STAT 415/615 **Exam 1** **Name:**

Part 1 (20 points): Concept problems

True or False. Justify your answer.

1. The sum of the residuals is equal to zero.

#Answer:

2. A significant positive correlation between X and Y implies that changes in X cause Y to change.

#Answer:

3. The residual is the difference between the observed value of the dependent variable and the predicted value of the dependent variable. In mathematical notation this is given by $Y - E\{Y\}$.

#Answer:

4. If MSR and MSE are of the same order of magnitude, this would suggest that $\beta_1 \neq 0$

#Answer:

5. When using simple regression analysis, if there is a strong correlation between the independent and dependent variable, then we can conclude that an increase in the value of the independent variable causes an increase in the value of the dependent variable.

#Answer:

6. The least squares regression line minimizes the sum of the squared differences between actual and predicted Y values

#Answer:

7. The correlation coefficient takes values between 0 and 1.

#Answer:

8. The coefficient of determination is interpreted as the proportion of observed variation in X that can be explained by the simple linear regression model.

#Answer:

9. One way to study the normality of the error is by histograms.

#Answer:

10. Draw a fitted versus residuals plot where we see that the constant variance assumption is not met and the linearity assumption is not violated.

Part 2 (80 points): Exercises

1. (Use R for data analysis)

The 1974 Motor Trend US magazine contained data on fuel consumption of 32 automobiles (1973-74 models). These data are in dataset *''mtcars''* which is already loaded in R. You can look at it with commands *attach(mtcars), names(mtcars), summary(mtcars), mtcars.* Your task is to study the effect of the number of carburetors (variable *carb*) on the fuel consumption in miles per gallon (variable *mpg*).

(a) Fit a linear regression model that can be used to predict miles per gallon based on the number of carburetors. Is the number of carburetors significant in this prediction? Report the estimated regression equation, the p-value testing significance of carburetors, and state your conclusion.

(b) Conduct a lack-of-fit test to decide whether the relation between the fuel consumption and the number of carburetors is linear. State the test statistic, the p-value, and your conclusion. What does this test statistic measure?

(c) Are there any outliers in this regression analysis? Test each residual keeping the **familywise error rate** at a 5% level. Explain how you did the test, report the numbers that lead to your conclusion.

2. (Use R for data analysis)

   The purpose of this experiment was to assess the influence of calcium in solution on the contraction of heart muscle in rats. The left auricle of 21 rat hearts was isolated and on several occasions a constant-length strip of tissue was electrically stimulated and dipped into various concentrations of calcium chloride solution, after which the shortening of the strip was accurately measured as the response.

   The data are stored in R package MASS. You can look at them with commands attach(muscle), names(muscle), summary(muscle), muscle. A linear regression model is used to predict the change in length of the strip (variable Length, in mm) based on the concentration of calcium chloride solution (variable Conc, in multiples of 2.2 mM).

(a) Calculate the equation of the sample regression line that predicts Length based on Conc.

(b) Complete the ANOVA table and estimate the variance of Length.

(c) Compute a 95% confidence interval for the regression slope β_1

(d) Test whether the slope is zero or not.

(e) Calculate the percent of total variation explained by this regression model.

(f) Compute a 90% confidence interval for the mean Length when the concentration of calcium is 2.5.

(g) Compute a 90% prediction interval for Length if the concentration of calcium is 2.5.

(h) Verify the standard regression assumptions - normality and homoscedasticity. Report p-values and state your conclusions.

(i) **(Graduate only)** Find the optimal Box-Cox transformation. Does it improve normality of residuals?

(j) **(Graduate only)** Test the model for the lack of fit.

3. (By hand: show all steps)
   A sample of size n = 100 contains two variables, X and Y . Sample statistics are: X_bar = 50, Y_bar = 10, S_X = 10, S_Y = 4, r_XY = 0.2.

(a) Calculate the equation of the sample regression line that predicts Y based on X.
(b) Complete the ANOVA table and estimate the variance of Y .
Include sum of squares, degrees of freedom, mean squares and the ANOVA F-statistic.
(c) Compute a 95% confidence interval for the regression slope β_1.
(d) Test whether the slope is zero or not.
(e) Calculate the percent of total variation explained by this regression model.
(f) Compute a 90% confidence interval for the mean response when $X = 35$.
(g) Compute a 90% prediction interval for the response Y_0 if the corresponding independent variable is $X_0 = 35$