

共 5 頁，20 題選擇題，每題 5 分，總分 100 分。
請用 2B 鉛筆作答於答案卡，並先詳閱答案卡上之「畫記說明」。

1. (5 分) For the following dot plots, which dataset has the largest standard deviation?

- A) 資料點均勻分布於 -100 到 100 之間
- B) 資料點集中於 -100 附近和 100 附近（兩端）
- C) 資料點分成三群： -100 附近、中間、 100 附近
- D) 資料點分成五群，均勻分布於 -100 到 100 之間
- E) 資料點集中於兩端（ -100 和 100 ）

解：答案：E

標準差衡量資料與平均值的離散程度。要使標準差最大，資料點應盡可能遠離平均值。
分析各選項：

- A)：資料均勻分布，標準差中等
- B)：資料集中於兩端，但中間也有一些點，標準差較大
- C)：資料分成三群，中間群會降低標準差
- D)：資料分成五群均勻分布，標準差中等
- E)：資料完全集中於兩個極端（ -100 和 100 ），平均值為 0 ，每個點距離平均值都是 100 ，標準差最大

當資料點都位於極端值（最遠離平均值的位置）時，標準差達到最大。

2. (5 分) There are three factories: A, B, and C, with 100, 1,000, and 10,000 workers, respectively. The mean wages are the same across all three factories. Which of the following statements is true?

- A. The standard deviations of wages in all three factories are equal.
- B. The ranges of wages in all three factories are equal.
- C. The variance of wages in Factory A is the smallest.
- D. The standard deviation of wages in Factory C is the smallest.
- E. None of the above.

解：答案：E (None of the above)

分析各選項：

- A) **錯誤**：標準差取決於資料的分散程度，與樣本大小或工人數量無直接關係。三個工廠的標準差可能不同。
- B) **錯誤**：全距 (Range) = 最大值 - 最小值。三個工廠的工資分布可能不同，全距也可能不同。

- C) **錯誤**：變異數與樣本大小沒有必然的大小關係。工廠 A 的變異數不一定最小。
- D) **錯誤**：工廠 C 有最多工人 (10,000 人)，但這不意味著其標準差最小。標準差取決於資料的分散程度，不是樣本大小。

題目只告訴我們三個工廠的平均工資相同，但沒有提供關於工資分布的任何資訊。因此，我們無法確定哪個工廠的標準差、變異數或全距較大或較小。

注意：如果題目是問「樣本平均數的標準誤」，那麼 $SE = \frac{\sigma}{\sqrt{n}}$ ，工廠 C 的標準誤會最小。但題目問的是「工資的標準差」，這是母體/樣本本身的特性，與樣本大小無關。

3. (5 分) Which of the following statements is correct in a negatively skewed distribution?
- The arithmetic mean is greater than the mode.
 - The arithmetic mean is greater than the median.
 - The difference between the third quartile and the median is equal to the difference between the first quartile and the median.
 - The difference between the third quartile and the median is less than the difference between the first quartile and the median.
 - None of the above.

解：答案：D

負偏（左偏）分配的特性：

- 分配的尾巴向左延伸（有較多極端小值）
- 關係：**平均數 < 中位數 < 眾數**

分析各選項：

- A) **錯誤**：在負偏分配中，平均數 < 眾數（不是大於）
- B) **錯誤**：在負偏分配中，平均數 < 中位數（不是大於）
- C) **錯誤**：這描述的是對稱分配的特性（ $Q_3 - \text{Median} = \text{Median} - Q_1$ ）
- D) **正確**：在負偏分配中，左尾較長，表示：

$$\text{Median} - Q_1 > Q_3 - \text{Median}$$

即 $Q_3 - \text{Median} < \text{Median} - Q_1$ （第三四分位數與中位數的差 < 中位數與第一四分位數的差）

- E) **錯誤**：D 是正確的

記憶方法：負偏 = 左偏 = 尾巴在左邊 = 左邊拉得比較長 = (Q_1 到 Median) 的距離 > (Median 到 Q_3) 的距離

4. (5 分) What does the intercept in a binary linear regression represent?

- The intercept indicates the strength of the relationship between x and y.

- B. The intercept is the expected value of the independent variable when the dependent variable is zero.
- C. The intercept is the expected value of the dependent variable when the independent variable is zero.
- D. The intercept is just a random constant.
- E. None of the above.

解：答案：C

簡單線性迴歸模型： $Y = \beta_0 + \beta_1 X + \epsilon$

其中：

- β_0 = 截距 (intercept)
- β_1 = 斜率 (slope)
- Y = 依變數 (dependent variable)
- X = 自變數 (independent variable)

截距的意義：當 $X = 0$ 時，

$$\mathbb{E}[Y|X = 0] = \beta_0 + \beta_1 \cdot 0 = \beta_0$$

因此，截距 β_0 代表當自變數為零時，依變數的期望值。

分析各選項：

- A) **錯誤：**關係強度由 R^2 或相關係數 r 表示，不是截距
- B) **錯誤：**選項把自變數和依變數搞反了
- C) **正確：**截距是當自變數為零時，依變數的期望值
- D) **錯誤：**截距不是隨機常數，它有明確的統計意義

5. (5 分) In a series of n games, the probability of winning each game is 0.5, and each game is independent. A player wins a prize if at least 60% of the games in the series are won. The possible values for n are 5, 20, and 100. Which value of n should the player choose to maximize the probability of winning a prize?
- A. $n = 5$
- B. $n = 20$
- C. $n = 100$
- D. It does not matter since the probabilities are all the same.
- E. The answer cannot be concluded based on the information provided.

解：答案：A ($n = 5$)

設 $X \sim \text{Binomial}(n, 0.5)$ 為贏得的場數。要贏得獎品，需要 $X \geq 0.6n$ 。

計算各情況的獲獎機率：

$n = 5$ ：需要贏 ≥ 3 場

$$\mathbb{P}(X \geq 3) = \binom{5}{3} \left(\frac{1}{2}\right)^5 + \binom{5}{4} \left(\frac{1}{2}\right)^5 + \binom{5}{5} \left(\frac{1}{2}\right)^5 = \frac{10 + 5 + 1}{32} = \frac{16}{32} = 0.5$$

$n = 20$ ：需要贏 ≥ 12 場

$$\mathbb{P}(X \geq 12) = \sum_{k=12}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20}$$

由對稱性， $\mathbb{P}(X \geq 12) = \mathbb{P}(X \leq 8) < 0.5$

使用常態近似： $X \approx N(10, 5)$ ， $Z = \frac{12-10}{\sqrt{5}} = 0.89$

$\mathbb{P}(X \geq 12) \approx \mathbb{P}(Z > 0.89) \approx 0.187$

$n = 100$ ：需要贏 ≥ 60 場

$$X \approx N(50, 25), \quad Z = \frac{60 - 50}{5} = 2$$

$\mathbb{P}(X \geq 60) \approx \mathbb{P}(Z > 2) \approx 0.0228$

比較： $0.5 > 0.187 > 0.0228$

結論： $n = 5$ 時獲獎機率最高。

直觀解釋：當 $p = 0.5$ 時，要達到超過 50% 的勝率越來越困難。 n 越大，由大數法則，勝率會越接近 50%，越難達到 60%。 n 較小時，變異較大，更容易偏離 50%。

6. (5 分) The dataset consists of four numbers with a mean of 27. If the mean of the smallest three numbers is 20 and the range of the dataset is 36, what is the mean of the largest three numbers?
- 31
 - 32
 - 33
 - 34
 - The answer cannot be concluded based on the information provided.

解：答案：D (34)

設四個數由小到大為 $a \leq b \leq c \leq d$ 。

已知條件：

- 四數平均 = 27 $\Rightarrow a + b + c + d = 108$
- 最小三數平均 = 20 $\Rightarrow a + b + c = 60$
- 全距 = 36 $\Rightarrow d - a = 36$

求解：

從條件 1 和 2：

$$d = 108 - 60 = 48$$

從條件 3：

$$a = d - 36 = 48 - 36 = 12$$

驗證： $a + b + c = 60$ ，所以 $b + c = 60 - 12 = 48$

計算最大三數的平均：

$$\text{最大三數平均} = \frac{b + c + d}{3} = \frac{48 + 48}{3} = \frac{96}{3} = 32$$

等等，讓我重新檢查...

$$b + c + d = (b + c) + d = 48 + 48 = 96$$

$$\frac{b+c+d}{3} = \frac{96}{3} = 32$$

但答案選項中 32 是 B... 讓我再確認一次。

重新計算：

- $a + b + c + d = 108$
- $a + b + c = 60 \Rightarrow d = 48$
- $d - a = 36 \Rightarrow a = 12$
- $b + c = 60 - 12 = 48$
- $b + c + d = 48 + 48 = 96$
- 平均 $= 96/3 = 32$

答案：B (32)

7. (5 分) If the variance of the first i natural numbers is 14, the variance of the first j even natural numbers is 16, and the variance of the first k odd natural numbers is also 16, which of the following statements is correct?

- A. $i = 11$
- B. $j = 7$
- C. $j \neq k$
- D. All of the above
- E. None of the above

解：答案：B ($j = 7$)

公式：前 n 個自然數 $1, 2, \dots, n$ 的變異數為：

$$\text{Var} = \frac{n^2 - 1}{12}$$

驗證選項 A：前 i 個自然數的變異數 $= 14$

$$\frac{i^2 - 1}{12} = 14 \Rightarrow i^2 - 1 = 168 \Rightarrow i^2 = 169 \Rightarrow i = 13$$

計算得 $i = 13$ ，不是 11。所以 A 錯誤。

驗證選項 B：前 j 個偶數 $2, 4, 6, \dots, 2j$ 的變異數 $= 16$

這等於 $2 \times (1, 2, 3, \dots, j)$ ，變異數 $= 4 \times \frac{j^2-1}{12} = \frac{j^2-1}{3}$

$$\frac{j^2-1}{3} = 16 \Rightarrow j^2-1 = 48 \Rightarrow j^2 = 49 \Rightarrow j = 7 \quad \checkmark$$

所以 B 正確。

驗證選項 C：前 k 個奇數 $1, 3, 5, \dots, 2k-1$ 的變異數 $= 16$

前 k 個奇數的變異數也等於 $\frac{k^2-1}{3}$ (可證明)

$$\frac{k^2-1}{3} = 16 \Rightarrow k = 7$$

所以 $j = k = 7$ ，意味著 $j \neq k$ 是錯誤的，C 錯誤。

結論：A 錯、B 對、C 錯。只有 B 正確，故選 B。

8. (5 分) A complex electronic device consists of three components: A, B, and C. The probability of failure for each component in a year is 0.03 for A, 0.05 for B, and 0.1 for C. If any component fails, the entire device will fail. Assuming the components fail independently, what is the closest probability that the device will not fail in a given year?

- A. 0.00015
- B. 0.82
- C. 0.83
- D. 0.95
- E. 0.97

解：答案：C (0.83)

設備不故障 = 三個元件都不故障

各元件不故障的機率：

- $\mathbb{P}(A \text{ 不故障}) = 1 - 0.03 = 0.97$
- $\mathbb{P}(B \text{ 不故障}) = 1 - 0.05 = 0.95$
- $\mathbb{P}(C \text{ 不故障}) = 1 - 0.1 = 0.90$

設備不故障的機率 (獨立事件)：

$$\mathbb{P}(\text{設備不故障}) = 0.97 \times 0.95 \times 0.90 = 0.82935 \approx 0.83$$

驗算： $0.97 \times 0.95 = 0.9215$ ， $0.9215 \times 0.90 = 0.82935$

最接近的選項是 C (0.83)。

9. (5 分) An exam uses the following scoring system to discourage students from guessing or choosing randomly on multiple-choice questions: 10 points for each correct answer, 3 points for each unanswered question, and 0 points for each incorrect answer. If each question has five answer choices, what is the minimum number of choices a student must eliminate before guessing among the remaining choices becomes more advantageous than leaving the question unanswered?

- A. 0
- B. 1
- C. 2
- D. 3
- E. 4

解：答案：C (2)

設定：

- 答對得 10 分，答錯得 0 分，不答得 3 分
- 共 5 個選項
- 假設學生能排除 k 個錯誤選項，剩下 $5 - k$ 個選項猜測

猜測的期望值：

若剩下 $5 - k$ 個選項（其中 1 個正確），猜中的機率為 $\frac{1}{5-k}$

$$\mathbb{E}[\text{猜測得分}] = \frac{1}{5-k} \times 10 + \frac{5-k-1}{5-k} \times 0 = \frac{10}{5-k}$$

比較猜測與不答：

猜測較有利當：

$$\begin{aligned}\frac{10}{5-k} &> 3 \\ 10 &> 3(5-k) = 15 - 3k \\ 3k &> 5 \\ k &> \frac{5}{3} \approx 1.67\end{aligned}$$

因此， $k \geq 2$ 時，猜測比不答有利。

驗證：

- $k = 0$ (5 選 1)： $\mathbb{E} = \frac{10}{5} = 2 < 3 \Rightarrow$ 不答較好
- $k = 1$ (4 選 1)： $\mathbb{E} = \frac{10}{4} = 2.5 < 3 \Rightarrow$ 不答較好
- $k = 2$ (3 選 1)： $\mathbb{E} = \frac{10}{3} = 3.33 > 3 \Rightarrow$ 猜測較好

答案是 C (必須至少排除 2 個選項)。

10. (5 分) The commuting time for a student traveling from home to a college campus follows a normal distribution with a mean of 25 minutes and a standard deviation of 5 minutes. If the student leaves

home at 8:40 AM, what is the closest probability that they will arrive at the campus later than 9:10 AM?

- A. 0.05
- B. 0.16
- C. 0.68
- D. 0.84
- E. 0.95

解：答案：B (0.16)

設定：

- 通勤時間 $X \sim N(25, 5^2)$
- 出發時間：8:40 AM
- 需在 9:10 AM 前到達，允許時間 = 30 分鐘

求： $\mathbb{P}(X > 30)$ (遲到的機率)

標準化：

$$Z = \frac{X - 25}{5} = \frac{30 - 25}{5} = 1$$

$$\mathbb{P}(X > 30) = \mathbb{P}(Z > 1) = 1 - \Phi(1)$$

查表： $\Phi(1) \approx 0.8413$

$$\mathbb{P}(Z > 1) = 1 - 0.8413 = 0.1587 \approx 0.16$$

答案是 B (0.16)。

記憶：標準常態分配的 68-95-99.7 法則

- $\mathbb{P}(|Z| < 1) \approx 68\% \Rightarrow \mathbb{P}(Z > 1) \approx 16\%$
- $\mathbb{P}(|Z| < 2) \approx 95\% \Rightarrow \mathbb{P}(Z > 2) \approx 2.5\%$

11. (5 分) In a hypothesis test for the equality of two population means with unknown but equal variances, the test statistic is based on:

- A. The pooled variance estimate.
- B. The difference between the sample variances.
- C. The sum of the sample means.
- D. The larger of the two sample variances.
- E. The square of the standard deviations.

解：答案：A (The pooled variance estimate)

等變異數假設下的兩樣本 t 檢定：

當假設兩母體變異數相等 ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) 但未知時，使用合併變異數估計量 (pooled variance estimate)：

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

檢定統計量為：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

分析各選項：

- A) 正確：使用合併變異數估計
- B) 錯誤：不是用樣本變異數的差
- C) 錯誤：是用樣本平均數的差，不是和
- D) 錯誤：不是取較大的變異數
- E) 錯誤：不是標準差的平方（雖然這等於變異數，但描述不正確）

12. (5 分) When performing a two-way ANOVA, a significant interaction effect implies:

- A. Neither factor has a significant main effect.
- B. The effects of the two factors are independent.
- C. The effect of one factor depends on the level of the other factor.
- D. Both factors have significant main effects.
- E. The dependent variable is not normally distributed.

解：答案：C

交互作用效應 (Interaction Effect) 的定義：

在雙因子變異數分析中，**顯著的交互作用效應**意味著一個因子對依變數的效果取決於另一個因子的水準。

換句話說，兩因子的效果不是簡單相加的（不可加性）。

分析各選項：

- A) 錯誤：交互作用顯著不代表主效果不顯著
- B) 錯誤：這是交互作用不顯著時的情況
- C) 正確：這正是交互作用的定義
- D) 錯誤：主效果顯著與否與交互作用無必然關係

- E) 錯誤：常態性是 ANOVA 的假設，與交互作用無關

例子：研究藥物（有/無）和性別（男/女）對血壓的影響。若交互作用顯著，可能意味著藥物對男性有效但對女性無效（藥物效果依性別而定）。

13. (5 分) Which of the following is NOT an assumption of ordinary least squares (OLS) regression?
- A. The residuals are normally distributed.
 - B. The relationship between the dependent and independent variables is linear.
 - C. The residuals are homoscedastic.
 - D. The independent variables are uncorrelated.
 - E. The residuals are independent of each other.

解：答案：D

OLS 迴歸的標準假設（Gauss-Markov 假設）：

1. 線性： $Y = X\beta + \epsilon$ （模型正確設定）
2. 嚴格外生性： $\mathbb{E}[\epsilon|X] = 0$
3. 無完全多重共線性： X 的秩等於 k
4. 同質變異數： $\text{Var}(\epsilon|X) = \sigma^2$ (homoscedasticity)
5. 無自相關： $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ （誤差項獨立）

加上常態性假設（用於小樣本推論）： $\epsilon \sim N(0, \sigma^2)$

分析各選項：

- A)：殘差常態分配—是假設（用於推論）
- B)：線性關係—是假設
- C)：殘差同質變異數—是假設
- D)：自變數不相關—不是假設！OLS 允許自變數之間相關，只要求沒有完全多重共線性
- E)：殘差相互獨立—是假設

注意：自變數之間可以有相關性，只要不是完全線性相關（完全多重共線性）就可以。高度相關會導致估計量的標準誤增大，但 OLS 仍然有效。

14. (5 分) In hypothesis testing, what does the power of a test represent?
- A. The probability of rejecting a true null hypothesis.
 - B. The probability of rejecting a false null hypothesis.
 - C. The significance level of the test.
 - D. The probability of failing to reject a true null hypothesis.

E. The probability of obtaining a Type I error.

解：答案：B

統計檢定力 (Power) 的定義：

$$\text{Power} = 1 - \beta = \mathbb{P}(\text{拒絕 } H_0 | H_1 \text{ 為真})$$

即當對立假設為真時，正確拒絕虛無假設的機率。

假設檢定的四種結果：

	H_0 為真	H_1 為真
拒絕 H_0	型一錯誤 (α)	正確 (檢定力)
不拒絕 H_0	正確	型二錯誤 (β)

分析各選項：

- A)：拒絕真的 H_0 —這是**型一錯誤** (α)
- B) **正確**：拒絕假的 H_0 —這是**檢定力** ($\text{Power} = 1 - \beta$)
- C)：顯著水準—這是 α ，不是檢定力
- D)：未能拒絕真的 H_0 —這是正確決策， $= 1 - \alpha$
- E)：型一錯誤的機率—這是 α

For Questions 15–17, please read the scenario and output below.

A behavioral scientist wants to investigate whether three different teaching methods (Method A, Method B, and Method C) lead to different academic performance in students. A total of 90 students are randomly assigned to one of the three methods, with 30 students per group. At the end of the semester, their test scores (out of 100) are recorded. The scientist performs a one-way ANOVA, and the results are summarized below:

Source	df	SS	MS	<i>F</i>	p-value
Between Groups	2	450.00	225.00	7.50	0.001
Within Groups	87	2610.00	30.00		
Total	89	3060.00			

15. (5 分) What is the null hypothesis for the one-way ANOVA conducted in this scenario?
- All groups have the same sample size.
 - The variances within all groups are equal.
 - The mean test scores are equal across all three teaching methods.
 - At least one group has a higher variance than the others.
 - The distribution of scores is normal for all three methods.

解：答案：C

單因子變異數分析的假設檢定：

$$H_0 : \mu_A = \mu_B = \mu_C \text{ (三種教學方法的平均成績相等)}$$

$$H_1 : \text{至少有一組的平均成績與其他組不同}$$

分析各選項：

- A) 錯誤：樣本大小相等是設計，不是假設
- B) 錯誤：變異數相等是 ANOVA 的前提假設，不是被檢定的虛無假設
- C) 正確：ANOVA 檢定的虛無假設是各組平均數相等
- D) 錯誤：這不是 ANOVA 的虛無假設
- E) 錯誤：常態性是前提假設，不是被檢定的假設

16. (5 分) What conclusion can be drawn from the one-way ANOVA output?
- All three teaching methods lead to the same academic performance.
 - At least one teaching method leads to a significantly different mean test score compared to the others.
 - The variances of the test scores differ significantly across the three groups.
 - Teaching method is not a significant factor in academic performance.
 - Post-hoc tests are unnecessary because the *p*-value is less than 0.05.

解：答案：B

解讀 ANOVA 結果：

- $F = 7.50$
- $p\text{-value} = 0.001 < 0.05$

由於 $p < 0.05$ ，我們拒絕虛無假設。

分析各選項：

- A) 錯誤：這是接受 H_0 的結論，但我們拒絕了 H_0
- B) 正確：拒絕 H_0 意味著至少有一種教學方法的平均成績與其他方法顯著不同
- C) 錯誤：ANOVA 檢定的是平均數，不是變異數
- D) 錯誤：這與結論相反，教學方法是顯著因素
- E) 錯誤：正因為 $p < 0.05$ 且拒絕 H_0 ，事後檢定 (post-hoc) 是必要的，以確定哪些組之間有差異

17. (5 分) Which of the following assumptions is critical for the validity of both the ANOVA and the post-hoc Tukey's HSD tests?

- A. The mean differences among groups must be large.
- B. The sample sizes must be equal in each group.
- C. The population variances must be equal across all groups.
- D. The sample sizes must be greater than 30 in each group.
- E. The F -statistic must be greater than 5.

解：答案：C

ANOVA 和 Tukey HSD 的共同假設：

1. 獨立性：觀測值相互獨立
2. 常態性：各組資料來自常態分配
3. 變異數同質性 (Homogeneity of Variance)：各組的母體變異數相等

分析各選項：

- A) 錯誤：平均差異大小是我們要檢定的，不是假設
- B) 錯誤：樣本大小相等有助於穩健性，但不是必要假設
- C) 正確：變異數同質性是 ANOVA 和 Tukey HSD 的關鍵假設
- D) 錯誤：沒有 $n > 30$ 的要求
- E) 錯誤： F 統計量大小是檢定結果，不是假設

注意：Tukey HSD 特別假設各組樣本大小相等或變異數相等。當違反變異數同質性假設時，可使用 Welch's ANOVA 或 Games-Howell 事後檢定。

For Questions 18–20, please read the scenario and output below.

A marketing analyst is investigating the impact of advertising expense and product pricing on sales revenue for a retail chain. Across the chain's 200 stores, the analyst collects data on the following variables and enters in a multiple regression model:

- Y : Sales revenue (in thousands of dollars),
- X_1 : Advertising expense (in thousands of dollars),
- X_2 : Product unit price (in dollars),
- X_3 : The interaction term of X_1 and X_2

The regression model output is as follows, with $R^2 = 0.72$, adjusted $R^2 = 0.71$, and standard error of the estimate = 8.5.

Predictor	Coefficient (β)	Standard Error	t -statistic	p -value
Constant	20	5	4.00	0.001
X_1	2.5	0.3	8.33	0.001
X_2	-1.2	0.5	-2.40	0.018
X_3	0.04	0.01	4.00	0.001

18. (5 分) What can be concluded based on the regression coefficients?

- Advertising expense has no significant effect on sales revenue.
- Advertising expense increases sales revenue at a constant rate regardless of product unit price.
- Advertising expense increases sales revenue, but the effect depends on the product unit price.
- Advertising expense reduces sales revenue when product unit price is high.
- Advertising expense reduces sales revenue regardless of product unit price.

解：答案：C

迴歸模型：

$$Y = 20 + 2.5X_1 - 1.2X_2 + 0.04X_1X_2$$

廣告支出對銷售收入的邊際效應：

$$\frac{\partial Y}{\partial X_1} = 2.5 + 0.04X_2$$

這個邊際效應取決於產品單價 X_2 。

分析各選項：

- A) 錯誤： X_1 的 p -value = 0.001 < 0.05，效果顯著
- B) 錯誤：由於有顯著的交互作用項 (X_3 的 $p = 0.001$)，效果不是固定的
- C) 正確：廣告支出增加銷售收入 ($\beta_1 = 2.5 > 0$)，但效果取決於產品單價 (因為交互作用項顯著)
- D) 錯誤： $\frac{\partial Y}{\partial X_1} = 2.5 + 0.04X_2 > 0$ (對合理的 X_2 值)，所以廣告總是增加銷售
- E) 錯誤：廣告增加 (不是減少) 銷售收入

解釋：當產品單價越高時 (X_2 越大)，廣告支出對銷售的正面效果越強 (邊際效應 $2.5 + 0.04X_2$ 越大)。

19. (5 分) Which of the following statements about R^2 is correct?

- A. 72% of the variability in advertising expense is explained by sales revenue.
- B. 72% of the variability in sales revenue is explained by advertising expense, price, and their interaction.
- C. 72% of the variability in sales revenue is explained by advertising expense alone.
- D. $R^2 = 0.72$ indicates the model overfits the data.
- E. 72% of the model's predictions are accurate.

解：答案：B

R^2 (判定係數) 的定義：

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} = \frac{SS_{regression}}{SS_{total}}$$

R^2 表示依變數的變異中，被自變數解釋的比例。

分析各選項：

- A) 錯誤：方向搞反了。不是「廣告支出被銷售收入解釋」，而是「銷售收入被自變數解釋」
- B) 正確：72% 的銷售收入變異被廣告支出、產品單價及其交互作用所解釋
- C) 錯誤：是被所有自變數 (X_1 、 X_2 、 X_3) 解釋，不只是廣告支出
- D) 錯誤： $R^2 = 0.72$ 不代表過度配適。過度配適通常在 R^2 非常高 (如 0.99) 且調整後 R^2 明顯較低時才考慮
- E) 錯誤： R^2 不是預測準確率的意思

20. (5 分) Which of the following assumptions must hold true for the p -values of the predictors to be valid?

- A. X_3 must be statistically significant.
- B. The residuals must have constant variance across all levels of the predictors.
- C. The residuals must be positively correlated with the predictors.
- D. The predictors must be uncorrelated with each other.
- E. The sample size must be greater than 100.

解：答案：B

迴歸分析中 p -value 有效性的假設：

為了使 t 統計量和對應的 p -value 有效，需要滿足以下假設：

1. 線性： Y 與 X 之間的關係是線性的
2. 獨立性：殘差相互獨立
3. 常態性：殘差服從常態分配
4. 同質變異數 (Homoscedasticity)：殘差的變異數在所有預測值水準上都相同

分析各選項：

- A) 錯誤： X_3 是否顯著是檢定結果，不是假設
- B) 正確：殘差必須有**常數變異數**（同質變異數），這是 OLS 推論有效的關鍵假設
- C) 錯誤：殘差應該與預測變數**不相關**，不是正相關
- D) 錯誤：預測變數之間可以相關，只要沒有完全多重共線性
- E) 錯誤：沒有 $n > 100$ 的硬性要求

補充：當違反同質變異數假設時，OLS 估計量仍然不偏，但標準誤估計會有偏，導致 p -value 不可靠。此時可使用穩健標準誤（robust standard errors）。