Name: ________________________________ ID #: ____________________

Q1: (5 pts.) If $\tanh^{-1} r \overset{appox.}{\sim} N\left(\tanh^{-1}\rho, \frac{1}{n-3}\right)$ 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.

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 $\hat{b} = 1$.

Q3: (6 pts.) For a multiple linear regression model $\boldsymbol{y}_{n\times 1} = \boldsymbol{X}_{n\times(k+1)}\boldsymbol{\beta}_{(K+1)\times 1} + \boldsymbol{\epsilon}_{n\times 1}$, we define the residuals $\boldsymbol{e} = (\boldsymbol{y} - \widehat{\boldsymbol{y}})$ obtained through the method of OLS. Mathematically prove that $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$ and from there show that the sum of residuals is zero.

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

Q4: Code ______

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 $\rho_{y.X_2}$. We are 99% confidence that the unknown quantity $\rho_{y.X_2}$ 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:

$\sum e_i =$ ________________

$\sum X_{1i}e_i =$ ________________

$\sum X_{2i}e_i =$ ________________

$\sum X_{3i}e_i =$ ________________

$\sum X_{4i}e_i =$ ________________

$\sum \hat{y}_i e_i =$ ________________

(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.

$R^2_{\text{part(b)}} =$ ________________

$R^2_{\text{part(d)}} =$ ________________

What are your observations about the fitted values of model from part (b) and the fitted values of model from part (d)?

(e) (3 pts.) Using the model from part (b), test the hypothesis that the rental rate increases by more than 0.5 due to an increase of one unit in vacancy rate, holding the other predictors i.e. $\beta_3 > 0.5$.

Test Statistic = ____________________

P-values = ____________________

Conclusion:

(f) (2 pts.) The estimated correlation coefficient between $\hat{\beta}_2$ and $\hat{\beta}_4$ is equal to _____________________.

Good Luck

Q4: Code ______

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 $\rho_{y.X_2}$. We are 99% confidence that the unknown quantity $\rho_{y.X_2}$ 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:

$\sum e_i =$ __________________

$\sum X_{1i}e_i =$ ________________

$\sum X_{2i}e_i =$ ________________

$\sum X_{3i}e_i =$ ________________

$\sum X_{4i}e_i =$ ________________

$\sum \hat{y}_i e_i =$ _________________

(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.

$R^2_{\text{part(b)}} =$ ________________

$R^2_{\text{part(d)}} =$ ________________

What are your observations about the fitted values of model from part (b) and the fitted values of model from part (d)?

(e) (3 pts.) Using the model from part (b), test the hypothesis that the rental rate increases by more than 0.5 due to an increase of one unit in vacancy rate, holding the other predictors i.e. $\beta_3 > 0.5$.

Test Statistic = ____________________

P-values = ____________________

Conclusion:

(f) (2 pts.) The estimated correlation coefficient between $\hat{\beta}_2$ and $\hat{\beta}_4$ is equal to ______________________.

Good Luck

Q4: Code ______

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 $\rho_{y.X_2}$. We are 99% confidence that the unknown quantity $\rho_{y.X_2}$ 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:

$\sum e_i =$ ___________________

$\sum X_{1i}e_i =$ _________________

$\sum X_{2i}e_i =$ _________________

$\sum X_{3i}e_i =$ _________________

$\sum X_{4i}e_i =$ _________________

$\sum \hat{y}_i e_i =$ _________________

(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.

$R^2_{\text{part(b)}} =$ _________________

$R^2_{\text{part(d)}} =$ _________________

What are your observations about the fitted values of model from part (b) and the fitted values of model from part (d)?

(e) (3 pts.) Using the model from part (b), test the hypothesis that the rental rate increases by more than 0.5 due to an increase of one unit in vacancy rate, holding the other predictors i.e. $\beta_3 > 0.5$.

Test Statistic = ____________________

P-values = ____________________

Conclusion:

(f) (2 pts.) The estimated correlation coefficient between $\hat{\beta}_2$ and $\hat{\beta}_4$ is equal to _____________________.

Good Luck

Q4: Code ______

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 $\rho_{y.X_2}$. We are 99% confidence that the unknown quantity $\rho_{y.X_2}$ 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:

$\sum e_i =$ __________________

$\sum X_{1i} e_i =$ ________________

$\sum X_{2i} e_i =$ ________________

$\sum X_{3i} e_i =$ ________________

$\sum X_{4i} e_i =$ ________________

$\sum \hat{y}_i e_i =$ _________________

(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.

$R^2_{\text{part(b)}} =$ ________________

$R^2_{\text{part(d)}} =$ ________________

What are your observations about the fitted values of model from part (b) and the fitted values of model from part (d)?

(e) (3 pts.) Using the model from part (b), test the hypothesis that the rental rate increases by more than 0.5 due to an increase of one unit in vacancy rate, holding the other predictors i.e. $\beta_3 > 0.5$.

Test Statistic = ____________________

P-values = ____________________

Conclusion:

(f) (2 pts.) The estimated correlation coefficient between $\hat{\beta}_2$ and $\hat{\beta}_4$ is equal to _____________________.

Good Luck

Q4: Code ______

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 $\rho_{y.X_2}$. We are 99% confidence that the unknown quantity $\rho_{y.X_2}$ 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:

$\sum e_i =$ ________________

$\sum X_{1i}e_i =$ ________________

$\sum X_{2i}e_i =$ ________________

$\sum X_{3i}e_i =$ ________________

$\sum X_{4i}e_i =$ ________________

$\sum \hat{y}_i e_i =$ ________________

(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.

$R^2_{\text{part(b)}} =$ ________________

$R^2_{\text{part(d)}} =$ ________________

What are your observations about the fitted values of model from part (b) and the fitted values of model from part (d)?

(e) (3 pts.) Using the model from part (b), test the hypothesis that the rental rate increases by more than 0.5 due to an increase of one unit in vacancy rate, holding the other predictors i.e. $\beta_3 > 0.5$.

Test Statistic = ____________________

P-values = ____________________

Conclusion:

(f) (2 pts.) The estimated correlation coefficient between $\hat{\beta}_2$ and $\hat{\beta}_4$ is equal to _____________________.

Good Luck