Regression Analysis Midterm – Solution Key

November 14, 2008

1. (a)

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}(X'\boldsymbol{y}) = \begin{pmatrix} 2.307 & 0.156 & -1.83 \\ 0.156 & 0.026 & -0.204 \\ -1.883 & -0.204 & 1.977 \end{pmatrix} \begin{pmatrix} 254 \\ 2280 \\ 483 \end{pmatrix} = \begin{pmatrix} 32.169 \\ 0.372 \\ 11.489 \end{pmatrix}$$

(b) $\hat{\sigma^2} = s^2 = 48.837$, $\boldsymbol{x}_0' = (1, 15, 2.5)$, $n = 15$, $p = 3$ and $t_{12;.025} = 2.179$.
Now $\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} = 66.4725$, so the 95% prediction interval of y_0 is $66.4715 \pm 2.179\sqrt{48.837}(1 + \boldsymbol{x}_0'(X'X)^{-1}\boldsymbol{x}_0)^{1/2}$. Note that

$$\boldsymbol{x}_0'(X'X)^{-1}\boldsymbol{x}_0 = (1\ 15\ 2.5)\begin{pmatrix} -0.0605 \\ 0.0360 \\ -0.0005 \end{pmatrix} = .47825,$$

we get

$$y_0 \in 66.4715 \pm 18.5142 = (47.957, 84.986).$$

(c) $\{\boldsymbol{\beta} : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(X'X)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})/3 \le s^2\, F_{3,12;\alpha}\}$ is a 95% confidence region for $\boldsymbol{\beta}$, where $F_{3,12;\alpha} = 3.49$.

(d) (i) Bonferroni's method: $l = 2$, $t_{12;\alpha/4} = 2.597$ and

$$\beta_1 \in \hat{\beta}_1 \pm 2.597\sqrt{48.837} \times \sqrt{.026} = (-2.554, 3.299),$$

and

$$\beta_2 \in 11.489 \pm 2.597 \times 6.988 \times \sqrt{1.977} = (-14.028, 37.006).$$

(ii) Scheffé's methods: $F_{2,12;.05} = 3.89$, so $\sqrt{2 \times F_{2,12;0.05}} = 2.789$. Thus

$$\beta_1 \in \hat{\beta}_1 \pm 2.789\sqrt{48.837} \times \sqrt{.026} = (-2.771, 3.515),$$

and

$$\beta_2 \in 11.489 \pm 2.789 \times 6.988 \times \sqrt{1.977} = (-15.918, 38.896).$$

(e) Test $H_0 : \beta_1 = \beta_2$. Use $T = (0, 1, -1)$, and $\gamma = 0$ so $T\hat{\boldsymbol{\beta}} = -11.117$ and $T(X'X)^{-1}T' = (2.039\ .23\ -2.181)T' = 2.411$. Thus,

$$F = \frac{(-11.117)^2/2.411}{48.837} = 1.0496 < F_{1,12;.05} = 4.75,$$

do not reject H_0 at $\alpha = 0.05$.

2. $n = 27$, $p = 4$ and $k = 3$.

(a) Full model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$.

$$F_0 = \frac{5506.3/3}{2363.8/23} = \frac{1835.433}{102.774} = 17.859 > F_{3,23;.05} = 3.03.$$

The regression model is significant at $\alpha = 5\%$.

(b) $R^2 = \frac{SS_R}{SS_T} = \frac{5506.3}{5506.3+2363.8} = \frac{5506.3}{7870.1} = 0.6996$.

(c) i. $SS(\beta_2|\beta_0, \beta_1) = 455.2$ and $SS_{Res} = 7870.1 - 5008.9 - 455.2 = 2406$. Thus

$$F_0 = \frac{455.2}{2406/24} = 4.541 > F_{1,24;0.05} = 4.26.$$

Reject $\beta_2 = 0$.

ii. The model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$. Hence,

$$R^2 = 1 - \frac{2406}{7870.1} = .6943 \text{ and } R^2_{adj} = 1 - \frac{2406/24}{7870.1/26} = 0.6688.$$

iii. $s^2 = 2406/24 = 100.25$.

(d) $SS(\beta_2, \beta_3|\beta_0, \beta_1) = 5506.3 - 5008.9 = 497.4$, so $F_0 = \frac{497.4/2}{2363.8/23} = 2.419 \sim F_{2,23;.05} = 4.28$, reject H_0.

3. $n = 32$, $\bar{y} = 144.53$, $\bar{x} = 0.53$; and $S_{yy} = \sum(y_i - \bar{y})^2 = [sd(y)]^2 \times 31 = 14.4^2 \times 31 = 6428.16$, $S_{xx} = 0.51^2 \times 31 = 8.0631$, and

$$S_{xy} = corr(y, x)\sqrt{S_{xx}S_{yy}} = 0.25 \times \sqrt{6428.16}\sqrt{8.0631} = 56.916.$$

Then $\hat{\beta}_1 = S_{xy}/S_{xx} = 56.916/8.0631 = 7.059$, $SS_R = \hat{\beta}_1 S_{xy} = 401.77$ and $SS_{Res} = 6428.16 - 401.77 = 6026.39$.

(a)

Source	SS	df
SS_R	401.77	1
SS_{Res}	6026.39	30
SS_T	6428.16	31

(b) The least squares estimate of the slope is $\hat{\beta}_1 = 7.059$ and that of the intercept is $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 144.53 - 7.059 \times .53 = 140.789$.

(c) $E(y|\text{smokers}) = E(y|x = 1) = \beta_0 + \beta_1$ is estimated by $\hat{\beta}_0 + \hat{\beta}_1 = 147.84$.
$E(y|\text{nonsmokers}) = E(y|x = 0) = \beta_0$ is estimated by $\hat{\beta}_0 = 140.789$.

(d) $E(y|\text{smokers}) = E(y|\text{nonsmokers})$ if and only if $\beta_1 = 0$, i.e. $H_0 : \beta_1 = 0$.

4. To fit the model $\boldsymbol{y} = X_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$, we have $\hat{\boldsymbol{\beta}}_1 = (X_1'X_1)^{-1}X_1'\boldsymbol{y}$.
   Now under the model $\boldsymbol{y} = X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$ with $E(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \sigma^2 I$, it yields that

$$E(\hat{\boldsymbol{\beta}}_1) = (X_1'X_1)^{-1}X_1'(X_1\boldsymbol{\beta}_1 + X_2\boldsymbol{\beta}_2) = \boldsymbol{\beta}_1 + (X_1'X_1)^{-1}X_1'X_2\boldsymbol{\beta}_2$$

   and

$$Var(\hat{\boldsymbol{\beta}}_1) = (X_1'X_1)^{-1}X_1'\sigma^2 I[(X_1'X_1)^{-1}X_1']' = (X_1'X_1)^{-1}\sigma^2.$$

   Therefore, $\hat{\boldsymbol{\beta}}_1$ is unbiased of $\boldsymbol{\beta}_1$ if and only if $X_1'X_2 = 0$, i.e. when X_1 and X_2 are orthogonal.