Topic 2
Multiple Linear Regression

1. A sample of 25 observations has been represented by a model of the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

where ε_i is a random error term with mean 0 and variance σ^2. Given that

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 188.9832 & 0.8578 & -28.0275 \\ 0.8578 & 0.2500 & -0.6 \\ -28.0275 & -0.6 & 5.0625 \end{bmatrix};\ MS_E = 0.0361;\ \hat{\boldsymbol{\beta}} = \begin{bmatrix} -4.04 \\ 0.14 \\ 0.45 \end{bmatrix}$$

(a) Determine the 95% confidence interval for β_1.

(b) Find the test statistics for testing $H_0 : \beta_2 = 2$.

(c) Construct a 95% confidence interval for $3\beta_0 + 5\beta_1 + 2\beta_2$.

2. The following regression has been estimated:

$$\hat{y} = 5 + 7X_1 + 3X_2$$

(1.6) (5) (9) ($t-$ ratio in parentheses)

$MS_E = 5,\ n = 43,\ R^2 = 0.75,\ Var(y_i) = 19.047$

a) Form and complete the ANOVA table of this regression.

b) Calculate the standard errors of the coefficients.

c) Given an out-of-sample observation $\mathbf{x}_h$ = (1 200 300)’, calculate the prediction of the corresponding Y value.

3. Suppose that we are trying to assess the effect of age (X_1) and gender (X_2 = 0 for males, X_2=1 for females on systolic blood pressure (Y). Three versions of models were fit to the data set containing age, gender, and systolic blood pressure information on 35 males and 27 females.

Model I: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \varepsilon$. SS_T = 20856, MS_E = 80.02

Model II: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$. SS_E = 4676

Model III: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$. SS_R = 13434

a) Test $H_0 : \beta_1 = \beta_2 = \beta_{12} = 0$ versus $H_1 :$ At least one of the β_j ’s is nonzero.

b) Test $H_0 : \beta_2 = \beta_{12} = 0$ versus $H_1 :$ At least one of the β_j ’s is nonzero.

c) Obtain R^2_{Adj} for the three models, which models should be most preferred? Explain.

d) Suppose a fit of Model II to the data yields

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.83 & -0.03 & 0.025 \\ -0.03 & 0.003 & -0.005 \\ 0.025 & -0.005 & 0.33 \end{bmatrix},\ \mathbf{X}'\mathbf{y} = \begin{bmatrix} 225 \\ 2539 \\ -18 \end{bmatrix}.$$

Give the prediction equation for males, estimate the mean systolic blood pressure for males age 35 and determine a 95% confidence interval for the mean systolic blood pressure of all males age 35.

e) Determine the prediction interval for the systolic blood pressure of a male age 35.

4. The following is a study to estimate the relationship between $X_1,...,X_7$ and Y. There are 50 observations in the data set. The investigators consider four models. The sum of squares of error (SS_E) are included for each model.

	SS_E	Model
1.	307.64	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \varepsilon$
2.	900.96	$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_5 X_5 + \varepsilon$
3.	941.25	$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_3 + \beta_6 X_6 + \varepsilon$
4.	707.42	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{22} X_2^2 + \beta_7 X_7 + \varepsilon$

a) Perform a test of whether Model 1 fits significantly better than Model 2, if this is possible from the available information. If not, explain what other information is needed. Use $\alpha = 0.05$.

b) Perform a test of whether Model 3 fits significantly better than Model 4, if this is possible from the available information. If not, explain what other information is needed. Use $\alpha = 0.05$.

c) Given that $Var(y_i)$ = 2000, compute R^2 of Model 2.

5. A multiple regression based on n = 20 data points and k = 3 variables (including the dependent variable) has R^2 = 0.5. The analysis is modified to include an additional independent variable (the original ones plus a new one), but based on the same 20 data points as before (with the values of the added independent variable). For the new model, it is found that R^2 = 0.52. Find the percentage change in the value of R^2_{Adj} from the original to the new model.

6. The following regressions models have been fitted to 40 observations:

Model A: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$.

Model B: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$.

Model C: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$.

Given that:

(i) SS_E = 220 for Model A

(ii) The $F-$ ratio for testing that $\beta_2 = 0$ in going from Model A to Model B is 30.83

(iii) The F-ratio for testing that $\beta_3 = 0$ in going from Model B to Model C is 12.

Find the test statistic for testing that $\beta_2 = \beta_3 = 0$.