

108-2 流行病學實習課期末上機考

2020 年 6 月 17 日

姓名：梁嫚芳

學號：b07801003

系級：公衛二

※注意事項

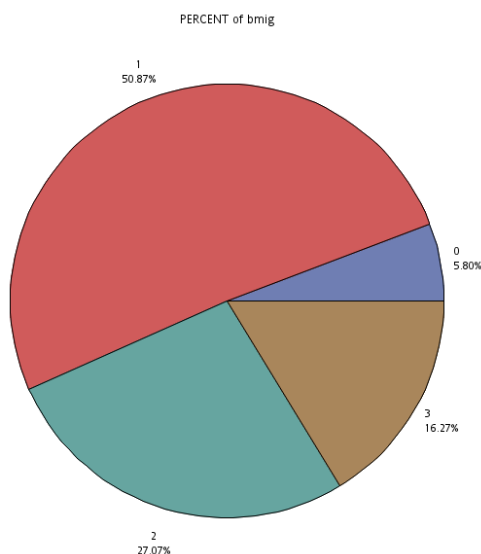
請於上方填上姓名、學號和系級。每題答案需闡釋解釋結果與結論，並附上 SAS code 與表格，最後上傳 PDF 檔至 NTU COOL 作業區。

六題任選四題作答，完成六題者其中兩題作為加分用。盡可能寫答案，依據寫的答案正確性斟酌給分。

1. 肥胖與心血管疾病是社區健康促進的重點預防疾病，為落實社區預防醫學的理念，應先了解社區身體質量指數(BMI)各組之盛行狀況，再配合適當的宣導介入措施。社區資料被儲存成兩個檔案，請協助將資料合併後，以圓餅圖呈現身體質量指數(BMI)各組的盛行狀況。(附上圓餅圖結果，並敘述 BMI 各組盛行率) (25%)

身體質量指數(BMI)各組分類標準：

- 過輕：BMI<18.5
- 正常：18.5≤BMI<24
- 過重：24≤BMI<27
- 肥胖：≥27



- 過輕：BMI<18.5 盛行率=5.8%
- 正常：18.5≤BMI<24 盛行率=50.87%

- 過重： $24 \leq \text{BMI} < 27$ 盛行率=27.07%
- 肥胖： ≥ 27 盛行率=16.27%

```

/*1*/
PROC SORT data=data1;
by ID;
RUN;
PROC SORT data=data2;
by ID;
RUN;

DATA data3;
merge data1 data2;
by ID;
run;

data datab;set data3;
bmi=weight/((height/100)**2);
if bmi<18.5 then bmig=0;
else if bmi<24 then bmig=1;
else if bmi<27 then bmig=2;
else bmig=3;
run;

proc gchart data=datab;
pie bmig/discrete type=percent;
run;
quit;

```

2. 除了調查身體質量指數各組的盛行率，公衛師想透過有、無罹患血管疾病的居民心其年齡和高血壓病史，以規劃健康促進計畫。依資料型態分別用平均值、標準差、樣本數及百分比呈現，並做檢定比較有無心血管疾病兩組是否具有統計學上差異。(不用呈現 SAS 分析圖表，完成下表並附上 SAS code 即可，四捨五入至小數第二位)(25%)

無心血管疾病 (n= 1399 , 93.27 %)	有心血管疾病 (n= 101 , 6.73 %)	p-value
----------------------------------	--------------------------------	---------

年齡, mean (s.d)	45.85	(15.14)	57.82 (16.90)	<0.0001
罹患高血壓, n (%)	205	(13.67)	40 (2.67)	<0.0001

```

/*2*/
proc freq data=datab;
table cvd;
run;

proc means data=datab mean std;
class cvd;
var age;
run;

proc freq data=datab;
table cvd*htn;
run;

proc ttest data=datab alpha=0.05;
class cvd;
var age;
run;

proc freq data=datab;
table cvd*htn/chisq expected fisher;
run;

```

3. 請協助瞭解該社區居民心血管疾病盛行狀況和 BMI (分組)的關係，就目前所知，**年齡、性別、高血壓**病史可能為干擾因子。心血管疾病為二分類變項，利用 logistic regression (羅吉斯回歸模型) 校正干擾因子以計算相關強度，回答以下問題：

(1) 解釋過輕和肥胖組的結果和結論。

(2) 解釋年齡的結果和結論。

請附上 SAS 分析回歸係數、OR 報表結果。BMI 分組請以”正常”做為參考組、性別以”女性”做為參考組，高血壓病史請以”正常”做為參考組進行分析。(25%)

1)

過輕:

H0:過輕組的迴歸係數=0

結論: $P=0.7577>\alpha(0.05)$ ，不拒絕 h_0 ，與 0 未達統計顯著差異。控制其他變項後，過輕組較正常組的罹患心血管疾病的 OR 平均增加 0.842 倍

(95%CI=(0.283,2.509)包含 1，OR 與 1 未達統計顯著差異)。

肥胖:

H0:肥胖組的迴歸係數=0

結論:P=0.6708>alpha(0.05)，不拒絕 h0，與 0 未達統計顯著差異。控制其他變項後，肥胖組較正常組的罹患心血管疾病的 OR 平均增加 0.879 倍

(95%CI=(0.484,1.595)包含 1，OR 與 1 未達統計顯著差異)。

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.6297	0.4175	122.9881	<.0001
bmig 0	1	-0.1719	0.5570	0.0952	0.7577
bmig 2	1	0.0178	0.2478	0.0051	0.9429
bmig 3	1	-0.1293	0.3041	0.1807	0.6708
age	1	0.0392	0.00744	27.7706	<.0001
sex 1	1	-0.4774	0.2212	4.6579	0.0309
htn	1	0.7642	0.2460	9.6500	0.0019

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	73.3	Somers' D	0.465
Percent Discordant	26.7	Gamma	0.466
Percent Tied	0.0	Tau-a	0.058
Pairs	141299	c	0.733

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
bmig 0 vs 1	1.0000	0.842	0.283	2.509
bmig 2 vs 1	1.0000	1.018	0.626	1.654
bmig 3 vs 1	1.0000	0.879	0.484	1.595
age	1.0000	1.040	1.025	1.055
sex 1 vs 0	1.0000	0.620	0.402	0.957
htn 1 vs 0	1.0000	2.147	1.326	3.478

2)

H0:年齡的迴歸係數=0

結論:P=<0.0001<alpha(0.05)，不拒絕 h0，與 0 達統計顯著差異。控制其他變項後，年齡每增加一歲罹患心血管疾病的 OR 平均增加 1.040 倍

(95%CI=(1.025,1.055)不包含 1，OR 與 1 達統計顯著差異)，年齡與罹患心血管疾病達統計顯著相關。

/*3*/

proc logistic data=datab;

class bmig(ref='1') htn(ref='0') sex(ref='0') /param=ref;

model cvd(event='1')=bmig age sex htn/risklimits;

4. run;腰圍代表腹部肥胖的狀況，它和身體質量指數(BMI)分別都是罹患慢性疾病的重要危險因子，請協助評估該社區居民腰圍和 BMI 之間的相關程

度，並以相關係數表示。(假設腰圍和 BMI 分布均服從常態分佈，請附上 SAS 報表及敘述分析結果及你的結論) (25%)

Pearson Correlation Coefficients, N = 1500 Prob > r under H0: Rho=0		
	waist	bmi
waist	1.00000	0.82965
waist		<.0001
bmi	0.82965	1.00000
	<.0001	

假設腰圍和 BMI 分布均服從常態分佈，pearson 相關係數為 0.82965， $p < 0.0001 < \text{ALPHA}(0.05)$ ，腰圍與 bmi 達統計顯著正向相關，腰圍會隨 bmi 增加而上升。

```
/*4*/
ods graphics on;
proc corr data=datab plots=scatter pearson;
var waist bmi;
run;
ods graphics off;
```

5. Dr. Epi 針對該社區進行另一個前瞻性研究，追蹤社區居民血壓偏高 (60 位) 及血壓正常 (240 位) 的受試者 15 年，希望了解日後發生中風的風險，研究結果發現，35% 血壓偏高者發生中風的狀況，15% 血壓正常者發生中風的狀況。請問血壓偏高是否會造成日後發生中風的風險增加？(計算相關強度，並敘述分析結果及結論，請附上 SAS 計算相關強度的報表) (25%)

Exposure	Disease		Total
	中風	無中風	
血壓偏高	21	39	60
血壓正常	36	204	240
Total	57	243	300

H0: 高血壓與中風無相關

結論： $P = 0.0004 < \text{ALPHA}(0.05)$ ，拒絕 H0。高血壓與中風有統計顯著相關。相關風險 RR 為 2.3333(95%CI=1.4762-3.6881)，暴露於血壓偏高的中風風險為正常的 2.3333 倍，且 95%信賴區間未包含 1，達統計顯著性。

Statistic	DF	Value	Prob
Chi-Square	1	12.4756	0.0004
Likelihood Ratio Chi-Square	1	11.1398	0.0008
Continuity Adj. Chi-Square	1	11.2099	0.0008
Mantel-Haenszel Chi-Square	1	12.4340	0.0004
Phi Coefficient		0.2039	
Contingency Coefficient		0.1998	
Cramer's V		0.2039	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	21
Left-sided Pr <= F	0.9998
Right-sided Pr >= F	0.0007
Table Probability (P)	0.0005
Two-sided Pr <= P	0.0008

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	3.0513	1.6122	5.7748
Relative Risk (Column 1)	2.3333	1.4762	3.6881
Relative Risk (Column 2)	0.7647	0.6304	0.9276

```

/*5*/
data a;
input d$ e$ count;
datalines;
1 1 21
1 0 36
0 1 39
0 0 204
;
proc freq data=a order=data;
weight count;
tables e*d/chisq relrisk;
run;

```

6. 假設血糖值為變異數相等的常態分佈。將血糖值視為反應變數，BMI 分組

視為解釋變數，擬合回歸模型。請問血糖值是否可以被 BMI 分組預測？請

解釋結果，並說明判定係數 (R^2)。(25%)

H0:bmi 分組的迴歸係數=0

結論:

BMI 分組能預測血糖值

Bmi 過輕: $p=0.1836 > \alpha(0.05)$ ，不拒絕 h_0 ，與 0 未達統計顯著差異，控制其他變相後，bmi 過輕較正常的血糖值平均增加 -3.26124945mg/dL。

Bmi 過重: $p < 0.0001 < \alpha(0.05)$ ，拒絕 h_0 ，與 0 達統計顯著差異，控制其他變相後，bmi 過輕較正常的血糖值平均增加 6.38242871mg/dL。

Bmi 肥胖: $p < 0.0001 < \alpha(0.05)$ ，拒絕 h_0 ，與 0 達統計顯著差異，控制其他變相後，bmi 過輕較正常的血糖值平均增加 7.97782158mg/dL。

判定係數 = 0.028947，此回歸模型中 BMI 分組能夠解釋血糖值總變異的 2.8947%。

The SAS System					
The GLM Procedure					
Dependent Variable: glu glu					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	20928.3588	6976.1196	14.87	<.0001
Error	1496	702058.0972	469.2902		
Corrected Total	1499	722986.4560			

R-Square	Coeff Var	Root MSE	glu Mean
0.028947	24.40528	21.66311	88.76400

Source	DF	Type I SS	Mean Square	F Value	Pr > F
bmig	3	20928.35876	6976.11959	14.87	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
bmig	3	20928.35876	6976.11959	14.87	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	85.92791612	B	0.78425710	109.57	<.0001
bmig 0	-3.26124945	B	2.45136678	-1.33	0.1836
bmig 2	6.38242871	B	1.33076907	4.80	<.0001
bmig 3	7.97782158	B	1.59322929	5.01	<.0001
bmig 1	0.00000000	B	.	.	.

```
/*6*/
```

```
proc glm data=datab;
class bmig(ref='1');
model glu=bmig/solution;
run;
```

所有 CODE

```
/*1*/  
PROC SORT data=data1;  
by ID;  
RUN;  
PROC SORT data=data2;  
by ID;  
RUN;  
  
DATA data3;  
merge data1 data2;  
by ID;  
run;  
  
data datab;set data3;  
bmi=weight/((height/100)**2);  
if bmi<18.5 then bmig=0;  
else if bmi<24 then bmig=1;  
else if bmi<27 then bmig=2;  
else bmig=3;  
run;  
  
proc gchart data=datab;  
pie bmig/discrete type=percent;  
run;  
quit;  
  
/*2*/  
proc freq data=datab;  
table cvd;  
run;  
  
proc means data=datab mean std;  
class cvd;  
var age;  
run;  
proc freq data=datab;
```



```

table cvd*htn;
run;
proc ttest data=datab alpha=0.05;
class cvd;
var age;
run;
proc freq data=datab;
table cvd*htn/chisq expected fisher;
run;

/*3*/
proc logistic data=datab;
class bmig(ref='1') htn(ref='0') sex(ref='0') /param=ref;
model cvd(event='1')=bmig age sex htn/risklimits;
run;

/*4*/
ods graphics on;
proc corr data=datab plots=scatter pearson;
var waist bmi;
run;
ods graphics off;

/*5*/
data a;
input d$ e$ count;
datalines;
1 1 21
1 0 36
0 1 39
0 0 204
;
proc freq data=a order=data;
weight count;
tables e*d/chisq relrisk;
run;

/*6*/
proc glm data=datab;

```

```
class bmig(ref='1');  
model glu=bmig/solution;  
run;
```