

※ 注意：請於試卷內之「非選擇題作答區」標明題號依序作答。

是非題 (每題 2 分)

1. 若比較臺大校內男女的血型分布是否相等，可以使用變異數分析(Analysis of variance, ANOVA)。
2. 小吳和小王針對同一個研究議題分別進行收集樣本及統計檢定，小吳得到 p 值為 0.00001，小王得到 p 值為 0.001，所以在顯著水準 0.05 的情況下，小吳的檢定結果比小王更顯著。
3. 兩組獨立樣本的 t 檢定，不管變異數是否相等，都可以利用變異數分析的 F 檢定得到相同的結論。
4. 兩個二元變項的隨機變數之間相關性的強度，可以利用卡方檢定(Chi-square test)來計算。
5. 信賴區間不是隨機的區間，不能視為是機率。
6. 型二誤差率(type II error rate)與檢力(power)的總和為 1。
7. 在假說檢定中，樣本數的大小會影響 p 值，樣本數越大，p 值就越大。
8. 檢定勝算比(odds ratio)是否等於 1，跟檢查其 95%信賴區間是否包含 1，會有非常相似的結論。
9. 型一誤差率(type I error rate)可以視為是拒絕虛無假說的機率。
10. 卡方檢定的結果也可以利用兩組獨立樣本比例的 Z 檢定(independent two proportion Z test)來完成。

單選題 (每題 3 分)

11. 某研究想要利用變異數分析(Analysis of variance, ANOVA)來探討收集的 K 組樣本，其 K 組母體的期望值是否相等，請問以下何者錯誤?
 - (A) 這 K 組的母體必須都服從常態分布。
 - (B) 這 K 組母體的變異數必須都相等。
 - (C) 這 K 組樣本，每組的樣本數必須都相等。
 - (D) 這 K 組樣本必須彼此都獨立。
12. 某高血壓藥物的配對病例對照研究(matched case-control study)給予病例組及對照組兩個不同治療藥物，並測量每個人服用藥物一個月的前後血壓差異值，請問以下何者為真?
 - (A) 因為已經測量前後血壓差異值，所以可使用兩獨立樣本 t 檢定(independent two-sample t test)。
 - (B) 不同治療組的受試者之間未必獨立，但在同一治療組內的不同受試者彼此之間必須獨立。
 - (C) 不同治療組的受試者人數一定會相等。
 - (D) 如果每位受試者服藥前後的血壓值都有紀錄，就可以使用變異數分析(Analysis of variance, ANOVA)檢定。
13. 陳博士想比較甲乙兩個國家在 2023 年的流感侵襲率，調查甲國 100 人中有 9 人得到流感，乙國 121 人中有 16 人得到流感，請問若甲乙兩國的流感侵襲率沒有差異，則兩國的流感侵襲率的估計值最接近以下何者?
 - (A) 0.09。
 - (B) 0.10。
 - (C) 0.11。
 - (D) 0.12。
14. (延續上題)請問上述估計值的變異數(即標準誤的平方， SE^2)最接近以下何者?
 - (A) 0.002。
 - (B) 0.02。
 - (C) 0.2。
 - (D) 0.1。
15. (延續上題)若甲乙兩國的流感侵襲率有差異，則此差異的 95%信賴區間可以表示為 $0.0422 \pm 1.96 \times S$ ，請問其中的 S 的平方(S^2)最接近以下何者?
 - (A) $(25/221) \times (221-25)/221$ 。

見背面

- (B) $\{(25/221) \times (221-25)/221\}/221$ 。
 (C) $0.09 \times 0.91/100 - 0.16 \times 0.84/121$ 。
 (D) $0.09 \times 0.91/100 + 0.16 \times 0.84/121$ 。

單選題 (每題 5 分)

16. 將一組共 1001 人的身高資料，每個人先加高 9 公分再乘以 4。請問下述何者錯誤？
 (A) 平均值(mean)為原先的 4 倍以上。
 (B) 標準差(standard deviation)恰好為原先的 4 倍。
 (C) 中位數(median)恰好為原先的 4 倍。
 (D) 四分位差(interquartile range)恰好為原先的 4 倍。
 (E) 變異係數(coefficient of variation)變小。
17. 有關常態分佈(normal distribution)，下述何者錯誤？
 (A) 常態分佈乘 1.96，仍為常態分佈。
 (B) 常態分佈減 1.96，仍為常態分佈。
 (C) 兩個獨立的(independent)常態分佈相加，仍為常態分佈。
 (D) 兩個獨立且相同的(independent and identical)常態分佈相減，仍為常態分佈。
 (E) 兩個獨立的標準常態分佈(standard normal distribution)的均方(mean square；平方後加總再除以 2)，為自由度(degree of freedom)為 1 的卡方分佈(chi-squared distribution)。
18. 有關布阿松分佈(Poisson distribution)，下述何者錯誤？
 (A) 布阿松分佈乘正整數，仍為布阿松分佈。
 (B) 兩個獨立的布阿松分佈相加，仍為布阿松分佈。
 (C) 布阿松分佈的平均值與變異數(variance)相等。
 (D) 布阿松分佈為右偏(positively skewed)分佈。
 (E) 平均值趨近無限大時，布阿松分佈趨近常態分佈。
19. 有關伯努立分佈(Bernoulli distribution)，下述何者錯誤？
 (A) 兩個獨立且相同的伯努立分佈相加，為二項式分佈(binomial distribution)。
 (B) 無限多個彼此獨立(mutually independent)且相同的平均值無限小的伯努立分佈的總和，趨近布阿松分佈。
 (C) 無限多個彼此獨立且相同的伯努立分佈的平均，趨近常態分佈。
 (D) 伯努立分佈的變異係數介於 0 和 1 間。
 (E) 兩個不獨立且不相同的伯努立分佈相乘，仍為伯努立分佈。
20. 有關抽樣分佈(sampling distribution)，下述何者錯誤？
 (A) 進行點估計(point estimation)時，估計量(estimator)的機率分佈(probability distribution)是一種抽樣分佈。
 (B) 進行假說檢定(hypothesis testing)時，檢定量(test statistic)的機率分佈是一種抽樣分佈。
 (C) 抽樣分佈的變異係數不隨樣本數(sample size)改變。
 (D) 抽樣分佈可以為離散型(discrete)機率分佈。
 (E) 抽樣分佈可以為非對稱型(non-symmetric)機率分佈。
21. 有關系統性抽樣(systematic sampling)，下述何者錯誤？
 (A) 當母體(population)的成員有現成編號時，系統性抽樣可以很便利的進行。

- (B) 系統性抽樣不是機率抽樣(probability sampling)，不需仰賴亂數(random number)進行抽樣。
- (C) 系統性抽樣無法求得變異數，通常以簡單隨機抽樣(simple random sampling)的變異數替代。
- (D) 即使母體中各成員的數值隨著編號呈現週期性變化，系統性抽樣依然能夠提供對母體平均值的不偏估計(unbiased estimation)。
- (E) 當母體中各成員的數值隨著編號呈現週期性變化時，通過系統性抽樣獲得的樣本在代表性(representativeness)方面可能存在不足。

問答題 (每題 5 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 (0-100 分，分數越高表示負荷越重、壓力越大)，並進行了一個回溯性的研究調查，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。現在研究者透過迴歸分析(regression analysis)來回答研究問題，結果如下表：

迴歸統計

R^2	0.5932
調整的 R^2	0.5911
標準誤	9.192
觀察值個數	1000

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	5	122454.7	24490.9	289.885	<0.0001
殘差	994	83978.2	84.5		
總和	999	206432.9			

Regression Model

變項	係數	標準誤	t 統計量	P-值	95% CI for β
截距	11.65	3.11	3.741	0.0002	(5.54, 17.76)
喘息服務使用時數 (小時)	-1.08	0.07	-16.50	<0.0001	(-1.21, -0.95)
男 vs. 女	1.13	3.49	0.323	<0.0001	(-5.72, 7.98)
照顧者年齡 (歲)	0.98	0.05	20.310	0.7466	(0.89, 1.08)
被照顧者失能程度	0.46	0.02	22.628	<0.0001	(0.42, 0.49)
喘息服務使用時數 (小時) × 性別 (男 vs. 女)	-0.27	0.12	-2.375	0.0177	(-0.50, -0.05)

註：「被照顧者失能程度」為 0-100 分，分數越高表示失能程度越嚴重。

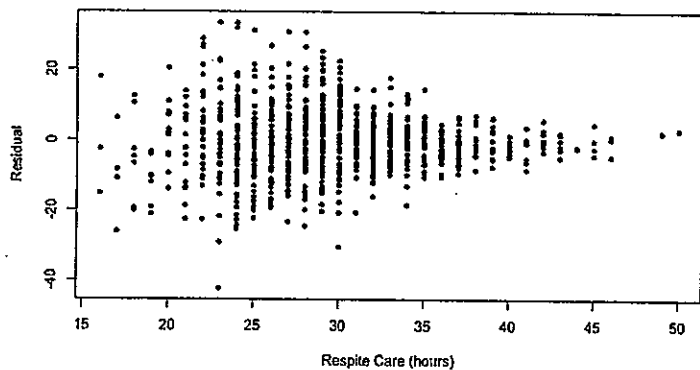
22. 請問，在使用迴歸分析時，下列敘述何者有誤？ (複選題)

- (A) 線性迴歸模型(linear regression model)中迴歸係數，可以利用最小平方法(least squared estimates, LSE)進行估計。
- (B) 模型需假設解釋變數(explanatory variable)服從常態分配(normal distribution)，意即 $X \sim N(\mu_X, \sigma^2)$ 。
- (C) 判定係數(coefficient of determination) R^2 是用來測量模型的解釋能力，越接近 1 代表解釋能力越佳。
- (D) 只要建立好迴歸模型，就可以說明解釋變數(explanatory variable)與反應變數(response variable)之間有因果關係。

見背面

(E) 線性迴歸分析(linear regression analysis)沒有樣本數與變數個數的限制，所以即便變數個數多於樣本數，也可以使用一般的線性迴歸模型。

23. 根據研究者欲探討的問題以及分析的結果，請問，下列敘述何者有誤？（複選題）
- (A) 模型中性別前的迴歸係數，解釋意義為男性相較於女性，照顧負荷與壓力程度的分數會多 1.13 分。
 - (B) 如果想探討不同性別(X)的照顧者對申請使用喘息服務(Y)的行為是否有差異 (Y為是否有申請)，在同時考量其他變數（如：年齡、社經條件、社會支持等）之下，也可以使用線性迴歸(linear regression)進行分析。
 - (C) 「喘息服務使用時數」的迴歸係數為負，代表此變項與「照顧負荷與壓力的程度」之間是負相關。
 - (D) 此模型的 $R^2 = 0.5932$ ，代表此迴歸模型只能解釋「照顧負荷與壓力的程度」59.32%的變異，此模型的配適(goodness-of-fit)程度差強人意，可能有重要的變數或交互作用項未納入，若能再進一步調整模型中放入的變數，有機會讓模型更好。
 - (E) 研究者若把照顧者年齡分組 (age_group: 0=40 歲以下、1=41-65 歲、2=65 歲以上)，想考慮年齡與性別 (sex: 0=女、1=男) 的交互作用項，在跑統計軟體時，可以直接將兩個變項相乘(age_group × sex)後放入模型，不需重新設定 dummy variable。
24. 請說明 ANOVA 表中 F 檢定的判讀方式與結論（包含虛無與對立假設、顯著水準、如何完成檢定及結論）。
25. 請說明「喘息服務使用的時數」的迴歸係數統計檢定的判讀方式與結論（包含虛無與對立假設、顯著水準、如何完成檢定及結論）。
26. 請用迴歸分析的結果解釋交互作用項（喘息服務使用的時數×性別）代表的意義。
27. 請解釋「喘息服務使用的時數」之迴歸係數的 95%信賴區間的意義。
28. 下面是迴歸分析的殘差圖（縱軸代表殘差項，橫軸代表解釋變數喘息服務使用的時數），請問此迴歸分析模型是否有存在任何的問題？



試題隨卷繳回