEDIAN

Median: The median of a set of quantitative data is a number which is greater than or equal to half of the data and which is less than or equal to half of the data.

Notation: $\tilde{x}$ denotes the median.

Even Number of Values: To find the median of an even number of values, sort the data and then average the middle two data values.

Example: To find the median of

$$3, 4, 2, 3, 5, 6, 7, 5, 4, 5$$

we first sort the data to get

$$2, 3, 3, 4, \mathbf{4}, \mathbf{5}, 5, 5, 6, 7.$$

We then average the middle two values 4 and 5 to get $\tilde{x} = 4.5$.

Odd number of Values: To find the median of an odd number of values, sort the data and the select the middle data value.

Example: To find the median of

$$3, 4, 2, 3, 5$$

we first sort the data to get

$$2, 3, \mathbf{3}, 4, 5.$$

The median is the middle value, $\tilde{x} = 3$.

Reliability: The median is not reliable in that medians of samples from the same population do not tend to be similar to each other or to the population median.

Resistence: The median is resistent in the sense that it is not affected by outliers.

Quantitative vs. Categorical: The median is only useful for quantitative data.

MODE

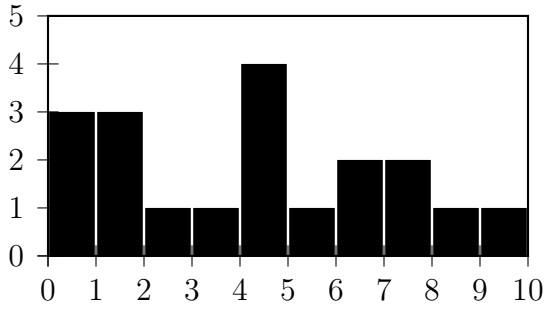
Mode: The mode of a data set is the most common data value.
Example: The mode of 2, 3, 3, 4, 4, **5, 5, 5**, 6, 7 is 5.
Example: The mode of a, a, a, **b, b, b, b**, c, c, c is b.

Number of Modes: If no data value is repeated (as in 1, 2, 3, 4, 5), then there is no mode. If two data values occur with the same greatest frequency (as in 1, 1, 1, 2, 2, 2, 3, 4, 5), then each is a mode, and the data is bimodal. If several data values occur with the same greatest frequency (as in 1, 1, 2, 2, 3, 3, 4, 5), then each is a mode. The data is multimodal.

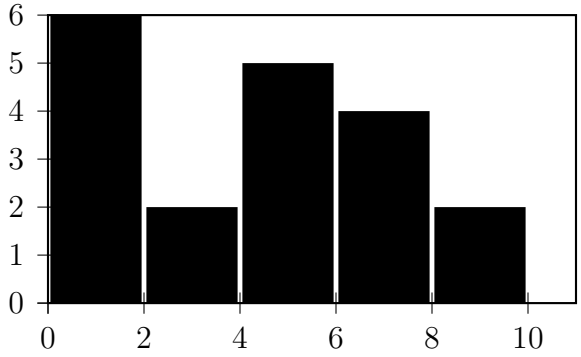
Histograms: The mode is often visible in histograms as the highest peak.
Example: For example, consider the data

$$0, 0, 0, 1, 1, 1, 2, 3, 4, 4, 4, 4, 5, 6, 6, 7, 7, 8, 9.$$

The mode of this data is 4. This is clear in this histogram which uses a class width of 1:



However this is a histogram of the same data with a class width of 2:



This histogram makes it appear as if the mode is less than 2.

Reliability: The mode is not reliable.

Resistence: The mode is resistent.

Quantitative vs. Categorical: The mode is useful for quantitative and categorical data.

MIDRANGE

Midrange: The midrange of a set of quantitative data is the average of the highest and lowest data values.

Example: The midrange of 2, 3, 3, 4, 4, 5, 5, 5, 6, 7 is $\dfrac{2+7}{2} = 4.5$.

Reliability: The midrange is not reliable.

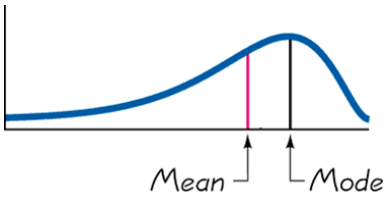
Resistence: The midrange is not resistent.

Quantitative vs. Categorical: The midrange is only useful for quantitative data.

SKEWNESS

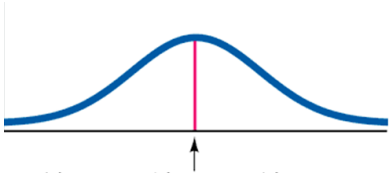
Measures of center can often be used to determine skewness.

Skew Left: For data that is skew left, the mean is usually less than the mode.



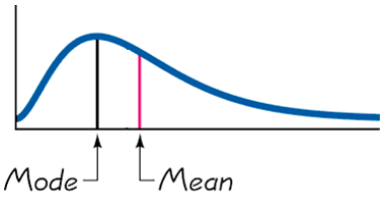
Mean $<$ Mode

Symmetric: For data that is symmetric, the mean is often equal to the mode and the median.



Mean = Mode = Median

Skew Right: For data that is skew right, the mean is often greater than the mode.



Mean $>$ Mode