* 流行病學原理：資料分析practice4

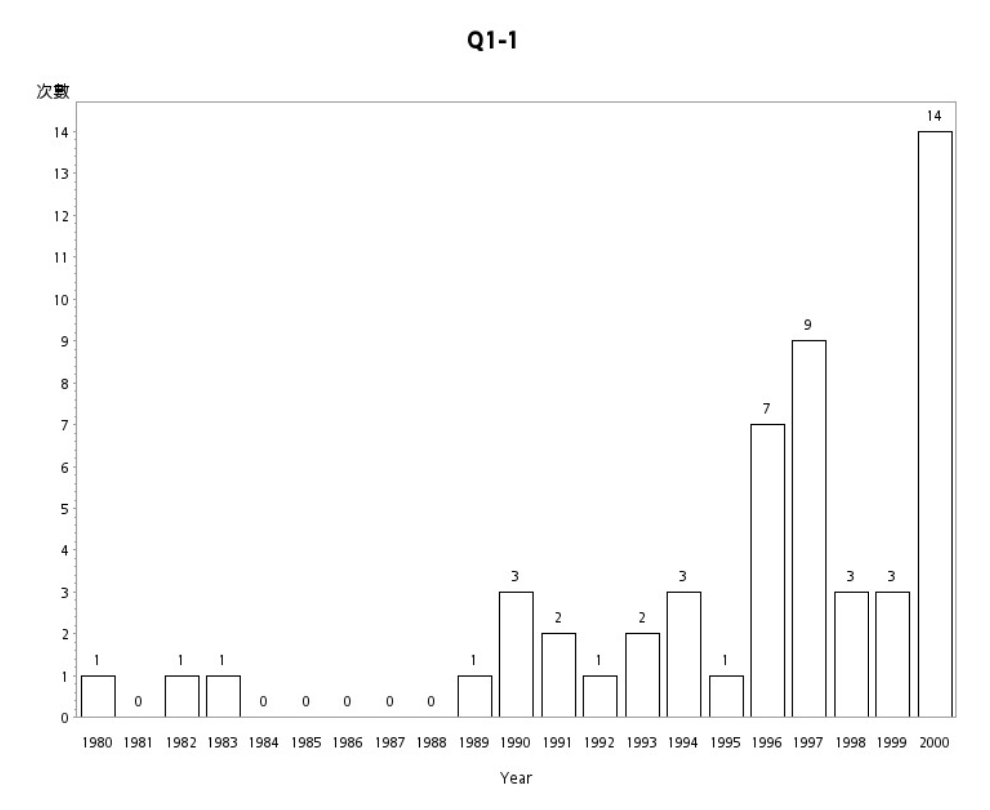
流行病學資料分析 – SAS

Practice 4

學號：b07401048

系級：醫學五

姓名：賴柏瑞

* 1. 請進行下列與時間相關的描述流行病學分析
     1. 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**;

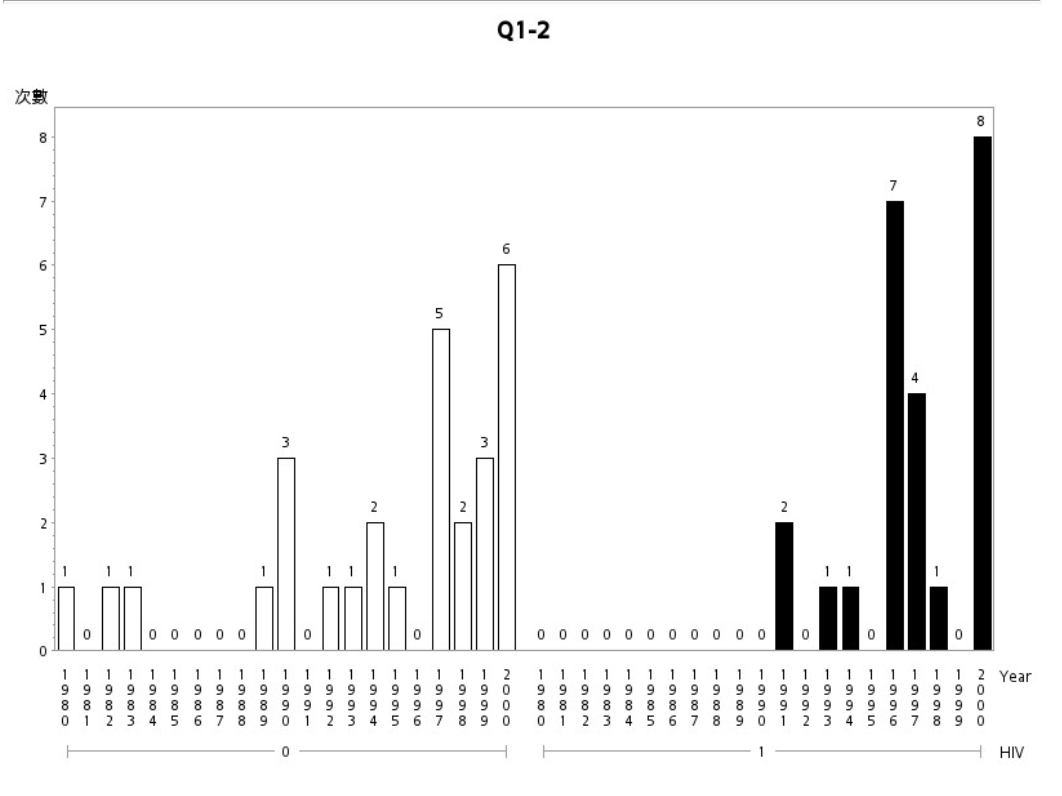
*/\*Q1-1 \*/*

title "Q1-1";

**proc** **gchart** data = dataset2;

vbar year / midpoints = 1980 to 2000 by 1 freq;

**run**;

* + 1. 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;

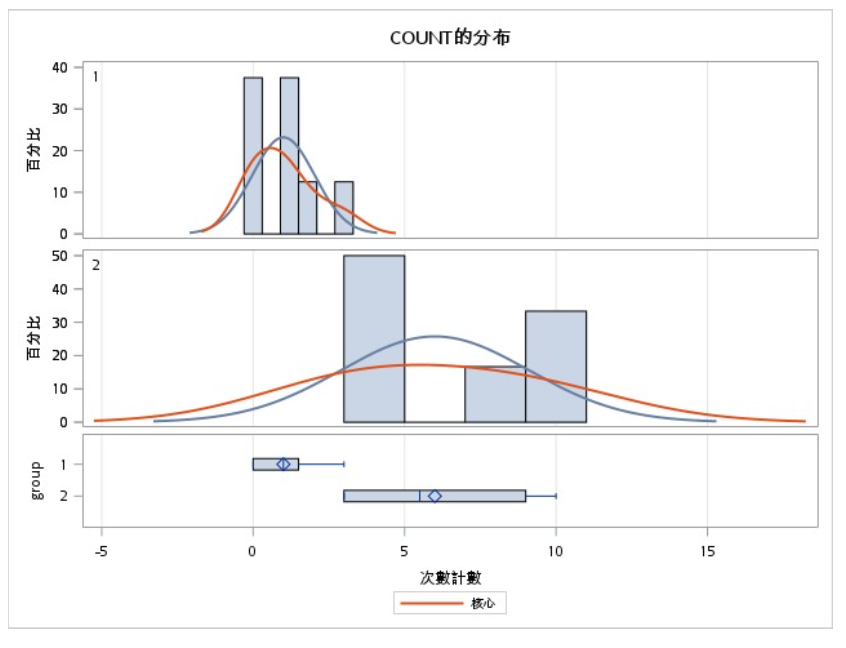
vbar year / midpoints = 1980 to 2000 by 1 freq group = hiv patternid = group;

pattern1 color = white;

pattern2 color = black;

**run**;

* + 1. 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（1980-1996）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：



* + - 1. 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 / nopercent 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**;

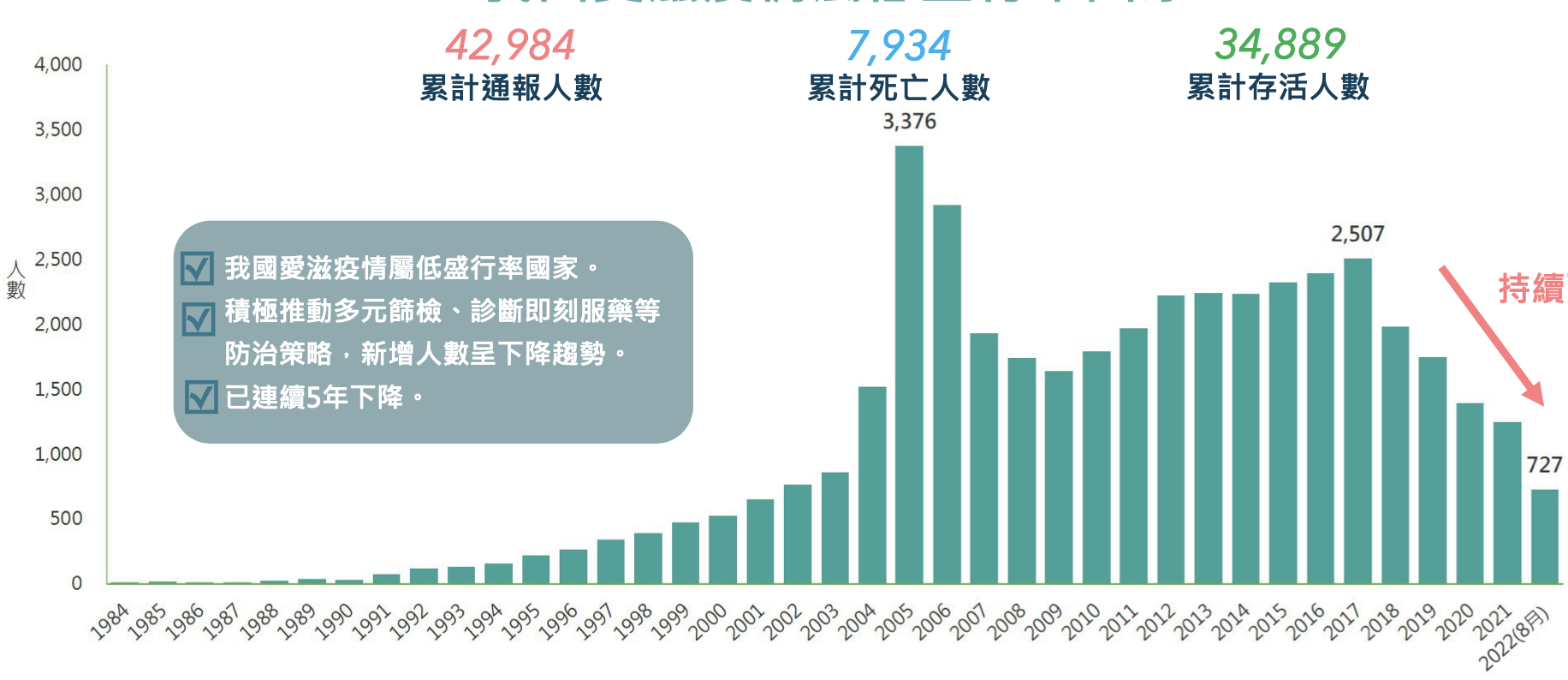
**proc** **ttest** data = year1996 h0 = 0 alpha = 0.05;

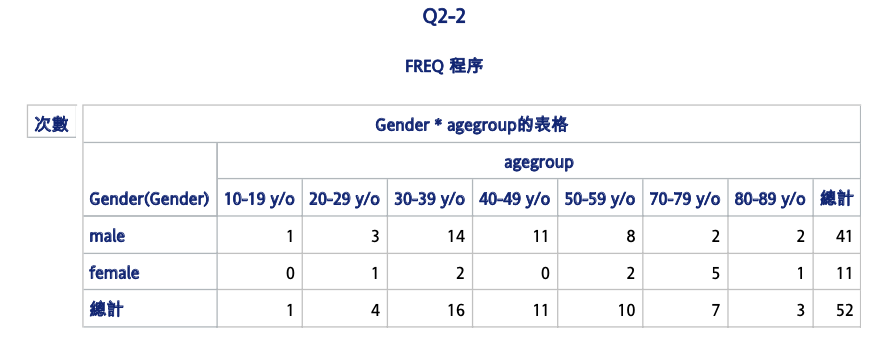
class group;

var count;

**run**;

* + 1. 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[[1]](#footnote-1)，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[[2]](#footnote-2)未發現1996年前後有HIV之爆發流行。



* 1. 人口學特徵相關的描述流行病學分析：
     1. 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：
        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 \*/*

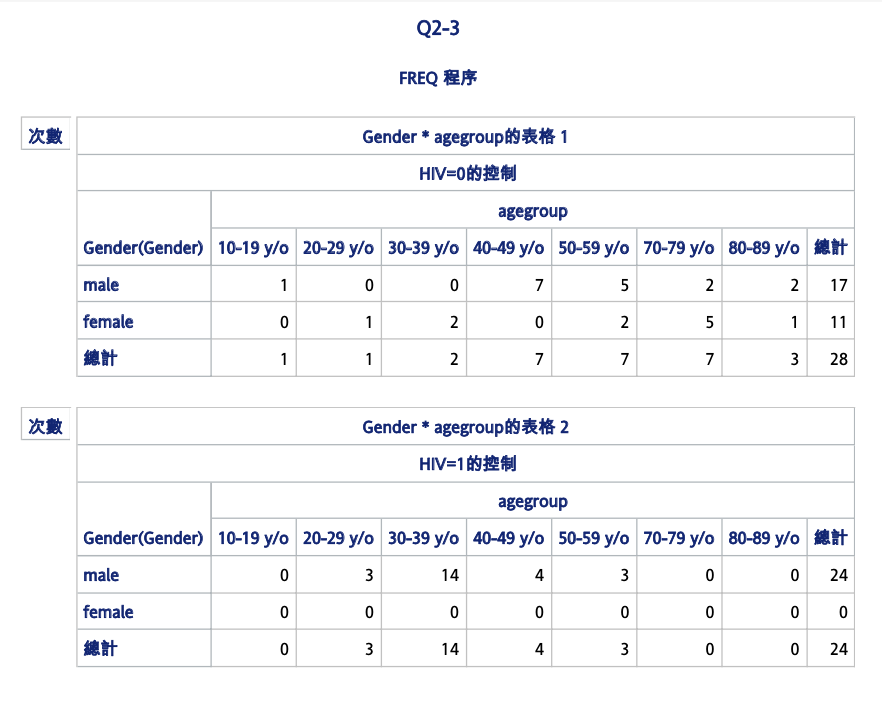
**proc** **freq** data = agegroup;

tables gender \* agegroup /norow nocol nopercent;

format gender **genderlabel.**

agegroup **agelabel.**;

**run**;

* + 1. 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：
          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;

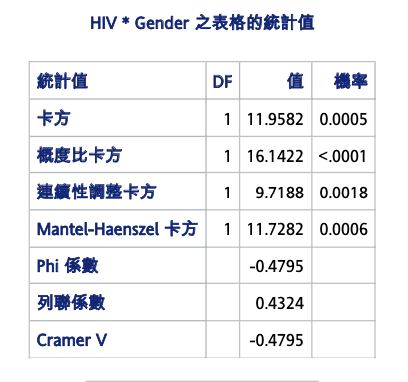
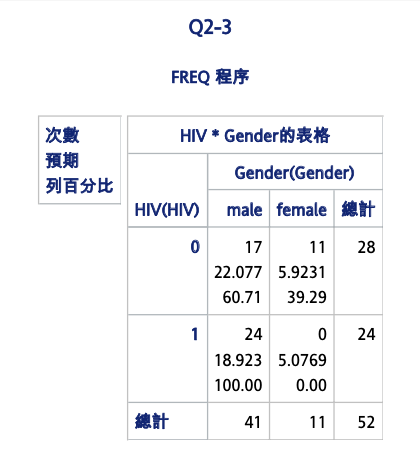
tables hiv\*gender\*agegroup / norow nocol nopercent;

format gender **genderlabel.**

agegroup **agelabel.**;

**run**;

* + 1. 比較HIV與非HIV組病人之1性別、2年齡分佈、3是否接受免疫治療以及4是否罹患肝硬化的比例是否有顯著差異。
       1. HIV vs. gender
          1. Figures：



* + - * 1. 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。

* + - * 1. Descriptions：

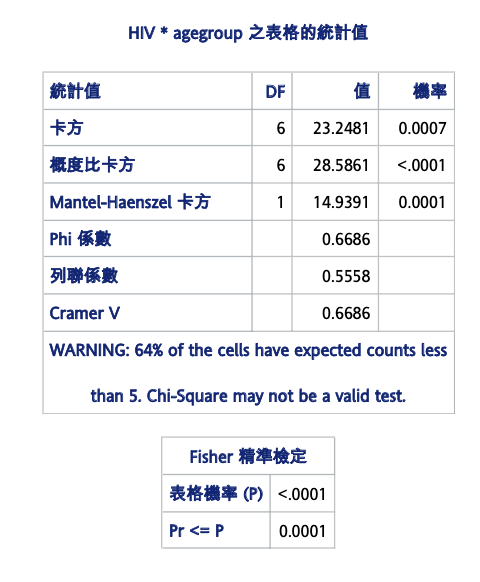
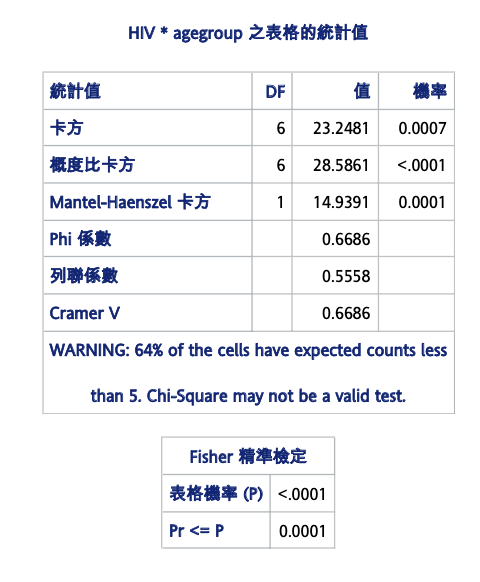
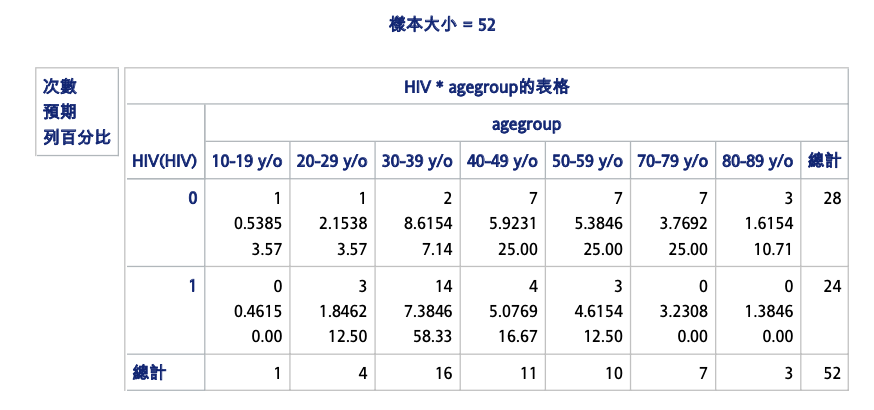
「是否為HIV感染者」及「男性或女性」兩者為類別變項，故以卡方檢定判斷兩者是否獨立。

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。

* + - 1. HIV vs. age：
         1. Figure：



* + - * 1. 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。

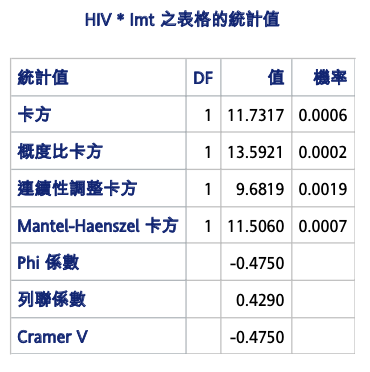
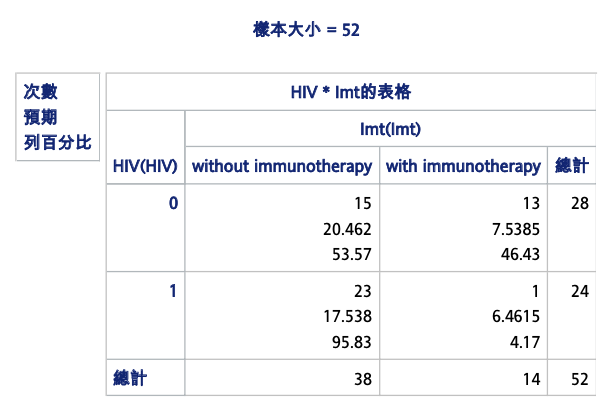
* + - * 1. Descriptions：

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。

* + - 1. HIV vs immunosuppressant ：
         1. Figures：



* + - * 1. 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。

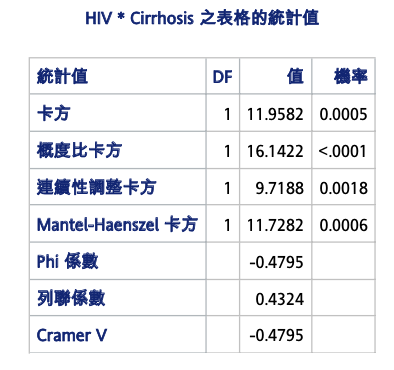
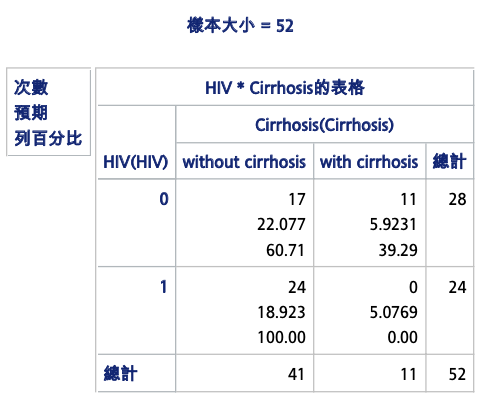
* + - * 1. Descriptions：

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。

* + - 1. HIV vs liver cirrhosis：
         1. Figures：



* + - * 1. Result：

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。

檢定結果p-value為0.0005，拒絕虛無假設，兩者非獨立變項。

* + - * 1. Descriptions：

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。

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。

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。

* + - * 1. Code：

*/\* Q2-3 \*/*

title "Q2-3";

*/\*chi-square \*/*

**proc** **freq** data = agegroup;

tables hiv\*( gender agegroup imt cirrhosis) / nopercent nocol chisq expected fisher;

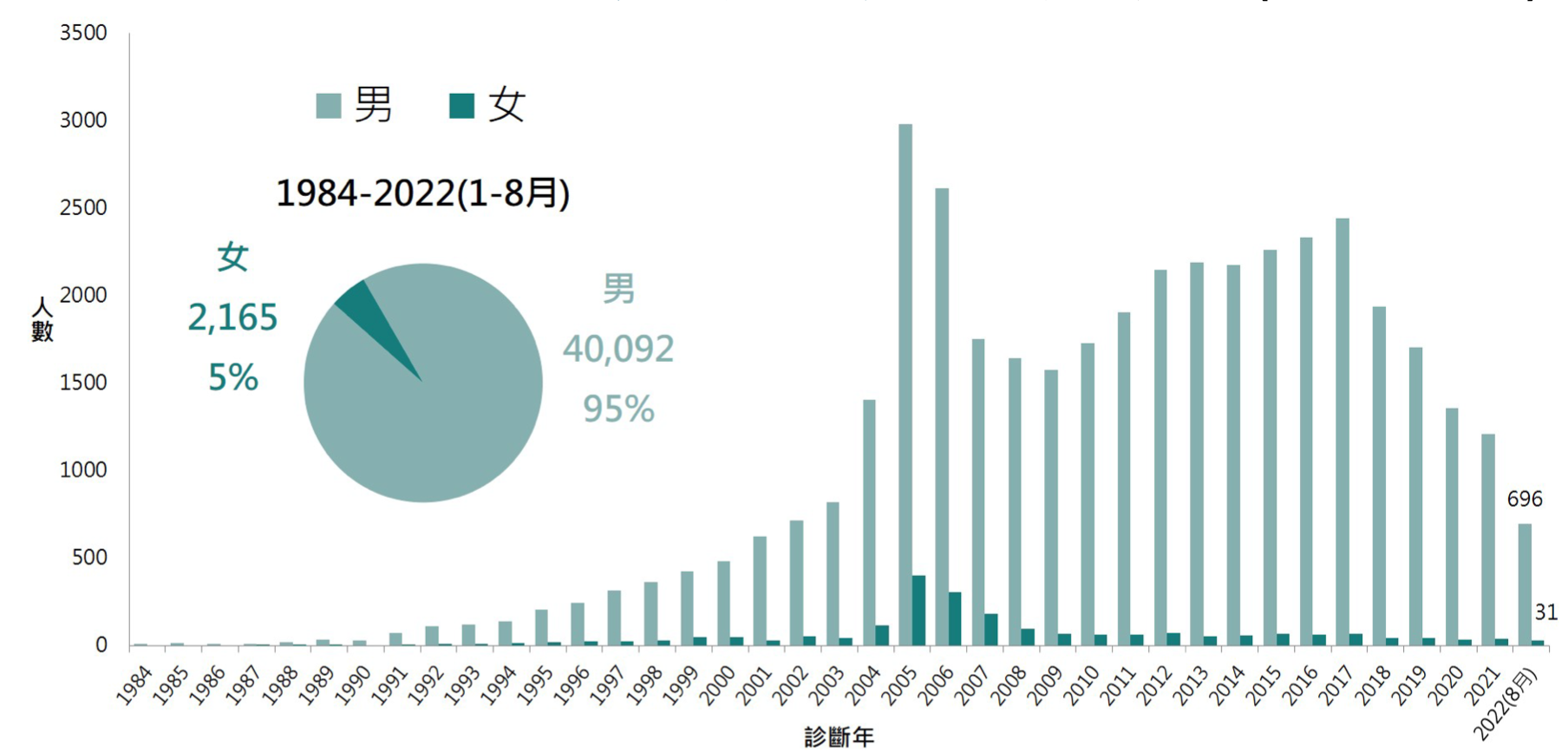
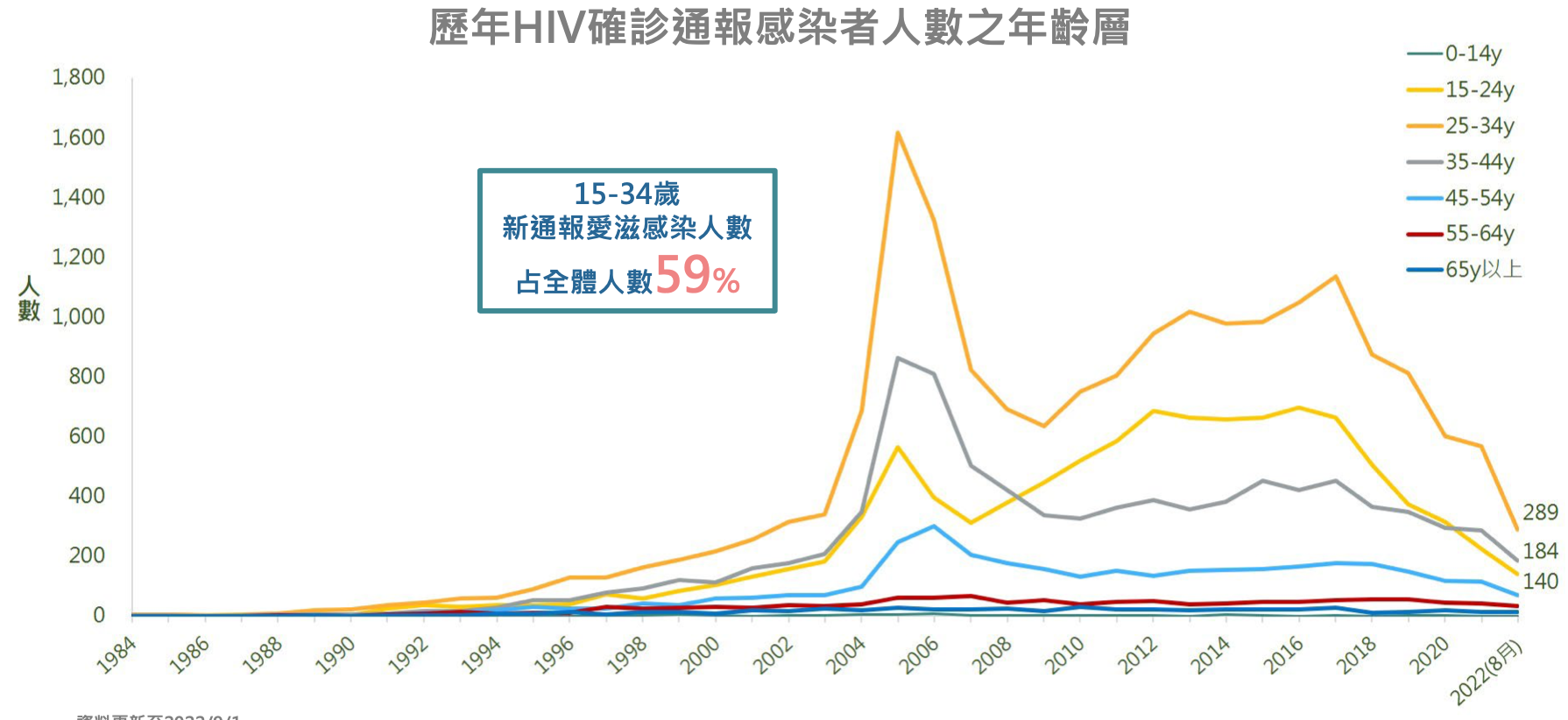
format gender **genderlabel.**

agegroup **agelabel.**

imt **imtlabel.**

cirrhosis **cirrhosislabel.**;

**run**;

* + 1. 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[[3]](#footnote-3)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[[4]](#footnote-4)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。
       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[[5]](#footnote-5)，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。
  1. 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行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

1. Robbins and Cotran Pathologic Basis of Disease 10e [↑](#footnote-ref-1)
2. 衛福部疾病管制署愛滋病統計資料https://www.cdc.gov.tw/Category/Page/rCV9N1rGUz9wNr8lggsh2Q [↑](#footnote-ref-2)
3. 衛福部疾病管制署愛滋病統計資料https://www.cdc.gov.tw/Category/Page/rCV9N1rGUz9wNr8lggsh2Q [↑](#footnote-ref-3)
4. 衛福部疾病管制署愛滋病統計資料https://www.cdc.gov.tw/Category/Page/rCV9N1rGUz9wNr8lggsh2Q [↑](#footnote-ref-4)
5. Robbins and Cotran Pathologic Basis of Disease 10th edition [↑](#footnote-ref-5)