Name: _____ ID #: _____

Q1: (5 pts.) If $\tanh^{-1} r \sim N\left(\tanh^{-1} \rho, \frac{1}{n-3}\right)$ where r is the sample correlation coefficient between X and Y, ρ is the population correlation and n is the sample size, then mathematically derive a

 $100(1 - \alpha)\%$ confidence interval for the unknown parameter ρ .

<u>STAT 510</u> <u>Regression Analysis</u> Q2: (6 pts.) An analyst fits model 1: $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_2 X_{3i} + \beta_2 X_{4i} + \epsilon_i$ to a sample of size *n* and obtains the fitted values $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i}$ through the method of ordinary least squares (OLS). The residuals of model 1 are defined as $e_i = (y_i - \hat{y}_i)$ and they have the following properties:

 $\sum e_{i} = \sum X_{1i} e_{i} = \sum X_{2i} e_{i} = \sum X_{3i} e_{i} = \sum X_{4i} e_{i} = \sum \hat{y}_{i} e_{i} = 0.$

After obtaining the fitted values (\hat{y}_i) from model 1, the analyst fits model 2: $y_i = a + b\hat{y}_i + \varepsilon_i$. Show that the OLS estimates for the unknown parameters of model 2 are given as: $\hat{a} = 0$ and $\hat{b} = 1$.

<u>STAT 510</u> Regression Analysis Q3: (6 pts.) For a multiple linear regression model $y_{n \times 1} = X_{n \times (k+1)} \beta_{(K+1) \times 1} + \epsilon_{n \times 1}$, we define the residuals $e = (y - \hat{y})$ obtained through the method of OLS. Mathematically prove that X'e = 0 and from there show that the sum of residuals is zero.

Note: If needed, you can use $\hat{\beta} = (X'X)^{-1}X'y$ without proving it.

Download the dataset from Blackboard and write down the code number in above blank.

A commercial real estate company evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data below are taken from 50 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. Shown in the data file are the age (X_1) , operating expenses and taxes (X_2) , vacancy rates (X_3) , total square footage (X_4) , and rental rates (Y).

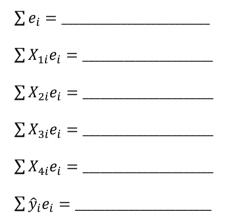
(a) (2 pts.) The Pearson's correlation coefficient between the y and X_2 is denoted by ρ_{y,X_2} . We are 99% confidence that the unknown quantity ρ_{y,X_2} is between:

[______ and _____]

(b) (1 pt.) Fit a regression line for predicting rental rates based on all 4 predictors. The fitted regression equation is

 $\hat{y} = ___X_1 + __X_2 + __X_3 + __X_4$

(c) (1 pt.) Save the fitted values (\hat{y}) and the residuals (e_i) and find the following sums:



(d) (2 pts.) Using the fitted values from part (b), fit another regression as $y_i = \gamma_0 + \gamma_1 \hat{y}_i + \delta_i$. Find the coefficient of determination for this model i.e. $R^2 = 1 - \frac{SSE}{SST}$. Compare this with the R^2 of full model.

 $R_{part(b)}^2 =$ _____ $R_{part(d)}^2 =$ _____

Test Statistic = _____

P-values = _____

Conclusion:

Download the dataset from Blackboard and write down the code number in above blank.

A commercial real estate company evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data below are taken from 50 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. Shown in the data file are the age (X_1) , operating expenses and taxes (X_2) , vacancy rates (X_3) , total square footage (X_4) , and rental rates (Y).

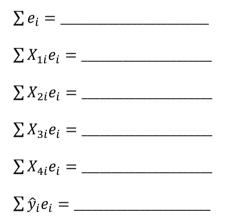
(a) (2 pts.) The Pearson's correlation coefficient between the y and X_2 is denoted by ρ_{y,X_2} . We are 99% confidence that the unknown quantity ρ_{y,X_2} is between:

[______ and _____]

(b) (1 pt.) Fit a regression line for predicting rental rates based on all 4 predictors. The fitted regression equation is

 $\hat{y} = ___X_1 + __X_2 + __X_3 + __X_4$

(c) (1 pt.) Save the fitted values (\hat{y}) and the residuals (e_i) and find the following sums:



(d) (2 pts.) Using the fitted values from part (b), fit another regression as $y_i = \gamma_0 + \gamma_1 \hat{y}_i + \delta_i$. Find the coefficient of determination for this model i.e. $R^2 = 1 - \frac{SSE}{SST}$. Compare this with the R^2 of full model.

 $R_{part(b)}^2 =$ _____ $R_{part(d)}^2 =$ _____

Test Statistic = _____

P-values = _____

Conclusion:

Download the dataset from Blackboard and write down the code number in above blank.

A commercial real estate company evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data below are taken from 50 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. Shown in the data file are the age (X_1) , operating expenses and taxes (X_2) , vacancy rates (X_3) , total square footage (X_4) , and rental rates (Y).

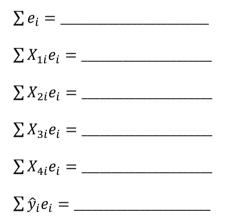
(a) (2 pts.) The Pearson's correlation coefficient between the y and X_2 is denoted by ρ_{y,X_2} . We are 99% confidence that the unknown quantity ρ_{y,X_2} is between:

[______ and _____]

(b) (1 pt.) Fit a regression line for predicting rental rates based on all 4 predictors. The fitted regression equation is

 $\hat{y} = ___X_1 + __X_2 + __X_3 + __X_4$

(c) (1 pt.) Save the fitted values (\hat{y}) and the residuals (e_i) and find the following sums:



(d) (2 pts.) Using the fitted values from part (b), fit another regression as $y_i = \gamma_0 + \gamma_1 \hat{y}_i + \delta_i$. Find the coefficient of determination for this model i.e. $R^2 = 1 - \frac{SSE}{SST}$. Compare this with the R^2 of full model.

 $R_{part(b)}^2 =$ _____ $R_{part(d)}^2 =$ _____

Test Statistic = _____

P-values = _____

Conclusion:

Download the dataset from Blackboard and write down the code number in above blank.

A commercial real estate company evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data below are taken from 50 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. Shown in the data file are the age (X_1) , operating expenses and taxes (X_2) , vacancy rates (X_3) , total square footage (X_4) , and rental rates (Y).

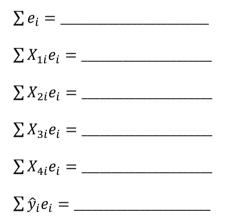
(a) (2 pts.) The Pearson's correlation coefficient between the y and X_2 is denoted by ρ_{y,X_2} . We are 99% confidence that the unknown quantity ρ_{y,X_2} is between:

[______ and _____]

(b) (1 pt.) Fit a regression line for predicting rental rates based on all 4 predictors. The fitted regression equation is

 $\hat{y} = ___X_1 + __X_2 + __X_3 + __X_4$

(c) (1 pt.) Save the fitted values (\hat{y}) and the residuals (e_i) and find the following sums:



(d) (2 pts.) Using the fitted values from part (b), fit another regression as $y_i = \gamma_0 + \gamma_1 \hat{y}_i + \delta_i$. Find the coefficient of determination for this model i.e. $R^2 = 1 - \frac{SSE}{SST}$. Compare this with the R^2 of full model.

 $R_{part(b)}^2 =$ _____ $R_{part(d)}^2 =$ _____

Test Statistic = _____

P-values = _____

Conclusion:

Download the dataset from Blackboard and write down the code number in above blank.

A commercial real estate company evaluates vacancy rates, square footage, rental rates, and operating expenses for commercial properties in a large metropolitan area in order to provide clients with quantitative information upon which to make rental decisions. The data below are taken from 50 suburban commercial properties that are the newest, best located, most attractive, and expensive for five specific geographic areas. Shown in the data file are the age (X_1) , operating expenses and taxes (X_2) , vacancy rates (X_3) , total square footage (X_4) , and rental rates (Y).

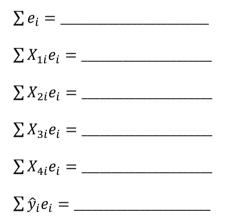
(a) (2 pts.) The Pearson's correlation coefficient between the y and X_2 is denoted by ρ_{y,X_2} . We are 99% confidence that the unknown quantity ρ_{y,X_2} is between:

[______ and _____]

(b) (1 pt.) Fit a regression line for predicting rental rates based on all 4 predictors. The fitted regression equation is

 $\hat{y} = ___X_1 + __X_2 + __X_3 + __X_4$

(c) (1 pt.) Save the fitted values (\hat{y}) and the residuals (e_i) and find the following sums:



(d) (2 pts.) Using the fitted values from part (b), fit another regression as $y_i = \gamma_0 + \gamma_1 \hat{y}_i + \delta_i$. Find the coefficient of determination for this model i.e. $R^2 = 1 - \frac{SSE}{SST}$. Compare this with the R^2 of full model.

 $R_{part(b)}^2 =$ _____ $R_{part(d)}^2 =$ _____

Test Statistic = _____

P-values = _____

Conclusion: