

I. 統計個案分析 (共 50 分，每題 5 分)

大木博士欲研究皮丘進化成皮卡丘的決定因子；隨機取得 20 隻皮丘和進化與否的資料，整理如下列表格。 $Y$  為一虛擬變數，若皮丘寶寶進化成皮卡丘，則  $Y=1$ ；反之  $Y=0$ 。 $X$  為皮丘和飼主過去一個月的訓練次數，假設  $X$  遵循常態分配。

編號	1	2	3	4	5	6	7	8	9	10
進化( $Y$ )	1	1	1	1	1	1	1	1	1	1
訓練次數( $X$ )	4	5	5	6	6	6	7	7	9	9
編號	11	12	13	14	15	16	17	18	19	20
進化( $Y$ )	0	0	0	0	0	0	0	0	0	0
訓練次數( $X$ )	1	2	3	3	3	4	4	4	6	6

請協助大木博士完成以下十個統計問題；如有需要，請檢附簡單的計算過程。

1. 請問  $Y$  的勝算率(odds ratio)是多少?
2. 若  $X$  的母體均數為  $\mu_X$ ，變異數為  $\sigma^2$ ；請完整寫下機率函數  $f(X)$ 。
3. 給定  $a$ 、 $b$  為常數，請問  $\text{var}(a+bX)$  為何?
4. 請畫出  $X$  變數的 Normal quintile plot(Q-Q plot)；需標記  $x$  軸和  $y$  軸數值。請解釋為何  $X$  看起來不像常態分配資料。(3+2 分)
5. 大木博士寫下虛無假設  $H_0: \sigma_X^2 = 1$ ，試以卡方檢定檢驗該假說。請問檢定統計量卡方值為何？虛無假說在 5% 顯著水準下是否成立？(提示： $\chi_{0.05, 19} = 30.144$ )
6. 大木博士寫下第二個虛無假設  $H_0: \sigma_{X|Y=0}^2 = \sigma_{X|Y=1}^2$ ，試以  $F$  檢定檢驗該假說。請問檢定統計量  $F$  值為何？虛無假說在 5% 顯著水準下是否成立？(提示： $F_{0.05, (9,9)} = 3.179$ )
7. 大木博士寫下第三個虛無假設  $H_0: \mu_{X|Y=0} = \mu_{X|Y=1}$ 。令  $\bar{X}_1$  為成功進化樣本的平均訓練次數， $\bar{X}_0$  為沒有進化樣本的平均訓練次數。請寫出檢定統計量  $\frac{\bar{X}_1 - \bar{X}_0}{SD(\bar{X}_1 - \bar{X}_0)}$  之實際分配(exact distribution)；需註記分配自由度的值。
8. 請以  $t$  檢定檢驗第三個假說。請問檢定統計量  $t$  值為何？虛無假說在 5% 顯著水準下是否成立?
9. 大木博士想了解  $t$  分配和標準常態分配  $Z$  之間的關係；請幫他寫出  $t$  分配如何由  $Z$  形成。也請你幫大木博士說明為何隨機樣本變異數未知時，檢定統計量  $\frac{\bar{X} - \mu_X}{\hat{S}_X}$  服從  $t$  分配。(3+2 分)
10. 如果大木博士想用統計軟體做單因子變異數分析(ANOVA)，檢定第三個虛無假說，請寫出對應的統計程式語法(寫出 SAS、Stata、R 任一統計軟體的語法即可，如所用語法為其他統計軟體亦請說明)。

II. 迴歸分析 (共 50 分，每題 5 分，本部分為選擇或填充題，作答時請直接寫出題號及答案，不需要附上計算過程。)

11. 假設公司執行長(CEO)的薪資( $wage$ ，單位為新台幣元)取對數後受到受教育的年數( $educ$ )，工作年資( $exper$ ，單位為年)，以及天生的能力( $abil$ )的影響：

$$\ln(wage)_i = \gamma_0 + \gamma_1 educ_i + \gamma_2 exper_i + \gamma_3 abil_i + u_i \quad (1)$$

其中  $\gamma_1$ ， $\gamma_2$ ，及  $\gamma_3$  皆為正值，且三個解釋變數與  $u_i$  皆無關。由於無法觀察到 CEO 的能力，小光使用古典最小平方法(ordinary least square, OLS)估計教育及年資對於 CEO 薪資的影響得到以下結果：

$$\ln(\widehat{wage})_i = \hat{\gamma}_0 + \hat{\gamma}_1 educ_i + \hat{\gamma}_2 exper_i \quad (2)$$

請問以下敘述何者正確？注意答案可能不只一個。

- (a) 若  $Cov(abil, educ) = Cov(abil, exper) = 0$ ，則  $\hat{\gamma}_1$  為  $\gamma_1$  的不偏估計式。
- (b) 若  $Cov(abil, educ) > 0$ ，但  $Cov(abil, exper) = 0$ ，則  $\hat{\gamma}_1$  將傾向高估教育對於薪資的真實影響( $\gamma_1$ )。
- (c) 若  $Cov(abil, educ) > 0$ ，但  $Cov(abil, exper) = 0$ ，則  $\hat{\gamma}_1$  將傾向低估教育對於薪資的真實影響。
- (d) 若  $Cov(abil, educ) > 0$ ，但  $Cov(abil, exper) = 0$ ，則  $\hat{\gamma}_2$  為  $\gamma_2$  的不偏估計式。

12. 承第 11 題，但假設小光使用 CEO 的智商( $IQ$ )作為  $abil$  的代理變數(proxy variable)，並以 OLS 估計下式：

$$\ln(\widehat{wage})_i = \hat{\gamma}_0 + \hat{\gamma}_1 educ_i + \hat{\gamma}_2 exper_i + \hat{\gamma}_3 IQ_i \quad (3)$$

假設  $E(abil|IQ) = \delta_0 + \delta_3 IQ$ ， $\delta_0 \neq 0$ ， $\delta_3 > 0$ 。請問以下敘述何者正確？注意答案可能不只一個。

- (a)  $\hat{\gamma}_0$  為  $\gamma_0$  的不偏估計式。
- (b)  $\hat{\gamma}_1$  為  $\gamma_1$  的不偏估計式。
- (c)  $\hat{\gamma}_2$  為  $\gamma_2$  的不偏估計式。
- (d)  $\hat{\gamma}_3$  為  $\gamma_3$  的不偏估計式。

13. 承第 12 題，但假設  $E(abil|IQ) = \delta_0 + \delta_1 educ + \delta_2 exper + \delta_3 IQ$ ， $\delta_0 \neq 0$ ， $\delta_1 > 0$ ， $\delta_2 < 0$ ， $\delta_3 > 0$ 。請問以下敘述何者正確？注意答案可能不只一個。

- (a)  $\hat{\gamma}_1$  將傾向高估教育對於薪資的影響。
- (b)  $\hat{\gamma}_1$  將傾向低估教育對於薪資的影響。
- (c)  $\hat{\gamma}_2$  將傾向高估年資對於薪資的影響。
- (d)  $\hat{\gamma}_2$  將傾向低估年資對於薪資的影響。

14. 承第 13 題，請寫出  $E(\hat{\gamma}_1)$ 。

15. 若小光在估計(3)式時將 *exper* 的衡量單位由年改為月，請問以下敘述何者正確？注意答案可能不只一個。

- (a)  $\hat{\gamma}_2$  將變大為原來的 12 倍。
- (b)  $\hat{\gamma}_2$  的標準誤(standard error)將變大為原來的 12 倍。
- (c) 檢定  $\hat{\gamma}_2$  是否為零的 *t* 值將不受影響。
- (d) 迴歸的判定係數( $R^2$ )將不受影響。

16. 若小光在估計(3)式時將 *wage* 的衡量單位由新台幣改為美元 (假設 1 美元 = 30 新台幣)，請問以下敘述何者正確？注意答案可能不只一個。

- (a)  $\hat{\gamma}_0$  將比原來小。
- (b)  $\hat{\gamma}_2$  將縮小為原來的 1/30。
- (c)  $\hat{\gamma}_2$  的標準誤(standard error)將縮小為原來的 1/30。
- (d) 迴歸的判定係數( $R^2$ )將不受影響。

17. 假設小光估計(3)式時想要同時檢測年資對於薪資的影響在男性與女性 CEO 之間是否相同。小光考慮創造兩個虛擬變數:  $D_{male}$  及  $D_{female}$ 。若 CEO 為男性則  $D_{male} = 1$ ,  $D_{female} = 0$ ; 若為女性, 則  $D_{male} = 0$ ,  $D_{female} = 1$ 。請問小光可以如何修改迴歸模型及進行檢定？注意答案可能不只一個。

- (a) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 educ_i \times D_{male,i} + \hat{\beta}_3 exper_i + \hat{\beta}_4 IQ_i$ , 檢定  $\hat{\beta}_2$  是否為 0?
- (b) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 educ_i \times D_{female,i} + \hat{\beta}_3 exper_i + \hat{\beta}_4 IQ_i$ , 檢定  $\hat{\beta}_2$  是否為 0?
- (c) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i \times D_{male,i} + \hat{\beta}_2 educ_i \times D_{female,i} + \hat{\beta}_3 exper_i + \hat{\beta}_4 IQ_i$ , 檢定  $\hat{\beta}_1$  是否等於  $\hat{\beta}_2$ ?
- (d) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 D_{male,i} + \hat{\beta}_3 D_{female,i} + \hat{\beta}_4 exper_i + \hat{\beta}_5 IQ_i$ , 檢定  $\hat{\beta}_2$  是否等於  $\hat{\beta}_3$ ?

18. 小光想比較 *educ* 與 *IQ* 何者對於 *wage* 的影響較大，請問可以使用以下哪種方法？注意答案可能不只一個。

- (a) 直接估計(3)式，比較  $\hat{\gamma}_1$  與  $\hat{\gamma}_3$  的大小。
- (b) 直接估計(3)式，比較  $\hat{\gamma}_1 \sigma_{educ}$  與  $\hat{\gamma}_3 \sigma_{IQ}$  的大小， $\sigma$  代表標準差。
- (c) 將 *educ* 與 *IQ* 分別標準化後(扣除平均數再除以標準差)再估計(3)式，比較  $\hat{\gamma}_1$  與  $\hat{\gamma}_3$  的大小。
- (d) 將  $\ln(wage)$ , *educ* 與 *IQ* 分別標準化後再估計(3)式，比較  $\hat{\gamma}_1$  與  $\hat{\gamma}_3$  的大小。

19. 小光覺得年資對於 CEO 薪資的影響可能為非線性，當年資較低時，年資增加對於薪資的正面幫助較大，但當年資已經很高時，再增加年資對於薪資的正面幫助較小。請問要如何修改(3)式來驗證以上的猜想？請同時根據此猜想寫出對於迴歸係數正負符號的預期。

見背面

20. 小光覺得在估計(3)式時應該要額外控制到 CEO 的薪資水準在不同產業間的差異。假設在樣本中的產業包括電子業，銀行業，及食品業三個產業，產業代碼分別為 1, 2, 3。請問小光可以怎麼修改模型以控制產業類別對於薪資水準的效果？注意答案可能不只一個。

(a) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 exper_i + \hat{\beta}_3 IQ_i + \hat{\beta}_4 Id_i$ ,  $Id$  為產業代碼。

(b) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 exper_i + \hat{\beta}_3 IQ_i + \hat{\beta}_4 D_{elec} + \hat{\beta}_5 D_{bank}$ , 若 CEO 在電子業服務則  $D_{elec} = 1$ , 否則  $D_{elec} = 0$ 。若 CEO 在金融業服務則  $D_{bank} = 1$ , 否則  $D_{bank} = 0$ 。

(c) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 exper_i + \hat{\beta}_3 IQ_i + \hat{\beta}_4 D_{elec} + \hat{\beta}_5 D_{food}$ , 若 CEO 在食品業服務則  $D_{food} = 1$ , 否則  $D_{food} = 0$ 。

(d) 估計  $\ln(\widehat{wage})_i = \hat{\beta}_0 + \hat{\beta}_1 educ_i + \hat{\beta}_2 exper_i + \hat{\beta}_3 IQ_i + \hat{\beta}_4 D_{elec} + \hat{\beta}_5 D_{bank} + \hat{\beta}_6 D_{food}$ 。

附錄

Standard Normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641