國立中正大學
115學年度碩士班招生考試
試題
[第3節]

科目名稱	統計學
系所組別	經濟學系國際經濟學-乙組

—作答注意事項—

※作答前請先核對「試題」、「試卷」與「准考證」之系所組別、科目名稱是否相符。

1. 預備鈴響時即可入場，但至考試開始鈴響前，不得翻閱試題，並不得書寫、畫記、作答。
2. 考試開始鈴響時，即可開始作答；考試結束鈴響畢，應即停止作答。
3. 入場後於考試開始40分鐘內不得離場。
4. 全部答題均須在試卷（答案卷）作答區內完成。
5. 試卷作答限用藍色或黑色筆（含鉛筆）書寫。
6. 試題須隨試卷繳還。

國立中正大學115學年度碩士班招生考試試題

第一部分：單選題（一共五題，每題2分）

1. 關於點估計量的「有效性（Efficiency）」，下列敘述何者「錯誤」？
   (A) 有效性是用來比較兩個「不偏估計量」優劣的準則，通常變異數較小者被認為較有效率
   (B) 若估計量$\widehat{\theta_1}$與$\widehat{\theta_2}$皆為θ的不偏估計量，且$Var(\widehat{\theta_1}) < Var(\widehat{\theta_2})$，則稱$\widehat{\theta_1}$相對於$\widehat{\theta_2}$具有相對有效性
   (C) 一個估計量的有效性越高，代表在相同的樣本數下，該估計量產生的估計值會更集中在母體參數真值附近
   (D) 有效性僅與估計量的不偏性（Unbiasedness）有關，只要一個估計量是不偏的，它在任何條件下都必定是最有效的估計量
   (E) 在常態母體下，樣本平均數$\bar{X}$是母體平均數μ的最小變異不偏估計量（UMVUE），因此它比樣本中位數更具備有效性
2. 某咖啡店宣稱其自動咖啡機平均每杯裝填量為200毫升。品質管制員隨機抽取16杯咖啡進行測試，發現樣本平均數為194毫升，樣本標準差為 12毫升。假設咖啡裝填量服從常態分配，在顯著水準$\alpha = 0.05$下，測試該咖啡機是否「低於」其宣稱的裝填量，試問下列何者正確？($Z_{0.05} = 1.645$、$Z_{0.025} = 1.96$、$t_{0.05}(15) = 1.753$、$t_{0.05}(16) = 1.746$)
   (A) 使用Z檢定，統計量$Z = -2.00$，拒絕域為$Z < -1.645$
   (B) 使用t檢定，統計量$t = -2.00$，拒絕域為$t < -1.746$
   (C) 使用Z檢定，統計量$Z = -0.50$，拒絕域為$Z < -1.645$
   (D) 使用t檢定，統計量$t = -2.00$，拒絕域為$t < -1.753$
   (E) 使用t檢定，統計量$t = -0.50$，拒絕域為$t < -1.753$
3. 某工程師欲比較兩家供應商（A與B）所提供的零件平均拉伸強度。他從兩家供應商中隨機抽取了大樣本進行測試，結果如下：
   供應商 A：樣本數$n_1 = 40$，樣本平均數$\overline{x_1} = 125kg/cm^2$，樣本標準差$s_1 = 12kg/cm^2$
   供應商 B：樣本數$n_2 = 50$，樣本平均數$\overline{x_2} = 120kg/cm^2$，樣本標準差$s_2 = 15kg/cm^2$
   在顯著水準$\alpha = 0.05$下，欲檢定兩家供應商零件的平均強度是否存在顯著差異，請問該檢定的檢定統計量值與結論為何？($Z_{0.05} = 1.645$、$Z_{0.025} = 1.96$、$t_{0.025}(88) = 1.987$)
   (A)使用t檢定，統計量$t = 1.74$，結論為不拒絕H_0
   (B)使用Z檢定，統計量$Z = 1.74$，結論為不拒絕H_0
   (C)使用t檢定，統計量$t = 1.74$，結論為拒絕H_0
   (D)使用Z檢定，統計量$Z = 1.72$，結論為不拒絕H_0
   (E)使用t檢定，統計量$t = 1.72$，臨界值使用$t_{0.025}(88) = 1.987$，結論為不拒絕H_0
4. 假設有兩個獨立隨機變數X與Y，已知$X \sim N(\mu X, \sigma_X^2)$且$Y \sim N(\mu Y, \sigma_Y^2)$。定義新的隨機變數$W = aX + bY$(其中$a, b$為非零常數）。關於$W$的性質，下列敘述何者錯誤？
   (A) 期望值具有線性特質，W的期望值$E[W]$必等於$a\mu X + b\mu Y$，此性質不論X, Y是否獨立皆成立
   (B) 若定義$D = X - Y$，則其變異數$Var(D)$等於$\sigma_X^2 + \sigma_Y^2$，而非兩變異數之差
   (C) 由於X與Y為獨立常態變數，W的動差母函數可寫為$M_W(t) = M_X(at) \cdot M_Y(bt)$
   (D) 若樣本數足夠大，則$W = aX + bY$之變異數的估計值將符合自由度為$n_X + n_Y - 2$的常

態分配

(E) 根據常態分配的特性，兩個獨立常態隨機變數的任何線性組合$W = aX + bY$依然會精確地服從常態分配，而不僅僅是近似

5. 關於單選變異數分析（One-Way ANOVA）的敘述，下列何者正確？
   (A) ANOVA 的虛無假設(H_0) 是各組樣本的平均值皆不相等
   (B) ANOVA 的基本原理是將總變異拆解為組間變異與組內變異，並藉此判斷各組平均數是否有顯著差異
   (C) 當F檢定值越接近 0 時，代表各組之間的平均值差異越顯著，越容易拒絕虛無假設
   (D) ANOVA 檢定結果若達到顯著水準($p < 0.05$)，則代表每一組與其他各組之間都有顯著差異
   (E) 執行 ANOVA 之前，不需要考慮資料是否符合常態分佈或變異數同質性，因為F檢定具有完全的穩健性(Robustness)

第二部分：複選題（一共有十題，每題3分）

6. 關於樣本變異數（Sample Variance）的描述，下列哪些選項是正確的？
   (A) 樣本變異數的數值永遠不會是負數
   (B) 樣本變異數的單位與原始資料的單位相同
   (C) 計算樣本變異數時，分母通常使用$n-1$而非樣本數n，這是為了得到母體變異數的不偏估計值
   (D) 若樣本中所有數值都增加 10，則樣本變異數也會隨之增加 10
   (E) 樣本變異數極易受極端值影響，因為計算過程中使用了離均差的絕對值
7. 關於盒鬚圖的描述，下列哪些選項是正確的?
   (A) 若盒鬚圖中的中位數線偏向盒子的左側，表示該數據呈現左偏分布
   (B) 盒子的長度被稱為四分位距，可用來衡量數據的離散程度
   (C) 盒子內部垂直線代表該組數據的中位數，而非平均數
   (D) 若盒鬚圖的鬚向右對稱，則表示平均數、中位數、眾數勢必在同一個位置上
   (E) 若將原始數據中的每一個數值都加上 100，則盒鬚圖中盒子的長度也會隨之增加 100
8. 關於機率論中古典機率、客觀機率與主觀機率之描述，下列哪些選項是正確的？
   (A) 主觀機率雖然依賴個人信念，但仍必須符合機率公設，例如互斥事件的機率必須具有可加性
   (B) 根據大數法則，當隨機實驗次數n趨近於無限大時，事件發生的相對次數必定會等於該事件的古典機率
   (C) 若一隨機實驗的樣本空間為無限不可數，則無法使用古典機率之原始定義進行計算
   (D) 若一隨機實驗結果只有發生與不發生兩種，則根據古典機率定義，該事件發生的機率必為 0.5
   (E) 在古典機率的架構下，若事件A與B互斥，則$P(A \cup B)$等於$P(A) + P(B)$；但在主觀機率下，此邏輯不一定成立
9. 關於獨立事件與互斥事件之性質描述，下列哪些選項是正確的？
   (A) 若A與B為兩個發生機率皆大於 0 的事件，且A與B互斥，則A與B勢必不獨立

(B) 對於任意兩個發生機率大於 0 的獨立事件，$A \cup B$發生的機率恆等於$P(A)+P(B)-P(A)P(B)$

(C) 某科技公司同時研發兩項不相關的技術A與B，已知技術A研發成功的機率為 0.6，技術B研發成功的機率為 0.5。由於兩項技術由不同團隊在獨立環境下研發，因此若技術A研發成功，則技術B研發成功的機率將會下降至 0

(D) 事件A與B獨立的充要條件為條件機率$P(A|B)=P(A|B^c)$，前提是相關機率皆存在且有意義

(E) 若事件A與B互相獨立，則其補集合A^c與B^c亦必相互獨立

10. 關於各離散型機率分配之情境與性質描述，下列哪些選項是正確的？

(A) 某抽獎箱中有 10 張彩券，其中 3 張有獎。若某人連續抽取 5 張彩券（採取出不放回方式），則中獎張數的分布適合以二項分配來描述

(B) 根據過去經驗，某路口平均每個月發生 2 次車禍，若政府新設紅綠燈後，我們欲觀察「直到發生下一次車禍為止所經過的時間」，此時間長度適合以泊松分配來模擬

(C) 一位保險業務員向潛在客戶進行推銷，已知成交率為 10%，若欲計算該業務員在遇到第一個成交客戶前，必須拜訪的客戶數之機率，適合使用幾何分配

(D) 在一家大型工廠中，產品的不良率固定為 2%，若隨機抽取 50 件產品進行檢測，則發現不良品的個數適合以二項分配來模擬

(E) 某大賣場平均每小時有 10 位顧客進入，若欲估計未來半小時內剛好有 3 位顧客進入的機率，適合使用泊松分配

11. 關於隨機變數X的動差母函數$M_X(t)=E[e^{tX}]$，下列敘述何者正確?

(A) 若動差母函數$M_X(t)$在包含原點的開區間內存在，則其$t=0$處的n階導數$M_X^n(0)$即為X的n階原始動差$E[X^n]$

(B) 對於任何隨機變數X，其動差母函數$M_X(t)$在整個實數數線上$(t \in \mathbb{R})$皆保證存在且有限

(C) 若隨機變數X與Y互相獨立，則其和$Z=X+Y$的動差母函數為兩者動差母函數的乘積，即$M_Z(t)=M_X(t)\cdot M_Y(t)$

(D) 若隨機變數X的動差母函數存在，則其一階導數在$t=0$的值$M_X{}'(0)$等於該隨機變數的變異數$Var(X)$

(E) 動差母函數$M_x(t)$永遠是一個關於t的奇函數(Odd function)，即滿足$M_X(-t)=-M_X(t)$

12. 關於超幾何分配、二項分配、泊松（Poisson）分配與常態分配之間的關係，下列敘述何者正確？

(A) 當超幾何分配中的母體總數N趨近於無窮大時，若抽取次數n保持不變，超幾何分配會趨近於泊松分配

(B) 對於二項分配$B(n,p)$，只要樣本數n足夠大，無論成功機率p的數值為何，皆應優先使用泊松分配來進行近似

(C) 泊松分配$P(\lambda)$的期望值與變異數並不相等，當λ增加時，其分佈會呈現嚴重的右偏

(D) 當二項分配$B(n,p)$的試驗次數n很大，且成功機率p未過於接近 0 或 1 時（例如滿足$np \geq 5$且$n(1-p)\geq 5$），可以用常態分配作為其近似

(E) 泊松分配可以視為二項分配在試驗次數n趨向無窮大、成功機率p趨向於 0，且乘積$np=\lambda$保持為常數時的極限形式

13. 關於隨機變數X與Y之間的獨立性與共變數之關係，下列敘述何者正確？
   (A) 若兩個隨機變數X與Y的共變數$Cov(X,Y)=0$，則可以推論X與Y必定相互獨立
   (B) 若隨機變數X與Y相互獨立，則它們的共變數$Cov(X,Y)$必定等於 0
   (C) 兩隨機變數的共變數$Cov(X,Y)$之數值越大，代表這兩個隨機變數的獨立性越強
   (D) 對於任意兩個隨機變數X與Y，其和的變異數公式為$Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$
   (E) 若X與Y不獨立，則其共變數$Cov(X,Y)$絕對不可能等於 0
14. 關於抽樣分配（Sampling Distribution）的特性，下列敘述何者正確？
   (A) 根據中央極限定理，不論母體分佈為何，只要樣本數n足夠大，樣本平均數$\bar{X}$的抽樣分配都會趨近於常態分配
   (B) 樣本平均數$\bar{X}$的標準差（又稱為標準誤，Standard Error）會隨著樣本數n的增加而變大
   (C) 若母體不服從常態分配，則樣本平均數的期望值$E[\bar{X}]$也不會等於母體平均數μ
   (D) 當母體服從常態分配時，無論樣本數n的大小為何，樣本平均數$\bar{X}$的抽樣分配必定也是常態分配
   (E) 樣本比例$\hat{p}$的抽樣分配之變異數，與樣本數n和母體比例p的數值有關
15. 在進行平均數的假設檢定時，關於何時「必須」使用t檢定而「不能」以Z檢定替代，下列敘述何者正確?
   (A) 只要母體分佈不服從常態分配，無論樣本數大小或母體變異數是否已知，都絕對不能使用Z檢定
   (B) 當樣本數n小於 30 且母體變異數σ^2為已知時，基於保守原則，仍必須使用t檢定以獲得更精確的臨界值
   (C) 當母體變異數σ^2未知，且樣本數n較小(通常小於 30)時，必須使用t檢定，因為此時樣本變異數S^2的估計誤差不可忽略
   (D) 在母體變異數未知的情況下，t分配會隨著自由度的增加而趨近於標準常態分配，但在小樣本下，使用Z分配會導致型一錯誤的機率被低估
   (E) 只有在樣本比例的檢定中，才可以使用Z檢定替代t檢定；對於平均數檢定，兩者在邏輯上完全不具備替代性

第三部分：是非題（一共五題，每題2分）

16. 型一錯誤與型二錯誤通常呈現互相消長的情況，若要減少α值勢必只能增加β值
17. 在進行兩獨立常態母體平均數差檢定時，若母體變異數未知但已知相等，可利用合併樣本變異數（Pooled Variance）估計共同變異數。然而，最終應選擇使用t檢定統計量或Z檢定統計量，取決於樣本數的大小
18. 在其他條件不變下，影響信賴區間長度的因素包含：點估計式抽樣分配的變異程度、樣本大小、信賴係數的大小，以及機率分配中上下限臨界值的選取方式
19. P值是指在虛無假設為真的條件下，出現樣本結果或更極端結果的機率。當P值大於顯著水準α時，代表拒絕虛無假設的證據非常充足
20. 在進行單因子變異數分析時，我們計算F檢定統計量的目的，是為了檢定各個樣本組別的變異數(Variance)是否具有顯著差異，進而判斷不同組別是否來自相同的母體

第四部分：填空題（每格 2.5 分，共 50 分）

注意事項:

(1) 此部分不須計算過程。

(2) 此部分請不要使用「選擇題作答區」作答。

(3) 此部分請自行於作答區的適當位置製作如下的填空題作答區:

(1)	(2)	(3)	(4)	(5)
(6)	(7)	(8)	(9)	(10)
(11)	(12)	(13)	(14)	(15)
(16)	(17)	(18)	(19)	(20)

1. (30%) 考慮一線性迴歸模型：$y_i = \beta x_i + e_i$，$i = 1,\cdots,n$。若對所有 i，x_i 之實現值皆為 1，則我們無法以普通最小平方 (OLS) 法估計參數 β，此敘述為____(1)____(此空格請填寫正確或錯誤或無法判斷)。給定動差條件 $\mathrm{E}[x_i(y_i - \beta x_i)] = 0$，則應用動差法 (method of moments) 所得參數 β 的估計式為 $\hat{\beta} =$____(2)____，且此估計式與 OLS 估計式____(3)____(此空格請填寫相等或不等或無法判斷)。令 $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$，$\hat{y}_i = \hat{\beta} x_i$，$\hat{e}_i = y_i - \hat{y}_i$，而 $R^2 = 1 - (\sum_{i=1}^{n} \hat{e}_i^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2)$，則我們可知 R^2 之值有可能為負值，此敘述為____(4)____(此空格請填寫正確或錯誤或無法判斷)。若 $\{(x_i, y_i): i = 1,\cdots,n\}$ 為隨機樣本 (random sample)，而 x_i 之實現值會隨 i 而變且 $\mathrm{E}[y_i|x_i] = 1 + x_i$，則 $\mathrm{E}[\hat{\beta}] =$____(5)____，故 $\hat{\beta}$ 具有不偏性 (unbiasedness)，此敘述為____(6)____(此空格請填寫正確或錯誤或無法判斷)。若 $\mathrm{var}(y_i|x_i) = x_i^2$，則我們可知 y_i 具有同質變異 (homoskedasticity)，此敘述為____(7)____(此空格請填寫正確或錯誤或無法判斷)。此時 $\mathrm{var}(\hat{\beta}|x_1,\cdots,x_n) =$____(8)____且 $\hat{\beta}$ 為最佳線性不偏估計式 (best linear unbiased estimator)，此敘述為____(9)____(此空格請填寫正確或錯誤或無法判斷)。現給定 4 筆觀察值: $y_1 = 1$，$y_2 = -1$，$y_3 = 2$，$y_4 = 0$ 而 $x_1 = 1$，$x_2 = 0$，$x_3 = 1$，$x_4 = 0$，則 $\hat{\beta}$ 的實現值為____(10)____，$\sum_{i=1}^{4} \hat{y}_i =$____(11)____，和 $\sum_{i=1}^{4} \hat{e}_i =$____(12)____。
2. (10%) 考慮一線性迴歸模型：$\mathrm{y} = \mathrm{X}\beta + e$，其中 y 和 e 為 $n \times 1$ 向量，X 為 $n \times k$ 矩陣而 β 為 $k \times 1$ 向量。欲以普通最小平方 (OLS) 法估計參數 β，我們須要求 rank(X) =____(13)____。在此條件下，OLS 估計式可表示為 $\hat{\beta} =$____(14)____。若 $\mathrm{E}[\mathrm{y}|\mathrm{X}] = \mathrm{X}\beta_0$，則 $\mathrm{E}[\hat{\beta}] =$____(15)____。若 $\mathrm{y}|\mathrm{X} \sim N(\mathrm{X}\beta_0, \sigma_0^2 \mathrm{I}_n)$，則 $\hat{\beta}$ 的條件分配 (conditional distribution) 為 $\hat{\beta}|\mathrm{X} \sim$____(16)____(此空格請寫下其分配名稱或常用的符號；若該分配有參數，則亦須正確寫出)。
3. (10%) 考慮一時間序列: $y_t = y_{t-1} + e_t$，$t = 1,2,\cdots$。令 $y_0 = 0$ 而 $e_t \sim \mathrm{i.i.d.}\, N(0,1)$，則我們可計算出 $\mathrm{E}[y_t] =$____(17)____，$\mathrm{var}(y_t) =$____(18)____，和 $\mathrm{cov}(y_t, y_{t-1}) =$____(19)____。由此可知，此時間序列為____(20)____(此空格請填寫定態 (stationary) 或非定態 (non-stationary) 或無法判斷)時間序列。