

國立中央大學97學年度碩士班考試入學試題卷

所別：資訊管理學系碩士班(甲乙組) 科目：統計學 共 2 頁 第 1 頁

\*請在試卷答案卷(卡)內作答

- (4分) 假說檢定 (Hypothesis Testing) 之基本觀念：
  - 假說檢定是統計推論的一種方法，也是人類科學演進過程中一個重要的思維，請說明  $H_0$  與  $H_1$  的內涵。
  - 天文學發展過程中， $H_0$ : 托勒密 (Ptolemy) 的地心說， $H_1$ : 哥白尼 (Copernicus) 的日心說。請描述其 Type II error。
- (4分) 假說檢定中之檢定法則與檢定能力：
  - 何謂檢定法則？(2) 用來評估一個檢定法則的鑑別力稱為檢定能力 (power function)，其中  $\text{Power} = p(\text{推翻 } H_0)$ ，在不同的世界去量會有不同的結果。而一個好的檢定法則，表示鑑別力很強，請問什麼是「鑑別力很強」呢？(提示：可以以條件機率的形式來說明)
- (4分) 當我們以樣本統計量 (statistic) 去推測 (estimate, 或翻譯成估計) 母體參數 (parameter) 時，通常會要求在「95%的信賴區間」或「95%的信心水準」下去作推測，請說明這裡所謂的「95%信賴區間」或「95%信心水準」的意義。

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

- (4分) 變異數  $\sigma^2$  之基本觀念：
  - 為什麼變異數可以用來表示母體資料分散的程度呢？
  - 假設  $E(X)=\mu$ ,  $V(X)=\sigma^2$ ,  $X_1, X_2, \dots, X_n$  iid, 請簡單證明  $\sigma_{\bar{X}}^2 = V(\bar{X}) = \frac{\sigma^2}{n}$
- (4分) 相關係數 (coefficient of correlation)  $\rho$  之基本觀念：
  - 為什麼  $\rho_{XY}$  可以用來衡量隨機變數 X 與隨機變數 Y 是否存在著線性關係呢？
  - 請分別說明  $\rho=1$  及  $\rho=0$  所代表的意義。
- (4分) 假設在中央大學校園違規停車的話，會收到一張違規罰單，根據校警統計：在所有違規停車的車輛中，80%的違規停車者會被開一張違規罰單。你認為在下學期的你的汽車會因為兩次中央大學的盛大活動而違規停車兩次，令 X=下學期你會收到的違規罰單的張數。
  - 請寫出隨機變數 X 的機率分配。(2) 用  $E(X)=?$
- (4分) 假設有三位同學進行網路線上 100 公尺賽跑，總比賽次數至少 30 次，每一次比賽只希望選出冠軍者，不排名次，每一次比賽的冠軍者可以不只一人，請描述比賽結果的樣本空間 (sample space)。

參考用

- (4分) C 大 ERP 中心培訓企業送訓人員數目如下表，計算 95 及 96 兩年度之平均成長率。

年度	94	95	96
人數	100	200	328

- (8分) 為將樣本數(n) 為 400 的原始觀察值分組作直方圖，常用的組數(k) 估計公式有以下 A, B, C 三種，試分析評論上述公式的適用性，說明妳(你)選用的組數及說明理由。註： $\log(400) = 2.60$ 。
  - 估計組數  $(k) \geq 1 + 3.32 \log(\text{樣本數})$ ，即  $k \geq 1 + 3.32 \log(n)$ 。
  - 估計組數  $(k) \geq \sqrt{\text{樣本數}}$ ，即  $k \geq \sqrt{n}$ 。
  - 估計組數  $(k) \geq 2^{\sqrt{\text{樣本數}}}$ ，即  $2^k \geq n$ 。
- (25分) C 大 ERP 中心為評核企業送訓人員培訓成效，隨機抽樣十位學員，訓練前後的測試成績如下：

學員	1	2	3	4	5	6	7	8	9	10
訓練前	66	48	84	63	71	81	53	81	56	54
訓練後	91	55	84	71	73	69	49	75	84	77

- (5分) 若訓練前後的測試成績皆呈常態分配，在  $\alpha=0.05$  下，為判斷訓練是否有顯著成效，由假說設定開始，試詳細說明妳(你)的分析步驟。
- (5分) 若訓練前後的測試成績未知何種分配，在  $\alpha=0.05$  下，若只考量每位學員訓練前後是否有進步，為判斷此訓練計畫是否有顯著成效，由假說設定開始，試詳細說明妳(你)的分析步驟。
- (5分) 若訓練前後的測試成績未知何種分配，在  $\alpha=0.05$  下，若考量每位學員訓練前後是進步大小幅度，為判斷此訓練計畫是否有顯著成效，由假說設定開始，試詳細說明妳(你)的分析步驟。
- (5分) 若訓練前後的測試成績皆呈常態分配，在  $\alpha=0.05$  下，為判斷訓練後的成績是否與訓練前的成績有相關性，由假說設定開始，試詳細說明妳(你)的分析步驟。
- (5分) 若訓練前後的測試成績未知何種分配，在  $\alpha=0.05$  下，為判斷訓練後的成績是否與訓練前的成績有相關性，由假說設定開始，試詳細說明妳(你)的分析步驟。註：查表值可以符號代入，但必須註明統計分布。
- (10分) 某廠商宣稱其所製的大燕麥片罐的平均重量為 3 磅以上，今自其中隨機抽取出 36 罐測定其重量，得出平均重量 2.93 磅，假設母體標準差為 0.36 磅，試檢驗此廠商的宣稱 (顯著水準  $\alpha=0.01$ ,  $z_{0.01}=2.33$ )。
  - 請寫出虛無假說。
  - 試以臨界值法檢定假說。
  - 試以信賴區間檢定假說。
  - 檢定結果之意義為何。

注：背面有試題

國立中央大學97學年度碩士班考試入學試題卷

所別：資訊管理學系碩士班(甲乙組) 科目：統計學 共 2 頁 第 2 頁

\*請在試卷答案卷(卡)內作答

12. (10分) 進行兩母體平均數比較時，有獨立樣本設計與相依樣本設計二種，請從標準誤的估計與自由度的角度討論二者的差異。
13. (5分) 請說明在二因子 ANOVA 結果分析中，交互作用顯著與不顯著如何影響資料分析？
14. (10分) 有一組共 13 名男性的心臟收縮血壓與體重及年齡之關係表，如表 1，複迴歸分析結果如表 2。  
 (1) 請寫出假說。(2) 請列出線性複迴歸方程式並說明其意義。(3) 請依數據佐證說明假說檢定結果。

表 1. 心臟收縮血壓與體重及年齡關係表

體重 (磅) $x_1$	年齡 $x_2$	血壓 $y$
152	50	120
183	20	141
171	20	124
165	30	126
158	30	117
161	50	129
149	60	123
158	50	125
170	40	132
153	50	123
174	40	132
190	40	155
185	20	147

表 2. 複迴歸電腦輸出表

THE REGRESSION EQUATION IS  
 $Y = -65.1 + 1.08X_1 + 0.425X_2$

	COLUMN	COEFFICIENT	ST. DEV. OF COEF.	T-RATIO= COEF/S. D.
	--	-65.09	14.94	-4.35
X1	C1	1.07710	0.07707	13.97
X2	C2	0.42541	0.07315	5.81

THE ST. DEV. OF Y ABOUT REGRESSION LINE IS  
 $S = 2.508$   
 WITH (13-3) = 10 DEGREES OF FREEDOM

R-SQUARED = 95.8 PERCENT

ANALYSIS OF VARIANCE

DUE TO	DF	SS	MS=SS/DF
REGRESSION	2	1423.837	711.918
RESIDUAL	10	62.931	6.293
TOTAL	12	1486.769	

注意：背面有試題