The California State University
WORKING FOR CALIFORNIA

Mark Van Selst

San Jose State University

Cozby and Bates: Methods in Behavioral Research

Chapter 12 Understanding Research Results: Description and Correlation

Summer 2014

Approaches to Research

DESCRIPTIVE

- Describes situational or personal characteristics at a particular place and time

CORRELATIONAL (Quasi-Experimental)

- GOAL: TO UNCOVER SYSTEMATIC RELATIONS
- Statistical analysis: correlation coefficient (pearsons r)
- r= +1 >r=0> r= -1
- Make predictions
- Test predictions (theory)
- Causal relations cannot be inferred on basis of correlation (requires temporal order, causal path, association)

EXPERIMENTAL

- Manipulate variables of interest (manipulate a given situation or experience for two or more groups of individuals, followed by a measurement of the effect of those experiences on thoughts, feelings, or behavior).
- Establishes Causality

Approaches to Research

Descriptive	Correlational (Quasi-Experimental)	Experimental
•Current state snapshot •Not about relations •Often “Real World” (e.g., focus group) •Allows complexity	•2+ variables •Is there a relationship •Not causal -> prediction (e.g., gpa/SAT/GRE) •Can test theory	•2+ variables •Causal relations •Manipulate variables •Often “lab”-based •Predict, understand, explain, control •Can lose complexity

Scales of Measurement

- Nominal
- Ordinal
- Interval
- Ratio

Different scales of measurement allow different types of comparisons to be performed. (e.g., frequency totals vs. group means)

Representing Frequency Distributions

- Frequency Distribution Table
- Pie Charts (proportions)
- Bar Graphs (categories)
- Frequency Polygons (line graphs)
- Frequency Histograms
- Scatterplots (Individual Score Pairings)

Descriptive vs. Inferential Statistics

- Descriptive Statistics DESCRIBE (the sample)
- Inferential Statistics allow INFERENCES to be drawn from the SAMPLE about the **POPULATION**

Descriptive Statistics

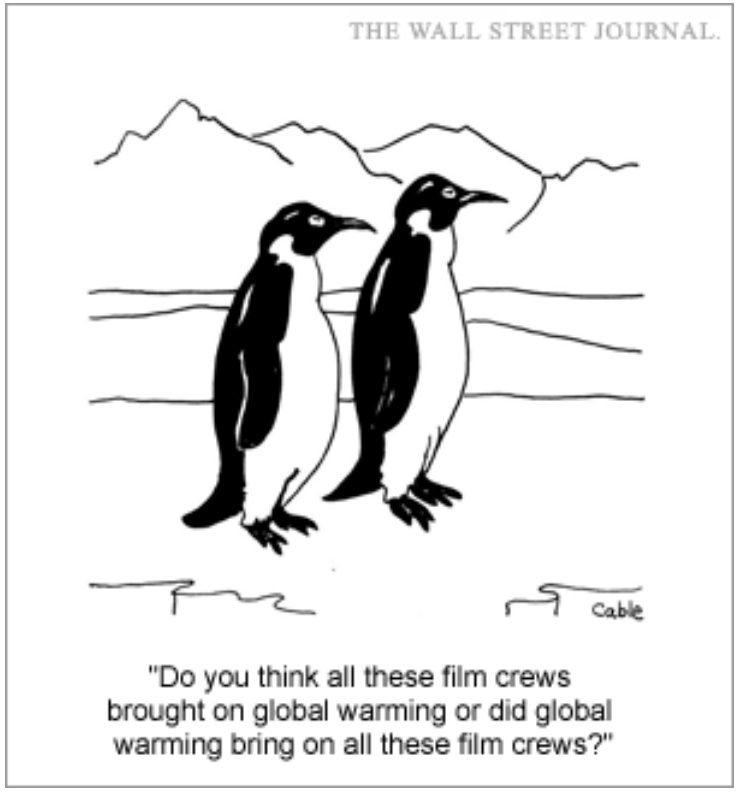
- Describe the sample
- Measures of Central Tendency;
  - Mean
  - Median
  - Mode,
- Measures of Dispersion;
  - Skew
  - Kurtosis,
  - Variability (variance, standard deviation, range)

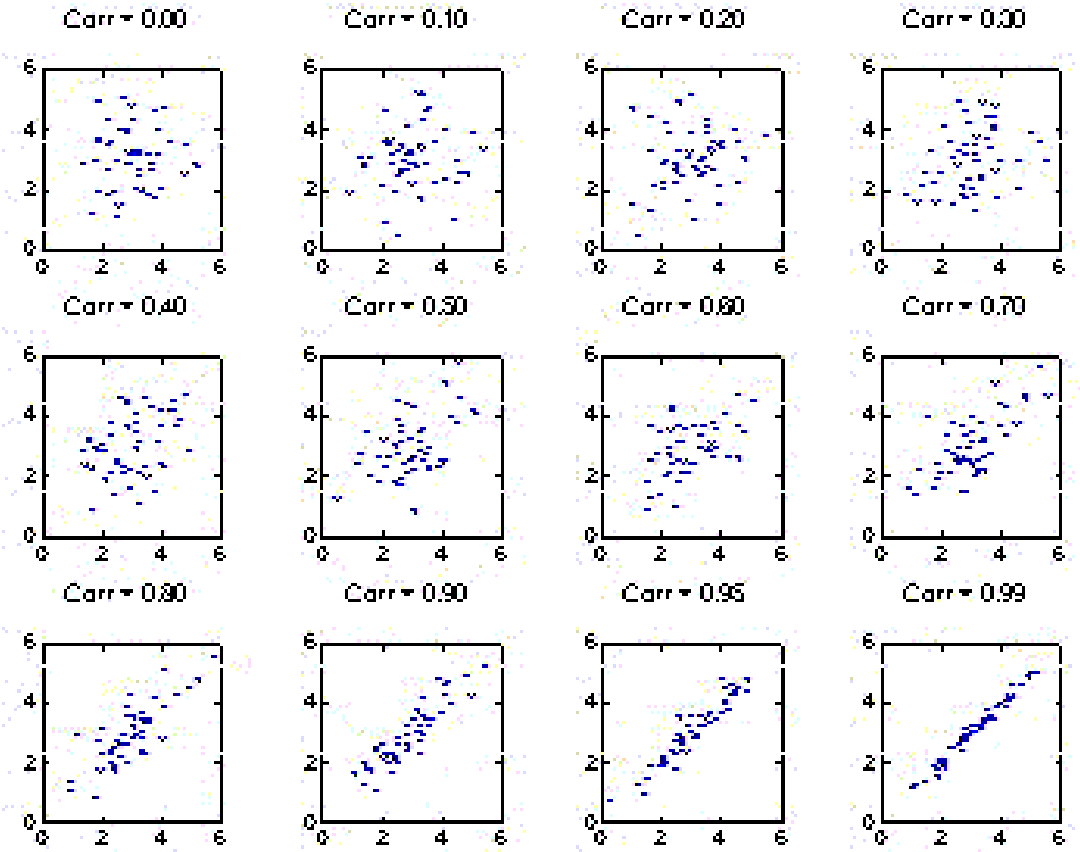
Correlation Coefficient

- Strength (0→1) and Direction (+ / -) of the relationship
- Constraints:
  - Restriction of range
    - Could reflect floor or ceiling effects or just the selection of a very restricted sample
  - Non-linear relationships

$$r = \frac{\sum_{i=1}^{n}\left(\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

Meaning, Strength (0→1), & Direction (+/-) of the Correlational Relationship





Effect Size

Effect size is a (important!) general term that refers to the strength of association between variables. In its most generic sense, the effect size indicates how much of the variability in one variable is attributable to its relation with particular other variables. Pearson's r is a measure of the effect size (recall: $-1.0 < r < 1.0$).

Small:	**/r/ <= 0 to .20**
Medium:	**/r/ <= .20 to .30**
Large:	**/r/ <= .30 to .40+**

r^2 is "variance accounted for"; that is how much (%) of the variation across Variable A is associated with variation in Variable B. The technical term is that r^2 is the "*coefficient of determination*".

Statistical Significance (Inferential Statistics)

- Would the relationship be present if we tested a different sample? Does this sample indicate a relationship in the larger population? (p-value)
- Whereas the correlation coefficient indicates the strength of the relationship, the regression equation provides the best linear fit (point estimation) for the data.
- Regression equations use the known value(s) (X) to predict the unknown value(s) (Y')
  - $Y' = a + bX$ (standard two-variable prediction equation)
  - $Y' = a + b_1X_1 + b_2X_2$ (Multiple Correlation / Multiple Regression equation)

Partial Correlation

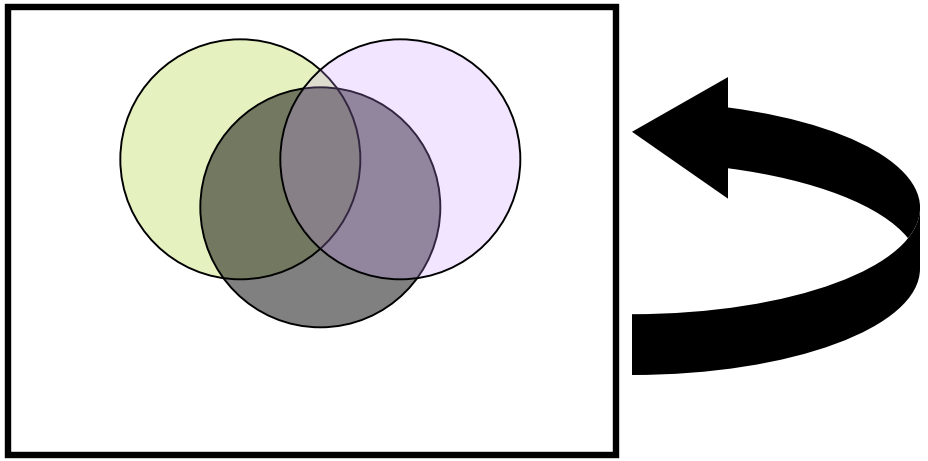
- Partial correlation allow one to “factor out” other variables by statistically “correcting” for their effects. Partial correlations approximate keeping the “undesired” variable constant (i.e., not influencing the result) without actually keeping the undesired variable constant.

The Third Variable Problem

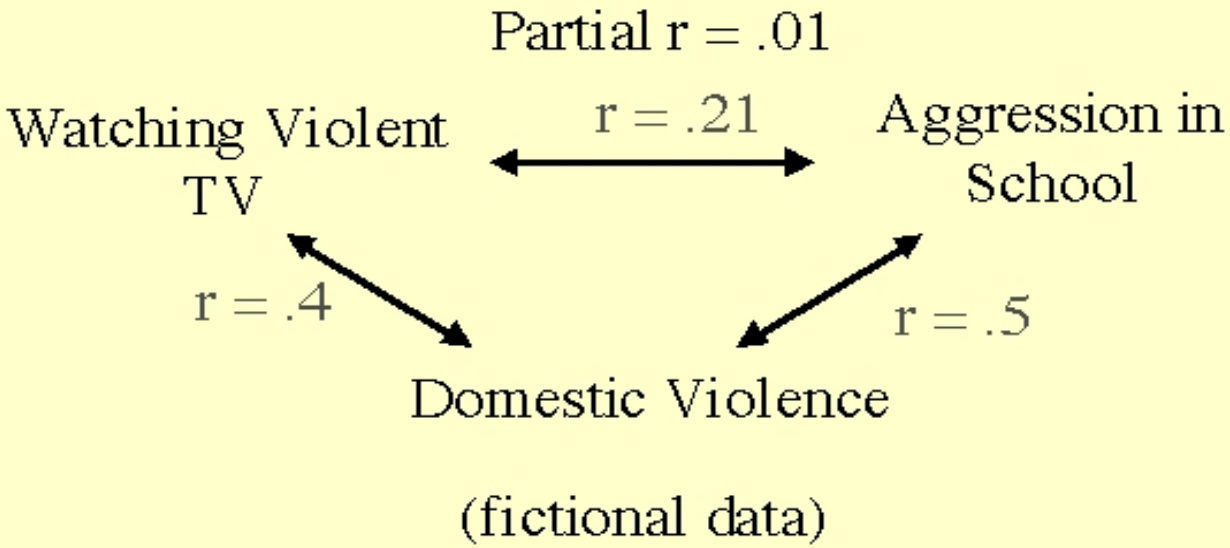
- Partial correlations (correlations having removed the effect of third variables) can indicate whether or not the third variable is “responsible” for the relationship.

(fictional) Partial Correlation Example

The overlap in variability (covariation) between the two variables (green, purple) that is not associated with the third variable (gray) yeilds the partial correlation of r=.01

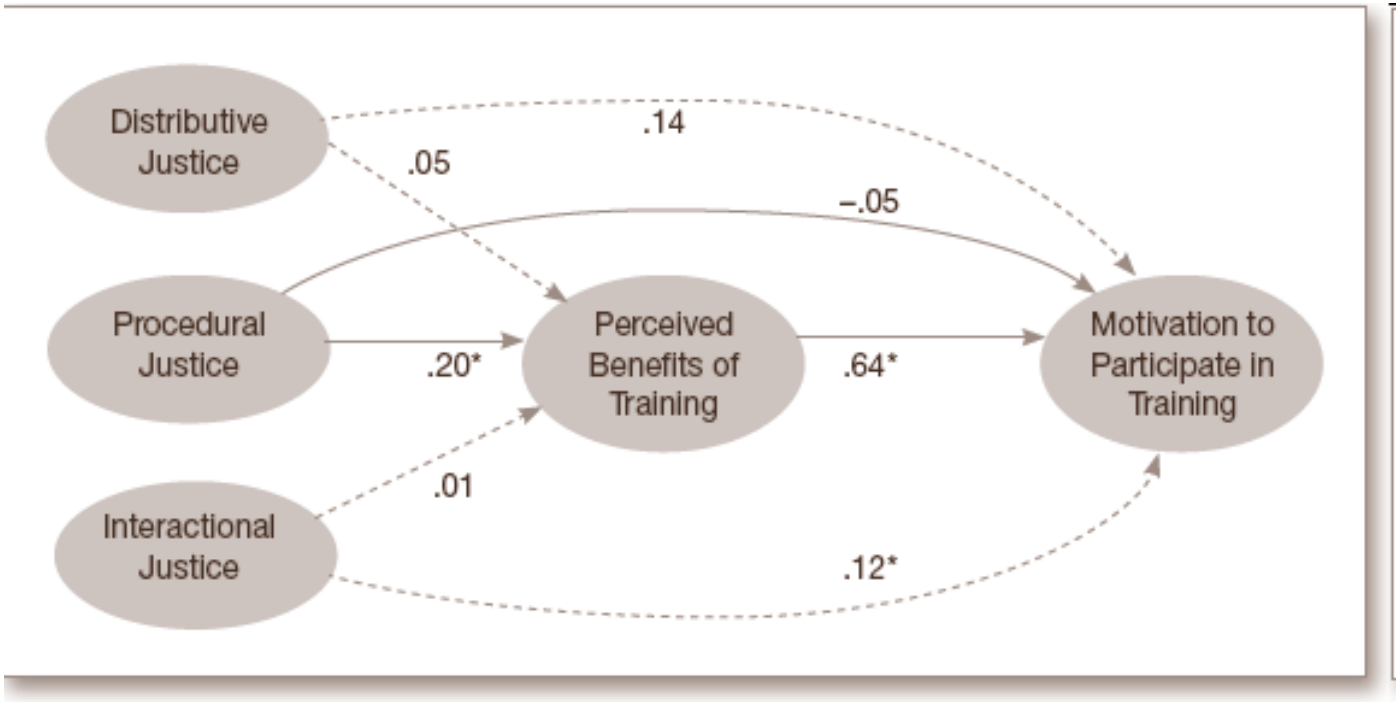


The effect of a third variable can be “partialed out” of the relationship between two other variables.



SEM (Structural Equation Modeling)

- General term for techniques that assess predicted relationships across variables
- *Path Analysis* depicts the relationships described in the models and show the "path" of the purported causal influences



Chapter 12 Terminology

- Bar graph
- Central tendency
  - Mean, median, mode
- Correlation coefficient
- Criterion variable
- Descriptive statistics
- Effect size
- Frequency distribution
- Frequency polygons
- Measurement scales
  - Nominal, ordinal, interval, ratio
- Multiple correlation
- Partial correlation
- Path analysis
- Pie chart
- Predictor variable
- Outcome measure
- Regression equation
- Restriction of range
- Scatterplot
- Standard deviation
- Statistical significance
- Structural equation modeling
- Variability
- Variance

www.calstate.edu

www.sjsu.edu/psych