Instructor: Ernesto F. L. Amaral, Assistant Professor, Department of Sociology
Date: May 7, 2018 (Monday)
Time: 3:30–5:30pm
Percent of final grade: 25%

The student can keep the pages of this exam. Turn in your Scantron testing form to the professor.

Multiple choice questions
Answer the following 75 multiple choice questions for a total of 25 points. Questions 1 to 50 are worth 0.3 points each. Questions 51 to 75 are worth 0.4 points each. Mark your responses on a grey and white 8.5”x11” Scantron testing form. Only No.2 pencils can be used to bubble in answers (not ink).

The correct answers are underlined.

1. In social research the purpose of statistics is to
   a. prove that the research theory is correct.
   b. validate the research project design.
   c. manipulate and analyze data.
   d. ensure acceptance by the scientific community.

2. A hypothesis states, in part, that "income increases as education increases". In this statement, education is
   a. the dependent variable.
   b. the independent variable.
   c. the hypothetical variable.
   d. the secondary variable.

3. The statement "Eight out of ten elderly residents fear victimization" is an example of
   a. univariate descriptive statistics.
   b. multivariate descriptive statistics.
   c. inferential statistics.
   d. inductive statistics.

4. The data reduction process of descriptive statistics
   a. allows a few meaningful numbers to summarize a large amount of data.
   b. eliminates incorrect data.
   c. simply lists all available information in order.
   d. is rarely used.

5. A researcher wants to know if there is a relationship between region of birth and political party affiliation. She should calculate a
   a. univariate descriptive statistic like the mean.
   b. qualitative measure of influence.
   c. measure of association.
   d. statistic that measures the non-relational differentiation between the two variables.

6. Which of the following survey items would generate a discrete variable?
   a. How old are you?
   b. How long does it take you to commute to work?
   c. How much did you pay in taxes last year?
   d. How many cars do you own right now?
7. A researcher has calculated the mean for a variable that is ordinal in level of measurement.
   a. This operation is a violation of level of measurement criterion and the results should be disregarded.
   b. This violation of level of measurement criterion is common and results should be treated with caution.
   c. No violation has occurred, this is a perfectly acceptable application of statistics.
   d. This is a mistake: means should never be calculated for ordinal variables because they are always continuous.

8. To find the ratio of smokers to non-smokers, you would
   a. divide the number of non-smokers by the number of smokers.
   b. add them together and divide by the number of smokers.
   c. divide the number of smokers by the number of non-smokers.
   d. multiply the number of smokers by the total number of people.

9. In Table 1, what percentage of Community A are Republicans?
   a. (103/264) x 100 = 39.02
   b. (103/17) x 100 = 605.88
   c. (264/328) x 100 = 80.49
   d. (103/135) x 100 = 76.30

Table 1. Political party membership in two communities

<table>
<thead>
<tr>
<th>Party</th>
<th>Community A</th>
<th>Community B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Republicans</td>
<td>103</td>
<td>17</td>
<td>120</td>
</tr>
<tr>
<td>Democrats</td>
<td>135</td>
<td>21</td>
<td>156</td>
</tr>
<tr>
<td>Independents</td>
<td>17</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>Socialists</td>
<td>9</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>264</td>
<td>64</td>
<td>328</td>
</tr>
</tbody>
</table>

Source: Fictitious data.

10. To calculate a proportion, the number of cases in any category (f) is divided by
    a. the total number of categories (k).
    b. the number of cases in all categories (N).
    c. the cases in that category (f).
    d. the number of cases in adjacent categories (k-1).

11. The midpoints of intervals for frequency distributions constructed with interval-ratio variables are found by
    a. adding the upper and lower class limits for each interval and then dividing by 2.
    b. multiplying the upper and lower class limits for each interval.
    c. dividing the range by 10.
    d. none of the above. Class intervals for interval-ratio variables do not have midpoints.

12. Cumulative frequencies and cumulative percentages allow a researcher to
    a. be sure the column totals are correct.
    b. tell at a glance how many cases fall above or below a given category.
    c. show the accuracy of his or her findings.
    d. all of the above.

13. A frequency distribution should reflect a balance of
    a. detail and conciseness.
    b. time and money.
    c. questions and answers.
    d. elegance and symmetry.

14. The crude birth rate of a city that has 250 births in a year and a population of 7500 would be found by using which of the following?
    a. crude birth rate = (250/7500) x 1000
    b. crude birth rate = 7500/250
    c. crude birth rate = (1000/7500) x 250
    d. crude birth rate = (250/1000) x 7500
15. A small town of 1,709 residents had one homicide in the past year. The homicide rate for this town
   a. cannot be determined from the information given.
   b. is \((1/1709) \times 100,000\).
   c. is rising.
   d. is \((1709:1) \times 1,000\).

16. When examining a single categorical variable with emphasis on the differences between two or more categories, it is best to use a
   a. histogram.
   b. column chart.
   c. line chart.
   d. none of these choices are correct.

17. Histograms and line charts or frequency polygons are used with data measured at the
   a. nominal level.
   b. ordinal level.
   c. interval-ratio level.
   d. any level.

18. A line chart or frequency polygon is based on
   a. the upper limits of each interval.
   b. the lower limits of each interval.
   c. the midpoint of each interval.
   d. any limit which the researcher selects.

19. The purpose of measures of central tendency is to describe what value of a distribution of scores is
   a. the most typical or representative.
   b. the most surprising or unexpected.
   c. the most significant or important.
   d. all of the above.

20. If the scores of an even number of cases are arranged from high to low, the median is
   a. the middle score.
   b. exactly halfway between the two middle values.
   c. the average of the highest and lowest scores.
   d. the same as the mode.

21. For ordinal level variables, the most appropriate measure of central tendency is generally
   a. the mode.
   b. the median.
   c. the mean.
   d. None of the above

22. If you subtracted the mean from each score in a distribution, squared the differences, and then added the squared differences, the sum would be
   a. zero.
   b. less than zero.
   c. a minimum.
   d. a maximum.

23. A distribution of income for a sample of 45 people that included the presidents of the five largest corporations in the United States and 40 assembly line workers would be
   a. unskewed.
   b. negatively skewed.
   c. positively skewed.
   d. symmetrical.
24. Measures of dispersion provide information about the
a. typical or most common score.
b. variety within the distribution of scores.
c. size of the sample.
d. adequacy of the selection criteria for the sample.

25. One problem with the range (R) as a measure of dispersion is that it
a. is very difficult to calculate.
b. ignores the most extreme scores.
c. can be used only for nominal level variables.
d. is based on only the most extreme scores.

26. Your score on the test is the same as the third quartile (Q3). You may conclude that
a. the distribution of the scores is skewed.
b. you scored higher than 75% of the people who took the test.
c. your score is "typical" since it is the same value as the median.
d. you scored higher than 25% of the people who took the test.

27. The income of a sample has been measured in dollars per year. Which of the following would be the preferred measure of the dispersion for this variable?
   a. The index of qualitative variation.
b. The average deviation.
c. The quartile deviation.
d. The standard deviation.

28. In terms of hypothesis testing, “significance” refers to the
   a. difference between an independent and dependent variable.
b. difference between the sample and population values.
c. difference between the two independent variables.
d. difference between our observed and our predicted outcomes.

29. In order to conduct a test of hypothesis with means or proportions, the sampling distribution must be
   a. skewed.
b. large.
c. bimodal.
d. normal.

30. Which assumption must be true in order to justify the use of hypothesis testing?
   a. Random sampling.
b. Interval-ratio level of measurement.
c. Very large samples.
d. Samples have been stratified.

31. The research and null hypotheses __________ each other.
   a. contradict
   b. complement
   c. amplify
   d. rescind

32. In tests of significance, if the test statistic falls in the critical region, we may conclude that
   a. the population distribution is normal.
b. the null hypothesis can be rejected.
c. the research hypothesis is true.
d. our sample size was too small.
33. A sample of people attending a professional football game averages 13.7 years of formal education while the surrounding community averages 12.1. The difference is significant at the .05 level. What could we conclude?
   a. The null hypothesis should be accepted.
   b. The research hypothesis should be rejected.
   c. The sample is significantly more educated than the community as a whole.
   d. The alpha level is too low.

34. In a one-tailed test of hypothesis, the entire __________ should be placed in either the upper or lower tail of the __________.
   a. critical area, sampling distribution
   b. sample mean, population distribution
   c. Z score, critical area
   d. sampling distribution, sample distribution

35. The central problem in the case of two-sample hypothesis test is to determine
   a. if the samples are random.
   b. if sample statistics are the same as those of the sampling distribution.
   c. if the parameters are representative of the populations.
   d. if two populations differ significantly on the trait in question.

36. When random samples are drawn so that the selection of a case for one sample has no effect on the selection of cases for another sample, the samples are
   a. dependent.
   b. independent.
   c. simple.
   d. systematic.

37. When testing for the significance of the difference between two samples, the null hypothesis states that the __________ are the same.
   a. sample means
   b. population means
   c. sampling distributions
   d. population standard deviations

38. Samples of Republicans and Democrats have been tested for their level of support for welfare reform and the null hypothesis has been rejected. What may we conclude?
   a. The difference is significant, there are differences between the parties on this issue.
   b. The difference is significant, the parties are the same on this issue.
   c. The difference is not significant.
   d. A Type I error has occurred.

39. When testing for the significance of the difference between two sample means, which of the following is almost always unknown?
   a. The population standard deviations.
   b. The sample standard deviations.
   c. Degrees of freedom.
   d. The sample means.

40. A researcher conducted a survey to determine if older people have different feelings about abortion than younger people. He used an alpha level of 0.05 (Z critical = ±1.96) to test for significance and found that his computed test statistic was 2.76. He may conclude that
   a. the difference occurred by random chance.
   b. the difference did not occur by random chance.
   c. the samples are not independent.
   d. the alpha level was too low.
41. When is the t distribution used instead of the Z distribution to test the significance of the difference between two sample means?
   a. Never.
   b. When sample size is very large.
   c. When one sample is much larger than the other.
   d. When N is small.

42. Since critical values of t vary by sample size, before using the t table we must first calculate
   a. the Z score.
   b. degrees of freedom.
   c. the population standard deviation.
   d. the alpha level.

43. ANOVA is appropriate for situations in which
   a. only nominal level variables are involved.
   b. we are comparing more than two samples.
   c. bivariate tables are used to organize the data.
   d. there are fewer than two samples.

44. In the ANOVA test, if the null hypothesis is true
   a. the sample standard deviations should be zero.
   b. the sample means should be roughly equal in value.
   c. the population means should be very different from each other.
   d. alpha should be less than zero.

45. The ANOVA test uses means and standard deviations to compare the amount of variation
   __________ with the amount of variation __________.
   a. within categories, between categories
   b. above categories, below categories
   c. within sample means, between sample means
   d. within sample standard deviations, between sample standard deviations

46. In the ANOVA test, the F ratio is equal to
   a. SST – SSB.
   b. the “mean square between” divided by the “mean square within.”
   c. the “mean square within” minus the “mean square between.”
   d. the total variance divided by the mean square between.

47. A researcher has analyzed differences in average College Board scores for random samples of students from four different colleges. The obtained F score is 0.45. What can be concluded about the null hypothesis?
   a. Nothing, more information is needed.
   b. Reject, the differences did not occur by random chance.
   c. Fail to reject, differences are not statistically significant.
   d. Fail to reject, differences are statistically significant.

48. The chi square test is frequently used because it is relatively easy to satisfy the model assumptions (step 1 of the five-step model). These assumptions require, in the case of chi square,
   a. a normal population distribution.
   b. no assumption about the shape of the sampling distribution.
   c. expected frequencies of equal value.
   d. a difference in expected and observed frequencies.

49. To conduct a chi square test, the variables must first be organized into a
   a. univariate table.
   b. bivariate table.
   c. list, which is then ranked.
   d. frequency distribution.
50. In the context of chi square, variables are independent if
   a. they are related.
   b. cause and effect can be proved.
   c. the obtained chi square falls in the critical region.
   d. the score of a case on one variable has no effect on the score of the case on the other variable.

51. In the chi square test, expected frequencies are computed by
   a. adding the observed frequency to $N$ and subtracting the number of cells.
   b. multiplying the observed frequencies by the row marginals and dividing by degrees of freedom.
   c. multiplying the proper row and column marginals for each cell and dividing by $N$.
   d. adding the proper row and column marginals together and dividing by $N$.

52. In the chi square test for independence, the research hypothesis is
   a. always one-tailed.
   b. that the variables are dependent.
   c. that the variables are independent.
   d. that expected frequencies are greater than observed frequencies.

53. Chi square has been used to test the relationship between gender and support for abortion. The null hypothesis has been rejected. What may be concluded?
   a. Support for abortion depends on gender.
   b. Gender and support for abortion are not related.
   c. Females are more supportive of abortion.
   d. Males are more supportive of abortion.

54. One limitation of the chi square test is that they cannot tell us if relationships between variables are
   a. significant.
   b. random.
   c. by chance.
   d. important.

55. Measures of association __________ the importance of a relationship between variables.
   a. nullify
   b. quantify
   c. prove
   d. maximize

56. Measures of association provide the researcher with information that
   a. indicates the strength of a relationship between variables.
   b. assists him or her in making predictions from one variable to another.
   c. facilitates his or her understanding of causal relationships between variables.
   d. All of the above

57. Conventionally, each column of a bivariate table represents
   a. the highest scores of each variable.
   b. a category of the independent variable (X).
   c. a category of the dependent variable (Y).
   d. the lowest scores on each variable.

58. The distribution of scores on the dependent variable for a specific category of the independent variable is called
   a. associated categories.
   b. the distribution correlation.
   c. the conditional distribution of Y.
   d. the marginal response category.
59. The closer the value of a measure of association is to one,  
a. the stronger the relationship.  
b. the more evenly the scores are distributed.  
c. the weaker the relationship.  
d. the more certain we are that the samples are representative.

60. Phi and Cramer’s V would be appropriate as measures of association for variables measured at the  
a. nominal level.  
b. ordinal level.  
c. interval-ratio level.  
d. continuous level.

61. For variables measured at the ordinal level, gamma measures  
a. the statistical significance of the relationship.  
b. the proportional reduction in error gained by predicting one variable while taking the other into account.  
c. the relative importance of each variable to the association.  
d. All of the above

62. For interval-ratio level variables, the scatterplot is analogous to __________ for nominal and ordinal level variables.  
a. chi square  
b. percentages  
c. row marginals  
d. bivariate tables

63. On a scatterplot, the regression line  
a. is drawn parallel to the horizontal axis.  
b. is always perpendicular to the vertical axis.  
c. comes as close as possible to touching every score.  
d. touches at least half of the scores.

64. The direction of a relationship between variables can be detected on a scatterplot by considering the angle of the regression line. The relationship is  
a. positive when the line slopes upward from left to right.  
b. positive when the line slopes downward from left to right.  
c. negative when the line slopes upward from left to right.  
d. negative when the line slopes downward from right to left.

65. The symbol Y’ (“Y-prime”) stands for  
a. the 1st Y score.  
b. the highest score on Y.  
c. the value of Y when X is zero.  
d. the predicted score on Y.

66. The conditional mean of Y for any value of X is found by  
a. adding the Y scores for the specific value of X and dividing by N.  
b. adding all the Y scores and dividing by X.  
c. subtracting the mean of Y from the overall mean of X.  
d. any of the above.

67. If the slope of a regression line is 1.5, then  
a. for every unit of change in X, there is a change of 1.5 units in Y.  
b. for every unit of change in Y, there is a change of 1.5 in X.  
c. the score of each case is 1.5 times higher on Y than on X.  
d. Y causes X.
68. The slope (b) is an awkward measure of the strength of a relationship because
a. it varies from 0 to 1.00.
b. it decreases in value as the relationship increases in strength.
c. it can have a negative sign when the relationship is actually positive.
d. its upper limit is greater than 1.00.

69. A researcher asked a sample of dual career families about the percentage of the family budget contributed by the wife's job (Y) and the total number of children (X). Pearson's r for this relationship is –0.34. Which of the following is an appropriate interpretation of these results?
a. For every dollar contributed by the wife, the number of children increases by .34.
b. For every additional child, the wife must work longer hours.
c. Every additional child lowers the economic wellbeing of the family.
d. As number of children increase, the percentage of the budget contributed by the wife decreases.

70. One reason to perform multivariate analysis is that this technique provides a way of
a. analyzing the strength and direction of bivariate relationships.
b. proving that variables are causally related.
c. analyzing possible causal relationships among variables.
d. showing that bivariate measures of association are actually unnecessary.

71. With two independent variables, the least-squares multiple regression equation would be
a. \( Y = a + bX_2 \)
b. \( Y = a + b + X_1 + X_2 \)
c. \( Y = b_1X_1 + b_2X_2 \)
d. \( Y = a + b_1X_1 + b_2X_2 \)

72. In the least-squares multiple regression equation, the unstandardized partial slopes show
a. the point where the line crosses the Y axis.
b. the combined effects of all independents on the dependent variable.
c. the amount of change in Y for a unit change in one independent variable while controlling for all other
   independent variables.
d. the residual effects of the independent after controlling for the effect of the dependent variable.

73. A researcher wants to predict score on personal happiness (Y) from score on income (X_1) and
number of friends (X_2). Which of the following would be the proper course of action?
a. Use the standardized partial slopes of X_1 and X_2 on Y.
b. Examine the unstandardized partial slopes of X_1 and X_2 on Y.
c. Compute the partial correlation coefficient for X_1 and X_2 controlling for Y.
d. Examine the partial slope of Y on each independent separately.

74. If the beta-weight of the first independent (X_1) is 0.45 and the beta-weight of the second
independent (X_2) is –0.50, then
a. X_1 has a stronger direct effect than X_2.
b. X_2 has a stronger direct effect than X_1.
c. Y_2 must equal zero.
d. the multiple correlation coefficient must be negative.

75. The coefficient of multiple determination (R^2) shows
a. the effect of the control variable after removing the independent from the equation.
b. essentially the same thing as a partial gamma.
c. the combined effects of all independents on the dependent.
d. the zero-order relationships after controlling for the independents.