Kevin Su

STA 206 Homework 5

1. Under the multiple regression model with X variables $X_1, \ldots, X_{p-1}$, show the following:

(a) The LS estimator of the regression intercept is:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \cdots - \hat{\beta}_{p-1} \bar{X}_{p-1}$$

where $\hat{\beta}_K$ is the LS estimator for β_K and $\bar{X}_K = \frac{1}{n} \sum_{i=1}^{n} X_{iK}$ $(k=1, \ldots, p-1)$

$$Q(b) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \cdots - \beta_{p-1} X_{i,p-1}\right)^2$$

$$= \sum_{i=1}^{n} \left(Y_i - \beta_0 - \hat{\beta}_1 X_{i1} - \cdots - \hat{\beta}_{p-1} X_{i,p-1}\right)^2$$

$$\min Q(b) = \min \sum \left(Y_i - \beta_0 - \hat{\beta}_1 X_{i1} - \cdots - \hat{\beta}_{p-1} X_{i,p-1}\right)^2$$

$$\min E_i\left[Y_i - \beta_0 - \hat{\beta}_1 X_{i1} - \cdots - \hat{\beta}_{p-1} X_{i,p-1}\right]$$

$$\Rightarrow \beta_0 = E_i\left[Y_i - \hat{\beta}_1 X_{i1} - \cdots - \hat{\beta}_{p-1} X_{i,p-1}\right]$$

$$\boxed{\beta_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \cdots - \hat{\beta}_{p-1} \bar{X}_{p-1}}$$

1. (b) SSE and R^2 remain the same if all variables are centered

$\Rightarrow$ SSE $= \min Q(b) = \min \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{i1} - \cdots - b_{p-1} X_{i,p-1})^2$

Note $\hat{Y}_i = b_0 + b_1 X_{i1} + \cdots + b_{p-1} X_{i,p-1}$

$= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)$

Centering all variables simply shifts by $\bar{Y}$ or $\bar{X}$, using solution from (a)

$\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \cdots - \beta_{p-1} \overline{X_{p-1}}$

$\Downarrow$

$\beta_0 = (\bar{Y} - \bar{Y}) - \beta_1 (\bar{X}_1 - \bar{X}_1) - \cdots - \beta_{p-1} (\bar{X}_{p-1} -$

$= 0$

$\beta_i = \bar{Y} - \bar{Y} - \beta_0^{\,\,0} -$

$i \neq 0$

2. a

$$X = \begin{bmatrix} 1 & .36 & 2.14 \\ 1 & .66 & .74 \\ 1 & .66 & 1.91 \\ 1 & -0.52 & -.41 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

b

$SSE = 26(1.02)^2 = 27.048$

$SST = 91.218$

$SSR = 64.17$

From ANOVA Table

c

$SSR(X_1 \cdot X_2 | X_1, X_2) = 0.448$

$SSR(X_1, X_2) = 63.722$

$SSR(X_2, X_1 \cdot X_2 | X_1) = 5.938$

$SSR(X_2 | X_1) = 5.490$

$SSR(X_1) = 58.232$

$SSE(X_1) = SST - SSR(X_1)$
$= 32.986$

$SSE(X_1, X_2) = SST - SSR(X_1, X_2)$
$= 27.496$

d

Predicted Value at $X_1 = 0$, $X_2 = 0$ is $\boxed{.9918}$ (Intercept

Prediction Standard error $= \sqrt{MSE(1 + X_h'(X'X)^{-1}X_h)}$

$X_h'(X'X)^{-1}X_h = .087$

$s(pred_h) = \sqrt{1.087(MSE)} = \boxed{1.0598}$

95% conf. Int $\Rightarrow$ $.9918 \pm (1.0598)(2.056) = \boxed{.9918 \pm 2.179}$

2(e)

Reduced Model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Full Model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

$H_0: \beta_3 = 0$

$H_A: \beta_3 \neq 0$

Reject if $F_{(1,26)} > 5.526$

$F^* = 0.4311 < 5.526$

$\therefore$ We fail to reject null for coefficient corresponding with $X_1 \cdot X_2$ at .01 level

$H_0: \beta_2 = 0$

$H_A: \beta_2 \neq 0$

Full Model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Reject if $F_{(1,27)} > 7.677$

$F^* = 5.2775 < 7.677$

$\therefore$ We fail to reject null for coefficient corresponding with X_2 at .01 level

Reduced Model: $Y = \beta_0 + \beta_1 X_1$

3. When executing in R, would order those two variables last and establish confidence intervals for them to see if the confidence intervals include β_{10} and β_{20}.