

等 別：二級考試
類 科：衛生行政（一般組）
科 目：流行病學與生物統計學研究
考試時間：2 小時

座號：_____

※注意：(一)可以使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

- 一、請定義何謂柏克森偏差（Berkson's Bias）？並舉例說明，以此例論述此誤差對結果的影響。（20 分）
- 二、最近有一研究探討有和沒有糖尿病的心肌梗塞病人接受保栓通（clopidogrel）治療和全死因的相關，此研究從 2002-2009 年住院檔中找出 58,851 位發生心肌梗塞存活且沒有接受心導管手術的病人，假如此國家中有糖尿病的心肌梗塞病人接受保栓通治療的盛行率為 50%，而沒有糖尿病的心肌梗塞病人接受保栓通治療的盛行率為 30%，下列表格為依照糖尿病狀態呈現接受或沒接受保栓通治療之心肌梗塞患者，其死亡人數、人年、粗發生率和調整風險比（Adjusted Hazard Ratio, HR）。

	Diabetes (N=7247)		No Diabetes (N=51604)	
	Clopidogrel (N=4078)	No Clopidogrel (N=3169)	Clopidogrel (N=31302)	No Clopidogrel (N=20302)
All-cause mortality				
Events/total, No. (%)	472/4078 (12)	753/3169 (24)	1816/31302 (6)	3661/20302 (18)
Person-years	3518	2566	28183	17156
Crude incidence rates	13.4 (12.8-14.0)	29.3 (28.3-30.4)	6.4 (6.3-6.6)	21.3 (21.0-21.7)
Adjusted HR (95%CI)	0.89 (0.79-1.00)	1.00	0.75 (0.70-0.80)	1.00
P for Interaction	0.001			

CI: confidence interval; HR: hazard ratio

- (一)根據表格粗發生率（crude incidence rates），以糖尿病狀態分層，計算接受保栓通治療的全死因相差危險性（risk difference）、需要被治療的病人數目（number needed to treat）和族群可歸因危險性（population attributable risk），並解釋結果。（10 分）
- (二)根據表格粗發生率，以糖尿病狀態分層，計算接受保栓通治療的全死因相對危險性（relative risk），並解釋結果。（10 分）
- (三)研究者進一步調整了傳統的危險因子後，得到上面表格中的調整風險比，試討論糖尿病所扮演的角色為干擾因子？還是交互作用（effect modifier）？（5 分）
- (四)根據表格結果，身為一位流行病學家，如何報導和解釋此研究結果？（5 分）

等 別：二級考試
類 科：衛生行政（一般組）
科 目：流行病學與生物統計學研究

三、某研究為探討某一種史塔汀類（statin）藥物其降低膽固醇的作用，所收集到的資料整理如下： $\bar{y}_1$ 、 $\bar{y}_2$ 分別表示為沒服用該藥物、有服用該藥物之樣本平均值， s_1 、 s_2 分別表示為沒服用該藥物、有服用該藥物之樣本標準差，此兩組樣本之樣本數相同，以 n 表示之，兩組樣本之膽固醇數值群體分布雖可能有不同之群體平均值，但此假設兩組樣本有相同之群體標準差，以 σ 表示之。考慮到此藥物其降低膽固醇的作用會因人而異，所以收集 n 位受試者測量其之前沒使用此藥物之膽固醇數值，以及之後使用此藥物長達半年之膽固醇數值的成對資料。此檢定之顯著水準為 0.05。

請回答下列問題：

- (一)依題意定義此研究之目標參數，寫出以此參數表示的虛無假說與對立假說。如上所述，當 $\bar{y}_1$ 、 $\bar{y}_2$ 、 s_1 、 s_2 、 n 、 σ 之數值均為已知訊息時，這些數值是否已提供足夠訊息來計算 p-value？為什麼？當答案為否時，說明需要那些額外資訊才能計算 p-value。而當已經有足夠資訊來計算 p-value 時，說明樣本數 n 與 p-value 之關係，以及說明檢定力函數（power function of a test）的意義。（20 分）
- (二)當原始資料之紀錄為 $\{(y_i, x_i), i=1, \dots, n, n+1, \dots, 2n\}$ ，其中，當 $i=1, \dots, n$ 時，設定 $x_i=0$ 及對應之 y_i 代表第 i 位受試者沒用藥時的膽固醇數值，而當 $i=n+1, \dots, 2n$ 時，設定 $x_i=1$ 及 y_i 代表第 $i-n$ 位受試者用藥時的膽固醇數值，寫出以 y_i 為反應變數及 x_i 為解釋變數之簡單線性迴歸模式，說明 x_i 的迴歸係數的意義，此簡單線性迴歸模式是否適合用來描述此成對資料以回答此研究問題「某種史塔汀類藥物是否有降低膽固醇的作用」？請說明統計理由。（10 分）

四、探討停經後婦女罹患乳癌是否與其生第一胎之年齡有關，從某社區停經婦女罹患乳癌（ $Y=1$ ）中隨機選取 50 位，其中有 15 位其生第一胎之年齡在 30 歲以前，同時從同一社區停經且沒有乳癌之婦女（ $Y=0$ ）隨機選取 50 位，其中有 35 位其生第一胎之年齡在 30 歲以前。在此，以 $X=0$ 表示生第一胎之年齡在 30 歲以前， $X=1$ 則表示生第一胎之年齡在 30 歲以後。

請回答以下問題：

- (一)將上述資料整理成一個 2×2 列聯表，寫出依題意與符號所對應之勝算比（Odds Ratio, 簡稱 OR）的意義，說明為何 OR 可以用來測量停經後婦女罹患乳癌與否與其生第一胎之年齡之相關性。計算此 OR 的近似 95% 信賴區間。當檢定停經後婦女罹患乳癌是否與其生第一胎之年齡有關之顯著水準設為 0.05 時，說明如何以此信賴區間來判斷檢定結果。（15 分）
- (二)若想要進一步推算出在 30 歲以後生第一胎之停經婦女組有乳癌的機率比在 30 歲以前生第一胎之停經婦女組有乳癌的機率的比值（稱為相對風險比值，Relative Risk, 簡稱 RR），除上述資料外，還需要提供何種資訊才可以估算出 RR 的數值？說明統計原因。（5 分）

（請接第三頁）

等 別：二級考試
 類 科：衛生行政（一般組）
 科 目：流行病學與生物統計學研究

附表

Z 分布										
$\Pr(Z \leq 1.96) = 0.975$ ； $\Pr(Z \leq 1.64) = 0.95$ ； $\Pr(Z \leq 1.28) = 0.90$										
t 分布 $\Pr(t_7 \leq 1.895) = 0.95$										
df	7	8	9	15	22	23	24	30	40	
$t_{0.95}$	1.895	1.860	1.833	1.753	1.717	1.714	1.711	1.697	1.684	
$t_{0.975}$	2.365	2.306	2.262	2.131	2.074	2.069	2.064	2.042	2.021	
χ^2 分布 $\Pr(\chi_1^2 \leq 3.84) = 0.95$					F 分布 $\Pr(F_{1,7} \leq 5.59) = 0.95$					
df	1	2	3	10	20	df_1, df_2	1,7	2,7	1,16	2,16
$\chi_{0.95}^2$	3.84	5.99	7.81	18.31	31.41	$F_{0.95}$	5.59	4.74	4.49	3.63
$\chi_{0.975}^2$	5.02	7.38	9.35	20.48	34.17	$F_{0.975}$	8.07	6.54	6.12	4.69