STAT3008 HW3sln

1. Solution:

(a)

$x = (1,12,-2)', \hat{y} = x'\hat{\beta} = 209.04, t_{5,0.05} = 2.015, se(\hat{y}) = \hat{\sigma}\sqrt{x'(X'X)^{-1}x} = 125.250,$

The 90% C.I. for the fitted value is: $\hat{y} \pm t_{5,0.05}\hat{\sigma}\sqrt{x'(X'X)^{-1}x} = [-42.439, 462.319]$;

(b)

$x = (1,12,-2)', \tilde{y} = x'\hat{\beta} = 209.04, t_{5,0.025} = 2.5706, se(\tilde{y}) = \hat{\sigma}\sqrt{1 + x'(X'X)^{-1}x} = 125.456,$

The 95% P.I. for the new observation is: $\tilde{y} \pm t_{5,0.025}\hat{\sigma}\sqrt{1 + x'(X'X)^{-1}x} = [-112.557, 532.437]$;

(c)

$x = (1,12,-2)', \hat{y} = x'\hat{\beta} = 209.04, F_{3,5,0.01} = 12.06, se(\hat{y}) = \hat{\sigma}\sqrt{x'(X'X)^{-1}x} = 125.250,$

The 99% C.B. for the regression line at (x_1, x_2) is:

$\hat{y} \pm \hat{\sigma}\sqrt{(p+1)F_{3,5,0.01}x'(X'X)^{-1}x} = [-543.436, 963.316]$;

(d)

$$8\beta_0^2 + 262\beta_1^2 + 265\beta_2^2 + 80\beta_0\beta_1 + 82\beta_0\beta_2 + 526\beta_1\beta_2 - 410.016\beta_0 - 2296.112\beta_1 - 2316.112\beta_2 + 4756.01 \leq 0.$$

2. Solution:

Consider the following regression models:
1: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$
2: $\hat{Y} = \hat{a}_0 + \hat{a}_1 X_2$
3: $\hat{X}_1 = \hat{b}_0 + \hat{b}_1 X_2$
From model 2 and model 3, we can compute $\hat{\alpha}_1$ from the model $\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_2$ by using their residuals.
Then we have $\hat{\alpha}_1 = \frac{S_{X1,Y} - \frac{S_{X1,X2}S_{X2,Y}}{S_{X2,X2}}}{S_{X1,X1} - \frac{S^2_{X1,X2}}{S_{X2,X2}}}$ and $\hat{\beta}_1 = \frac{S_{X1,Y}}{S_{X1,X1}}$.

where
$S_{X_1,Y} = \sum(X_{1i} - \bar{X}_1)(Y_i - \bar{Y})$,
$S_{X_1,X_2} = \sum(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)$,
$S_{X_2,Y} = \sum(X_{2i} - \bar{X}_2)(Y_i - \bar{Y})$,
$S_{X_2,X_2} = \sum(X_{2i} - \bar{X}_2)^2$,
$S_{X_1,X_1} = \sum(X_{1i} - \bar{X}_1)^2$.

(i) Therefore, $\hat{\beta}_1 = \hat{\alpha}_1$ if $S_{X1,X2} = 0$. For example, $Y = 1, 2, 3, X1 = 1, 3, 9, X2 = 2, 2, 2$.
(ii) $\hat{\alpha}_1/\hat{\beta}_1 < 0$ For example, $Y = 1, 2, 3, X1 = 1, 2, 2, X2 = 6, 4, 3$.

Note: demonstrate $\sum \hat{e}_{i\{Y|X_2\}}\hat{e}_{i\{X_1|X_2\}} = S_{X_1,Y} - \frac{S_{X_1,X_2}S_{X_2,Y}}{S_{X_2,X_2}}$.

$$\sum \hat{e}_{i\{Y|X_2\}}\hat{e}_{i\{X_1|X_2\}} = \hat{e}'_{\{Y|X_2\}}\hat{e}_{\{X_1|X_2\}},$$

$$\hat{e}_{\{Y|X_2\}} = Y - \hat{a}_0\mathbf{1} - \hat{a}_1 X2 = Y - (\bar{Y} - \frac{S_{X_2,Y}}{S_{X2,X2}}\bar{X}_2)\mathbf{1} - \frac{S_{X_2,Y}}{S_{X2,X2}}X2,$$

$$\hat{e}_{\{X_1|X_2\}} = Y - (\bar{X}_1 - \frac{S_{X_1,X_2}}{S_{X2,X2}}\bar{X}_2)\mathbf{1} - \frac{S_{X_1,X_2}}{S_{X2,X2}}X2,$$

$$\begin{aligned}
\hat{e}'_{\{Y|X_2\}}\hat{e}_{\{X_1|X_2\}} &= (Y' - (\bar{Y}\mathbf{1})' - \frac{S_{X_2,Y}}{S_{X2,X2}}(X_2 - \bar{X}_2)')(X_1 - \bar{X}_1\mathbf{1} - \frac{S_{X_1,X_2}}{S_{X2,X2}}(X_2 - \bar{X}_2)) \\
&= (Y' - (\bar{Y}\mathbf{1})')(X_1 - \bar{X}_1\mathbf{1}) - 2\frac{S_{X_1,X_2}S_{X_2,Y}}{S_{X_2,X_2}} + \frac{S_{X_1,X_2}S_{X_2,Y}}{S_{X_2,X_2}} \\
&= S_{X_1,Y} - \frac{S_{X_1,X_2}S_{X_2,Y}}{S_{X_2,X_2}}.
\end{aligned}$$

3. Solution:

(i)

$$X=\begin{bmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{bmatrix}, Y=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix}$$

$\hat{\alpha}=(X'X)^{-1}X'Y, X'X=\begin{bmatrix}n & \sum x_i\\ \sum x_i & \sum x_i^2\end{bmatrix}, (X'X)^{-1}=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{bmatrix}\sum x_i^2 & -\sum x_i\\ -\sum x_i & n\end{bmatrix}$

$X'Y=\begin{bmatrix}\sum y_i\\ \sum x_i y_i\end{bmatrix}$

$\therefore \hat{\alpha}=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{bmatrix}\sum x_i^2\sum y_i-\sum x_i\sum x_i y_i\\ -\sum x_i\sum y_i+n\sum x_i y_i\end{bmatrix}=\begin{bmatrix}\bar{Y}-\bar{X}SXY/SXX\\ SXY/SXX\end{bmatrix}.$

(ii)

$$E(Y)=X_2\beta, Var(Y)=\sigma_0^2 I \text{ where } \beta=(\beta_0,\beta_1)' \text{ and } X_2=\begin{bmatrix}1 & x_1^2\\ 1 & x_2^2\\ \vdots & \vdots\\ 1 & x_n^2\end{bmatrix}.$$

$\therefore E(\hat{\alpha})=(X'X)^{-1}X'E(Y)=(X'X)^{-1}X'X_2\beta.$

$Var(\hat{\alpha})=(X'X)^{-1}X'Var(Y)X(X'X)^{-1}=\sigma_0^2(X'X)^{-1}.$

(iii)

$E(\hat{\alpha})=(X'X)^{-1}X'E(Y)=(X'X)^{-1}X'X_2\beta$

$X'X_2=\begin{bmatrix}n & \sum x_i^2\\ \sum x_i & \sum x_i^3\end{bmatrix}.$

$(X'X)^{-1}X'X_2=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{bmatrix}n\sum x_i^2-(\sum x_i)^2 & (\sum x_i^2)^2-(\sum x_i)(\sum x_i^3)\\ 0 & n\sum x_i^3-(\sum x_i)(\sum x_i^2)\end{bmatrix}$

$\therefore E(\hat{\alpha})=\begin{bmatrix}E(\hat{\alpha}_0)\\ E(\hat{\alpha}_1)\end{bmatrix}=\frac{1}{n\sum x_i^2-(\sum x_i)^2}\begin{bmatrix}(n\sum x_i^2-(\sum x_i)^2)\beta_0+((\sum x_i^2)^2-(\sum x_i)(\sum x_i^3))\beta_1\\ (n\sum x_i^3-(\sum x_i)(\sum x_i^2))\beta_1\end{bmatrix}$

$=\begin{bmatrix}\beta_0+\frac{(\sum x_i^2)^2-(\sum x_i)(\sum x_i^3)}{n\sum x_i^2-(\sum x_i)^2}\beta_1\\ \frac{n\sum x_i^3-(\sum x_i)(\sum x_i^2)}{n\sum x_i^2-(\sum x_i)^2}\beta_1\end{bmatrix}$

(iv)

When $n\to\infty$,

$E(\hat{\alpha}_0)=\beta_0+\frac{(\sigma_x^2)^2-\mu_x\eta_x}{\sigma_x^2-\mu_x^2}\beta_1$ and $E(\hat{\alpha}_1)=\frac{\eta_x-\mu_x\sigma_x^2}{\sigma_x^2-\mu_x^2}\beta_1.$

$(\hat{\alpha_0},\hat{\alpha_1})$ is not a consistent estimator of $(\hat{\beta}_0,\hat{\beta}_1)$.

4. Solution:

(i) $w_i=\frac{1}{x_i^2}$

(ii) $w_i=\frac{1}{\sqrt{x_i}}$

(iii) $w_i = \frac{1}{x_i}$
(iv) $w_i = n_i$

5. Solution:

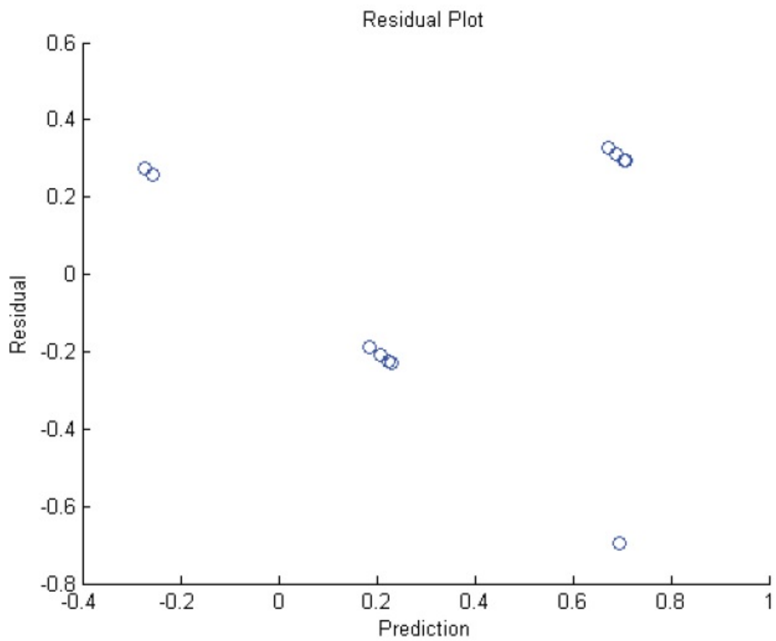
(i)

β_i can be interpreted as the rate of change of Y with respect to X_i.
Therefore, β_1 can be interpreted as the increase in chance of being hired by increasing one year of education.
β_2 can be interpreted as the increase in chance of being hired by increasing one year of experience.
β_3 can be interpreted as the increase in chance of being hired if you are a male.
β_0 can be interpreted as the baseline chance of a female being hired without considering education and experience.

(ii)

$\hat{Y} = -1.2245 + 0.2435X_1 - 0.0056X_2 + 0.4705X_3$.

(iii)



The residual plot seems not to be okay as it doesn't look like a null plot.

(iv)

$H_0: E(Y|X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ v.s. $H_1: E(Y|X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
$SSE_{H_0} = 1.7339, df_{H_0} = 12 - 3 = 9,$
$SSE_{H_1} = 1.2245, df_{H_1} = 12 - 4 = 8,$
obs.T.S.= $\frac{(1.7339-1.2245)/(9-8)}{1.2245/8} = 3.3281$
$F_{1,8,0.05} = 5.3177 obs.T.S. < F_{1,8,0.05}$ $\therefore$ The null hypothesis is not rejected at $\alpha = 0.05$.

Code:

```
y<-c(0,0,1,1,0,1,0,0,0,1,0,0)
x1<-c(6,4,6,6,4,8,4,4,6,8,4,8)
x2<-c(2,6,0,3,1,3,2,4,1,9,2,5)
x3<-c(0,1,1,1,0,0,1,0,0,0,1,0)
fit<-lm(y∼x1+x2+x3)
plot(fit$fitted.values,fit$residuals)
fit0<-lm(y∼x1+x2)
sum(fit$residualsˆ2)
sum(fit0$residualsˆ2)
```