BUS 310, Final Exam A, Part II, Summer 2019 Name _____

Section _____

Instructions: This exam is in two parts: Part I is to be completed partly at home using the materials posted on Blackboard for Part I, and you will answer questions about that work in class below; Part II is to be completed entirely in class using your computer.

1. You may not use cell phones, and you may only access internet resources you are specifically directed to use: You may access your data file for Part I of the exam in Blackboard. You may access the data files posted to Blackboard for the Exam part II.
2. Be sure you are using the data file that matches the exam version you are given.
3. It is a violation of the honor code to communicate with other students in or out of the class during the exam, by any means. Students whose exams show evidence of coordination will be reported.
4. Show all work to support your reasoning. Primarily, this can be done in Excel. Deletion of evidence of your logical process can result in loss of credit. A significant amount of credit goes toward process, reasoning and interpretation.
5. When rounding, do not over-round. In general, do not report dollar amounts beyond the penny. Means should be rounded to one digit more than the original data; standard deviations to two digits more. Do not report fractions rounded to single digit expressions: $\frac{131}{256} \neq \frac{1}{2}$, and do not round decimals or percents to a single digit: $0.57846 \ldots \neq 60\%$ *or* 0.6. Report a minimum of two digits, up to four, unless otherwise specified in the problem.
6. If a problem asks for an explanation, state the solution clearly, then interpret or explain in addition to stating the solution, not in place of. Explanations without solutions, just as solutions without explanations, will not be awarded full credit.

Part I: At Home

This part was completed at home. You can upload the Excel file for Part I to the Part I folder in Blackboard for use during the Exam period. However, this submission will **not** be graded in this location, it must be submitted to the "**to be graded** folder" to receive credit.

Part II: In Class

1. Use the work done at home to answer the Part I questions.
2. Open the file from the in-class portion of the final posted on Blackboard that corresponds to the version of the exam you have. This is Exam A.
3. Answer the questions corresponding to the data file, and any additional calculation in Excel required. Be sure to sign the honor code statement on the next page.
4. When you have finished answering questions on the exam, and all your answers have been recorded on the paper test for grading, upload **both** the take home Excel file **and** the in-class Excel file to the same in-class Exam folder in Blackboard for grading. Only those files submitted to the Submission/To-Be-Graded Folder will be graded. (If in doubt, put all work in one Excel file.)
5. Turn in your paper copy of the exam to your instructor.

Honor Code Statement:

I, _____ (print your name), agree to abide by the George Mason Honor Code and Academic Integrity Pledge: *To promote a stronger sense of mutual responsibility, respect, trust, and fairness among all members of the George Mason University Community and with the desire for greater academic and personal achievement, I, a student member of the university community, pledge not to cheat, plagiarize, steal, or lie in matters related to academic work.* Furthermore, I have read and I agree to follow the guidelines laid out in the instructions for this exam above. I also agree not to participate in the efforts of other students to circumvent these guidelines, or to assist in their violations of the code, and will report such efforts in a timely manner.

_____ _____

Student Signature and G# Today's Date

Part I:

The following questions refer to problem #1 from Part I:

1. Is the model of units vs. Labor Hours linear or non-linear? Explain. Use the residual graphs in your explanation, and a discussion of the long-term trend in your explanation. [Hint: is there a point where the models may predict labor hours values that make no sense? Note at least three factors in your decision.] (10 points)

nonlinear since log model explains 30% more of variability than linear model

2. What is the equation and R^2 value of the model that best fits the data? (8 points)

$Y = -4.03 \ln X + 20.148$

$R^2 = 0.76$

3. State the null and alternative hypotheses for a multiple regression analysis, for the full model. (6 points)

H_0: all $\beta_i = 0$ or $\rho = 0$ $\Leftrightarrow$ no relationship between units & hours
H_a: all $\beta_i \neq 0$ or $\rho \neq 0$
$F = 151.966$ P-value $= 1.75 \times 10^{-16} < 0.05$ reject null

The following questions refer to problem #2 from Part I:

4. Based on your analysis of the selling price of homes in the data set, which variables appear to have a negligible effect on the price? Explain your reasoning. (6 points)

rooms, attached garage, age
all removed from the final model
not able to establish coeffs. were non-zero

5. Give the final regression equation produced from your analysis along with the R^2 value. (8 points)

$Y = 84.01 X$

$R^2 = 0.9916$

6. Based on your best equation, interpret the slope coefficient of the size variable in context. (6 points)

for each additional square foot of house size, the selling price goes up by an average of $84

7. Interpret the R^2 value obtained in context. (6 points)

99% of the variability in selling price can be explained by the relationship to size of house

8. Interpret the standard error of the test. (4 points)

The average distance from an observation to the predicted value on the regression line is $19,833.

9. Calculate a 95% prediction interval for the rating of a single employ with 3100 square feet (size), 8 rooms, no attached garage and 12 years old. (10 points)

($217,890 , $302,966)

The following questions refer to problem #3 from Part I:

10. For the data on property taxes by neighborhood, state the null and alternative hypotheses for this test, along with the test-statistic and P-value. What is the result of the test in context? (10 points)

H_0: all means the same
H_a: at least one mean is different
$F = 107.37$
P-value: $7.29 \times 10^{-50} < 0.05$ reject null
at least one neighborhood pays a different amount of property tax than others

11. Are all the assumptions of the ANOVA test satisfied? Explain. (5 points)

doubtful
Since N1 and N6 are more than 4 times the
balance of N4

The following questions refer to problem #1 from Part I:

12. How many of each type of hammock should Treetop Hammocks make with all constraints included? (6 points)

150 double, 200 single

13. State the maximum profit obtained, and the objective function (an algebraic equation) used to obtain that value. (8 points)

$65x + 55y = \text{Profit} = \$20,750$

14. State the value of the shadow price for ~~Labor Hours~~ Production Time and interpret it in context. (6 points)

20.3125
for each additional unit of production time
available, profit increases $20.31

15. Describe what happened when you removed the constraint on single hammocks. (6 points)

Double went to 0, and we should make
400 singles for max profit of $22,000

Calculations in Excel: (1) 20 points, (2) 40 points, (3) 25 points, (4) 25 points.

Part II:

16. Use the data provided on Cholesterol levels and exercise to conduct a two-sample T-test to determine if exercise reduces cholesterol levels. State the null and alternative hypothesis clearly. Is there enough evidence to support the conclusion that exercise reduces cholesterol? Is the test dependent or independent? (10 points)

independent

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 < \mu_2$

$T = -1.952$

p-value = 0.0259 $<$ .05 reject null

yes, there is enough evidence to support claim that exercise reduces cholesterol

17. Explain how poorly worded questions in a survey can contribute to non-sampling bias. Provide a specific example to illustrate your point. (5 points)

questions may be leading and bias the data or subjects may not understand them and so you may not collect the data you think you are

examples will vary

18. A car dealership wants to determine if their sales staff is discriminating against their female customers. The owner sends in husband and wife pairs shopping for a car who are purchasing the car jointly. The owner sends the wife to one salesperson, and the husband to the same staff member on a different day with the same financial information and vehicle of interest. The owner collects the day and compares the price and financing offers made to each couple. Is the data the owner collected dependent or independent? (5 points)

dependent

19. Explain the procedures for dealing with an outlier. (6 points)

identify them. determine if they are reasonably likely to occur given the size of the data set. Copy the data & remove the outliers. Reanalyze the data to determine the impact of removal on the model. If possible, explain the source of the excess variability. Then given all the analysis, determine if its reasonable to omit the suspect values (some or all).

20. The data file includes data on the proportion of employees for a particular company who exercised before a health and fitness center was installed in the office building, and afterwards. The company wants to determine if installing the fitness center changed the likelihood that employees were to exercise. Conduct a test of proportions, using the proportion from the "Before" condition as the null hypothesis for the "After" condition. What can you conclude? (10 points)

$H_0: p = 32\%$

$H_a: p \neq 32\%$ vs. $p > 32\%$

Z-stat: 1.715

P-value (2 tailed) = 0.086... > 0.05 fail to reject null

one-tailed = 0.043 < 0.05 reject null

if we expect the program to increase exercise, we have sufficient; if we are looking change, then we don't

21. The data file contains data on the lifetime hours of batteries. Calculate a confidence interval for both sets of batteries. (8 points)

(99.5, 100.65) Battery 1

(98.3, 100.81) Battery 2

22. Based on the calculated confidence intervals, what conclusion can you come to about how the lifetimes of the batteries compare? (6 points)

they are basically the same since there is substantial overlap

23. Create a pivot table of Day of the Week vs. Time of Day for a sample of shoppers at a particular store. Conduct an appropriate test to determine if Time of Day is independent of Day of the Week. State the null and alternative hypothesis, test statistic and/or P-value, and the conclusions of the test in context. (10 points)

H_0: variables are independent

H_a: variables are dependent

$\chi^2 = 9.787$

P-value = 0.6346 > 0.05 fail to reject null

we have insufficient evidence to conclude that Day of Week & Time of Day are dependent

Upload your completed Excel files to the Exam #2 submission box in Blackboard and submit your completed paper exam to your instructor. You may not modify anything once the exam is submitted.

Standard errors: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ $s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

$s_{x_1-x_2} = s_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

Sample sizes: $n > \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$ $n > \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$ $m = n = \frac{4z_{\alpha/2}^2(\sigma_1^2+\sigma_2^2)}{w^2}$

Confidence intervals:

One sample: $\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Two samples (independent): $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,n-1}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ $(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Test statistics:

One sample: $z \text{ or } t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

Two samples: dependent: $z \text{ or } t = \frac{\bar{d}_0-\delta}{\frac{s_d}{\sqrt{n}}}$

Independent: $z \text{ or } t = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ $z = \frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}$

Degrees of freedom (two samples, unpooled) $\nu = \frac{\left(\frac{s_1^2}{m}+\frac{s_2^2}{n}\right)^2}{\frac{\left(\frac{s_1^2}{m}\right)^2}{m-1}+\frac{\left(\frac{s_2^2}{n}\right)^2}{n-1}}$

χ^2Tests: $\chi^2 = \sum_{all\ cells}\frac{(obs-\exp)^2}{exp}$