

※ 考生請注意：本試題不可使用計算機。請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

1. 大數據時代已經來臨，大數據的各項技術能快速擷取、處理和分析大量資料，成為可解讀有用的資訊，以供理解現象、預測趨勢和決策之參考。判斷以下大數據推論的敘述為正確或錯誤，答案必須說明判斷的理由和依據。
 - (1) 若隨機樣本個數夠多時，根據弱大數法則 (Weak Law of Large Number)，樣本平均數會等於母體平均數 (5%)
 - (2) 若隨機樣本個數夠多時，依據中央極限定理 (Central Limit Theorem) 保證觀察值會服從常態分配 (5%)
2. 某大品牌的某 I 型號手機市占率極高，2019 上半年累計銷量達 26900000 支。其所使用之電池是由兩家工廠供應，高知名度 A 工廠供應 80% 電池，低知名度 B 工廠為 20%。消費者常會關心如果買到不良品會是由低知名度 B 工廠做出來的機率。已知 A 工廠供應的電池不良率為 2%，低知名度 B 工廠供應的電池不良率則有 5%。
 - (1) 今有一顧客買此 I 型號手機產品因電池不良而退貨，而此產品是由低知名度 B 工廠供應的機率為何? (5%)
 - (2) 假設此品牌忠實的顧客接受了退貨，並換回一支相同型號新手機，該顧客又遭遇因電池不良而退貨的機率為何? (5%)
3. 某種特殊汽車事故保險過去 30 年每年平均發生理賠之件數很少，只有 4 件。去年只發生 3 件。
 - (1) 明年理賠件數為 0 的機率為何? 請敘述理賠件數分布假設的基礎及其合理性。(5%)
 - (2) 去年理賠件數為 0 的機率為何? (5%)
 - (3) 明年兩相繼理賠事件相隔時間超過半年的機率為何? (5%)
4. 一家雜誌社欲知其讀者的平均年齡，以作為雜誌內容走向的參考，根據其對訂閱戶抽查所得的讀者平均年齡為 36 歲。
 - (1) 當隨機樣本數為 25，假設母體分布為常態，樣本標準差為 6 歲，求該雜誌讀者平均年齡的 90% 信賴區間。(5%)
 - (2) 當隨機樣本數為 100，假設母體分布未知，但知母體標準差為 6 歲，求該雜誌讀者平均年齡的 95% 信賴區間。(5%)
 - (3) 為何(1)和(2)的計算信賴區間的分配不同，理由為何? (5%)
 - (4) 假設訂閱戶數量為 100000，請建議適當的抽樣方法和樣本數量給雜誌社。(5%)
5. 職業安全預防輔導及諮詢服務是勞動部職業安全衛生署的工作項目之一。對於特定的中小企業要給予職業安全預防輔導。在某家中型工廠的抽查中，由三年內受過燙傷和割傷的員工中隨機抽出

100 名。結果抽到燙傷過的員工 60 名，割傷過的員工 40 名，並沒有員工三年內同時受過燙傷和割傷。年齡的分布小於 25 歲者有 70 名，其中 46 名燙傷過，24 名割傷過。年齡的分布大於等於 25 歲有 30 名，其中 14 名燙傷過，16 名割傷過。職業安全衛生署想知道此中型工廠的員工受傷型態和年齡是否有關，然後再決定是否依照 25 歲分組進行職業安全教育。職業安全衛生署欲使用統計檢定回答此問題。

- (1) 請列出適當的虛無假設和對立假設。(5%)
- (2) 列出適當的檢定統計量計算式 (其數值接近 3.175) (5%)
- (3) 假設在型一錯誤為 0.05 時，此檢定統計量分配的臨界值為 3.84，在型一錯誤為 0.1 時，此檢定統計量分配的臨界值為 2.76。請敘述檢定結果。(5%)
- (4) 請定義此檢定的型二錯誤，並簡述計算的方法。(5%)

6. 以下資料表為 Kingstone 公司在台灣售出的某類型每一季的隨身碟數量 (Sales 單位為 1000 個)

year	第一季	第二季	第三季	第四季
2016	260	398	252	260
2017	256	429	376	256
2018	228	496	425	228
2019	248	576	426	248

該公司想了解各季之間的銷售差異，並預估 2020 的銷售數量。

該公司的資料分析師建立了簡單線性模型

$$\text{Sales} = \beta_0 + \beta_1 T + \beta_2 Q_1 + \beta_3 Q_2 + \beta_3 Q_3 + \varepsilon$$

T=1, 2, 3, ..., 16. T=1 當 Sales 是 2016 年第一季，到 T=16 當 Sales 是在 2019 年第四季;
 Q1=1 當 T 是第一季；Q1=0 當 T 不是第一季;
 Q2=1 當 T 是第二季；Q2=0 當 T 不是第二季;
 Q3=1 當 T 是第三季；Q3=0 當 T 不是第三季。

使用以下所附 SAS 的輸出結果回答問題

- (1) 請檢定第一季和第四季的平均銷售差異是否顯著 (5%)
- (2) 請用此線性回歸模型的輸出報表估計第二季和第一季的平均銷售差異 (5%)
- (3) 請列出 2020 年第三季銷售預測值與 95% 預測區間 (5%)
- (4) 以下所附輸出報表能否足夠協助判定隨機誤差項的一般假設是否成立？為何？(5%)
- (5) 隨機誤差項的一般假設的成立是否如何影響(1)、(2)和(3)的推論？(5%)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	157126	39282	16.37	0.0001
Error	11	26399	2399.9		
Corrected Total	15	183526			

Root MSE	48.98915	R-Square	0.8562
Dependent Mean	335.12500	Adj R-Sq	0.8038
Coeff Var	14.61817		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	182.75000	36.74186	4.97	0.0004
T	1	6.52500	2.73858	2.38	0.0363
Q1	1	19.57500	35.60149	0.55	0.5934
Q2	1	239.80000	35.07089	6.84	<.0001
Q3	1	128.27500	34.74864	3.69	0.0036

Output Statistics						
Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Predict	Residual	
1	260.0000	208.8500	29.4954	82.9907 334.7093	51.1500	
2	398.0000	435.6000	29.4954	309.7407 561.4593	-37.6000	
3	252.0000	330.6000	29.4954	204.7407 456.4593	-78.6000	
4	260.0000	208.8500	29.4954	82.9907 334.7093	51.1500	
5	256.0000	234.9500	25.0995	113.7974 356.1026	21.0500	
6	429.0000	461.7000	25.0995	340.5474 582.8526	-32.7000	
7	376.0000	356.7000	25.0995	235.5474 477.8526	19.3000	
8	256.0000	234.9500	25.0995	113.7974 356.1026	21.0500	
9	228.0000	261.0500	25.0995	139.8974 382.2026	-33.0500	
10	496.0000	487.8000	25.0995	366.6474 608.9526	8.2000	
11	425.0000	382.8000	25.0995	261.6474 503.9526	42.2000	
12	228.0000	261.0500	25.0995	139.8974 382.2026	-33.0500	
13	248.0000	287.1500	29.4954	161.2907 413.0093	-39.1500	
14	576.0000	513.9000	29.4954	388.0407 639.7593	62.1000	
15	426.0000	408.9000	29.4954	283.0407 534.7593	17.1000	
16	248.0000	287.1500	29.4954	161.2907 413.0093	-39.1500	
17	.	313.2500	36.7419	178.4695 448.0305	.	
18	.	540.0000	36.7419	405.2195 674.7805	.	
19	.	435.0000	36.7419	300.2195 569.7805	.	
20	.	313.2500	36.7419	178.4695 448.0305	.	