CHAPTER 12

UNDERSTANDING RESEARCH RESULTS: DESCRIPTION AND CORRELATION

LEARNING OBJECTIVES

- ✓ Contrast three ways of describing results:
  - ✓ Comparing group percentages
  - ✓ Correlating scores
  - ✓ Comparing group means
- ✓ Describe a frequency distribution, including the various ways to display a frequency distribution
- ✓ Describe the measures of central tendency and variability
- ✓ Define a correlation coefficient
- ✓ Define effect size
- ✓ Describe the use of a regression equation and a multiple correlation to predict behavior
- ✓ Discuss how a partial correlation addresses the third-variable problem
- ✓ Summarize the purpose of structural equation models

SCALES OF MEASUREMENT: A REVIEW

- Whenever a variable is studied, the researcher must create an operational definition of the variable and determine what type of scale will be used to analyze the variable.
  - Do you want to measure your variables using a yes/no scale
  - will it be an overall score acquired through scale rating questions on a survey measuring a psychological construct
  - or will it be a physiological measurement?

SCALES OF MEASUREMENT: A REVIEW

- The scales of the variable can be described using one of four scales of measurement:
  - nominal,
  - ordinal,
  - interval,
  - ratio.
- The scale used determines the types of statistics to be used.

SCALES OF MEASUREMENT: A REVIEW

- **Nominal**
  - No numerical, quantitative properties
  - **Levels represent different categories or groups**
  - Most independent variables in experiments are nominal.
- **Ordinal**
  - exhibit minimal quantitative distinctions
  - **Rank the levels from lowest to highest**
- **Interval**
  - Intervals between levels are equal in size
  - **Can be summarized using mean or arithmetic average**
  - No absolute zero
- **Ratio**
  - The variables have both equal intervals and an absolute zero point that indicates the absence of the variable being measured.
  - **Can be summarized using mean or arithmetic average**
  - **Time, weight, length, and other physical measures** are the best examples of ratio scales.
- Interval & Ration Variables are treated the same in SPSS and are called **Scale** variables

DESCRIBING RESULTS: Reporting for Non-Significance Testing Descriptives & Frequencies

- Depending on the way that the variables are studied, there are three basic ways of describing the results:
  - (1) comparing group percentages,
  - (2) correlating scores of individuals on two variables,
  - (3) comparing group means.

DESCRIBING RESULTS: Reporting for Significance Testing

- **Comparing group percentage**
  - Here the focus is on percentages because the variable is **nominal**:
    - for example, liking and disliking are simply two different categories.
  - After describing the data, the next step would be to perform a statistical analysis to determine whether there is a statistically significant difference between two groups
    - **Chi-square**: can be tested in Crosstabs in SPSS

DESCRIBING RESULTS: Reporting for Significance Testing

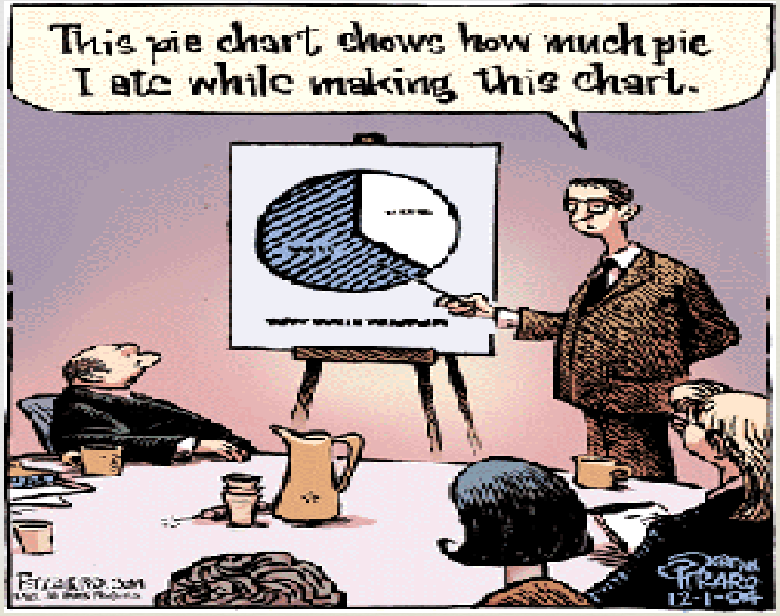
- **Correlating individual scores**
  - A second type of analysis is needed when one does not have distinct groups of subjects.
  - Instead, individuals are measured on two variables, and each variable has a **range (interval/ratio)** of numerical values.
    - Means are used for continuous variables (interval/ratio)
    - **Correlations** (Pearson's-*r*) & **Regression**

DESCRIBING RESULTS: Reporting for Significance Testing

- **Comparing group means**
  - Much research is designed to compare the mean responses of participants in two or more groups.
  - Therefore, the Independent variable is the study's groups (nominal/ordinal) which are measured for differences in the Dependent variable **(interval/ratio)**
    - T-Tests & ANOVA (F-Tests)

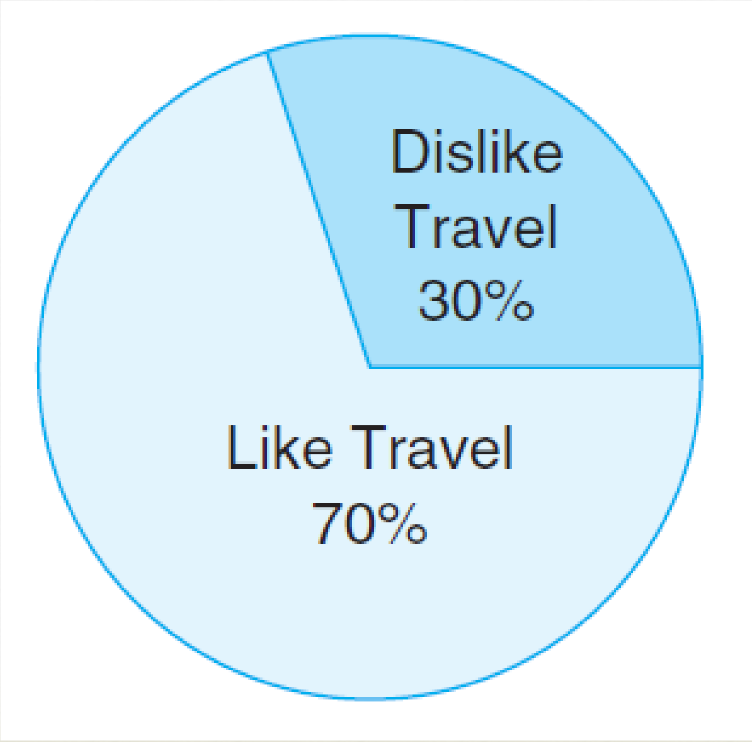
FREQUENCY DISTRIBUTIONS

- A **frequency distribution** indicates the frequency or count of the number of occurrences of values for a particular variable.
- **Graphing frequency distributions can be displayed using charts or graphs such as:**
  - **Pie charts**
  - **Bar graphs**
  - **Histograms**
  - **Frequency polygons**



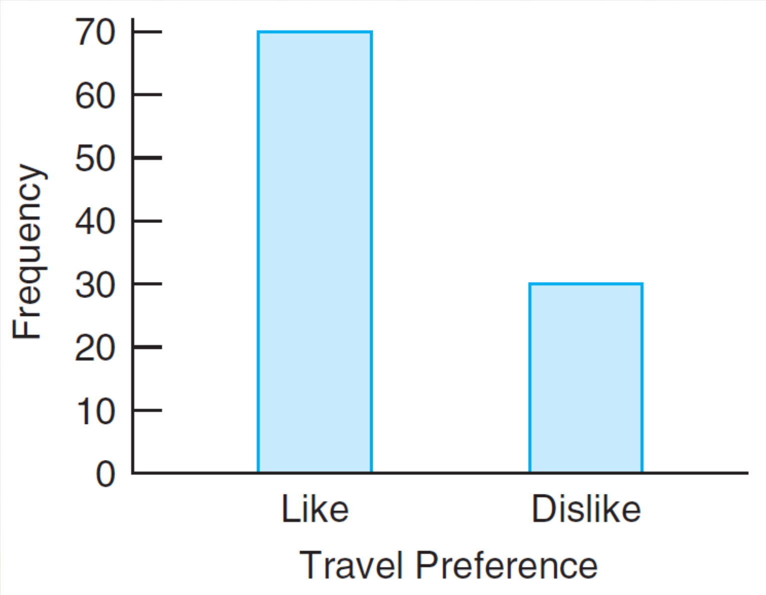
PIE CHART

✓**Pie charts** divide a whole circle, or “pie,” into “slices” that represent relative **percentages**.



BAR GRAPH DISPLAYING DATA OBTAINED IN TWO GROUPS

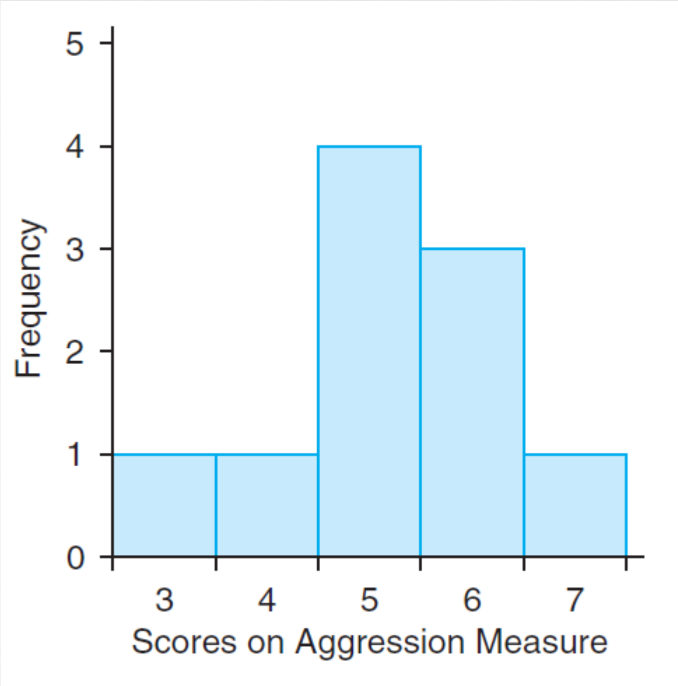
✓**Bar graphs** use a **separate and distinct** bar for each piece of information. They are used when the values on the x axis (independent variable) are nominal (categorical) variables



HISTOGRAM SHOWING FREQUENCY OF RESPONSES

✓**Histograms** use bars to display a frequency distribution for a quantitative variable.

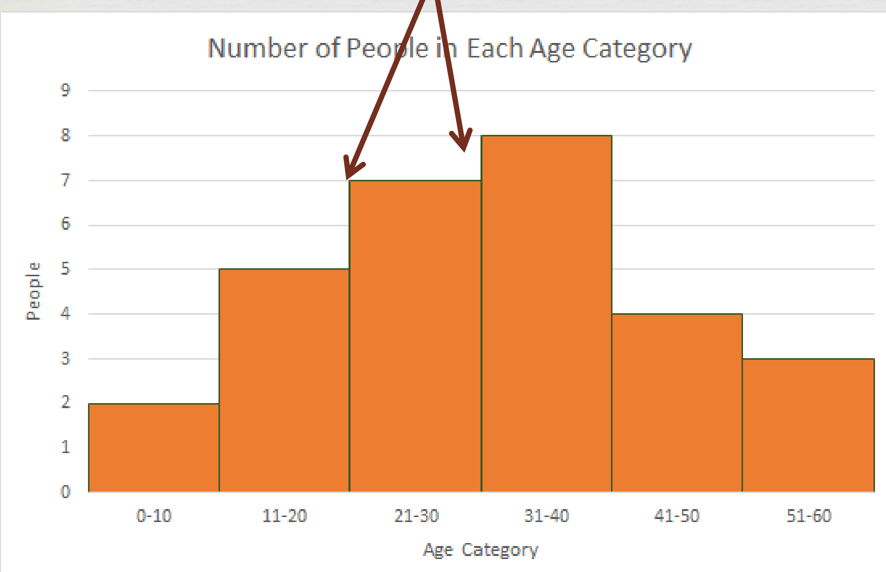
✓In this case, the scale values are **continuous**.



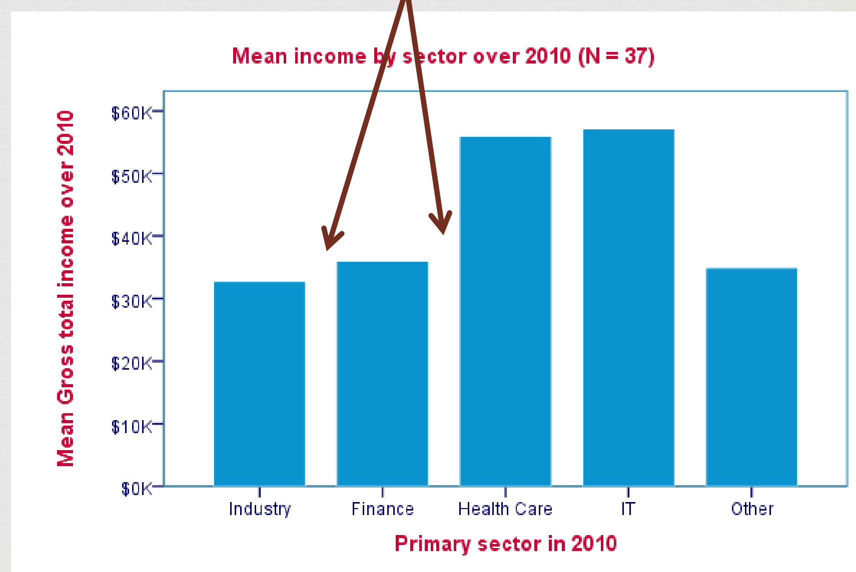
Histogram vs. Bar Chart

Histograms have no gaps to indicate continuous or ordinal data. Bar Graphs have gaps do indicate separate categories for nominal or categorical data.

Histogram: No Gaps



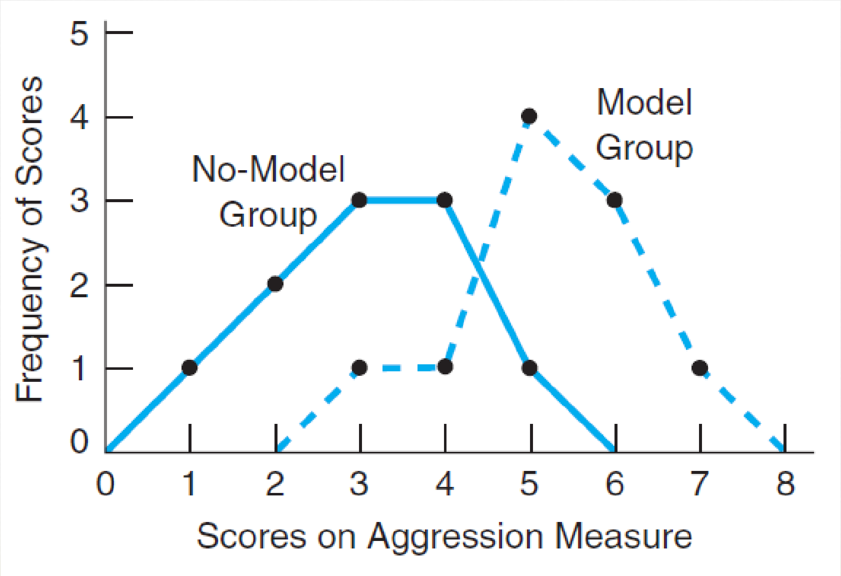
Bar Chart: Gaps



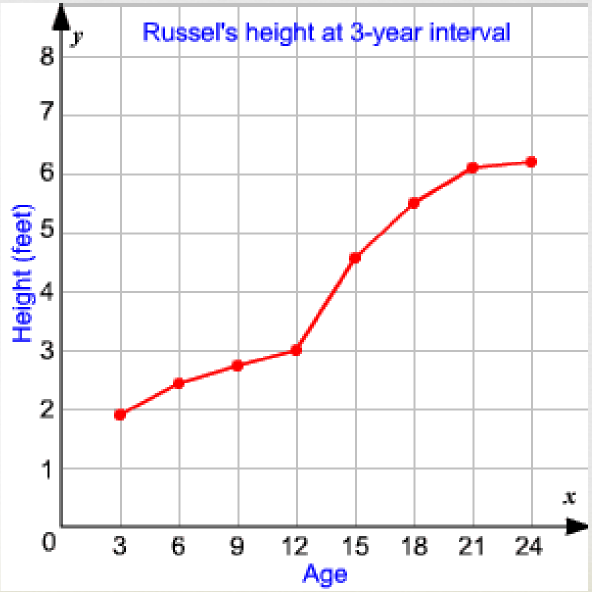
FREQUENCY POLYGONS VS LINE GRAPHS

✓**Frequency polygons** use a line(s) to represent the distribution of frequencies of scores.
✓**Line graphs** - Used when the values on the x axis (independent variable) are ratio/interval (continuous) numbers

Frequency Polygon



Line Graph



SCORES ON AGGRESSION MEASURE IN MODELING AND AGGRESSION

When reporting results, the individual score counts along with Standard Deviations, Sum of Squares, Means, and total numbers are often given.

Model group	No-model group
3	1
4	2
5	2
5	3
5	3
5	3
6	4
6	4
6	4
7	5
$\Sigma X = 52$	$\Sigma X = 31$
$\overline{X} = 5.20$	$\overline{X} = 3.10$
$s^2 = 1.29$	$s^2 = 1.43$
$s = 1.14$	$s = 1.20$
$n = 10$	$n = 10$

DESCRIPTIVE STATISTICS

- A **Central tendency** statistic tells us what the sample as a whole, or on the average, is like. There are three measures of central tendency—the mean, the median, and the mode.
  - **Mean** (*M*)
    - Obtained by adding all the scores and dividing by the number of scores
    - Indicates central tendency when scores are measured on an **interval** or **ratio** scales (continuous variables)
  - **Median** (In scientific reports, the median is abbreviated as *Mdn*)
    - Score that divides the group in half
    - Indicates central tendency when scores are measured on an **ordinal**, **interval**, and **ratio** scales (categorical and continuous variables)
  - **Mode**
    - Most frequent score
    - The only measure of central tendency that is appropriate if a **nominal** scale is used (categorical variables)
    - The mode does not use the actual values on the scale, but simply indicates the most frequently occurring category.

DESCRIPTIVE STATISTICS

- **Variability** is the amount of spread in the distribution of scores.
  - **Variance** (s^2)
    - The standard deviation is derived by first calculating the **variance**, symbolized as s^2
    - The standard deviation is the square root of the variance

Sample Variance

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

DESCRIPTIVE STATISTICS

- **Standard deviation** (*s*)
  - Abbreviated as SD in scientific reports
  - Indicates the *average deviation* of scores from the mean.

Sample Standard Deviation

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

- **Range**
  - Another measure of variability is the range, which is simply the difference between the highest score and the lowest score.

CORRELATION COEFFICIENTS: STRENGTH OF RELATIONSHIPS

- **Correlation coefficient**: Describes how strongly variables are related to one another.
  - The Pearson product-moment correlation coefficient is commonly used for this measure.

"Ignore the lifeboats in the parking lot.
Everything is fine."

CORRELATION COEFFICIENTS: STRENGTH OF RELATIONSHIPS

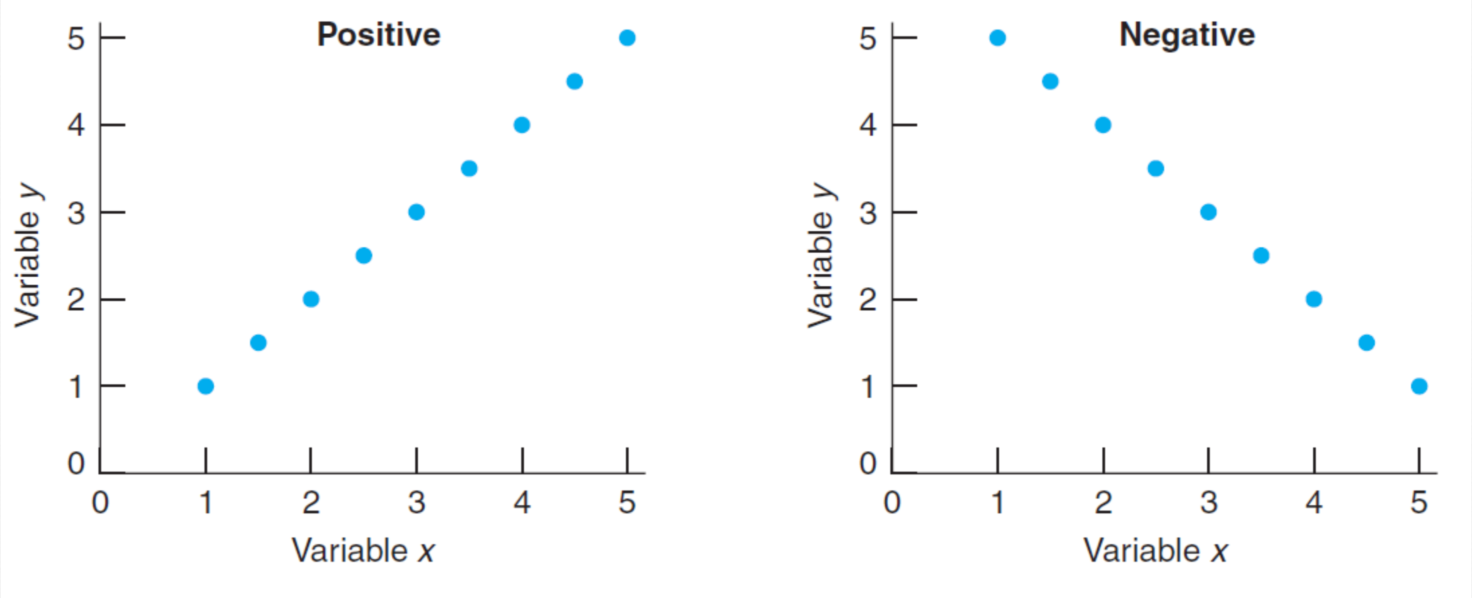
- **Pearson product-moment correlation coefficient** (Pearson *r*): Used when both variables have interval or ratio scale properties
  - Values of a Pearson *r* can range from 0.00 to ±1.00
  - Provides information about the strength and the direction of relationship

CORRELATION COEFFICIENTS: STRENGTH OF RELATIONSHIPS

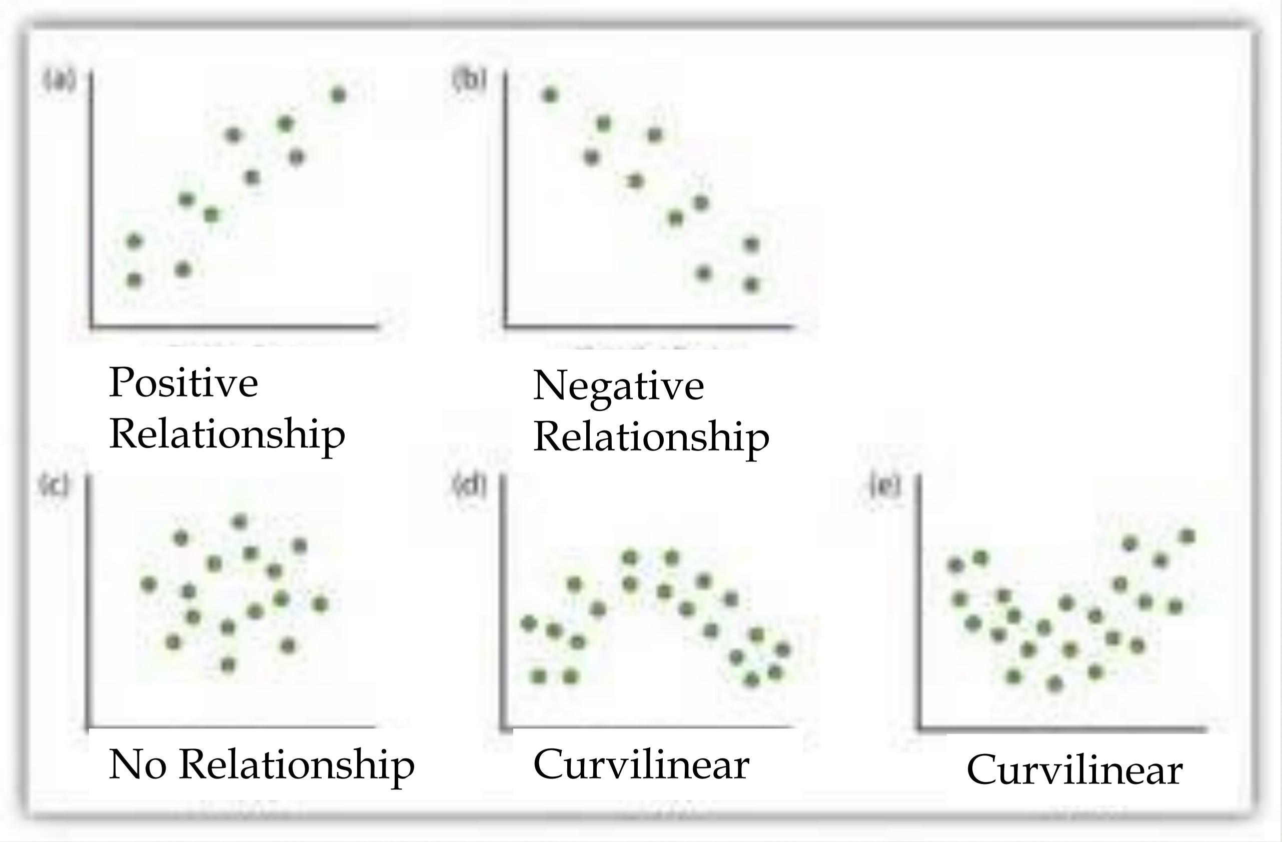
- **Pearson *r*** (continued):
  - A correlation of 0.00 indicates that there is no relationship between the variables.
  - The nearer a correlation is to 1.00 (plus or minus), the stronger the relationship.
  - The relationship between variables can be described visually using scatterplots.
    - Scatterplots can often show at a glance whether there is a relationship between variables.

SCATTERPLOTS OF PERFECT (±1.00) RELATIONSHIPS

✓ An upward trending line indicates a positive relationship, and a downward trending line indicates a negative relationship.

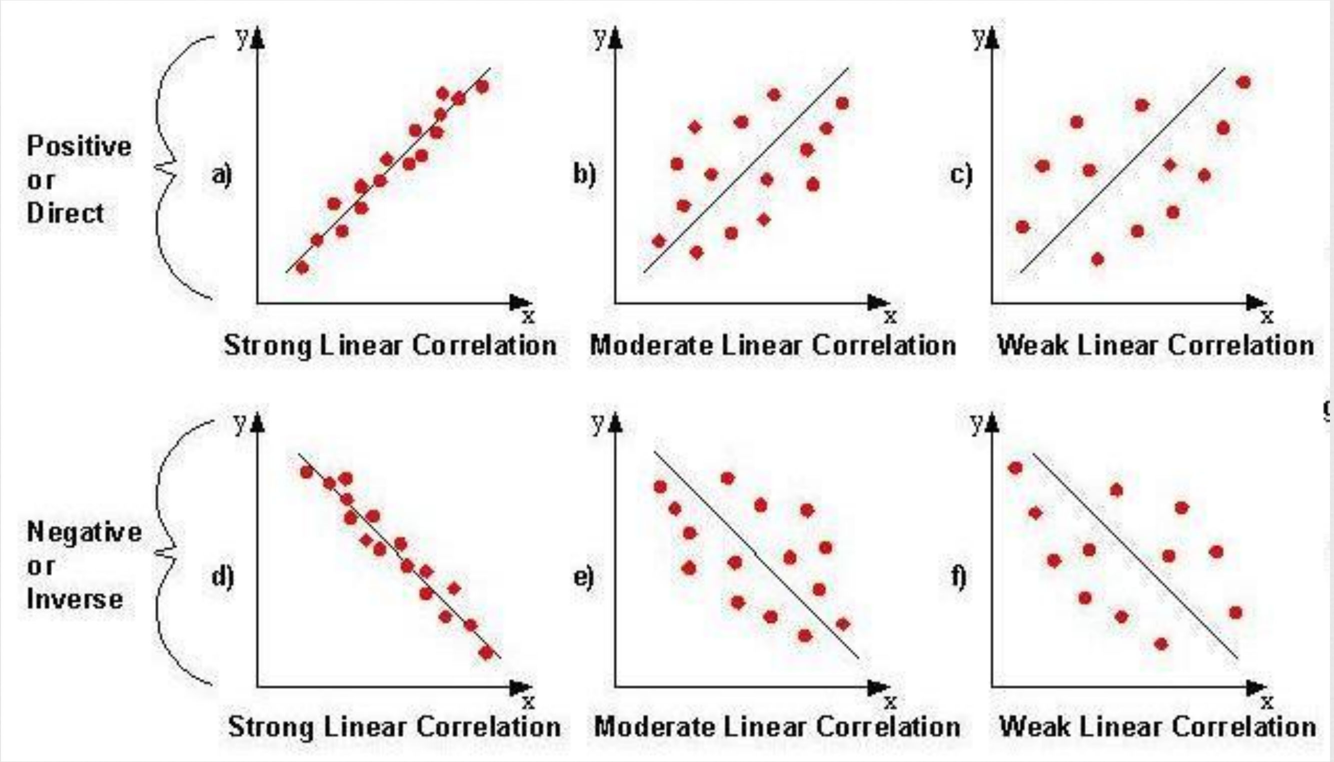


SCATTERPLOT PATTERNS OF CORRELATION



Strong vs Weak Linear Relationships

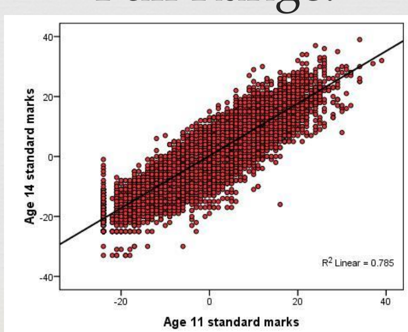
Strong vs Week Relationships & the Line of Best Fit:



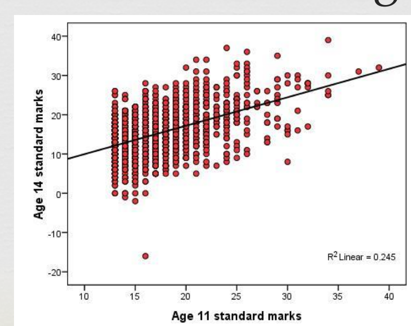
IMPORTANT CONSIDERATIONS

- Some important things to consider:
  - **Restriction of range** occurs when the individuals in your sample are very similar on the variable you are studying.
    - If one is studying age as a variable, for instance, testing only 6- and 7-year-olds will reduce one's chances of finding age effects.
    - if you reduce the range of values of the variables in your analysis than you restrict your ability to detect relationships within the wider population.

Full Range:

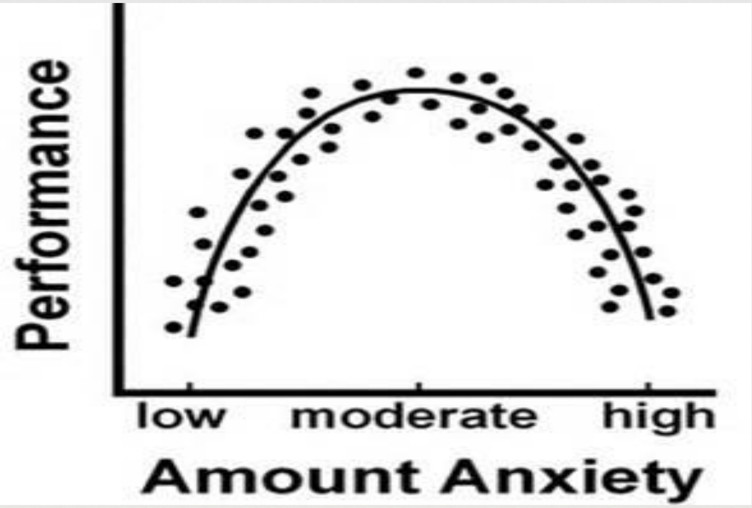


Restriction of Range:



IMPORTANT CONSIDERATIONS

- Important things to consider:
  - **Curvilinear relationship** The Pearson product-moment correlation coefficient (*r*) is designed to detect only linear relationships. If the relationship is curvilinear, the correlation coefficient will not indicate the existence of a relationship.



EFFECT SIZE

- **Effect Size**: Refers to the strength of association between variables
- **Pearson** ***r*** correlation coefficient is one indicator of effect size
  - It indicates the strength of the linear association between two variables
- Advantage of reporting effect size - Provides a scale of values that is consistent across all types of studies

EFFECT SIZE

- These reflect the differences in effect sizes for correlations:
- **Differences in effect sizes**
  - Small effects near r = .15
  - Medium effects near r = .30
  - Large effects above r = .40
- Squared value of the coefficient (r^2) transforms the value of r to a percentage
  - When you report effect size for correlations, you report r^2 as a percentage of variance that your two variables share
    - Hence, r^2 - is the Percent of shared variance between the two variables

REGRESSION EQUATIONS

- Calculations used to predict a person's score on one variable when that person's score on another variable is already known
  - $Y = a + bX$
    Y = Score one wishes to predict (DV)
    X = Score that is known (IV)
    a = Constant (y-intercept)
    b = Weighing adjustment factor (Slope)
  - To predict **criterion variable (X)** on the basis of **predictor variable (Y)**, demonstrate that there is a reasonably high correlation between the two

- FYI:
  - The **constant term** in linear regression analysis is also known as the y intercept, it is simply the value at which the fitted line crosses the y-axis.
  - While the concept is simple, there's a lot of confusion about interpreting the constant.
  - That's not surprising because the value of the constant term is almost always meaningless!
  - Paradoxically, while the value is generally meaningless, it is crucial to include the constant term in most regression models!

Regression Equation vs Linear Equation

Regression Equation

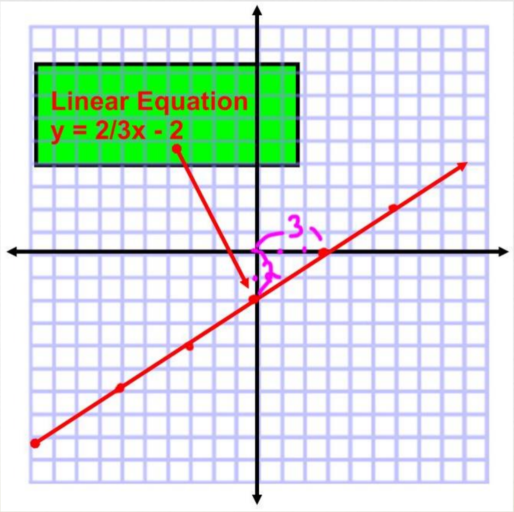
- Does this formula look familiar? **Y = a + bX**
  - Y = Score one wishes to predict
  - X = Score that is known
  - a = Constant (y-intercept)
  - b = Weighing adjustment factor (rise/run)
- To predict **criterion variable (X; DV)** on the basis of **predictor variable (Y; IV)**, one must demonstrate that there is a reasonably high correlation between the two

Linear Equation from Algebra Class:

- **Slope Intercept Form**:

  Y = mx + b

  - m = slope (rise/run)
  - b = y-intercept



MULTIPLE CORRELATION

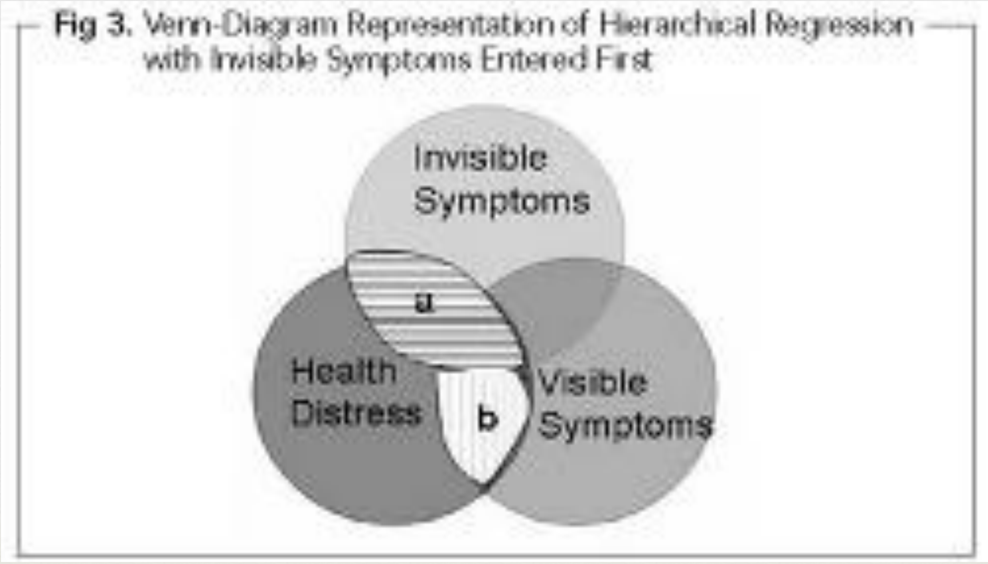
- A technique called **multiple correlation** is used to combine a number of predictor variables to increase the accuracy of prediction of a given criterion or outcome variable
- A multiple correlation (symbolized as *R* to distinguish it from the simple *r*) is the correlation between a combined set of predictor variables and a single criterion variable.
- Symbolized as *R*
  - $Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n$
    - Y = Criterion variable (DV)
    - X_1 to X_n = Predictor variables (IV)
    - a = Constant (y-intercept)
    - b_1 to b_n = Weights multiplied by scores on the predictor variables (slope)

PARTIAL CORRELATION AND THE THIRD-VARIABLE PROBLEM

- **Third-variable problem** - An uncontrolled third variable may be responsible for the relationship between two variables of interest
- **Partial correlation**: Provides a way of statistically controlling for third variables
  - A partial correlation is a correlation between the two variables of interest, with the influence of the third variable removed from, or "partialed out of," the original correlation.
  - This provides an indication of what the correlation between the primary variables would be if the third variable were held constant.
  - This is not the same as actually keeping the variable constant, but it is a useful approximation.
  - The outcome depends on the magnitude of the correlations between the third variable and the two variables of primary interest

PARTIAL CORRELATION AND THE THIRD-VARIABLE PROBLEM

- **Hierarchical Regression:**
  - Deals with the third-variable problem by statistically controlling for the effects of a third variable (a) and only looking at the effects of the variable of interest on the dependent variable (b)



STRUCTURAL EQUATION MODELING (SEM)

- ✓ **Structural equation modeling (SEM)** is a general term to refer to these statistical techniques.
- ✓ The methods of SEM are beyond the scope of this class, but one will likely encounter some research findings that use SEM. Therefore, it is worthwhile to provide an overview.
- ✓ SEM Describes expected pattern of relationships among quantitative non-experimental variables
  - ✓ After data have been collected, statistical methods describe how well the data fits the model

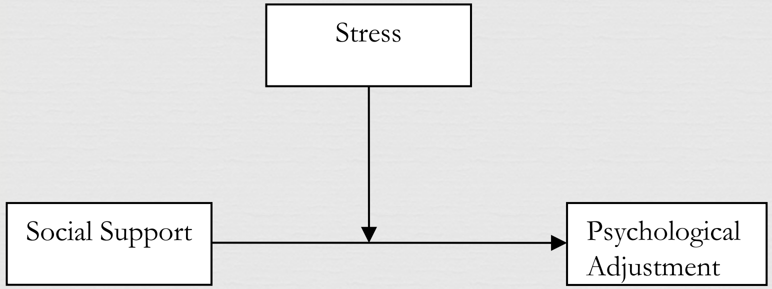
STRUCTURAL EQUATION MODELING (SEM)

- **Structural equation modeling (SEM)** uses statistical techniques known as mediation and moderation using regression techniques to identify the direction and influence that two or more variables have on another variable
  - A **moderator** is a variable that affects the direction and/or strength of relationship between an independent/predictor variable and a dependent/criterion variable
  - A **mediator** to the extent that it accounts for the relationship between the predictor and the criterion

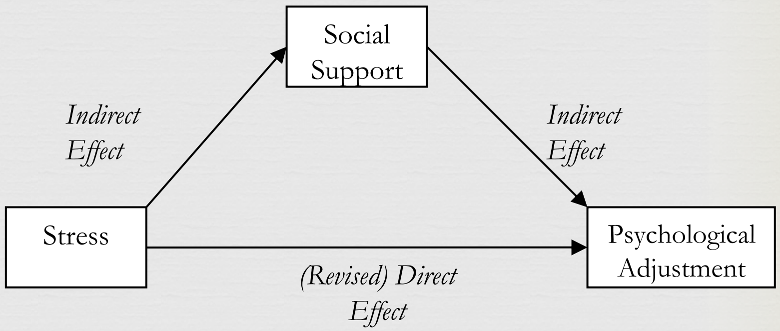
STRUCTURAL EQUATION MODELING (SEM)

- **Structural equation modeling (continued)**
  - Mediators explain how external physical events take on internal psychological significance.
  - Whereas moderator variables specify when certain effects will hold,
  - mediators speak to how or why such effects occur

Moderation (a.k.a. Interaction)



Mediation

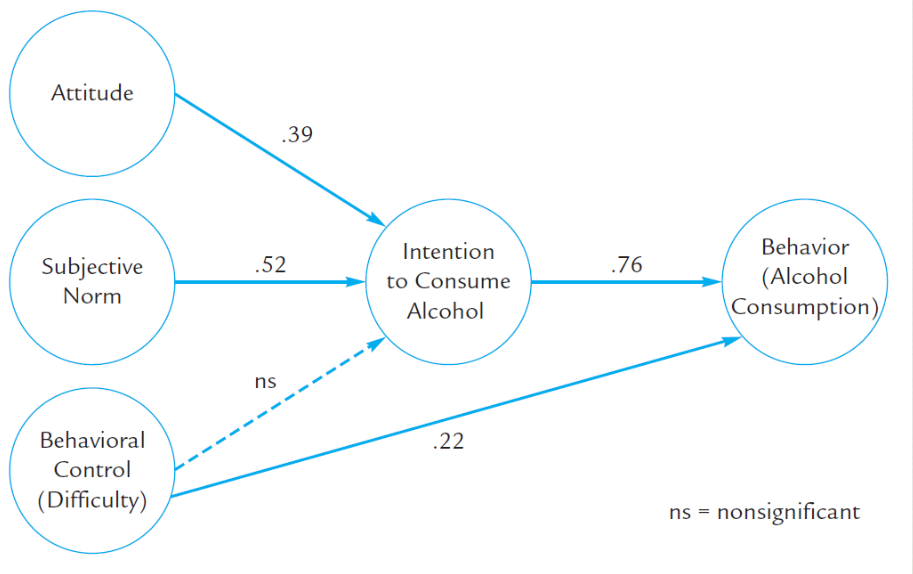


STRUCTURAL EQUATION MODELING (SEM)

- Path diagrams
  - Visual representation of the model being tested
  - Show theoretical causal paths among the variable
  - Used to study modeling
  - Arrows lead from variable to variable
  - Statistics provide path coefficients
    - Similar to standardized weights in regression equations
    - Indicate the strength of relationship between variables in the path

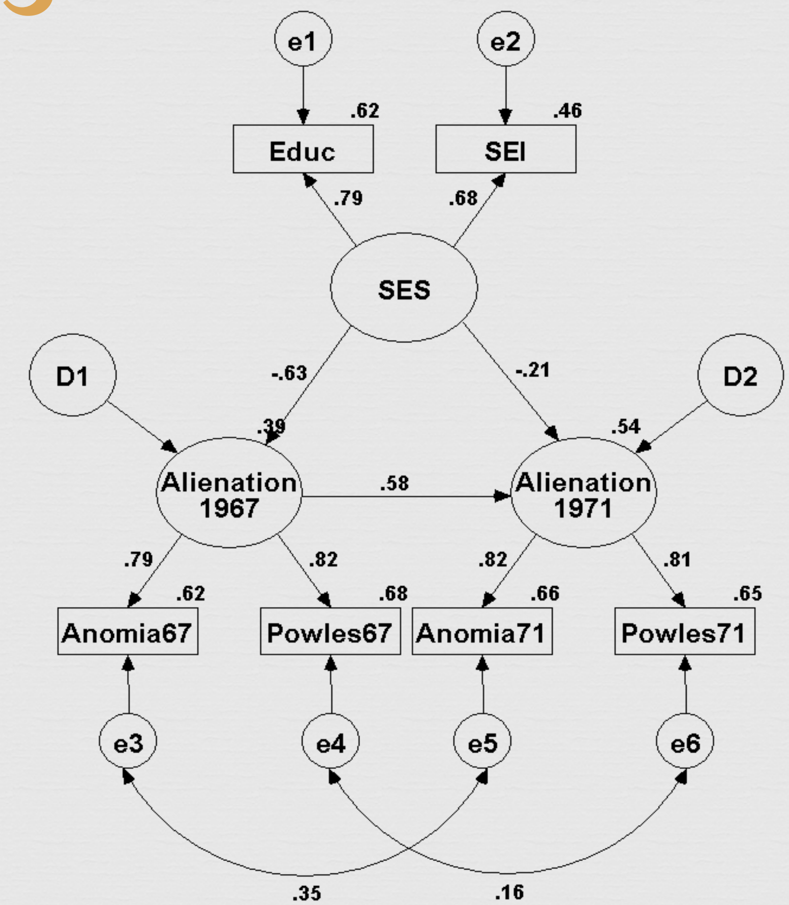
STRUCTURAL MODEL BASED ON DATA FROM HUCHTING, LAC, AND LABRIE

- ✓ Here's an example of a structural equation model.
  - ✓ The direction of the arrows indicate the direction of a variable's influence upon another variable
  - ✓ Notice the values along arrow pathways (recall +/-1 indicates a perfect relationship)



STRUCTURAL EQUATION MODELING (SEM)

- ✓ **Resulting Structural Equation Models (SEM)** can become quite intricate:



LAB

- Description & Correlation
- Describing Correlation Coefficients

(Due before class next week)

- Work on Research Projects and Final Research Papers (Due Friday, April 27th)