CPH 682 Fall 2016
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 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 five questions of your choice. **If you complete *both* the contingency table and linear regression exercises, you may use the "W" symbol to excuse yourself from up to 15 questions.**

_____ 1. How large must the sample size n be before I can make a 95% confidence interval for a population mean using TINV, if the population is normal ?

A. $n \geq 2$
B. $n \geq 30$
C. $n \geq 200$
D. I can never use TINV in this case.

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

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

_____ 3. How large must the sample size n be before I can make a 95% confidence interval for a population proportion using TINV ?

A. $n \geq 30$
B. $n \geq 200$
C. $n \geq 5/(p\hat{p}(1-p\hat{p}))$
D. I can never use TINV in this case.

_____ 4. When does a type II 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 power when the sample size increases ?

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

_____ 6. What does α ordinarily denote ?

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

_____ 7. Suppose a 99% 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 1% significance level
B. Reject H_0 at 5% significance level
C. Reject H_0 at 1% significance level
D. Both B and C are correct

_____ 8. What are the possible values for an odds ratio ?

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 0.06. 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 0.06. 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. $n \geq 5$
B. $n \geq 20$
C. Any value in the observed table should be at least 5
D. Any value in the expected table should be at least 5

_____ 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. population size is less than 5% of n
D. n is less than 5% of population size

_____ 13. How may paired samples occur ?

A. Each person in one sample is matched to a similar person in the other sample
B. Before and after scores on the same people
C. All of the above
D. None of the above

_____ 14. Suppose the p-value is 0.053 and the critical value with $\alpha = 5\%$ is 2.10. If this is a T test, what is the value of the test statistic ?

A. Slightly less than 2.10
B. Slightly more than 2.10
C. Slightly more than -2.10
D. Either A or C is possible

_____ 15. Suppose the critical value with $\alpha = 5\%$ is 2.10. If this is a T test and the test statistic equals 2.05, 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 1 standard deviation below the mean
B. Skew is greater than 2
C. Both A and B are correct
D. Neither A nor B is correct

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

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

_____ 18. What is a residual in linear regression ?

A. Predicted value of the outcome variable
B. Difference between actual and predicted values of the outcome variable
C. Predicted value of the error term
D. Both B and C are correct

_____ 19. If the predictor accounts for 64% of the variation in the outcome, what is the correlation between the predictor and the outcome ?

A. 0.64
B. 0.80
C. -0.80
D. Either B or C is possible based on the information given

_____ 20. Upon observing that an estimated line has a slope of 4, a person says that, for every one 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 is confusing the independent and dependent variables
C. No, because the person should say “actual” instead of “predicted”
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 150. What is R^2 ?

A. 25%
B. 33.3%
C. 50%
D. 300%

_____ 22. Suppose that the regression 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 ?

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. Create a normal probability plot of the residuals
B. Make a scatter plot of the outcome variable against the predictor variable
C. Decide whether the assumption of normally distributed errors is tenable
D. Both A and C should precede estimation of a regression line

_____ 24. What is the rationale for the name "analysis of variance" ?

A. We are analyzing whether variances differ from one sample to another
B. We are analyzing whether variances differ from one population to another
C. We are inferring whether means differ from one sample to another, by analyzing between-sample and within-sample variability
D. We are inferring whether means differ from one population to another, by analyzing between-sample and within-sample variability

_____ 25. Within 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. Between sum of squares

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

A. Inversion of a confidence interval
B. Compare F 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 normally distributed
B. Error terms are independent
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 incorrect rejections of true post-hoc null hypotheses
B. Reduce incorrect acceptances of false post-hoc null hypotheses
C. Increase power
D. Compensate for non-normality of data

_____ 30. A Type I 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

Comparing group means. Suppose that "Group 1 Data" represent the fevers for nine persons with the flu, in degrees Fahrenheit above 98.6. For example, the first person has a fever of 2.8 degrees, which corresponds to a body temperature of 101.4 degrees. Suppose that "Group 2 Data" represent the fevers for the same nine persons two hours after taking a fever-reducing medication. For instance, the first person now has a fever of 1.0 degrees, which corresponds to a body temperature of 99.6 degrees. You are interested in comparing the Group 1 mean to the Group 2 mean. Choose only one answer for each of the items below.

1. Will you assume that the data arise from underlying normal distributions ?

[ ] Yes [ ] No, I'll apply the Central Limit Theorem [ ] No, normality is never needed to compare means

2. Which version of the T test will you use for comparing means ?

[ ] Paired [ ] Homoscedastic [ ] Heteroscedastic

3. Do you accept or reject the null hypothesis that the underlying population means are equal ?

[ ] Accept [ ] Reject [ ] Cannot be determined from the information given

4. What is the average reduction in fever, among all of the subjects ?

[ ] 0.9 [ ] 1.0 [ ] Cannot be determined from the information given

Group 1 Data	Group 2 Data	Syntax	Note about Syntax	Result of Syntax
2.8	1.0	FTEST(A10:A18,B10:B18)		0.031
1.3	2.0	TEST(A10:A18,B10:B18,2,1)	Paired	0.046
0.4	0.5	TEST(A10:A18,B10:B18,2,2)	Homoscedastic	0.045
3.2	0.8	TEST(A10:A18,B10:B18,2,3)	Heteroscedastic	0.053
1.5	1.6			
3.4	1.7			
1.9	0.4			
0.6	0.8			
3.9	1.2			

You are required to complete either the contingency table exercise or the linear regression exercise but not both. If you choose to complete both, you can waive 15 multiple choice questions instead of five.

Contingency table. Below are 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.

Observed	Day	Night	
Emergency	32	36	68
Other	28	4	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 odds ratio $\{ p_1 / (1-p_1) \} / \{ p_2 / (1-p_2) \}$. In words, we are trying to estimate the factor by which the odds of a non-emergency visit are 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 ?

5. Is the p-value less than or greater than 0.05 ?

Linear regression. Reconsider the scenario in the comparing group means exercise. The output below pertains to a linear regression model in which Y := fever after medication and X := fever before medication. Please fill in blanks {1} through {6} with the appropriate numbers.

{1} _____

{2} _____

{3} _____

{4} _____

{5} _____

{6} _____

Regression Statistics	
Multiple R	0.204301
R Square	{1}
Adjusted R Square	-0.09516
Standard Error	0.581358
Observations	9

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.103049	0.103049	0.3049	0.598006
Residual	7	{4}	{5}		
Total	{3}	2.468889			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.921756	0.393891	{2}	0.051835	-0.00965	1.853161	-0.00965	1.853161
Before	0.089695	{6}	0.552177	0.598006	-0.29441	0.4738	-0.29441	0.4738