MTH 325, Exam #1, Spring 2024 Name _____

Instructions: Answer each question thoroughly. For questions in Part 1, use the work you did at home to answer the questions. Be sure to answer each part of each question. In Part 2, report exact answers unless directed to round.

Part I:
Use the work you did at home to answer these questions about tax paid and the neighborhoods in our dataset.

1. Based on your correlation table, identify the variable that has the highest correlation with Salary. What is the correlation value?

2. Based on your correlation table (or graphs), which variables (other than Salary) appear to have potential collinearity problems?

3. What is the simple linear regression equation your found relating Number Supervised to Annual Salary?

4. Interpret the slope in the context of the problem.

5. What percent of the variability in Salary can be explained by the relationship with Number Supervised?

6. Compare your machine found model with your final backwards selection model. Describe any differences in your models (variables included), any errors generated in the selection, etc.

7. Answer this question and the remaining questions in Part 1 using the **backward selection model** you found by hand. Write the equation of your model that describes your multiple regression model.

8. Construct a prediction interval for an employee with gender 1, 4 years of education, 15 years of previous experience, 15 years employed, department 3 and supervises 5 people.

9. Interpret the meaning of the Department coefficient in the context of the problem.

10. Construct a confidence interval for the Department variable coefficient.

11. Test your model assumptions using your residual plots and other diagnostic plots. Do they appear to be approximately satisfied? Identify any potential outliers.

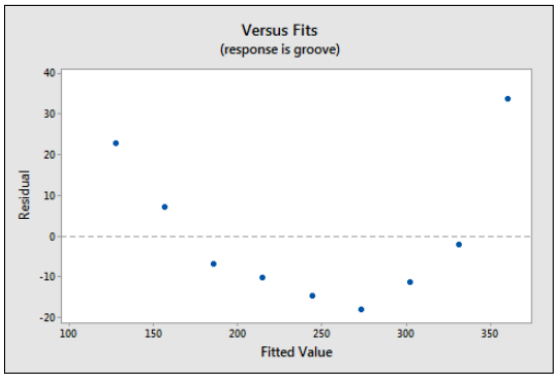
12. Based on your best model, interpret the meaning of the R^2 value.

13. Are there any potential problems with treating department like a numerical variable in this context? Explain.

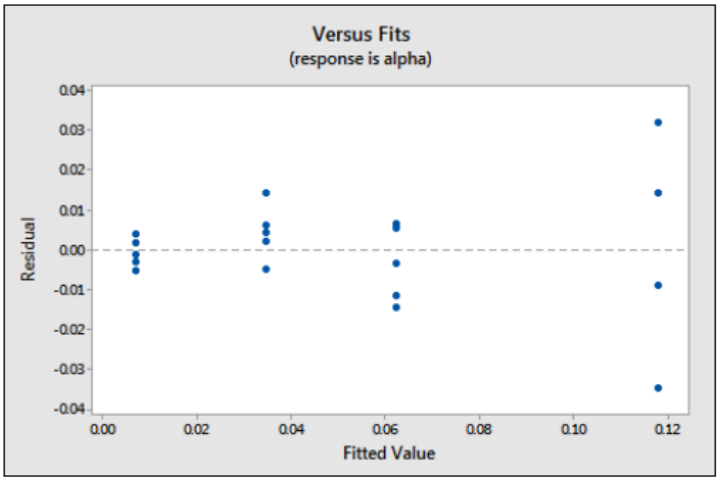
Part II:

14. Examine the residual plots below. Identify any problems associated with each plot in terms of potential problems for the standard assumptions made for a linear regression model.

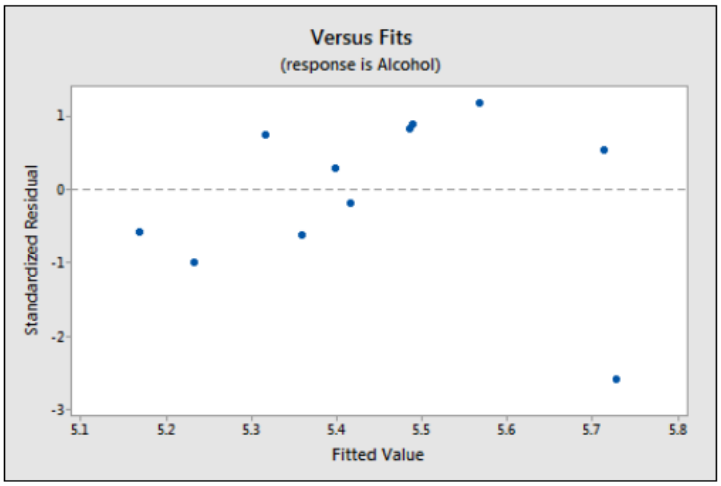
a.



b.



c.



15. State the null and alternative hypothesis for a multiple regression model (for the full model and any other tests conducted to assess model quality).

16. Recall that $Cov(X,Y) = E(XY) - E(X)E(Y)$. For the probability density function $f(x,y) = \frac{5}{384}x^4\left(x+\sqrt{y}\right), x \in [0,2], y \in [0,4]$, find the covariance.

17. Consider the small data set $\{(10,1),(8,3),(4,7)\}$. Find the value of the regression coefficients for $y = \beta_0 + \beta_1 x$, using the normal equation $(A^T A)^{-1}A^T Y = B$. Write the coefficients you find in the equation.