

※ 注意：請於試卷內之「非選擇題作答區」標明題號依序作答。

第一部分：問答題，每題 6 分，共 24 分

1. 請解釋在做假設檢定(hypothesis testing)的時候，p 值為 0.03 ($p\text{-value} = 0.03$)的意義。
2. 假設我們要做一個臨床試驗比較藥物 A 和藥物 B 之間的療效，藥物 B 是目前最常被使用的藥物，而我們試驗目的只要證明藥物 A 和藥物 B 之間的療效差不多就算成功了，請寫出這個臨床試驗的虛無假設(null hypothesis)和對立假設(alternative hypothesis)。
3. 請解釋 analysis of variance (ANOVA)是如何檢定來多組獨立樣本的平均值差異。
4. 請說明在進行線性迴歸分析 (regression analysis) 時，如果解釋變數的共線性(multicollinearity)很高，會對模型的參數估計產生甚麼樣的影響。

第二部分：簡答題，共 26 分

1. 一個探討出生嬰兒體重和母親年齡的關係的研究，共搜集了 189 個剛出生嬰兒和母親的相關資料，包括嬰兒體重是否過低 (1 是或 0 否)、母親年齡(以年為單位)、母親是否有高血壓(1 有或 0 沒有)以及母親抽菸習慣(1 有或 0 沒有) 等等。logistic 迴歸分析的結果如下：

	Coefficient	Standard error	P-value
截距	-0.017	0.769	0.982
母親年齡	-0.050	0.032	
母親有高血壓	1.234	0.621	
母親有抽菸習慣	0.701	0.326	

- (1)請用上面 logistic 迴歸分析的結果解釋嬰兒體重是否過低和母親年齡、母親是否有高血壓以及母親抽菸習慣之間的關係 (5 分)。
 - (2)設 type-I error rate 為 0.05，請問母親年齡、母親是否有高血壓以及母親抽菸習慣這三個變數在統計上是否顯著 (5 分)。
2. 假設某一數學考試總共有 10 題選擇題，每一題都有 5 個可能答案，而每一題都只有一個是正確答案。如果某學生純粹以猜答案的方式來答題，而隨機變數 Y 代表此學生猜對的題數。則
 - (1) 隨機變數 Y 的機率分配為_____ (2 分)。
 - (2) 隨機變數 Y 的平均數=_____ (2 分)及變異數=_____ (2 分)。
 - (3) $P(Y \geq 3) =$ _____ (3 分)。
 3. X 為一隨機變數，服從均勻分配(Uniform distribution)，其機率密度函數如下，
$$f(x) = \begin{cases} 1/10 & 20 \leq x \leq 40 \\ 0 & \text{其他} \end{cases}$$
 - (1) 則 X 之平均數=_____ (2 分)與變異數=_____ (2 分)。
 - (2) $P(26 < X < 33) =$ _____ (3 分)。

見背面

第三部分：每題 10 分，共 50 分

社區 A 和 B 其第二型糖尿病(Type 2 Diabetes, 簡稱 T2DM)盛行率分別為 5% 及 10%, 今有一新穎檢驗方法可以早期偵測 T2DM, 以此工具診斷後得到盛行率分別為 23% 及 26%。

1. 請以適當統計分佈敘述 T2DM 盛行率之隨機變數特性。
2. 假設此新檢驗工具正確性不會因社區不同而有差異，請計算以此樣本所得之 $P(\text{檢驗陽性}|\text{T2DM})$ 及 $P(\text{檢驗陰性}|\text{非 T2DM})$ 。
3. 計算 A 社區及 B 社區檢驗後陽性者得 T2DM 的機率(分別以 $P_A(\text{T2DM}|\text{檢驗陽性})$ 及 $P_B(\text{T2DM}|\text{檢驗陽性})$ 表示之)。
4. 請以統計觀念說明為何 $P_A(\text{T2DM}|\text{檢驗陽性})$ 及 $P_B(\text{T2DM}|\text{檢驗陽性})$ 會比原來盛行率高。
5. 如果 $P_A(\text{T2DM}|\text{檢驗陽性}) = 17.4\%$ 且 $P_B(\text{T2DM}|\text{檢驗陽性}) = 14.3\%$, 請以統計觀念解釋為何在新的檢驗工具施行之下結果和原來兩社區盛行率比較結果(5% vs 10%)相反?

試題隨卷繳回