STAT 524
HW5

Satoshi Ido
34788706
10/24/2023

7.2

7.2. Given the data

z_1	10	5	7	19	11	18
z_2	2	3	3	6	7	9
y	15	9	3	25	7	13

fit the regression model

$$Y_j = \beta_1 z_{j1} + \beta_2 z_{j2} + \varepsilon_j, \qquad j = 1, 2, \ldots, 6.$$

to the *standardized* form (see page 412) of the variables y, z_1, and z_2. From this fit, deduce the corresponding fitted regression equation for the original (not standardized) variables.

$$\bar{Z}_1 = \frac{1}{6}(10 + 5 + 7 + 19 + 11 + 18) = 11.667$$

$$\bar{Z}_2 = \frac{1}{6}(2 + 3 + 3 + 6 + 7 + 9) = 5$$

$$\bar{y} = \frac{1}{6}(15 + 9 + 3 + 25 + 7 + 13) = 12$$

$$Sd_{Z_1} = \frac{1}{\sqrt{5}}\sqrt{(10 - 11.667)^2 + \cdots + (18 - 11.667)^2} = 5.715$$

$$Sd_{Z_2} = \frac{1}{\sqrt{5}}\sqrt{(2-5)^2 + \cdots + (9-5)^2} = 2.757$$

$$Sd_y = \frac{1}{\sqrt{5}}\sqrt{(15-12)^2 + \cdots + (13-12)^2} = 7.668$$

Standardized $\quad Z_{j1}^* = \dfrac{Z_{j1} - \bar{Z}_1}{Sd_{Z_1}} \qquad j = 1, 2, \ldots, 6$

$$Z_{j2}^* = \frac{Z_{j2} - \bar{Z}_2}{Sd_{Z_2}} \qquad j = 1, 2, \ldots, 6$$

$$y_j^* = \frac{y_j - \bar{y}}{Sd_y} \qquad j = 1, 2, \ldots, 6$$

$$\therefore \quad Z_{11} = -0.292 \qquad Z_{12} = -1.088 \qquad Z_{1y} = 0.391$$
$$Z_{21} = -1.166 \qquad Z_{22} = -0.725 \qquad Z_{2y} = -0.391$$
$$Z_{31} = -0.816 \qquad Z_{32} = -0.725 \qquad Z_{3y} = -1.174$$
$$Z_{41} = 1.283 \qquad Z_{42} = 0.363 \qquad Z_{4y} = 1.695$$
$$Z_{51} = -0.117 \qquad Z_{52} = 0.725 \qquad Z_{5y} = -0.652$$
$$Z_{61} = 1.108 \qquad Z_{62} = 1.451 \qquad Z_{6y} = 0.130$$

$$y_j^* = \beta_1 Z_{j1}^* + \beta_2 Z_{j2}^* + \varepsilon_j \qquad j = 1, 2, \cdots, 6$$

$$Z^{*T} = \begin{bmatrix} -0.292 & -1.166 & -0.816 & 1.283 & -0.117 & 1.108 \\ -1.088 & -0.725 & -0.725 & 0.363 & 0.725 & 1.451 \end{bmatrix}$$

$$y^{*T} = \begin{bmatrix} 0.391 & -0.391 & -1.174 & 1.695 & -0.652 & 0.130 \end{bmatrix}$$

$$\hat{\beta} = (Z^T Z)^{-1} Z^T y \qquad = \begin{bmatrix} 1.329 \\ -0.787 \end{bmatrix} \quad //$$

$$y_j^* = 1.329\, Z_{j1}^* - 0.787\, Z_{j2}^*$$

$$\frac{y_j - \bar{y}}{sd_y} = 1.329 \frac{Z_{j1} - \bar{Z}_1}{sd_{Z1}} - 0.787 \frac{Z_{j2} - \bar{Z}_2}{sd_{Z2}}$$

$$\frac{y_j - 12}{7.668} = 1.329 \frac{Z_{j1} - 11.667}{5.715} - 0.787 \frac{Z_{j2} - 5}{2.757}$$

$$y_j = 1.783\, Z_{j1} - 2.189\, Z_{j2} + 2.148 \quad //$$

7.3.

7.3. (*Weighted least squares estimators.*) Let

$$\underset{(n\times 1)}{\mathbf{Y}} = \underset{(n\times(r+1))}{\mathbf{Z}}\ \underset{((r+1)\times 1)}{\boldsymbol{\beta}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}$$

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ but $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{V}$, with $\mathbf{V}(n \times n)$ known and positive definite. For $\mathbf{V}$ of full rank, show that the *weighted least squares* estimator is

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Y}$$

If σ^2 is unknown, it may be estimated, unbiasedly, by

$$(n - r - 1)^{-1} \times (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_W)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}_W).$$

Hint: $\mathbf{V}^{-1/2}\mathbf{Y} = (\mathbf{V}^{-1/2}\mathbf{Z})\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\varepsilon}$ is of the classical linear regression form $\mathbf{Y}^* = \mathbf{Z}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^*$, with $E(\boldsymbol{\varepsilon}^*) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}^*\boldsymbol{\varepsilon}^{*\prime}) = \sigma^2\mathbf{I}$. Thus, $\hat{\boldsymbol{\beta}}_W = \hat{\boldsymbol{\beta}}^* = (\mathbf{Z}^{*\prime}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{Y}^*$.

$$Y = Z\beta + \varepsilon$$

$$\cancel{Y^*} = Z^*\beta_w + \varepsilon^*$$

$$\sigma^2 V = \begin{bmatrix} w_1\sigma_1 & & & \\ & w_2\sigma_2 & & \\ & & \ddots & \\ & & & w_n\sigma_n \end{bmatrix}$$

Let $Y^* = V^{-1/2}Y$, $Z^* = V^{-1/2}Z$, and $\varepsilon^* = V^{-1/2}\varepsilon$

$$\rightarrow \quad V^{-1/2}Y = (V^{-1/2}Z)\beta_w + V^{-1/2}\varepsilon$$

$$Y^* = Z^*\beta_w + \varepsilon^* \qquad \text{where } E(\varepsilon^*) = 0 \text{ and } E(\varepsilon^*\varepsilon^{*T}) = \sigma^2 I$$

$$E[Y^*] = \hat{Y}^*$$

$$= E[V^{-1/2}\hat{Y}]$$

$$= (V^{-1/2}Z)\hat{\beta}_w$$

$$\therefore\ Y^* - \hat{Y}^* = V^{-1/2}Y - (V^{-1/2}Z)\hat{\beta}_w$$

$$\varepsilon^* = V^{-1/2}Y - (V^{-1/2}Z)\hat{\beta}_w$$

$$\varepsilon^*\varepsilon^{*T} = (V^{-1/2}Y - (V^{-1/2}Z)\hat{\beta}_w)(V^{-1/2}Y - (V^{-1/2}Z)\hat{\beta}_w)^T$$

$$= (Y - Z\hat{\beta}_w)\, V^{-1} (Y - Z\hat{\beta}_w)^T$$

$$E(\varepsilon^{*T}\varepsilon^*) = \sigma^2 I \sim \sigma^2\chi^2_{n-r-1}$$

$\therefore$ The weighted least squares estimator

$$\hat{\beta}_W = (Z^{*T} Z^*)^{-1} Z^{*T} Y^*$$

$$= (V^{-\frac{1}{2}} Z V^{-\frac{1}{2}} Z)^{-1} (V^{-\frac{1}{2}} Z)^T (V^{-\frac{1}{2}} Y)$$

$$= (Z V^{-1} Z)^{-1} Z^T V^{-1} Y \quad //$$

Given the hint, let $Y^* = V^{-\frac{1}{2}} Y$

$$Z^* = V^{-\frac{1}{2}} Z$$

$$\varepsilon^* = V^{-\frac{1}{2}} \varepsilon$$

From the classical linear regression model, the MLE of β is

$$\hat{\beta}^* = (Z^{*T} Z^*)^{-1} Z^{*T} Y^*$$

$$\therefore \quad \hat{\beta}^* = (V^{-\frac{1}{2}} Z^T V^{-\frac{1}{2}} Z)^{-1} V^{-\frac{1}{2}} Z^T V^{-\frac{1}{2}} Y$$

$$= (Z^T V^{-1} Z)^{-1} Z^T V^{-1} Y$$

given $Y = ZB + \varepsilon$

From the hint, the weighted least squares estimator for β is

$$\hat{\beta}_W = \hat{\beta}^* = (Z^T V^{-1} Z)^{-1} Z^T V^{-1} Y \quad //$$

7.7.

7.7. Suppose the classical regression model is, with rank $(\mathbf{Z}) = r + 1$, written as

$$\underset{(n\times 1)}{\mathbf{Y}} = \underset{(n\times(q+1))}{\mathbf{Z}_1}\ \underset{((q+1)\times 1)}{\boldsymbol{\beta}_{(1)}} + \underset{(n\times(r-q))}{\mathbf{Z}_2}\ \underset{((r-q)\times 1)}{\boldsymbol{\beta}_{(2)}} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}$$

where rank$(\mathbf{Z}_1) = q + 1$ and rank$(\mathbf{Z}_2) = r - q$. If the parameters $\boldsymbol{\beta}_{(2)}$ are identified beforehand as being of primary interest, show that a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}_{(2)}$ is given by

$$(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)})'[\mathbf{Z}_2'\mathbf{Z}_2 - \mathbf{Z}_2'\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Z}_2](\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)}) \le s^2(r-q)F_{r-q,n-r-1}(\alpha)$$

Hint: By Exercise 4.12, with 1's and 2's interchanged,

$$\mathbf{C}^{22} = [\mathbf{Z}_2'\mathbf{Z}_2 - \mathbf{Z}_2'\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Z}_2]^{-1}, \quad \text{where } (\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix}$$

Multiply by the square-root matrix $(\mathbf{C}^{22})^{-1/2}$, and conclude that $(\mathbf{C}^{22})^{-1/2}(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)})/\sigma^2$ is $N(\mathbf{0}, \mathbf{I})$, so that

$$(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)})'(\mathbf{C}^{22})^{-1}(\hat{\boldsymbol{\beta}}_{(2)} - \boldsymbol{\beta}_{(2)}) \text{ is } \sigma^2\chi^2_{r-q}.$$

$$\beta = \begin{bmatrix} \beta_{(1)} \\ \beta_{(2)} \end{bmatrix} = (Z^TZ)^{-1}Z^TY \sim N_{r+1}(\beta, \sigma^2(Z^TZ)^{-1})$$

And β is distributed independently of the residuals

$$\hat{\varepsilon} = Y - Z\hat{\beta}$$

Further, $n\hat{\sigma}^2 = \hat{\varepsilon}^T\hat{\varepsilon} = (n-r-1)s^2 \sim \sigma^2\chi^2_{n-r-1}$

$$Z = [\underset{n\cdot(q+1)}{Z_{(1)}} \ \vdots\ \underset{n\cdot(r-q)}{Z_{(2)}}]$$

$$C^{22} = [Z_2^TZ_2 - Z_2^TZ_1(Z_1^TZ_2)^{-1}(Z_1^TZ_2)]^{-1}$$

where $(Z^TZ)^{-1} = \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{bmatrix}$

Set $V = (C^{22})^{-1/2}(\hat{\beta}_{(2)} - \beta_{(2)})$ and note that $E(V) = 0$

$$\mathrm{Cov}(V) = (C^{22})^{-1/2}\,\mathrm{Cov}(\hat{\beta}_2)\,(C^{22})^{-1/2} = \sigma^2 I$$

$$\Leftrightarrow \frac{\mathrm{Cov}(V)}{\sigma^2} \sim I$$

$\therefore$ We can conclude that

$$E\left((C^{22})^{-1/2}\right) = 0 \qquad E\left(\frac{(C^{22})^{-1/2}(\hat{\beta}_{(2)} - \beta_{(2)})}{\sigma^2}\right) = I$$

$$\therefore V^T V = (\hat{\beta}_{(2)} - \beta_{(2)})^T (C^{22})^{-1} (\hat{\beta}_{(2)} - \beta_{(2)}) \sim \sigma^2 \chi^2_{r-q}$$

Given that $n\hat{\sigma}^2 = \hat{\varepsilon}^T\hat{\varepsilon} = (n-r-1)S^2 \sim \sigma^2\chi^2_{n-r-1}$,

independently of $\hat{\beta}_2$ and, of V as well,

it satisfies the definition of $F_{r,var}$ since the ratio of

two independent $\chi^2_{r,var}$ divided by their df.

$\therefore$ The $(100-\alpha)\%$ confidence region for $\beta_{(2)}$ is given by

$$(\hat{\beta}_{(2)} - \beta_{(2)})^T \left[Z_2^T Z_2 - Z_2^T Z_1 (Z_1^T Z_1)^{-1} Z_1^T Z_2 \right] (\hat{\beta}_{(2)} - \beta_{(2)})$$

$$\leq S^2 (r-q) F_{r-q,\, n-r-1}(\alpha)$$

//

7.9.

7.9. Consider the following data on one predictor variable z_1 and two responses Y_1 and Y_2:

z_1	-2	-1	0	1	2
y_1	5	3	4	2	1
y_2	-3	-1	-1	2	3

Determine the least squares estimates of the parameters in the bivariate straight-line regression model

$$Y_{j1} = \beta_{01} + \beta_{11}z_{j1} + \varepsilon_{j1}$$
$$Y_{j2} = \beta_{02} + \beta_{12}z_{j1} + \varepsilon_{j2}, \qquad j = 1,2,3,4,5$$

Also, calculate the matrices of fitted values $\hat{\mathbf{Y}}$ and residuals $\hat{\boldsymbol{\varepsilon}}$ with $\mathbf{Y} = [\mathbf{y}_1 \mid \mathbf{y}_2]$. Verify the sum of squares and cross-products decomposition

$$\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

$$y_1^T = \begin{bmatrix} 5 & 3 & 4 & 2 & 1 \end{bmatrix}$$

$$y_2^T = \begin{bmatrix} -3 & -1 & -1 & 2 & 3 \end{bmatrix}$$

$$\Rightarrow Y^T = \begin{bmatrix} 5 & 3 & 4 & 2 & 1 \\ -3 & -1 & -1 & 2 & 3 \end{bmatrix}$$

$$Z^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix}$$

$$Z^TZ = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

$$(Z^TZ)^{-1} = \frac{1}{5\cdot 10}\begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{10} \end{bmatrix} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}$$

$$\therefore \hat{\beta}_1 = (Z^TZ)^{-1}Z^Ty_1 = \begin{bmatrix} 3 \\ -0.9 \end{bmatrix}$$

$$\therefore \hat{\beta}_2 = (Z^TZ)^{-1}Z^Ty_2 = \begin{bmatrix} 0 \\ 1.5 \end{bmatrix}$$

$$\Rightarrow \hat{\beta} = \begin{bmatrix} 3 & 0 \\ -0.9 & 1.5 \end{bmatrix}$$

//

$$\hat{Y} = Z\hat{\beta}$$

$$= \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ -0.9 & 1.5 \end{bmatrix} = \begin{bmatrix} 4.8 & -3 \\ 3.9 & -1.5 \\ 3 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3 \end{bmatrix} //$$

$$\hat{\varepsilon} = Y - \hat{Y}$$

$$= \begin{bmatrix} 5 & -3 \\ 3 & -1 \\ 4 & -1 \\ 2 & 2 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} 4.8 & -3 \\ 3.9 & -1.5 \\ 3 & 0 \\ 2.1 & 1.5 \\ 1.2 & 3 \end{bmatrix} = \begin{bmatrix} 0.2 & 0 \\ -0.9 & 0.5 \\ 1 & -1 \\ -0.1 & 0.5 \\ -0.2 & 0 \end{bmatrix} //$$

$$\hat{\varepsilon}^T\hat{\varepsilon} = \begin{bmatrix} 1.9 & -1.5 \\ -1.5 & 1.5 \end{bmatrix}$$

$$\hat{Y}^T Y = \begin{bmatrix} 53.1 & -13.5 \\ -13.5 & 22.5 \end{bmatrix}$$

$$Y^T Y = \begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix}$$

$$\hat{Y}^T Y + \varepsilon^T \varepsilon = \begin{bmatrix} 53.1 & -13.5 \\ -13.5 & 22.5 \end{bmatrix} + \begin{bmatrix} 1.9 & -1.5 \\ -1.5 & 1.5 \end{bmatrix} = \begin{bmatrix} 55 & -15 \\ -15 & 24 \end{bmatrix}$$

$$\therefore Y^T Y = \hat{Y}^T Y + \varepsilon^T \varepsilon //$$

7.12.

7.12. Given the mean vector and covariance matrix of Y, Z_1, and Z_2,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_Y \\ \boldsymbol{\mu}_Z \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{YY} & \boldsymbol{\sigma}'_{ZY} \\ \boldsymbol{\sigma}_{ZY} & \boldsymbol{\Sigma}_{ZZ} \end{bmatrix} = \begin{array}{c|ccc} & Y & Z_1 & Z_2 \\ \hline Y & 9 & 3 & 1 \\ Z_1 & 3 & 2 & 1 \\ Z_2 & 1 & 1 & 1 \end{array}$$

$Z_1 \quad Z_2$

determine each of the following.

(a) The best linear predictor $\beta_0 + \beta_1 Z_1 + \beta_2 Z_2$ of Y

(b) The mean square error of the best linear predictor

(c) The population multiple correlation coefficient

(d) The partial correlation coefficient $\rho_{YZ_1 \cdot Z_2}$

$$\begin{bmatrix} \sigma_{YY \cdot Z_2} & \sigma_{YZ_1 \cdot Z_2} \\ \sigma_{YZ_1 \cdot Z_2} & \sigma_{Z_1Z_1 \cdot Z_2} \end{bmatrix}$$

a) $\beta = \Sigma_{ZZ}^{-1} \sigma_{ZY}$

$$= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$\beta_0 = \mu_Y - \beta^T \mu_Z$

$$= 4 - \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = -4$$

$\therefore$ The best linear predictor is $\beta_0 + \beta^T Z = \underline{-4 + 2Z_1 - Z_2}$ //

b) The mean square error is

$\sigma_{YY} - \sigma_{ZY}^T \Sigma_{ZZ}^{-1} \sigma_{ZY}$

$$= 9 - \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \underline{4} \text{ //}$$

c) The population multiple correlation coefficient is

$$\rho_{Y(Z)} = \sqrt{\frac{\sigma_{ZY}^T \Sigma_{ZZ}^{-1} \sigma_{ZY}}{\sigma_{YY}}} = \sqrt{\frac{5}{9}} = \frac{\sqrt{5}}{3}$$

d) The partial correlation coefficient between Y and Z_1, eliminating Z_2 is

$$\rho_{YZ_1 \cdot Z_2} = \frac{\sigma_{YZ_1 \cdot Z_2}}{\sqrt{\sigma_{YY \cdot Z_2}}\sqrt{\sigma_{Z_1Z_1 \cdot Z_2}}}$$

$$\Sigma = \left[\begin{array}{cc|c} 9 & 3 & 1 \\ 3 & 2 & 1 \\ \hline 1 & 1 & 1 \end{array}\right] = \left[\begin{array}{c|c} \sigma_{YZ_1,YZ_1} & \sigma_{YZ_1Z_2} \\ \hline \sigma_{YZ_1Z_2}^T & \sigma_{Z_2Z_2} \end{array}\right]$$

The covariance of Y and Z_1 given Z_2 is

$$S_{YZ_1YZ_1} - S_{YZ_1Z_2} S_{Z_2Z_2}^{-1} S_{Z_2YZ_1}$$

$$= \begin{bmatrix} 9 & 3 \\ 3 & 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot (1)^{-1} \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 8 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_{YZ_1YZ_1} & \sigma_{YZ_1Z_2} \\ \sigma_{YZ_1Z_2} & \sigma_{Z_2Z_2} \end{bmatrix}$$

$$\therefore \rho_{YZ_1 \cdot Z_2} = \frac{2}{\sqrt{8}\sqrt{1}} = \frac{\sqrt{2}}{2}$$

7.25

うつ藥

7.25. Amitriptyline is prescribed by some physicians as an antidepressant. However, there are also conjectured side effects that seem to be related to the use of the drug: irregular heartbeat, abnormal blood pressures, and irregular waves on the electrocardiogram, among other things. Data gathered on 17 patients who were admitted to the hospital after an amitriptyline overdose are given in Table 7.6. The two response variables are

$$Y_1 = \text{Total TCAD plasma level (TOT)}$$
$$Y_2 = \text{Amount of amitriptyline present in TCAD plasma level (AMI)}$$

The five predictor variables are

$$Z_1 = \text{Gender: 1 if female, 0 if male (GEN)}$$
$$Z_2 = \text{Amount of antidepressants taken at time of overdose (AMT)}$$
$$Z_3 = \text{PR wave measurement (PR)}$$
$$Z_4 = \text{Diastolic blood pressure (DIAP)}$$
$$Z_5 = \text{QRS wave measurement (QRS)}$$

Table 7.6 Amitriptyline Data

y_1 TOT	y_2 AMI	z_1 GEN	z_2 AMT	z_3 PR	z_4 DIAP	z_5 QRS
3389	3149	1	7500	220	0	140
1101	653	1	1975	200	0	100
1131	810	0	3600	205	60	111
596	448	1	675	160	60	120
896	844	1	750	185	70	83
1767	1450	1	2500	180	60	80
807	493	1	350	154	80	98
1111	941	0	1500	200	70	93
645	547	1	375	137	60	105
628	392	1	1050	167	60	74
1360	1283	1	3000	180	60	80
652	458	1	450	160	64	60
860	722	1	1750	135	90	79
500	384	0	2000	160	60	80
781	501	0	4500	180	0	100
1070	405	0	1500	170	90	120
1754	1520	1	3000	180	0	129

Source: See [24].

(a) Perform a regression analysis using only the first response Y_1.
  (i) Suggest and fit appropriate linear regression models.
  (ii) Analyze the residuals.
  (iii) Construct a 95% prediction interval for Total TCAD for $z_1 = 1$, $z_2 = 1200$, $z_3 = 140$, $z_4 = 70$, and $z_5 = 85$.

(b) Repeat Part a using the second response Y_2.

(c) Perform a multivariate multiple regression analysis using both responses Y_1 and Y_2.
  (i) Suggest and fit appropriate linear regression models.
  (ii) Analyze the residuals.
  (iii) Construct a 95% prediction ellipse for both Total TCAD and Amount of amitriptyline for $z_1 = 1$, $z_2 = 1200$, $z_3 = 140$, $z_4 = 70$, and $z_5 = 85$. Compare this ellipse with the prediction intervals in Parts a and b. Comment.

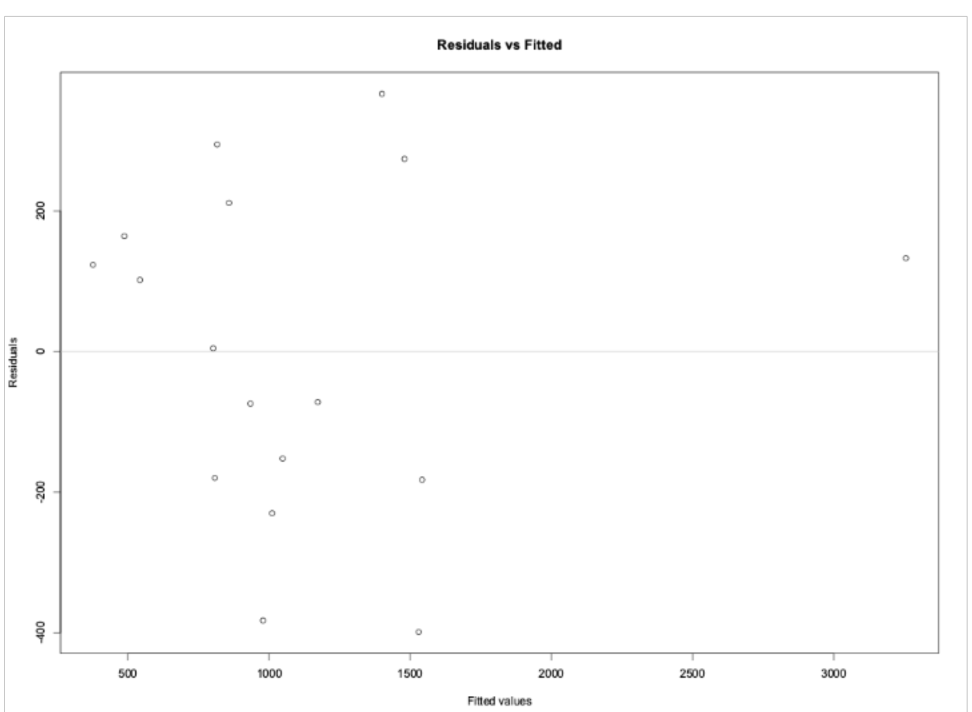
(a) i) I run ANOVA model selection and found no significant difference between the full model ($Z_1 - Z_5$) and the simpler models ($Z_1 - Z_2$); ($Z_1 - Z_3$). Therefore, I will accept the simpler model here.

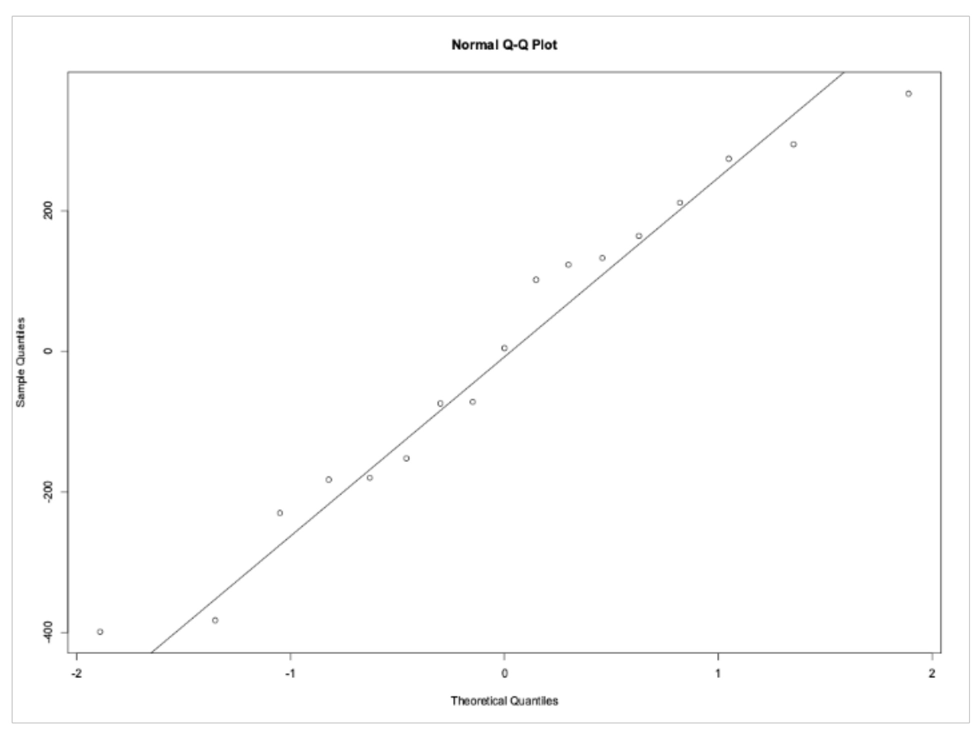
$\hat{y}_1 = 56.72 + 507.07 z_1 + 0.329 z_2$ //

```
Call:
lm(formula = y1 ~ z1 + z2, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-756.05 -190.68  -59.83  203.32  560.84

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  56.72005  206.70337   0.274   0.7878
z1          507.07308  193.79082   2.617   0.0203 *
z2            0.32896    0.04978   6.609 1.17e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 358.6 on 14 degrees of freedom
Multiple R-squared:  0.7664,    Adjusted R-squared:  0.733
F-statistic: 22.96 on 2 and 14 DF,  p-value: 3.8e-05
```

```
> anova(mod3, mod2, mod1)
Analysis of Variance Table

Model 1: y1 ~ z1 + z2
Model 2: y1 ~ z1 + z2 + z3
Model 3: y1 ~ z1 + z2 + z3 + z4 + z5
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     14 1800356
2     13 1459222  1    341134 4.3131 0.06204 .
3     11  870008  2    589214 3.7249 0.05817 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ii) From the fitted vs residual plot and Q-Q plot, there a little abnormality, but we can say the data is mostly normally distributed. Hence, there is no need to adjust the data itself.





(iii) $z_0^T = [1, 1, 1200]$

By the formula,

$$z_0^T \hat{\beta}_{(i)} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right) \cdot \sqrt{(1 + z_0^T (Z^T Z)^{-1} z_0)\, MSE} \qquad \alpha = 0.05$$

$$\hat{y}_1 = 56.72 + 507.07 z_1 + 0.329 z_2 \quad \text{with } s = 358.$$

The 95 % PI for z_0 = $[154.040, 1763.054]$ //

```
> predicted_y1
       fit      lwr      upr
1 958.5473 154.0402 1763.054
```

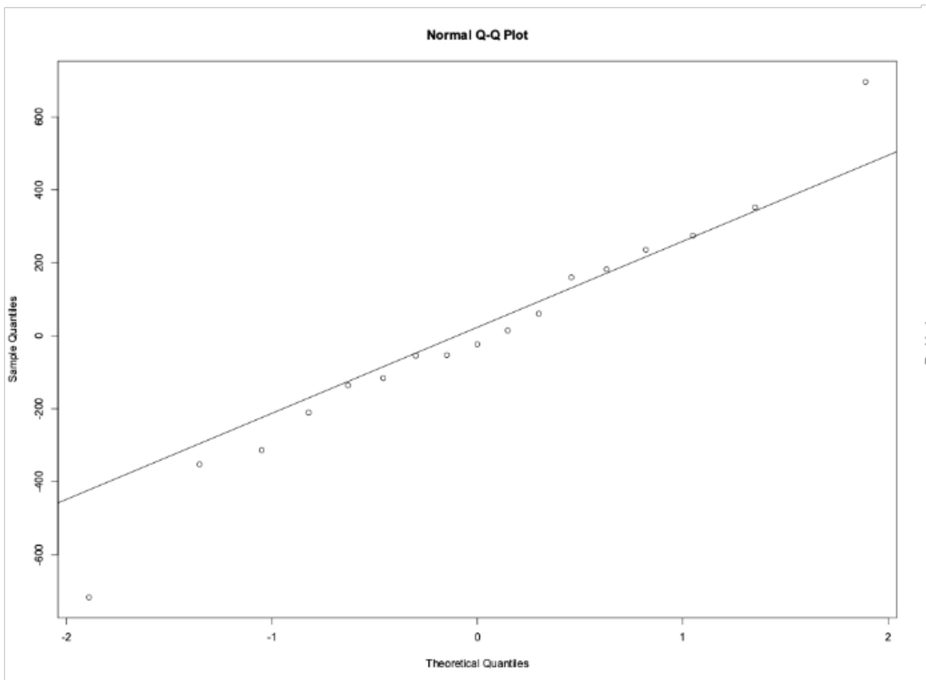
(e) i) From the ANOVA test, I accept the simpler model here as well.

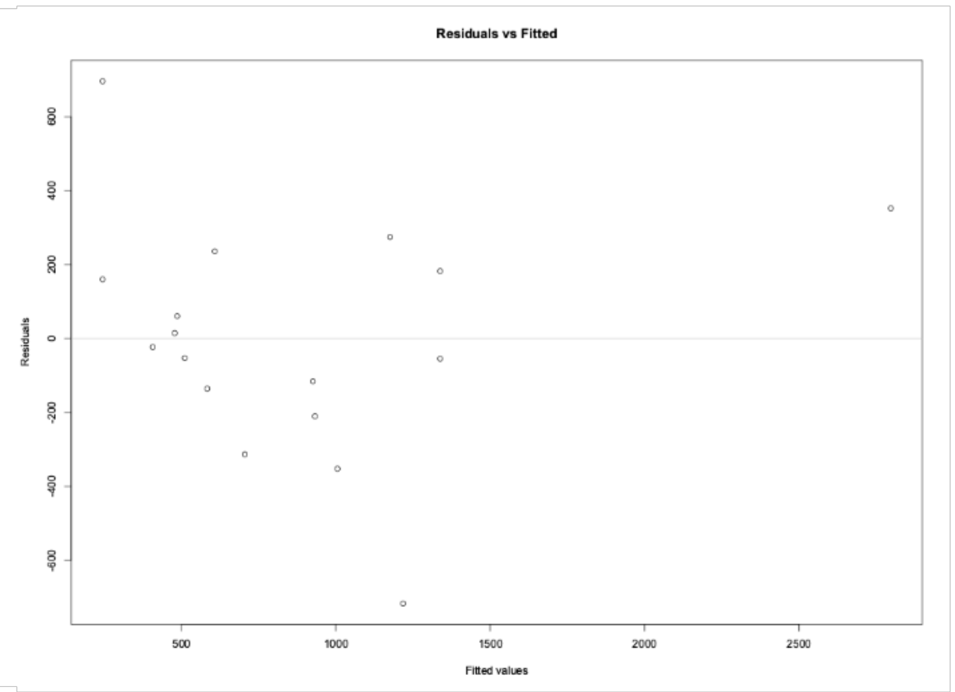
$$\hat{y}_2 = -241.35 + 606.31 z_1 + 0.32 z_2 \quad //$$

```
> anova(mod3, mod2, mod1)
Analysis of Variance Table

Model 1: y2 ~ z1 + z2
Model 2: y2 ~ z1 + z2 + z3
Model 3: y2 ~ z1 + z2 + z3 + z4 + z5
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     14 1620657
2     13 1393645  1    227012 2.6545 0.1315
3     11  940709  2    452936 2.6482 0.1151
```

```
> summary(mod3)

Call:
lm(formula = y2 ~ z1 + z2, data = df)

Residuals:
    Min      1Q  Median      3Q     Max 
-716.80 -135.83  -23.16  182.27  695.97 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -241.34791  196.11640  -1.231  0.23874    
z1           606.30967  183.86521   3.298  0.00529 ** 
z2             0.32425    0.04723   6.866 7.73e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 340.2 on 14 degrees of freedom
Multiple R-squared:  0.787,	Adjusted R-squared:  0.7566 
F-statistic: 25.87 on 2 and 14 DF,  p-value: 1.986e-05
```

ii) From the fitted vs residual plot and Q-Q plot, there a little abnormality, but we can say the data is mostly normally distributed. Hence, there is no need to adjust the data itself.





iii) $z_0^T = [1, 1, 1200]$

By the formula,

$$z_0^T \hat{\beta}_{(i)} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right) \cdot \sqrt{(1 + z_0^T (Z^T Z)^{-1} z_0)\, MSE} \qquad \alpha = 0.05$$

$$\hat{y}_2 = -241.35 + 606.31\, z_1 + 0.32\, z_2 \text{ with } s = 340.2$$

The 95 % PI for z_0 = $[-9.234,\ 1517.37]$ //

```
       fit       lwr      upr
1 754.0677 -9.234071 1517.369
```

(c) i) To run multivariate multiple regression test, I use MANOVA with Phillai's trace. From R-output, I drop Z3 - Z5 since they are insignificant.

```
> Anova(mod_manova, test.statistic="Pillai", type="II")

Type II MANOVA Tests: Pillai test statistic
   Df test stat approx F num Df den Df   Pr(>F)
z1  1   0.65521   9.5015      2     10 0.004873 **
z2  1   0.69097  11.1795      2     10 0.002819 **
z3  1   0.34649   2.6509      2     10 0.119200
z4  1   0.32381   2.3944      2     10 0.141361
z5  1   0.29184   2.0606      2     10 0.178092
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

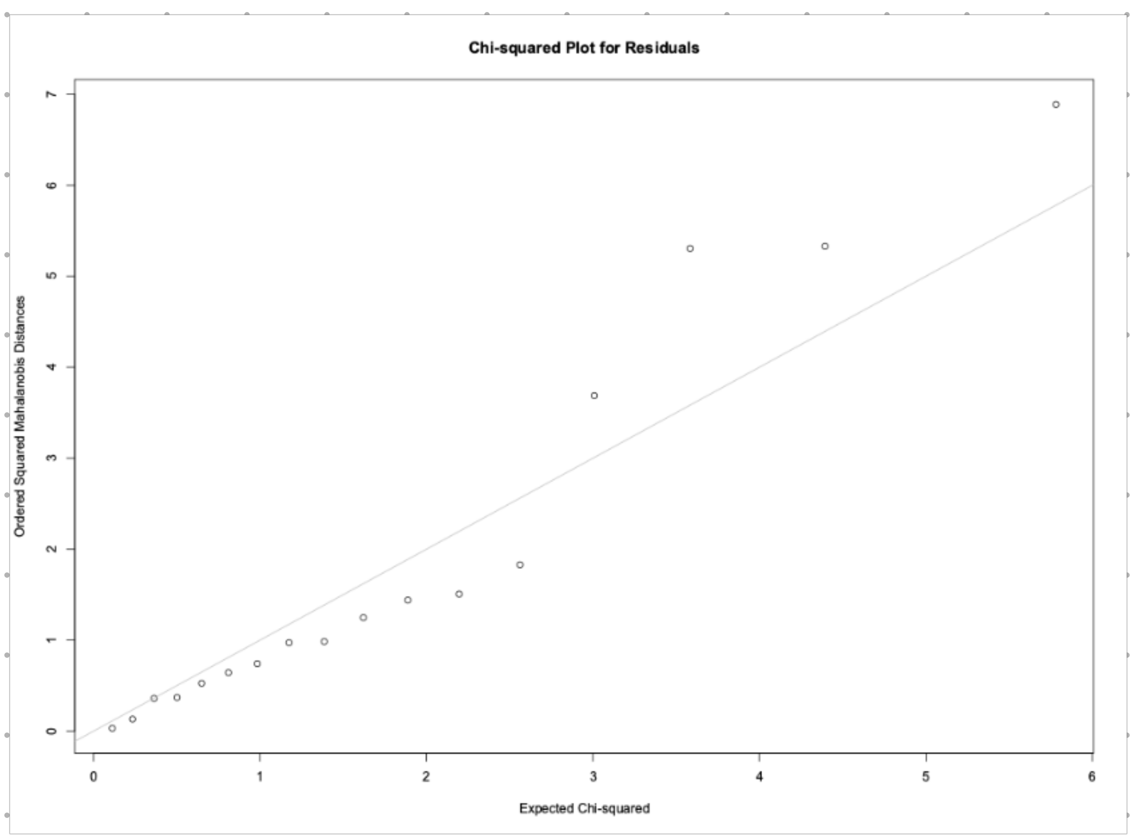
```
> Anova(mod_manova2, test.statistic = "Pillai", type = "II")

Type II MANOVA Tests: Pillai test statistic
   Df test stat approx F num Df den Df    Pr(>F)
z1  1   0.45366   5.3974      2     13   0.01966 *
z2  1   0.77420  22.2866      2     13 6.298e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

∴ The appropriate model here is

$$\hat{y}_1 = 56.72 + 507.07 z_1 + 0.329 z_2$$

$$\hat{y}_2 = -241.35 + 606.31 z_1 + 0.32 z_2$$

ii) By using chi-squared plot for residuals, we see some trend among data. Hence it possibly violates the normality, and we want to transform the data a little.



iii) $z_0^T = [1, 1, 1200]$

By the formula

$$z_0 \hat{\beta}_{(i)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right) F_{m,n-r-m}(\alpha)} \sqrt{\left(1 + z_0^T (Z^T Z)^{-1} z_0\right)\left(\frac{n}{n-r-1} \hat{\sigma}_{ii}\right)} \quad i = 1, 2$$

$\alpha = 0.05$, $m = 2$, $n-r-m = 17-2-2 = 13$

From the ellipsoide a P.I, we can say that the ellipsoide is wider than the Prediction intervals in both part (a) and (b)