**108-2** 流行病學實習課作業三更新

梁嫚芳

B07801003

公衛二

繳交日期: 2020.06.**05** 提醒：

* 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

題組：第一題至第三題

作業檔案 data1 為一個 Cohort study 研究資料，此檔案包含受試者的性別(sex)、年齡(age)、收縮壓(sbp)、舒張壓(dbp)、BMI(bmi)、中風(stroke)、追蹤時間(followup) 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。

請利用此檔案進行以下分析：

1. 請問在此 Cohort study 資料中有多少高血壓的盛行個案？ (高血壓標準：收縮壓140mmHg 或舒張壓90mmHg) （13%）

高血壓的盛行個案有740個，盛行率為37%

/\*1\*/

/\*將血壓分組\*/

**data** data1\_1 ;

set data1 ;

if sbp >= **140** or dbp >= **90** then bp = **1** ;

else bp = **0** ;

**run** ;

/\*呈現表格\*/

**PROC** **FREQ** DATA=data1\_1;

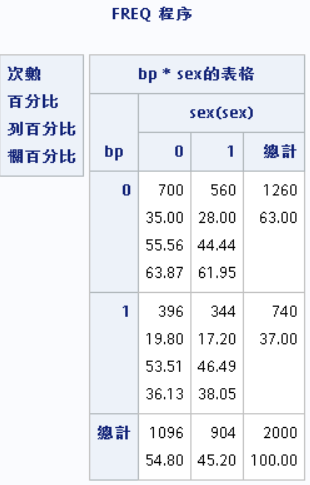
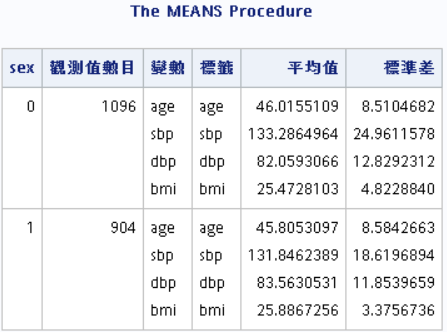
TABLE bp ;

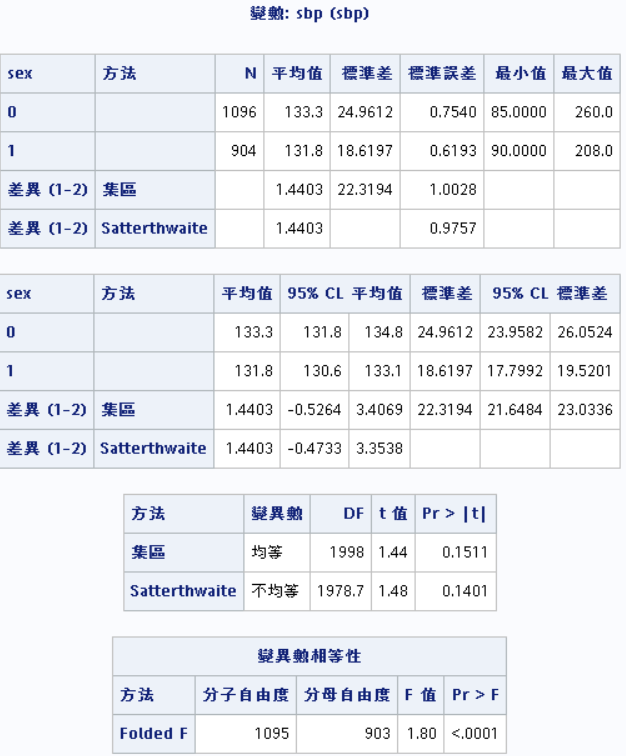
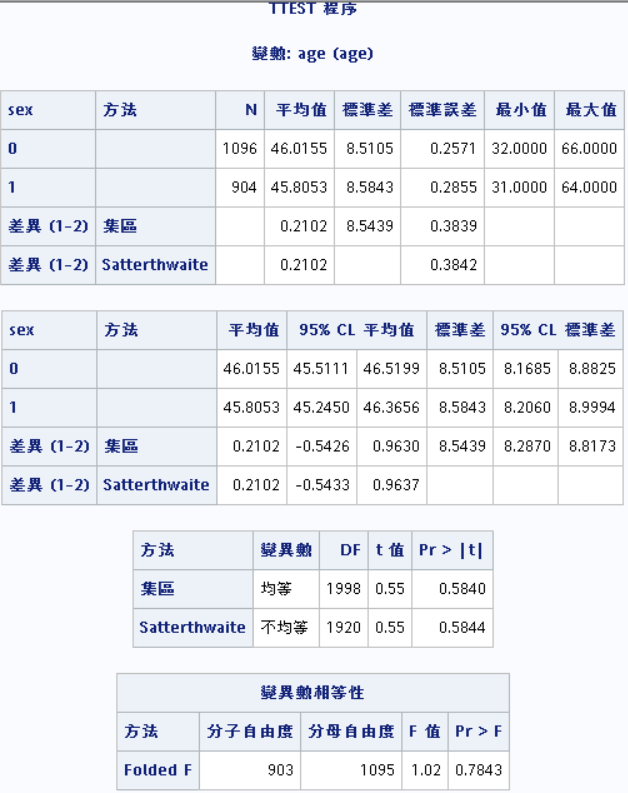
**RUN**;

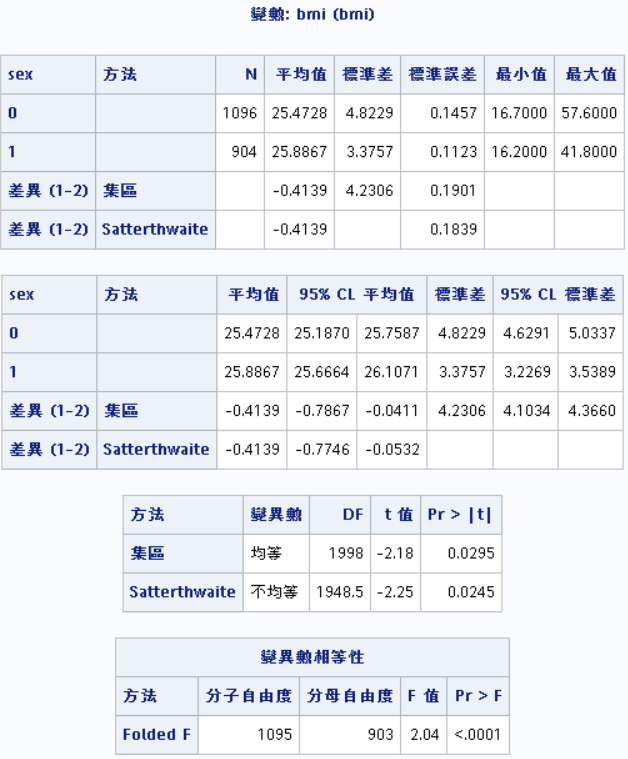
1. 下表欲呈現該 Cohort study 的受試者男性與女性基本人口學資料，請協助完成。請依資料型態分別用平均值、標準差、樣本數及百分比呈現，並檢定不同性別該變項是否有顯著差異。(分析變項：年齡、收縮壓、舒張壓、

BMI、高血壓)（27%）

|  |  |  |  |
| --- | --- | --- | --- |
|  | 男性  (n=904, 45.2%) | 女性  (n= 1096, 54.8%) | p-value |
| 年齡 (mean, s.d)  收縮壓 (mean, s.d)  舒張壓 (mean, s.d)  BMI (mean, s.d)  罹患高血壓 (n, %) | 45.80, 8.58  131.85, 18.62  83.56, 11.85  25.89, 3.38  344, 46.49 | 46.02, 8.51  133.29, 24.96  82.06, 12.83  25.47, 4.82  396, 53.51 | 0.5840  0.1401  0.0066  0.0245  0.3759 |







H0: 男性的變項平均值=女性的變項平均值

變項分別代入：年齡、收縮壓、舒張壓、BMI、高血壓

結論：根據t-test，變項舒張壓、bmi的p-value分別為0.0066、0.0245，皆小於α (0.05)，拒絕虛無假說，因此舒張壓、bmi在不同性別間具有統計顯著的差異。其餘變項年齡、收縮壓、罹患高血壓之p-value皆大於α (0.05) ，不拒絕虛無假說，因此在不同性別間無統計顯著的差異。

/\*2\*/

/\*呈現表格\*/

**PROC** **MEANS** DATA=data1\_1 mean std;

class sex;

VAR age sbp dbp bmi;

**RUN**;

**PROC** **FREQ** DATA=data1\_1;

TABLE bp\*sex ;

**RUN**;

/\*檢定\*/

**PROC** **TTEST** DATA=data1\_1;

CLASS sex;

VAR age sbp dbp bmi bp;

**RUN**;

1. 以存活分析中的 Kaplan-Meier method 估計性別不同其罹患中風 (stroke) 的狀況，並利用 Log-rank test 檢定不同性別的存活曲線是否有不同？需呈現存活曲線(survival curve)圖、列出虛無假設、檢定結果與針對圖表結果加以闡釋。(中風：stroke=1，無中風：stroke=0) （25%）

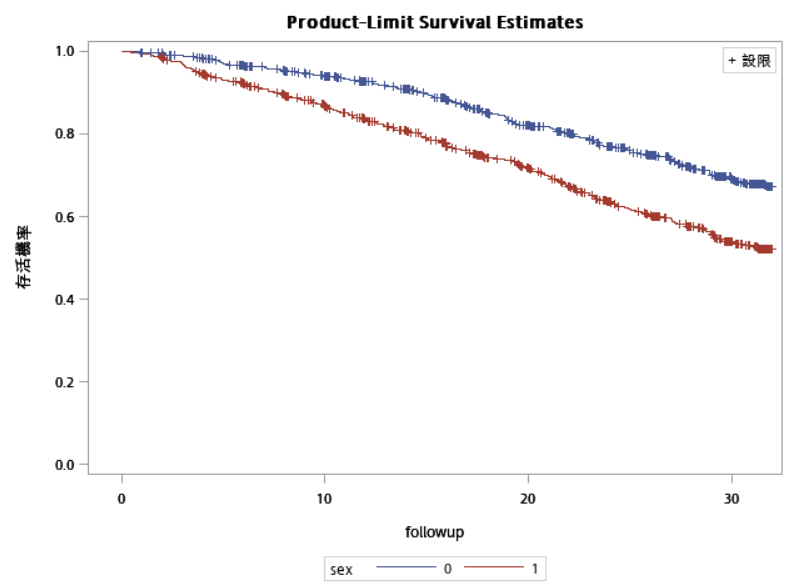


H0: S0(t)=S1(t) S0(t): sex=0

H1: S0(t)≠S1(t) S1(t): sex=1

結論：根據Log-rank test 檢定不同性別的存活曲線，p-value<0.0001<α(0.05)

拒絕H0，S0(t)≠S1(t)，不同性別間的存活曲線達統計顯著的差異。



女性的存活曲線皆高於男性的的存活曲線，即曲線下面積大於男性，女性中風的風險較男性低，性別不同有效影響中風情形。

/\*3\*/

/\*Kaplan-Meier analysis and Log-Rank test\*/

ods graphics on;

**proc** **lifetest** data=data1\_1 method=pl plots=(survival);

/\*method:Kaplan-Meier estimates\*/

time followup\*stroke(**0**);

strata sex;

**run**;

ods graphics off;

1. 利用 data2 進行分析。假設總膽固醇為變異數相等的常態分佈。BMI 依照過輕(<18.5)、正常(18.5-24)、過重(≥24)分為三組。將膽固醇視為反應變

數，BMI 分組視為解釋變數，擬合回歸模型。請問膽固醇是否可以被 BMI 分組預測？請解釋結果，並說明判定係數 (R2)。 (35%)



R-square = 0.038331，此迴歸模型中bmi分組可解釋膽固醇總變異的3.8331%。

|  |  |  |
| --- | --- | --- |
| BMI | X1(1) | X1(2) |
| 過輕 | 1 | 0 |
| 正常 | 0 | 1 |
| 過重 | 0 | 0 |

Y=β0+β1(1)X1(1)+ β1(2)X1(2)+ϵ

Y: 膽固醇

ϵ: model的誤差

β0: model的截距

β1(1): bmi過輕相較於過重對膽固醇的影響。當固定其他變項後，bmi過輕的膽固醇平均較過重減少27.14mg/dL

β1(2): bmi正常相較於過重對膽固醇的影響。當固定其他變項後，bmi正常的膽固醇平均較過重減少10.57mg/dL

Y=185.50-27.14\*X1(1)-10.57\*X1(2)+ϵ

H0 : βi =0 ; i=1(1) , 1(2)

H1 : βi ≠0 ; i=1(1) , 1(2)

結論：β1(1)與β1(2)的p-value<0.0001<α(0.05)，皆拒絕H0，β1(1)與β1(2)與0達統計顯著的差異，因此膽固醇可以被 BMI 分組預測。

/\*4\*/

libname hw3 'C:\Users\user\OneDrive - 國立台灣大學\108-2\流病\期末考\SAS HW\SAS Homework 3\_for student' ;

/\*bmi類別化\*/

**data** data2\_1; set hw3.hw3\_data2;

bmi = weight / (height/**100**)\*\***2** ;

if bmi < **18.5** then bmi\_g = **0** ;

else if bmi < **24** then bmi\_g = **1** ;

else if bmi >= **24** then bmi\_g = **2** ;

else bmi\_g= **.** ;

**run** ;

/\*linear regression\*/

**proc** **glm** data=data2\_1;

class bmi\_g; /\*categorical variable\*/model

chol=bmi\_g/solution; /\*solution: to illustrate regression coefficients\*/

**run**;

1. 假設我們對於社區居民糖尿病與抽菸狀況的關係有興趣。請將血糖值進行分組，血糖值≥126 mg/dL視為糖尿病。目前已知與糖尿病相關的因子分別

有：年齡、性別、體重和運動習慣。請針對以下題目進行解釋與說明結論。

* 1. 請問糖尿病與抽菸之間的關係為何？（校正已知的干擾因子）(15%)

1. After adjusting for age, sex, weight, exercise habits,   
   the OR of smoke=1 vs 0 is <0.001, p-value=0.9813, it’s not significantly different from 0 at the 0.05 level;  
   the OR of smoke=2 vs 0 is 4.483, p-value=0.0374, it’s significantly different from 0 at the 0.05 level;  
   the OR of smoke=3 vs 0 is 1.493, p-value= 0.1564, it’s not significantly different from 0 at the 0.05 level.
2. After adjusting for age, sex, weight, exercise habits,

the risk of smoke=1(僅嘗試吸過幾次而已) having diabetes was <0.001-fold higher than smoke=0(沒有吸過), and it is significantly different from 0 at the 0.05 level;

the risk of smoke=2(有吸過，從以前到現在沒有吸超過 5 包（ 100 支）菸 ) having diabetes was 4.483-fold higher than smoke=0(沒有吸過), and it is significantly different from 0 at the 0.05 level;

the risk of smoke=3(有吸過，從以前到現在有吸超過 5 包（ 100支）菸) having diabetes was <1.493-fold higher than smoke=0(沒有吸過), and it is not significantly different from 0 at the 0.05 level.

/\*5\*/

/\*糖尿病event\*/

**data** hw3.hw3\_data2\_2; set hw3.hw3\_data2;

if glu >= **126** then diabetes = **1** ;

else diabetes = **0** ;

**run**;

/\*fit logistic model\*/

**proc** **logistic** data=hw3.hw3\_data2\_2;

class smoke(ref='0') sex(ref='0') exercise(ref='0')/ param=ref;

model diabetes(event='1')=smoke age sex weight exercise/ risklimits;

**run**;

* 1. 請問哪些干擾因子可能會影響糖尿病與抽菸之間的關係 (15%)

Age, weigh之係數的p-value分別為<0.0001, 0.0003 ，皆<α (0.05)，拒絕H0 ，與0達統計顯著的差異，因此會對model造成影響，而影響糖尿病與抽菸間的關係。