

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

Questions 1–4 共用題幹：

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 _{<i>t</i>}	r_x (%)	r_m (%)	r_f (%)
1	10	10	2
2	0	10	2
3	10	10	3
4	0	10	3
5	-10	-4	3
6	-20	-4	3
7	0	6	2
8	6	6	2

1. (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. $\hat{b}_0^{OLS} = -0.058$, $\hat{b}_1^{OLS} = 0.944$
 B. $\hat{b}_0^{OLS} = -0.073$, $\hat{b}_1^{OLS} = 0.944$
 C. $\hat{b}_0^{OLS} = -0.058$, $\hat{b}_1^{OLS} = 1.437$
 D. $\hat{b}_0^{OLS} = -0.073$, $\hat{b}_1^{OLS} = 1.437$
 E. $\hat{b}_0^{OLS} = -0.084$, $\hat{b}_1^{OLS} = 1.443$

解答：首先計算超額報酬（以百分比表示）：

t	$y_t = r_{xt} - r_{ft}$	$x_t = r_{mt} - r_{ft}$
1	$10 - 2 = 8$	$10 - 2 = 8$
2	$0 - 2 = -2$	$10 - 2 = 8$
3	$10 - 3 = 7$	$10 - 3 = 7$
4	$0 - 3 = -3$	$10 - 3 = 7$
5	$-10 - 3 = -13$	$-4 - 3 = -7$
6	$-20 - 3 = -23$	$-4 - 3 = -7$
7	$0 - 2 = -2$	$6 - 2 = 4$
8	$6 - 2 = 4$	$6 - 2 = 4$

計算平均值：

$$\bar{y} = \frac{8 - 2 + 7 - 3 - 13 - 23 - 2 + 4}{8} = \frac{-24}{8} = -3$$

$$\bar{x} = \frac{8 + 8 + 7 + 7 - 7 - 7 + 4 + 4}{8} = \frac{24}{8} = 3$$

計算 $\sum(x_t - \bar{x})(y_t - \bar{y})$ 和 $\sum(x_t - \bar{x})^2$ ：

t	$x_t - \bar{x}$	$y_t - \bar{y}$	$(x_t - \bar{x})(y_t - \bar{y})$	$(x_t - \bar{x})^2$
1	5	11	55	25
2	5	1	5	25
3	4	10	40	16
4	4	0	0	16
5	-10	-10	100	100
6	-10	-20	200	100
7	1	1	1	1
8	1	7	7	1
Σ			408	284

$$\hat{b}_1 = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} = \frac{408}{284} = 1.437$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} = -3 - 1.437 \times 3 = -3 - 4.31 = -7.31$$

轉換為小數形式（原數據以百分比表示）： $\hat{b}_0 = -0.0731 \approx -0.073$

答案：(d)

2. (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. $E[\hat{\beta}] \neq E[\hat{b}_1]$.
- II. $\sum_t \hat{\epsilon}_t = 0$.
- III. $\sum_t (r_{mt} - r_{ft}) \hat{\epsilon}_t = 0$.
- IV. Model (2) should yield a higher centered- R^2 than model (1).
 - A. I
 - B. I and II.
 - C. I and III.
 - D. I, II, and III.
 - E. I, II, and IV.

解答：分析各陳述：

I. $E[\hat{\beta}] \neq E[\hat{b}_1]$ ：

- Model (1) 有截距項， \hat{b}_1 在正確設定下是不偏的
- Model (2) 無截距項，如果真實模型有截距， $\hat{\beta}$ 會有偏誤
- 即使真實模型無截距，兩個估計量的期望值形式也不同
- 正確

II. $\sum_t \hat{\epsilon}_t = 0$:

- 在無截距的 OLS 迴歸中，殘差和不一定為零
- 只有有截距的模型才保證殘差和為零
- 不正確

III. $\sum_t (r_{mt} - r_{ft}) \hat{\epsilon}_t = 0$:

- 這是 OLS 的正規方程式（一階條件）： $\sum x_t \hat{\epsilon}_t = 0$
- 無論有無截距項都成立
- 正確

IV. Model (2) 的 centered- R^2 高於 Model (1) :

- 限制模型（無截距）的解釋能力通常低於非限制模型
- Centered- R^2 對無截距模型可能為負
- 不正確

正確的陳述：I 和 III

答案：(c)

3. (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 ρ_1 denote the sample correlation coefficient between $\widehat{(r_{xt} - r_{ft})_1}$ and $r_{xt} - r_{ft}$, and ρ_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?

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

解答：在有截距的簡單線性迴歸中：

$$\rho_1^2 = R_1^2$$

這是因為預測值與實際值的相關係數平方等於決定係數。

對於無截距模型，centered- R^2 的計算方式不同， $\rho_2^2 \neq R_2^2$ （一般而言）。

此外，由於 Model (2) 是 Model (1) 的限制版本， $R_1^2 \geq R_2^2$ （通常嚴格不等）。

因此沒有選項完全正確。

答案：(e)

4. (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. $\text{Var}(e_t) = \frac{\sigma^2}{1-\rho}$
- 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 *BUE* but just *BLUE*.
- E. None of the above choices (a)-(d).

解答： e_t 遵循 AR(1) 過程。

(a) \hat{b}_1^{OLS} 是有偏的：

- 在嚴格外生條件 $E[e_t|r_{mt}, r_{ft}] = 0$ 下，OLS 仍是不偏的
- 自相關影響的是效率，不是不偏性
- 不正確

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

- 正確公式： $\text{Var}(e_t) = \frac{\sigma^2}{1-\rho^2}$ (注意是 ρ^2)
- 不正確

(c) $\text{Cov}(e_t, e_{t-1}) = \text{Cov}(e_t, e_{t-2}) = 0$ ：

- AR(1) 過程有自相關： $\text{Cov}(e_t, e_{t-1}) = \rho \text{Var}(e_t) \neq 0$
- 不正確

(d) 不再是 BUE 但仍是 BLUE：

- BUE (Best Unbiased Estimator) 是所有不偏估計量中最有效的
- BLUE (Best Linear Unbiased Estimator) 是線性不偏估計量中最有效的
- 當誤差有自相關時，OLS 不再是 BLUE (GLS 才是)
- 不正確

所有選項都不正確。

答案：(e)

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

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

where $\lambda > 0$. The expected value of T is $E(T) = \frac{1}{\lambda}$. 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. $\widehat{\lambda}_{MLE} = \frac{\sum_{i=1}^n t_i}{n}$.
- B. $\widehat{\lambda}_{MLE}$ is an unbiased estimator.
- C. $\widehat{\lambda}_{MLE}$ is a consistent estimator.
- D. $\widehat{\lambda}_{MLE}$ is unbiased and consistent.
- E. None of the above choices (a)-(d).

解答：指數分布 $\text{Exp}(\lambda)$ 的 MLE：

概似函數：

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum t_i}$$

對數概似：

$$\ell(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n t_i$$

一階條件：

$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n t_i = 0$$

$$\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^n t_i} = \frac{1}{\bar{t}}$$

檢驗各選項：

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

(b)： $E[\hat{\lambda}_{MLE}] = E[1/\bar{T}]$ 。由 Jensen 不等式， $E[1/\bar{T}] > 1/E[\bar{T}] = \lambda$ ，所以有偏。不正確。

(c)：由大數法則， $\bar{T} \xrightarrow{p} E[T] = 1/\lambda$ ，因此 $\hat{\lambda}_{MLE} = 1/\bar{T} \xrightarrow{p} \lambda$ 。正確（一致估計量）。

(d)：有偏但一致，不是「不偏且一致」。不正確。

答案：(c)

6. (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 μ_{men} and women as μ_{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	\$1,200	\$250	20
Women	\$900	\$300	15

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 2.861.
- C. The t -statistic for $H_0 : \mu_{men} = \$1,000$ is 3.578.
- D. The t -statistic for $H_0 : \mu_{women} = \$1,000$ is -1.789 .
- E. We cannot 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 $H_0 : \mu_{men} = \mu_{women}$ 。表述錯誤。不正確。

(b)：雙樣本 t 檢定 (Welch's t -test)：

$$t = \frac{\bar{Y}_{men} - \bar{Y}_{women}}{\sqrt{\frac{S_{men}^2}{n_{men}} + \frac{S_{women}^2}{n_{women}}}} = \frac{1200 - 900}{\sqrt{\frac{250^2}{20} + \frac{300^2}{15}}} \\ = \frac{300}{\sqrt{\frac{62500}{20} + \frac{90000}{15}}} = \frac{300}{\sqrt{3125 + 6000}} = \frac{300}{\sqrt{9125}} = \frac{300}{95.53} = 3.14$$

不是 2.861。不正確。

(c)：單樣本 t 檢定 (男性)：

$$t = \frac{\bar{Y}_{men} - 1000}{S_{men}/\sqrt{n_{men}}} = \frac{1200 - 1000}{250/\sqrt{20}} = \frac{200}{55.9} = 3.578$$

正確。

(d)：單樣本 t 檢定 (女性)：

$$t = \frac{\bar{Y}_{women} - 1000}{S_{women}/\sqrt{n_{women}}} = \frac{900 - 1000}{300/\sqrt{15}} = \frac{-100}{77.46} = -1.291$$

不是 -1.789 。不正確。

(e)：從 (b) 計算 $t \approx 3.14$ ，在 5% 顯著水準下 (臨界值約 2.03)，可以拒絕虛無假設。不正確。

答案：(c)

7. (5 分) There are machines I, II, and III in a certain factory all producing the same product. Of their production, machines I, II, and III produce 1%, 2% and 3% defective products respectively. Of the total production in the factory, machine I produces 30%, machine II produces 30%, and machine III produces 40%. If one product is selected at random from the total product produced in a day, which of the following is true?

- A. The probability that it is defective is 2%
- B. If the selected product is defective, the conditional probability that it was produced by machine I is $\frac{1}{7}$.
- C. If the selected product is defective, the conditional probability that it was produced by machine II is $\frac{3}{7}$.

- D. If the selected product is defective, the conditional probability that it was produced by machine III is $\frac{5}{7}$.
- E. None of the above choices (a)-(d).

解答：設 D = 瑕疵品事件

計算 $P(D)$ ：

$$\begin{aligned} P(D) &= P(D|I)P(I) + P(D|II)P(II) + P(D|III)P(III) \\ &= 0.01 \times 0.30 + 0.02 \times 0.30 + 0.03 \times 0.40 \\ &= 0.003 + 0.006 + 0.012 = 0.021 = 2.1\% \end{aligned}$$

(a)： $P(D) = 2.1\%$ ，不是 2% 。不正確。

(b)：

$$P(I|D) = \frac{P(D|I)P(I)}{P(D)} = \frac{0.003}{0.021} = \frac{3}{21} = \frac{1}{7}$$

正確。

(c)：

$$P(II|D) = \frac{P(D|II)P(II)}{P(D)} = \frac{0.006}{0.021} = \frac{6}{21} = \frac{2}{7}$$

不是 $\frac{3}{7}$ 。不正確。

(d)：

$$P(III|D) = \frac{P(D|III)P(III)}{P(D)} = \frac{0.012}{0.021} = \frac{12}{21} = \frac{4}{7}$$

不是 $\frac{5}{7}$ 。不正確。

答案：(b)

8. (5 分) Bharadwaj, Løken, and Neilson (2013, *AER*) are interested in the effects of early life health interventions and provide the following figures. It is widely known that the birth weight of an infant is negatively associated with the mortality rate and positively associated with future health condition. The researchers found a natural experiment in both Chile and Norway that the governments require hospitals to provide additional care for infants born less than 1,500 grams. Below are the days stayed in the hospital, mortality rates, and the math performance of the infants when they turn to first to eighth grade across their birth weights. Which of the following statement is **incorrect**?
- Infants born right below 1,500 grams stay longer in the hospital than infants born right above 1,500 grams.
 - The negative slopes of infant mortality rates and birth weights below and above the 1,500-gram birth weight are what the researchers are interested in.
 - Infants born right below 1,500 grams tend to perform better in math than infants born right above 1,500 grams when they grow up.
 - The findings suggest that the relationship between early life health care and future intellectual development may be causal.
 - Except for the discontinuity at exactly 1,500 grams, there seems to be a positive relationship between the birth weights and future math performance.

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：住院天數在 1,500 克處有跳躍（低於 1,500 克的嬰兒住院更久）
- 中圖：死亡率在 1,500 克處有跳躍（低於 1,500 克的嬰兒死亡率較低）
- 右圖：數學表現在 1,500 克處有跳躍（低於 1,500 克的嬰兒表現較好）

題目問哪個陳述是**不正確**的。

- (a)：正確。額外照護導致住院天數增加。
(b)：研究者感興趣的是**斷點處的跳躍**（處理效果），而非斜率。**這個陳述不正確**。
(c)：正確。額外照護導致未來數學表現較好。
(d)：正確。RDD 可以支持因果推論。
(e)：正確。除斷點外，體重與數學表現呈正相關。

答案：(b)

9. (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?

- A. \hat{b}^{OLS} is an unbiased estimator for γ .
B. \hat{b}^{OLS} is biased since there is no intercept in the data generating process for Y_t .
C. \hat{b}^{OLS} is biased because u_t is not stationary.
D. $\text{Cov}(Y_{t-1}, u_t) = \frac{\phi\sigma^2}{(1-\gamma\phi)(1-\phi)}$
E. 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 $u_t = \phi u_{t-1} + \epsilon_t$ ，而 Y_{t-1} 依賴於 u_{t-1}, u_{t-2}, \dots ，所以 $\text{Cov}(Y_{t-1}, u_t) \neq 0$ 。OLS 估計量是有偏的。**不正確**。

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

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

由 $Y_t = \gamma Y_{t-1} + u_t$ ，在穩態下：

$$\text{Cov}(Y_{t-1}, u_{t-1}) = \text{Var}(u_t)/(1 - \gamma^2) \text{ 的某種形式}$$

經過推導（使用 Yule-Walker 方程），正確公式涉及 $(1 - \gamma\phi)(1 - \phi^2)$ ，而非 $(1 - \gamma\phi)(1 - \phi)$ 。 **不正確**。

答案：(e)

10. (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^* :

$$X_{i1} = X_i^* + u_{i1}, \quad X_{i2} = X_i^* + u_{i2}$$

We know that $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, $X_i^* \sim N(\mu_X, 4)$, $u_{i1} \sim N(0, 1)$, and $u_{i2} \sim N(0, 2)$. u_{i1} and u_{i2} are also uncorrelated to X_i^* , Y_i , ϵ_i , and each other. We now construct two additional regressors:

$$X_{i3} = \frac{1}{2}X_{i1} + \frac{1}{2}X_{i2}, \quad X_{i4} = \frac{1}{4}X_{i1} + \frac{3}{4}X_{i2}$$

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

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

解答：測量誤差模型的衰減偏誤：

$$\text{plim}(\hat{\beta}_n) = b \times \frac{\text{Var}(X^*)}{\text{Var}(X_{in})} = b \times \frac{4}{4 + \text{Var}(\text{測量誤差})}$$

計算各代理變數的測量誤差變異數：

- X_{i1} : $\text{Var}(u_{i1}) = 1$, $\text{Var}(X_{i1}) = 4 + 1 = 5$, $\text{plim}(\hat{\beta}_1) = b \times \frac{4}{5} = 0.8b$
- X_{i2} : $\text{Var}(u_{i2}) = 2$, $\text{Var}(X_{i2}) = 4 + 2 = 6$, $\text{plim}(\hat{\beta}_2) = b \times \frac{4}{6} = 0.667b$
- $X_{i3} = \frac{1}{2}X_{i1} + \frac{1}{2}X_{i2} = X^* + \frac{1}{2}u_{i1} + \frac{1}{2}u_{i2}$:

$$\text{Var}\left(\frac{1}{2}u_{i1} + \frac{1}{2}u_{i2}\right) = \frac{1}{4}(1) + \frac{1}{4}(2) = 0.75$$

$$\text{Var}(X_{i3}) = 4 + 0.75 = 4.75 \text{ , } \text{plim}(\hat{\beta}_3) = b \times \frac{4}{4.75} = 0.842b$$

- $X_{i4} = \frac{1}{4}X_{i1} + \frac{3}{4}X_{i2} = X^* + \frac{1}{4}u_{i1} + \frac{3}{4}u_{i2}$:

$$\text{Var}\left(\frac{1}{4}u_{i1} + \frac{3}{4}u_{i2}\right) = \frac{1}{16}(1) + \frac{9}{16}(2) = \frac{19}{16} = 1.1875$$

$$\text{Var}(X_{i4}) = 4 + 1.1875 = 5.1875, \text{plim}(\hat{\beta}_4) = b \times \frac{4}{5.1875} = 0.771b$$

當 $b < 0$ 時，比較各估計量（乘以負數後，絕對值最小的變最大，絕對值最大的變最小）：

衰減因子排序： $0.5 < 0.667 < 0.771 < 0.8 < 0.842$

乘以負的 b ： $0.5b > 0.667b > 0.771b > 0.8b > 0.842b$

所以最小值是 $\hat{\beta}_3 = 0.842b$ （因為 $b < 0$ ，乘以最大因子得最小值）。

答案：(d)

11. (5 分) Let $\{Y_i\}_{i=1}^n$ be a sequence of independent and $N(0, 1)$ -distributed random variables. Denote $\hat{\mu}_2 := n^{-1} \sum_{i=1}^n Y_i^2$. By Chebyshev's inequality, we obtain the result:

$$P(0.5 \leq \hat{\mu}_2 \leq 1.5) \geq \alpha,$$

for some $\alpha \in (0, 1)$. Which of the following is right when $n = 100$?

- A. $\alpha = 0.88$
- B. $\alpha = 0.90$
- C. $\alpha = 0.92$
- D. $\alpha = 0.94$
- E. None of the above choices (a)-(d).

解答： $Y_i \sim N(0, 1)$ ，所以 $Y_i^2 \sim \chi^2(1)$ 。

$\chi^2(1)$ 的性質： $E[Y_i^2] = 1$ ， $\text{Var}(Y_i^2) = 2$

對於 $\hat{\mu}_2 = \frac{1}{n} \sum Y_i^2$ ：

$$E[\hat{\mu}_2] = 1, \quad \text{Var}(\hat{\mu}_2) = \frac{2}{n} = \frac{2}{100} = 0.02$$

Chebyshev 不等式： $P(|X - \mu| \geq k) \leq \frac{\text{Var}(X)}{k^2}$

$$P(0.5 \leq \hat{\mu}_2 \leq 1.5) = P(|\hat{\mu}_2 - 1| \leq 0.5)$$

$$P(|\hat{\mu}_2 - 1| > 0.5) \leq \frac{0.02}{0.5^2} = \frac{0.02}{0.25} = 0.08$$

$$\text{因此：} P(|\hat{\mu}_2 - 1| \leq 0.5) \geq 1 - 0.08 = 0.92$$

答案：(c)

12. (5 分) Let $\{Y_i\}_{i=1}^n$ be a sequence of independent and $\chi^2(1)$ -distributed random variables. Consider a linear regression:

$$Y_i = \beta + e_i,$$

where β is a parameter, and e_i is the error term with $\mathbb{E}[e_i] = 0$. Let $\hat{\beta}$ be the least squares estimator of β . Which of the following is right?

- A. $\mathbb{E}[\hat{\beta}] = 1$ and $\text{Var}[\hat{\beta}] = \frac{2}{n}$
- B. $\mathbb{E}[\hat{\beta}] = \frac{1}{n}$ and $\text{Var}[\hat{\beta}] = \frac{3}{n}$
- C. $\mathbb{E}[\hat{\beta}] = 1$ and $\text{Var}[\hat{\beta}] = \frac{3}{n}$
- D. $\mathbb{E}[\hat{\beta}] = \frac{1}{n}$ and $\text{Var}[\hat{\beta}] = \frac{4}{n}$
- E. None of the above choices (a)-(d).

解答：模型 $Y_i = \beta + e_i$ ，OLS 估計量是樣本平均： $\hat{\beta} = \bar{Y}$

$Y_i \sim \chi^2(1) : E[Y_i] = 1, \text{Var}(Y_i) = 2$

因此：

$$E[\hat{\beta}] = E[\bar{Y}] = E[Y_i] = 1$$

$$\text{Var}(\hat{\beta}) = \text{Var}(\bar{Y}) = \frac{\text{Var}(Y_i)}{n} = \frac{2}{n}$$

答案：(a)

13. (5 分) Let $\{(Y_{1i}, Y_{2i})\}_{i=1}^n$ be a sequence of IID random vectors with the distribution:

$$\begin{bmatrix} Y_{1i} \\ Y_{2i} \end{bmatrix} \sim N \left(\begin{bmatrix} 72 \\ 28 \end{bmatrix}, \begin{bmatrix} 1 & 0.25 \\ 0.25 & 0.5 \end{bmatrix} \right).$$

Denote $\bar{Y}_1 = n^{-1} \sum_{i=1}^n Y_{1i}$ and $\bar{Y}_2 = n^{-1} \sum_{i=1}^n Y_{2i}$. Which of the following is right when $n = 50$?

- A. $\text{Var}[\bar{Y}_1 - \bar{Y}_2] = 0.015$
- B. $\text{Var}[\bar{Y}_1 - \bar{Y}_2] = 0.020$
- C. $\text{Var}[\bar{Y}_1 - \bar{Y}_2] = 0.025$
- D. $\text{Var}[\bar{Y}_1 - \bar{Y}_2] = 0.030$
- E. None of the above choices (a)-(d).

解答：

$$\text{Var}(\bar{Y}_1) = \frac{\text{Var}(Y_{1i})}{n} = \frac{1}{50} = 0.02$$

$$\text{Var}(\bar{Y}_2) = \frac{\text{Var}(Y_{2i})}{n} = \frac{0.5}{50} = 0.01$$

$$\text{Cov}(\bar{Y}_1, \bar{Y}_2) = \frac{\text{Cov}(Y_{1i}, Y_{2i})}{n} = \frac{0.25}{50} = 0.005$$

$$\begin{aligned} \text{Var}[\bar{Y}_1 - \bar{Y}_2] &= \text{Var}(\bar{Y}_1) + \text{Var}(\bar{Y}_2) - 2\text{Cov}(\bar{Y}_1, \bar{Y}_2) \\ &= 0.02 + 0.01 - 2(0.005) = 0.03 - 0.01 = 0.02 \end{aligned}$$

答案：(b)

14. (5 分) Let $\{Y_i\}_{i=1}^n$ be a sequence of independent and $N(0, 1)$ -distributed random variables. Denote $X_n := \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i + \frac{1}{n-n_1} \sum_{i=n_1+1}^n Y_i$, for some $n_1 \leq n$. Let $M_n(t)$ be the moment generating function of X_n with t denoting a real number. Which of the following is right?

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

解答： $X_n = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i + \frac{1}{n-n_1} \sum_{i=n_1+1}^n Y_i = \bar{Y}_1 + \bar{Y}_2$

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

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

常態分布 $N(0, \sigma^2)$ 的動差生成函數： $M(t) = e^{\frac{\sigma^2 t^2}{2}}$

所以：

$$M_n(t) = \exp\left(\frac{1}{2} \cdot \frac{n}{n_1(n-n_1)} \cdot t^2\right)$$

$$\ln M_n(t) = \frac{t^2}{2} \left(\frac{n}{n_1(n-n_1)} \right)$$

答案：(b)

15. (5 分) Let $\{Y_i\}_{i=1}^n$ be a sequence of independent and $N(0, 1)$ -distributed random variables. Consider the following variable:

$$X_i := \begin{cases} Y_1, & i = 1, \\ Y_i - \alpha Y_{i-1}, & i > 1, \end{cases}$$

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

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

解答： $X_1 = Y_1$, $X_i = Y_i - \alpha Y_{i-1}$ for $i > 1$

$$nZ_n = \sum_{i=1}^n X_i = Y_1 + \sum_{i=2}^n (Y_i - \alpha Y_{i-1}) = Y_1 + \sum_{i=2}^n Y_i - \alpha \sum_{i=2}^n Y_{i-1} = \sum_{i=1}^n Y_i - \alpha \sum_{i=1}^{n-1} Y_i$$

$$\text{令 } S = nZ_n = Y_n + (1 - \alpha) \sum_{i=1}^{n-1} Y_i$$

$$\mathbb{E}[S] = 0$$

$$\text{Var}(S) = \mathbb{E}[S^2] = \text{Var}(Y_n) + (1 - \alpha)^2 \text{Var}\left(\sum_{i=1}^{n-1} Y_i\right) + 2(1 - \alpha) \text{Cov}\left(Y_n, \sum_{i=1}^{n-1} Y_i\right)$$

由於 Y_i 獨立：

$$\text{Var}(S) = 1 + (1 - \alpha)^2(n - 1) + 0 = 1 + (1 - \alpha)^2(n - 1)$$

$$\mathbb{E}[Z_n^2] = \frac{\text{Var}(S)}{n^2} = \frac{1}{n^2}(1 + (1 - \alpha)^2(n - 1))$$

答案：(a)

16. (5 分) Let $\{Y_i\}_{i=1}^n$ be a sequence of independent and $U(0, 1)$ -distributed random variables with $\sigma^2 := \text{Var}[Y_i]$, $\bar{Y} := n^{-1} \sum_{i=1}^n Y_i$ and $\hat{\sigma}^2 := n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$. By a suitable large-sample method, it can be shown that $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ has the limiting distribution $N(0, v)$, as $n \rightarrow \infty$, for some $v > 0$. Which of the following is right?

A. $v = \frac{1}{360}$

B. $v = \frac{1}{280}$

C. $v = \frac{1}{240}$

D. $v = \frac{1}{180}$

E. None of the above choices (a)-(d).

解答： $Y_i \sim U(0, 1)$ ：

- $E[Y] = 0.5$
- $\text{Var}(Y) = \sigma^2 = \frac{1}{12}$
- $E[Y^2] = \frac{1}{3}$
- $E[Y^3] = \frac{1}{4}$
- $E[Y^4] = \frac{1}{5}$

第四中央動差： $\mu_4 = E[(Y - 0.5)^4]$

對於 $U(0, 1)$ ，經計算： $\mu_4 = \frac{1}{80}$

樣本變異數的漸近變異數：

$$\begin{aligned} v = \mu_4 - \sigma^4 &= \frac{1}{80} - \left(\frac{1}{12}\right)^2 = \frac{1}{80} - \frac{1}{144} \\ &= \frac{144 - 80}{80 \times 144} = \frac{64}{11520} = \frac{1}{180} \end{aligned}$$

答案：(d)

17. (5 分) Let $\{(e_i, X'_i)\}$ be a sequence of independent and $N(0, I_{k+1})$ -distributed random vectors for some $k := \dim(X_i) > 1$. Assume that Y_i is a random variable generated by the formula:

$$Y_i = X'_i X_i + e_i.$$

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

- A. $\text{Var}[h^*(X_i)] = k$
- B. $\text{Var}[h^*(X_i)] = 3k$
- C. $\text{Var}[h^*(X_i)] = 6k$
- D. $\text{Var}[h^*(X_i)] = 9k$
- E. None of the above choices (a)-(d).

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

$$h^*(X_i) = \mathbb{E}[Y_i|X_i] = \mathbb{E}[X_i'X_i + e_i|X_i] = X_i'X_i$$

(因為 e_i 與 X_i 獨立且 $\mathbb{E}[e_i] = 0$)

$h^*(X_i) = X_i'X_i = \sum_{j=1}^k X_{ij}^2$ ，其中 $X_{ij} \sim N(0, 1)$ 獨立。

所以 $h^*(X_i) \sim \chi^2(k)$ 。

$\chi^2(k)$ 的變異數： $\text{Var}(h^*(X_i)) = 2k$

但 $2k$ 不在選項中。

答案：(e)

18. (5 分) Let $\{(Y_i, X_i)\}$ be a sequence of IID random vector, in which Y_i is a Bernoulli random variable, and X_i is a random variable with the distribution $U(a, b)$ for some $0 < a < b < 1$. Consider the following regression:

$$Y_i = \beta X_i + e_i,$$

for some $0 < \beta < \frac{1}{b}$, and e_i is an error term with the property $\mathbb{E}[e_i|X_i] = 0$. Which of the following is right?

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

解答：由 $\mathbb{E}[e_i|X_i] = 0$ ，我們有 $\mathbb{E}[Y_i|X_i] = \beta X_i$ 。

由於 Y_i 是 Bernoulli 隨機變數， $P(Y_i = 1|X_i) = \beta X_i$ 。

$$\mathbb{E}[Y_i] = \mathbb{E}[\mathbb{E}[Y_i|X_i]] = \mathbb{E}[\beta X_i] = \beta \cdot \frac{a+b}{2}$$

$$\text{令 } p = \beta \cdot \frac{a+b}{2}$$

對於 Bernoulli 隨機變數， $\text{Var}(Y_i) = p(1 - p)$ ：

$$\text{Var}(Y_i) = \frac{\beta(a+b)}{2} \left(1 - \frac{\beta(a+b)}{2}\right)$$

答案：(b)

19. (5 分) Let $\{(e_i, X'_i)\}_{i=1}^n$ be a sequence of independent and $N(0, I_{k+1})$ -distributed random vectors for some $k := \dim(X_i) > 1$. Consider the following regression:

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

where β is a vector of regression coefficients, and e_i is an error term. Let R^2 be the “centered 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. $R_*^2 = \frac{\beta' \beta}{1+2\beta' \beta}$
- B. $R_*^2 = \frac{\beta' \beta}{1+(\beta' \beta)^2}$
- C. $R_*^2 = \frac{2\beta' \beta}{1+2\beta' \beta}$
- D. $R_*^2 = \frac{\beta' \beta}{1+2(\beta' \beta)^2}$
- E. None of the above choices (a)-(d).

解答： $Y_i = X'_i \beta + e_i$ ，其中 $X_i \sim N(0, I_k)$ ， $e_i \sim N(0, 1)$ ，獨立。

$$\text{Var}(Y_i) = \text{Var}(X'_i \beta) + \text{Var}(e_i) = \beta' \text{Var}(X_i) \beta + 1 = \beta' I_k \beta + 1 = \beta' \beta + 1$$

$$\text{Var}(X'_i \beta) = \beta' \beta$$

$$R_*^2 = \frac{\text{Var}(X'_i \beta)}{\text{Var}(Y_i)} = \frac{\beta' \beta}{\beta' \beta + 1}$$

這不完全符合任何選項的形式。

答案：(e)

20. (5 分) Assume that $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$ are two independent sequences of IID random variables with finite third moments. Consider the following two regressions:

$$X_i = \alpha_X + e_i \quad \text{and} \quad Y_i = \alpha_Y + u_i,$$

where α_X and α_Y are parameters, and

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

Denote $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ and $\bar{Y} := \frac{1}{n} \sum_{i=1}^n Y_i$. 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 $\alpha_X - \alpha_Y$ implied by a suitable large-sample method when $n = 100$, $\bar{X} = 0.56$ and $\bar{Y} = 0.53$?

- A. $(-0.166, 0.226)$
- B. $(-0.188, 0.244)$
- C. $(-0.206, 0.262)$
- D. $(-0.264, 0.288)$
- E. None of the above choices (a)-(d).

解答：估計量： $\hat{\alpha}_X = \bar{X} = 0.56$ ， $\hat{\alpha}_Y = \bar{Y} = 0.53$

點估計： $\hat{\alpha}_X - \hat{\alpha}_Y = 0.56 - 0.53 = 0.03$

變異數計算：

$$\text{Var}(\bar{X}) = \frac{\text{Var}(e_i)}{n} = \frac{1}{100} = 0.01$$

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

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

$$\begin{aligned} \text{Var}(\bar{X} - \bar{Y}) &= \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) - 2\text{Cov}(\bar{X}, \bar{Y}) \\ &= 0.01 + 0.01 - 2(0.005) = 0.02 - 0.01 = 0.01 \end{aligned}$$

$$\text{SE}(\bar{X} - \bar{Y}) = \sqrt{0.01} = 0.1$$

95% 信賴區間：

$$0.03 \pm 1.96 \times 0.1 = 0.03 \pm 0.196 = (-0.166, 0.226)$$

答案：(a)