CPH 682-001 Fall 2017
Final Examination Dr. Charnigo

Name: _____

Self Assessment. Full credit will be awarded for presumed honest answers.

1. About how many hours did you spend studying for this final examination ? _____

2. Without looking them up now, how many of the three emphasized competencies would you recall ? _____

Multiple Choice Questions. Identify the best answer in the space to the left of a question. You may use the "W" symbol to excuse yourself from up to 10 questions of your choice.

_____ 1. How large should the sample size n be before I can make a 95% confidence interval for a population mean using TINV, if the population is not normal ?

A. $n \geq 2$
B. $n \geq 30$
C. $n \geq 200$
D. I won't use TINV

_____ 2. If I wish to make a 95% confidence interval for a population proportion, how shall I estimate the standard error ?

A. Sample proportion
B. Sample standard deviation divided by square root of sample size
C. Sample proportion divided by square root of sample size
D. None of the above

_____ 3. A positive study is which of the following ?

A. One in which the null hypothesis is accepted.
B. One in which the null hypothesis is rejected.
C. One in which a statistically significant result is obtained.
D. Both B and C are correct.

_____ 4. When does a type I error occur ?

A. H_0 is true and I accept it
B. H_0 is true and I reject it
C. H_0 is false and I accept it
D. H_0 is false and I reject it

_____ 5. Other things being equal, what happens to Type II error probability when the sample size increases ?

A. Increases
B. Remains constant
C. Decreases
D. One cannot say, in general

_____ 6. What does α denote ?

A. Probability of Type I error
B. Significance level
C. Probability of Type II error
D. Both A and B are correct

_____ 7. Suppose a 90% confidence interval for a population mean is 123 to 135. What can be said about a test of the null hypothesis that the population mean is 122 ?

A. Accept H_0 at 10% significance level
B. Reject H_0 at 5% significance level
C. Reject H_0 at 10% significance level
D. Both B and C are correct

_____ 8. What are the possible values for a risk difference ?

A. Any positive or negative number
B. Any positive number
C. Any number between -1 and 1
D. Any number between 0 and 1

_____ 9. Suppose I obtain a p-value of 0.003. At which significance level(s) do I reject H_0 ?

A. 1%
B. 5%
C. 10%
D. All of the above

_____ 10. Suppose I obtain a p-value of 0.003. Of what is this the probability ?

A. Making a Type I error
B. Correctly rejecting the null hypothesis
C. Null hypothesis being true
D. Obtaining data more opposed to H_0 than current data, if H_0 is true and you repeat the study

_____ 11. What sample size is required to perform the chi-square test for association based on data in a two-by-two contingency table ?

A. $n \geq 30$
B. $n \geq 200$
C. Any row total multiplied by any column total divided by the grand total must be at least 5
D. Any entry in the table must be at least 10

_____ 12. Implicitly assumed for the one sample T test in Chapter 9 is which of the following ?

A. $n \geq 30$
B. $n \geq 200$
C. Any row total multiplied by any column total divided by the grand total must be at least 5
D. n is less than 5% of population size

_____ 13. Suppose that someone is performing a T test for a null hypothesis of equality of two population means and is using sample sizes of 50 and 15 respectively. Without any further information, what may you conclude ? (You may assume competence of the person doing the testing.)

A. Populations are approximately normal
B. Populations have a common variance
C. Samples are independent
D. Both A and C are correct

_____ 14. Suppose the p-value is 0.047 and the critical value with $\alpha = 5\%$ is 3.84. If this is for a chi-square test for association, what is the value of the test statistic ?

A. Slightly less than 3.84
B. Slightly more than 3.84
C. Slightly more than -3.84
D. Either A or C is possible

_____ 15. Suppose the critical value with $\alpha = 5\%$ is 3.84. If this is for a chi-square test for association and the test statistic equals 3.75, what is the p-value ?

A. Slightly more than 0.05
B. Slightly less than 0.05
C. Statistically significant
D. Both B and C are correct

_____ 16. When will you conclude that a population is not approximately normal, with respect to some nonnegative quantity such as cost ?

A. Minimum observation is 3 standard deviations below the mean
B. Skew is greater than 0
C. Both A and B warrant such a conclusion
D. Neither A nor B warrants such a conclusion

_____ 17. What can the FTEST function in Excel help you accomplish ?

A. Decide whether population means differ
B. Decide how to test whether population means differ
C. Decide whether population variances differ
D. Both B and C are correct

_____ 18. Which of the following is not generally assumed for linear regression ?

A. Error terms are normally distributed
B. Error terms are independent
C. Error variances are equal
D. Error variances increase with larger values of the predictor variable

_____ 19. If the predictor is positively associated with and accounts for 25% of the variation in the outcome, what is the correlation between the predictor and the outcome ?

A. 0.50
B. 0.25
C. -0.25
D. Either B or C is possible

_____ 20. Upon observing that an estimated line has a slope of 3 and an intercept of 4, a person says that, for every 1-point increase in X, the predicted value of Y will go up by 4 points. Is this statement correct ?

A. Yes
B. No, because the person should say "the predicted value of X" instead of "X"
C. No, because the person should say "the actual value of Y" instead of "the predicted value of Y"
D. No, because the person is confusing the slope and the intercept

_____ 21. Suppose that the regression sum of squares is 50 and that the residual sum of squares is 100. What is R^2 ?

A. 25%
B. 33.3%
C. 50%
D. 66.7%

_____ 22. Suppose that the residual sum of squares is 100 and that the sample size is 27. What is the estimated standard deviation of the error term in the regression model ? (Assume that there is one predictor.)

A. 1.92
B. 2.00
C. 4.00
D. Answer cannot be determined from the information given

_____ 23. Which of the following should you do before estimating a regression line ?

A. Make a scatter plot of the residuals against the predictor variable
B. Make a scatter plot of the outcome variable against the predictor variable
C. Make a normal probability plot of the residuals
D. Both B and C should precede estimation of a regression line

_____ 24. In linear regression with one predictor variable, consider testing the null hypothesis of no association between the outcome variable and the predictor variable. What test statistic could be used ?

A. T
B. F
C. Either T or F
D. Chi-square

_____ 25. Between sum of squares in ANOVA is most nearly analogous to what in regression ?

A. Regression sum of squares
B. Residual sum of squares
C. Total sum of squares
D. Within sum of squares

_____ 26. Consider testing the null hypothesis in ANOVA. What approach could you use ?

A. Inversion of a confidence interval
B. Compare T test statistic to critical value
C. Look at p-value
D. Either B or C is correct

_____ 27. What is generally assumed for ANOVA ?

A. Error terms are not normally distributed
B. Error terms are dependent
C. Error variance is constant
D. All of the above

_____ 28. If the null hypothesis is rejected in ANOVA, what is usually done next ?

A. Conduct post-hoc tests to see which pairs of groups are different
B. Fit a linear regression model with the same data set
C. Nothing, because all groups are similar
D. Determine an appropriate sample size

_____ 29. What is the purpose of a Bonferroni adjustment ?

A. Reduce risk of Type I errors
B. Reduce risk of Type II errors
C. Increase power
D. Compensate for non-normality of data

_____ 30. A Type II error in ANOVA does what ?

A. Concludes that groups are similar when they really are
B. Concludes that groups are similar when they really are not
C. Concludes that groups are not similar when they really are
D. Concludes that groups are not similar when they really are not

Inference concerning a population average. Suppose that a public health nurse is concerned about whether the nutritional needs of children in a population are being met and arranges for a sample of 200 six-year-old boys to have their heights recorded. Suppose, moreover, that the public health nurse is prepared to assert that the nutritional needs are being met if a 95% confidence interval suggests that the population average height is at least 47 inches and to assert that the nutritional needs are not being met if a 95% confidence interval suggests that the population average height is less than 47 inches. If the nurse is unable to make either assertion based on a 95% confidence interval, then she will state that the situation is inconclusive. (This exercise is based on an example in Section 7.3 of the text.)

1. Suppose that the average height in the sample is 45.52 inches and that the standard deviation in the sample is 10.03 inches. Given that TINV(0.05,199) = 1.972 and TINV(0.025,199) = 2.258, find a 95% confidence interval for the population average height.

2. Based on your 95% confidence interval, will the nurse assert that the nutritional needs are being met, assert that they are not being met, or state that the situation is inconclusive ?

You are required to complete both the contingency table exercise and the linear regression exercise.

Contingency table. Below are (fictional) observed and expected values (under a null hypothesis of no association) for the numbers of emergency room visitors with true emergencies versus other causes for their visits, based on whether they arrived during the day or at night. (This exercise is based on an example in Section 8.1 of the text.)

Observed	Day	Night	
Emergency	35	33	68
Other	25	7	32
	60	40	100

Expected	Day	Night	
Emergency	40.8	27.2	68
Other	19.2	12.8	32
	60	40	100

1. Let p_1 := P(Other | Day) and p_2 := P(Other | Night). Obtain a point estimate of the relative risk p_1 / p_2. In words, we are trying to estimate the factor by which the risk of a non-emergency visit is multiplied when the person arrives during the day versus at night. (Note: You may want to re-organize the table of observed values so that it "fits" the formulas you were given.)

2. Obtain a 95% confidence interval for the same.

3. Calculate the chi-square test statistic for testing a null hypothesis of no association between nature of visit and time of arrival.

4. Should the aforementioned null hypothesis be rejected at the 5% significance level ?

Linear regression. Below are results from a (fictional) linear regression analysis of whether Essential Public Health Service 2 scores for local health departments are related to (the natural log of) population size. Please fill in each blank with the correct number. (This exercise is based on an example in my notes for Chapter 11.)

{1} _____

{2} _____

{3} _____

{4} _____

{5} _____

{6} For every doubling of the population size, the predicted score on EPHS 2 increases by _____.

{7} Extra credit: Consult the last slide of my Chapter 1 notes and briefly comment on the appropriateness of linear regression analysis.

_____

{8} Extra credit: What is Essential Public Health Service 2 ?

_____

Regression Statistics	
Multiple R	0.906394771
R Square	{1}
Adjusted R Square	0.799245415
Standard Error	4.734486742
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	825.5770823	825.5771	36.83086	0.0002996
Residual	8	179.3229177	{2}		
Total	{3}	1004.9			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	41.32578766	{5}	8.364438	3.16E-05	29.93262019	52.7189551
Logpop	6.903657656	1.137557446	{4}	0.0003	4.280445481	9.52686983