Regression Assignment Two

Terry Liu 刘骏杰 l630005038
Section: Dr.Hua-Jun YE

2018 年 10 月 8 日

Question One Answer: (a) According to the basic SLR model, we can see that $\beta_1 = \frac{S_{xy}}{S_{xx}}$ and now we are given $\sum x$ and $\sum y$:

$$
\begin{aligned}
S_{xy} &= \sum (x_i - \overline{x})(y_i - \overline{y}) \\
&= \frac{\sum (x_i) \sum (y_i) - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = 1 \\
S_{xx} &= \sum (x_i - \overline{x})^2 \\
&= \sum x^2 - \frac{(\sum x_i)^2}{n}
\end{aligned}
$$

We can get $\beta_1 = 1, \beta_0 = -1$ the LSR line is: $\hat{y} = 1 + 1x$

(b)

$\beta_1 = 1$ is the estimated change in the average value of Y as a result of a one-unit change in X

$\beta_0 = 1$ represents the Y is also influenced by other factor(s).

(c)

$SS_{regression} = \sum (\hat{y}_i - y_i)^2 = b_1^2 \sum (x_i - \overline{x})^2 = 14$

$SS_{total} = \sum_i (y_i - \overline{y})^2 = \sum y_i^2 - 2\overline{y} \sum y_i + \sum \overline{y}^2 = 18$

$SS_{error} = SS_{total} - SS_{regression} = 18$

The Sum of Squares Regression (SSR) is the sum of the squared differences between the prediction for each observation and the population mean.

(d)

ANOVA Table				
Source	Sum of Squares	Degress of Freedom	Mean Squares	F
Regression	14	1	14	28
Residual	4	8	0.5	
Total	18	9		

(e)

$\sigma = \sqrt{\frac{SS_{error}}{n-2}} = 0.71$

(f)

We assume the hypothesis is $H_0 : \beta_1 = 0$, so the $H_1 : \beta_1 \neq 0$, the significance level is $\alpha = 0.05$, we can get that the t-test: $t_{\frac{\alpha}{2},n-2} = 2.306$. The t-value is $\frac{\hat{\beta_1}\sqrt{S_xx}}{s} = \frac{\sqrt{14}}{0.71} = 5.27 > t_{\frac{\alpha}{2},n-2} = 2.306$

(g)

The Significance level is $\alpha = 0.05$ so we can get the confidence interval: $\hat{\beta_1} \pm t_{\frac{\alpha}{2},n-2}\frac{s}{\sqrt{S_xx}} = 1 \pm 2.306 \times \frac{0.71}{\sqrt{14}} \Rightarrow 95\%$ confidence interval $[0.56, 1.44]$

(h)

$$R^2 = \frac{S_xy^2}{S_xxS_yy} = \frac{196}{252} = 0.78$$

$\mathbf{R^2}$ relects how strong the relationship of two variables there are. In this case, $R^2 = 0.78$ is close to 1, which means the SS_{error} is small and the relationship between x and y is strong.

(i)

$$r = b_1\sqrt{\frac{S_xx}{S_xy}} = 0.88$$

(j)

Significance level is $\alpha = 0.05$ if $x' = 4 \Rightarrow y(\hat{x'}) = \hat{\beta_0} + \hat{\beta_1}x' = 5$ and $t_{\frac{\alpha}{2},n-2} = 2.306$.

$y(\hat{x'}) \pm t_{\frac{\alpha}{2},n-2}s\sqrt{\frac{1}{n} + \frac{(x'-\overline{x})^2}{S_xx}} = 5 \pm 2.306 \times 0.71 \times \sqrt{\frac{1}{10} + \frac{(4-3)^2}{14}}$

So the 95% confidence interval is $[4.32, 5.68]$

(k)

Significance level is $\alpha = 0.05$ if $x'' = 4 \Rightarrow y(\hat{x'}) = \hat{\beta_0} + \hat{\beta_1}x' = 5$ and $t_{\frac{\alpha}{2},n-2} = 2.306$.

$y(\hat{x''}) \pm t_{\frac{\alpha}{2},n-2}s\sqrt{1 + \frac{1}{n} + \frac{(x'-\overline{x})^2}{S_xx}} = 5 \pm 2.306 \times 0.71 \times \sqrt{1 + \frac{1}{10} + \frac{(4-3)^2}{14}}$

So the 95% confidence interval is $[3.23, 6.77]$

(l)

We assume the hypothesis is $H_0 : \beta_1 = 0$, so the $H_1 : \beta_1 \neq 0$, the significance level is $\alpha = 0.05$, we can get that the t-test:

$$t_{\frac{\alpha}{2},n-2} = 2.306$$

$$r = 0.88 \quad r^2 = 0.78$$

t-value equals:$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = 0.88 \times \frac{\sqrt{8}}{\sqrt{1-0.78}} = 5.31 > 2.306$ and we reject the null hypothesis.

Question Two Answer: Using the matrix, we can notice that $y = \mathbf{X}\beta_1 + \epsilon \Rightarrow \beta_1 = (\mathbf{X^T X})^{-1}\mathbf{X^T}y$

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

we can get $\beta_1 = (4.54, 0, 29)^T$

(d)

X	30	40	50	80	30	40	60	70	70	70	30	80	70	70
Y	13	17	20	29	12	15	22	25	23	27	15	27	24	26
Fitted Value	13.3	16.3	19.2	28.0	13.3	16.3	22.1	25.0	25.0	25.0	13.3	28.0	25.0	25.0
Residual	-0.3	0.7	0.8	1.0	-1.3	-1.3	-0.1	0.0	-2.0	2.0	1.7	-1.0	-1.0	1.0

ANOVA Table				
Source	Sum of Squares	Degress of Freedom	Mean Squares	F
Regression	405.44	1	1405.44	249.69
Residual	19.49	12	1.62	
Total	424.93	13		

$$Cov(\hat{\beta}) = \begin{pmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_0) \end{pmatrix} = \begin{pmatrix} 1.21 & -0.02 \\ -0.02 & 0.00 \end{pmatrix}$$

$$Var(\hat{y}|x = 65) = \sigma^2[\frac{1}{n} + \frac{(x_0 - \overline{x})}{S_x x}] = 0.14$$

$$y(\hat{x_0}) \pm t_{\frac{\alpha}{2}, n-2} s \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_x x}} = 23.58 \pm 0.82$$

So, the 95% confidence interval is $[22.76, 24.40]$

Question Three Answer: According to the data from question:

(a) T-statistic:

$$\frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{S_x x}}} = -\frac{210.35}{24.19} = -8.70$$

$$t_{\frac{\alpha}{2}, n-2} = 4.14$$

$$8.70 > 4.14$$

So, we have to reject the $H_0: \ \beta_1 = 0$

(b)

$\hat{\beta_1} = -210.35$ $\hat{\beta_0} = 5566.1$ and $x = 8.5, \hat{y} = 3778.125$

(c)

$$SS_{error} = (n-2) \times MSE = 14 \times 52439 = 734146$$

(d)

The proportion of unexplained variance in the model is $F = 75.59$

(e)

$x_0 = 8.5, \ y(\hat{x_0}) = 3778.125$

By given value $\sum x = 163.65, \ sumx^2 = 1763.418$

So, we can get $\overline{x} = 10.20 \Rightarrow S_{xx} \sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x)^2 \approx 89.59$

and by $\sqrt{\frac{\sigma^2}{S_{xx}}} = 24.19$ we can get $\sigma^2 \approx 52424.14$

We can get: $\hat{y}(x_o) \pm t_{\frac{\alpha}{2}, n-2} = 2.145$ so the 95% prediction interval is $[36.26.94, 3929.31]$

(f)

The hypothesis $H_0: \rho = 0, \ H_1: \rho \neq 0$ and significance level $\alpha = 0.01$

$$r^2 = \frac{SS_{regression}}{SS_{regression} + SS_{error}} = \frac{3963719}{3963719 + 734146} = 0.84$$

So, $r = \sqrt{r^2} \approx 0.92$ $T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.92\sqrt{16-2}}{\sqrt{1-0.84}} \approx = 6.75$

Because $6.75 > 2.977$, we can reject $H_0: \rho = 0$

Question Four Answer: From the problem we can get:

$$
\begin{aligned}
Q(\omega) &= \sum_{i=1}^{3}(y_i-\omega_i)^2 = (y_1-\omega_1)^2+(y_2-\omega_2)^2+(y_3-\omega_3)^2\\
\frac{\partial Q}{\partial \omega_1} &= 2(y_1-\omega_1)=0\\
\frac{\partial Q}{\partial \omega_2} &= 2(y_2-\omega_2)=0\\
\frac{\partial Q}{\partial \omega_3} &= 2(y_3-\omega_3)=0\\
\Rightarrow (\hat{\omega_1},\hat{\omega_2},\hat{\omega_3}) &= (3.06,1.94,1.05) \Rightarrow Cov(\hat{\omega}) = Cov(y) = \sigma^2 I_n
\end{aligned}
$$

(b)

$$
\begin{aligned}
Q(\omega) = (z_1-\omega_1-\omega_2-\omega_3)^2 + (z_2-\omega_1+\omega_2)^2 + (z_3-\omega_2+\omega_3)^2\\
z = \begin{pmatrix} 1 & 1 & 1\\ 1 & -1 & 0\\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix}\omega_1\\ \omega_2\\ \omega_3\end{pmatrix} + \epsilon = X\omega+\epsilon\\
\hat{\omega} = (X^TX)^{-1}X^Tz = \begin{pmatrix}3.01\\ 2.02\\ 1.00\end{pmatrix}\\
Cov(\hat{\omega}) = \begin{pmatrix} \frac{2}{3} & 0 & -\frac{1}{3}\\ \frac{1}{3} & \frac{1}{3} & 0\\ -\frac{1}{3} & 0 & \frac{2}{3}\end{pmatrix}\sigma^2
\end{aligned}
$$

(c)

The second one is better since its $(\hat{\omega})$ is smaller.

Question Five Answer: The proof:

$$
\begin{aligned}
E(x^TAx) &= E(tr(x^TAx) \\
&= E(tr(Axx^T)) \text{ Because } trace(xy) = trace(yx) \\
&= tr(AExx^T) = tr(A(Cov(x) + E(x)E(x)^T)) \\
&= tr(A + (\mu^TA\mu)
\end{aligned}
$$

So the equation has been proved.

Question Six Answer: (a)

Because $E(x) = 1,\ E(Y) = 2.\ Var(x) = 3,\ Var(Y) = 4,\ Cov(x, y) = 2$

$$
\begin{aligned}
Z = 2x + Y\ W = x - 2Y \\
E(Z) = E(2x + Y) = E(2X) + E(Y) = 4 \\
Var(Z) = Var(2x + Y) = Var(2x) + Var(Y) + 2Cov(2x, Y) = 24 \\
E(W) = E(x - 2Y) = E(x) + E(2Y) = -3 \\
Var(W) = Var(x - 2Y) = Var(x) + Var(2Y) - 2Cov(x, 2Y) = 11
\end{aligned}
$$

(b)

Because from the section (a) we can notice that: $Var(x) = 3,\ Var(Y) = 4,\ Cov(x, y) = 2,\ Z = 2x + Y,\ W = x - 2Y$, that is:

$$
\begin{aligned}
Cov(x, Z) = Cov(x, 2x + Y) = 2Var(x) + Cov(x, Y) = 8 \\
Cov(w, Y) = Cov(x - 2Y, Y) = Cov(x, Y) - 2Var(Y) = -6 \\
Cov(Z, w) = Cov(2x + Y, x - 2Y) = 2Var(x) - 3Cov(x, Y) - 2Var(Y) = -8
\end{aligned}
$$

(c)

$Cov(x,Y) = \frac{Cov(x,Y)}{\sqrt{Var(x)Var(Y)}} = \frac{2}{\sqrt{2}} = \frac{\sqrt{3}}{3} \approx 0.58$

(d)

$$Var(Z) = Var(2x+Y) = Cov(2x+Y, 2x+Y) = 4Var(x) + 4Cov(x,Y) + Var(Y) = 24$$
$$Var(W) = Var(x-2Y) = Var(x) - 4Cov(x,Y) - 4Var(Y) = 11$$
$$Corr(Z,W) = \frac{Cov(Z,W)}{\sqrt{Cov(Z)Cov(W)}} = \frac{-8}{\sqrt{24}\sqrt{11}} = -0.49$$

Question Seven Answer: $E(x) = \begin{pmatrix} 1.1 \\ 2.3 \\ 3.2 \\ 1.7 \end{pmatrix}$, $S = Cov(x) = \begin{pmatrix} 1.0 & 0.5 & 0.4 & 0.3 \\ 0.5 & 2.0 & 0.5 & 0.4 \\ 0.4 & 0.5 & 3.0 & 0.6 \\ 0.3 & 0.4 & 0.6 & 1.5 \end{pmatrix}$

(a)

$X_1 = (x_1, x_3)^T = \begin{pmatrix} 1.1 \\ 3.2 \end{pmatrix}$,so the $Cov(X_1) = Cove \begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.3 \\ 0.5 & 0.6 \end{pmatrix}$

(b)

(c)

$$E(z_1) = E(x_1) + E(2x_2) + E(3x_3) + E(4x_4) = 22.1$$
$$E(z_2) = E(4x_1) - E(3x_2) - E(2x_3) + E(x_4) = -7.2$$
$$z = (z_1, z_2)^T$$
$$Cov(z) = \begin{pmatrix} 22.1 \\ -7.2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 + 3x_3 + 4x_4 \\ 4x_1 - 3x_2 - 2x_3 + x_4 \end{pmatrix} = \begin{pmatrix} 93.6 & -21.9 \\ -21.9 & 32.7 \end{pmatrix}$$

(d)

$Cov(x_1-2x_2,\ x_2-3x_3+1.5x_4) = Cov(x_1,x_2)-3Cov(x_1,x_3)+1.5Cov(x_1,x_4)-2Var(x_2) + 6Cov(x_2,x_3) - 3Cov(x_2,x_4) = -2.45$

Question Eight Answer: (a)

From the LSE we can notice that:

$$\hat{y} = x\hat{\beta} = \mathbf{H}y$$
$$\text{where } \hat{y} = \begin{pmatrix} \hat{y_1} \\ \vdots \\ \hat{y_n} \end{pmatrix},\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.\ H \text{ is a hat matrix.}$$
$$\begin{aligned} Var(\hat{y}) &= Var(\mathbf{H}y) = H^T Var(y) H = H^T \sigma^2 I_n H = sigma^2 H^T H \\ &= \sigma^2 (X(X^TX)^{-1}X^T)^T X^T X (X^TX)^{-1} X^T \\ &= X(X^TX)^{-1}X^T = H \end{aligned}$$
$$(I-H)^T = (I-H)(I+H) = I - 2H + H^2 d$$
$$\text{since } H^2 = H \Rightarrow (I-H)^2 = I - H$$

So, H and $I-H$ are independent.

Question Nine Answer: (a)

!

$$P_1^T = P_1 \Rightarrow P_1^2 = P_1$$
$$P_2^T = P_2 \Rightarrow P_2^2 = P_2$$
From the question: $P_1P_2 = P_2$
$$\text{so}(P_1P_2)^T = P_2^T \Rightarrow P_2^TP_1^T = P_2^T \Rightarrow P_2P_1 = P_2$$
Then we are going to proof: $(P_1 - P)^2 = P_1 - P_2$
$$(P_1 - P_2)^2 = P_1^2 - P_1P_2 - P_2P_1 + P_2^2 = P_1 - P_2$$
There for $P_1 - P_2$ is a projection matrix.

Question Ten Answer: (a)

$$E(MS_{Res}) = E(\frac{SSE}{n-2}) = \frac{1}{n-2}E(\sum_{i=1}^{n}(y_i - \hat{h}_i)^2)$$

$$\begin{aligned}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 &= (y_i - \hat{y}_i)^T)(y - \mathbf{H}y) \\ &= (y - \mathbf{H}y)^T(y - \mathbf{H}y) \\ &= y^T(I - \mathbf{H})y\end{aligned}$$

Because $y = \begin{pmatrix} y_1 \\ vdots \\ y_n \end{pmatrix}$, $\mathbf{H}$is a hat matrix: E$(\sum_{i=1}^{2}(y_i - \hat{y}_i)^2) = E(y^T(I - \mathbf{H})y)$
$= (n-2)\sigma^2$
$E(MS_{Res}) = \frac{1}{n-2}(n-2)\sigma^2 = \sigma^2$
(b)
$E(MS_{Reg}) = E(\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2) = E(\sum_{i=1}^{n}(b_1(x_i - \overline{x}))^2) = S_{xx}(\frac{\sigma^2}{S_{xx}} + \beta_1^2) = \sigma^2 + \beta_1^2S_{xx}$