Regression modeling 2018 solution

1(a) (i)

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu, \alpha_1, \alpha_2 \end{pmatrix}^T + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

(ii)

$$\begin{pmatrix} \\ \\ \\ \end{pmatrix} = \begin{pmatrix} 1 & x_{11}-\bar{x} & (x_{11}-\bar{x})^2 \\ 1 & & \\ 1 & & \\ 1 & x_{1n_1}-\bar{x} & \\ 1 & & \\ 1 & & \\ 1 & x_{2n_2}-\bar{x} & (x_{2n_2}-\bar{x})^2 \end{pmatrix} \begin{pmatrix} \beta_0, \beta_1, \beta_2 \end{pmatrix}^T + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{2n_2} \end{pmatrix}$$

(iii)

$$\begin{pmatrix} \\ \\ \\ \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 1 & x_{11} \\ 1 & & 1 & \\ 1 & & 1 & \\ 1 & x_{1n_1} & 1 & x_{1n_2} \\ 1 & & 0 & 0 \\ 1 & & 0 & 0 \\ 1 & & 0 & 0 \\ & x_{2n_2} & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha & \beta & \gamma & \delta \end{pmatrix}^T + \underset{\sim}{\varepsilon}$$

1(b)

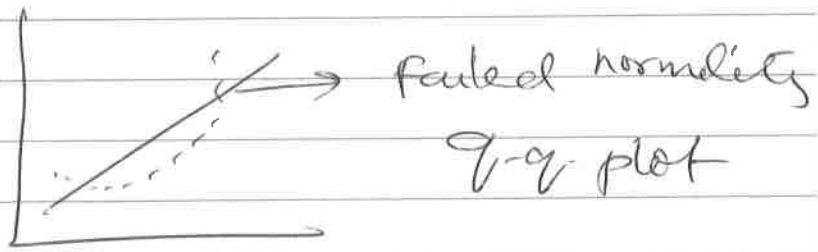
$$\hat{Y} = (X^TX)^{-1}X^T\underline{y}$$

$(X^TX)^{-1}X^T$ is a linear combination.

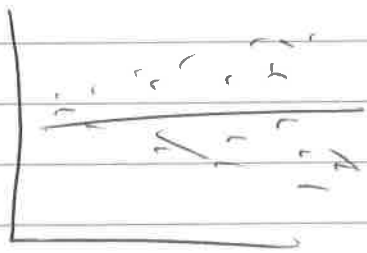
(c) $$\hat{\sigma}^2 = \cancel{(X^TX)}\frac{SSE}{n-p} = \frac{\sum(y_i - \hat{y})^2}{n-p}$$

p is the number of parameters

(d) Q-Q plot for determining normality



Variance → funnel shape



(e) Include non-linear term,

Include other variables

Apply transformations.

2 (a) ~~A~~ Response → BP (blood pressure)
Covariates → Age, Wt, BMI

Start with Full Model

BP = Age + Wt + BMI

~~Take~~ Drop one covariate at a time.
evaluate AIC / C_p / RSS.

Stop when you are not doing better than the larger model at that stage

(b) (i) a = 5492.5
C = 3932.5
b = 131.29

(ii) R^2 = .5828

58% of the variation is explained.

~~iii)~~

(iii) 95% CI of β_1 =

$$(0.689, .977)$$

Calculation needs to be shown.

(iv) 95% PI for husband 160 cm

$$(152.24, 178.25)$$

3. (a) Simplified models

Parallel => $Y_{ij} = \alpha_i + \beta(x_{ij} - \bar{x}_i) + \varepsilon_{ij}$

Single => $Y_{ij} = \alpha + \beta(x_{ij} - \bar{x}_i) + \varepsilon_{ij}$

Two separate => Model A

(b)
$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{2n_1} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & x_{11}-\bar{x}_1 & 0 \\ 1 & & & \\ 1 & 0 & & 0 \\ 1 & 0 & x_{1n_1}-\bar{x}_1 & 0 \\ 0 & 1 & 0 & x_{21}-\bar{x}_2 \\ 0 & 1 & 0 & \vdots \\ 0 & 1 & 0 & x_{2n}-\bar{x}_2 \\ & 1 & & \end{pmatrix} (\alpha_1\ \alpha_2\ \beta_1\ \beta_2) + \underset{\sim}{\varepsilon}.$$

$\hat{\beta} = (X^T X)^{-1} X^T Y.$

Can be simplified.

(c) ii

(d) iv

(e) i

(f) i