

第一題 (共 50 分)： 注意：請於試卷內之「非選擇題作答區」依序作答，並應註明作答之大題及小題題號。

- (10 分) 某工廠有 50 位作業員，欲確保抽到暴露矽塵濃度前 10% 作業員 (至少一位) 的機率高達 90%，最少需隨機抽樣多少位作業員？(可不計算出最後結果，但需寫出想法與推導過程)
- (10 分) 請說明二項分佈(Binomial distribution)與卜瓦松分布(Poisson distribution)之關聯，二者何時可以近似，需寫出推導過程。
- (10 分) 某醫學中心收到 360 個「病例親本三元體」(從染病小孩收樣，並邀請其父母親一同參與，父母親疾病狀態未知) 某「單核苷酸多態性」(Single-Nucleotide Polymorphism, SNP) 的基因型如下，請以統計方法檢定該 SNP 與疾病是否有顯著關聯。顯著水準設為 0.05。請寫出虛無假說、對立假說、檢定統計量、虛無假說下該檢定統計量之分布、拒絕域。如有自設符號請定義。

親本一基因型	親本二基因型	染病小孩基因型	家庭數
CC	CT	CT	100
CC	TT	CT	150
CT	TT	CT	50
CT	TT	TT	60

- 研究三種運動種類 (慢跑、游泳、重量訓練) 對降低體脂肪率的功效，找來若干位 30 歲左右的女性，每週進行三次每次 30 分鐘的運動，於三個月後，三組女性體脂肪率 (單位：%) 變化如下 (變化量是後測減前測)：

慢跑組	游泳組	重訓組
-4	-3	-4
-3	-1	-1
+1	+1	+2
-5	+3	+1
-2	+2	-2
+2	-4	-3
-1	+1	+2
-3	-3	-1
-2	+2	
+3		

(A) (10 分) 請以「變異數分析」(Analysis of variance, ANOVA) 檢定運動種類是否影響體脂肪率降低程度。顯著水準設為 0.05。請寫出虛無假說、對立假說、檢定統計量、虛無假說下該檢定統計量之分布、拒絕域。如有自設符號請定義。

(B) (10 分) 請說明執行「變異數分析」需滿足哪些條件與假設，及如何檢驗各個假設是否成立。

見背面

第二題 (共 50 分):

某研究想探討穿戴式裝置是否在受訓期間能輔助跆拳道選手提升自身技巧與降低運動傷害，招募了 800 位跆拳道選手參與為期半年的實驗，選手們被隨機分派為兩組，兩組人數相等，對照組選手不使用穿戴式裝置，維持傳統的訓練模式；實驗組選手於訓練時皆會配戴智慧手環，手環會自動記錄選手執行動作時的軌跡與狀態，用以判斷動作的完成度、正確度、強度等，選手與教練於實驗期間會依據手環提供的數據調整動作、訓練內容與策略，在實驗開始前、實驗進行中的每個月底、實驗結束後，都會請專業人員評估選手的身體狀況與技能表現，以了解選手技能進步與運動傷害的情況。請根據此研究回答下列問題：

1. (10 分) 此手環的訊號紀錄會因選手流汗量而受到影響，偶有感應不良的狀況發生。假設在選手流汗量較少的情況下，手環平均一個月會發生 3 次紀錄缺失，在選手流汗量較多的情況下，手環平均一個月會發生 10 次紀錄缺失。如今一位實驗組的跆拳道選手，在一般訓練下他大爆汗的機率為 0.6，不會大爆汗的機率為 0.4。請問，此選手在半年的實驗期間，手環紀錄資料發生缺失的機率模型為何？期望值與標準差又各為多少？
2. (10 分) 對照組的選手有 192 位發生運動傷害，而實驗組的選手僅 144 位發生運動傷害，請利用 95% 的信賴區間(confidence interval)估計兩組選手發生運動傷害現象的差異，並解釋此區間的意義。
3. (10 分) 選手在使用智慧手環輔助訓練的實驗期間，手環紀錄下的動作表現的平均分數為 86.25 分，標準差為 8.2 分。除手環紀錄外，同時請教練於訓練期間評估選手動作表現，教練給出的平均分數為 80.64 分，標準差為 15.3 分。若看每對教練與選手的配對組合，動作表現分數的差異的平均值為 5.61 分，標準差為 22.44 分。請問手環紀錄下的動作表現分數和教練給出的分數是否一致？(若利用假設檢定的方式回答，必須說明虛無與對立假設、顯著水準、檢定統計量的計算及結論)
4. (5 分) 研究者想利用線性迴歸分析(linear regression)探討穿戴式裝置的效用，迴歸模型如下：

$$Y = \beta_0 + \beta_1 \times \text{Group} + \beta_2 \times \text{Time} + \varepsilon$$

其中，Y 代表技能進步分數，Group 代表實驗組別 (0 為對照組，1 為實驗組組)，Time 代表每週平均的訓練時間(小時)。分析報表如下：

	df	SS	MS	F	p-value
Regression	2	4505.692	2252.846	9.336	9.82E-05
Residuals	797	192322.882	241.309		
Total	799	196828.575			

	$\hat{\beta}$	SE	t	P-value
intercept	53.525	2.281	23.467	5.44E-93
Group	4.123	1.100	3.750	0.0002
Time	0.285	0.144	1.977	0.0483

請問 ANOVA table 中的 F 統計量，與迴歸係數報表中的 t 統計量，兩個檢定的虛無假說(null hypothesis)各是什麼？

5. (5 分) 承上題，請解釋組別與每週平均訓練時間和技能進步分數間的相關性。
6. (10 分) 承上題，使用最小平方方法所得之迴歸係數估計值為 BLUE (best linear unbiased estimator)，請說明何謂 BLUE 並證明之。