

題號： 288
科目： 基礎統計學
節次： 6

國立臺灣大學 114 學年度碩士班招生考試試題

題號：288

共 3 頁之第 1 頁

1. (50 分，共 5 題)

※ 注意：請於試卷內之「非選擇題作答區」作答，並應註明作答之題號。

因連日低溫，寒流與院前心肺功能停止(Out-of-Hospital Cardiac Arrest, OHCA)相關的新聞頻繁出現。某研究者想驗證寒流發生與 OHCA 的發病率之間是否有關係，因此收集了臺灣地區 10 年的資料，請依據下面的敘述回答問題：

(1-1) (15 分) 根據蒐集的資料最近十波寒流持續天數的資料如下：

3, 4, 5, 7, 1, 5, 2, 3, 2, 2

請問寒流持續天數的機率模型為何？同時請導出此隨機變數的 moment-generating function，並利用 method of moment 估計此隨機變數的期望值與變異數。

(1-2) (5 分) 延續上題，請問 first moment 得到的估計量，是否為此機率模型中參數的充分統計量(sufficient statistic)？請證明之。

(1-3) (5 分) 在資料收集時有將臺灣劃分為四個地理區域（北部、中部、南部、東部），請問若要檢定這十年來每次寒流發生時四個地理區域的 OHCA 發生數是否有差異，可以如何做？請說明統計檢定完整的流程，並說明要如何下結論。

(1-4) (5 分) 延續上題，使用此檢定方法是否有任何假設？若有，請說明。

(1-5) (20 分，共 7 小題)

研究者利用迴歸分析探討寒流持續天數用 OHCA 發病數的關係，結果如下：

Regression model	迴歸係數 $\hat{\beta}$	標準誤 $se(\hat{\beta})$	t statistic	p-value	95% CI
截距	-13.48	4.40	-3.06	0.0037	(-22.35, -4.61)
寒流持續天數	7.54	3.54	2.13	0.0390	(0.40, 14.86)
寒流持續天數 x 東部 (參考組)	--	--	--	--	--
寒流持續天數 x 南部	4.47	3.56	1.26	0.2158	(-2.70, 11.64)
寒流持續天數 x 中部	2.73	3.39	0.81	0.4249	(-4.10, 9.56)
寒流持續天數 x 北部	7.22	3.35	2.16	0.0364	(0.48, 13.97)

ANOVA table

	自由度	SS	MS	F	p-value
迴歸	4	19416.41	4854.10	21.99	4.1264E-10
殘差	45	9934.97	220.78		
總和	49	29351.38			

(a) (2 分) 請問此迴歸模型為何？

(b) (8 分) 請問此模型的迴歸係數該如何估計？請寫出推導過程。

(c) (2 分) 請問 ANOVA table 中 F 統計檢定的虛無假說(null hypothesis)與對立假說(alternative hypothesis)是什麼？

(d) (2 分) 請解釋截距項迴歸係數的意義。

(e) (2 分) 請解釋寒流持續天數之迴歸係數的 95%信賴區間的意義。

(f) (2 分) 此模型中包含地區與寒流持續天數的交互作用項，請說明北部地區的 OHCA 發病數與寒流持續天數的關係該如何解讀。

見背面

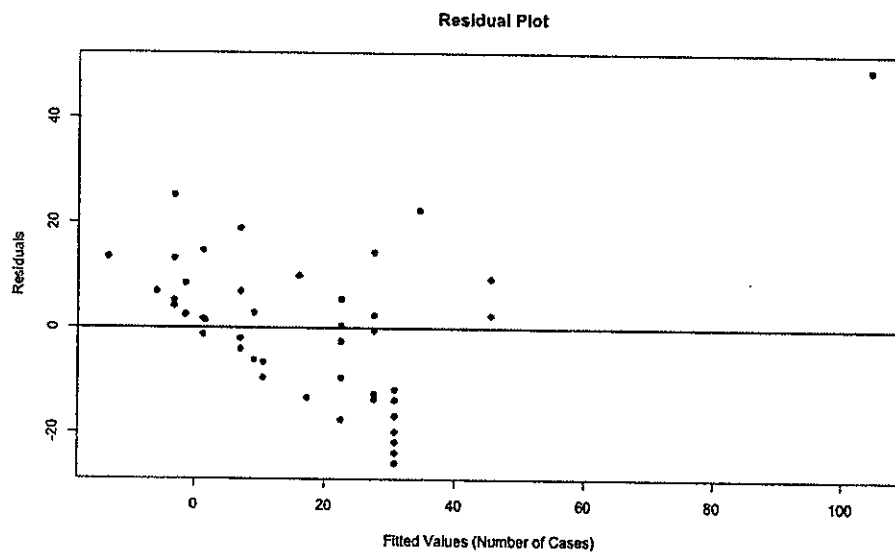
題號： 288
 科目： 基礎統計學
 節次： 6

國立臺灣大學 114 學年度碩士班招生考試試題

題號：288
 共 3 頁之第 2 頁

(續 (1-5))

(g) (2 分) 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 OHCA 發病數), 請問此迴歸分析模型是否有存在任何的問題?



2. (50 分, 共 5 題)

欲研究某老化基因上的「單核苷酸多態性」(Single-Nucleotide Polymorphism, 簡稱 SNP) 基因型是否影響民眾之「生理年齡」(biological age), 於是自全台隨機抽取若干成年民眾, 檢測其基因型, 並以甲基化資料來推估其生理年齡, 資料如下表:

生理年齡 (單位: 年)	SNP 基因型	實足年齡 (單位: 年)
60	AA	63
65	AG	69
72	GG	65
58	AA	63
40	AG	43
32	GG	26
58	AA	64
36	AG	39
46	GG	42
68	AA	73
70	AG	71
66	GG	63
23	AA	30
45	AG	46
78	GG	68
71	AA	76
66	AG	68
52	GG	50
30	AG	26
43	AA	36

接次頁

題號： 288
科目： 基礎統計學
節次： 6

國立臺灣大學 114 學年度碩士班招生考試試題

題號：288
共 3 頁之第 3 頁

(續 2.)

- (2-1) (10 分) 首先，為求得個體老化加速的程度，將「生理年齡」(biological age, BA) 對「實足年齡」(chronological age, CA) 作一般線性迴歸，模式為 $BA = \beta_0 + \beta_1 CA + \varepsilon$ ，其中誤差項 ε 服從平均數 0 變異數 σ^2 的常態分布，請推導出 σ^2 的最大概似估計量 (maximum likelihood estimate)。
- (2-2) (10 分) 承上一小題， σ^2 的最大概似估計量 (maximum likelihood estimate) 是否為 σ^2 的不偏估計量？如果是，請證明之。如果不是，請推導證明出 σ^2 的不偏估計量 (unbiased estimate)。
- (2-3) (10 分) 模式 $BA = \beta_0 + \beta_1 CA + \varepsilon$ 之殘差 (e) 稱為「老化加速程度」，將 e 對 SNP 基因型作一般線性迴歸，由於基因型有三種，所以需要設置兩個虛擬變項 (dummy variables)，請寫出迴歸模式及虛擬變項 (dummy variables) 之定義，並推導出三個迴歸係數 (截距項與兩個虛擬變項之係數) 的最大概似估計量 (maximum likelihood estimate)。
- (2-4) (10 分) 承上一小題，三個迴歸係數的最大概似估計量 (maximum likelihood estimate) 是否為各係數的不偏估計量？如果是，請證明之。如果不是，請推導證明出三個迴歸係數的不偏估計量 (unbiased estimate)。
- (2-5) (10 分) 若「老化加速程度」(e) 不呈常態分布，請以無母數的方法來檢定該 SNP 基因型是否影響民眾之「老化加速程度」(e)。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。如有自設符號請定義。

試題隨卷繳回