Assignment3 — Qintao Ying

1. (a) $\sigma_i^2 = \sigma^2 X_i \quad (X_i > 0)$, $\varepsilon \sim MVN(0, \sigma^2 \Sigma_1)$, $\Sigma = \begin{pmatrix} x_1 & & \\ & x_2 & \\ & & \ddots \; x_n \end{pmatrix}$

$\Sigma^{-\frac{1}{2}} \in \mathbb{R}^{n \times n}$, $\Sigma^{-\frac{1}{2}} Y = \Sigma^{-\frac{1}{2}} X\beta + \Sigma^{-\frac{1}{2}} \varepsilon$, set $\varepsilon_{new} = \Sigma^{-\frac{1}{2}} \varepsilon$.

$\varepsilon_{new} \sim N(0, \sigma^2 I_n)$, $\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$

$E(\tilde{\beta}) = (X^T \cdot \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \cdot (X\beta + \varepsilon) = \beta$

$Var(\tilde{\beta}) = (X^T \cdot \Sigma^{-1} X)^{-1} \cdot X^T \Sigma^{-1} \cdot \Sigma_1 \cdot \Sigma^{-1} \cdot X \cdot (X^T \cdot \Sigma^{-1} X)^{-1} \cdot \sigma^2$

$= (X^T \cdot \Sigma^{-1} \cdot X)^{-1} \sigma^2$

$\tilde{\beta} \sim N(\beta, (X^T \Sigma^{-1} X)^{-1} \sigma^2)$

(b) $Q(\beta) = (Y - X \cdot \beta)^T (Y - X\beta) = Y^T - Y^T X\beta - \beta^T X^T \cdot Y - \beta^T X^T X \beta$

$\frac{\partial Q(\beta)}{\partial \beta} = 0 \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$

(c) $E(\hat{\beta}) = (X^T X)^{-1} \cdot X^T (X\beta + \varepsilon) = \beta$

$Var(\hat{\beta}) = (X^T X)^{-1} \cdot X^T \cdot Var(\varepsilon) \cdot X \cdot (X^T X)^{-1}$, $Var(\varepsilon) = \sigma^2 \Sigma$

where $\Sigma_1 = \begin{pmatrix} x_1 & & 0 \\ & x_2 & \\ 0 & \ddots & x_n \end{pmatrix}$, $Var(\hat{\beta}) = (X^T X)^{-1} X^T \cdot \sigma^2 \cdot \Sigma_1 \cdot X (X^T X)^{-1}$

$\hat{\beta} \sim N(\beta, (X^T X)^{-1} X^T \Sigma \cdot X (X^T X)^{-1} \sigma^2)$

$\hat{e}_i = Y_i - \hat{Y}_i = Y_i - X_i \cdot \hat{\beta}$

(4) Here $\hat{\beta} = (X^T X)^{-1} X^T Y$, $\tilde{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$

denote $A = (X^T X)^{-1} X^T$, $B = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}$

$\hat{\beta} - \tilde{\beta} = (A - B) Y$, let $C = A - B \Rightarrow \hat{\beta} = \tilde{\beta} + CY$

$Var(\hat{\beta}) = Var(\tilde{\beta} + CY) = Var(\tilde{\beta}) + Var(CY) + 2\,Cov(\tilde{\beta}, Y) \cdot C^T$

Here $CX = \left((X^T X^{-1}) X^T - (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}\right) X = I_n - I_n = 0$

$Cov(\tilde{\beta}, Y) = (X^T \Sigma^{-1} X)^{-1} X^T \cdot \Sigma^{-1} \cdot \sigma^2 \cdot \Sigma = (X^T \cdot \Sigma^{-1} \cdot X)^{-1} X^T \cdot \sigma^2$

$Cov(\tilde{\beta}, Y) \cdot C^T = 0$ (because $C \cdot X = 0$)

$\Rightarrow Var(\hat{\beta}) = Var(\tilde{\beta}) + Var(CY) \geq Var(\tilde{\beta})$ ($Var(CY)$ is positive semi-definite matrix)

Here we can conclude that $Var(\hat{\beta}_1) \geq Var(\tilde{\beta}_1)$

2.(b) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ for $i = 1, 2, \cdots, n$ $n = 15$, $\beta = (\beta_0, \beta_1)^T$

Y_i is the ith response of concentration, X_i is the ith observation over time, $Var(\varepsilon_i) = \sigma^2$, $Cov(\varepsilon_i, \varepsilon_j) = 0$ $(i \neq j)$ $\varepsilon_i \sim N(0, \sigma^2)$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad Y = X\beta + \varepsilon$$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = 0.2249605$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{pmatrix} 2.573 \\ -0.3240 \end{pmatrix}$$

$$sd(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} = 0.0433$$

$$sd(\hat{\beta}_0) = \sqrt{Var(\hat{\beta}_0)} = \hat{\sigma} \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} = 0.2487$$

$$\hat{R}^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 0.8116$$

(c) Now we want to know whether time will affect concentration.

which is equal to testing:

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

Here $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2}} \sim t_{n-2}$, then $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2}} = -7.483$

p-value $= 4.61 \times 10^{-6} < 0.05$, hence we have reason to believe that β_1 is not equal to 0, H_0 is rejected. Therefore, time will affect concentration.

(f) $Z_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, for $i=1,2,\dots,n$, $n=15$, $\beta = \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$

Z_i is the ith response of concentration after log-transformation, others the same.

$Z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$. $Z = X\beta + \varepsilon$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X^T X)^{-1} X^T Z = \begin{pmatrix} 1.50792 \\ -0.44993 \end{pmatrix} \quad \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Z_i - \hat{Z}_i)^2 = 0.013225$$

$$sd(\hat{\beta}_1) = \hat{\sigma} \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 0.01049, \quad sd(\hat{\beta}_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = 0.06028$$

$$\hat{R}^2 = 1 - \frac{\sum_{i=1}^{n} (Z_i - \hat{Z}_i)^2}{\sum_{i=1}^{n} (Z_i - \bar{Z})^2} = 0.993$$

3. Here we have a constrained least square estimate.

Now, $\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \cdots (\mathrm{I})$. $\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$ $(i=1,2,3)$ and $\beta_3 = \beta_1 + \beta_2$

$\Rightarrow \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ 0 \end{bmatrix} \cdots (\mathrm{II})$, here we let β^* be the estimator of (I), $\hat{\beta}$ be the estimator of (II).

take $R = [1, 1, -1]$, $\beta^* = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

We have $\hat{\beta} = \beta^* - (I_3^T \cdot I_3)^{-1} \cdot R^T \cdot \{R (I_3^T \cdot I_3)^{-1} \cdot R^T\}^{-1} \cdot R\beta^*$

Hence $\hat{\beta} = \beta^* - \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \cdot \frac{1}{3} \cdot [1, 1, -1]\, \beta^*$

$$= \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \right) \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \frac{2}{3}b_1 - \frac{1}{3}b_2 + \frac{1}{3}b_3 \\ -\frac{1}{3}b_1 + \frac{2}{3}b_2 + \frac{1}{3}b_3 \\ \frac{1}{3}b_1 + \frac{1}{3}b_2 + \frac{2}{3}b_3 \end{bmatrix}$$

$$Var(\beta^*) = \sigma^2 [I_3 - R^T \cdot \frac{1}{3} \cdot R] = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \cdot \sigma^2$$

$E(\hat{\beta}_1) = \frac{2}{3}b_1 - \frac{1}{3}b_2 + \frac{1}{3}b_3$, $E(\hat{\beta}_2) = -\frac{1}{3}b_1 + \frac{2}{3}b_2 + \frac{1}{3}b_3$

$E(\hat{\beta}_3) = \frac{1}{3}b_1 + \frac{1}{3}b_2 + \frac{2}{3}b_3$, $sd(\hat{\beta}_1) = sd(\hat{\beta}_2) = sd(\hat{\beta}_3) = \sigma \cdot \sqrt{\frac{2}{3}} = \frac{\sqrt{6}\,\sigma}{3}$

4. (a) $Y = X\beta + \varepsilon$, now $\varepsilon \sim N(0, \sigma^2 I_n)$

Here $\hat{\beta} = (X^T X)^{-1} \cdot X^T Y$. Then $\sum_{j=0}^{p-1} c_j \hat{\beta}_j = c^T \cdot \hat{\beta}$

$$E(c^T \hat{\beta}) = c^T \cdot (X^T \cdot X^{-1}) \cdot X^T \cdot (X\beta + \varepsilon) = c^T \cdot \beta$$

$$Var(c^T \hat{\beta}) = c^T \cdot (X^T X)^{-1} \cdot c \cdot \sigma^2$$

$$\therefore c^T \hat{\beta} \sim N(c^T \cdot \beta, \ c^T (X^T X)^{-1} c \cdot \sigma^2)$$

(b) Now we have hypothesis:

$$H_0: c^T \cdot \beta = h \quad vs \quad H_1: c^T \cdot \beta = h$$

$$\hat{\sigma}^2 = \frac{Y^T (I - X(X^T X)^{-1} X^T) Y}{n-p}$$

Then here our statistic is $T = \dfrac{c^T \beta - h}{\hat{\sigma} \cdot \sqrt{c^T (X^T X)^{-1} c}} \sim t_{n-p}$

And the rejection region is $\{ |T| > t_{n-p, \frac{\alpha}{2}} \}$

(c) Here we have $\hat{Y}_{n+1} = z \cdot \hat{\beta}$, $Y_{n+1} = z\beta + \varepsilon_{n+1}$

$$E(Y_{n+1} - \hat{Y}_{n+1}) = z \cdot \beta - z\beta = 0$$

$$Var(Y_{n+1} - \hat{Y}_{n+1}) = \sigma^2 \cdot (1 + z^T (X^T X)^{-1} z)$$

$$Y_{n+1} - \hat{Y}_{n+1} \sim N(0, \ \sigma \sqrt{1 + z^T (X^T X)^{-1} z})$$

(d) $MSE(\hat{Y}_{n+1}) = E[(\hat{Y}_{n+1} - Y_{n+1})^2] = Var(\hat{Y}_{n+1}) + (E(\hat{Y}_{n+1}) - Y_{n+1})^2$

$$= \sigma^2 z^T (X^T X)^{-1} z + \varepsilon_{n+1}^2 = \sigma^2 (1 + z^T (X^T X)^{-1} z) > \sigma^2, \quad QED.$$

(e) here we have $\dfrac{\hat{Y}_{n+1} - Y_{n+1}}{\hat{\sigma} \cdot \sqrt{1 + z^T (X^T X)^{-1} z}} \sim t_{n-p}$, $\hat{\sigma} = \frac{1}{n-p} (Y^T \cdot (I_n - X (X^T X)^{-1} X^T) Y)$

Now the $1-\alpha$ confidence interval is

$$\left[\hat{Y}_{n+1} - t_{n-p, \frac{\alpha}{2}} \cdot \hat{\sigma} \cdot \sqrt{1 + z^T (X^T X)^{-1} z}, \ \hat{Y}_{n+1} + t_{n-p, \frac{\alpha}{2}} \cdot \hat{\sigma} \cdot \sqrt{1 + z^T (X^T X)^{-1} z} \right] \quad (\hat{Y} = z \cdot \hat{\beta})$$

$$\Rightarrow \left[z (X^T X)^{-1} X^T Y, - t_{n-p, \frac{\alpha}{2}} \cdot \hat{\sigma} \cdot \sqrt{1 + z^T (X^T X)^{-1} z}, \ z (X^T X)^{-1} X^T Y + t_{n-p, \frac{\alpha}{2}} \cdot \hat{\sigma} \cdot \sqrt{1 + z^T (X^T X)^{-1} z} \right]$$

(f), Now, denote $a = \begin{pmatrix} a_0 \\ \vdots \\ a_{p-1} \end{pmatrix}$ as the least-squares projection of x_p onto the subspace of $\mathbb{R}^n$ spanned by X. Here:

$\hat{a} = (X^T X)^{-1} X^T \cdot x_p$, $\quad r_p = x_p - \hat{x}_p = x_p - X (X^T X)^{-1} X^T x_p$

(g) set $X' = [X, x_p]$, $\gamma = \begin{bmatrix} \gamma_0 \\ \vdots \\ \gamma_p \end{bmatrix}$, then $\gamma = (X'^T X')^{-1} X'^T Y$

$$\gamma = \left(\begin{bmatrix} X^T \\ x_p^T \end{bmatrix} [X \ \ x_p] \right)^{-1} \cdot \begin{bmatrix} X^T Y \\ x_p^T Y \end{bmatrix}$$

$$= \begin{pmatrix} (X^T X)^{-1} + (X^T X)^{-1} \cdot X^T x_p (x_p^T x_p - x_p^T X (X^T X)^{-1} X^T x_p)^{-1} x_p^T X \cdot (X^T X)^{-1} & -(X^T X)^{-1} X^T x_p (x_p^T x_p - x_p^T X (X^T X)^{-1} X^T x_p)^{-1} \\ -(x_p^T x_p - x_p^T X (X^T X)^{-1} X^T x_p)^{-1} x_p^T \cdot X (X^T X)^{-1} & (x_p^T x_p - x_p^T X (X^T X)^{-1} X^T x_p)^{-1} \end{pmatrix} \begin{pmatrix} X^T Y \\ x_p^T Y \end{pmatrix}$$

$$= \begin{pmatrix} \hat{\beta} - \dfrac{\hat{a} x_p^T N_X Y}{x_p^T N_X x_p} \\[2ex] \dfrac{x_p^T N_X Y}{x_p^T N_X x_p} \end{pmatrix}$$

Here $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_{p-1} \end{pmatrix}$, $\hat{a} = \begin{pmatrix} \hat{a}_0 \\ \vdots \\ \hat{a}_{p-1} \end{pmatrix}$

$$N_X = I - X(X^T X)^{-1} X^T$$

And $N_X x_p = r_p$, $x_p^T N_X x_p = r_p^T \cdot r_p$

Hence $\begin{pmatrix} \hat{\gamma}_0 \\ \vdots \\ \hat{\gamma}_p \end{pmatrix} = \begin{pmatrix} \hat{\beta} - \dfrac{\hat{a} r_p^T \cdot Y}{r_p^T r_p} \\[2ex] \dfrac{r_p^T Y}{r_p^T r_p} \end{pmatrix}$