Chapter Outlines for:
Frey, L., Botan, C., & Kreps, G. (1999). *Investigating communication: An introduction to research methods.* (2nd ed.) Boston: Allyn & Bacon.

Chapter 14: Analyzing Relationships Between Variables

I. Introduction
   A. This chapter examines how two or more variables may be related: It starts by considering the relationship between two variables (**bivariate association**) and then expands to consider more variables.
   B. The chapter examines the types of possible relationships between variables, explains how relationships are analyzed statistically, shows how relationship analysis is used to make predictions, and introduces some advanced statistical relationship analyses used in communication research.
II. Types of Relationships
   A. A **scatter plot** (**scattergram** or **scatter diagram**) is a visual image of the ways in which variables may or may not be related.
   B. Two variables can be associated in one of three ways: unrelated, linear, or nonlinear.
   1. **Unrelated variables** have no systematic relationship; changes in one variable simply are not related to the changes in the other variable.
   2. **Linear relationships** between variables can generally be represented and explained by a straight line on a scatter plot.
   a. There are two types of linear relationships: positive and negative
   i. **Positive relationship**: Two variables move, or change, in the *same* direction.
   ii. **Negative relationship**: Two variables move in opposite directions.
   3. **Nonlinear relationships** between variables can be represented and explained by a line on a scatter plot that is not straight, but curved in some way.
   a. A **curvilinear relationship** is described by a polynomial equation, which means that it takes at least one curve, or turn, to represent the data on a scatter plot.
   i. A **quadratic relationship** is a curvilinear relationship that has only one curve in it, while cubic, quartic, and quintic relationships describe even more complex relationships between variables.
b. A **U-shaped curvilinear relationship** means that two variables are related negatively until a certain point and then are related positively.
   c. An **inverted U-shaped curvilinear relationship** means that two variables are related positively to a certain point and then are related negatively.
III. Correlations: Statistical relationships between variables
   A. A statistical relationship between variables is referred to as a **correlation**
   1. A correlation between two variables is sometimes called a **simple correlation**.
   2. The term **measure of association** is sometimes used to refer to any statistic that expresses the degree of relationship between variables.
   3. The term **correlation ratio** (**eta**) is sometimes used to refer to a correlation between variables that have a curvilinear relationship.

B. To determine the statistical correlation between two variables, researchers calculate a *correlation coefficient* and a *coefficient of determination*.
   1. Correlation coefficient: A **correlation coefficient** is a numerical summary of the type and strength of a relationship between variables.
      a. A correlation coefficient takes the form: r_{ab} = +/-x, where r stands for the correlation coefficient, a and b represent the two variables being correlated, the plus or minus sign indicates the direction of the relationship between the variables (positive and negative, respectively), and x stands for some numerical value.
         i. The first part is the sign (+ or -), which indicates the *direction* (positive or negative) of the relationship between the variables of interest.
         ii. The second part is a numerical value that indicates the *strength* of the relationship between the variable; this number is expressed as a decimal value that ranges from +1.00 (a perfect positive relationship) to -1.00 (a perfect negative relationship); a correlation coefficient of 0.00 means two variables are unrelated, at least in a linear manner.
      b. Interpreting correlation coefficients: interpreting the importance of or strength of a correlation coefficient depends on many things, including the purpose and use of the research and sample size.
      c. Calculating correlation coefficients: Researchers use a variety of statistical procedures to calculate correlation coefficients between two variables, depending on how the two variables are measured.
         i. Relationships between ratio/interval variables can be assessed in the following ways.
            (a) The **Pearson product moment correlation** calculates a correlation coefficient for two variables that are measured on a ratio or interval scale.
            (b) The **point biserial correlation** (r_{pb}) is used when researchers measure one variable using a ratio/interval scale and the other variable using a nominal scale.
         ii. Relationships between ordinal variables can be assessed in the following ways.
            (a) The **Spearman rho correlation** can be used to compare two sets of ranked scores for the same group of research participants, or the ranked scores of various items by two different groups might be compared.
            (b) **Kendall's correlation** (**tau**), which refers to three measures of association and is used in lieu of a Spearman rho correlation coefficient, typically when a researcher has a pair of ranks for each of several individuals.
         iii. The procedures for computing a correlation coefficient between nominal variables, such as **Cramer's V**, are based on the chi-square value associated with the two-variable chi-square test.
      d. Correlation matrices: A **correlation matrix** lists all the relevant variables across the top and down the left side of a matrix where the respective rows and columns meet, researchers indicate the **bivariate correlation coefficient** for those two variables and whether it is significant by using stars (such as one star for significance at the .05 level) or in a note that accompanies the matrix.
      e. Causation and correlation: Correlation is one of the criteria used to determine causation, but causation cannot necessarily be inferred from a correlation coefficient.
         i. Researchers can sometimes use the sequencing of events in time to infer causation.
         ii. Two variables may also be correlated, but their relationship is not necessarily meaningful.
   2. Coefficient of Determination: A **coefficient of determination** (r-squared) is a numerical indicator that tells how much of the variance in one variable is associated, explained, or determined by another variable.
      a. A coefficient of determination rages from 0.00 to 1.00 and is found by squaring the correlation coefficient.

  b. Researchers must pay careful attention to the correlation of determination when interpreting the results of the correlation coefficients found in their studies.

 3. Multiple correlation: A **multiple correlation** is computed when researchers want to assess the relationship between the variable they wish to explain, the **criterion variable**, and two or more other independent variables working together; and the procedure yields two types of statistics.

  a. A **multiple correlation coefficient** ® is just like a correlation coefficient, except that it tells researchers how two or more variables *working together* are related to the criterion variable of interest.

   i. A multiple correlation coefficient indicates both the direction and the strength of the relationship between a criterion variable and the other variables.

   ii. Takes the form $R_{a.b.x}$ = +/-x, read "The multiple correlation of variables b and c with variable a (the criterion variable is . . . ."

  b. A **coefficient of multiple determination** (**R-squared**, R^2) expresses the amount of variance in the criterion variable that can be explained by the other variables acting together; it is computed by squaring the multiple correlation coefficient.

   i. The **coefficient of nondetermination** is that part of the variance in the criterion variable that is left unexplained by the independent variables, symbolized and calculated as $1-R^2$.

 4. Partial Correlation: A **partial correlation** explains the relationship between two variables while statistically controlling for the influence of one or more other variables (sometimes called **effects analysis** or **elaboration**).

  a. A partial correlation coefficient takes the form $r_{ab.c}$ = +/-x, read, "The partial correlation between variable a and variable b with variable c controlled for is . . . . ."

  b. A First-order partial correlation controls for one other variable; **higher-order partial correlation** controls for two or more variables; **zero-order correlation** is a correlation between two variables with no variable being controlled.

  c. A **semi-partial correlation** partials out a variable from one of the other variables being correlated.

   i. A semi-partial correlation coefficient takes the form $r_{a(b.c)}$ = +/-x, read, "The semipartial correlation coefficient of variables a and b after variable c has been partialed out from variable b is . . . . "

IV. Regression Analysis

 A. Used to *predict or explain* how people are likely to score on a criterion, or outcome variable on the basis of their scores on another variable, called a **predictor variable** (also called a **regression**).

 B. Statistical procedures used to make such predictions are referred to as **regression analysis**.

 C. **Linear regression** (**simple regression**): used to predict or explain scores on a criterion variable on the basis of obtained scores on a predictor variable and knowledge of the relationship between the two variables.

  1. The **regression line** (**line of best fit**) is denoted by a straight line through the data on a scatter plot.

  2. Regression analysis is accomplished by constructing a **regression equation** (also called a **prediction equation** or **regression model**), which is an algebraic equation expressing the relationship between variables.

  a. The typical regression equation for two variables is: y = a + bc, where y is the criterion variable, a is the intercept, b is the slope, and x is the predictor variable.

   i. The **intercept** is how far up the y-axis the regression line crosses it.

   ii. The **slope** denotes how many units the variable Y is increasing for every unit increase in X; it depends on the correlation between the two variables.

b. A **regression coefficient**, which is part of a regression equation, is a statistical measure of the relationship between the variables (a **bivariate regression coefficient** references two variables).

c. Significance tests are applied to a regression equation to determine whether the predicted variance is significant.

d. The extent to which any model or equation, such as a regression line, summarizes or "fits" the data is referred to as the **goodness of fit**.

C. Multiple linear regression: allows researchers to predict or explain scores on a criterion variable on the basis of obtained scores on two or more predictor variables and knowledge of the relationships among all the variables.

1. There are different ways to do multiple linear regression with respect to how the predictor variables are entered into the regression analysis to see how much variance they explain in the criterion variable.

a. **Hierarchical regression analysis**: the researcher determines, on the basis of previous theory and research, the order of the variables entered into the regression equation.

b. **Stepwise regression**: the computer is instructed to enter the predictor variables in various combinations and orders until a "best" equation is found.

2. Multiple linear regression provides researchers with at least three important pieces of information

a. A *multiple correlation coefficient* ® that tells researchers the relationship between the criterion variable and all the predictor variables.

b. A *coefficient of multiple determination* (R^2) that expresses the amount of variance in the criterion variable that can be explained by the predictor variables acting together.

i. An **adjusted R2 ($*R^2$)** takes into account the number of independent variables studied.

c. How much each of the predictor variables contributes toward explaining the criterion variable by providing a *regression coefficient*, a **beta coefficient** (often called a **beta weight**, **regression weight**, or sometimes standardized regression coefficient), that indicates the extent to which, or relative weight that, each predictor variable contributes to explaining the scores on the criterion variable, while controlling for the other predictor variable.

3. Researchers must be aware of the potential problem of **collinearity** (**or multicollinearity**), the extent to which the predictor variables are correlated with one another.

a. A correlation between independent/predictor variables is called an **intercorrelation**, as compared to a correlation between an independent/predictor and dependent/criterion variable.

4. Because multiple linear regression assesses the relationships between numerous predictor variables and a criterion variable, researchers frequently use this procedure to capture the complexity of events, including communication processes, called **polynomial regression analysis** (or **curvilinear regression**).

V. Advanced Relationship Analysis

A. There are more complex multivariate analytic procedures that assess relationships among three or more variables (see Figure 14.8).

1. **Canonical correlation analysis** (**Rc**) is a form of regression analysis used to examine the relationship between multiple independent and dependent variables.

2. **Path analysis** examines hypothesized relationships among multiple variables (usually independent, mediating, and dependent) for the purpose of helping to establish causal connections and inferences by showing the "paths" the causal influences take.

3. **Discriminant analysis** is a form of regression analysis that classifies, or discriminates,

individuals on the basis of their scores on two or more ratio/interval independent variables into the categories of a nominal dependent variable.

    4. **Factor analysis** examines whether a large number of variables can be reduced to a smaller number of factors (a set of variables).

    5. **Cluster analysis** explains whether multiple variables or elements are similar enough to be placed together into meaningful groups or clusters that have not been predetermined by the researcher.

    6. **Multidimensional scaling** (**MDS**) plots variables or elements in two or more dimensions to see the statistical similarities and differences between and among them.

VI. Conclusion

  A. Like the relationships that exist between people, relationships between variables range from positive to neutral to negative.

  B. Relationships among variables quickly become numerous and complex, especially when the goal is to make a prediction about something.

  C. Caution must be exercised about the difference between correlation and causation; researchers and readers of research must be careful interpreting statistical relationships between variables.