Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

Chapter 2: Advanced Data Screening, Outliers, and Missing Values

Test Bank

Multiple Choice

1. Prior to beginning to analyze data collected during a study, researchers should _____.
a. report confidentiality violations
b. plan secondary analyses
c. conduct extensive data screening
d. notify involved parties about anticipated results
Ans: C
Cognitive Domain: Knowledge
Answer Location: 2.1: Introduction
Difficulty Level: Easy

2. Empty cells in an SPSS worksheet indicate _____.
a. data entry errors
b. missing values
c. standard errors
d. outliers
Ans: B
Cognitive Domain: Knowledge
Answer Location: 2.2.2: Codes for Missing Values
Difficulty Level: Easy

3. In a dataset including participant age in years and cognitive ability scores that can range from 50–145, an entry of 999 for several participants indicates a(n) _____.
a. a missing value
b. an outlier
c. a data entry error
d. a replication error
Ans: A
Cognitive Domain: Knowledge
Answer Location: 2.2.2: Codes for Missing Values
Difficulty Level: Easy

4. For which of the following items would an iresearcher who has administered an inventory designed to measure work-related stress need to use reverse scoring ?
a. Work interferes with my family life.
b. My responsibilities are much different than I had anticipated.

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

c. I do not get the recognition I believe I deserve.
d. Colleagues at work contribute their fair share.
Ans: D
Cognitive Domain: Analysis
Answer Location: 2.2.4: Use Different Variable Names to Keep Track of Modifications
Difficulty Level: Hard

5. Researchers should screen datasets carefully so that the _____.
a. probability of rejecting the null hypothesis is increased
b. results of their analyses will not be biased
c. overlap of confidence intervals will be minimized
d. results will have maximum practical significance
Ans: B
Cognitive Domain: Comprehension
Answer Location: 2.2.5: Save SPSS Syntax
Difficulty Level: Medium

6. Over- or under-estimation of a statistic is _____.
a. error variance
b. plagiarism
c. bias
d. miscoding
Ans: C
Cognitive Domain: Knowledge
Answer Location: 2.3: Sources of Bias
Difficulty Level: Easy

7. When participants in a study can influence the behavior of other participants through such actions as cooperation or competition, the resulting bias may _____.
a. make the values of *t* or *F* too large
b. reduce the risk of Type I errors
c. increase the size of the group standard deviations
d. increase the practical significance of statistically significant results
Ans: A
Cognitive Domain: Comprehension
Answer Location: 2.3: Sources of Bias
Difficulty Level: Medium

8. When an analysis based on the general linear model is planned, an important assumption is that _____.
a. residuals for all variables are dependent on each other
b. are skewed to one tail of the distribution, but not both tails
c. have equal variances for all values of the predictor variables
d. have a mean of 1.00 for all values of the predictor variables
Ans: C
Cognitive Domain: Comprehension

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

Answer Location: 2.3: Sources of Bias
Difficulty Level: Medium

9. In order to reduce or remove skewness from a distribution, a researcher can consider _____.
a. adjusting the expected *p* values
b. adding a minimum of five points to each score
c. ignoring residuals
d. removing outliers
Ans: D
Cognitive Domain: Comprehension
Answer Location: 2.3: Sources of Bias
Difficulty Level: Medium

10. A researcher who eliminates participants in the comparison group whose scores on an important variable overlap scores in the experimental group in order to obtain a *p* value < .05 is practicing _____.
a. HARKing
b. power analysis
c. *p*-hacking
d. replication
Ans: C
Cognitive Domain: Comprehension
Answer Location: 2.4: Screening Sample Data
Difficulty Level: Medium

11. Obtained *p* value estimates are not seriously biased when_____.
a. $SD \leq 10$ for each group
b. 1-tailed tests are applied
c. outliers are retained
d. $n \geq 30$ per group
Ans: D
Cognitive Domain: Comprehension
Answer Location: 2.4.3: Data Screening for Comparison of Group Means
Difficulty Level: Medium

12. The most appropriate transformation for a distribution in which the smallest score is 10 and the largest score is 1,000 is to _____.
a. square each score
b. use a log transformation
c. use an arcsine transformation
d. convert *Fisher r* to *Z*
Ans: B
Cognitive Domain: Application
Answer Location: 2.5: Possible Remedy for Skewness: Nonlinear Data Transformations
Difficulty Level: Medium

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

13. If a dataset consisting of proportions is seriously skewed and requires transforming, the appropriate transformation is _____.
a. square each score
b. use a log transformation
c. use an arcsine transformation
d. *Fisher r* to *Z*
Ans: C
Cognitive Domain: Application
Answer Location: 2.5: Possible Remedy for Skewness: Nonlinear Data Transformations
Difficulty Level: Medium

14. If a dataset consisting of correlations is seriously skewed and requires transforming, the appropriate transformation is _____.
a. square each score
b. use a log transformation
c. use an arcsine transformation
d. *Fisher r* to *Z*
Ans: D
Cognitive Domain: Application
Answer Location: 2.5: Possible Remedy for Skewness: Nonlinear Data Transformations
Difficulty Level: Medium

15. In order to quantify the distance that an outlier lies from the cloud that contains most of the data in a bivariate distribution, a researcher can calculate the _____.
a. Mahalanobis distance
b. standard deviation
c. interquartile range
d. coefficient of nondetermination
Ans: A
Cognitive Domain: Knowledge
Answer Location: 2.6.2: Bivariate and Multivariate Outliers
Difficulty Level: Easy

16. The test used to determine the statistical significance for Mahalanobis distances is _____.
a. t
b. F
c. R^2
d. X^2
Ans: D
Cognitive Domain: Knowledge
Answer Location: 2.6.2: Bivariate and Multivariate Outliers
Difficulty Level: Easy

17. A researcher who has trimmed a dataset has _____.

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

a. Winsorized the dataset
b. removed all extreme values from the dataset
c. dealt with bivariate outliers
d. recorded the square root of all outliers
Ans: B
Cognitive Domain: Knowledge
Answer Location: 2.7.2: Handling Univariate Outliers
Difficulty Level: Easy

18. If a researcher replaces a score that is one standard higher than the next highest score with the next highest score, the researcher is _____.
a. truncating the dataset
b. trimming the dataset
c. Winsorizing the dataset
d. hacking the dataset
Ans: C
Cognitive Domain: Knowledge
Answer Location: 2.7.2: Handling Univariate Outliers
Difficulty Level: Easy

19. The most direct way to assess whether a linear relationship exists between two variables is to _____.
a. examine a bivariate scatterplot
b. calculate the Mahalanobis distance scores
c. regress X on Y
d. regress X^2 on Y
Ans: A
Cognitive Domain: Knowledge
Answer Location: 2.8: Testing Linearity Assumptions
Difficulty Level: Easy

20. When determining prior to conducting a regression analysis, the statistic of interest is _____.
a. F
b. X^2
c. R^2
d. t
Ans: C
Cognitive Domain: Comprehension
Answer Location: 2.8: Testing Linearity Assumptions
Difficulty Level: Medium

21. If adding X^2 as predictor in addition to X results in a significant increase in R^2, the trend is _____.
a. linear
b. quadratic

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

c. cubic
d. polynomial
Ans: B
Cognitive Domain: Comprehension
Answer Location: 2.8: Testing Linearity Assumptions
Difficulty Level: Medium

22. If adding X^2 and X^3 as predictors in addition to X results in a significant increase in R^2, the trend is _____.
a. linear
b. quadratic
c. cubic
d. polynomial
Ans: C
Cognitive Domain: Comprehension
Answer Location: 2.8: Testing Linearity Assumptions
Difficulty Level: Medium

23. If a researcher constructing a correlation matrix based on six variables eliminates any participant who does not have a score for all six variables, the deletion method is said to be _____.
a. Winsorizing
b. truncatation
c. listwise
d. pairwise
Ans: C
Cognitive Domain: Knowledge
Answer Location: 2.10.1: Why Missing Values Create Problems
Difficulty Level: Easy

24. If a researcher constructing a correlation matrix based on six variables includes all participants who have scores for both variables in any given pair of variables, the deletion method is said to be _____.
a. Winsorizing
b. truncatation
c. listwise
d. pairwise
Ans: D
Cognitive Domain: Knowledge
Answer Location: 2.10.1: Why Missing Values Create Problems
Difficulty Level: Easy

25. Using listwise deletion to handle missing data may not result in serious problems if the amount of missing data is _____.
a. < 5%
b. < 7%

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

c. < 9%
d. <10%
Ans: A
Cognitive Domain: Knowledge
Answer Location: 2.10.1: Why Missing Values Create Problems
Difficulty Level: Easy

26. If a university office that initially contacted 500 students about their satisfaction with dining options on campus only received 200 responses, how should the office describe the results in the report to the university president?
a. only report the number of actual responses
b. use imputation to replace missing values
c. directly address the possibility that the sample is unrepresentative
d. assume that the respondents are representative of all students
Ans: C
Cognitive Domain: Application
Answer Location: 2.11: How Missing Data Arise
Difficulty Level:

27. Missing data that can be attributed to the reluctance of people with certain personal and social characteristics to respond to some research conditions of questions can be described as _____.
a. response aversion
b. type A missingness
c. type B missingness
d. nonresponse tendency
Ans: B
Cognitive Domain: Comprehension
Answer Location: 2.12.1: Type A and Type B Missingness
Difficulty Level: Medium

28. Missing responses on an anxiety inventory that can be attributed to the respondent's high level of anxiety can be described as _____.
a. response aversion
b. type A missingness
c. type B missingness
d. nonresponse tendency
Ans: C
Cognitive Domain: Comprehension
Answer Location: 2.12.1: Type A and Type B Missingness
Difficulty Level: Medium

29. A factor that should be considered before using listwise deletion is _____.
a. statistical power is reduced
b. the sample composition varies unsystematically
c. obtaining a statistically significant result increases unrealistically

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

d. generalizations to populations of interest become more accurate
Ans: A
Cognitive Domain: Comprehension
Answer Location: 2.14: Possible Remedies for Missing Data
Difficulty Level: Medium

30. Multiple imputation may be described as comparable to _____.
a. nonrandom replacement
b. replacement after follow-up
c. random replacement
d. replacement by regression
Ans: D
Cognitive Domain: Knowledge
Answer Location: 2.15: Empirical Example: Multiple Imputation to Replace Missing Values
Difficulty Level: Easy

31. A reader of a journal article interested in a description of how the sample used in a study should find this information in the _____.
a. Introduction
b. Method section
c. Results section
d. Discussion section
Ans: B
Cognitive Domain: Knowledge
Answer Location: 2.17: Reporting Guidelines
Difficulty Level: Easy

32. A reader of a journal article interested in a description of how outliers were detected and handled used in a study should find this information in the _____.
a. Introduction
b. Method section
c. Results section
d. Discussion section
Ans: C
Cognitive Domain: Knowledge
Answer Location: 2.17: Reporting Guidelines
Difficulty Level: Easy

33. A reader of a journal article interested in learning about any imputation methods used in a study should find this information in the _____.
a. Introduction
b. Method section
c. Results section
d. Discussion section
Ans: C

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

Cognitive Domain: Knowledge
Answer Location: 2.17: Reporting Guidelines
Difficulty Level: Easy

34 A reader of a journal article interested in learning about the ways in which missing data may have affected the generalizability of a study should find this information in the _____.
a. Introduction
b. Method section
c. Results section
d. Discussion section
Ans: D
Cognitive Domain: Knowledge
Answer Location: 2.17: Reporting Guidelines
Difficulty Level: Easy

True/False

1. Numbers used to identify participants in research studies should never be removed.
Ans: F
Cognitive Domain: Knowledge
Answer Location: 2.2.1: Case Identification Numbers
Difficulty Level: Easy

2. Researchers who have stored data collected during a study over time are encouraged to file names like ‘Final’.
Ans: F
Cognitive Domain: Knowledge
Answer Location: 2.2.3: Keeping Track of Files
Difficulty Level: Easy

3. Statistical procedures based on the General Linear Model assume that the scores are independent of each other.
Ans: T
Cognitive Domain: Knowledge
Answer Location: 2.3: Sources of Bias
Difficulty Level: Easy

4. The assumption that variables are normally distributed is more important when a researcher plans to use advanced statistical procedures.
Ans: T
Cognitive Domain: Knowledge
Answer Location: 2.3: Sources of Bias
Difficulty Level: Easy

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

5. Researchers should examine datasets carefully before deciding how to handle outliers.
Ans: F
Cognitive Domain: Knowledge
Answer Location: 2.4: Screening Sample Data
Difficulty Level: Easy

6. If less than 5% of the observations in a dataset are missing, missing data may not be an issue.
Ans: T
Cognitive Domain: Knowledge
Answer Location: 2.4.1: Data Screening Need in All Situations
Difficulty Level: Easy

7. If a researcher plans to compare two groups, outliers should be examined based on the distribution created by combining the groups.
Ans: F
Cognitive Domain: Knowledge
Answer Location: 2.6.1: Univariate Outliers
Difficulty Level: Easy

8. When the univariate distributions of variables include no outliers, there is no reason to test for multivariate outliers.
Ans: F
Cognitive Domain: Knowledge
Answer Location: 2.7.3: Handling Bivariate and Multivariate Outliers
Difficulty Level: Easy

9. Researchers can safely assume that missing values occur randomly.
Ans: F
Cognitive Domain: Knowledge
Answer Location: 2.10.1: Why Missing Values Create Problems
Difficulty Level: Easy

10. Dropping variables with high percentages of missing values after an analysis has been conducted is acceptable.
Ans: F
Cognitive Domain: Knowledge
Answer Location: 2.10.3: Decisions Based on Amount of Missing Data
Difficulty Level: Easy

11. List-wise deletion is considered to be poor practice.
Ans: T
Cognitive Domain: Knowledge
Answer Location: 2.10.3: Decisions Based on Amount of Missing Data | 2.14: Possible Remedies for Missing Data

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

Difficulty Level: Easy

Essay

1. List five objectives of data screening.
Ans: (1) Correct errors, (2) become knowledgeable about aspects of the dataset like the distribution shapes, (3) determine whether assumptions of proposed analyses met, (4) correct assumption violations where possible, and (5) identify and correct problems like outliers, skewness, and missing data.
Cognitive Domain: Knowledge
Answer Location: 2.1: Introduction
Difficulty Level: Easy

2. Give an example of an item on an inventory that uses reverse-wording and explain how it affects data preparation and analysis.
Ans: Examples will vary but should be something like the following item on an inventory designed to measure work-related stress: “My productivity has increased.” If respondents use a 4-point scale to respond to the items in the inventory and higher scores indicate higher stress, responses to this item should be entered (or recoded) so that 1 = 4, 2 = 3, 3 = 2, and 4 = 1. If recoding is used, variable names should clearly distinguish the original responses from the recoded responses.
Cognitive Domain: Application
Answer Location: 2.2.4: Use Different Variable Names to Keep Track of Modifications
Difficulty Level: Medium

3. Define bias and list two possible sources of bias in research data.
Ans: Definition – over-or under-estimation of statistics. Sources – (1) assumptions of analyses violated, (2) outliers present, (3) data are missing for some measures
Cognitive Domain: Knowledge
Answer Location: 2.3: Sources of Bias
Difficulty Level: Easy

4. Describe the possible ways that researchers can handle univariate outliers.
Ans: (1) conduct the analysis with outliers included, (2) discard all outliers, (3) replace all outliers with the next largest score value not considered an outlier, and (4) conduct the analysis with and without outliers and report the results of both analyses.
Cognitive Domain: Knowledge
Answer Location: 2.7.2: Handling Univariate Outliers
Difficulty Level: Easy

5. Describe the effect ignoring missing values on the *N*s variables in a study, and explain its potential effect on readers.
Ans: Ignoring missing values will result in statistics based on different *N*s for different variables, which may make readers wonder why participants did not provide information about some or all variables. In addition, results may not be comparable across analyses

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner

because they are bases on different samples.
Cognitive Domain: Application
Answer Location: 2.10.1: Why Missing Values Create Problems
Difficulty Level: Medium

6. Provide four explanations for why missing data may arise and briefly explain why each may cause problems.
Ans: (1) refusal to participate, which may lead to unrepresentative samples because nonresponse is unlikely to be random; (2) attrition in longitudinal studies may result in smaller, less representative samples at the conclusion of a study; (3) missing data may be planned by asking questions designed to eliminate participants at various points in surveys will intentionally result in smaller samples; as will (4) using missing values to replace outliers.
Cognitive Domain: Comprehension
Answer Location: 2.11: How Missing Data Arise
Difficulty Level: Medium

7. Imagine a group of students is being surveyed to determine their levels of college debt. Briefly describe why type B missingness may occur and describe three methods that might be used to evaluate whether it is occurring.
Ans: Type b missingness might occur because students with higher levels of debt may not respond because they do not want to acknowledge their financial situation. In order to determine whether it is occurring, researchers can (1) follow up with nonresponding students to try to obtain missing data, (2) try to locate other sources of information about these students' debt, or (3) examine the distribution and range of debt in the sample and compare these values to those of students across the university.
Cognitive Domain: Application
Answer Location: 2.12.4: Detection of Type B Missingness
Difficulty Level: Medium

Access Complete Test Bank for Applied Statistics II Multivariable and Multivariate Techniques 3rd Warner