

1. (共 50 分，每小題 10 分) 關於運動習慣與糖尿病的研究，某研究員蒐集了 2000 人的資料，以下列出前十位：

| 運動習慣 | 性別 | 年齡 (歲) | 甲基化 位點 1 | 甲基化 位點 2 | 空腹血糖值 (單 位：mg/dL) |
|------|----|-----------|-------------|-------------|----------------------|
| 有 | 男 | 59 | 0.491 | 0.786 | 92 |
| 有 | 男 | 34 | 0.556 | 0.762 | 87 |
| 無 | 女 | 45 | 0.533 | 0.776 | 93 |
| 有 | 女 | 64 | 0.585 | 0.842 | 103 |
| 無 | 女 | 53 | 0.641 | 0.836 | 84 |
| 無 | 男 | 66 | 0.327 | 0.808 | 103 |
| 無 | 男 | 30 | 0.486 | 0.772 | 88 |
| 無 | 男 | 45 | 0.641 | 0.784 | 93 |
| 有 | 男 | 32 | 0.601 | 0.738 | 93 |
| 無 | 女 | 69 | 0.557 | 0.805 | 87 |

- (1) (10 分) 已知空腹血糖值可能受到性別、年齡的影響，請提出一個可以檢定運動習慣是否會影響空腹血糖值的方法。顯著水準設為 0.05，需寫出虛無假說、對立假說、檢定統計量、虛無假說下該檢定統計量之分布及自由度、拒絕域。為能清楚寫出檢定統計量，請自行設定符號並定義之。
- (2) (10 分) 執行(1)小題的方法，資料需要滿足哪些條件？您如何檢定這些條件是否成立？為能清楚寫出各個假設，請自行設定符號並定義之。
- (3) (10 分) 已知甲基化位點 1 分布範圍在 0 到 1 之間，其可能受到性別、年齡的影響，請提出一個可以檢定運動習慣是否會影響甲基化位點 1 的方法。顯著水準設為 0.05，需寫出虛無假說、對立假說、檢定統計量、虛無假說下該檢定統計量之分布及自由度、拒絕域。為能清楚寫出檢定統計量，請自行設定符號並定義之。
- (4) (10 分) 已知空腹血糖值、甲基化位點 1、甲基化位點 2 都可能受到性別、年齡的影響，請提出一個可以檢定運動習慣是否藉由甲基化位點 1、甲基化位點 2 來影響空腹血糖值的方法。顯著水準設為 0.05，需寫出虛無假說、對立假說、檢定統計量、虛無假說下該檢定統計量之分布及自由度、拒絕域。為能清楚寫出檢定統計量，請自行設定符號並定義之。
- (5) (10 分) 執行(4)小題的方法，請說明如何在考慮性別、年齡的情形下，將運動習慣對空腹血糖值的總效應區分為「直接效應」(運動習慣 \rightarrow 空腹血糖值) 與「間接效應」(運動習慣 \rightarrow 甲基化位點 1 \rightarrow 空腹血糖值，運動習慣 \rightarrow 甲基化位點 2 \rightarrow 空腹血糖值)，以及如何檢定「直接效應」與「間接效應」是否存在。為能清楚寫出檢定統計量，請自行設定符號並定義之。
2. (10 分) 某藥廠開發出一種新的心血管疾病藥物 P，可以降低粥狀動脈硬化及粥狀動脈栓塞，且宣稱其效果比現有的心血管疾病藥物 A 為佳。某醫師找來 100 對高血壓的同卵雙胞胎，每對雙胞胎中有一人服用 A 藥、另一人服用 P 藥，一年後，100 個服用 A 藥的人之中有 8 名發生粥狀動脈硬化及粥狀動脈栓塞，另外 92 人則沒有成功；其餘 100 名服用 P 藥的人當中有 16 人發生粥狀動脈硬化及粥狀動脈栓塞，剩下 84 人沒有。或者說 100 對雙胞胎中，有 80 對雙胞胎兩人都沒有發生粥狀動脈硬化及粥狀動脈栓塞；有 12 對雙胞胎中服 A 藥的沒有發生、但服 P 藥的有發生；有 4 對雙胞胎中服 A 藥的有發生、但服 P 藥的沒有發生；剩下的 4 對雙胞胎則是 A、B 藥都有發生粥狀動脈硬化及粥狀動脈栓塞。請檢定這二種藥物是否同樣有效？

見背面

3. (共 20 分) 某長期照顧研究學者想了解喘息服務的使用是否可以減輕家庭照顧者的照顧負荷與壓力，他設計了一份量表測量照顧負荷與壓力的程度，並進行了一個前瞻性的研究(prospective study)。在此研究中，研究者搜集了 1000 位家庭照顧者的資料，了解這些研究個案過去一年間的照顧負荷與壓力的狀況，並在研究期間每兩個月持續關心家庭照顧者，並提供相關長照資源使用的建議，同時了解這些家庭照顧者使用情況，在研究滿一年後，重新測量照顧負荷與壓力的程度，並進行相關的分析。請回答下列問題：

(1) (7 分) 研究者想知道經濟狀況是否會影響照顧負荷與壓力的程度 (0 - 100 分，分數越高表示負荷越重、壓力越大)，所搜集到的資料如下：

| | 經濟無虞 | 堪堪打平 | 經濟拮据 |
|--------------------------|--------|--------|--------|
| 照顧負荷與壓力的平均分數 \bar{X}_k | 41.03 | 41.86 | 44.07 |
| 樣本標準差 S_k | 15.82 | 15.71 | 15.94 |
| 樣本變異數 S_k^2 | 250.36 | 246.81 | 254.04 |
| 樣本數 N | 429 | 322 | 249 |

請完成以下變異數分析(Analysis of Variance, ANOVA)的表格，並說明您的結論 (定顯著水準 $\alpha = 0.1$)

| | Sum of Squares | d.f. | MSE | F | P-value |
|---------|----------------|------|-----|---|---------|
| Between | 1466 | ② | ③ | ⑤ | 0.0538 |
| Within | ① | 997 | ④ | | |
| total | 250846 | 999 | | | |

(2) (3 分) 請問前一小題所使用的統計方法是否有任何的假設？若有，假設為何？

(3) (10 分) 請問變異數分析是否可以用來檢定兩獨立樣本的母體平均值是否相等？若可以，請說明與哪一種兩獨立樣本的檢定結果會相等，並證明兩者檢定統計量(test statistic)之關係；若不行，請說明理由，並指出應該進行哪一種兩獨立樣本的檢定，並證明兩者的檢定統計量沒有數學上的關係。

4. (共 20 分) 某研究想要探討性別對幼兒專注力的影響。現隨機抽樣男女幼童各 50 人，資料如下表：

| | 女童 | 男童 | 合計 |
|-------|----|----|-----|
| 專注力高 | 25 | 15 | 40 |
| 專注力低 | 25 | 35 | 60 |
| total | 50 | 50 | 100 |

(1) (4 分) 請問女童相較於男童專注力高的勝算比(odds ratio, OR)為何？並解釋其意義。

(2) (6 分) 請估計「高專注力女童的比例」比「高專注力男童的比例」高出多少？提出這個差異的 95%信賴區間(confidence interval)，並解釋其意義。

(3) (5 分) 請證明樣本比例差 $\hat{p}_G - \hat{p}_B$ 、以及隨機抽取一位女童和一位男童的觀察值差 $X_i^{(G)} - X_j^{(B)}$ 都是不偏估計量(unbiased estimate)。(註： \hat{p}_G 代表樣本中高專注力女童的比例， \hat{p}_B 代表樣本中高專注力男童的比例； $X_i^{(G)}$ 代表女童的觀察值， $X_j^{(B)}$ 代表男童的觀察值)

(4) (5 分) 請說明上一小題的兩個估計值中，哪一個估計值是個比較好的估計值？為什麼？

試題隨卷繳回