Assignment 2: Solution

1. (a)

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e \rightarrow y' = \beta_0' + \beta_1 x_1' + \beta_2 x_2' \\
\text{Under Centered Model} & \\
\left(\boldsymbol{X_c}^\top \boldsymbol{X_c}\right)^{-1} &= \begin{pmatrix} 0.05 & 0 & 0 \\ 0 & 0.039152 & -0.000796 \\ 0 & -0.000796 & 0.001617 \end{pmatrix} \\
\boldsymbol{X_c}^\top \boldsymbol{Y} &= \begin{pmatrix} 219 \\ 60.9 \\ 606.35 \end{pmatrix} \\
\hat{\boldsymbol{\beta}}' &= \begin{pmatrix} 10.95 \\ 1.9017 \\ 0.9320 \end{pmatrix} \\
\hat{\boldsymbol{\beta}} &= \begin{pmatrix} 1.3809 \\ 1.9017 \\ 0.9320 \end{pmatrix} \\
\hat{y} &= 1.3809 + 1.9017 x_1 + 0.9320 x_2
\end{aligned}
$$

(b)

$$
\begin{aligned}
RSS &= \boldsymbol{Y}^\top \boldsymbol{Y} - \hat{\boldsymbol{\beta}}'^\top \boldsymbol{X_c}^\top \boldsymbol{Y} = 12.0183 \\
\hat{\sigma}^2 &= \frac{RSS}{n - p'} = 12.0183/(20-3) = 0.7070
\end{aligned}
$$

(c)

Source of variation	Sum of Squares	D.F.	Mean Square	F value
Regression	680.9317	2	340.4659	481.5641
Residual	12.0183	17	0.707	
Total	692.95	19		

$F_{obs} = 481.5641 > F_{0.05,2,17} = 3.59$
So reject null hypothesis.

(d) i. t test:
s.e. of $(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{0.7070 * (0.039152 + 0.001617 - 2 * (-0.000796)} = 0.1731$
$\text{t} = \dfrac{\hat{\beta}_1 - \hat{\beta}_2}{s.e.of \hat{\beta}_1 - \hat{\beta}_2} = \dfrac{1.9017 - 0.9320}{0.1731} = 5.6020$
Critical value $= t_{0.25,17} = 2.11$
Since$|t_{obs}| > t_{17,0.025}$, we reject the null hypothesis.

ii. "Increase in Regression Sum of Squares". The reduced model is $y = \beta_0 + \beta_2(x_1 + x_2) + e_i$
Take $x' = x_1 + x_2$, and under reduced model,
$\hat{\beta}_2 = \frac{S_{x'y}}{S_{x'x'}} = \frac{S_{x_1y} + S_{x_2y}}{S_{x_1x_1} + S_{x_2x_2} + 2S_{x_1x_2}}$ $RSS_{reduced} = S_{yy} - \hat{\beta}_2^2 S_{x'x'} = 34.0931$
Increase in Regression Sum of Squares $= RSS_{reduced} - RSS_{full} = 34.0931 - 12.0183 = 22.0784$
$F_{obs} = \frac{RSS_{reduced} - RSS_{full}}{1 * \hat{\sigma}^2} = \frac{22.0784}{1 * 0.7070} = 31.3824 > F_{1,17,0.5} = 4.45$
So reject the null hypothesis.

iii. F test for testing $H_0 : C\beta = d$. $C =$ [0 1 -1]
d = 0
C$\hat{\beta}$ - d = 0.9697
$[C(X_c^T X_c)^{-1} C^T]$ =0.042361
$[C(X_c^T X_c)^{-1} C^T]^{-1}$ =23.6066
Value of test statistic $= \dfrac{(0.9697)(0.042361)^{-1}(0.9697)/1}{0.7070} = 31.3824$

Critical value $= F_{0.05,1,17} = 4.45$
Since $F_{obs} > F_{1,3,0.05}$, Reject the null hypothesis.

2. Ten men were studied during a maximal exercise treadmill test. The dependent and independent variables are: $y = \text{VO}_{2max}$, $x_1 = \text{weight}$, $x_2 = \text{HR}_{max}$, $x_3 = \text{SV}_{max}$. The table of parameter estimates, standard error and covariance matrix is given below:

			Covariance Matrix			
Variable	$\hat{\beta}_i$	St. Error	Intercept	x_1	x_2	x_3
Intercept	-1.4545	22.2144	493.4780	-2.1663	-1.5222	-0.4450
x_1	-0.6985	0.1281	-2.1663	0.01641	0.004525	0.0001291
x_2	0.2895	0.07810	-1.5222	0.004525	0.006099	0.0008443
x_3	0.4481	0.05110	-0.4450	0.0001291	0.0008443	0.002611

(a) Find the t-value for testing the statistical significance of $\beta_3 = 0$. Do we reject $\beta_3 = 0$ at the 5% significance level?
Solution:

$t = \dfrac{0.4481}{0.0511} = \mathbf{8.7691}$
critical value: $t_{0.025,6} = 2.447 \quad \Rightarrow \quad$ reject $\beta_3 = 0$

(b) Construct a 95% confidence interval for β_1.
Solution:
95% C.I. for β_1: $-0.6985 \pm t_{0.025,6} * 0.1281 \quad \Rightarrow \quad$ **(-1.0120,-0.3850)**

(c) Test whether the ratio of the regression coefficient of x_2 to that of x_3 is equal to 0.5 at the 5% significance level. Write down your test statistic, critical value and your conclusions clearly.
Solution:
H_0: $\dfrac{\beta_2}{\beta_3} = 0.5 \quad \Rightarrow \quad H_0\text{: } 2\beta_2 - \beta_3 = 0$
$pt. = 2 \times 0.2895 - 0.4481 = 0.1309$
$Var(2\hat{\beta}_2 - \hat{\beta}_3) = 4Var(\hat{\beta}_2) + Var(\hat{\beta}_3) - 4Cov(\hat{\beta}_2, \hat{\beta}_3)$
$s.e.(2\hat{\beta}_2 - \hat{\beta}_3) = \sqrt{4 \times 0.00781^2 + 0.0511^2 - 4 \times 0.0008443} = 0.02363^2 = 0.1538$
$t = \dfrac{0.1309}{0.1538} = \mathbf{0.8511} < 2.447 = t_{0.025,6} \quad \Rightarrow \quad$ can't reject H_0.

(d) Fill in the missing values in the analysis of variance table below. Is the regression significant at the 5% significance level?

Source	Sum of Squares	D.F.	Mean Squares	F value
Regression	**1249.1073**	**3**	**416.3691**	**44.6360**
Residual	55.9687	**6**	**9.3281**	—
Total	1305.0760	**9**	—	—

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$
$F = \mathbf{44.6360} > F_{0.05,3,6} \quad \Rightarrow \quad$ reject H_0

3. (a) It is the centered model in multiple regression model, with 3 unknown parameters.

For Model A:$(X_A^T X_A)^{-1}$= $\begin{pmatrix} n & 0 & 0 \\ 0 & S_{xx} & S_{xz} \\ 0 & S_{xz} & S_{zz} \end{pmatrix}^{-1} = \begin{pmatrix} 1/n & 0 & 0 \\ 0 & \frac{S_{zz}}{S_{xx}S_{zz}-S_{xz}^2} & \frac{-S_{xz}}{S_{xx}S_{zz}-S_{xz}^2} \\ 0 & \frac{-S_{xz}}{S_{xx}S_{zz}-S_{xz}^2} & \frac{-S_{xx}}{S_{xx}S_{zz}-S_{xz}^2} \end{pmatrix}$

The least square estimate $\hat{\beta}$= $\begin{pmatrix} 0 & \frac{S_{zz}}{S_{xx}S_{zz}-S_{xz}^2} & \frac{-S_{xz}}{S_{xx}S_{zz}-S_{xz}^2} \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum(x_i - \bar{x})y_i \\ \sum(z_i - \bar{z})y_i \end{pmatrix} = \frac{S_{zz}S_{xy}-S_{xz}S_{yz}}{S_{xx}S_{zz}-S_{xz}^2}$

Taking the variance on the estimator simply yields Var($\hat{\beta}$)=$\frac{S_{zz}\sigma^2}{S_{xx}S_{zz}-S_{xz}^2}$

(b) The model is the same as the simple linear regression in centered form.

For Model B:$(X_B^T X_B)^{-1}$= $\begin{pmatrix} n & 0 \\ 0 & S_{xx} \end{pmatrix}^{-1} = \begin{pmatrix} 1/n & 0 \\ 0 & 1/S_{xx} \end{pmatrix}$

The least square estimate $\tilde{\beta}$= $\begin{pmatrix} 0 & 1/S_{xx} \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum(x_i - \bar{x})y_i \end{pmatrix} = S_{xy}/S_{xx}$

Taking the variance on the estimator simply yields $Var\tilde{\beta}$=$\frac{\sigma^2}{S_{xx}}$

$S_{xz}^2 \geq 0 <=> \frac{S_{zz}}{S_{xx}S_{zz}-Sxz^2} \geq \frac{1}{S_{xx}}$ and multiplying by σ^2 then the desire result follows.

$Var\tilde{\beta} \leq Var\hat{\beta}$ and the equality holds if and only if $S_{xz}^2 = 0$

4.

$$\frac{y_i - \hat{y}_i}{s\sqrt{1 - \frac{1}{n} - \frac{(x_i-\bar{x})^2}{S_{xx}}}} \qquad \text{where } s = \sqrt{\sum_{i=1}^{n} \hat{e}_i^2/(n-2)}, \quad \hat{e}_i = y_i - \hat{y}_i$$

$$Var(y_i - \hat{y}_i) = \sigma^2 \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right]$$

(a)

$$\begin{aligned} \sum_{i=1}^{n} Var(y_i - \hat{y}_i)/\sigma^2 &= \sum_{i=1}^{n} \sigma^2 \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}\right]/\sigma^2 \\ &= n - 1 - \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{S_{xx}} \\ &= n - 1 - 1 \\ &= n - 2 \end{aligned}$$

(b) $\hat{e}_i$ & $\sum_{i=1}^{n} \hat{e_i}^2 = (n-2)s^2$ are not independent.

$\therefore$ Studentized residual does not follow t-distribution.

5. Fitted model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$

Residual mean squares = $\frac{\text{Res.S.S.}}{n-3}$ where Res.S.S. = $\underset{\sim}{Y}^T(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{Y}$, $\underset{\sim}{X}$ is a $n \times 3$ matrix and $\underset{\sim}{\beta}^T = (\beta_0, \beta_1, \beta_2)$

In fact, $\beta_2 = 0$, i.e. E($\underset{\sim}{Y}$) = $\underset{\sim}{\mu} = \underset{\sim}{X}_1\underset{\sim}{\beta}_1$ and Var = $\sigma^2\underset{\sim}{I}$, where $\underset{\sim}{X}_1$ is a $n \times 2$ matrix and $\underset{\sim}{\beta}_1^T = (\beta_0, \beta_1)$.

Then,

$$\begin{aligned} \text{E(Res.S.S.)} &= \sigma^2\text{trace}(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T) + \underset{\sim}{\mu}^T(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{\mu} \\ &= (n-3)\sigma^2 + \underset{\sim}{\beta}_1^T(\underset{\sim}{X}_1^T\underset{\sim}{X}_1 - \underset{\sim}{X}_1^T\underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T\underset{\sim}{X}_1)\underset{\sim}{\beta}_1 \end{aligned}$$

$$
\begin{aligned}
\underset{\sim}{X}_1^T \underset{\sim}{X}(\underset{\sim}{X}^T \underset{\sim}{X})^{-1} \underset{\sim}{X}^T \underset{\sim}{X}_1 &= \begin{pmatrix} n & 0 & 0 \\ 0 & S_{X_1X_1} & S_{X_1X_2} \end{pmatrix} \begin{pmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{S_{X_2X_2}}{D} & -\frac{S_{X_1X_2}}{D} \\ 0 & -\frac{S_{X_1X_2}}{D} & \frac{S_{X_1X_1}}{D} \end{pmatrix} \begin{pmatrix} n & 0 & 0 \\ 0 & S_{X_1X_1} & S_{X_1X_2} \end{pmatrix}^T \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} n & 0 \\ 0 & S_{X_1X_1} \\ 0 & S_{X_1X_2} \end{pmatrix} \\
&= \underset{\sim}{X}_1^T \underset{\sim}{X}_1
\end{aligned}
$$

where $D = S_{X_1X_1} S_{X_2X_2} - S_{X_1X_2}^2$. Thus, residual mean squares from the fitted model is still an unbiased estimator of σ^2.