

PSYCH 611 Test #1 (Spring 2020)

INSTRUCTIONS: There are 50 multiple-choice questions, worth 2 points each. Please read all questions and answers carefully! Note: For the “choose the right analysis” questions I’m talking about a *single* analysis that could be done – not a series of the same analysis with different configurations. Additionally, any mention of logistic regression in this test refers to “binary logistic regression” even if not specifically mentioned as such. [Note: As with many exams, my inspiration for the ‘story problems’ comes from recent articles in journals that publish on a range of psychological topics. This test was inspired by recent publications in *Psychological Science*, although I pretty much just used the titles of each article to generate my own ideas.]

1. Gender bias in language is nothing new. However, I’m interested in whether people differ in whether their language in different areas of life is biased (such as talking about politics vs. talking about job performance vs. talking about social media vs. talking about music, etc.). I write a 3-item measure to assess gender bias. I rewrite these items ten times each (30 items total), so that each set of three items assesses gender bias in a different domain. I want to see if people really do differentiate between these sets of items or whether gender bias may instead just be one unitary thing. Which type of technique should I use?
 - a. Dependence
 - b. Interdependence

2. Suppose that someone actually wrote a valid measure of learning styles (note: this is complete fiction). Each of the three common learning styles has is measured with 10 questions each. I want to know whether there are different types of people in terms of how they respond to these questions, so I give a large sample my 30-question measure. Which of the following questions would I need to answer in order to determine what type of analysis to use?
 - a. How many dependent variables are there?
 - b. Am I looking for patterns/clusters of variables or of people?
 - c. Are the dependent variables metric or nonmetric?
 - d. Are the predictor variables metric or nonmetric?

3. Some kids are better than others at delayed gratification. I want to know what accounts for this difference. I measure a bunch of potential predictor variables, and then also record (as my DV) whether or not (yes or no) the child was able to delay gratification on a particular task. In this example, the dependent variable is
 - a. Metric
 - b. Non-metric

4. Does the expectation of getting a better/worse/same outcome affect happiness? I do an experiment where I randomly assign people to three different conditions: expecting a better outcome than I eventually give them, the same outcome I give them, or a worse outcome. I then measure four different variables: happiness, positive affect, excitement, and dejection. How many dependent variables are there in this study?
 - a. 2
 - b. 3
 - c. 4
 - d. This is not the kind of study that has dependent variables

PSYCH 611 Test #1 (Spring 2020)

5. Is religiosity related to violent crime, and if so, what affects that relationship? To find out, I get a sample of 176 countries (each country is a single 'case' in the analysis, so nesting is not a problem) and measure each one's violent crime rate using a continuous scale. I also measure religiosity using a continuous measure. Finally, I think that the relationship might depend on intelligence, so I get a continuous measure of average IQ for each country in the sample. Which of the following would be the most appropriate technique to use?
 - a. Multiple regression
 - b. Logistic regression
 - c. None of the above

6. Speed dating is...well... fast, and misperceptions abound. Sometimes you're accurate about whether the other person liked you, and sometimes you're not. What predicts this accuracy? To find out, I get a sample of 300 speed daters and examined a single encounter for each, recording whether the dater was right (yes/no) about whether the potential date felt some attraction (note: none of the 'potential dates' were part of the sample so nesting is not a problem). To see whether I can predict accuracy, I measure gender, intelligence, and experience dating. Which of the following would be the most appropriate technique to use?
 - a. Multiple regression
 - b. Logistic regression
 - c. None of the above

7. Does linguistic similarity relate to whether friendships are formed? I randomly pair up 400 people and measure the similarity of their linguistic styles using a continuous measure (note that each dyad constitutes a case in my dataset, so you can assume that there is no problem with nesting). I also record other variables such as similarity in height, age, etc. Finally, after having them interact for a while, I record (yes/no) whether or not they agreed that they became friends. Which of the following would be the most appropriate technique to use?
 - a. Multiple regression
 - b. Logistic regression
 - c. None of the above

8. What predicts racial bias among physicians? I get a sample of physicians and measure their racial bias using a continuous scale. I then measure physicians' gender, extent of interracial contact in daily life, and personality, all using continuous measures. Which of the following would be the most appropriate technique to use?
 - a. Multiple regression
 - b. Logistic regression
 - c. None of the above

PSYCH 611 Test #1 (Spring 2020)

9. Assume for a moment that there is no relationship between my predictor variable and my dependent variable. If I ran a single study, what is the probability that I'll incorrectly conclude that there is indeed a relationship?
- Alpha
 - Beta
 - Delta
 - Gamma
10. Cohen's d is a measure of
- Dispersion
 - Central tendency
 - Effect size
 - Variate
11. If I wanted to describe the shape of a frequency distribution, which of the following concepts would be most likely to play a role?
- Residual
 - Missingness
 - Leverage
 - Histogram
12. Suppose I want to know the effects of taking multivariate statistics and social psychology on quality of life. I randomly assign half of my students to take PSY 611 (or not). Furthermore, within each of those groups, I randomly assign half to take PSY 525 (or not). I measure quality of life at the end of the semester. Which of the following would best describe this design?
- Skewed
 - Platykurtic
 - Composite
 - Factorial
13. Suppose I ran a study and found a correlation of $r=1.25$ between extraversion and GPA. Which of the following would be the correct interpretation of this relationship?
- As extraversion increases, GPA increases
 - As extraversion increases, GPA decreases
 - There is no relationship between extraversion and GPA
 - I must have done something wrong to get such a value

PSYCH 611 Test #1 (Spring 2020)

14. Suppose I regressed anxiety on a set of predictors. I got a regression equation and used it to predict the anxiety of a single person based on that person's scores on the predictors. However, although I predicted that this person would be anxious, the person really wasn't. What would you call the difference between my prediction and this person's actual anxiety?
- Skew
 - Tolerance
 - Loglikelihood
 - A residual
15. Picture a regression. If the residuals from the regression are equally spread out from each other when participants are low vs. high vs. in the middle on a predictor variable, this is known as
- Homoscedasticity
 - Heteroscedasticity
 - Positive skew
 - Negative skew
16. Boxplots can be used to look for all of the following **EXCEPT:**
- Outliers
 - Normality
 - Homogeneity of variance (if you have groups)
 - Independence of errors
17. If a distribution looks flatter than a normal distribution should, this is known as
- Positive skew
 - Negative skew
 - Leptokurtosis
 - Platykurtosis
18. Although many would prefer a significance test, a common 'rule-of-thumb' cutoff for skew is an absolute value of
- 1
 - 2
 - 7
 - 10
19. One way of identifying multivariate outliers is using
- Adjusted R^2
 - Mahalanobis D^2
 - R^2D^2
 - BB^8

PSYCH 611 Test #1 (Spring 2020)

20. A z -score is a common way to look for
- Heterogeneity of variance
 - Multicollinearity
 - Independence of errors
 - Univariate outliers
21. If a Q-Q plot shows a curved line, it indicates the violation of which assumption?
- Independence of errors
 - Homogeneity
 - Normality
 - Linearity
22. Squaring, cubing, taking the square root of a variable, or raising the variable to the power of e , are all ways to
- Reduce multicollinearity
 - Remedy violations of the normality or homoscedasticity assumptions
 - Perform hierarchical regression
 - Account for nesting in the dataset
23. [Note: you should not need to do any math for this problem. Read all of the answers carefully] Suppose that the Tolerance for a variable is .049. This would mean that
- The variable exceeds the rule of thumb for multicollinearity
 - The variable has a significant relationship with the dependent variable
 - The variable is positively skewed
 - The variable has at least one outlier
24. Suppose that **last week** I created a dataset with students' grades over the past few years. Chris, one of the students, has missing data. Upon investigating, I find that Chris is missing a "final grade" from PSY 101 from Spring 2020. Based on what you know, this particular missing datum is
- Not ignorable
 - Ignorable
 - Missing Completely at Random
 - Missing at Random
25. Suppose that I'm missing about 10% of the data from a particular variable. However, I cannot find any other variable in my dataset that predicts whether people are missing data or not. From this perspective, my data could be considered
- MCAR
 - MAR
 - MARS
 - TWIX

PSYCH 611 Test #1 (Spring 2020)

26. I'm missing a decent amount of data on variable X . To fix my missing data problem, suppose I took all of the people with complete data and regressed X onto all of my other predictors. I then use the prediction equation from that output to predict the X values for each person who were missing X data. What procedure have I used?
- Hot deck imputation
 - Cold deck imputation
 - Regression imputation
 - Case substitution
27. What is true of semipartial/partial/part correlations in regression?
- They relate to the idea that the b value for each predictor includes only its unique effects on the DV
 - You use them if you've violated the assumption of linearity, but use regular correlations otherwise
 - You use them if you want to include only one predictor in your model
 - They are only relevant in logistic regression – not multiple regression
28. The coefficient of determination is a measure of
- Multicollinearity
 - Generalizability
 - Effect size
 - Factorability
29. Suppose I want to do an experiment on the effects of lighting on mood. I'm afraid that hunger might also affect mood and I don't want the people in my low-light and high-light conditions to differ in hunger, so I randomly assign the people in each of my lighting conditions to either have a meal before the study or to come hungry. The addition of my meal manipulation to this study is an example of
- Experimental control
 - Statistical control
 - Ground control
 - Major Tom
30. In multiple regression, nested models (i.e., hierarchical regressions) involve
- Adding each DV one at a time
 - Comparing the VIF values for each model
 - Making sure that the Adjusted R^2 is above 1.
 - Comparing the R^2 values between models in which predictor variables have been added or deleted
31. In multiple regression, if a predictor has a β (beta) value of 0.75, it means that
- As the predictor increases one point, the DV increases .75 points
 - As the predictor increases one point, the DV decreases .75 points
 - As the predictor increases one SD , the DV increases .75 SD 's
 - As the predictor increases .75 SD 's, the DV increases one SD

PSYCH 611 Test #1 (Spring 2020)

32. A power analysis is better, but which of the following is a common rule of thumb regarding a minimum sample size-to-variable ratio?
- 2
 - 3
 - 5
 - 100
33. Suppose that I have a small sample size and a lot of variables in my study, and I'm concerned about how my prediction equation might function if I applied the equation to a new sample. Specifically, I'm worried that my effect size would be lower in a new sample. What statistic might I look at to get a sense of how bad things might get?
- Tolerance
 - Leverage
 - Adjusted R^2
 - Studentized residuals
34. In multiple regression, the method of entry that allows the computer to either add or delete variables depending on what it thinks works best is known as
- All possible subsets
 - Stepwise
 - Simultaneous
 - Enter
35. In a multiple regression, if I had a theoretically-derived order in which I wanted to order variables (e.g., specific covariates first, followed by a group of variables, and maybe followed by another group of variables), which entry procedure would I use?
- Setwise
 - Stepwise
 - Forward
 - Backward
36. Which of the following would be a sign that there might be a problem with multicollinearity in a multiple regression?
- A VIF value of 1.02
 - A leverage value of .49
 - A condition index value of 150
 - A Mahalanobis D^2 value of 17
37. In multiple regression, the heteroscedasticity assumption is typically evaluated through
- The Mahalanobis D^2 statistic
 - Graphs of residuals against predictors
 - Evaluation of the DFBeta's
 - Hierarchical regression

PSYCH 611 Test #1 (Spring 2020)

38. A person (perhaps an outlier) who had a strong influence on the regression would be denoted by
- A low VIF
 - A high Cook's D
 - A low studentized residual
 - A high R^2
39. In multiple regression, dummy coding is a way to
- Evaluate the normality assumption
 - Transform a skewed variable
 - Test for a curvilinear effect
 - Include a categorical predictor
40. The big difference between multiple regression and logistic regression is
- The homogeneity assumption
 - The potential for outliers
 - The need to evaluate multicollinearity
 - The nature of the DV
41. If the logit for my 'writing an easy test' is -0.45, it means that
- I'm probably going to write an easy test
 - I'm probably going to write a hard test
42. In a logistic regression, a significant Hosmer-Lemeshow test means that
- My first step is better than Model 0 (the null model)
 - I have too much multicollinearity
 - My model fits the data well
 - My model is a poor fit for the data
43. In logistic regression, kappa is used to
- Test for linearity
 - Test for the independence of errors assumption
 - Rescale the DV so that it's continuous
 - Evaluate the effect size of the model
44. Nested model comparisons in logistic regression involve the comparison of
- Leverage values
 - Mahalanobis D^2 values
 - 2LL values
 - Exponentiated coefficients
45. In logistic regression, better fitting models have
- Higher exponentiated coefficients
 - Lower logistic coefficients
 - Lower -2LL values
 - Lower Kappas

PSYCH 611 Test #1 (Spring 2020)

46. In logistic regression, a Nagelkerke value would indicate
- Multicollinearity
 - Direction of a variable's relationship
 - Linearity among the predictors
 - Effect size
47. In logistic regression, a Wald statistic
- Indicates whether a variable has high multicollinearity
 - Indicates whether a variable has a significant relationship with group membership
 - Is a test for outliers
 - Evaluates the homoscedasticity assumption
48. In logistic regression, a variable that has a significantly positive relationship with group membership (i.e., as scores on the variable increase, you're more likely to be in the group given a "1") would have a
- 2LL less than 0
 - Kappa of 2 or higher
 - Positive 'raw' / logit-scale logistic regression coefficient
 - Negative 'raw' / logit-scale logistic regression coefficient
49. If a logistic regression predictor variable had a significant exponentiated coefficient of 0.25, it would mean that
- Higher predictor scores make it more likely that someone is in the group labeled as "1"
 - Lower predictor scores make it more likely that someone is in the group labeled as "1"
 - There is no relationship between the predictor and group membership
 - Something is wrong with the analysis – this is an impossible value
50. Which of the following is true regarding assumptions in multiple regression vs. logistic regression?
- Logistic regression doesn't have any assumptions
 - Logistic regression has completely different assumptions than multiple regression does
 - Aside from the normality and heterogeneity assumptions, the two techniques have very similar assumptions
 - The assumptions are the same but the 'flip sides,' such that the assumptions for the DV in multiple regression are actually the assumptions for the predictors in logistic regression