

## § SAS : homework 9 :

### 一. Univariable analysis of age and BMI

學號：b07401048 醫學五

姓名：賴柏瑞

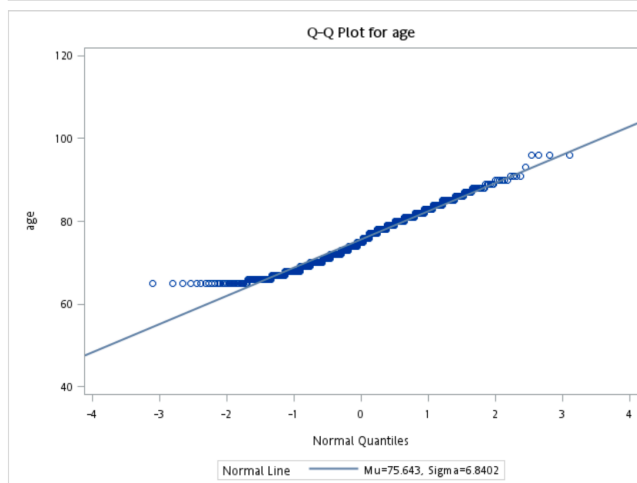
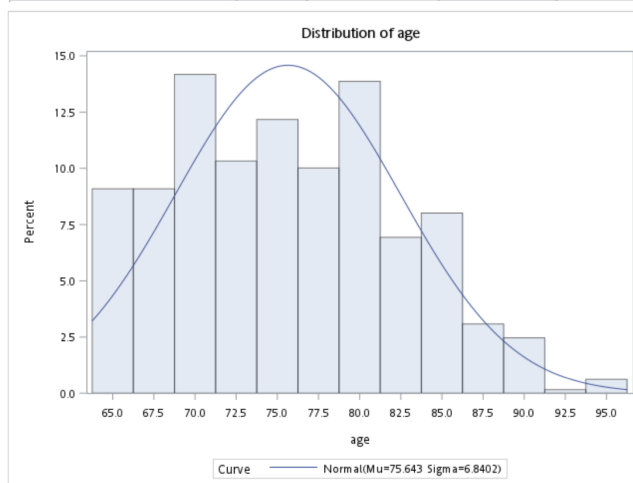
#### (一) Age :

##### 1. Result :

- (1) 以 histogram 大致觀察 age 這個變項的分佈，大致符合常態分佈
- (2) 以 Q-Q plot 大致觀察 age 這個變項分佈，樣本點大致落於斜線附近，僅在 age 較小部分稍微偏離。
- (3) 以 goodness-of-fit test 檢定該變項是否符合常態分佈。檢定結果拒絕虛無假說，表示該變項不符合常態分佈。然而根據課程所述，該檢定過於嚴格，容易受到樣本數影響。

##### 2. Figure :

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.08494189	Pr > D	<0.010
Cramer-von Mises	W-Sq	0.73862550	Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	4.95680165	Pr > A-Sq	<0.005



## (二) BMI：

### 1. Result：

(1) BMI 為自訂變項，利用原始資料中的 height 與 weight 計算而得。公式為

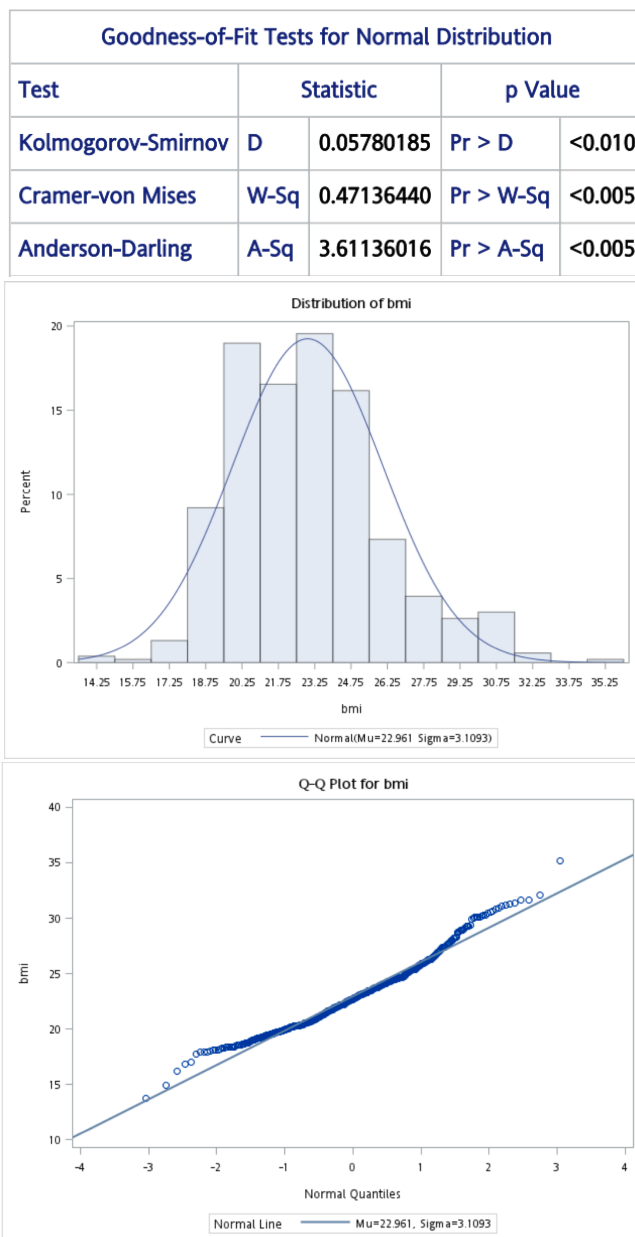
$$\frac{\text{weight (kg)}}{\text{height (m)}^2}。$$

(2) 以 histogram 大致觀察 BMI 這個變項的分佈，大致符合常態分佈

(3) 以 Q-Q plot 大致觀察 BMI 這個變項分佈，樣本點大致落於斜線附近，僅在極端值部分稍微偏離。

(4) 以 goodness-of-fit test 檢定該變項是否符合常態分佈。檢定結果拒絕虛無假說，表示該變項不符合常態分佈。然而根據課程所述，該檢定過於嚴格，容易受到樣本數影響。

### 2. Figure：



### (三) Code :

```
/*homework 9*/
dm "odsresult" clear;
dm "log"clear;
/* import data of sasdataset*/
libname data "\\Mac\Home\Desktop\";
/* q1 */
title "q1";
/* univariable analysis (age, bmi)*/
data ad;
  set data.ad_dataset_new;
  bmi = wt40 / (ht / 100) ** 2;
run;

proc univariate data = ad ;
  var age bmi;
  histogram age bmi / normal;
  qqplot age bmi / normal ( mu = est sigma = est);
run;

/* excluding unsuitable data */
proc print data = ad;
  where bmi > 60;
  var ht wt40 bmi age;
run;

data ad;
  set data.ad_dataset_new;
  bmi = wt40 / (ht / 100) ** 2;
  if bmi < 60;
run;

proc univariate data = ad ;
  var age bmi;
  histogram age bmi / normal;
  qqplot age bmi / normal ( mu = est sigma = est);
run;
```

## 二. E ( education status) → DZ ( AD ), stratified by APOE e4 status

### (一) Stratum specific ORs

#### 1. Result :

(1) Non-APOE e4 carrier stratum specific OR : 7.0911 ( 95% CI : 4.5459 – 11.0613)

(2) APOE e4 carrier stratum specific OR : 5.2275 ( 95% CI : 2.5410 – 10.7542)

#### 2. Figure :

Statistics for Table 1 of edugp by caco

Controlling for apo4car=0

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	7.0911	4.5459	11.0613
Relative Risk (Column 1)	1.9148	1.6124	2.2738
Relative Risk (Column 2)	0.2700	0.1978	0.3686

Statistics for Table 2 of edugp by caco

Controlling for apo4car=1

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	5.2275	2.5410	10.7542
Relative Risk (Column 1)	2.8667	1.7463	4.7058
Relative Risk (Column 2)	0.5484	0.4169	0.7214

#### 3. Descriptions

(1) 將 educational status 分為小於 12 年與大於等於 12 年分為兩組，並定義小於 12 年為 exposure，大於等於 12 年為 non-exposure。

(2) 以是否帶有 APOE e4 (potential confounder) 將所有樣本分層，觀察兩層中 educational status (exposure) 與發生 AD (disease) 之間的關係。

(3) 在有 APOE e4 的層中的 odd ratio 為 5.2275 (95% CI : 2.5410 – 10.7542)。在抽樣與暴露相互獨立的前提下，可由 OR 推論 RR。故在有 APOE e4 層中，education 小於 12 年的組發生 AD 的風險為 education 大於 12 年的 5.2275 倍，且達統計上顯著。

(4) 在沒有 APOE e4 的層中的 odd ratio 為 7.0911 (95% CI : 4.5459 – 11.0613)。在抽樣與暴露相互獨立的前提下，可由 OR 推論 RR。故在沒有 APOE e4 層中，education 小於 12 年的組發生 AD 的風險為 education 大於 12 年的 7.0911 倍，且達統計上顯著。

## (二) Breslow-Day homogeneity test

### 1. Result :

(1) P-value 為 0.4803，無法拒絕虛無假說，兩層 OR 沒有顯著差異。

### 2. Figure :

Breslow-Day Test for Homogeneity of Odds Ratios	
Chi-Square	0.4981
DF	1
Pr > ChiSq	0.4803

### 3. Descriptions

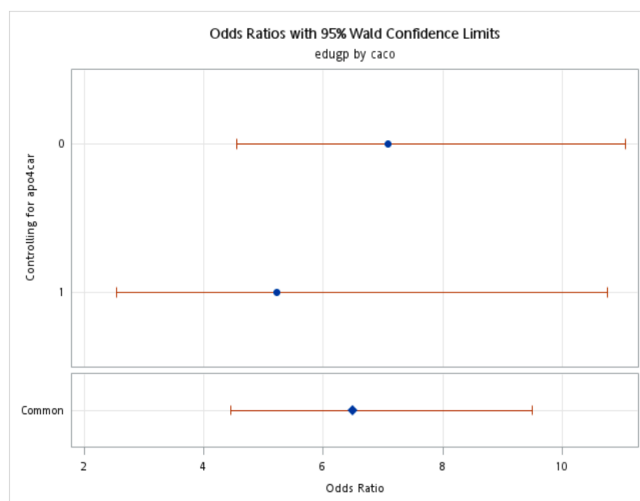
- (1) 若要以 Mantel-Haenszel 計算 pooling 的 OR，需達成兩層 OR 沒有顯著差異的前提。("The effect being estimated is constant across the strata")
- (2) 故以 Breslow-Day test 檢定兩層的同質性。Breslow-Day test 的虛無假說為兩層 OR 相等。
- (3) 檢定結果 p-value 為 0.4803，未達統計上顯著，無法拒絕虛無假設，表示兩層 OR 沒有顯著差異。

## (三) Mantel-Haenszel OR

### 1. Result :

(1) Common OR : 6.5029 (95% CI : 4.4535 – 9.4955)。

### 2. Figure :



Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel-Haenszel	6.5029	4.4535	9.4955
	Logit	6.5201	4.4656	9.5200
Relative Risk (Column 1)	Mantel-Haenszel	2.0438	1.7341	2.4089
	Logit	1.9995	1.6999	2.3520
Relative Risk (Column 2)	Mantel-Haenszel	0.3649	0.2957	0.4504
	Logit	0.4024	0.3276	0.4943

### 3. Descriptions :

- (1) 前題已經以 Breslow-Day test 檢定兩層具有同質性，故本題以 Mantel-Haenszel test 計算 pooling 的 OR。
- (2) Mantel-Haenszel 檢定結果 common odds ratio 為 6.5029 (95% CI : 4.4535 – 9.4955)。信賴區間不包含 1，達統計上顯著。
- (3) 未分層前 crude OR 為 5.8989，分層後 unbiased OR 為 6.5029。1 < crude OR < unbiased OR，可見 bias toward null。
- (4) 此外， $\left| \frac{6.5029 - 5.8989}{5.8989} \right| = 10.23\% > 10\%$ 。根據 10% rule，APOE e4 為一個 confounder。

#### (四) Code :

```
/* q2 educational status (E) --> AD (DZ), stratified by APOE status */
title "q2";
/* grouping educational status , <12 and >=12*/
proc univariate data = ad;
    var eduyr;
    histogram eduyr / normal;
run;

data ad2;
    set ad;
    if eduyr = . then edugp = .;
    else if eduyr < 12 then edugp = 1;
    else edugp = 0;
run;

proc print data = ad2;
    var eduyr edugp;
run;

/* crude OR */
proc freq data = ad2;
    tables edugp * caco / or;
run;

/* stratum specific OR */
proc freq data = ad2;
    tables apo4car * edugp * caco / cmh or ;
run;
```

### 三. E ( education status) → DZ ( AD ), logistic regression

#### (一) Unadjusted logistic regression :

1. Result :

(1) OR : 5.899 (95% CI : 4.164 – 8.354)

2. Figure :

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
edugp 1 vs 0	5.899	4.164	8.357

3. Descriptions :

(1) 本題以 logistic regression 計算結果基本上與上題未分層的 crude OR 是一樣的。

(2) Education 小於 12 年的人發生 AD 的機會機會是 education 大於 12 年的人的 5.899 倍。

#### (二) Adjusted logistic regression :

1. Result :

(1) OR : 6.113 (95% CI : 3.571 – 10.465)

2. Figure :

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
edugp 1 vs 0	6.113	3.571	10.465
age	1.214	1.164	1.266
sex 1 vs 0	0.925	0.535	1.599
bmi	0.922	0.849	1.002
apo4car 1 vs 0	5.439	3.084	9.590

3. Descriptions :

(1) 本題為控制 age, sex, BMI 與 APOE e4 status (四個 potential confounder) 之後進行 logistic regression。

(2) Education 小於 12 年的人發生 AD 的機會機會是 education 大於 12 年的人的 6.113 倍。

(3)  $1 (H_0) < 5.899$  (unadjusted OR)  $< 6.113$  (adjusted OR), 可見 bias toward  $H_0$ 。

(三) Code :

```
/* logistic regression educational status (E) --> AD(DZ) */
title "q3" ;
/* unadjusted */
proc logistic data = ad2;
    class edugp (ref = "0") /param = ref;
    model caco (event = "1") = edugp;
run;

proc logistic data = ad2;
    class edugp (ref = "0") sex (ref = "0") apo4car (ref = "0") / param = ref;
    model caco ( event = "1") = edugp age sex bmi apo4car ;
run;

proc logistic data = ad2;
    model caco(event = "1") = eduyr;
run;
```

四. In a stratified analysis of data arising in a case-control study, which of the following types of strata contribute information to the Mantel-Haenszel odds ratio:

- (a) strata in which all cases and all controls are exposed
- (b) strata in which all cases and all controls are unexposed
- (c) strata in which all cases are exposed and all controls are unexposed
- (d) strata in which there are only controls and no cases