Correlational Research

Chapter Fifteen

Bring folder of readings

Correlational Research

Chapter Fifteen

TABLE 15.2 *Teacher Expectation of Failure and Amount of Disruptive Behavior for a Sample of 12 Classes*

Class	Teacher Expectation of Failure (Ratings)	Amount of Disruptive Behavior (Ratings)
1	10	11
2	4	3
3	2	2
4	4	6
5	12	10
6	9	6
7	8	9
8	8	6
9	6	8
10	5	5
11	5	9
12	7	4

The Nature of Correlational Research

- Correlational Research is also known as Associational Research.
- Relationships among two or more variables are studied without any attempt to influence them.
- Investigates the possibility of relationships between two variables.
- There is no manipulation of variables in Correlational Research.

Correlational studies describe the variable relationship via a correlation coefficient

Three Sets of Data Showing Different Directions and Degrees of Correlation

(Table 15.1)

(A) $r = +1.00$		(B) $r = -1.00$		(C) $r = 0$	
X	Y	X	Y	X	Y
5	5	5	1	2	1
4	4	4	2	5	4
3	3	3	3	3	3
2	2	2	4	1	5
1	1	1	5	4	2

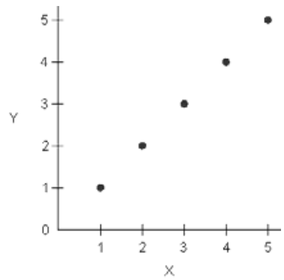
Purpose of Correlational Research

- Correlational studies are carried out to explain important human behavior or to predict likely outcomes (identify relationships among variables).
- If a relationship of sufficient magnitude exists between two variables, it becomes possible to predict a score on either variable if a score on the other variable is known (Prediction Studies).
- The variable that is used to make the prediction is called the <u>predictor variable.</u>

Purpose of Correlational Research (cont.)

- The variable about which the prediction is made is called the <u>criterion variable</u>.
- Both <u>scatterplots</u> and <u>regression lines</u> are used in correlational studies to predict a score on a criterion variable
- A predicted score is never exact. Through a prediction equation (see p. 585), researchers use a predicted score and an index of prediction error (standard error of estimate) to conclude if the score is likely to be incorrect.

Scatterplot Illustrating a Correlation of +1.00 (Figure 15.1)

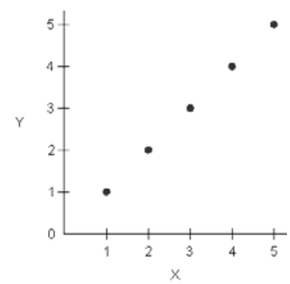


Prediction Using a Scatterplot
(Figure 15.2)

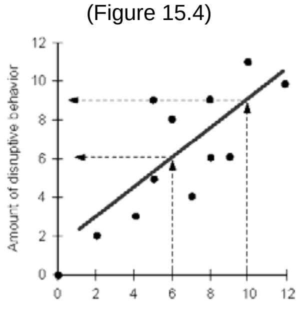
More Complex Correlational Techniques

- Multiple Regression
  - Technique that enables researchers to determine a correlation between a criterion variable and the best combination of two or more predictor variables
- Coefficient of multiple correlation (R)
  - Indicates the strength of the correlation between the combination of the predictor variables and the criterion variable
- Coefficient of Determination
  - Indicates the percentage of the variability among the criterion scores that can be attributed to differences in the scores on the predictor variable
- Discriminant Function Analysis
  - Rather than using multiple regression, this technique is used when the criterion value is categorical
- Factor Analysis
  - Allows the researcher to determine whether many variables can be described by a few factors
- Path Analysis
  - Used to test the likelihood of a causal connection among three or more variables
- Structural Modeling
  - Sophisticated method for exploring and possibly confirming causation among several variables

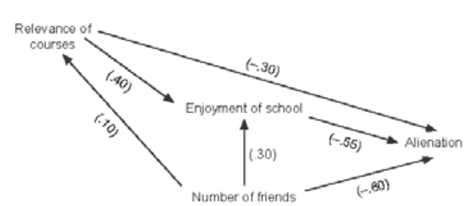
Scatterplot Illustrating a Correlation of +1.00 (Figure 15.3)



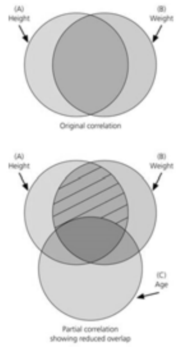
Prediction Using a Scatterplot
(Figure 15.4)



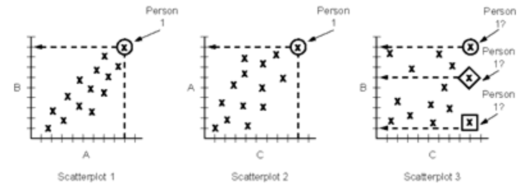
Path Analysis Diagram
(Figure 15.5)



Partial Correlation (Figure 15.6)

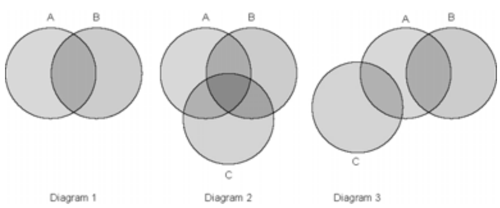


Scatterplots Illustrating How a Factor (C) May Not be a Threat to Internal Validity (Figure 15.7)



Circle Diagrams Illustrating Relationships Among Variables

(Figure 15.8)



Basic Steps in Correlational Research

- Problem selection
- Choosing a sample
- Selecting or choosing proper instruments
- Determining design and procedures
- Collecting and analyzing data
- Interpreting results

What Do Correlational Coefficients Tell Us?

- The meaning of a given correlation coefficient depends on how it is applied.
- Correlation coefficients below .35 show only a slight relationship between variables.
- Correlations between .40 and .60 may have theoretical and/or practical value depending on the context.
- Only when a correlation of .65 or higher is obtained, can one reasonably assume an accurate prediction.
- Correlations over .85 indicate a very strong relationship between the variables correlated.

Threats to Internal Validity in Correlational Research

- Subject characteristics
- Mortality
- Location
- Instrument decay
- Testing
- History
- Data collector characteristics
- Data collector bias

The following must be controlled to reduce threats to internal validity

Statistical Inference

The larger the sample size (n) the more confident you can be that your sample mean is a good representation of the population mean. In other words, the "n" justifies the means.

~ Ancient Kung Foole Proverb

One- and Two Tailed Probabilities

- One-tailed
  - The probability that an observation will occur at one end of the sampling distribution.
- Two-tailed
  - The probability that an observation will occur at either extreme of the sampling distribution.

Hypothesis Testing

- Conceptual (Research) Hypothesis
  - A general statement about the relationship between the independent and dependent variables
- Statistical Hypothesis
  - A statement that can be shown to be supported or not supported by the data.

Statistical Significance Testing

- Indirect Proof of a Hypothesis
- Null Hypothesis
  - A statement that specifies no relationship or difference on a population parameter.
- Alternative Hypothesis
  - A statement that specifies some value other than the null hypothesis is true.

Examples of the Null and Alternative Hypotheses

Nondirectional Test	Directional Test
Ho: $\mu = 5$ Ha: $\mu \neq 5$	Ho: $\mu \geq 5$ or $\mu \leq 5$ Ha: $\mu < 5$ or $\mu > 5$

Rejecting the Null

- Alpha Level
  - The level of significant set by the experimenter. It is the confidence with which the researcher can decide to reject the null hypothesis.
- Significance Level
  - The probability value used to conclude that the null hypothesis is an incorrect statement. Common significance levels are .05, .01 and .001.

Two Types of Error

- Type I
  - When a researcher rejects the null hypothesis when in fact it is true. The probability of a type I error is α.
- Type II
  - An error that occurs when a researcher fails to reject a null hypothesis that should be rejected. The probability of a Type II error is β.

Type 1 Error & Type 2 Error

Scientist's Decision

	Reject null hypothesis	Fail to reject null hypothesis
Null hypothesis is true	Type 1 Error probability = α	Correct Decision Probability = 1- α
Null hypothesis is false	Correct decision probability = 1 - β	Type 2 Error probability = β

Type 1 Error = α

Cases in which you reject null hypothesis when it is really true

Type 2 Error = β

Cases in which you fail to reject null hypothesis when it is false

The OJ Trial

For a nice tutorial go to:
http://www.socialresearchmethods.net/OJtrial/ojhome.htm

Statistical Significance Testing

The Problems with SST

- We misunderstand what it *does* tell us.
- It does *not* tell us what we want to know.
- We often overemphasize SST.

Four Important Questions

1. Is there a real relationship in the population?
   Statistical Significance
2. How large is the relationship?
   Effect Size or Magnitude
3. Is it a relationship that has important, powerful, useful, meaningful implications?
   Practical Significance
4. Why is the relationship there?
   ??????

SST is all about . . .

- **Sampling Error**
  - The difference between what I see in my sample and what exists in the target population.
  - *Simply because I sampled, I could be wrong.*
  - This is a threat to Internal Validity

How it works:

1. Assume sampling error occurred; there is no relationship in the population.
2. Build a statistical scenario based on this *null hypothesis*
3. How likely is it I got the sample value I got when the null hypothesis is true? (This is the fabled *p*-value.)

How it works (cont'd):

- How unlikely does my result have to be to rule out sampling error? alpha (α).
- If $p < \alpha$, then our result is statistically rare, is unlikely to occur *when there isn't a relationship in the population.*

What it *does* tell us

- What is the probability that we would see a relationship in our sample when there is no relationship in the population?
- Can we rule out sampling error as a competing hypothesis for our finding?

What it *does not* tell us

- Whether the null hypothesis is true.
- Whether our results will replicate.
- Whether our research hypothesis is true.
- How big the effect or relationship is.
- How important the results are.
- Why there is a relationship.