

Pre-Processing Data



- I. nothing ... the raw data are what's needed
- 2. "condensed" score ... a variety of different raw measures are combined (in some way) to create what's needed
- 3. "summary" score ... a set of identical measures are combined (almost always by averaging) to create what's needed

Descriptive vs Inferential Stats

- Descriptive Stats summarize a given set of data the set of data is usually a sample, not the entire population; because these are "just" summaries, they can't be wrong
- Inferential Stats go beyond a given set of data and make a (probabilistic) statement about the population from which the sample was taken
 - the statement is a best guess about the sampled population; because it's a guess (of some sort), it can be wrong

Summarizing Univariate Data

- Function: to pass lots of information in few words
- Center, Spread, and Shape of Distribution

Center: mean, median, or mode

- Spread: standard deviation (variance), IQR, or range
- Shape: skew & kurtosis or name

Summarizing Univariate Data

Center: (arithmetic) mean

$$\overline{x} = \frac{x_1 + x_2 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

Spread: standard deviation

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

Summarizing Univariate Data

Shape: name

