STAT 420 Spring 2014

Homework 4: Solutions

Exercise 1

(a) The least-squares regression coefficients are given by

$$\begin{aligned}
\hat{\mathbf{b}} &= \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= \begin{pmatrix} 0.95 & -0.15 & -0.1 \\ -0.15 & 0.05 & 0 \\ -0.1 & 0 & 0.025 \end{pmatrix} \begin{pmatrix} 620 \\ 1960 \\ 2600 \end{pmatrix} \\
&= \begin{pmatrix} 0.95(620) - 0.15(1960) - 0.1(2600) \\ -0.15(620) + 0.05(1960) + 0(2600) \\ -0.1(620) + 0(1960) + 0.025(2600) \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{35} \\ \mathbf{5} \\ \mathbf{3} \end{pmatrix}
\end{aligned}$$

(b) The hypotheses are $H_0 : b_1 = b_2 = 0$ versus $H_1 : b_j \neq 0$ for some $j \in \{1, 2\}$

First note that $SSE = \sum_{i=1}^{10}(y_i - \hat{y})^2 = 896$ and $SST = \sum_{i=1}^{10}(y_i - \bar{y})^2 = 1756$, which implies that $SSR = SST - SSE = 1756 - 896 = 860$

$MSR = SSR/2 = 860/2 = 430$ and $MSE = SSE/7 = 896/7 = 128$

The F statistic is given by $F = MSR/MSE = 430/128 = 3.359375$

The F critical value is given by $F_{(2,7;\ 0.1)} = 3.257442$

$F_{(2,7;\ 0.1)} = 3.257442 < 3.359375 = F \implies$ **Reject $\mathbf{H_0}$**

(c) The covariance matrix for $\hat{\mathbf{b}}$ is given by

$$\begin{aligned}
\hat{V}(\hat{\mathbf{b}}) &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\
&= 128 \begin{pmatrix} 0.95 & -0.15 & -0.1 \\ -0.15 & 0.05 & 0 \\ -0.1 & 0 & 0.025 \end{pmatrix}
\end{aligned}$$

which implies that the standard error of $\hat{b}_1$ is given by $\hat{\sigma}_{\hat{b}_1} = \sqrt{128(0.05)} = 2.529822$

The critical T value is $t_{(7;\ .05)} = 1.894579$

So the 90% CI is given by: $\hat{b}_1 \pm t_{(7;\ .05)}\hat{\sigma}_{\hat{b}_1} = 5 \pm (1.894579)(2.529822) = [\mathbf{0.2070531};\ \mathbf{9.7929469}]$

(d) Using the result from 1(c), the 90% CI is given by $[0.2070531;\ 9.7929469]$

Note that $0 \notin [0.2070531, 9.7929469] \Longrightarrow$ **Reject $\mathbf{H_0}$**

Or $t_{(7;\ .05)} = 1.894579 < 1.976424 = 5/2.529822 = T \Longrightarrow$ **Reject $\mathbf{H_0}$**

(e) Using 1(c), the standard error of $\hat{b}_2$ is given by $\hat{\sigma}_{\hat{b}_2} = \sqrt{128(0.025)} = 1.788854$

The test statistic is $T = (3-5)/1.788854 = -1.118034$

Note that $T \sim t_7$, which implies that $P(T < -1.118034) = 0.1502244 \Longrightarrow$ **Retain $\mathbf{H_0}$**

Or $t_{(7;\ .9)} = -1.414924 < -1.118034 = T \Longrightarrow$ **Retain $\mathbf{H_0}$**

(f) Given $x_1 = 2$ years experience and $x_2 = 4$ publications, the expected salary would be $\hat{y} = \hat{b}_0 + \hat{b}_1 2 + \hat{b}_2 4 = 35 + 5(2) + 3(4) = 57$

The variance of the predicted observations would be

$$\begin{aligned}
\hat{\sigma}^2_{\hat{y}} &= \hat{\sigma}^2 \left[1 + \begin{pmatrix} 1 & 2 & 4 \end{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \right] \\
&= 128 \left[1 + \begin{pmatrix} 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 0.95 & -0.15 & -0.1 \\ -0.15 & 0.05 & 0 \\ -0.1 & 0 & 0.025 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \right] \\
&= 147.2
\end{aligned}$$

So the 90% PI is given by: $\hat{y} \pm t_{(7;\ .05)}\hat{\sigma}_{\hat{y}} = 57 \pm (1.894579)\sqrt{147.2} = [\mathbf{34.01383};\ \mathbf{79.98617}]$

Exercise 2

(a) The hypotheses are $H_0 : b_1 = b_2 = b_3 = 0$ versus $H_1 : b_j \neq 0$ for some $j \in \{1, 2, 3\}$

First note that $SSE = \sum_{i=1}^{8}(y_i - \hat{y})^2 = 2000$ and $SST = \sum_{i=1}^{8}(y_i - \bar{y})^2 = 19700$, which implies that $SSR = SST - SSE = 19700 - 2000 = 17700$

$MSR = SSR/3 = 17700/3 = 5900$ and $MSE = SSE/4 = 2000/4 = 500$

The F statistic is given by $F = MSR/MSE = 5900/500 = 11.8$

The F critical value is given by $F_{(3,4;\ 0.05)} = 6.591382$

$F_{(3,4;\ 0.05)} = 6.591382 < 11.8 = F \Longrightarrow$ **Reject $\mathbf{H_0}$**

(b) The covariance matrix for $\hat{\mathbf{b}}$ is given by

$$\begin{aligned}\hat{V}(\hat{\mathbf{b}}) &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= 500 \begin{pmatrix} 17.08333 & -12.5 & -0.33333 & 2.75 \\ -12.5 & 10 & 0 & -2.5 \\ -0.33333 & 0 & 0.53333 & 0 \\ 2.75 & -2.5 & 0 & 0.75 \end{pmatrix}\end{aligned}$$

which implies that the standard error of $\hat{b}_2$ is given by $\hat{\sigma}_{\hat{b}_2} = \sqrt{500(0.53333)} = 16.32988$

The test statistic is $T = \hat{b}_2/\hat{\sigma}_{\hat{b}_2} = 40/16.32988 = 2.449497$

The critical T values are $t_{(4;\ .975)} = -2.776445$ and $t_{(4;\ .025)} = 2.776445$

$t_{(4;\ .025)} = 2.776445 > 2.449497 = T \Longrightarrow$ **Retain $\mathbf{H_0}$**

(c) First note that $\sum_{i=1}^{8}(x_{i3} - \bar{x}_3)^2 = \sum_{i=1}^{8} x_{i3}^2 - (\sum_{i=1}^{8} x_{i3})^2/8 = 80 - (24^2)/8 = 8$

Similarly note that $\sum_{i=1}^{8}(x_{i3} - \bar{x}_3)(y_i - \bar{y}) = \sum_{i=1}^{8} x_{i3}y_i - (\sum_{i=1}^{8} x_{i3})(\sum_{i=1}^{8} y_i)/8 = 5860 - (24)(1840)/8 = 340$

For the null model $y_i = b_0 + b_3x_{i3} + e_i$, we have $\hat{b}_3 = \frac{\sum_{i=1}^{8}(x_{i3}-\bar{x}_3)(y_i-\bar{y})}{\sum_{i=1}^{8}(x_{i3}-\bar{x}_3)^2} = 340/8 = 42.5$

The SSR for the null model is $SSR_{null} = \sum_{i=1}^{8}(\hat{y}_i - \bar{y})^2 = \hat{b}_3^2 \sum_{i=1}^{8}(x_{i3} - \bar{x}_3)^2 = 14450$

So, the SSE for the null model is $SSE_{null} = SST - SSR_{null} = 19700 - 14450 = 5250$

Thus, the F statistic is $F = \frac{(SSE_{null}-SSE)/(df_{null}-df)}{SSE/df} = \frac{(5250-2000)/(6-4)}{2000/4} = \frac{1625}{500} = 3.25$

The critical F statistic is $F_{(2,4;\ .05)} = 6.944272$

$F_{(2,4;\ .05)} = 6.944272 > 3.25 = F \Longrightarrow$ **Retain $\mathbf{H_0}$**

(d) Given $x_1 = 2$ thousand square feet, $x_2 = 1$ backyard, and $x_3 = 3$ bedrooms, the expected selling price would be $\hat{y} = \hat{b}_0 + \hat{b}_1 2 + \hat{b}_2 1 + \hat{b}_3 3 = 15 + 50(2) + 40(1) + 30(3) = 245$ thousand dollars

The variance of the predicted observations would be

$$
\begin{aligned}
\hat{\sigma}^2_{\hat{y}} &= \hat{\sigma}^2 \left[1 + \begin{pmatrix} 1 & 2 & 1 & 3 \end{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} \right] \\
&= 500 \left[1 + \begin{pmatrix} 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 17.08333 & -12.5 & -0.33333 & 2.75 \\ -12.5 & 10 & 0 & -2.5 \\ -0.33333 & 0 & 0.53333 & 0 \\ 2.75 & -2.5 & 0 & 0.75 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \\ 3 \end{pmatrix} \right] \\
&= 600
\end{aligned}
$$

So the 95% PI is given by: $\hat{y} \pm t_{(4;\ .05)} \hat{\sigma}_{\hat{y}} = 245 \pm (2.776445)\sqrt{600} = [\mathbf{176.9913};\ \mathbf{313.0087}]$

Exercise 3

(a) The hypotheses are $H_0 : b_1 = \ldots = b_8 = 0$ versus $H_1 : b_j \neq 0$ for some $j \in \{1, \ldots, 8\}$

First note that $SSE = \sum_{i=1}^{36} (y_i - \hat{y})^2 = 108$ and $SST = \sum_{i=1}^{36} (y_i - \bar{y})^2 = 204$, which implies that $SSR = SST - SSE = 204 - 108 = 96$

$MSR = SSR/8 = 96/8 = 12$ and $MSE = SSE/27 = 108/27 = 4$

The F statistic is given by $F = MSR/MSE = 12/4 = 3$

The F critical value is given by $F_{(8,27;\ 0.05)} = 2.305313$

$F_{(8,27;\ 0.05)} = 2.305313 < 3 = F \Longrightarrow$ **Reject $\mathbf{H_0}$**

(b) The hypotheses are $H_0 : b_2 = b_4 = b_8 = 0$ versus $H_1 : b_j \neq 0$ for some $j \in \{2, 4, 8\}$

First note that $SSE_{null} = \sum_{i=1}^{36} (y_i - \hat{y})^2 = 138$

The F statistic is given by $F = \frac{(SSE_{null} - SSE)/(df_{null} - df)}{SSE/df} = \frac{(138-108)/(30-27)}{108/27} = \frac{10}{4} = 2.5$

The F critical value is given by $F_{(3,27;\ 0.05)} = 2.960351$

$F_{(3,27;\ 0.05)} = 2.960351 > 2.5 = F \Longrightarrow$ **Retain $\mathbf{H_0}$**