

流行病學實習課 SAS 作業一

繳交日期: 2020.04.14

公衛二
梁嫚芳
B07801003

- 1 某社區隨機抽取 50 人的健康調查資料，該資料分別被存於 HW1_basic 和 HW1_health 檔案中。請利用 SAS 程式語言將兩個檔案 HW1_basic 和 HW1_health 合併後存於一個永久資料庫中。(8%)

建立永久資料庫“A”，將HW1_basic和hw1_health匯入，分別命名為hw1_basic與hw1_health，將兩者水平合併並存於永久資料庫“A”中，命名為hw1_data1(如圖)

ID	sex	age	drink	smoke	exercise	height	weight	waist	sbp	dbp	htn	dm	hi	hba
1	1	74	0	0	0	146.5	46.5	84	125	85	0	0	0	0
2	1	38	1	3	0	170.6	84	92	118	79	0	0	1	8
3	0	30	1	0	0	167	52	63.5	92	60	0	0	0	0
4	0	38	0	0	0	161.4	66.5	79.7	107	66	0	0	0	0
5	0	81	0	0	1	139.5	52.5	68.8	115	77	0	0	0	0
6	0	54	1	0	0	163.5	55	71	80	59	0	0	0	0
7	1	52	1	3	1	176.5	71.5	87	112	74	0	0	1	0
8	1	58	0	2	1	168	88.5	94.5	147	87	1	0	0	0
9	0	58	0	0	1	156.3	55	77	110	67	0	0	1	0
10	1	51	0	3	1	166	66.5	85	112	78	0	0	0	0
11	1	33	0	3	0	176.8	82	86	137	92	8	8	8	8
12	0	37	1	1	0	163.6	53	69	113	77	0	0	0	0
13	0	36	1	0	0	163.1	47	59	91	52	0	0	8	8
14	1	39	0	3	0	174	70	79	124	90	0	0	0	1
15	0	41	0	0	0	158	49.5	61.5	110	78	0	0	0	0
16	0	45	0	0	0	154.5	45	59.3	105	71	0	0	0	0
17	0	42	0	0	1	154	49.5	66	100	70	0	0	0	0
18	0	74	0	3	0	148.6	60	87.7	146	72	1	0	1	1
19	0	53	0	0	0	146.1	48.3	74.5	107	79	0	8	8	8
20	0	27	1	0	0	157.5	68	78	111	67	0	0	8	0
21	0	53	0	0	0	162.3	71.5	79.5	148	94	8	0	8	0
22	1	23	0	0	0	181.7	116	110	132	83	0	0	1	1
23	1	53	0	0	0	165	77.5	88	121	84	0	0	0	0
24	0	79	0	0	0	146.7	56	86.5	182	92	1	1	0	0
25	1	42	0	0	1	173	69	82.5	122	79	0	0	0	0
26	0	69	1	0	0	146.7	55	77.4	125	71	1	0	1	0
27	1	36	1	3	0	169.5	94.5	106	138	93	1	0	1	0
28	1	72	0	0	1	164.7	71	87.5	113	71	1	1	0	1
29	1	28	0	3	0	184.5	58.5	71.4	123	90	0	0	0	0
30	1	71	0	0	1	169.5	68	87.6	136	83	1	1	0	1

```

/*1*/
/*建立永久資料庫'a'*/
libname a 'D:\SAS';
/*SORT*/
PROC SORT data=a.Hw1_basic;
  by ID;
RUN;
PROC SORT data=a.Hw1_health;
  by ID;
RUN;
/*合併兩組資料*/
DATA a.hw1_data1;
merge a.Hw1_basic a.Hw1_health;
by ID;
RUN;

```

- 2 為方便檢視某社區居民之健康狀況，將每位居民的 BMI、血壓依照衛服部之標準分組檢視。BMI 分為四組：過輕(<18.5)、正常($18.5 \leq \text{BMI} < 24$)、過重($24 \leq \text{BMI} < 27$)、肥胖(≥ 27)。血壓分為正常($\text{SBP} < 140 \text{mmHg}$ 和 $\text{DBP} < 90 \text{mmHg}$)與異常兩組。(20%)

建立新資料檔“Hw1_data2”，計算bmi(體重除以身高平方)後建立新變項“BMI”，將BMI分為四組(過輕:0、正常:1、過重:2、肥胖:3)建立為新變項“bmi_g”，將血壓分為兩組(正常:0、異常:1)建立為新變項“pp”。(如圖)

	dm	hl	hs	stroke	cvd	chd	glu	ua	creat	chol	trig	hdl	ldl	BMI	bmi_g	pp
1	0	0	0	1	0	0	78	4.7	0.8	202	128	44.1	140	21.66594635	1	0
2	0	1	8	0	0	0	80	8.1	0.9	215	515	47	100	28.86165509	3	0
3	0	0	0	0	0	0	80	4.4	0.9	167	65	54.5	94	18.64534404	1	0
4	0	0	0	0	0	0	100	4.5	0.8	162	48	63.3	88	25.52787029	2	0
5	0	0	0	0	0	0	90	4.7	1.1	146	102	51	78	20.89780971	1	0
6	0	0	0	0	0	0	77	5.3	0.8	260	83	88.8	119	20.57439983	1	0
7	0	1	0	0	0	0	86	7.6	1.2	384	85	66.2	258	22.55179320	1	0
8	0	0	0	0	0	0	98	7.1	1.2	192	115	45.9	122	31.35629251	3	1
9	0	1	0	0	0	0	82	4.1	0.7	203	95	59.4	119	22.51358894	1	0
10	0	0	0	0	0	0	90	5.2	1	188	149	40.1	142	24.13267527	2	0
11	8	8	8	0	0	8	89	6.2	1.3	125	85	45.9	57	26.23308286	2	1
12	0	0	0	0	0	0	94	3.8	0.9	77	35	46.7	16	19.80200979	1	0
13	0	8	8	0	0	0	87	5.5	0.9	193	68	68.8	83	17.66810354	0	0
14	0	0	1	0	0	0	79	4.3	1	169	82	53.5	113	23.12062359	1	1
15	0	0	0	0	0	0	81	3.9	0.8	141	58	65.6	61	19.82855311	1	0
16	0	0	0	0	0	0	78	4.4	0.8	186	61	67	101	18.85191818	1	0
17	0	0	0	0	1	0	78	4.5	0.8	173	71	57.7	89	20.87198515	1	0
18	0	1	1	0	0	1	81	3.5	0.7	155	112	44.5	101	27.17150108	3	1
19	8	8	8	0	8	0	83	3.8	0.6	171	226	45.3	86	22.62802755	1	0
20	0	8	0	0	0	0	84	5.3	0.7	255	138	51.3	181	27.41244644	3	0
21	0	8	0	0	0	0	109	4.6	0.7	184	128	60	112	27.14369721	3	1
22	0	1	1	0	0	0	71	6.1	0.9	190	173	37.1	130	35.13566151	3	0

```

/*2*/
DATA a.hw1_data2; set a.hw1_data1;
/*bmi分組*/
BMI=weight/((height/100)**2);
if BMI<18.5 then bmi_g='0';
else if BMI<24 then bmi_g='1';
else if BMI<27 then bmi_g='2';
else bmi_g='3';
/*血壓分組*/
if sbp<140 and dbp<90 then pp='0';
else pp='1';

RUN;

```

- 3 推行預防醫學，社區健康中心之公衛護理師需透過該社區居民的年齡、體重、BMI 和腰圍來評估居民的肥胖狀況，請協助計算上述變項的平均值、標準差及變異係數。(12%)

Output:

The MEANS Procedure

Variable	Mean	Std Dev	Coeff of Variation
age	51.0200000	15.7499465	30.8701422
weight	62.1360000	15.0088513	24.1548398
BMI	23.5599444	4.4587806	18.9252596
waist	78.5460000	11.9734436	15.2438617

	平均值	標準差	變異係數
年齡	51.02	15.7499465	30.8701422
體重	62.136	15.0088513	24.1548398
BMI	23.5599444	4.4587806	18.9252596
腰圍	78.546	11.9734436	15.2438617

```

/*3*/
/*計算平均值 標準差 變異數*/
PROC MEANS DATA=a.hw1_data2 mean std cv;
VAR age weight BMI waist;
RUN;

```

- 4 接著，預計推動健康促進介入計畫，公衛護理師調查該社區男性與女性的糖尿病病史分布情形，我們想知道性別與患有糖尿病是否有相關。請檢定虛無假設並說明你的結論為何？(20%)

H0: 性別與是否患有糖尿病無相關
H1: 性別與是否患有糖尿病有相關



```

/*4*/
/*刪除不知道是否患有糖尿病的資料*/
DATA a.Hw1_data4; set a.Hw1_data2;
if dm=8 then delete;
run;
/*卡方檢定*/
PROC FREQ DATA=a.Hw1_data4;
TABLE sex*dm /expected chisq;
RUN;

```

- 5 根據上述的調查結果，社區健康中心想進一步確認男性和女性血糖之平均值、標準差、變異係數。請計算上述數值，並確認平均血糖值在不同性別間是否一樣呢？(30%)

Output:

Analysis Variable : glu				
sex	N Obs	Mean	Std Dev	Coeff of Variation
0	28	88.7500000	19.7945466	22.3037144
1	22	86.5454545	10.4367814	12.0593063

	平均值	標準差	變異係數
女血糖	88.7500000	19.7945466	22.3037144
男血糖	86.5454545	10.4367814	12.0593063

平均血糖值就上表數值而言在不同性別間不相同。

但當進行兩組獨立樣本t-test，H0:平均血糖值在不同性別間相等，根據以下報表 p-value(變異數不均等 → Welch-Satterthwaite t-test)=0.6151> $\alpha(0.05)$ ，因此不拒絕H0，平均血糖值在不同性別間無統計顯著的差異。

變異: glu							
sex	方法	N	平均值	標準差	標準誤差	最小值	最大值
0		28	88.7500	19.7945	3.7408	73.0000	180.0
1		22	86.5455	10.4368	2.2251	71.0000	107.0
差異 (1-2)	集區		2.2045	16.3724	4.6645		
差異 (1-2)	Satterthwaite		2.2045		4.3526		

方法	變異數	DF	t 值	Pr > t
集區	均等	48	0.47	0.6386
Satterthwaite	不均等	42.625	0.51	0.6151

變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	27	21	3.60	0.0038

/*5*/

```
/*以性別分組計算平均值 標準差 變異數*/
```

```
PROC MEANS DATA=a.Hw1_data2 mean std cv;
```

```
CLASS sex;
```

```
VAR glu ;
```

```
RUN;
```

```
/*兩組獨立樣本t-test*/
```

```
PROC TTEST DATA=a.Hw1_data2;
```

```
CLASS sex;
```

```
VAR glu;
```

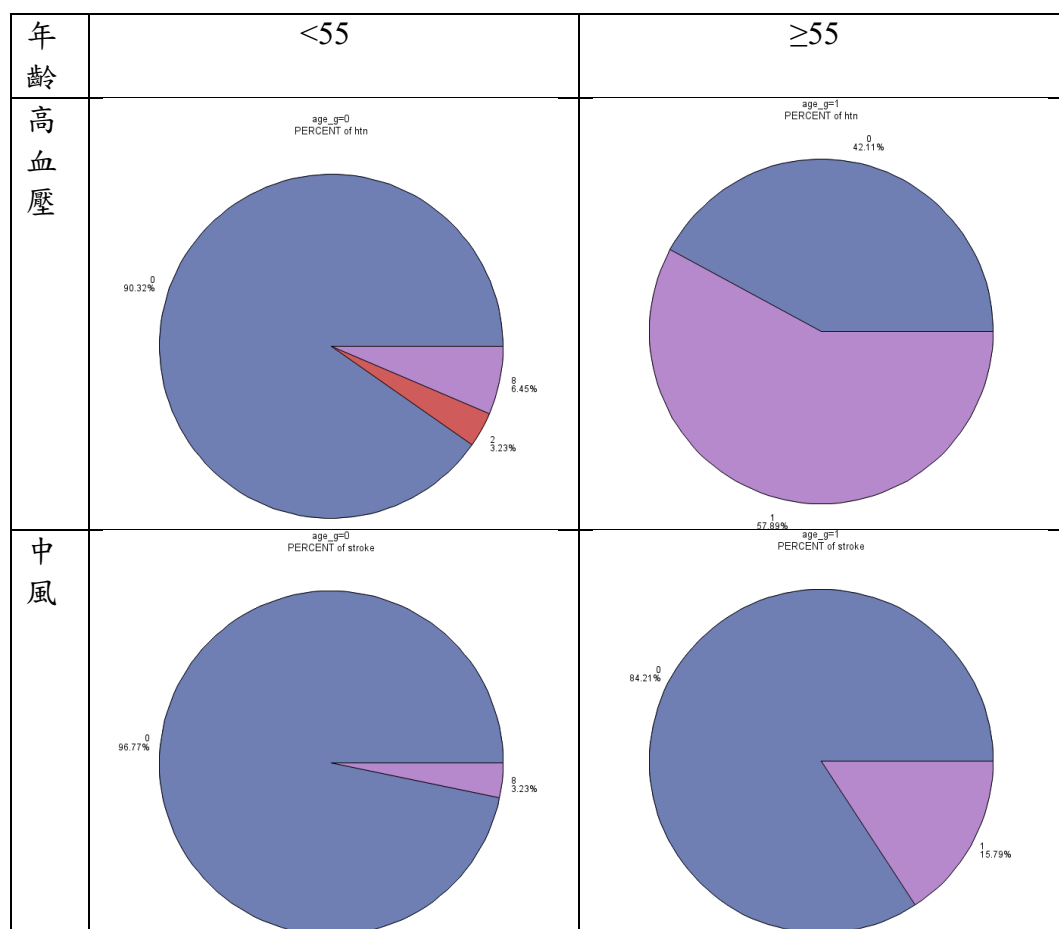
```
RUN;
```

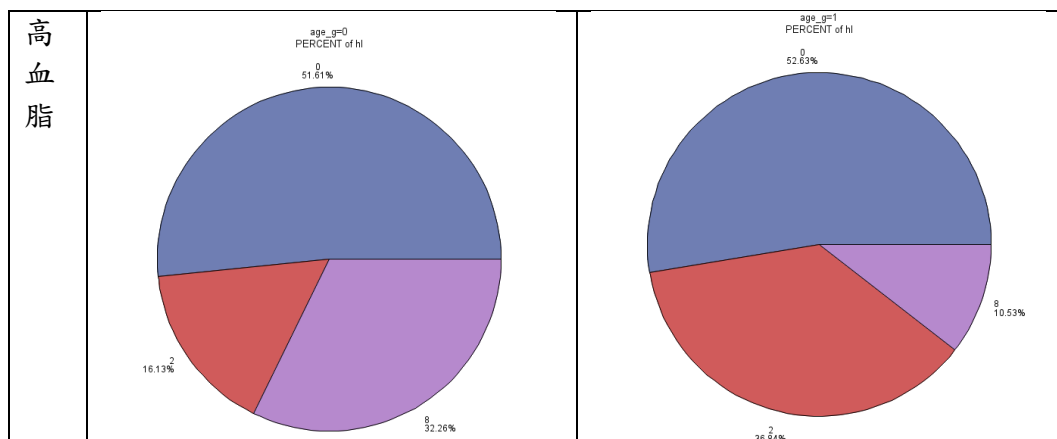
- 6 社區健康中心乃重視除糖尿病以外的其他慢性疾病，請協助計算高血壓、中風、高血脂在年齡小於 55 歲和大於等於 55 歲之盛行率為何？以一種合適的圖形表示，並列出盛行率。(10%)

Output:

Frequency Row Pct	Table of age_g by htn					Frequency Row Pct	Table of age_g by stroke					Frequency Row Pct	Table of age_g by hl				
	htn						stroke						hl				
	age_g	0	1	8	Total		age_g	0	1	8	Total		age_g	0	1	8	Total
	0	28	1	2	31		0	30	0	1	31		0	16	5	10	31
		90.32	3.23	6.45				96.77	0.00	3.23				51.61	16.13	32.26	
	1	8	11	0	19		1	16	3	0	19		1	10	7	2	19
		42.11	57.89	0.00				84.21	15.79	0.00				52.63	36.84	10.53	
	Total	36	12	2	50		Total	46	3	1	50		Total	26	12	12	50

	年齡	高血壓	中風	高血脂
盛行率(%)	<55	3.23	0	16.13
	≥55	57.89	15.79	36.84





```

/*6*/
/*將年齡分組*/
DATA a.Hw1_data6; set a.Hw1_data2;
if age<55 then age_g=0;
else age_g=1;
RUN;
/*計算盛行率*/
PROC FREQ DATA=a.Hw1_data6;
TABLE age_g*htn age_g*stroke age_g*hl /nopercent nocol;
RUN;
/*將資料以年齡組別排序*/
PROC SORT DATA=a.Hw1_data6;
by age_g;
RUN;
/*圓餅圖*/
PROC GCHART DATA=a.Hw1_data6;
By age_g;
PIE htn stroke hl/type = percent ;
RUN;
QUIT;

```