EXERCISE 1

$Y_i \sim N(\beta_1 + \beta_2 e^{z_i}, 1)$ indep $i = 1, \dots, 120$

$Y_i \sim N(\beta_1 + \beta_3 e^{z_i^2}, 1)$ indep $i = 121, \dots, 200$

z_i known constants

a) yes: 1. normality, homoscedasticity, independence

2. the model is linear in $\beta_1, \beta_2, \beta_3$

3. the covariates are linearly independent.

b) sample space : $\mathcal{Y} = \mathbb{R}^{200}$

parameter space : $\Theta = \mathbb{R}^3$ (space of $(\beta_1, \beta_2, \beta_3)$)

σ^2 is known

c) the model can be written as

$Y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$ $\quad i = 1, \dots, 200$

where we define

$\varepsilon_i \sim N(0, 1)$ iid

x_{i2} covariate $\quad x_{i2} = \begin{cases} e^{z_i} & \text{for } i = 1, \dots, 120 \\ 0 & \text{otherwise} \end{cases}$

x_{i3} covariate $\quad x_{i3} = \begin{cases} e^{z_i^2} & i = 121, \dots, 200 \\ 0 & \text{otherwise} \end{cases}$

in matrix form we get

$\underline{Y} = [Y_1, \dots, Y_{120}, Y_{121}, \dots, Y_{200}]^T$ vector of random variables (dim: 200×1)

$\underline{Y} \sim N_{200}(X\underline{\beta}, I_{200})$

X $(n \times p) = (200 \times 3)$ matrix of known constants

$$X = [\underline{1} \;\; \underline{x}_2 \;\; \underline{x}_3] = \begin{bmatrix} 1 & e^{z_1} & 0 \\ 1 & e^{z_2} & 0 \\ \vdots & \vdots & \vdots \\ 1 & e^{z_{120}} & 0 \\ 1 & 0 & e^{z_{121}^2} \\ \vdots & \vdots & \vdots \\ 1 & 0 & e^{z_{200}^2} \end{bmatrix}$$

$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$ vector of unknown constants

$\underline{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_{120}, \varepsilon_{121}, \dots, \varepsilon_{200}]^T$ vector of random variables

$\underline{\varepsilon} \sim N_{200}(\underline{0}, I)$ $\quad I$ 200×200 identity matrix

d)

$$X^T X = \begin{bmatrix} \underline{1}^T \\ \underline{x}_2^T \\ \underline{x}_3^T \end{bmatrix} \cdot [\underline{1} \;\; \underline{x}_2 \;\; \underline{x}_3] = \begin{bmatrix} \underline{1}^T\underline{1} & \underline{1}^T\underline{x}_2 & \underline{1}^T\underline{x}_3 \\ \underline{x}_2^T\underline{1} & \underline{x}_2^T\underline{x}_2 & \underline{x}_2^T\underline{x}_3 \\ \underline{x}_3^T\underline{1} & \underline{x}_3^T\underline{x}_2 & \underline{x}_3^T\underline{x}_3 \end{bmatrix}$$

$$= \begin{bmatrix} 200 & \sum_{i=1}^{120} e^{z_i} & \sum_{i=121}^{200} e^{z_i^2} \\ \sum_{i=1}^{120} e^{z_i} & \sum_{i=1}^{120} e^{2z_i} & 0 \\ \sum_{i=121}^{200} e^{z_i^2} & 0 & \sum_{i=121}^{200} e^{2z_i^2} \end{bmatrix}$$

$$\underset{3\times 200 \;\; 200\times 1}{X^T\underline{y}} = \begin{bmatrix} \underline{1}^T \\ \underline{x}_2^T \\ \underline{x}_3^T \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_{200} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{200} y_i \\ \sum_{i=1}^{120} e^{z_i} y_i \\ \sum_{i=121}^{200} e^{z_i^2} y_i \end{bmatrix}$$

the MLE $\hat{\underline{\beta}}$ is found as $\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{y}$

e) the distribution of $\hat{\underline{\beta}}(\underline{Y})$ is $\hat{\underline{\beta}}(\underline{Y}) \sim N_3(\underline{\beta}, \underbrace{(X^TX)^{-1}}_{(X^TX)^{-1}\sigma^2})$ $\quad \sigma^2 = 1$

f) $\underline{e} = \underline{y} - X\hat{\underline{\beta}}$

$\sum_{i=1}^{200} e_i = 0$ yes: the model includes the intercept

$\sum_{i=1}^{200} e_i z_i = 0$ no: $[z_1, \dots, z_{200}]^T$ does not belong to $C(X)$ (column space of X)

$\sum_{i=1}^{200} e_i z_i^2 = 0$ no

$\sum_{i=1}^{200} e_i e^{z_i} = 0$ no

$\sum_{i=1}^{200} e_i e^{z_i^2} = 0$ no

$\sum_{i=1}^{120} e_i e^{z_i} = 0$ yes