Q2

a. We have

(i)

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\therefore n\bar{x} = \sum_{i=1}^{n} x_i$$

$$\therefore X^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix}$$

$$X^TX = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$= \begin{bmatrix} 1\times1 + 1\times1 + \dots + 1\times1 & x_1 + x_2 + \dots + x_n \\ x_1 + x_2 + \dots + x_n & x_1^2 + x_2^2 + \dots + x_n^2 \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

and $n\bar{x} = \sum_{i=1}^{n} x_i$

$$\therefore X^TX = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

(ii) We have:

$$\bullet\ \det(X^TX) = \begin{vmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{vmatrix}$$

$$= n\sum_{i=1}^{n} x_i^2 - (n\bar{x})(n\bar{x})$$

$$= n\sum_{i=1}^{n} x_i^2 - n^2\bar{x}^2$$

$$= n\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

and $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$

$$\therefore \det(X^TX) = nS_{xx}$$

ⓑ. ~~Bas~~ According to Lemma 6, if X ($n \times 2$) is a matrix with linearly independent columns then the symmetric, 2×2 matrix X^TX is invertible.

. ~~We need~~ We have:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$ ~~with~~

and for X^TX ~~is~~ to be invertible, $\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ and $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ are linearly independent

. Hence, we can not express $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ in form of $\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$

. Therefore, there has to be at least one of x_i's (for $i = 1, 2, \dots n$) that is different from the ~~remaining~~ remaining.

ⓒ. We have

$$.X^TX = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

$$.X^Ty = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} y_1 + y_2 + \dots + y_n \\ x_1y_1 + x_2y_2 + \dots + x_ny_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_iy_i \end{bmatrix}$$ and

and $n\bar{y} = \sum_{i=1}^{n} y_i$

$$\therefore X^Ty = \begin{bmatrix} n\bar{y} \\ \sum_{i=1}^{n} x_iy_i \end{bmatrix}$$

X^TX is invertible

$$\therefore (X^TX)^{-1} = \frac{1}{n\sum_{i=1}^{n} x_i^2 - n^2\bar{x}^2} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$= \frac{1}{n\,S_{xx}} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

(from ⓐ (ii))

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\therefore n\bar{y} = \sum_{i=1}^{n} y_i$$

$$\cdot\ (X^T X)^{-1} X^T y = \frac{1}{n S_{xy}} \begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} \begin{bmatrix} n\bar{y} \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

$$= \frac{1}{n S_{xx}} \begin{bmatrix} n\bar{y}\sum_{i=1}^{n} x_i^2 - n\bar{x}\sum_{i=1}^{n} x_i y_i \\ -n^2\bar{x}\bar{y} + n\sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

$$= \frac{1}{S_{xx}} \begin{bmatrix} \bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \end{bmatrix}$$

and $S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$

$$\therefore (X^T X)^{-1} X^T y = \frac{1}{S_{xx}} \begin{bmatrix} \bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i \\ S_{xy} \end{bmatrix} \qquad (1)$$

$$\cdot\ \bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i = \bar{y}\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\bar{y} + n\bar{x}^2\bar{y} - \bar{x}\sum_{i=1}^{n} x_i y_i$$

$$= \bar{y}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) - \bar{x}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)$$

and $S_{xx} = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$; $S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$

$$\therefore \bar{y}\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i y_i = \bar{y} S_{xx} - \bar{x} S_{xy} \qquad (2)$$

. From (1) and (2).

$$\therefore (X^T X)^{-1} X^T y = \frac{1}{S_{xx}} \begin{bmatrix} \bar{y} S_{xx} - \bar{x} S_{xy} \\ S_{xy} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x} \\ \frac{S_{xy}}{S_{xx}} \end{bmatrix}$$

Q1

a. We have:

$. H = \sqrt{n-3}\,(z - \operatorname{arctanh}(p))$

$$. E[H] = E\left[\sqrt{n-3}\,(z - \operatorname{arctanh}(p))\right]$$

$$= \sqrt{n-3}\, E[z - \operatorname{arctanh}(p)]$$

$$= \sqrt{n-3}\left(E[z] - E[\operatorname{arctanh}(p)]\right)$$

and $z \sim N\left(\operatorname{arctanh}(p), \frac{1}{n-3}\right)$

$$\therefore E[H] = \sqrt{n-3}\,(\operatorname{arctanh}(p) - \operatorname{arctanh}(p))$$

$$= 0$$

$$. \operatorname{Var}(H) = \operatorname{Var}\left(\sqrt{n-3}\,(z - \operatorname{arctanh}(p))\right)$$

$$= (n-3)\operatorname{Var}(z - \operatorname{arctanh}(p))$$

$$= (n-3)\left(\operatorname{Var}(z) + \operatorname{Var}(\operatorname{arctanh}(p))\right)$$

$$= (n-3)\left(\frac{1}{n-3} + 0\right)$$

$$= \frac{n-3}{n-3} = 1.$$

$\therefore H \sim N(0,1)$

. Since ~~[illegible]~~ $H \sim N(0,1)$, which is a known distribution that does not depend on p, and $H = \sqrt{n-3}\,(z - \operatorname{arctanh}(p))$ ~~[illegible]~~

$\therefore$ H is a pivotal quantity for p

b. We have:

The symmetric $(1-\alpha)100\%$ confidence interval for p:

$$(1-\alpha)100\% = P(L \leq p \leq U)$$

$$= P(\operatorname{arctanh}(L) \leq \operatorname{arctanh}(p) \leq \operatorname{arctanh}(U))$$

$$= P(z - \operatorname{arctanh}(L) \geq z - \operatorname{arctanh}(p) \geq z - \operatorname{arctanh}(U))$$

$$= P(\sqrt{n-3}(z - \operatorname{arctanh}(L)) \geq \sqrt{n-3}(z - \operatorname{arctanh}(p)) \geq \sqrt{n-3}(z - \operatorname{arctanh}(U)))$$

$$= P(\sqrt{n-3}(z - \operatorname{arctanh}(L)) \geq H \geq \sqrt{n-3}(z - \operatorname{arctanh}(U)))$$

where $H \sim N(0,1)$

• We want a symmetric confidence interval:

$P(p \leq L) = \frac{\alpha}{2}$ and $P(p \geq U) = \frac{\alpha}{2}$

• And we know:

$P(H \geq Z_{\alpha/2}) = \frac{\alpha}{2}$ and $P(H \leq -Z_{\alpha/2}) = \frac{\alpha}{2}$ where $H \sim N(0,1)$

• This implies:

• $Z_{\alpha/2} = \sqrt{n-3}\,(z - \operatorname{arctanh}(L))$

$\therefore \frac{1}{\sqrt{n-3}} Z_{\alpha/2} = z - \operatorname{arctanh}(L)$

$\therefore \operatorname{arctanh}(L) = z - \frac{Z_{\alpha/2}}{\sqrt{n-3}}$

$\therefore L = \tanh\left(z - \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right)$

• $-Z_{\alpha/2} = \sqrt{n-3}\,(z - \operatorname{arctanh}(U))$

$\therefore \frac{-1}{\sqrt{n-3}} Z_{\alpha/2} = z - \operatorname{arctanh}(U)$

$\therefore \operatorname{arctanh}(U) = z + \frac{Z_{\alpha/2}}{\sqrt{n-3}}$

$\therefore U = \tanh\left(z + \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right)$

$\therefore$ The symmetric $(1-\alpha)100\%$ confidence ~~interval~~ interval for p:

$$\left(\tanh\left(z - \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right),\ \tanh\left(z + \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right)\right)$$

ⓒ. ~~We need to calc~~ Based on the given dataset, we can calculate 95% confidence interval for the true correlation p between X and Y (using R code)

~~data~~

data ← read.csv("bivariate_normal_data.csv")

s_x = sd(data\$X) ≈ 0.994

s_y = sd(data\$Y) ≈ 1.044

s_{xy} = cov(data\$X, data\$Y) ≈ 0.522

$\therefore r = \frac{s_{xy}}{s_x s_y} = \frac{0.522}{(0.994)(1.044)} \approx 0.503$

$\therefore z = \operatorname{arctanh}(r) = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right) = \frac{1}{2}\log\left(\frac{1+0.503}{1-0.503}\right) \approx 0.553$

• Since it is a 95% confidence interval, the confidence level is 5%

$\therefore \alpha = 0.05$

$Z_{\alpha/2} = Z_{0.05/2} = Z_{0.025}$ = qnorm(0.025, lower.tail = FALSE) ≈ 1.960

and $n = 100$

. Therefore, the 95% confidence interval for p:

$$\left(\tanh\left(z - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right), \tanh\left(z + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right)\right)$$

$$= \left(\tanh\left(0.553 - \frac{1.960}{\sqrt{100-3}}\right), \tanh\left(0.553 + \frac{1.960}{\sqrt{100-3}}\right)\right)$$

$$\simeq (0.340 \, , \, 0.636)$$

ⓓ. We have:

$H_0: p = \frac{1}{2}$

and $\alpha = 0.05$

. Since $p = \frac{1}{2}$ is in the 95% confidence interval for p $(0.340, 0.636)$, there is insufficient evidence to reject the null hypothesis at the 5% level

. ~~In context, this says it is unreasonable to assume~~

. In context, we do not have sufficient evidence to say that the true correlation p between X and Y is not equal to $\frac{1}{2}$