

Overview of Basic Concepts: Part 1

Two types of statistics

1. **Descriptive** = describes the data with graphs, words, & numbers.
 - a. Central tendency & variation (mean, median, mode, range, standard deviation, variance)
 - b. Graphical representations (bar graph, histogram, pie charts, line graph)
 - c. Variable inter-relationships (correlations)
2. **Inferential** = Uses what we know about a sample to *infer* about the larger population, using the concept of probability.

Two types of variables

1. **Qualitative** – quality of something. In SPSS – string variables
2. **Quantitative** – has a value, and levels indicate different amounts
 - a. We can do statistical analyses on quantitative variables. Qualitative variables are very informative, but we must change them into quantitative values to work with them

Scales of measurement

1. **Nominal**: numbers are used as names or categories, and have no value (e.g., sex, ethnicity)
2. **Ordinal**: numbers have an order (e.g., place in a race)
3. **Interval**: equal intervals between the variables. 0 has no value
 - a. Many of our scales in psychology are ordinal, but we treat them as an interval scale
4. **Ratio**: has a true zero point, and variables are a ratio of one another (e.g., weight, income)

Some definitions:

- **Statistics** involve the numbers that we get, how we manipulate those numbers, and the conclusions we draw.
- **Experimental design** involves all the things that influence how we get those numbers.
- **Experiment**: A study that manipulates one or more variable(s), hoping to find a specific outcome.
- **Correlational (observational)**: A study about the relation between variables (variables are not manipulated).
- **Quasi-Experiment**: An experiment done in a real life setting.
- **Independent variables (IV)**: The variable you manipulate or try to predict from
- **Dependent variables (DV)**: Outcome variable – the thing we are measuring
- **Extraneous variables**: Variables other than the IV that may affect the DV

Part 1: Describing data

Frequency distributions

- **Raw scores** = scores obtained by observation or from an experiment
- **Frequency** = how often each score occurs

Central tendency (the middle of the data)

- **Mean**: “the average”- the value after adding up all values in a distribution and dividing that value by the number of values that were added together.
- **Median**: “the middle” – the number that cuts the distribution in half.
- **Mode**: “the most”- the number that is repeated the most number of times in a distribution.

Variability (the spread of the data)

- **Range:** Highest score minus the lowest score.
- **Interquartile range:** Range of scores that captures the middle 50% of the data
- **Percentile score:** Point at which a specified percentage of the distribution falls below this score
- **Standard deviation** = the dispersion of scores around the mean
- **Variance** = squared standard deviation

Describing distributions:

- **Center:** Where is the middle of the distribution?
- **Spread:** Are the values close to the center or spread out?
- **Shape:** What does the distribution look like?
 - Symmetrical (normal or bell-shaped), positive or negative skew, J shape, rectangular, bimodal

Descriptive reports:

- Brings together all of the important descriptive statistics we have talked about
- Most commonly include:
 - Means (medians if skewed data; mode with categorical variables)
 - Standard deviation
 - Include an effect size
 - Graphical representation of things (e.g., box & whiskers plot, frequency polygon)
- Tells a story about the data!

Correlational research

- Considers the relation between different variables
- Non-experimental – can't determine cause and effect
- Used a lot in social sciences

Correlations:

- The degree to which two variables are related.
- Important note: Correlation \neq causation!
- Correlation indicates the degree to which two variables are related.
- The variables are related in a linear fashion
- Correlations can be positive or negative
 - **Positive correlation:** As one variable goes up, the other goes up
 - **Negative correlation:** As one goes up, the other goes down
 - **Zero correlation:** No relation between the two variables

Pearson r correlation coefficient

- Descriptive statistic that expresses the degree of relation between two variables.
- **Properties:**
 - Ranges from -1 to 1
 - Comes from a bivariate distribution
 - There is no IV and DV per se, but we assign one variable as X and one as Y
- **A rough guide** (according to Cohen)
 - .10 (or -.10) = small effect
 - .30 (or -.30) = moderate effect

- .50 (or -.50) = strong effect
- But it depends on the variables!!!
- In explaining correlations:
 - **Strength**- how big (far away from zero in either direction) or small (close to zero) is the correlation?
 - **Direction**- is the correlation positive or negative?
 - **Explain** the relation between the two variables in words.

Linear regression

- A technique that uses data from two variable to create a straight line, and then use this line to make predictions.
 - **Correlation = non directional.** In correlation, we had two variables, arbitrarily defined as X and Y.
 - **Regression = directional.** In regression, X is the predictor (IV) and Y is the outcome (DV).
- We are inferring that X leads to Y (BUT, this doesn't mean X causes Y!)

Computer programs

- **Excel**
 - Advantages: Offers an easy way to do a lot of basic descriptive statistics; Makes some nice graphs and charts for displaying results; Used a lot in business and related fields
 - Disadvantages: No more complex stats (that's what we have special programs for)
- **SPSS**
 - Advantages: Very popular in the social sciences; Point and click icons and menus make it easy to do a lot of powerful analyses
 - Disadvantages: Not as good at large-scale data management; Special program – expensive to buy (though student version are cheaper)
- **SAS**
 - Advantages: very popular in social sciences. Good for powerful analyses and large, complicated datasets
 - Disadvantages: syntax based, not user friendly
- And a host of other programs, ranging from basic statistics to VERY advanced stats (e.g., Amos, EQS, Lisrel, Mx, Mplus, Systat)
 - Other programs are more specialized – depends on what analyses you are trying to do.
- **Computers are a great tool, BUT, you need to know what numbers are being entered, what analyses are being done, and what any results you get mean.**