

東吳大學 111 學年度碩士班研究生招生考試試題

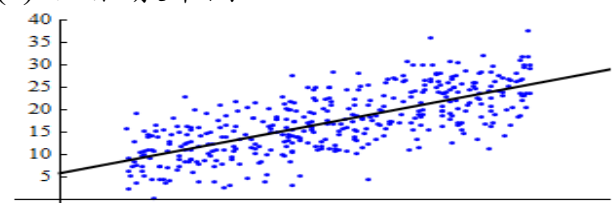
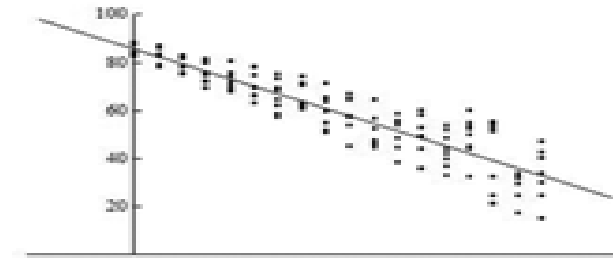
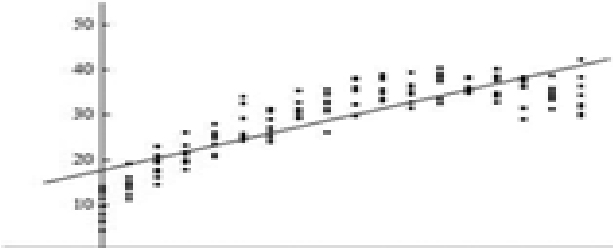
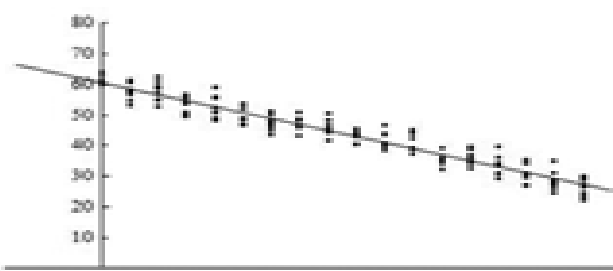
第1頁，共7頁

系級	資料科學系碩士班	考試時間	100 分鐘
科目	統計學	本科總分	100 分

※一律作答於答案卷上(題上作答不予計分)；並務必標明題號，依序作答。

一、配合題：(共 20 小題，每小題 2 分，佔總分 40%)

1. 線性迴歸是統計學的基本模型之一，過往都會透過對預測變數之殘差圖來判斷線性迴歸函數是否適於呈現 X 變項與 Y 變項之間的關係模式，請先看過右側 **a 到 h** 的所有選項、根據左側的題目、挑選適當的右側答案，可複選與重複使用：(每小題 2 分)

題目	答案選項(請填小寫英文字母)
1.1 請問相對應於(a)描繪 X 變項與 Y 變項的點狀散布圖的殘差圖是哪一個 <u>(1)</u>	(a) 點狀散布圖 
1.2 呈上題，為什麼 <u>(2)</u>	
1.3 請問相對應於(b)描繪 X 變項與 Y 變項的點狀散布圖的殘差圖是哪一個 <u>(3)</u>	(b) 點狀散布圖 
1.4 呈上題，為什麼 <u>(4)</u>	
1.5 請問相對應於(c)描繪 X 變項與 Y 變項的點狀散布圖的殘差圖是哪一個 <u>(5)</u>	(c) 點狀散布圖 
1.6 呈上題，為什麼 <u>(6)</u>	
1.7 請問相對應於(d)描繪 X 變項與 Y 變項的點狀散布圖的殘差圖是哪一個 <u>(7)</u>	(d) 點狀散布圖 
1.8 呈上題，為什麼 <u>(8)</u>	
1.9 請問哪些殘差圖無法符合同質變異 (indetical)的假設 <u>(9)</u>	
1.10 請問哪些殘差圖可以表明，當我們在解釋變量的範圍內從左向右移動時，基於線性迴歸線的預測會導致更大的誤差 <u>(10)</u>	(e) 因為此點狀散布圖右側的點比左側的點更遠離回歸線，乃對應著殘差圖顯示了隨著從左到右移動而呈扇形散開的模式。

東吳大學 111 學年度碩士班研究生招生考試試題

系級	資料科學系碩士班	考試時間	100 分鐘
科目	統計學	本科總分	100 分

1.11 請問哪些殘差圖顯示著殘差隨機散落的樣態 (11)

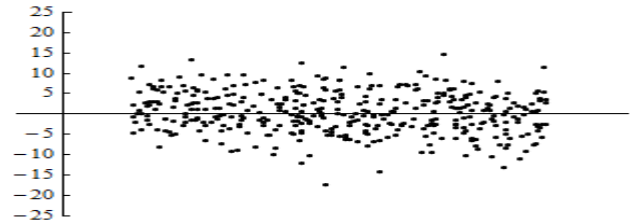
1.12 請問哪兩個點狀散布圖之迴歸線未能捕捉到變量之間的重要關係 (12)

(f) 因為此殘差圖顯示了從負到正再到負的曲線模式，這與該點狀散布圖一致。

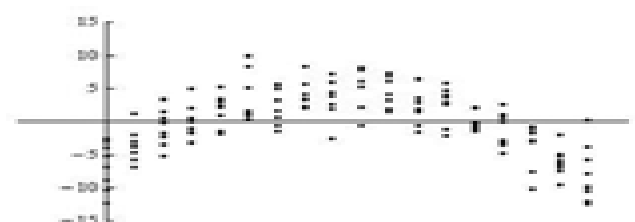
(g) 因為殘差圖曲線的，這與點狀散布圖類似。

(h) 因為殘差圖是矩形的，這與點狀散布圖一致，圖中線和點之間的距離在我們從左到右移動時保持相當穩定，最大距離約為 15 個單位

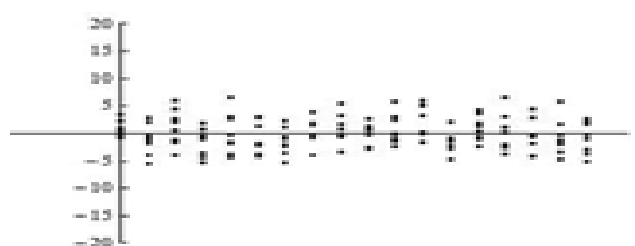
(e) 殘差圖



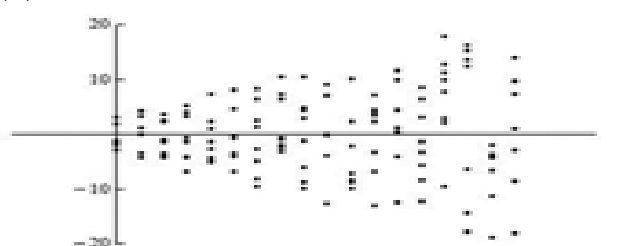
(f) 殘差圖



(g) 殘差圖



(h) 殘差圖



東吳大學 111 學年度碩士班研究生招生考試試題

第3頁，共7頁

系級	資料科學系碩士班	考試時間	100 分鐘
科目	統計學	本科總分	100 分

2. 統計學概念與名詞解釋，請根據下列左側的題目、挑選適當的右側答案、請用大寫英文字母、可複選、不能重複使用：**(每小題 2 分)**

題目	答案選項(請填大寫英文字母)
2.1 請問什麼是顯著水準 (significance level) <u>(13)</u>	A. 避免 type I error 變大，不易推翻假說
2.2 請問什麼是 p 值 (p-value) <u>(14)</u>	B. 此方法為簡單的成對比較、可處理 sample size 不同的組別
2.3 請問下列針對自由度(degree of freedom)的描述，何者為真 <u>(15)</u>	C. 在抽樣完，計算出樣本平均值後，大小為 n 的樣本中只要 n-1 個數確定了，第 n 個數就只有一個能使樣本符合樣本平均值的數值。也就是說，樣本中只有 n-1 個數可以自由變化，只要確定了這 n-1 個數，標準差也就確定了
2.4 當 ANOVA 分析結果顯著時，我們要進行事後檢定 (post-hoc test)進行多組之間的比較，此時會浮現什麼問題 <u>(16)</u>	D. 以樣本的統計量來估計母體的參數時，樣本中獨立或能自由變化的數據的個數
2.5 當發生上述的問題時，可能的解決方法為 <u>(17)</u>	E. 此一名詞可用來表示為一個隨機向量的維度數，也就是一個向量能被完整描述、所需要的最少單位向量數
2.6 請問上述解決方法的顯著水準該如何運算 <u>(18)</u>	F. 多重比較問題 (multiple comparisons problem)
2.7 請問 Student-Newman-Keuls 的優點為 <u>(19)</u>	G. $0.05 * C^k_2$
2.8 當各個實驗組內出現異質現象時，會將同質的小群體聚集成一個個的 Block，再進行 ANOVA 分析，此研究設計為哪一種設計 <u>(20)</u>	H. 有較大的 Power 檢測出差異
	I. 隨機集區設計(Randomized Block Design)
	J. $0.05 / C^k_2$
	K. Type I error 變大
	L. 邦費羅尼校正 (Bonferroni correction)
	M. $\Pr(\text{推翻 } H_0 \mid H_0 \text{ 為真})$
	O. 想要了解「五個類別的快樂程度」與「三個類別的肥胖程度」之間的關聯時，採用 5*3 的列聯表、卡方檢定，其自由度為 $(5-1)*(3-1) = 1$
	P. 在虛無假設為真的條件下，比樣本的觀察值更極端的機率總和 = $2 * \min\{\Pr(X \leq x \mid H_0), \Pr(X \geq x \mid H_0)\}$

東吳大學 111 學年度碩士班研究生招生考試試題

第4頁，共7頁

系級	資料科學系碩士班	考試時間	100 分鐘
科目	統計學	本科總分	100 分

二、台灣從 2021 年 3 月 22 日至 2022 年 1 月 19 日止，國內施打三劑疫苗的比率分別為 $P(\text{第一劑})=0.8091$ 、 $P(\text{第二劑})=0.7287$ 、 $P(\text{追加劑})=0.1181$ 。假設台灣只有施打 AZ、BNT、Moderna、高端四種疫苗，請根據疾病管制署所提供之數據，回答下列之問題：**(共 4 小題，每小題 4 分，佔總分 16%)**

	第 1 劑	第 2 劑	追加劑
AZ	0.4250	0.4139	0.0009
BNT	0.3236	0.3166	0.1585
Moderna	0.2094	0.2277	0.8125
高端	0.0420	0.0418	0.0281

現在，有一民眾曾接種過疫苗，

1. 某民眾曾接種 AZ 疫苗的機率為多少? (21)
2. 假設某民眾曾接種 AZ 疫苗，請問他是第二劑施打 AZ 疫苗的機率為何? (22)
3. 上述兩小題的機率是否不同? (23)
4. 其意涵為何? (24)

(請翻頁、完成最後一大題)

東吳大學 111 學年度碩士班研究生招生考試試題

系級	資料科學系碩士班	考試時間	100 分鐘
科目	統計學	本科總分	100 分

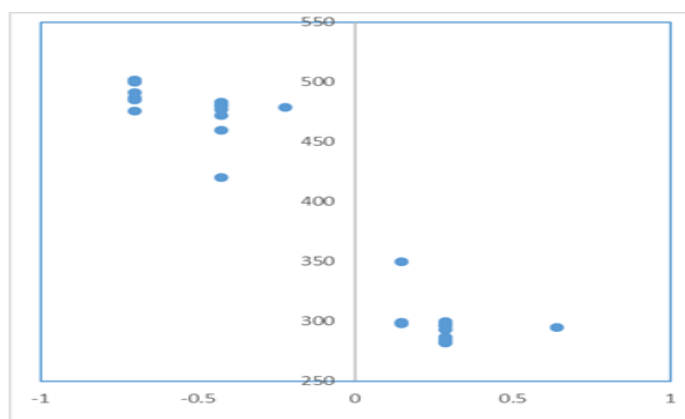
三、目前國內外的金融銀行為達到資金流通運用的最高效益、多在積極發展適當的「信用評分演算模型」，可迎合客戶的即時需求、加速貸款流程、進行風險管控、與及早偵測信用欺詐行為；如今，你身為資料分析師，將根據以下假設「過去發生的現象，未來亦會重複出現；例如，負債程度高者，其違約率亦較高」來進行模型的建置。請根據下列的數據與圖表來回答下列問題：**(共 22 小題，每小題 2 分，佔總分 44%)**

表一、數據資料表

序號	信用評分	手續費之 WOE	手續費	有違約行為	序號	信用評分	手續費之 WOE	手續費	有違約行為
1	282	0.2855	4758	1	14	460	0.146	15773	0
2	283	0.2855	6243	1	15	472	-0.4256	1135	0
3	285	0.2855	8571	1	16	476	-0.702	840	0
4	287	0.2855	6947	1	17	477	-0.4256	1894	0
5	293	-0.4256	1364	1	18	479	-0.2209	2842	0
6	295	-0.4256	1784	1	19	480	-0.4256	1456	0
7	296	0.2855	7128	1	20	481	-0.4256	1086	0
8	297	0.6406	3778	1	21	483	-0.4256	1444	0
9	298	0.2855	5569	1	22	485	-0.702	0	0
10	299	0.2855	8453	1	23	487	-0.702	374	0
11	300	0.146	16887	1	24	491	-0.702	446	0
12	350	0.146	24745	1	25	500	-0.702	184	0
13	420	0.2855	7845	1	26	502	-0.702	781	0

註：信用評分是整合多種指標所計算出來的分數，分數越高、表示信用越好；手續費是根據不同的貸款項目與額度、會有不同的費用額度；WOE 是將手續費切成六個級距後、計算每個級距會有違約機率 (probability of bad) 的勝算比(odds ratio)之自然對數值；違約行為表示曾經逾期還款。

圖一、點狀散布圖



表二、迴歸統計表

迴歸統計	
R 的倍數	0.95
R 平方	0.91
調整的 R 平方	0.90
標準誤	29.49
觀察值個數	26

東吳大學 111 學年度碩士班研究生招生考試試題

第6頁，共7頁

系級	資料科學系碩士班	考試時間	100 分鐘
科目	統計學	本科總分	100 分

1. 請問，圖一的 X 軸為哪一個變項名稱 (25) 、Y 軸為哪一個變項名稱 (26)
2. 請根據圖一的點狀散布圖的樣態來推測 Y 的數值資料是近似哪一種分布 (27)
3. 根據上題 Y 的分布型態(distribution)，您覺得哪一種迴歸分析較為適當 (28) ；
為什麼？ (29) ；當您想驗證 Y 變數並非常態分佈時，可以作什麼圖形來當作證據 (30) (列舉任何一個即可)
4. 在此，一位初學者採用了簡單線性迴歸進行分析，請您接續請完成下列迴歸分析的 ANOVA Table，填寫表三與表四(31)到(35)的數值：

表三、ANOVA Table

	DF	Sum of Square	Mean Square	F value	Pr>F
Model	1	199422.8	199422.8	<u> (34) </u>	8.83E-14
Error	<u> (31) </u>	<u> (32) </u>	<u> (33) </u>		
Corrected Total	25	220294.5			

表四、迴歸係數表

Parameter	Parameter Estimate	Standard Error	t value	Pr> t	下限 95%	上限 95%
Intercept	362.69	6.15	58.94	1.76E-27	349.99	375.39
WOE	-204.49	13.50	<u> (35) </u>	8.83E-14	-232.36	-176.62

- 5.請寫出檢定 β_1 是否為 0 的假設檢定、95%信賴區間、以及結論：

 (36)、(37)、(38)

東吳大學 111 學年度碩士班研究生招生考試試題

第7頁，共7頁

系級	資料科學系碩士班	考試時間	100 分鐘
科目	統計學	本科總分	100 分

6. 請依照表二估計 X 變項與 Y 變項之間的相關係數： (39) ，請論述此簡單線性迴歸模型可能有什麼問題： (40) ，針對此問題，應該還要查探哪個指標： (41)
7. 請說明線性迴歸模型(linear regression)的三大基本假設： (42)、(43)、(44)
8. 你會如何評估迴歸模型的好壞： (45) 請列舉一個
9. 當你日後進入資料科學的領域下，分析的樣本數目越大、P-value 愈容易小於 0.05，當你隨便放入許多特徵值去做迴歸分析、P-value 都小於 0.05 時，你該怎麼進一步處置？
 (46)