

醒：

- 第一頁上方註明姓名、學號、系級
- 請以 PDF 檔案繳交
- 除了回答問題，結果還需包含文字敘述和 SAS 程式碼

作業檔案中 HW2_cohort 為一個長期追蹤研究資料，此檔案中包含受試者的性別(sex)、年齡(age)、收縮壓(sbp)、舒張壓(dbp)、BMI(bmi)、中風情形(stroke)、追蹤時間(follow-up)等資料。除了追蹤時間與中風情形是追蹤研究結束後得知，其餘資料都是在受試者剛開始參與追蹤研究時由測量或填寫問卷得知。請利用此檔案進行第一題分析。

1. 高血壓是造成中風重要的危險因子之一，本研究判藉由模型控制高血壓的影響，單純探討不同性別對於罹患中風危險性是否不同？(高血壓標準：收縮壓 ≥ 140 mmHg 或舒張壓 ≥ 90 mmHg) 請解釋結果。(20%)

```
/*分類高血壓=1*/  
data datac2; set datac ;  
    if sbp >= 140 or dbp >= 90 then bp = 1 ;  
    else bp = 0 ;  
run ;  
  
/*建立cox ph model*/  
proc phreg data = datac2 ;  
model followup*stroke(0)=sex bp/risklimits;  
/*risklimit: to calculate the confidence interval*/  
run;
```

最大概度估計值分析								
參數	DF	參數估計值	標準誤差	卡方	Pr > ChiSq	危險率	95% 危險率信賴界限	標籤
sex	1	0.51790	0.08027	41.6238	<.0001	1.678	1.434 1.964	sex
bp	1	0.72693	0.08005	82.4563	<.0001	2.069	1.768 2.420	

控制高血壓的影響後，男性較女性的中風 Hazard ratio 為 1.678，95%信賴區間為 1.434-1.964，達統計顯著的差異($p < 0.0001 < \alpha(0.05)$)

題組：第二題及第三題

某一金融公司調查全體員工健康情形，調查資料存為兩個檔案(HW2_data1, HW2_data2)，公司主管希望知道員工的健康情形。請利用此兩個檔案協助公司主管了解該公司的員工健康。

2. 肥胖與心血管疾病是職場健康促進主要希望預防疾病之一，為落實提早預防的理念，該公司協助調查員工有代謝症候群的盛行率為何。員工資料被儲存成兩個檔案，請協助將資料合併後，再進行分析。(30%) 代謝症候群分類標準：
- 腰圍過大：男性 ≥ 90 公分、女性 ≥ 80 公分
 - 血壓異常：收縮壓 $\geq 130\text{mmHg}$ 或舒張壓 $\geq 85\text{mmHg}$
 - 高密度脂蛋白：男性 $< 40\text{mg/dl}$ 、女性 $< 50\text{mg/dl}$
 - 血糖值過高：空腹血糖 $\geq 100\text{mg/dl}$
 - 三酸甘油酯過高：三酸甘油酯 $\geq 150\text{mg/dl}$ 此項指標中，達三項異常就為代謝症候群。

```
/*資料合併*/
DATA data2c;
    merge data2_1 data2_2;
    by ID;
RUN;

/*分類代謝症候群(達三項異常)*/

data data2c2; set data2c ;

    /*腰圍過大：男性 $\geq 90$ 公分、女性 $\geq 80$ 公分 */
    if sex = 1 and waist >= 90 then waistbig = 1 ;
    else if sex = 0 and waist >= 80 then waistbig = 1 ;
    else waistbig = 0 ;

    /*血壓異常：收縮壓 $\geq 130\text{mmHg}$ 或舒張壓 $\geq 85\text{mmHg}$  */
    if (sbp >= 130) or (dbp >= 85) then bphigh = 1 ;
    else bphigh = 0 ;

    /*高密度脂蛋白：男性 $< 40\text{mg/dl}$ 、女性 $< 50\text{mg/dl}$  */
    if sex = 1 and hdl < 40 then proteinless = 1 ;
    else if sex = 0 and hdl < 50 then proteinless = 1 ;
    else proteinless = 0 ;

    /*血糖值過高：空腹血糖 $\geq 100\text{mg/dl}$  */
    if glu >= 100 then gluhigh = 1 ;
    else gluhigh = 0 ;

    /*三酸甘油酯過高：三酸甘油酯 $\geq 150\text{mg/dl}$  */
    if tri >= 150 then trihigh = 1 ;
    else trihigh = 0 ;

    /*達三項*/
    total = waistbig + bphigh + proteinless + gluhigh + trihigh ;
    if total >= 3 then metabolic = 1 ;
    else metabolic = 0 ;

run ;
```

```
PROC FREQ DATA= data2c2;
  TABLE metabolic;
RUN;
```

metabolic	次數	百分比	累積 次數	累積 百分比
0	1690	84.50	1690	84.50
1	310	15.50	2000	100.00

代謝症候群的盛行率=15.5%

3. 若腰圍和 BMI 是代謝症候群重要的危險因子，金融公司主管想了解腰圍和 BMI 之間的相關性。(標示統計方法、相關係數，並解釋結果。)(20%)

```
/*計算BMI*/
data data2c3 ; set data2c ;
  BMI = weight / (( height / 100 ) ** 2 ) ;
run;

/*to test the normal distribution*/
proc univariate data=data2c3 normal;
  var waist bmi;
run;

/*Spearman's rank correlation coefficient*/
ods graphics on;
proc corr data=data2c3 plots=matrix spearman;
var waist bmi;
run;
ods graphics off;
```

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.982451	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.04492	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.975318	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	6.551186	Pr > A-Sq	<0.0050

常態性檢定				
檢定	統計值		p 值	
Shapiro-Wilk	W	0.965158	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.050086	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1.738604	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	12.0258	Pr > A-Sq	<0.0050

H0: waist 符合常態分布

H1: waist 不符合常態分布

p-value<0.05，拒絕 H0，腰圍未符合常態分佈。

H0: BMI 符合常態分布

H1: BMI 不符合常態分布

p-value<0.05，拒絕 H0，BMI 未符合常態分佈。

waist 和 BMI 均不為常態分佈，因此使用 Spearman's rank correlation coefficient

Pearson 相關係數, N = 2000 Prob > r (低於 H0): Rho=0			Spearman 相關係數, N = 2000 Prob > r (低於 H0): Rho=0		
	waist	BMI		waist	BMI
waist	1.00000	0.83162	waist	1.00000	0.82297
waist		<.0001	waist		<.0001
BMI	0.83162	1.00000	BMI	0.82297	1.00000
BMI		<.0001	BMI		<.0001

waist 與 BMI 的相關係數約為 0.823。 $p < 0.0001 < \alpha (0.05)$ ，拒絕 H_0 ，兩變數間高度相關。BMI 隨著 waist 增加而上升，具統計顯著的差異。

4. Dr. Epi 進行一項高血壓與危險因子的研究，想調查高血壓患者可能的危險因子。此研究招募 60 位高血壓患者和 240 位健康人，同時詢問兩組飲食習慣後發現，高血壓患者有 40% 飲食偏向重口味、60% 飲食偏向清淡，健康對照組則有 25% 飲食偏向重口味、75% 飲食偏向清淡。請問高血壓與飲食口味的相關性和相關強度為何？請劃出 2*2 table (可參考實習課講義)、計算相關強度，並下結論解釋之。(30%)

	飲食		
	重口味=1	清淡口味=0	
高血壓=1	60*40%=24	60*60%=36	60
健康者=0	240*25%=60	240*75%=180	240

```

/*4*/
/*=Strength of association=*/
/*=建立資料=*/
data data4;
input htn$ diet$ count; /*高血壓患者=1 重口味=1 比例*/
datalines;
1 1 24
1 0 36
0 1 60
0 0 180
;

proc freq data = data4 order=data ;
weight count;
tables htn*diet /chisq relrisk;
run;

```

FREQ 程序				
次數 百分比 列百分比 欄百分比	htn * diet的表格			
	htn	diet		
		1	0	總計
	1	24	36	60
		8.00	12.00	20.00
		40.00	60.00	
		28.57	16.67	
	0	60	180	240
		20.00	60.00	80.00
		25.00	75.00	
71.43		83.33		
總計	84	216	300	
	28.00	72.00	100.00	

htn * diet 之表格的統計值			
統計值	DF	值	機率
卡方	1	5.3571	0.0206
擬度比卡方	1	5.0897	0.0241
連續性調整卡方	1	4.6389	0.0313
Mantel-Haenszel 卡方	1	5.3393	0.0209
Phi 係數		0.1336	
列聯係數		0.1325	
Cramer V		0.1336	

勝算比和相對風險			
統計值	值	95% 信賴界限	
勝算比	2.0000	1.1049	3.6203
相對風險 (欄 1)	1.6000	1.0947	2.3386
相對風險 (欄 2)	0.8000	0.6426	0.9960

P-value=0.0206 < $\alpha(0.05)$ ，拒絕 H0。飲食口味和高血壓有統計顯著的相關。

高血壓與飲食口味的勝算比為 2.00 (95% C.I.: 1.10-3.62)