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?] (10 points)

nonlinear

residuals of linear shows a pattern (not random)
log model much better R^2.

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

$y = -4.037 \ln x + 20.148$

$R^2 = 0.76$

3. To what extent can you improve the model by removing the first value (a possible outlier)? Explain. (6 points)

The linear model goes from $R^2 = .465$ to .581
The log model goes from $R^2 = .76$ to .754.
removing first point improves linear fit and reduces log fit.

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, age, attached garage

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

$y = 84.01x$ (setting intercept to zero makes $R^2 = .99$)

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

for each additional square foot of home, the price increases by $84.01 on average.

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

99% of the variability in selling price can be explained by the change in size.

8. 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? (12 points)

$H_0: \mu_i = \mu_j$ for all i, j (all means equal)
$H_a: \mu_i \neq \mu_j$ for at least one $i \neq j$

$F = 107.366$ p-value $= 7.288 \times 10^{-50}$ reject H_0

There is good reason to think the means are not the same

9. Are all the assumptions of the ANOVA test satisfied? Explain. (6 points)

Not really. one of the variances is much larger than the smallest one (by a factor of almost 10). The assumption would be better satisfied if we removed neighborhood 4 (or 5).

10. Using the information provided on the manufacture of chairs and tables, what is the maximum revenue the company can produce under these constraints? (6 points)

1,460,000

11. What production levels of chairs and tables will the company need to produce to obtain the maximum revenue? (6 points)

3000 oak chairs
700 pine tables

12. Describe the sensitivity of the model to modifying the amount of oak available (between 10,000 and 20,000 board feet). At what point does the production model substantially change? Explain. (12 points)

it does not change much in this range. Each additional 1,000 board feet changes the profit by 60,000. The relationship is stable and linear.

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

Part II:

13. 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? (15 points)

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 < \mu_2$

$t = -1.95$
p-value (one-tailed) $= 0.0259 < .05$

Yes, there is enough evidence to support the conclusion that exercise reduces cholesterol.

independent (sample size are not the same so it cannot be dependent)

14. 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? (15 points)

$H_0: p = .32$
$H_a: p \neq .32$

$$z = \frac{.4 - .32}{\sqrt{\frac{.32(1-.32)}{100}}} = 1.714986$$

p-value = 0.086348

fail to reject

This is not strong evidence that behavior has changed.

15. The data file contains data on the lifetime hours of batteries. Calculate a confidence interval for both sets of batteries. (6 points)

#1 (99.53, 100.65)

#2 (98.31, 100.81)

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

The variance of battery 2 is larger, but since the entire battery #1 interval is contained inside battery #2 interval, we cannot conclude the means are different.

17. Create a pivot table of Region ~~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 Region. State the null and alternative hypothesis, test statistic and/or P-value, and the conclusions of the test in context. (15 points)

H_0: Region and Time of Day are Independent

H_a: They are not independent

P-value for χ^2-test is 0.7435 fail to reject H_0

the two variables cannot be shown to be dependent

18. Explain the meaning of a Type I and Type II error on the context of the test in #17. (6 points)

A Type I error occurs when we think Region and Time of Day are dependent, but they are independent.

A Type II error occurs when we think Region and Time of Day are independent, but they are dependent.

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