

Multiple Choice Questions. 每題 5 分，共 100 分。

1. (5 分) Let  $\{(e_i, X'_i)\}$  be a sequence of independent and  $N(0, I_{k+1})$ -distributed random vectors for some  $k := \dim(X_i) > 1$ , and  $X_{ij}$  be the  $j$ th element of  $X_i$  for  $j = 1, \dots, k$ . Assume that  $Y_i$  is a random variable defined in the following way:

$$Y_i = \sum_{j=1}^k X_{ij} + \left( \sum_{j=1}^k X_{ij}^2 \right)^2 + e_i.$$

Let  $g(X_i)$  be an arbitrary transformation of  $X_i$  such that  $\mathbb{E}[(Y_i - g(X_i))^2]$  is defined, and  $g^*(X_i)$  be the optimal choice of  $g(X_i)$  which minimizes  $\mathbb{E}[(Y_i - g(X_i))^2]$ . Which of the following is right when  $k = 10$ ?

- A.  $\mathbb{E}[g^*(X_i)] = 60$
- B.  $\mathbb{E}[g^*(X_i)] = 80$
- C.  $\mathbb{E}[g^*(X_i)] = 160$
- D.  $\mathbb{E}[g^*(X_i)] = 180$
- E. None of the above choices (a)-(d).

**解答：**最小化  $\mathbb{E}[(Y_i - g(X_i))^2]$  的最優解是**條件期望**：

$$g^*(X_i) = \mathbb{E}[Y_i \mid X_i]$$

由於  $e_i$  與  $X_i$  獨立且  $\mathbb{E}[e_i] = 0$ ：

$$g^*(X_i) = \mathbb{E} \left[ \sum_{j=1}^k X_{ij} + \left( \sum_{j=1}^k X_{ij}^2 \right)^2 + e_i \mid X_i \right] = \sum_{j=1}^k X_{ij} + \left( \sum_{j=1}^k X_{ij}^2 \right)^2$$

計算  $\mathbb{E}[g^*(X_i)]$ ：

$$\mathbb{E}[g^*(X_i)] = \mathbb{E} \left[ \sum_{j=1}^k X_{ij} \right] + \mathbb{E} \left[ \left( \sum_{j=1}^k X_{ij}^2 \right)^2 \right]$$

第一項： $\mathbb{E} \left[ \sum_{j=1}^k X_{ij} \right] = \sum_{j=1}^k \mathbb{E}[X_{ij}] = 0$  (因為  $X_{ij} \sim N(0, 1)$ )

第二項：令  $W = \sum_{j=1}^k X_{ij}^2 \sim \chi^2(k)$ ，則

$$\mathbb{E}[W^2] = \text{Var}(W) + (\mathbb{E}[W])^2 = 2k + k^2$$

(因為  $\chi^2(k)$  的期望值為  $k$ ，變異數為  $2k$ )

當  $k = 10$ ：

$$\mathbb{E}[g^*(X_i)] = 0 + (2 \times 10 + 10^2) = 20 + 100 = 120$$

120 不在選項 (a)-(d) 中。

**答案：**(e)

2. (5 分) Let  $\{(X_{1i}, X_{2i})\}_{i=1}^n$  be a sequence of IID random vectors with the distribution:

$$\begin{bmatrix} X_{1i} \\ X_{2i} \end{bmatrix} \sim N \left( \begin{bmatrix} 99 \\ 89 \end{bmatrix}, \begin{bmatrix} 2 & 0.5 \\ 0.5 & 4 \end{bmatrix} \right).$$

Denote  $\bar{X}_1 = n^{-1} \sum_{i=1}^n X_{1i}$  and  $\bar{X}_2 = n^{-1} \sum_{i=1}^n X_{2i}$ . Which of the following is right when  $n = 100$ ?

- A.  $\text{Var}[\bar{X}_1 + \bar{X}_2] = 0.160$
- B.  $\text{Var}[\bar{X}_1 + \bar{X}_2] = 0.155$
- C.  $\text{Var}[\bar{X}_1 + \bar{X}_2] = 0.150$
- D.  $\text{Var}[\bar{X}_1 + \bar{X}_2] = 0.070$
- E. None of the above choices (a)-(d).

**解答：**對於 IID 樣本：

$$\text{Var}(\bar{X}_1) = \frac{\text{Var}(X_{1i})}{n} = \frac{2}{100} = 0.02$$

$$\text{Var}(\bar{X}_2) = \frac{\text{Var}(X_{2i})}{n} = \frac{4}{100} = 0.04$$

$$\text{Cov}(\bar{X}_1, \bar{X}_2) = \frac{\text{Cov}(X_{1i}, X_{2i})}{n} = \frac{0.5}{100} = 0.005$$

因此：

$$\begin{aligned} \text{Var}[\bar{X}_1 + \bar{X}_2] &= \text{Var}(\bar{X}_1) + \text{Var}(\bar{X}_2) + 2\text{Cov}(\bar{X}_1, \bar{X}_2) \\ &= 0.02 + 0.04 + 2(0.005) = 0.02 + 0.04 + 0.01 = 0.07 \end{aligned}$$

**答案：**(d)

3. (5 分) Let  $\{Y_i\}_{i=1}^n$  be a sequence of independent and  $t(4)$ -distributed random variables. Consider a linear regression:

$$Y_i = \alpha + e_i,$$

where  $\alpha$  is a parameter, and  $e_i$  is the error term. Let  $\hat{\alpha}$  be the least squares estimator of  $\alpha$ . Which of the following is right?

- A.  $\mathbb{E}[\hat{\alpha}] = 0$  and  $\text{Var}[\hat{\alpha}] = \frac{1}{4n}$
- B.  $\mathbb{E}[\hat{\alpha}] = 0$  and  $\text{Var}[\hat{\alpha}] = \frac{1}{n}$
- C.  $\mathbb{E}[\hat{\alpha}] = 0$  and  $\text{Var}[\hat{\alpha}] = \frac{1}{2n}$
- D.  $\mathbb{E}[\hat{\alpha}] = 0$  and  $\text{Var}[\hat{\alpha}] = \frac{2}{n}$
- E. None of the above choices (a)-(d).

**解答：**在模型  $Y_i = \alpha + e_i$  中，最小二乘估計量是樣本平均數：

$$\hat{\alpha} = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

$Y_i \sim t(4)$ ，其性質：

- $\mathbb{E}[Y_i] = 0$  (當自由度  $> 1$ )
- $\text{Var}(Y_i) = \frac{\nu}{\nu-2} = \frac{4}{4-2} = \frac{4}{2} = 2$  (當自由度  $\nu > 2$ )

因此：

$$\begin{aligned}\mathbb{E}[\hat{\alpha}] &= \mathbb{E}[\bar{Y}] = \mathbb{E}[Y_i] = 0 \\ \text{Var}[\hat{\alpha}] &= \text{Var}(\bar{Y}) = \frac{\text{Var}(Y_i)}{n} = \frac{2}{n}\end{aligned}$$

答案：(d)

4. (5 分) Let  $\{X_i\}_{i=1}^n$  be a sequence of IID random variables with the probability density function:

$$f(x) = \exp(-x), \quad \text{for } x \in \mathbb{R}.$$

Denote  $\sigma^2 := \text{Var}[X_i]$ ,  $\bar{X} := n^{-1} \sum_{i=1}^n X_i$  and  $\hat{\sigma}^2 := n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ . Using a suitable large-sample method, we obtain that  $\sqrt{n}(\hat{\sigma} - \sigma)$  has the limiting distribution  $N(0, \kappa)$ , as  $n \rightarrow \infty$ , for some  $\kappa > 0$ . Which of the following is right?

- A.  $\kappa = 1$
- B.  $\kappa = 2$
- C.  $\kappa = 4$
- D.  $\kappa = 8$
- E. None of the above choices (a)-(d).

**解答：**  $f(x) = e^{-x}$  對於  $x \in \mathbb{R}$  不是正確的機率密度函數（應該是  $x > 0$ ）。假設題目意指指數分布  $f(x) = e^{-x}$  for  $x > 0$ 。

指數分布的性質：

- $\mathbb{E}[X] = 1$
- $\text{Var}(X) = \sigma^2 = 1$
- $\mathbb{E}[X^2] = 2, \mathbb{E}[X^3] = 6, \mathbb{E}[X^4] = 24$

根據 Delta 方法，對於  $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ ：

$$\sqrt{n}(\hat{\sigma} - \sigma) \xrightarrow{d} N\left(0, \frac{1}{4\sigma^2} \cdot \text{Var}(\hat{\sigma}^2) \cdot n\right)$$

樣本變異數的漸近變異數：

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$$

其中  $\mu_4 = \mathbb{E}[(X - \mu)^4]$  是第四中央動差。

對於指數分布， $\mu_4 = 9$ （可計算得出），所以

$$\mu_4 - \sigma^4 = 9 - 1 = 8$$

應用 Delta 方法：

$$\kappa = \frac{1}{4\sigma^2} \times 8 = \frac{8}{4 \times 1} = 2$$

答案：(b)

5. (5 分) Assume that  $\{Y_i\}_{i=1}^n$  and  $\{Z_i\}_{i=1}^n$  are two independent sequences of IID random variables with finite fourth moments. Consider the following two regressions:

$$Y_i = \alpha_Y + e_i \quad \text{and} \quad Z_i = \alpha_Z + u_i,$$

where  $\alpha_Y$  and  $\alpha_Z$  are parameters, and

$$\begin{bmatrix} e_i \\ u_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right).$$

Denote  $\bar{Y} := \frac{1}{n} \sum_{i=1}^n Y_i$  and  $\bar{Z} := \frac{1}{n} \sum_{i=1}^n Z_i$ . Let  $\Phi^{-1}(\cdot)$  be the quantile function of  $N(0, 1)$ , and set  $\Phi^{-1}(0.9) = 1.282$ ,  $\Phi^{-1}(0.95) = 1.645$ ,  $\Phi^{-1}(0.975) = 1.96$  and  $\Phi^{-1}(0.99) = 2.326$ . Suppose that we estimate these two regressions using the least squares method. Which one of the following is the 95% confidence interval of  $\alpha_Y + \alpha_Z$  implied by a suitable large-sample method when  $n = 100$ ,  $\bar{X} = 0.4$  and  $\bar{Y} = 0.6$ ?

- A. (0.822, 1.626)
- B. (0.716, 1.426)
- C. (0.608, 1.392)
- D. (0.416, 2.266)
- E. None of the above choices (a)-(d).

**解答：**估計量： $\hat{\alpha}_Y = \bar{Y} = 0.6$ ， $\hat{\alpha}_Z = \bar{Z} = 0.4$ （假設題目中  $\bar{X} = 0.4$  應為  $\bar{Z} = 0.4$ ）

點估計： $\hat{\alpha}_Y + \hat{\alpha}_Z = 0.6 + 0.4 = 1.0$

變異數計算：

$$\text{Var}(\bar{Y}) = \frac{\text{Var}(e_i)}{n} = \frac{2}{100} = 0.02$$

$$\text{Var}(\bar{Z}) = \frac{\text{Var}(u_i)}{n} = \frac{1}{100} = 0.01$$

$$\text{Cov}(\bar{Y}, \bar{Z}) = \frac{\text{Cov}(e_i, u_i)}{n} = \frac{0.5}{100} = 0.005$$

$$\text{Var}(\bar{Y} + \bar{Z}) = 0.02 + 0.01 + 2(0.005) = 0.04$$

$$\text{SE}(\bar{Y} + \bar{Z}) = \sqrt{0.04} = 0.2$$

95% 信賴區間：

$$1.0 \pm 1.96 \times 0.2 = 1.0 \pm 0.392 = (0.608, 1.392)$$

**答案：**(c)

6. (5 分) Let  $\{X_i\}_{i=1}^n$  be a sequence of independent and  $\chi^2(1)$ -distributed random variables. Denote  $\bar{X} := n^{-1} \sum_{i=1}^n X_i$ . By Chebyshev's inequality, we have the result:

$$P(\bar{X} \leq 2) \geq \beta,$$

for some  $\beta \in (0, 1)$ . Which of the following is right when  $n = 50$ ?

- A.  $\beta = 0.95$
- B.  $\beta = 0.96$

- C.  $\beta = 0.97$   
 D.  $\beta = 0.98$   
 E. None of the above choices (a)-(d).

**解答：**  $X_i \sim \chi^2(1) : \mathbb{E}[X_i] = 1, \text{Var}(X_i) = 2$

對於  $\bar{X} : \mathbb{E}[\bar{X}] = 1, \text{Var}(\bar{X}) = \frac{2}{n} = \frac{2}{50} = 0.04$

Chebyshev 不等式：

$$P(|\bar{X} - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

我們要找  $P(\bar{X} \leq 2)$ 。注意  $\bar{X} \leq 2$  等價於  $\bar{X} - 1 \leq 1$ 。

$$P(\bar{X} > 2) = P(\bar{X} - 1 > 1)$$

設  $\sigma_{\bar{X}} = \sqrt{0.04} = 0.2$ ，則  $1 = 5 \times 0.2 = 5\sigma_{\bar{X}}$

Chebyshev 不等式給出：

$$P(|\bar{X} - 1| \geq 1) \leq \frac{\text{Var}(\bar{X})}{1^2} = 0.04$$

因此：

$$P(\bar{X} \leq 2) \geq P(|\bar{X} - 1| < 1) \geq 1 - 0.04 = 0.96$$

**答案：**(b)

7. (5 分) Let  $\{(e_i, X'_i)\}_{i=1}^n$  be a sequence of IID random vectors with the distribution:

$$\begin{bmatrix} e_i \\ X_i \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} 1 & 0_{1 \times k} \\ 0_{k \times 1} & \Sigma \end{bmatrix} \right),$$

for some  $k := \dim(X_i) > 1$ . Consider the following regression:

$$Y_i = X'_i \gamma + e_i,$$

where  $\gamma$  is a vector of regression coefficients, and  $e_i$  is an error term. Let  $R^2$  be the “uncentered  $R^{2n}$ ” of this regression, based on the least squares method, and  $R_*^2$  be the probability limit of  $R^2$  as  $n \rightarrow \infty$ . Which of the following is right?

- A.  $R_*^2 = \frac{\gamma' \gamma}{1 + 2\gamma' \gamma}$   
 B.  $R_*^2 = \frac{\gamma' \Sigma \gamma}{1 + \gamma' \Sigma \gamma}$   
 C.  $R_*^2 = \frac{\gamma' (\mu \mu' + \Sigma) \gamma}{1 + \gamma' \Sigma \gamma}$   
 D.  $R_*^2 = \frac{\gamma' \Sigma \gamma}{1 + \gamma' (\mu \mu' + \Sigma) \gamma}$   
 E. None of the above choices (a)-(d).

**解答：** Uncentered  $R^2$  的定義：

$$R^2 = \frac{\sum \hat{Y}_i^2}{\sum Y_i^2} = 1 - \frac{\sum \hat{e}_i^2}{\sum Y_i^2}$$

在大樣本下， $R_*^2 = \frac{\mathbb{E}[\hat{Y}^2]}{\mathbb{E}[Y^2]}$

$Y_i = X_i'\gamma + e_i$ ，所以：

$$\mathbb{E}[Y_i^2] = \mathbb{E}[(X_i'\gamma)^2] + \mathbb{E}[e_i^2] + 2\mathbb{E}[X_i'\gamma \cdot e_i]$$

由於  $e_i$  與  $X_i$  獨立且  $\mathbb{E}[e_i] = 0$ ：

$$\mathbb{E}[Y_i^2] = \mathbb{E}[(X_i'\gamma)^2] + 1 = \gamma'\mathbb{E}[X_i X_i']\gamma + 1$$

$$\mathbb{E}[X_i X_i'] = \text{Cov}(X_i) + \mathbb{E}[X_i]\mathbb{E}[X_i]' = \Sigma + \mu\mu'$$

$$\text{所以 } \mathbb{E}[Y_i^2] = \gamma'(\Sigma + \mu\mu')\gamma + 1$$

預測值  $\hat{Y}_i = X_i'\hat{\gamma}$ ，在大樣本下  $\hat{\gamma} \rightarrow \gamma$ ，所以

$$\mathbb{E}[\hat{Y}_i^2] \rightarrow \gamma'\mathbb{E}[X_i X_i']\gamma = \gamma'(\Sigma + \mu\mu')\gamma$$

因此：

$$R_*^2 = \frac{\gamma'(\mu\mu' + \Sigma)\gamma}{1 + \gamma'(\mu\mu' + \Sigma)\gamma}$$

這不完全符合任何選項。檢視選項 (c)，分母應為  $1 + \gamma'(\mu\mu' + \Sigma)\gamma$ 。

答案：(e)

8. (5 分) Let  $\{(Y_i, Z_i)\}$  be a sequence of IID random vector, in which  $Y_i$  is a Bernoulli random variable, and  $Z_i$  is a random vector with the distribution  $U(\alpha, \beta)$  for some  $0 < \alpha < \beta < 1$ . Consider the following regression:

$$Y_i = \gamma Z_i + e_i,$$

for some  $0 < \gamma < \frac{1}{\beta}$ , and  $e_i$  is an error term with the property  $\mathbb{E}[e_i|Z_i] = 0$ . Also, let  $S$  be the skewness coefficient of  $Y_i$ . Which of the following is right?

- A.  $S = \frac{1-3\left(\frac{\gamma(\alpha+\beta)}{2}\right)+2\left(\frac{\gamma(\alpha+\beta)}{2}\right)^2}{\left(1-\frac{\gamma(\alpha+\beta)}{2}\right)\left(\frac{\gamma(\alpha+\beta)}{2}\left(1-\frac{\gamma(\alpha+\beta)}{2}\right)\right)^{1/2}}$
- B.  $S = \frac{1-3\left(\frac{\gamma(\alpha+\beta)}{2}\right)+2\left(\frac{\gamma(\alpha+\beta)}{2}\right)^2}{\left(1-\frac{\gamma(\alpha+\beta)}{2}\right)^{1/2}\left(\frac{\gamma(\alpha+\beta)}{2}\left(1-\frac{\gamma(\alpha+\beta)}{2}\right)\right)^{1/2}}$
- C.  $S = \frac{1-3\left(\frac{\gamma(\alpha+\beta)}{2}\right)+2\left(\frac{\gamma(\alpha+\beta)}{2}\right)^2}{\left(1+\frac{\gamma(\alpha+\beta)}{2}\right)\left(\frac{\gamma(\alpha+\beta)}{2}\left(1-\frac{\gamma(\alpha+\beta)}{2}\right)\right)^{1/2}}$
- D.  $S = \frac{1-3\left(\frac{\gamma(\alpha+\beta)}{2}\right)+2\left(\frac{\gamma(\alpha+\beta)}{2}\right)^2}{\left(1+\frac{\gamma(\alpha+\beta)}{2}\right)^{1/2}\left(\frac{\gamma(\alpha+\beta)}{2}\left(1-\frac{\gamma(\alpha+\beta)}{2}\right)\right)^{1/2}}$
- E. None of the above choices (a)-(d).

解答：  $Y_i$  是 Bernoulli 隨機變數， $\mathbb{E}[Y_i|Z_i] = \gamma Z_i$  (由  $\mathbb{E}[e_i|Z_i] = 0$ )

令  $p = \mathbb{E}[Y_i] = \mathbb{E}[\mathbb{E}[Y_i|Z_i]] = \mathbb{E}[\gamma Z_i] = \gamma \cdot \frac{\alpha+\beta}{2}$

設  $p = \frac{\gamma(\alpha+\beta)}{2}$

Bernoulli 分布的偏態係數：

$$S = \frac{\mathbb{E}[(Y - \mathbb{E}[Y])^3]}{(\text{Var}(Y))^{3/2}} = \frac{1 - 2p}{\sqrt{p(1 - p)}}$$

但由於  $Y_i$  的分布受  $Z_i$  影響，需要更仔細計算。對於 Bernoulli( $p$ )：

$$S = \frac{1 - 2p}{\sqrt{p(1-p)}}$$

展開  $1 - 2p = 1 - 2 \cdot \frac{\gamma(\alpha+\beta)}{2} = 1 - \gamma(\alpha + \beta)$

這與選項中的形式不完全吻合。經過仔細比對，選項 (b) 的分子形式  $1 - 3p + 2p^2 = (1-p)(1-2p)$  可化簡。

Bernoulli 的偏態公式為  $\frac{1-2p}{\sqrt{p(1-p)}}$ ，經整理選項 (b) 最接近。

**答案：(b)** (需進一步驗證)

9. (5 分) Let  $\{X_i\}_{i=1}^n$  be a sequence of independent and  $N(0,1)$ -distributed random variables. Denote  $Y_n := \frac{1}{n_1} \sum_{i=1}^{n_1} X_i - \frac{1}{n-n_1} \sum_{i=n_1+1}^n X_i$ , for some  $n_1 \leq n$ . Let  $K_n(t)$  be the cumulant generating function of the statistic, with  $t$  denoting a real number. Which of the following is right?

- A.  $\ln K_n(t) = \ln 4 + 2 \ln t + \ln \left( \frac{n}{2n_1(n-n_1)} \right)$   
 B.  $\ln K_n(t) = 2 \ln t + \ln \left( \frac{n}{n_1(n-n_1)} \right)$   
 C.  $\ln K_n(t) = \ln \frac{1}{2} + 2 \ln t + \ln \left( \frac{n}{n_1(n-n_1)} \right)$   
 D.  $\ln K_n(t) = \ln \frac{1}{4} + \frac{1}{2} \ln t + \ln \left( \frac{n}{2n_1(n-n_1)} \right)$   
 E. None of the above choices (a)-(d).

**解答：**  $Y_n = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i - \frac{1}{n-n_1} \sum_{i=n_1+1}^n X_i = \bar{X}_1 - \bar{X}_2$

其中  $\bar{X}_1 \sim N(0, \frac{1}{n_1})$ ， $\bar{X}_2 \sim N(0, \frac{1}{n-n_1})$ ，且獨立。

$Y_n \sim N\left(0, \frac{1}{n_1} + \frac{1}{n-n_1}\right) = N\left(0, \frac{n}{n_1(n-n_1)}\right)$

常態分布  $N(0, \sigma^2)$  的累積量生成函數：

$$K(t) = \frac{\sigma^2 t^2}{2}$$

所以：

$$K_n(t) = \frac{1}{2} \cdot \frac{n}{n_1(n-n_1)} \cdot t^2$$

$$\ln K_n(t) = \ln \frac{1}{2} + \ln \frac{n}{n_1(n-n_1)} + 2 \ln t$$

這與選項 (c) 相符。

**答案：(c)**

10. (5 分) Let  $\{X_i\}_{i=1}^n$  be a sequence of independent and  $N(0,1)$ -distributed random variables. Consider the following variable:

$$Y_i := \begin{cases} X_1, & i = 1, \\ X_i + \beta X_{i-1}, & i > 1, \end{cases}$$

for some  $\beta \in (0, 1)$ , and denote the statistic  $W_n := \frac{1}{n} \sum_{i=1}^n Y_i$ . Which of the following is right?

- A.  $\text{Var}[W_n] = \frac{1}{n^2}(1 + (1 + \beta)(n - 1))^2$
- B.  $\text{Var}[W_n] = \frac{1}{n^2}(1 + (1 + \beta)^2(n - 1))$
- C.  $\text{Var}[W_n] = \frac{1}{n^2}(1 + (1 + \beta)^2(n - 1) + 2\beta(n - 1)^2)$
- D.  $\text{Var}[W_n] = \frac{1}{n^2}(1 + (1 + \beta)(n - 1) + (1 + \beta)^2(n - 1)^2)$
- E. None of the above choices (a)-(d).

**解答：**  $Y_1 = X_1$  ,  $Y_i = X_i + \beta X_{i-1}$  for  $i > 1$

$$W_n = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} (X_1 + \sum_{i=2}^n (X_i + \beta X_{i-1}))$$

整理：

$$nW_n = X_1 + \sum_{i=2}^n X_i + \beta \sum_{i=2}^n X_{i-1} = \sum_{i=1}^n X_i + \beta \sum_{i=1}^{n-1} X_i = (1 + \beta) \sum_{i=1}^{n-1} X_i + X_n$$

令  $S = nW_n$ ：

$$\text{Var}(S) = (1 + \beta)^2(n - 1) + 1 + 2(1 + \beta)\text{Cov}\left(\sum_{i=1}^{n-1} X_i, X_n\right)$$

由於  $X_i$  獨立， $\text{Cov} = 0$ ：

$$\text{Var}(S) = (1 + \beta)^2(n - 1) + 1$$

$$\text{Var}(W_n) = \frac{1}{n^2} \text{Var}(S) = \frac{1}{n^2}(1 + (1 + \beta)^2(n - 1))$$

**答案：**(b)



**Questions 11–14:** Piske and Usagi made a big fortune by selling Line stickers. They are planning to invest the money in the stock market. They are now studying the property of a certain stock  $X$ . They have collected data of monthly returns of  $X$  in the past 8 months ( $r_x$ ) and the corresponding market returns ( $r_m$ ) as well as the risk-free rates ( $r_f$ ).

Month <sub><math>t</math></sub>	$r_x$ (%)	$r_m$ (%)	$r_f$ (%)
1	3	5	2
2	5	5	2
3	3	5	2
4	0	-5	2
5	-2	-5	1
6	-6	-5	1
7	10	10	1
8	15	10	1

11. (5 分) Piske is interested in the systematic risk of stock  $X$  and would like to estimate the following regression:

$$r_{xt} - r_{ft} = b_0 + b_1(r_{mt} - r_{ft}) + e_t \quad (1)$$

Assume that  $e_t \sim IID(0, \sigma^2)$  and that  $E[e_t | r_{mt}, r_{ft}] = 0$ . Please help Piske estimate the ordinary least square (OLS) estimates  $\hat{b}_0^{OLS}$  and  $\hat{b}_1^{OLS}$ .

- A.  $\hat{b}_0^{OLS} = 0.011$ ,  $\hat{b}_1^{OLS} = 0.927$   
 B.  $\hat{b}_0^{OLS} = 0.025$ ,  $\hat{b}_1^{OLS} = 0.927$   
 C.  $\hat{b}_0^{OLS} = 0.011$ ,  $\hat{b}_1^{OLS} = 0.955$   
 D.  $\hat{b}_0^{OLS} = 0.025$ ,  $\hat{b}_1^{OLS} = 0.955$   
 E. None of the above choices (a)-(d).

**解答：**計算超額報酬：

$t$	$y_t = r_{xt} - r_{ft}$	$x_t = r_{mt} - r_{ft}$
1	1	3
2	3	3
3	1	3
4	-2	-7
5	-3	-6
6	-7	-6
7	9	9
8	14	9

$$\bar{y} = \frac{1+3+1-2-3-7+9+14}{8} = \frac{16}{8} = 2$$

$$\bar{x} = \frac{3+3+3-7-6-6+9+9}{8} = \frac{8}{8} = 1$$

$$\sum (x_t - \bar{x})(y_t - \bar{y}) = (2)(-1) + (2)(1) + (2)(-1) + (-8)(-4) + (-7)(-5) + (-7)(-9) + (8)(7) + (8)(12) = -2 + 2 - 2 + 32 + 35 + 63 + 56 + 96 = 280$$

$$\sum (x_t - \bar{x})^2 = 4 + 4 + 4 + 64 + 49 + 49 + 64 + 64 = 302$$

$$\hat{b}_1 = \frac{280}{302} = 0.927$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = 2 - 0.927 \times 1 = 1.073 \text{ (以百分比計)} = 0.01073 \approx 0.011$$

**答案：**(a)

12. (5 分) Usagi learned from his investment class that there should be no intercept in a Capital Asset Pricing Model. Thus, he plans to perform the following regression using the Ordinary Least Square (OLS) method:

$$r_{xt} - r_{ft} = \beta(r_{mt} - r_{ft}) + \epsilon_t \quad (2)$$

Which of the following statement(s) is (are) correct?

- I.  $E[\hat{\beta}^{OLS}] = b_1$ .
- II.  $\sum_t \hat{\epsilon}_t = 0$ .
- III.  $\sum_t (r_{mt} - r_{ft})\hat{\epsilon}_t = 0$ .
- IV. Model (2) should yield a lower centered- $R^2$  than model (1).
- A. I and IV.
- B. II and III.
- C. III and IV.
- D. II and IV.
- E. II, III, and IV.

**解答：**分析各陳述：

**I.**  $E[\hat{\beta}^{OLS}] = b_1$ ：只有當真實模型確實沒有截距項時才成立。如果真實模型有截距， $\hat{\beta}^{OLS}$  會有偏誤。不一定正確。

**II.**  $\sum_t \hat{\epsilon}_t = 0$ ：在無截距模型中，殘差和不一定為零（這是有截距模型的性質）。不正確。

**III.**  $\sum_t (r_{mt} - r_{ft})\hat{\epsilon}_t = 0$ ：這是 OLS 的正規方程式之一，無論有無截距都成立。正確。

**IV.** Model (2) 的 centered- $R^2$  低於 Model (1)：限制模型（無截距）的 centered- $R^2$  通常低於非限制模型。正確。

**答案：**(c) III and IV

13. (5 分) Piske and Usagi are interested in assessing the goodness-of-fit of the Ordinary Least Squares (OLS) estimations for the two models above. Let  $R_1^2$  represent the centered- $R^2$  for the first model, and  $R_2^2$  represent the centered- $R^2$  for the second model. In addition, let  $\widehat{(r_{xt} - r_{ft})_1}$  denote the predicted value of the first model, and  $\widehat{(r_{xt} - r_{ft})_2}$  denote the predicted value of the second model. Lastly, let  $\rho_1$  denote the sample correlation coefficient between  $\widehat{(r_{xt} - r_{ft})_1}$  and  $r_{xt} - r_{ft}$ , and  $\rho_2$  denote the sample correlation coefficient between  $\widehat{(r_{xt} - r_{ft})_2}$  and  $r_{xt} - r_{ft}$ . Which of the following statements is correct?

- A.  $\rho_1^2 = \rho_2^2 = R_1^2$
- B.  $\rho_1^2 = \rho_2^2 = R_2^2$
- C.  $\rho_1^2 = R_1^2 = R_2^2$
- D.  $\rho_2^2 = R_1^2 = R_2^2$
- E. None of the above choices (a)-(d).

**解答：**在有截距的 OLS 迴歸中， $\rho^2 = R^2$ （預測值與實際值的相關係數平方等於  $R^2$ ）。

對於 Model (1)（有截距）： $\rho_1^2 = R_1^2$

對於 Model (2)（無截距）： $\rho_2^2$  不一定等於  $R_2^2$ （centered- $R^2$ ）

但由於兩模型使用相同的自變數，且 Model (2) 是 Model (1) 的限制版本， $R_1^2 \geq R_2^2$ 。經分析， $\rho_1^2 = R_1^2$ ，但  $\rho_2^2$  和  $R_2^2$  的關係較複雜。

答案：(e) (需進一步計算驗證)

14. (5 分) Let's come back to model (1). Assume that  $E[e_t|r_{mt}, r_{ft}] = 0$ , but the error term ( $e_t$ ) follows the distribution below:

$$e_t = \rho e_{t-1} + \varepsilon_t, \quad |\rho| < 1$$

$$\varepsilon_t \sim IID(0, \sigma^2)$$

Which of the following statement is correct?

- A.  $\hat{b}_1^{OLS}$  is biased.
- B.  $\text{Var}(e_t) = \frac{\sigma^2}{1-\rho^2}$
- C.  $\text{Cov}(e_t, e_{t-1}) = \text{Cov}(e_t, e_{t-2}) = 0$
- D.  $\hat{b}_0^{OLS}$  and  $\hat{b}_1^{OLS}$  will no longer be BLUE but just BLUE.
- E. None of the above choices (a)-(d).

解答： $e_t$  遵循 AR(1) 過程。

(a)：OLS 估計量在嚴格外生條件下仍是不偏的 ( $E[e_t|r_{mt}, r_{ft}] = 0$ )，即使有自相關。不正確。

(b)：AR(1) 過程的變異數：

$$\text{Var}(e_t) = \text{Var}(\rho e_{t-1} + \varepsilon_t) = \rho^2 \text{Var}(e_t) + \sigma^2$$

$$\text{Var}(e_t)(1 - \rho^2) = \sigma^2 \Rightarrow \text{Var}(e_t) = \frac{\sigma^2}{1 - \rho^2}$$

正確。

(c)： $\text{Cov}(e_t, e_{t-1}) = \rho \text{Var}(e_t) \neq 0$ 。不正確。

(d)：當誤差項有自相關時，OLS 不再是 BLUE (不是最有效的線性不偏估計量)。不正確 (應該說不再是 BLUE)。

答案：(b)

15. (5 分)  $T$  is a random variable with the following probability density function:

$$f(t) = \beta e^{-\beta t}, \text{ for } t > 0$$

where  $\beta > 0$ . The expected value of  $T$  is  $E(T) = \frac{1}{\beta}$ . A sample consisting of  $n$  independent realizations of  $T$  ( $\{t_1, t_2, \dots, t_n\}$ ) was collected. Which of the following statement is correct?

- A.  $\widehat{\beta}_{MLE} = \frac{\sum_{i=1}^n t_i}{n}$ .
- B.  $\widehat{\beta}_{MLE}$  is an unbiased estimator.
- C.  $\widehat{\beta}_{MLE}$  is a consistent estimator.
- D.  $\widehat{\beta}_{MLE}$  is neither unbiased nor consistent.
- E. None of the above choices (a)-(d).

解答：指數分布的 MLE：

概似函數： $L(\beta) = \prod_{i=1}^n \beta e^{-\beta t_i} = \beta^n e^{-\beta \sum t_i}$

對數概似： $\ell(\beta) = n \ln \beta - \beta \sum t_i$

$$\frac{d\ell}{d\beta} = \frac{n}{\beta} - \sum t_i = 0$$

$$\hat{\beta}_{MLE} = \frac{n}{\sum t_i} = \frac{1}{\bar{t}}$$

(a)： $\hat{\beta}_{MLE} = \frac{n}{\sum t_i}$ ，不是  $\frac{\sum t_i}{n}$ 。不正確。

(b)： $E[\hat{\beta}_{MLE}] = E[1/\bar{T}] \neq \beta$  (Jensen 不等式)。不正確。

(c)：由大數法則， $\bar{T} \xrightarrow{p} 1/\beta$ ，所以  $\hat{\beta}_{MLE} = 1/\bar{T} \xrightarrow{p} \beta$ 。正確。

答案：(c)

16. (5 分) Lisa and Gaspard are interested in the NTU Master program in Finance and collect the salary information of their recent alumni. To investigate potential gender discrimination in the finance industry in Taiwan, they group the observations by gender and estimate the statistics of their annual salary (in thousand NTD). Denote the population mean salary of men as  $\mu_{men}$  and women as  $\mu_{women}$ . Assume that the samples of men and women are distributed independently. Please help Lisa and Gaspard to conduct hypotheses testing using the sample information below.

	sample average salary ( $\bar{Y}$ )	sample standard deviation ( $S_Y$ )	$n$
Men	\$1,080	\$270	30
Women	\$960	\$360	20

Which of the following statement is correct?

- A. To examine potential gender discrimination, we may establish a null hypothesis of  $H_A : \mu_{men} \neq \mu_{women}$ .
- B. The  $t$ -statistic for  $H_0 : \mu_{men} = \mu_{women}$  is 1.271.
- C. The  $t$ -statistic for  $H_0 : \mu_{men} = \$1,000$  is 2.623.
- D. The  $t$ -statistic for  $H_0 : \mu_{women} = \$1,000$  is  $-1.497$ .
- E. We can reject the null hypothesis that there is no gender discrimination at 5% significance level.

解答：(a)： $H_A$  是對立假設，虛無假設應是  $H_0 : \mu_{men} = \mu_{women}$ 。錯誤表述。

(b)：兩獨立樣本  $t$  檢定（假設不等變異數）：

$$\begin{aligned} t &= \frac{\bar{Y}_{men} - \bar{Y}_{women}}{\sqrt{\frac{S_{men}^2}{n_{men}} + \frac{S_{women}^2}{n_{women}}}} = \frac{1080 - 960}{\sqrt{\frac{270^2}{30} + \frac{360^2}{20}}} \\ &= \frac{120}{\sqrt{2430 + 6480}} = \frac{120}{\sqrt{8910}} = \frac{120}{94.4} = 1.271 \end{aligned}$$

正確。

(c)： $t = \frac{1080-1000}{270/\sqrt{30}} = \frac{80}{49.3} = 1.62$ ，不是 2.623。不正確。

(d) :  $t = \frac{960-1000}{360/\sqrt{20}} = \frac{-40}{80.5} = -0.497$  , 不是  $-1.497$  。 **不正確** 。

(e) :  $t = 1.271$  , 雙尾檢定 5% 臨界值約 2.01 ,  $1.271 < 2.01$  , 無法拒絕 。 **不正確** 。

**答案 : (b)**

17. (5 分) There are machines A, B, and C in a factory all producing screws. Of their production, machines A, B, and C produce 1%, 2% and 4% defective screws respectively. Of the total production in the factory, machine A produces 50%, machine B produces 25%, and machine C produces 25%. If one screw is selected at random from the entire screws produced in a day, which of the following is true?
- A. The probability that it is defective is 3%.
  - B. If the selected product is defective, the conditional probability that it was produced by machine A is  $\frac{1}{4}$ .
  - C. If the selected product is defective, the conditional probability that it was produced by machine B is  $\frac{2}{4}$ .
  - D. If the selected product is defective, the conditional probability that it was produced by machine C is  $\frac{3}{4}$ .
  - E. None of the above choices (a)-(d).

**解答 :** 設  $D$  = 瑕疵品事件

**計算**  $P(D)$  :

$$\begin{aligned} P(D) &= P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C) \\ &= 0.01 \times 0.50 + 0.02 \times 0.25 + 0.04 \times 0.25 \\ &= 0.005 + 0.005 + 0.01 = 0.02 = 2\% \end{aligned}$$

(a) :  $P(D) = 2\%$  , 不是  $3\%$  。 **不正確** 。

(b) :  $P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0.005}{0.02} = 0.25 = \frac{1}{4}$  。 **正確** 。

(c) :  $P(B|D) = \frac{P(D|B)P(B)}{P(D)} = \frac{0.005}{0.02} = 0.25 = \frac{1}{4}$  , 不是  $\frac{2}{4}$  。 **不正確** 。

(d) :  $P(C|D) = \frac{P(D|C)P(C)}{P(D)} = \frac{0.01}{0.02} = 0.5 = \frac{2}{4}$  , 不是  $\frac{3}{4}$  。 **不正確** 。

**答案 : (b)**

18. (5 分) Chen, Huang, Lin, and Shen (2022, *MS*) are interested in the causal relationship between credit supply and growth stability of small enterprises in China. They use the data from Alibaba and provide the following figures. The horizontal axis shows the credit scores of the entrepreneurs. The vertical axes are (from left to right) the probability that the entrepreneurs are granted with credit, the sales growth volatility one year *before* the event year, and the sales growth volatility one year *after* the event year. Which of the following statement is incorrect?
- A. There is a sudden jump in credit access when one's credit score surpasses 480.
  - B. There is a sudden drop in the sales growth volatility one year *before* the borrowing when one's credit score surpasses 480.
  - C. There is a sudden drop in the sales growth volatility one year *after* the borrowing when one's credit score surpasses 480.

- D. The findings suggest that the relationship between credit access and future growth stability may be causal.
- E. If there is a discontinuity in the sales growth volatility one year *before* the credit access, we cannot make a causal inference between credit access and future growth stability.

**解答：**這是斷點迴歸設計 (Regression Discontinuity Design, RDD) 的題目。

根據圖形描述：

- 左圖：信用分數超過 480 時，獲得信用的機率有跳躍（斷點）
- 中圖：借款前一年的銷售成長波動性
- 右圖：借款後一年的銷售成長波動性

RDD 的識別假設：在斷點處，除了處理變數外，其他變數應該是連續的。

(a)：信用分數超過 480 時，信用取得有跳躍。觀察左圖可確認。**正確。**

(b)：借款前的波動性在 480 處有突降。如果這是真的，就違反了 RDD 的假設（事前變數應連續）。根據中圖，借款前的波動性在 480 處應該是**連續**的。**不正確的陳述。**

(c)：借款後的波動性在 480 處有突降。這是我們要檢驗的因果效果。**正確。**

(d)：RDD 的發現支持因果關係。**正確。**

(e)：如果事前變數有斷點，就無法做因果推論。**正確。**

題目問哪個陳述是**不正確**的。

**答案：(b)**

19. (5 分) Consider the following model:

$$Y_t = \gamma Y_{t-1} + u_t$$

$$u_t = \phi u_{t-1} + \epsilon_t$$

$$\epsilon_t \sim IID(0, \sigma^2)$$

It is known that  $|\gamma| < 1$  and  $|\phi| < 1$ . We also have a reasonably large  $T$  number of observations. Now, you are planning to run an Ordinary Least Square (OLS) regression of  $\{Y_t\}$  on its lag term  $\{Y_{t-1}\}$ :

$$Y_t = a + bY_{t-1} + e_t$$

Which of the following statements is correct?

- $\hat{b}^{OLS}$  is an unbiased estimator for  $\gamma$ .
- $\hat{b}^{OLS}$  is biased since there is no intercept in the data generating process for  $Y_t$ .
- $\hat{b}^{OLS}$  is biased because  $u_t$  is not stationary.
- $\text{Cov}(Y_{t-1}, u_t) = \frac{\phi\sigma^2}{(1-\gamma\phi)(1-\phi^2)}$
- None of the above choices (a)-(d).

**解答：**真實模型是  $Y_t = \gamma Y_{t-1} + u_t$ ，其中  $u_t$  遵循 AR(1) 過程。

(a)：由於  $u_t$  與  $Y_{t-1}$  相關 ( $u_t$  依賴  $u_{t-1}$ ，而  $Y_{t-1}$  依賴  $u_{t-1}$ )，OLS 估計量  $\hat{b}$  是有偏的。**不正確**。

(b)：偏誤不是因為沒有截距，而是因為  $\text{Cov}(Y_{t-1}, u_t) \neq 0$ 。**不正確**。

(c)： $u_t$  是平穩的 (因為  $|\phi| < 1$ )。**不正確**。

(d)：計算  $\text{Cov}(Y_{t-1}, u_t)$ ： $u_t = \phi u_{t-1} + \epsilon_t$ ，所以  $\text{Cov}(Y_{t-1}, u_t) = \phi \text{Cov}(Y_{t-1}, u_{t-1})$

這需要進一步推導。 $\text{Cov}(Y_{t-1}, u_{t-1})$  可通過狀態空間方法計算。

最終結果為  $\text{Cov}(Y_{t-1}, u_t) = \frac{\phi \sigma^2}{(1-\gamma\phi)(1-\phi^2)}$ 。**正確**。

**答案：**(d)

20. (5 分) Suppose that a data generating process is as the following:

$$Y_i = a + bX_i^* + \epsilon_i$$

However,  $X_i^*$  cannot be directly observed. There are two observable proxies for  $X_i^*$ :

$$X_{i1} = X_i^* + u_{i1}$$

$$X_{i2} = X_i^* + u_{i2}$$

We know that  $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ ,  $X_i^* \sim N(\mu_X, 6)$ ,  $u_{i1} \sim N(0, 1)$ , and  $u_{i2} \sim N(0, 3)$ .  $u_{i1}$  and  $u_{i2}$  are also uncorrelated to  $X_i^*$ ,  $Y_i$ ,  $\epsilon_i$ , and each other. We now construct two additional regressors:

$$X_{i3} = \frac{2}{3}X_{i1} + \frac{1}{3}X_{i2}$$

$$X_{i4} = \frac{1}{5}X_{i1} + \frac{4}{5}X_{i2}$$

For  $n = 1, 2, 3$ , or  $4$ , denote  $\hat{\beta}_n$  as the OLS coefficient when we regress  $Y$  on  $X_{in}$ . For a positive  $b$ , which of the following has the largest value?

- A.  $\frac{1}{2}b$
- B.  $\hat{\beta}_1$
- C.  $\hat{\beta}_2$
- D.  $\hat{\beta}_3$
- E.  $\hat{\beta}_4$ .

**解答：**測量誤差模型中，OLS 估計量的機率極限：

$$\text{plim}(\hat{\beta}_n) = b \times \frac{\text{Var}(X^*)}{\text{Var}(X_{in})} = b \times \frac{6}{6 + \text{Var}(u_{in})}$$

這稱為衰減偏誤 (attenuation bias)。

計算各代理變數的變異數：

- $\text{Var}(X_{i1}) = 6 + 1 = 7$ ， $\text{plim}(\hat{\beta}_1) = b \times \frac{6}{7} = 0.857b$

- $\text{Var}(X_{i2}) = 6 + 3 = 9$  ,  $\text{plim}(\hat{\beta}_2) = b \times \frac{6}{9} = 0.667b$

- $X_{i3} = \frac{2}{3}X_{i1} + \frac{1}{3}X_{i2} = X^* + \frac{2}{3}u_{i1} + \frac{1}{3}u_{i2}$

$$\text{Var}(\text{error of } X_{i3}) = \frac{4}{9}(1) + \frac{1}{9}(3) = \frac{4}{9} + \frac{3}{9} = \frac{7}{9}$$

$$\text{Var}(X_{i3}) = 6 + \frac{7}{9} = \frac{61}{9} \text{ , } \text{plim}(\hat{\beta}_3) = b \times \frac{6}{61/9} = b \times \frac{54}{61} = 0.885b$$

- $X_{i4} = \frac{1}{5}X_{i1} + \frac{4}{5}X_{i2} = X^* + \frac{1}{5}u_{i1} + \frac{4}{5}u_{i2}$

$$\text{Var}(\text{error of } X_{i4}) = \frac{1}{25}(1) + \frac{16}{25}(3) = \frac{1}{25} + \frac{48}{25} = \frac{49}{25}$$

$$\text{Var}(X_{i4}) = 6 + \frac{49}{25} = \frac{199}{25} \text{ , } \text{plim}(\hat{\beta}_4) = b \times \frac{6}{199/25} = b \times \frac{150}{199} = 0.754b$$

比較： $0.5b < 0.667b < 0.754b < 0.857b < 0.885b$

最大值是  $\hat{\beta}_3 \approx 0.885b$ 。

答案：(d)