STAT 3008: Applied Regression Analysis
2019-20 Term 2
Assignment #2 Solutions

Problem 1:

(a)

Coefficient Table				
Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	-23.4325	12.74	-1.839	0.0824
X	1.2713	0.1528	8.320	1.396E-07

ANOVA Table					
Source	df	SS	MS	F	p-value
Regression	1	1848.76	1848.760	69.222	1.396E-07
Residuals	18	480.74	26.708		
Total	19	2329.50			

(b) R2 = SSreg/SStotal = 1848.76/2329.5 = 79.36%, $r = (0.7936)^{1/2} = 0.8909$

(c) **Hypotheses**: H_0: $\beta_o = -10$ vs H_1: $\beta_o \neq -10$

Test Statistic: $t_0 = (-23.4325-(-10))/12.74 = -1.054$

Decision: Since p-value = $2\Pr(t_{18}>1.054) = 2(0.1529131) = 0.3058 > \alpha = 0.05$, we do not reject H_0 at $\alpha = 0.05$.

Conclusion: We do not have sufficient evidence that β_o is different from -10.0.

Problem 2:

(a)
$$\begin{aligned}
E(\hat{\mathbf{Y}}'\hat{\mathbf{Y}}) &= tr\left(E\left(\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right]'\left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right]\right)\right) \\
&= tr(E(\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})) \\
&= tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}\mathbf{Y}')) \\
&= tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'+\mathbf{e}\mathbf{e}'+2\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}')) \\
&= tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'+\sigma^2\mathbf{I_n}+\mathbf{0})) \\
&= tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}') + tr(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I_n}) + 0 \\
&= \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + tr(\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) = \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (p+1)\sigma^2
\end{aligned}$$

(b) $E(\mathbf{Y}'\mathbf{Y}) = \mathrm{tr}(E(\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}+\mathbf{e}'\mathbf{e}+2\mathbf{X}'\boldsymbol{\beta}'\mathbf{e})) = \mathrm{tr}(\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}+\sigma^2\mathbf{I_n}+0) = \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}+\mathrm{tr}(\sigma^2\mathbf{I_n}) = \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}+n\sigma^2$

Hence,

$$\sum_{i=1}^{n} E(y_i^2) = E(\mathbf{Y}'\mathbf{Y}) = \beta XX\beta + (p+1)\sigma^2 + (n-p-1)\sigma^2 = E(\hat{\mathbf{Y}}'\hat{\mathbf{Y}}) + E(\hat{\mathbf{e}}'\hat{\mathbf{e}}) = \sum_{i=1}^{n} E(\hat{y}_i^2) + \sum_{i=1}^{n} E(\hat{e}_i^2).$$

Problem 3:

(a) From the R Codes, SYY=716.8889, RSS=200.2901, SSreg=516.5988, $\hat{\sigma}^2$=33.38168. R^2= 0.7206

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 11.6819 \\ 0.323155 \\ 2.15267 \end{pmatrix},\ \hat{\mathbf{Y}} = \begin{pmatrix} 19.10941 \\ 33.96438 \\ 24.06107 \\ 19.10941 \\ 11.35878 \\ 29.01272 \\ 33.64122 \\ 21.58524 \\ 14.15776 \end{pmatrix},\ \hat{\mathbf{e}} = \begin{pmatrix} 1.89059 \\ -8.96438 \\ -3.06107 \\ 4.89059 \\ -2.35878 \\ 6.98728 \\ 2.35878 \\ 2.41476 \\ -4.15776 \end{pmatrix},\ \widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} 13.0524 & 4.8983 & -6.7103 \\ 4.8983 & 21.4617 & -21.4900 \\ -6.7103 & -21.4900 & 21.9147 \end{pmatrix}$$

(b) $\mathbf{x}_* = (1,-1,1), \tilde{y} = \mathbf{x}_*\hat{\boldsymbol{\beta}} = 13.51145$, $t_{6,0.025}$ =2.4469, $\mathrm{sepred}(y\,|\,\mathbf{x}_*) = \hat{\sigma}\sqrt{1+x_*'(X'X)^{-1}x_*} = 10.46773$

A 95% PI for the response is $\tilde{y} \pm t_{6,0.025}\mathrm{sepred}(y\,|\,\mathbf{x}_*)$ =(-12.10217 , 39.12507)

(c) **Hypotheses** H_o: E(Y|**X**) = $\beta_o + \beta_1 x_1$ vs H_1: E(Y|**X**) = $\beta_o + \beta_1 x_1 + \beta_2 x_2$

Source	df	SS	MS	F_0	p-value
Regression	1	0.16	0.16	0.00479	0.9470
Residual	6	200.29	33.38		
Total	7	200.45			

Decision: Since p-value = 0.9470 > α = 0.05, we do not reject H_o at α = 0.05.
Conclusion: We do not have sufficient evidence that E(Y|**X**) = $\beta_o + \beta_1 x_1 + \beta_2 x_2$ is the appropriate mean function vs E (Y|**X**) = $\beta_o + \beta_1 x_1$.

Note on Problem 3: Since the vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are close to each other. Multicollinearity exists and therefore the diagonal elements of $\hat{\mathrm{V}}\mathrm{ar}(\hat{\boldsymbol{\beta}})$ are all large. Hence, (i) the PI in part(b) is WIDE, and (ii) the ANOVA table in part (c) suggests that it's preferred to use only one EV (namely $\mathbf{X}_2$) instead of both.

<u>R Codes for Problem #3</u>

Problem 3(a)

```
y<-c(21,25,21,24,9,36,36,24,10); x1<-c(3,9,5,3,-1,7,8,4,1)
x2<-c(3,9,5,3,0,7,9,4,1)
n<-length(y); x<-rep(1,n); X<-cbind(x,x1,x2)
beta.hat<-solve(t(X)%*%X)%*%t(X)%*%y; beta.hat
yhat<-X%*%beta.hat; yhat
res<-y-yhat; res
SYY<-sum(y^2)-n*mean(y)^2; SYY
RSS <- as.numeric(t(y)%*%y-t(y)%*%X%*%solve(t(X)%*%X)%*%t(X)%*%y); RSS
SSreg<-SYY-RSS;SSreg
sigma2.hat<-RSS/(n-2-1); sigma2.hat
var.hat.beta.hat<-sigma2.hat*solve(t(X)%*%X); var.hat.beta.hat
R2<-1-RSS/SYY; R2
```

Problem 3(b)

```
xstar<-c(1,-1,1)
xstar%*%beta.hat
xstar%*%beta.hat+c(-1,1)*qt(0.975,length(y)-2-1)*sqrt(sigma2.hat)*sqrt(1+t(xstar)%*%solv
e(t(X)%*%X)%*%xstar)
```

Problem 3(c) ### (NOT Required – but the ANOVA function will provide the answers right away)

```
fit0<-lm(y~x2)
anova(fit0)
Analysis of Variance Table
Response: y
           Df  Sum Sq   Mean Sq   F value Pr(>F)
x2          1  516.44   516.44    18.035  0.003809 **
Residuals   7  200.45   28.64
```

```
fit1<-lm(y~x1+x2)
anova(fit0,fit1)
Analysis of Variance Table
Model 1: y ~ x2
Model 2: y ~ x1 + x2
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      7 200.45
2      6 200.29  1   0.16243 0.0049 0.9467
```

Problem 4: (a) $\mathbf{X'X} = \begin{bmatrix} n & 0 & 0 \\ 0 & SUU & 0 \\ 0 & 0 & SVV \end{bmatrix}$. Hence

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \\ \hat\beta_2 \end{bmatrix} = (\mathbf{X'X})^{-1}\mathbf{X'Y} = \begin{bmatrix} n & 0 & 0 \\ 0 & SUU & 0 \\ 0 & 0 & SVV \end{bmatrix}^{-1} \begin{bmatrix} \sum y_i \\ SUY \\ SVY \end{bmatrix} = \begin{bmatrix} \bar y \\ SUY/SUU \\ SVY/SVV \end{bmatrix}$$

(b) $\hat{\boldsymbol{\alpha}} = \begin{bmatrix} \hat\alpha_0 \\ \hat\alpha_1 \end{bmatrix} = \begin{bmatrix} \bar y - SUY/SUU(\bar x) \\ SUY/SUU \end{bmatrix} = \begin{bmatrix} \bar y \\ SUY/SUU \end{bmatrix}$, which are the same as those in part (a).

Problem 5:

(a) Since $\hat{\boldsymbol{\alpha}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$, $E(\hat{\boldsymbol{\alpha}}) = (\mathbf{X'X})^{-1}\mathbf{X'}E(\mathbf{Y}) = (\mathbf{X'X})^{-1}\mathbf{X'X_2\boldsymbol{\beta}}$

(b) Since $\mathbf{X'X}_2 = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar x \\ n\bar x & n\overline{x^2} \end{bmatrix}$ and $\mathbf{X'X}_2 = \begin{bmatrix} n & \sum x_i^2 \\ \sum x_i & \sum x_i^3 \end{bmatrix} = \begin{bmatrix} n & n\overline{x^2} \\ n\bar x & n\overline{x^3} \end{bmatrix}$

$$(\mathbf{X'X})^{-1}\mathbf{X'X_2\boldsymbol{\beta}} = \frac{1}{n(\overline{x^2}-\bar x^2)}\begin{bmatrix} \overline{x^2} & -\bar x \\ -\bar x & 1 \end{bmatrix}\begin{bmatrix} n & n\overline{x^2} \\ n\bar x & n\overline{x^3} \end{bmatrix}\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \frac{1}{\overline{x^2}-\bar x^2}\begin{bmatrix} (\overline{x^2}-\bar x^2)\beta_0 + ((\overline{x^2})^2 - \bar x\overline{x^3})\beta_1 \\ (\overline{x^3}-\bar x\overline{x^2})\beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \dfrac{(\overline{x^2})^2 - \bar x\overline{x^3}}{\overline{x^2}-\bar x^2}\beta_1 \\ \left(\dfrac{\overline{x^3}-\bar x\overline{x^2}}{\overline{x^2}-\bar x^2}\right)\beta_1 \end{bmatrix} = \begin{bmatrix} E(\hat\alpha_0) \\ E(\hat\alpha_1) \end{bmatrix}$$

(c) As $n \to \infty$,

$$E(\hat\alpha_0) = \beta_0 + \frac{(\overline{x^2})^2 - \bar x\overline{x^3}}{\overline{x^2}-\bar x^2}\beta_1 \to \beta_0 + \frac{\sigma_x^4}{\sigma_x^2}\beta_1 = \beta_0 + \sigma_x^2\beta_1 \neq \beta_0$$

$$E(\hat\alpha_1) = \left(\frac{\overline{x^3}-\bar x\overline{x^2}}{\overline{x^2}-\bar x^2}\right)\beta_1 \to \frac{\kappa_x\sigma_x^3}{\sigma_x^2}\beta_1 = \kappa_x\sigma_x\beta_1 \neq \beta_1$$