**Chapter 14: More Powerful Statistical Methods**

**Multiple Choice**

1. If you are studying the impact of training on salesperson performance as indicated on an exam taken by salespersons at the end of training, salesperson sales would be the \_\_\_\_\_\_\_\_\_\_\_\_ variable, and the salesperson exam score would be the \_\_\_\_\_\_\_\_\_\_\_\_\_ variable.

a. criterion, predictor

b. covariate, independent

c. independent, predictor

d. dependent, independent

e. covariate, dependent

Answer: d

1. Which of the following techniques requires that both variables are ratio or interval data?

a. bivariate regression

b. ANOVA

c. *t* test

d. Pearson’s product-moment correlation

e. All are appropriate.

Answer: a

1. The percentage of total variation in the dependent variable that is described by the independent variable is expressed by\_\_\_\_\_\_\_\_\_.

a. coefficient of determination

b. correlation coefficient

c. coefficient of covariation

d. regression coefficient

e. none of these

Answer: a

1. In a regression equation, sales would typically be the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ variable.

a. independent

b. coefficient

c. dependent

d. none of these

Answer: c

1. \_\_\_\_\_\_\_\_\_\_\_\_ is the mathematical technique for fitting a line that best describes the relationship between two variables.

a. Pearson’s product-moment correlation

b. Least squares procedure

c. Chi-square test

d. Significance test

e. none of these

Answer: b

1. A regression equation that estimates the number of hours of study that impact a test score would most likely produce what type of relationship?

a. positive

b. negative

c. curvilinear

d. none of these

Answer: a

1. Research on fear appeal in advertising suggests that a moderate amount of fear is most effective at achieving persuasion. That is, using not enough fear or too much fear in advertising will not be effective. This is an example of what type of relationship that regression might find?

a. positive linear

b. negative linear

c. curvilinear

d. no relationship

e. none of these

Answer: c

1. Which of the following is the appropriate technique for correlation analysis, which involves metric (interval or ratio) data?

a. Spearman’s rank-order correlation

b. Jensen’s analysis of covariance

c. Kendall’s coefficient of concordance

d. Pearson’s product-moment correlation

e. Edward’s metric analysis

Answer: d

1. What are the possible values in correlation analysis?

a. 0 to 1

b. -1 to +1

c. -1 to 0

d. -.99 to + .99

e. -∞ to +∞

Answer: b

1. When two variables are not correlated at all, the correlation coefficient would be \_\_\_\_\_\_\_.

a. -1

b. 0

c. 1

d. -∞

e. +∞

Answer: b

1. When two variables are perfectly positively correlated, the correlation coefficient would be \_\_\_\_\_\_\_.

a. -1

b. 0

c. 1

d. -∞

e. +∞

Answer: c

1. The explained variation by the regression is referred to as the:

a. total variation.

b. bivariate.

c. sum of squares due to regression.

d. none of these

Answer: c

1. If a marketer wants to correlate an ordinal ranking of TV sets with the nominal scale variable gender, the marketer would have to use\_\_\_\_\_\_\_\_\_\_\_\_\_.

a. ANOVA

b. Pearson’s product-moment correlation

c. z-test

d. chi-square test

e. none of these

Answer: e

1. Which test statistic is used to test the significance of the results of a regression analysis?

a. F

b. t

c. Z

d. chi-square

e. least square

Answer: a

1. In regression, the higher the \_\_\_\_\_\_\_\_\_ value, the more likely the relationship between variables is significant.

a. F

b. p

c. Z

d. t

e. none of these

Answer: a

1. Bivariate regression *cannot* demonstrate:

a. when the two variables are linear.

b. when the two variables are strongly inversely related.

c. when the two variables are strongly positively related.

d. when the two variables are causally related.

Answer: d

1. Correlation *cannot* demonstrate:

a. when the two variables are linear.

b. when the two variables are strongly inversely related.

c. when the two variables are strongly positively related.

d. when the two variables are causally related.

Answer: d

1. A correlation analysis between sales and sales training scores results in R = +.98. Which of the following best interprets the relationship between sales and sales training?

a. Exactly 98% of the salespeople taking the test have higher sales.

b. The correlation between sales and sales training is very weak and insignificant.

c. The correlation between sales and sales training is strong and positive, indicating that higher sales training scores are closely associated with higher sales and vice-versa.

d. Exactly 98% of the variation in sales is explained by variations in sales training scores.

e. none of these

Answer: c

1. When the value of one variable increases at exactly the same rate as another variable decreases, this is said to be a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_relationship.

a. strong positive linear

b. positive linear

c. perfect negative linear

d. parabolic

e. strong curvilinear association

Answer: c

1. In a regression output, R2 represents:

a. the independent variable.

b. the dependent variable.

c. the estimated slope of the regression line.

d. the coefficient of determination.

Answer: d

1. An R2 of .05 would indicate:

a. a significant relationship between X and Y.

b. a negative relationship between X and Y.

c. a very weak relationship between X and Y.

d. none of these

Answer: c

1. In a regression equation, which symbol represents the dependent variable?

a. *Y*

b. *X*

c.e

d. r

e. p

Answer: a

1. Which of the following is *not* included in the equation for the coefficient of determination?

a. explained variation

b. total variation

c. estimated variation

d. unexplained variation

Answer: c

1. \_\_\_\_\_\_\_\_\_\_\_ are statistical methods of analyzing the relationship between two variables.

a. Bivariate techniques

b. Univariate techniques

c. Multivariate techniques

d. None of these

Answer: a

1. The dependent variable is also called the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

a. measure

b. criterion

c. principle

d. standard

e. predictor

Answer: b

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ refers to analysis is one of strength of the linear relationship between two variables when one is considered the independent variable and the other the dependent variable

a. Multivariate regression

b. Univariate regression

c. Bivariate regression

d. Trivariate regression

e. None of these

Answer: c

1. What is the variation not explained by the regression?

a. Sum of squares due to regression

b. Error sum of squares

c. Neither sum of squares due to regression or Error sum of squares

d. Both Sum of squares due to regression and Error sum of squares

Answer: b

1. The degree to which changes in one variable (the dependent variable) are associated with changes in another is which of the following?

a. correlation

b. consideration

c. regression

d. association

e. none of these

Answer: a

1. What is multivariate analysis?

a. a procedure for managing large database systems.

b. synonymous with decision support analysis.

c. a group of statistical procedures used to analyze simultaneously multiple measurements of the individual or object being measured.

d. the measurement procedure for monitoring how several marketing variables function in multiple market situations.

Answer: c

1. Which of the following is a procedure that is useful for analyzing the relationship among two or more predictor/independent variables and one metric criterion/dependent variable?

a. multiple discriminant analysis

b. multiple regression analysis

c. conjoint analysis

d. cluster analysis

Answer: b

1. Which of the following would not be an application best suited for multiple regression analysis?

a. estimating the relationship between various demographic and lifestyle variables and the attitude toward a particular product

b. determining which variables best predict the sales of a particular product or service

c. estimating the effect of several marketing mix variables on product sales

d. reducing a large number of potential predictor variables to a smaller set of composite variables

Answer: d

1. The coefficient of determination is normally associated with which of the following multivariate procedures?

a. multiple regression analysis

b. factor analysis

c. conjoint analysis

d. cluster analysis

Answer: a

1. Which of the following multivariate procedures does not include a dependent variable in its analysis?

a. multiple regression analysis

b. cluster analysis

c. multiple discriminant analysis

d. All of these procedures include a dependent variable.

Answer: b

1. Which of the following techniques might be most useful in helping a restaurant determine how to position itself in the marketplace?

a. factor analysis

b. multiple regression analysis

c. conjoint analysis

d. perceptual mapping

Answer: d

1. A dependent variable is coded 1=respondent did purchase and 0=respondent did not purchase. Independent variables include various demographic and lifestyle characteristics of the respondents. The goal of the analysis is to determine how respondents who did purchase are different from respondents who did not purchase. Which of the following procedures goes with the preceding description?

a. multiple discriminant analysis

b. multiple regression analysis

c. cluster analysis

d. perceptual mapping

Answer: a

1. A company that is trying to determine which three distinct target markets to focus its marketing efforts on might use which statistical tool?

a. ANOVA

b. multiple regression analysis

c. cluster analysis

d. bivariate regression

Answer: c

1. Cluster analysis is particularly valuable for what type of marketing strategy?

a. product differentiation

b. positioning

c. segmentation

d. cost leadership

Answer: c

1. Suppose an analyst wanted to determine whether or not dollars spent on advertising, number of sales people, number of new products introduced, and dollars spent on research and development were contributing to the growth in company market share. Which of the following procedures would be most appropriate?

a. conjoint analysis

b. cluster analysis

c. multiple regression analysis

d. perceptual mapping

Answer: c

1. Which of the following is *not* a step in conjoint analysis?

a. collecting trade-off data

b. estimating factor loadings

c. estimating preference structures

d. predicting consumers’ choice

Answer: b

1. An insurance company wants to know whether the subscribers to its HMO, PPO, or traditional single indemnity plan exhibit better health characteristics. They have the following data for each of their subscribers: % of body weight over optimum weight; blood pressure; cholesterol and hours per week of exercise. The analysis will group subscribers as either healthy, somewhat healthy or not healthy. Which of the following would be the best procedure for making such a determination?

a. factor analysis

b. multiple discriminant analysis

c. multiple regression analysis

d. cluster analysis

Answer: d

1. Summarizing a number of survey questions into a single concept could be achieved with which type of analysis?

a. factor analysis

b. multiple discriminant analysis

c. multiple regression analysis

d. cluster analysis

Answer: a

1. Which of the following techniques is often used for concept testing of new products?

a. conjoint analysis

b. multiple regression analysis

c. perceptual mapping

d. cluster analysis

Answer: a

1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_ is a measure of the percentage of the variation in the dependent variable that is explained by the variation in the independent variables.

a. Discriminant weights

b. Factor loadings

c. Coefficient of determination

d. Coefficient of correlation

Answer: c

1. Dummy variables created to include nominally scaled variables in an analysis are commonly used in:

a. multiple discriminant analysis.

b. factor analysis.

c. conjoint analysis.

d. multiple regression analysis.

Answer: d

1. Independent variables are important components in multiple regression analysis and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

a. multiple discriminant analysis

b. factor analysis

c. conjoint analysis

d. cluster analysis

Answer: a

1. Which of the following is multiple discriminant analysis most similar to?

a. multiple regression analysis

b. factor analysis

c. conjoint analysis

d. cluster analysis

Answer: a

1. Cluster analysis produces mutually exclusive and exhaustive groups such that the individuals or objects grouped are \_\_\_\_\_\_\_\_\_\_\_\_\_\_ within and \_\_\_\_\_\_\_\_\_\_\_\_\_\_ between groups.

a. homogeneous; heterogeneous

b. homogeneous; homogeneous

c. heterogeneous; homogeneous

d. heterogeneous; heterogeneous

Answer: a

1. If the goal is to classify business travelers into distinct groups based on their responses to 20 questions on preferences to mode of transportation, hotel accommodation, and ethnic food, which of the following techniques would be most appropriate?

a. multiple regression analysis

b. conjoint analysis

c. cluster analysis

d. factor analysis

Answer: c

1. The purpose of factor analysis can be best described as:

a. data simplification by summarizing the information contained in a large number of variables into summary measures.

b. presenting the relationship between multiple categorical measures in a low-dimensional diagram.

c. assessing the significance of group differences across many metric dependent variables simultaneously.

d. assessing the degree of correlation of a group of categorical measures.

Answer: a

1. This analytical technique helps determine what features a new product or service should have and how it should be priced.

a. factor analysis

b. perceptual mapping

c. multiple discriminant analysis

d. conjoint analysis

Answer: d

1. In factor analysis, the correlation between factor scores and the original variables is known as the factor \_\_\_\_\_\_\_\_.

a. loading

b. mapping

c. attribute rating

d. conjoint

Answer: a

1. All of the following are techniques for multivariate analysis procedure except \_\_\_\_\_.

a. Cluster analysis

b. Factor analysis

c. Conjoint analysis

d. Bivariate Regression analysis

Answer: d

1. In the multiple regression formula, which of the following represents the dependent variable?

a. Y

b. a

c. *b*1

d. *X*1

Answer: a

1. In the case of multiple regression analysis, the dependent variable must be \_\_\_\_\_\_\_\_\_\_\_; in multiple discriminant analysis, the dependent variable is\_\_\_\_\_\_\_\_\_\_ in nature.

a. metric; nonmetric

b. metric; nominal

c. nominal; categorical

d. categorical; nonmetric

Answer: b

1. An adjusted *R* 2 corrects the coefficient of determination based on the relationship between the number of predictor variables and which of the following?

a. overall sample size

b. coefficient of correlation

c. regression results

d. population

Answer: a

1. Which of the following is not an example of a cluster technique?

a. K-means

b. Nearest neighbor

c. Decision trees

d. BIRCH

e. All of these are examples

Answer: e

1. Hadoop, an open-source platform, is most closely related to which of the following?

a. Factor Analysis

b. Cluster Analysis

c. Multiple Discriminant Analysis

d. Conjoint Analysis

e. Big Data

Answer: e

1. Which of the following is not a part of using predictive analysis?

a. Acquiring a Data Set

b. Pre-processing

c. Modeling

d. Validating Results

e. All of these are a part

Answer: e

True/False

1. Dependent variables are also known as predictor variables.

Answer: False

1. Bivariate statistical techniques are often used to establish cause and effect relationships between two or more variables.

Answer: False

1. In correlation analysis, the closer the value of R is to –1 or +1, the stronger the correlation between the two variables in question.

Answer: True

32. A perfect correlation indicates that two variables are causally related.

Answer: False

1. An R = -.95 is always weaker than +.95, because positive relationships are more significant than negative relationships.

Answer: False

1. In bivariate regression, both the independent and dependent variables are measured in the form of nominal and/or ordinal data.

Answer: False

1. The variation explained by the regression equation is called SSR or sum of squares due to the regression.

Answer: True

1. The least squares estimation procedure often results in a perfectly straight line.

Answer: False

1. The R2 statistic ranges from -1 to 1.

Answer: False

1. The coefficient of determination is used in correlation analysis.

Answer: False

1. A scatter plot compares actual data to a predicted relationship between two variables.

Answer: True

1. An R2 of 0 indicates no relationship between variables.

Answer: True

1. One way to study the relationship between the dependent and independent variable is to plot the data in a scatter diagram.

Answer: True

1. One way to study the relationship between the dependent and independent variable is to plot the data in a bar chart.

Answer: False

1. In the equation for regression analysis, the letter X represents the independent variable.

Answer: True

1. The correct formula for the analysis of variance used to test the significance of results is F = MSR / MSE.

Answer: True

1. In regression, the unexplained variation is referred to as error sum of squares (SSE).

Answer: True

1. The coefficient of correlation describes the degree of association between X and Y.

Answer: True

1. Multivariate techniques analyze relationships between more than two variables.

Answer: True

1. In multiple regression analysis the dependent variable is nominal or categorical.

Answer: False

1. If the goal of an analysis is to group respondents into mutually exclusive and collectively exhaustive subgroups, the preferred procedure would be cluster analysis.

Answer: True

1. Multiple regression analysis can group respondents into subgroups when there are at least 4 classification variables.

Answer: False

1. The independent variables in multiple regression should not be correlated with each other.

Answer: True

1. There is no dependent variable in a factor analysis.

Answer: True

1. Multiple regression enables the researcher to show that two variables are causally related.

Answer: False

1. In order to make valid comparisons of regression coefficients, the coefficients must be scaled in the same units, or be standardized.

Answer: True

1. An analyst will always prefer that the R2 be close to zero.

Answer: False

1. The goal of factor analysis is to identify as many factors as possible.

Answer: False

1. Collinearity is a situation in which independent variables are heavily correlated with each other.

Answer: True

1. Multiple discriminant analysis is useful when the dependent variable is metric.

Answer: False

1. Cluster analysis is useful for clustering groups of questions together into factors.

Answer: False

1. If a company was interested in the impact that advertising and sales price had on sales, cluster analysis would be an appropriate tool to use.

Answer: False

1. In multiple regression, the value of ***R2*** is influenced by the number of predictor variables relative to sample size.

Answer: True

1. In factor analysis, naming factors is a somewhat subjective step, combining intuition and knowledge of the variables with an inspection of the variables that have high loadings on each factor.

Answer: True

1. The type of conjoint approach used impacts how the exercise is presented and what statistical procedures are most appropriate for analyzing the results.

Answer: True

1. One of the most common approaches to conducting conjoint analysis is the use of a discrete choice.

Answer: True

1. If key attributes or popular options within key attributes are excluded from the study, demand estimates could be severely impacted in conjoint analysis.

Answer: True

1. Predictive analytics do not apply to big data.

Answer: False

Essay Questions

1. Some of the results of a regression analysis are as follows:  
   SSE = 52,500 SST = 485,200  
     
   Compute the Coefficient of Determination (R2) and interpret.

Answer: R2 = (SST -- SSE) / SST  
R2 = .8918 which means that over 89% of the variation in the dependent variable can be explained by the independent variable.

1. Interpret the correlation coefficient for the following data:

|  |  |
| --- | --- |
| **Sales** | **Sales Training Score** |
| 45,000 | 98 |
| 34,500 | 74 |
| 23,750 | 57 |
| 51,450 | 99 |
| 41,000 | 85 |

The resulting correlation coefficient is +.98 and the probability of insignificance is 1%.

Answer: There is a strong positive correlation between sales and sales training scores. Hence, higher sales are associated with higher sales training scores and vice-versa.

1. Draw a graph depicting a strong but inverse correlation between X (Sales) and number of times the sales force did not reach its quota. Interpret the relationship.

Answer: The interpretation of the relationship would be that as the number of times the sales force did not reach its quota increases, sales decrease. The resulting graph will be linear and down sloping to the right (negative slope).

Given the correlation matrix below, answer the following three questions.

|  |  |  |  |
| --- | --- | --- | --- |
|  | AGE | INCOME | EDUCATION |
| AGE | 1.00  .000  400 | -.65  .000  400 | -.10  .221  400 |
| INCOME | -.65  .000  400 | 1.00  .000  400 | .62  .000  400 |
| EDUCATION | -.10  .221  400 | .62  .000  400 | 1.00  .000  400 |

Assume that respondents answered the questions with ratio scale responses.

1. Which of the relationships in the matrix is the strongest?

Answer: The strongest relationship is between AGE and INCOME, and it indicates that in the sample results that younger persons consistently had higher incomes than older people, and that we can be 100% confident that this trend will also occur in the population from which the sample was drawn.

1. How would you interpret the relationship between Education and Age?

Answer: The relationship between EDUCATION and AGE is insignificant, as the results in the matrix indicate a 21% chance that the trend in the sample data would not necessarily occur in the population from which was drawn.

1. How many pairs of responses were in this correlation analysis?

Answer: The results in the matrix indicate that there were 400 pairs of results for each of the correlated results.

1. Explain the meaning of the coefficient of determination. What does this coefficient tell the researcher about the nature of the relationship between the dependent and independent variables?

Answer: The coefficient of determination tells the marketing researcher how much variation in the dependent variable can be explained by variation in the independent variable. It is a measure of the strength of the relationship. If the coefficient of determination is low, the independent variable does not have significant explanatory power in predicting changes in the dependent variable.

1. An analyst is trying to group together persons who attend theatre performances and who prefer a dinner to accompany the theatre performance. Hence, the two pieces of information are 1) frequency of attending theatre performances and 2) preference for a dinner-theatre format. What type of analysis should be undertaken to determine if those persons preferring the dinner-theatre format are those who attend theatre performances more frequently?

Answer: This problem would be best solved using cluster analysis, which would simultaneously group persons preferring the dinner-theatre format with how frequently they attend theatre performances.

1. An advertising agency has been doing work for a client selling widgets. The three-month campaign has produced a low correlation between advertising expenditures and sales for its client. Hence, the client is considering firing the ad agency. The ad agency counters that consumer sales are not a fair assessment of the effectiveness of the ad campaign after only three months. They counter with an analysis of advertising expenditures in relation to number of requests for information about the widgets; number of distributors stocking widgets; and number of retailers requesting shipments of widgets. The ad agency has a database with such information. What kind of analysis would best assist the ad agency in making their case for the effectiveness of their ad campaign?

Answer: Since the ad agency wants to show that advertising expenditures are strongly related to the requests for information about widgets, number of distributors stocking widgets, and number of retailers requesting shipments of widgets, they should invoke a multiple regression analysis.

1. A client has a data set that would be appropriate for a linear model. He wants the resulting model to predict whether or not a person would buy a particular product. The client has been advised that a multiple regression analysis would be the best approach. What would you suggest?

Answer: Multiple regression analysis would be appropriate if the goal is to predict what sales would be given a set of predictor variables. However, if the goal was to simply predict whether or not a particular type of purchase behavior would occur, then the proper method would be multiple discriminant analysis.

1. Suppose a marketer was trying to predict the sales (Y) of a product for a given level of advertising (X1) and sales performance (X2). The resulting output was as follows:  
     
   Y = 2,300 + 34(X1) + 15.5(X2) with a R2 = .12  
     
   Given the above results, what would you suggest to the marketer?

Answer: The resulting regression equation depicts a positive relationship between Sales and both Advertising and Salesperson Performance, which means greater inputs of advertising and higher salesperson performance result in higher sales. However, the R2 is only .12, meaning that only 12% of the variation in the dependent variable Sales (Y) is explained jointly by Advertising (X1) and Salesperson Performance (X2). There must be other performance variables that would explain more of the variation in Sales. The consultant should suggest additional analysis of the company database to find additional variables that would explain more of the variation in Sales.

1. Describe the potential problem of collinearity and multiple regression. How might a researcher test for collinearity? If collinearity is a problem, what should the researcher do?

Answer: Collinearity refers to the condition when a significant correlation exists between two or more independent variables. This condition reduces the statistical power of significance tests for the regression coefficients. One can test for collinearity by examining the correlation matrix. If there is a value higher than .30, the researcher should consider corrective action. This correction might be accomplished be dropping one of the correlated variables, or collapsing the correlated variables into a single variable.