Final B

Summer 2019C

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 R^2 is higher by 0.30
better fits data
residual plots have fewer extreme outliers & patterns

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

$R^2 = 0.76$
$y = -4.03 \ln x + 20.148$

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 btwn units & hours
H_a: all $\beta_i \neq 0$ or $\rho \neq 0$
$F = 151.966$ $P\text{-value} = 1.75 \times 10^{-16} < 0.05$ 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 coefficients were non-zero

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

$y = 84.01x$
$R^2 = 0.9916$

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

For each additional square foot, the selling price goes up by $84 on average.

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

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

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

the average distance of 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 1200 square feet (size), 5 rooms, an attached garage and 32 years old. (10 points)

($58,273, $143,349)

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 $= 7.29 \times 10^{-30} \ll 0.05$ 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 variance 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 Total (hammocks) and interpret it in context. (6 points)

0
the constraint is not satisfied so adjusting the level won't change the profit

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 $20,000

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)

$H_0: \mu_1 = \mu_2$ independent

$H_a: \mu_1 < \mu_2$

$T = -1.952$

p-value $= 0.0259 < 0.05$ 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. If possible, explain the source of the excess variability. Then given all the analysis, determine if its reasonable to omit the 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)

$H_0: p = 32\%$

$H_a: p \neq 32\%$ or $p > 32\%$

Z-stat: 1.715

P-value =

2-tailed = 0.086 > 0.05 fail to reject

one tailed = 0.043 < 0.05 reject null

if we expect the program to increase exercise, then there is sufficient evidence; if we test just change, then we do not

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 they are substantially overlapping

23. Perform a χ^2 test of independence to determine if Score is depending on Upper class standing. You will need to recode the scores as letter grades (<60=F, 60-60=D, 70-79=C, 80-89=B, 90-100=A). Use a 10% significance level. Clearly state your hypotheses and all test-statistics or P-values. Interpret your results in a sentence that includes the context of the test. (10 points)

H_0: variables are independent

H_a: variables are dependent

$\chi^2 = 1.39$ P-value = 0.845 > 0.05 fail to reject

letter does not affect upper class standing