SACRAMENTO STATE
Leadership begins here.

Correlational Research

Stephen E. Brock, Ph.D., NCSP
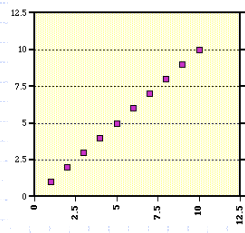California State University, Sacramento

1

Correlational Research

◆ A quantitative methodology used to determine whether, and to what degree, a relationship exists between two or more variables within a population (or a sample).
◆ The degree of relationships are expressed by correlation coefficients.
  ▪ Coefficients range from +1.00 to -1.00
  ▪ Higher correlations (coefficients closer to +1.00 or -1.00) indicate stronger relationships.
◆ Positive correlations indicate that as the values associated with one variable go up, so do the values associated with the other.
  ▪ e.g., higher grades are associated with higher ???.
◆ Negative correlations indicate that as the values associated with one variable go up, the values associated with the other go down
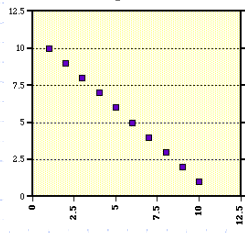 ▪ e.g., higher grades are associated with lower ???.

2

Correlational Research:
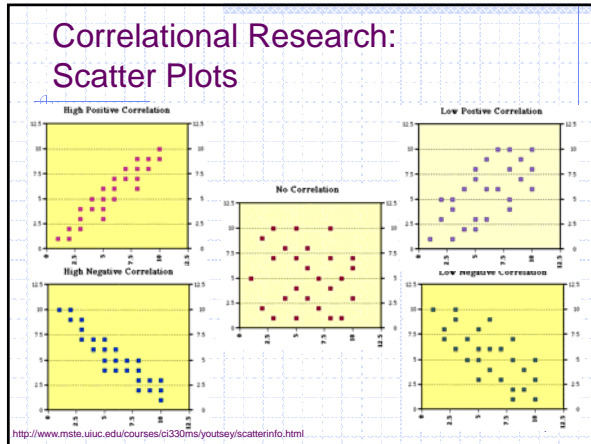# Scatter Plots

Perfect Positive Correlation   Perfect Negative Correlation

http://www.mste.uiuc.edu/courses/ci330ms/youtsey/scatterinfo.html

3

Correlational Research: Scatter Plots

High Positive Correlation

Low Positive Correlation

No Correlation

High Negative Correlation

Low Negative Correlation

http://www.mste.uiuc.edu/courses/ci330ms/youtsey/scatterinfo.html

## Portfolio Activity # 7: Mini-proposal 2

◆ Briefly describe a correlational research project relevant to one of your identified research topics.

◆ Small group discussion

5

Correlation vs. Causation

◆ A researcher found that there was a +0.85 correlation between the variable of height and Mental Age among a random sample of 100 individuals.

◆ From these data the researcher determines that taller people are smarter than shorter people.

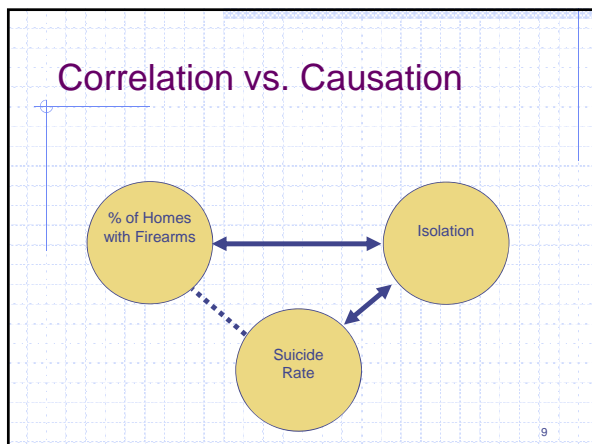◆ What do you think?

◆ Interpret this finding.

6

Correlation vs. Causation

Mental Age

Chronological Age

Height

7
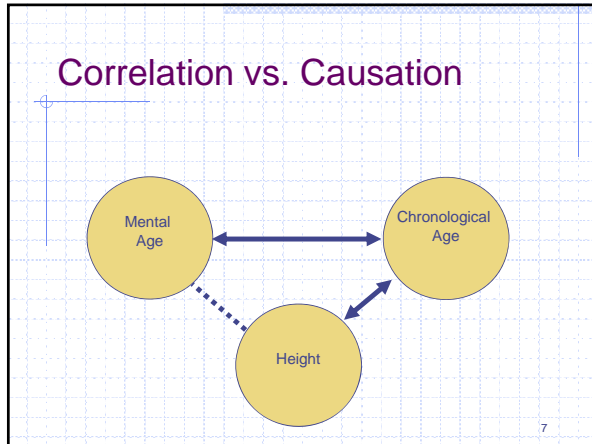
Let's conduct a Correlational Study

◆ Open an excel spread sheet
◆ Label 3 columns

State	% of Homes with Firearm	2017 Suicide Rate
Alabama	57.2	16.65
Alaska	60.6	27.11
Arizona	36.2	18.14
Arkansas	58.3	20.72
…	…	…
Wisconsin	44.2	15.45
Wyoming	62.8	26.72

◆ Enter Firearm data in one column
◆ Enter suicide data in the third column
  ▪ https://www.cdc.gov/injury/wisqars/fatal.html

8

Correlation vs. Causation

% of Homes with Firearms

Isolation

Suicide Rate

9

Correlation vs. Causation

- ◆ Even a perfect correlation does not necessarily imply a causal connection between variables.
  - ▪ For example, in a recent CDE study, the number of support staff in school districts was positively correlated with poor attendance.
- ◆ An educational research example:
  - ▪ Attention span is highly correlated with reading comprehension test scores.
  - ▪ But both are also correlated with basic reading skill.
  - ▪ The correlation may be the result of a mutual association with these other variables.

10

Correlation vs. Causation
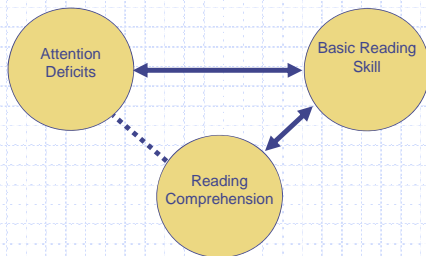
11

Correlation vs. Causation

- ◆ A statistically significant relationship (correlation) is a necessary, but not sufficient condition when determining causation.
- ◆ Must be able to document that the causal variable occurred first and that all other factors are accounted for.
- ◆ Experiments are typically necessary to determine causation.

12

Types of Correlational Studies

◆**Descriptive**
- Used to simply describe relationships.
- Often a precursor to the experimental study.
  - Variables suggested to be related would be the subject of further study.
  - Also helps to identify variables that need to be controlled during an experiment.
    - e.g., basic reading skill in a study of the effects of ADHD on reading comprehension.
- Hypotheses, if offered, are often non-directional.

◆**Predictive**
- Hypotheses are directional

13

The Correlational Research Process

1. Problem selection
   - Variables to be correlated should be selected based on
     a) A logical relationship
     b) Theoretical grounds
     c) Personal experience
     - *What are some examples of problems (or questions) that are consistent with these three bases for correlational research?*
   - Correlational "treasure hunts" (AKA "the shotgun approach") are strongly discouraged.
     - What does $r = .50$, $\boldsymbol{p = .05}$ mean?
     - r = strength of the relationship (actually it is r^2 or 25% of variance)
     - p = significance of the relationship (how unlikely a given r value will occur given NO relationship in the population, 5% chance of an r of .50 if there is no real relationship between variable in the population) 14

The Correlational Research Process

2. Select/Obtain Participants
   - Sample from the population so as to maximize generalizability.
     - What are examples of preferred sampling techniques?
   - At least 30 participants.
     - *If you are dealing with a **sample** and the correlation between two variables is* r = .05, p = .50, *what would you say about the relationship?*
     - *If you are dealing with a **population** and the correlation between two variables is* r = .05, *what would you say about the relationship?*

15

The Correlational Research Process

3. Select Measures
   - How to quantify the variables under study.
     - How might your quantify ADHD, Reading Achievement, phonological processing?
   - If the measures lack reliability a larger sample will be required.
4. Specify Procedures
   - How is data assessing the variables are obtained and correlated?

16

The Correlational Research Process

5. Conduct Data Analysis
   - Statistically significant correlations
     - p values (p. 582 of the text).
       - Whether the obtained coefficient is really different from zero – or- the probability that the correlation represents a true relationship or a chance occurrence.
       - In larger samples, lower correlations are required to reach statistical significance. **Why?**
       - Tests of significance are not required if the entire population has been assessed. **Why?**
       - What do the levels of significance (e.g., $p = .1$, $p = .05$, $p = .01$, $p = .001$) mean?
       - **Is a significant relationship necessarily an important relationship?**
         - Compare these two results:
           (1) ADHD correlates with Reading ($r = .75$, $p = .05$);
           (2) ADHD correlates with Reading ($r = .25$, $p = .001$)

17

The Correlational Research Process

5. Conduct Data Analysis (continued)
   - Determining statistically significant correlations
     - http://www.danielsoper.com/statcalc3/calc.aspx?id=44
   - If you are using a Table, the number of degrees of freedom is **two less** than the number of pairs.

18

The Correlational Research Process

5. Conduct Data Analysis (continued)
    - Correlation's significance vs. its strength.
        - Just because a correlation is significant does not mean it is high enough to reflect an important relationship.
    - Variance (the correlation coefficient squared)
        - "When two or more variables are correlated, each variable will have a range of scores. Each variable will have some variance; that is not everyone will get the same score. Common or shared variance indicates the extent to which variables vary in a systematic way" (pp. 314-315).
            - r^2 is the amount of variance explained (or accounted for) by the correlation coefficient.
            - Determine the amount of variance accounted for by the following r values: 1.0, .95, .75, .50, .25
                - http://www.calculator.org/jcalc98.html

19

Relationship Studies

- Often used to study complex variables before beginning an experiment.
    - To identify variables (other than the independent variable) that correlate with the dependent measure. When relationships are identified these variables are then controlled for.
        - For example, before studying how a given IV (like ADHD symptom severity) influences reading comprehension you would want to identify other variables (such as word reading, word attack, vocabulary, background knowledge) that also affect reading comprehension and then control for them.
        - How would this be done?

20

Relationship Studies

- Why is it important to be selective when identifying variables to be correlated?
    - What problems might arise if you used a "shotgun approach" and obtained correlations among 100 randomly selected variables and you used a p value of .05?
    - Chances are that 5 of the obtained coefficients will not reflect a true relationship greater than zero.

21

Prediction Studies

◆ Regression Analysis
- A method of analyzing the variability of a criterion variable by examining information available on one or more predictor variables.
- When only one predictor variable is used, the analysis is referred to as simple regression. When more than one predictor is used, the analysis is referred to as multiple regression

22

Simple Regression

◆ A college football coach wishes to use the scores on one variable to predict the scores on another variable.

◆ He wishes to determine the best prediction equation for the grade-point averages of potential freshmen recruits.

◆ SAT test scores for the current group of recruits as well as their grade point averages are available.

◆ From the available SAT and GPA scores for this year's class, the prediction equation for next year's class can be calculated.

23

Simple Regression

◆ A teacher wishes to determine the effects of hours of study (the predictor variable) on vocabulary test performance (the criterion variable).

◆ When vocabulary test means associated with different amounts of study differ from each other and lie on a straight line, it is said that there is a simple linear regression of vocabulary test performance on hours of study

24

Hours	Score A	Score B	
1	3	1	**Hours of study & vocabulary tests**
1	5	2	
1	6	3	
1	9	0	
2	4	3	
2	6	4	
2	7	3	
2	10	3	**Excel Data Sheet**
3	4	4	
3	6	5	
3	8	6	**pp. 609-610 of text for SPSS screen shots**
3	10	4	
4	5	5	
4	7	6	
4	9	7	
4	12	5	
5	6	7	
5	7	8	
5	10	9	25
5	12	6	

Multiple Regression

◆ A GATE program administrator wishes to determine the best prediction equation for rapid learning among a group of ELL elementary students. The available predictor variables are:
  1. SAT-9 scores (X_1)
  2. Changes in scores on the English LAS (X_2)
  3. Scores on a non-verbal reasoning test (X_3)
  4. Primary language vocabulary test scores (X_4)
◆ Y The criterion variable (the one the administrator wishes to predict) are achievement test gains made between 2nd to 3rd grades ($Y_{observed}$).
◆ From the regression of the predictor variables ($X_{1, 2, 3,}$ and $_4$) on the criterion variable ($Y_{observed}$), a prediction equation can be developed.

26

Types of Correlation Coefficients

		Variable 1		
		Continuous	Rank	Dichotomous
Variable 2	Continuous	Pearson r Correlation ratio		
	Rank		Spearmans rho Kendall's tau	
	Dichotomous	Biserial Point Biserial		Tetrachoric Phi coefficient

Scales of Measurement: Ratio or Interval (Continuous, e.g., scores on a test); Ordinal (Rank, e.g., class rankings); and Nominal or Categorical (Dichotomous, e.g., gender)

Notes:
Correlation ratio is used for nonlinear relationships (i.e., curvilinear correlations)
Biserial and *Tetrachoric* are used when the dichotomy is artificial
Point Biserial and *Phi coefficient* are used when the dichotomy is genuine or true.

27

Types of Statistical Correlations

◆ Partial Correlations
  ▪ If you have three variables and you wish to know how highly two of them are related when the mutual relationships with the third variable are taken out ("partialed out"), use partial correlation.

Height ←——— $r = .835$ ———→ M.A.
                (A spurious correlation?)

                    Age

After partialing out Age this correlation dropped to .219. This is an example of statistical control. What might be another variable to partial out? What might be another way to control for the Age?

28

Types of Statistical Correlations

◆ Multiple Correlations
  ▪ If you have three variables and you wish to know how highly two of them, taken together, are related to the third, use multiple correlation.

Reading Achievement ←———————→ Reading Speed + Phonological awareness

In a typical multiple-correlation study, the first set of numbers represents measures of a criterion variable (e.g., reading achievement) and the other two sets of numbers are measures of predictors (e.g., reading speed and sound awareness). The multiple-correlation coefficient between the criterion variable and the two predictor variables will give an indication of the degree to which the two predictors, taken together, actually predict the criterion

Questions: What is the difference between prediction and causation?

29

Next Meeting:

◆ Causal Comparative and Experimental Research

◆ Read *Educational Research* Chapters 10 & 11.

◆ Portfolio Element #8 Due: Mini-proposal 3

Week after next, April 2

◆ Portfolio Element #9 Due: Mini-proposal 4

30

Portfolio Activity # 8
Mini-proposal 3

◆ Students will briefly describe a causal comparative research project relevant to one of their identified research topics.
  - Chapter 10 provide guidance necessary to complete this mini-proposal.

31