

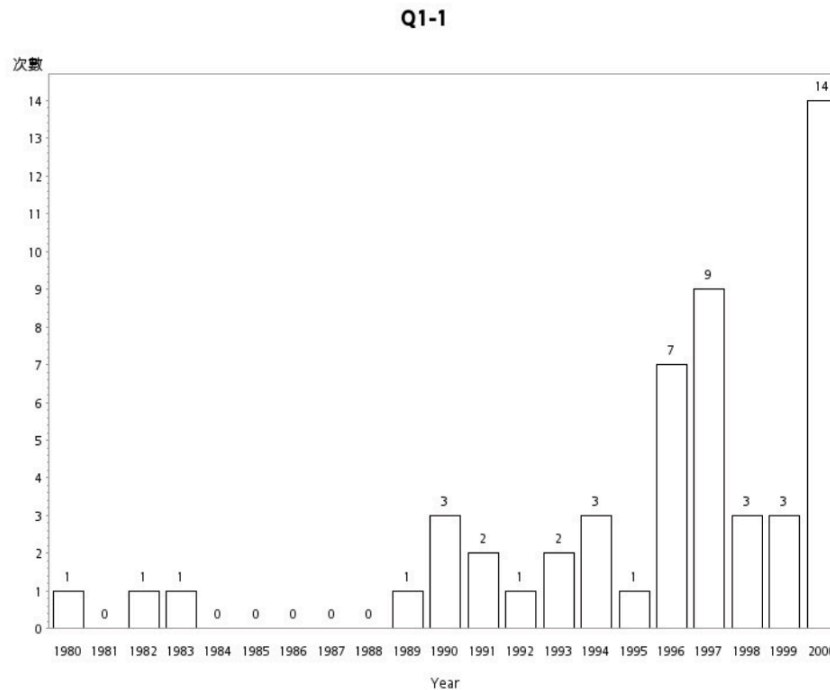
## § 流行病學原理：資料分析 practice4

流行病學資料分析 –  
SAS  
Practice 4  
學號：b07401048  
系級：醫學五

### 一. 請進行下列與時間相關的描述流行病學分析

(一)流行曲線圖（所有個案）：橫坐標為年，縱座標為個案數，以長條圖呈現。

#### 1. figures：



#### 2. Description：

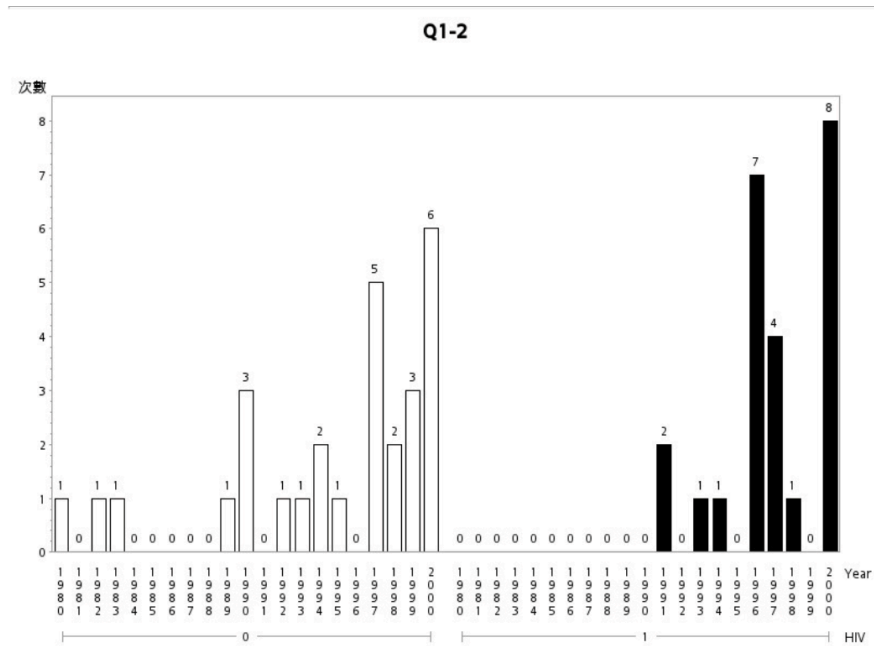
- (1) 由長條圖觀察，1990 年以前，每年個案數維持 1 例以下；1996 年以前，每年個案數維持 3 例以下。
- (2) 1996 年以後，每年個案數皆在 3 例或以上，以 2000 年的 14 例為最高。然而 1996 年後是否爆發流行尚待驗證，見 Q1-3。

#### 3. Code：

```
/* HW4 descriptive epidemiology */  
proc import datafile = "C:\Users\User\Desktop\Dataset 2.xlsx"  
    dbms = xlsx  
    out = work.dataset2  
    replace;  
run;  
  
/*Q1-1 */  
title "Q1-1";  
proc gchart data = dataset2;  
    vbar year / midpoints = 1980 to 2000 by 1 freq;  
run;
```

(二)流行曲線圖，HIV 個案以黑色長條圖表示，非 HIV 個案以白色表示？

1. figures：



2. Description：

- (1) 上圖左白色長條圖為 non-HIV 個案的流行曲線圖，上圖右為 HIV 個案的流行曲線圖。
- (2) 觀察上圖可知，自 1991 年開始才有觀察到有隱球菌菌血症的 HIV 個案；1991 年以前共觀察 7 名個案，皆無記載為 HIV 感染者。

3. Code：

```
/* Q1-2 */
title "Q1-2";

proc gchart data = dataset2;
    vbar year / midpoints = 1980 to 2000 by 1 freq group = hiv patternid
= group;
    pattern1 color = white;
    pattern2 color = black;
run;
```

(三)收案期間的最後 6 年（1996-2001）是否出現爆發流行？如何認定出現爆發流行？

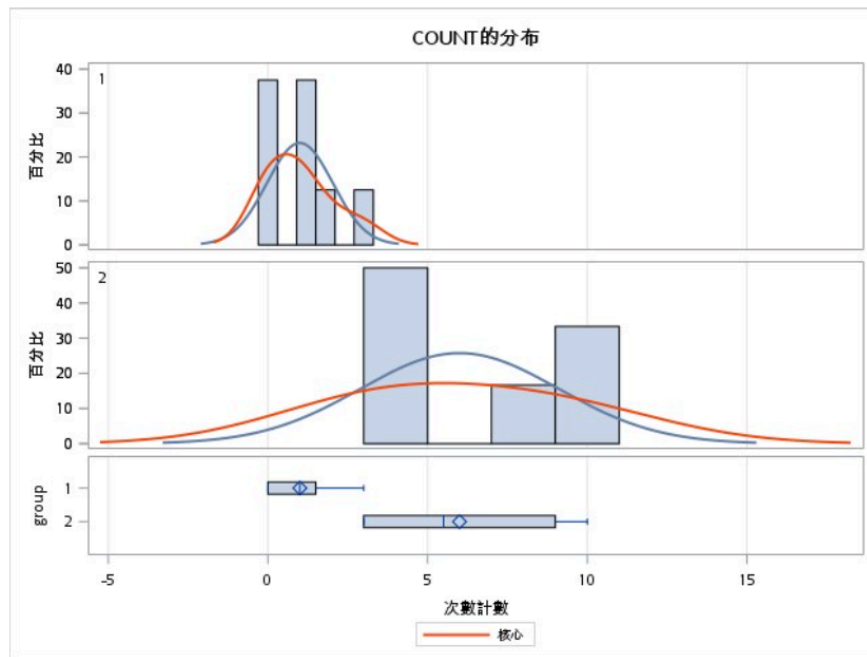
1. Result :

- (1) 認定方法：以 1996 年為切點，判斷 1996 年以前（1980-1995）及 1996 年以後（1996-2001）兩組之「每年個案出現次數的平均值」是否相等。
- (2) 結果：以兩組樣本 t 檢定判斷，1996 年以後「每年個案出現次數平均值」顯著高於 1996 年以前，足以認定 1996 年後確實出現爆發流行。

2. Description :

- (1) 將資料分為 1996 年以前（1980-1995）及 1996 年以後（1996-2001）兩組，並計算兩組「個案每年出現次數的平均值」。
- (2) 以「每年個案出現次數的平均值是否相等」作為認定 1996 年以後是否出現爆發流行的依據。故以「兩組個案每年出現次數的平均值」相減為 0 為虛無假設作 two-sample t-test。
- (3) 首先 folded F method 判斷兩組變異數是否相等，以兩組變異數相等為虛無假說。檢定結果 p-value 為 0.0008，拒絕虛無假說，兩組變異數不相等。
- (4) 之後以假定兩組變異數不相等的 Satterthwaite method 判斷兩組平均值是否相等，以兩組平均數相等為虛無假說。檢定結果 p-value 為 0.01，拒絕虛無假說，兩組變異數不相等。
- (5) 爆發流行（epidemic）之定義為“occurrence of more cases of disease than normally expected in given area over a particular period of time in a population”，即當疾病在特定時空的新個案數超過正常情況的期望值，則可稱之為爆發流行。檢定結果顯示 1996 年以後「個案每年出現平均次數」顯著高於 1996 年以前，可視為 1996 年以後新病例數超過正常情況的期望值，故可推論 1996 年以後有爆發流行。

### 3. Figures :



變異數相等性				
方法	分子自由度	分母自由度	F 值	Pr > F
Folded F	5	15	9.00	0.0008

方法	變異數	DF	t 值	Pr >  t
集區	均等	20	-5.84	<.0001
Satterthwaite	不均等	5.4222	-3.87	0.0100

### 4. Code :

```

/* Q1-3 */
title "Q1-3";
/* obtaining freq of all year */
data datayear;
  do year = 1980 to 2001;
    output;
  end;
run;

data allyear;
  set dataset2(in = a keep = year) datayear(in= b);
  if a then weight = 1;
  if b then weight = 0;
run;

proc freq data = allyear;
  tables year / nopercnt nocum out = all_year(drop = percent);
  weight weight / zeros;
run;

/* grouping and result*/
data year1996;
  set all_year;
  if year < 1996 then group = 1;
  else group = 2;
run;

```

```
proc ttest data = year1996 h0 = 0 alpha = 0.05;
  class group;
  var count;
run;
```

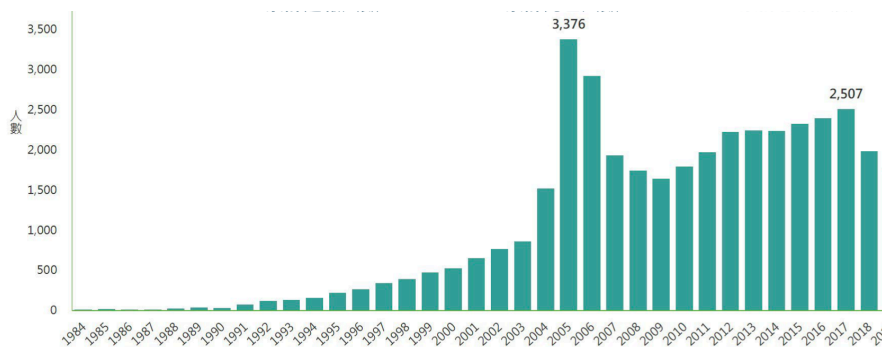
(四)請提出合理假說，解釋收案期間的最後 6 年（1996-2001）隱球菌菌血症個案數增加的現象。

1. 假說一：全民健保實施。

- (1) 台灣全民健保自民國 84（1995 年）實施，時間點與本題呈現的 1996 年相符。健保的實施可能導致民眾就醫意願及醫療行為的改變，以下分述。
- (2) 全民健保的實施預期可使民眾就醫所需負擔的醫藥費下降，而可能使民眾求醫的意願相對健保未實施前為高。因此，台大醫院能在健保實施後求診人次上升，也因此可記錄到更多隱球菌菌血症的病人個案數，進而反映在本題資料中。
- (3) 此外，健保實施後提出相對應的給付規則，可能影響特定醫療行為及醫療常規。台大醫院的醫師可能在健保實施後，更傾向對門診或急診病人進行抽血檢驗，並在遇到疑似症狀時加入隱球菌作為檢驗項目。如上醫療行為的改變會使隱球菌菌血症的檢出率上升，進而反映在本題資料中。

2. 假說二：愛滋病流行

- (1) 隱球菌為常見的環境微生物<sup>1</sup>，一般免疫正常的人不易感染隱球菌，更不常見造成隱球菌菌血症等嚴重症狀。然而，當人處於免疫低下或免疫不全的狀態時，隱球菌即可能感染人體並導致症狀。
- (2) 倘若台灣在 1996 年前後曾有愛滋病流行，且在當時缺乏針對 HIV 感染者完善的藥物治療策略或藥物補助政策，可能導致當時 HIV 感染盛行率上升，且由於沒有完善的藥物治療，多數 HIV 感染者處於免疫低下的狀態。
- (3) 若以上情形發生，可能反映在隱球菌菌血症的感染人數。
- (4) 查閱文獻<sup>2</sup>未發現 1996 年前後有 HIV 之爆發流行。



<sup>1</sup> Robbins and Cotran Pathologic Basis of Disease 10e

<sup>2</sup> 衛福部疾病管制署愛滋病統計資料 <https://www.cdc.gov.tw/Category/Page/rCV9N1rGUz9wNr8lggsh2Q>

## 二. 人口學特徵相關的描述流行病學分析：

(一)性別與年齡分佈：以 2 by N 表格呈現，縱欄為性別，橫列為不同年齡層，年齡層以 10 年為單位。

### 1. Result：

Q2-2

FREQ 程序

次數	Gender * agegroup的表格									
	Gender(Gender)	agegroup								
		10-19 y/o	20-29 y/o	30-39 y/o	40-49 y/o	50-59 y/o	70-79 y/o	80-89 y/o	總計	
		male	1	3	14	11	8	2	2	41
		female	0	1	2	0	2	5	1	11
	總計	1	4	16	11	10	7	3	52	

### 2. Descriptions：

- (1) 觀察上圖，總共觀察 52 名個案，不論年齡，男性 41 人，女性 11 人。
- (2) 不論性別，以 30-39 歲的區間最多人（16 人），40-49 歲次之。
- (3) 若考慮性別及年齡，以 30-39 歲男性最多（14 人），40-49 歲男性次之（11 人）。

### 3. Code：

```
/*Q2-1 */
title "Q2-1";
/* grouping */
data agegroup;
set dataset2;
if 10 <= age <= 19 then agegroup = 1;
else if age <= 29 then agegroup = 2;
else if age <= 39 then agegroup = 3;
else if age <= 49 then agegroup = 4;
else if age <= 59 then agegroup = 5;
else if age <= 69 then agegroup = 6;
else if age <= 79 then agegroup = 7;
else if age <= 89 then agegroup = 8;
else if age <= 99 then agegroup = 9;
else if age > 100 then agegroup = 10;

if imt = 1 then immunotherapy = 1;
else if imt = 0 then immunotherapy = 0;
run;

proc print data = agegroup(obs = 30);
var age agegroup;
run;

/* labeling */
proc format;
value agelabel
1 = "10-19 y/o"
```

```

2 = "20-29 y/o"
3 = "30-39 y/o"
4 = "40-49 y/o"
5 = "50-59 y/o"
6 = "60-69 y/o"
7 = "70-79 y/o"
8 = "80-89 y/o"
9 = "90-99 y/o"
10 = ">100 y/o";

value genderlabel
0 = "male"
1 = "female";

value imtlabel
0 = "without immunotherapy"
1 = "with immunotherapy";

value cirrhosislabel
0 = "without cirrhosis"
1 = "with cirrhosis";
run;

/* output */
proc freq data = agegroup;
tables gender * agegroup /norow nocol nopercents;
format gender genderlabel.
agegroup agelabel.;
run;

```

(二)性別與年齡分佈：分為 HIV 與非 HIV 兩組，以 2 by N 表格呈現，縱欄為性別，橫列為不同年齡層，年齡層以 10 年為單位。

1. Result :
2. Descriptions :

FREQ 程序

次數	Gender * agegroup的表格 1							
	HIV=0的控制							
	agegroup							
Gender(Gender)	10-19 y/o	20-29 y/o	30-39 y/o	40-49 y/o	50-59 y/o	70-79 y/o	80-89 y/o	總計
male	1	0	0	7	5	2	2	17
female	0	1	2	0	2	5	1	11
總計	1	1	2	7	7	7	3	28

次數	Gender * agegroup的表格 2							
	HIV=1的控制							
	agegroup							
Gender(Gender)	10-19 y/o	20-29 y/o	30-39 y/o	40-49 y/o	50-59 y/o	70-79 y/o	80-89 y/o	總計
male	0	3	14	4	3	0	0	24
female	0	0	0	0	0	0	0	0
總計	0	3	14	4	3	0	0	24

- (1) 觀察上圖，上圖上半為 non-HIV 組，下半為 HIV 組。Non-HIV 組共 28 名個案，HIV 組則有 24 名。
- (2) 首先觀察年齡，non-HIV 組個案集中在 40-49, 50-59, 70-79 歲三個區間（各有 7 名個案）；HIV 組則集中在 30-39 歲的區間（共有 14 名個案）。HIV 組感染者有比 non-HIV 組更為年輕的趨勢。
- (3) 接著觀察性別，HIV 組全部的個案皆為男性（24 名個案），non-HIV 組則較為平均，男性略多於女性（17 名男性，11 名女性）。

3. Code :

```
/*Q2-2 */
title"Q2-2";

proc freq data = agegroup;
  tables hiv*gender*agegroup / norow nocol nopercnt;
  format gender genderlabel.
         agegroup agelabel.;
run;
```



(三)比較 HIV 與非 HIV 組病人之 <sup>1</sup> 性別、<sup>2</sup> 年齡分佈、<sup>3</sup> 是否接受免疫治療以及 <sup>4</sup> 是否罹患肝硬化的比例是否有顯著差異。

## 1. HIV vs. gender

### (1) Figures :

## Q2-3

### FREQ 程序

次數  
預期  
列百分比

HIV * Gender的表格				
HIV(HIV)	Gender(Gender)			總計
	male	female		
0	17	11		28
	22.077	5.9231		
	60.71	39.29		
1	24	0		24
	18.923	5.0769		
	100.00	0.00		
總計	41	11		52

### HIV \* Gender 之表格的統計值

統計值	DF	值	機率
卡方	1	11.9582	0.0005
極度比卡方	1	16.1422	<.0001
連續性調整卡方	1	9.7188	0.0018
Mantel-Haenszel 卡方	1	11.7282	0.0006
Phi 係數		-0.4795	
列聯係數		0.4324	
Cramer V		-0.4795	

### (2) Result :

- 以卡方檢定「是否為 HIV 感染者」及「男性或女性」兩類別變項是否獨立，即男女比例在 HIV 或 non-HIV 兩組是否有顯著差異。
- 檢定結果為顯著，男女比例在 HIV 或 non-HIV 兩組有顯著差異。

### (3) Descriptions :

- 「是否為 HIV 感染者」及「男性或女性」兩者為類別變項，故以卡方檢定判斷兩者是否獨立。
- 虛無假說為兩變項獨立，男女比例在 HIV 組跟 non-HIV 組沒有顯著差異。
- 統計結果 p-value 為 0.0005，拒絕虛無假說，「是否為 HIV 感染者」及「男性或女性」兩者非獨立變項，即男女比例在 HIV 或 non-HIV 兩組具有顯著差異。
- HIV 組男性的比例為 100% (24/24)，non-HIV 組男性比例為 60.71% (17/28)，HIV 組男性比例顯著高於 non-HIV 組。

## 2. HIV vs. age :

(1) Figure :

樣本大小 = 52

次數 預期 列百分比	HIV * agegroup 的表格							
	agegroup							總計
	HIV(HIV)	10-19 y/o	20-29 y/o	30-39 y/o	40-49 y/o	50-59 y/o	70-79 y/o	
0	1	1	2	7	7	7	3	28
	0.5385	2.1538	8.6154	5.9231	5.3846	3.7692	1.6154	
	3.57	3.57	7.14	25.00	25.00	25.00	10.71	
1	0	3	14	4	3	0	0	24
	0.4615	1.8462	7.3846	5.0769	4.6154	3.2308	1.3846	
	0.00	12.50	58.33	16.67	12.50	0.00	0.00	
總計		1	4	16	11	10	7	52

HIV \* agegroup 之表格的統計值

統計值	DF	值	機率
卡方	6	23.2481	0.0007
擬度比卡方	6	28.5861	<.0001
Mantel-Haenszel 卡方	1	14.9391	0.0001
Phi 係數		0.6686	
列聯係數		0.5558	
Cramer V		0.6686	

Fisher 精準檢定	
表格機率 (P)	<.0001
Pr <= P	0.0001

WARNING: 64% of the cells have expected counts less than 5. Chi-Square may not be a valid test.

(2) Result :

- 以 Fisher exact test 檢定「是否為 HIV 感染者」與「屬於哪個年齡區間」兩類別變項是否為獨立變項。
- 檢定結果 p-value 為 0.0001，拒絕虛無假設，兩者非獨立變項。

(3) Descriptions :

- 檢定「是否為 HIV 感染者」與「屬於哪個年齡區間」兩類別變項是否為獨立變項。由於以卡方檢定將有 64% 的細格的期望值小於 5，故不適用卡方檢定的結果，以 Fisher exact test 為準。
- 虛無假設為兩變項為獨立變項，年齡區間的比例在 HIV 組及 non-HIV 組相同。
- Fisher exact test 檢定結果 p-value 為 0.0001，拒絕虛無假設，年齡區間的比例在 HIV 組及 non-HIV 組不相同。

### 3. HIV vs immunosuppressant :

#### (1) Figures :

樣本大小 = 52

次數 預期 列百分比	HIV * lmt的表格			
	HIV(HIV)	lmt(lmt)		總計
		without immunotherapy	with immunotherapy	
0		15	13	28
		20.462	7.5385	
		53.57	46.43	
1		23	1	24
		17.538	6.4615	
		95.83	4.17	
總計		38	14	52

HIV \* lmt 之表格的統計值

統計值	DF	值	機率
卡方	1	11.7317	0.0006
概度比卡方	1	13.5921	0.0002
連續性調整卡方	1	9.6819	0.0019
Mantel-Haenszel 卡方	1	11.5060	0.0007
Phi 係數		-0.4750	
列聯係數		0.4290	
Cramer V		-0.4750	

#### (2) Result :

- 以卡方檢定判斷「是否為 HIV 感染者」與「是否接受免疫抑制劑治療」兩類別變項是否為獨立變項。
- 檢定結果 p-value 為 0.0006，拒絕虛無假設，兩者非獨立變項。

#### (3) Descriptions :

- 以卡方檢定判斷「是否為 HIV 感染者」與「是否接受免疫抑制治療」兩類別變項是否為獨立變項。虛無假設為兩變項為獨立變項，接受免疫抑制治療的比例在 HIV 組及 non-HIV 組相同。
- 卡方檢定結果 p-value 為 0.0006，拒絕虛無假設，接受免疫抑制治療的比例在 HIV 組及 non-HIV 組不相同。
- Non-HIV 組接受免疫抑制治療的比例為 46.43% (13/28)，HIV 組接受免疫抑制治療的比例為 4.17% (1/24)，non-HIV 組顯著高於 HIV 組。

### 4. HIV vs liver cirrhosis :

#### (1) Figures :

樣本大小 = 52

次數 預期 列百分比	HIV * Cirrhosis的表格			
	HIV(HIV)	Cirrhosis(Cirrhosis)		總計
		without cirrhosis	with cirrhosis	
0		17	11	28
		22.077	5.9231	
		60.71	39.29	
1		24	0	24
		18.923	5.0769	
		100.00	0.00	
總計		41	11	52

HIV \* Cirrhosis 之表格的統計值

統計值	DF	值	機率
卡方	1	11.9582	0.0005
概度比卡方	1	16.1422	<.0001
連續性調整卡方	1	9.7188	0.0018
Mantel-Haenszel 卡方	1	11.7282	0.0006
Phi 係數		-0.4795	
列聯係數		0.4324	
Cramer V		-0.4795	

#### (2) Result :

(a) 以卡方檢定判斷「是否為 HIV 感染者」與「是否有肝硬化」兩類別變項是否為獨立變項。

(b) 檢定結果 p-value 為 0.0005，拒絕虛無假設，兩者非獨立變項。

(3) Descriptions :

(a) 以卡方檢定判斷「是否為 HIV 感染者」與「是否有肝硬化」兩類別變項是否為獨立變項。虛無假設為兩變項為獨立變項，有肝硬化的個案的比例在 HIV 組及 non-HIV 組相同。

(b) 卡方檢定結果 p-value 為 0.0005，拒絕虛無假設，有肝硬化的比例在 HIV 組及 non-HIV 組不相同。

(c) Non-HIV 組有肝硬化的比例為 39.29% (11/28)，HIV 組有肝硬化的比例為 0% (0/24)，non-HIV 組顯著高於 HIV 組。

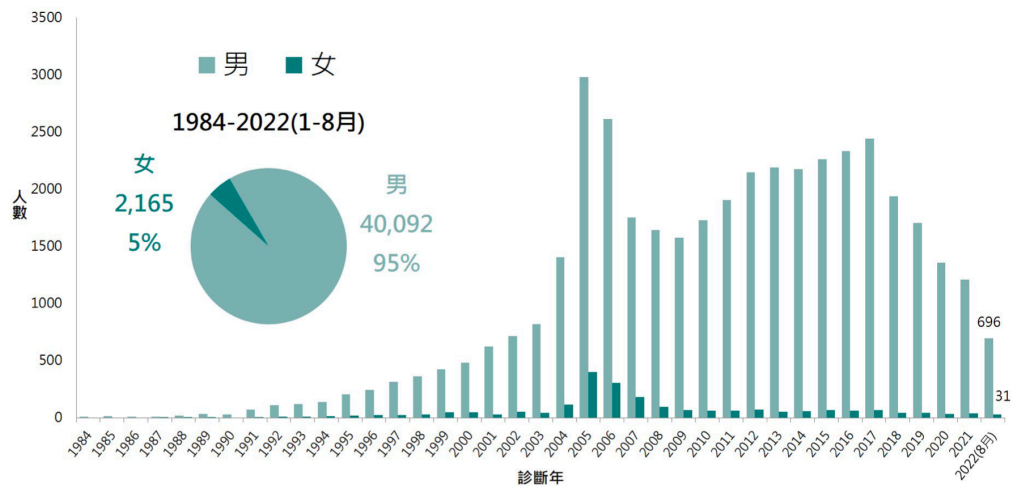
(4) Code :

```
/* Q2-3 */
title "Q2-3";
/*chi-square */
proc freq data = agegroup;
    tables hiv*( gender agegroup imt cirrhosis) / nopercnt nocol
chisq expected fisher;
    format gender genderlabel.
           agegroup agelabel.
           imt imtlabel.
           cirrhosis cirrhosislabel.;
run;
```

(四)請提出合理的假說，解釋上述分析中呈現的人口學特徵。

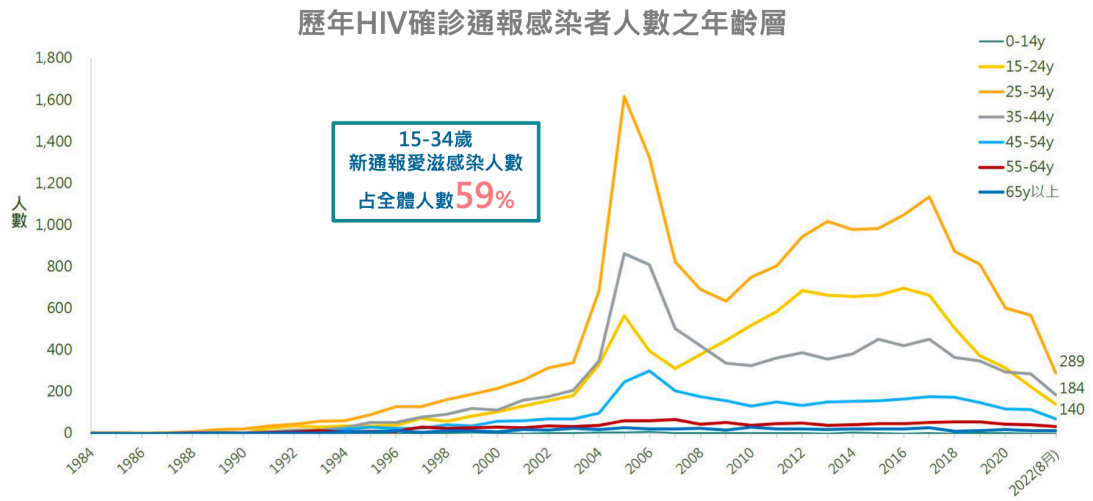
1. HIV vs gender :

- (1) HIV 組男性的比例為 100% (24/24)，non-HIV 組男性比例為 60.71% (17/28)，HIV 組男性比例顯著高於 non-HIV 組。
- (2) 查閱資料<sup>3</sup>可知，台灣 HIV 感染者男性遠多於女性 (19:1)，或可解釋「有隱球菌菌血症的 HIV 感染者」男性遠多於女性。



2. HIV vs age :

- (1) Fisher exact test 檢定結果 p-value 為 0.0001，拒絕虛無假設，年齡區間的比例在 HIV 組及 non-HIV 組不相同。
- (2) 查閱資料<sup>4</sup>可知，新通報的 HIV 感染者中 15-34 歲佔全部新通報的 HIV 感染者的 59%，其中又以 25-34 歲的區間為最多。或可解釋「有隱球菌菌血症的 HIV 感染者」之年齡分佈。



<sup>3</sup> 衛福部疾病管制署愛滋病統計資料 <https://www.cdc.gov.tw/Category/Page/rCV9N1rGUz9wNr8lggsh2Q>

<sup>4</sup> 衛福部疾病管制署愛滋病統計資料 <https://www.cdc.gov.tw/Category/Page/rCV9N1rGUz9wNr8lggsh2Q>

### 3. HIV vs immunosuppressant

- (1) Non-HIV 組接受免疫抑制治療的比例為 46.43% (13/28)，HIV 組接受免疫抑制治療的比例為 4.17% (1/24)，non-HIV 組顯著高於 HIV 組。
- (2) 隱球菌為常見的環境微生物<sup>5</sup>，一般免疫正常的人不易感染隱球菌，更不常見造成隱球菌菌血症等嚴重症狀。然而，當人處於免疫功能低下的狀態時，隱球菌即可能感染人體並導致症狀。HIV 感染（若沒有妥善的抗病毒藥物治療）為導致免疫低下的因素之一，而免疫抑制劑也是導致免疫低下的因素之一。
- (3) 推測使 HIV 組產生免疫功能低下的主因可能為 HIV 感染，而 non-HIV 組由於沒有 HIV 感染，使該組病人產生免疫低下尚有他因，服用免疫抑制劑或可為其一。

### 4. HIV vs liver cirrhosis :

- (1) Non-HIV 組有肝硬化的比例為 39.29% (11/28)，HIV 組有肝硬化的比例為 0% (0/24)，non-HIV 組顯著高於 HIV 組。
- (2) 肝臟為人體產生抗體、補體及免疫相關蛋白的重要器官，當肝硬化時肝功能下降，會使人體處於免疫低下的狀態。
- (3) 推測使 HIV 組產生免疫功能低下的主因可能為 HIV 感染，而 non-HIV 組由於沒有 HIV 感染，使該組病人產生免疫低下尚有他因，有肝硬化或可為其一。

### 5. 結論：

- (1) HIV 組之人口學特徵或可由 HIV 感染者的人口學特徵解釋。
- (2) Non-HIV 感染者服用免疫抑制劑及肝硬化的比例顯著高於 HIV 組，兩者或與非 HIV 感染者產生免疫低下狀態，進而導致隱球菌菌血症相關。

---

<sup>5</sup> Robbins and Cotran Pathologic Basis of Disease 10<sup>th</sup> edition

三. 從上述描述的流行病學分析，是否可以直接認定 HIV、接受免疫治療椅及肝硬化就是隱球菌菌血症的危險因子？為什麼？

HIV, 免疫抑制劑及肝硬化是否為隱球菌菌血症的危險因子不可由以上描述流行病學的資料就直接推論。以上資料僅針對疾病間共同的特點進行描述並產生危險因子的假說，由於缺乏對照組，假說不可直接推論為真。假說是否成立，尚待 cohort study 或 case-control study 建立對照組進行驗證