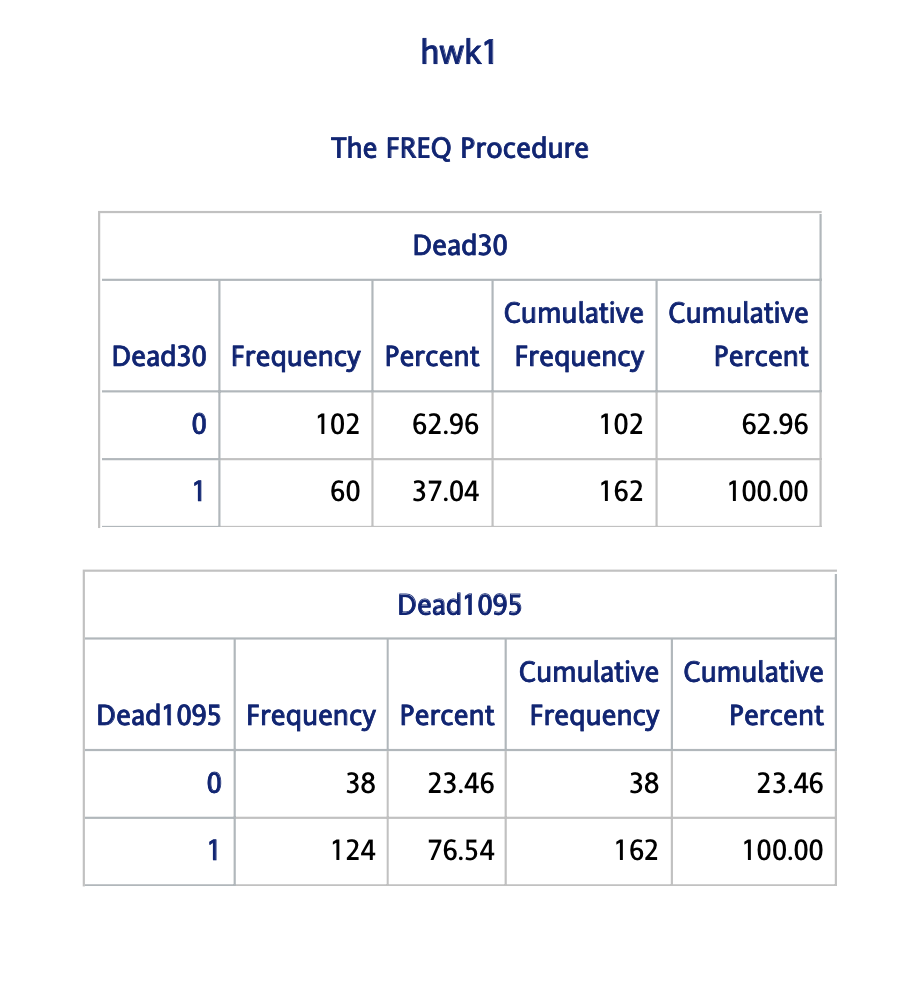
* 流行病學原理：資料分析 課堂作業1 （data set 3 ）
  1. Risk 的計算與作圖
     1. 請計算 Dataset 3 抗藥性金黃色葡萄球菌(Methicillin-resistant Staphylococcus aureus, MRSA)菌血症病人 30 天內死亡之 Risk 與 1095 天內死亡之 Risk 。
        1. Results：
           1. 30天內死亡之risk：0.3704
           2. 1095天內死亡之risk：0.7654
        2. Descriptions：
           1. 被診斷為MRSA bacteremia的病人自第一次血液培養陽性起的30天內發生死亡的累積發生率為0.3704，即37.04%。換言之，100個被診斷為MRSA bacteremia的病人中，有37.04個會在第一次血液培養陽性後的30天內死亡。
           2. 在被診斷為MRSA bacteremia的病人中，自第一次血液培養陽性起的1095天內發生死亡的累積發生率為0.7654，即76.54%%。換言之，100個被診斷為MRSA bacteremia的病人中，有76.54個會在第一次血液培養陽性後的1095天內死亡。
           3. 
        3. Figures：
        4. SAS code：

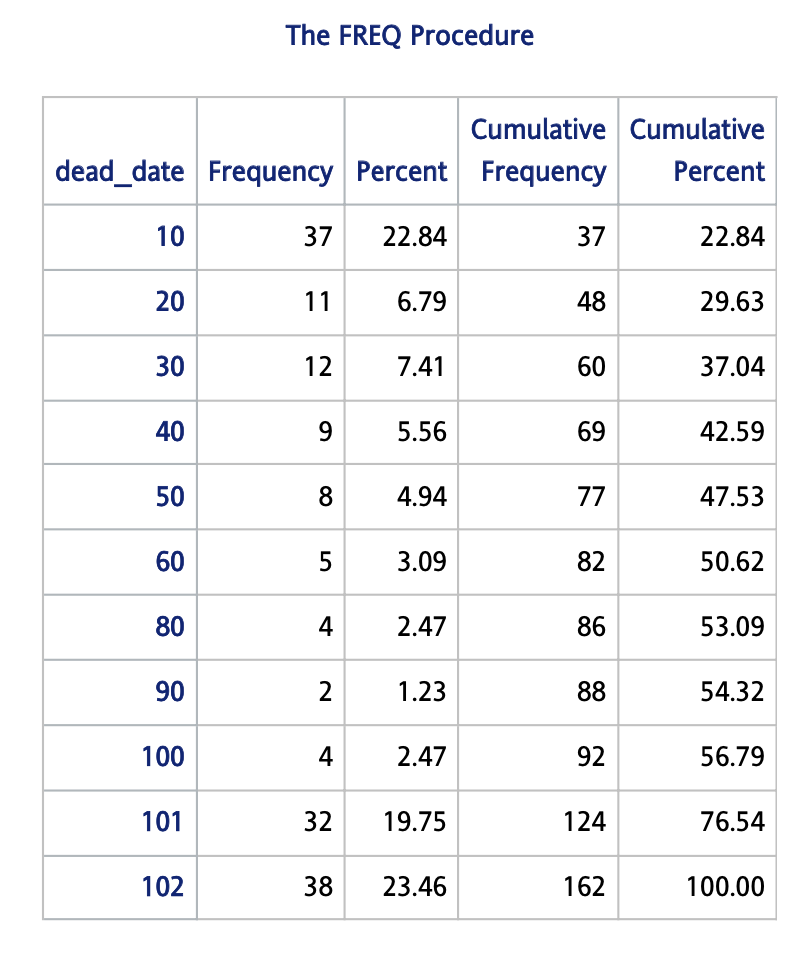
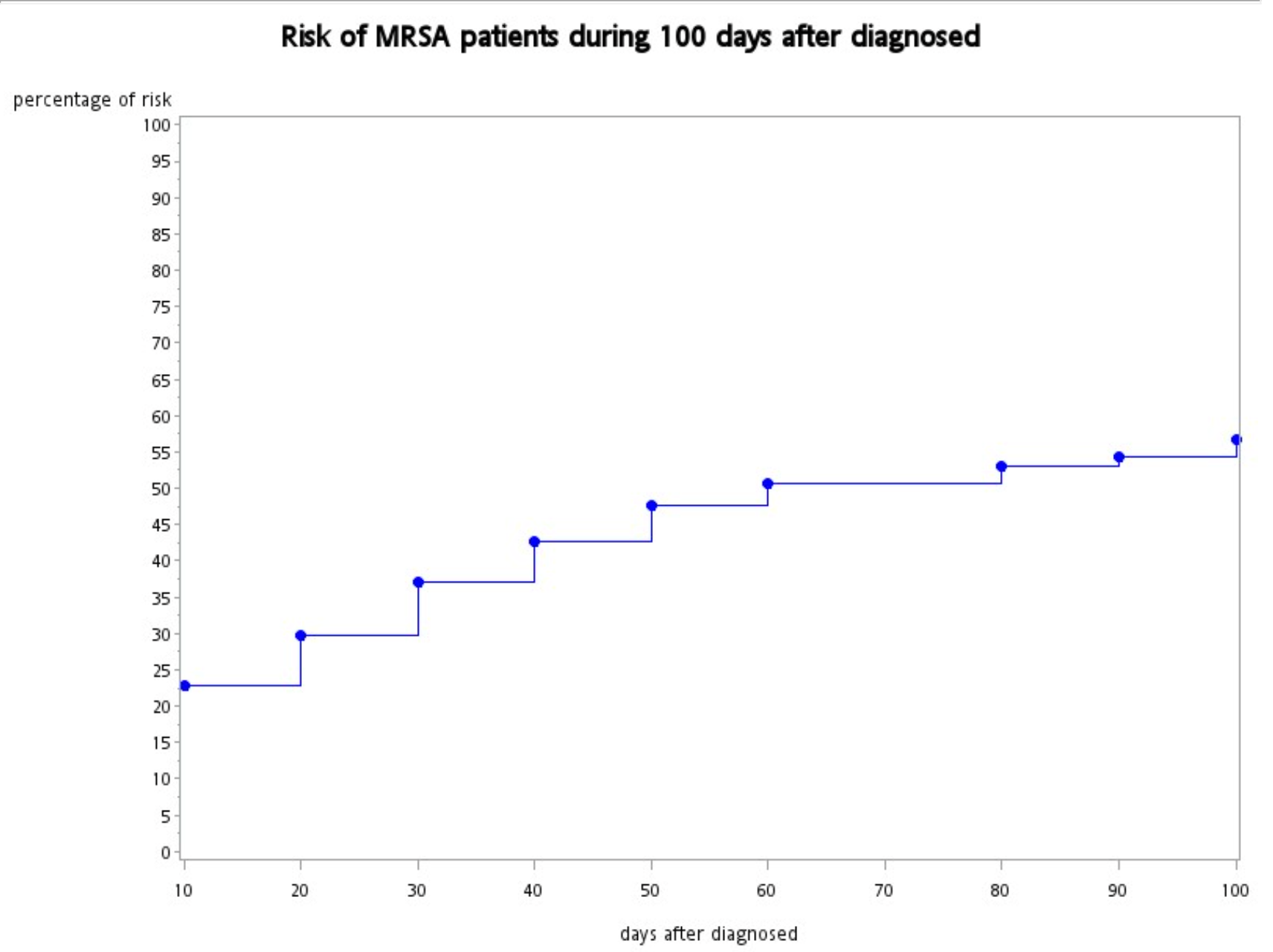
/\* hwk1 \*/

/\* hwk1-1 calculate risk of day30 and day1095 \*/

proc freq data = hwk1.dataset3;

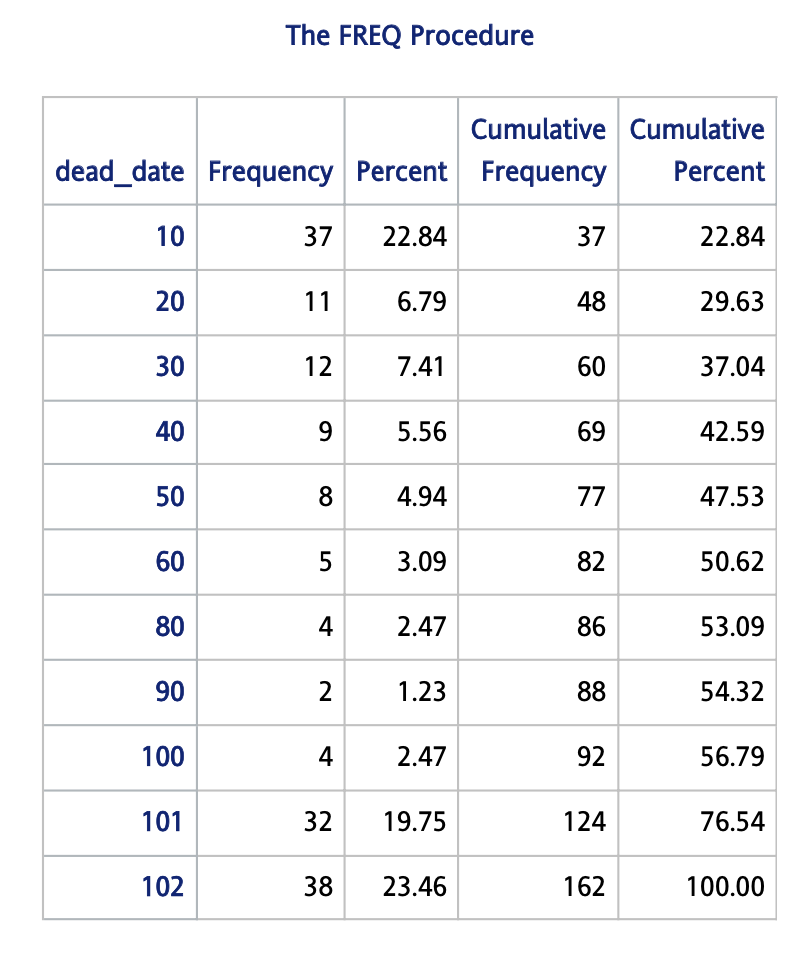
tables dead30 dead1095 / out = hwk1.hwk1\_1;

run;

* + 1. 延續上題，請計算 MRSA 菌血症病人 10 天, 20 天, 30 天, 40 天, 50 天, 60 天, 70 天, 80 天, 90 天, 及 100 天內死亡之 Risk，然後對時間作圖，並說明病人死亡之 Risk 隨時間變化的趨勢。
       1. Results：
          1. 10天內死亡之risk為22.84%
          2. 20天內死亡之risk為29.63%
          3. 30天內死亡之risk為37.04%
          4. 40天內死亡之risk為42.59%
          5. 50天內死亡之risk為47.53%
          6. 60天內死亡之risk為50.62%
          7. 70天內死亡之risk為50.62%
          8. 80天內死亡之risk為53.09%
          9. 90天內死亡之risk為55.32%
          10. 100天內死亡之risk為56.79%
       2. Description：
          1. Risk 的解釋如Q1-1。
          2. 觀察數據可發現，若以30天為區間，0至30天死亡的累積發生率上升了37.04%，31至60天上升了13.58%。61至90天上升了4.70%。可見隨時間離第一次血液培養陽性越遠，累積發生率上升幅度越小，意即單位時間內發生死亡的次數越低。（如附圖所標示）
       3. Figures：

**13.58%**

**37.04%**



* + - 1. SAS code：

/\* hwk1-2 risk in day(10, 100, by = 10) \*/

/\* grouping \*/

data hwk1.hwk1\_2;

set hwk1.dataset3;

if dead1095 = 1 then

if day1095 <= 10 then dead\_date = 10;

else if day1095 <= 20 then dead\_date = 20;

else if day1095 <= 30 then dead\_date = 30;

else if day1095 <= 40 then dead\_date = 40;

else if day1095 <= 50 then dead\_date = 50;

else if day1095 <= 60 then dead\_date = 60;

else if day1095 <= 70 then dead\_date = 70;

else if day1095 <= 80 then dead\_date = 80;

else if day1095 <= 90 then dead\_date = 90;

else if day1095 <= 100 then dead\_date = 100;

else dead\_date = 101;

else dead\_date = 0;

run;

/\* calculate risk \*/

proc freq data = hwk1.hwk1\_2;

tables dead\_date / out = hwk1\_2\_risk outcum;

run;

/\* plot x = dead date, y = cumulative risk \*/

proc gplot data = hwk1.hwk1\_2\_risk(where = (dead\_date));

plot cum\_pct \* dead\_date / haxis = axis1 vaxis = axis2;

axis1

label = ("days after diagnosed")

order = (10 to 100 by 10);

axis2

label = ("percentage of risk")

order = (0 to 100 by 5);

symbol

interpol = steplj

line = 1

width = 2

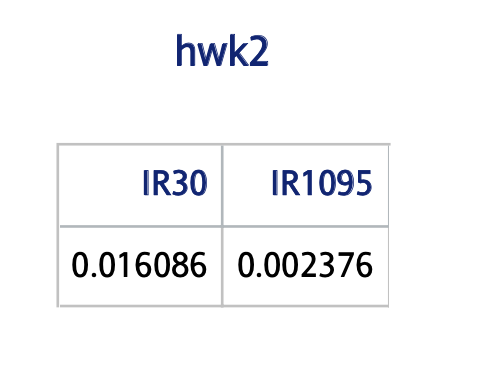
color = blue

value = dot;

title "Risk of MRSA patients during 100 days after diagnosed";

run;

quit;

* 1. Incident rate的計算與作圖
     1. 請計算 Dataset 3 MRSA 菌血症病人 30 天內死亡之 Incidence Rate 與 1095 天內死亡之 Incidence Rate(單位請用:日-1)
        1. Result：
           1. 30天內死亡的incident rate：IR30 = 0.016086（1/day）, 1 / IR30 = 62.167
           2. 1095天內死亡的incident rate：IR1095 = 0.002376（1/day）, 1/IR1095 = 420.88
        2. Description：
           1. 在被診斷為MRSA bacteremia 的病人中，自第一次血液培養陽性起的30天內，發生死亡的incident rate為0.016086。若將incident rate予以倒數，可解釋為「觀察1個MRSA bacteremia病人，每62.167天就會有一人因發生死亡」，抑或「觀察5人，每12.4334天就會有一人發生死亡」。
           2. 在被診斷為MRSA bacteremia 的病人中，自第一次血液培養陽性起的1095天內，發生死亡的incident rate為0.002376。若將incident rate予以倒數，可解釋為「觀察1個MRSA bacteremia病人，每420.88天就會有一人因發生死亡」，抑或「觀察5人，每84.176天就會有一人發生死亡」
           3. 由IR30 ≠ IR1095可知死亡的發生率並非始終維持不變，即死亡的病人數與時間並非線性關係。
           4. 由IR30 > IR1095可知，在30 天內死亡的發生率高於30至1095天死亡的發生率。或可推測MRSA bacteremia病人於30天內仍屬危險期，發生死亡的風險較高。
        3. figures：
        4. SAS code：

/\* hwk2\*/

/\*hwk2-1 calculate incident rate in day30 and day1095 \*/

proc sql;

select sum(dead30) / sum(day30) as IR30,

sum(dead1095) /sum(day1095) as IR1095

from hwk1.dataset3;

quit;

* + 1. 延續上題，請計算 MRSA 菌血症病人 0 到 30 天天內死亡之 Incidence Rate, 30 到 60 天天內死亡之 Incidence Rate, 及 60 到 90 天內死亡之 Incidence Rate。請問上述三個不同時段，病人死亡的 Incidence Rate 是否相同?
       1. Result：
          1. 0到30天內死亡的incident rate：IR30 = 0.016086（1/day）, 1 / IR30 = 62.167（day）
          2. 30到60天內死亡的incident rate：IR30-60 = 0.008209（1/day）, 1 / IR30-60 = 121.82
          3. 60到90天內死亡的incident rate：IR60-90 = 0.002581（1/day），1 / IR60- 90 = 387.446
       2. Descriptions：
          1. Incidence rate的解釋如2-1
          2. 由IR30 , IR30-60, IR60-90互不相等可知死亡的發生率並非始終維持不變，即死亡的病人數與時間並非線性關係。
          3. 由IR30 > IR30-60 > IR60-90可知，MRSA bacteremia 的病人在診斷後的30日內死亡發生率最高，且離診斷日越久，死亡發生率越低。
          4. 或可推測，以30日為區間，MRSA bacteremia在診斷的30日內仍屬危險期，病情不易控制，病況不穩定，發生死亡的機率最高；在30-60日，60-90日死亡發生率逐漸降低，推測病情越趨穩定。
       3. Figures：
       4. SAS code：

/\* hwk2-2 calculate incident rate in day30, day30-60, day60-90 \*/

/\* IR-0-30-60-90 \*/

data hwk1.IR0\_30\_60\_90;

set hwk1.dataset3;

if day1095 <=30 and dead1095 = 1 then dead0\_30 = 1;

else dead0\_30 = 0;

if 30 < day1095 <= 60 and dead1095 = 1 then dead30\_60 = 1;

else dead30\_60 = 0;

if 60 < day1095 <= 90 and dead1095 = 1 then dead60\_90 = 1;

else dead60\_90 = 0;

if day1095 > 30 then day0\_30 = 30;

else day0\_30 = day1095 ;

if day1095 > 60 then day30\_60 = 30;

else if 30 < day1095 <= 60 then day30\_60 = (day1095 - 30);

else day30\_60 = 0;

if day1095 > 90 then day60\_90 = 30;

else if 60 < day1095 <= 90 then day60\_90 = (day1095 - 60);

else day60\_90 = 0;

run;

/\* just to check \*/

proc print data = hwk1.ir0\_30\_60\_90;

var day1095 dead0\_30 dead30\_60 dead60\_90 day0\_30 day30\_60 day60\_90;

run;

/\* results \*/

proc sql;

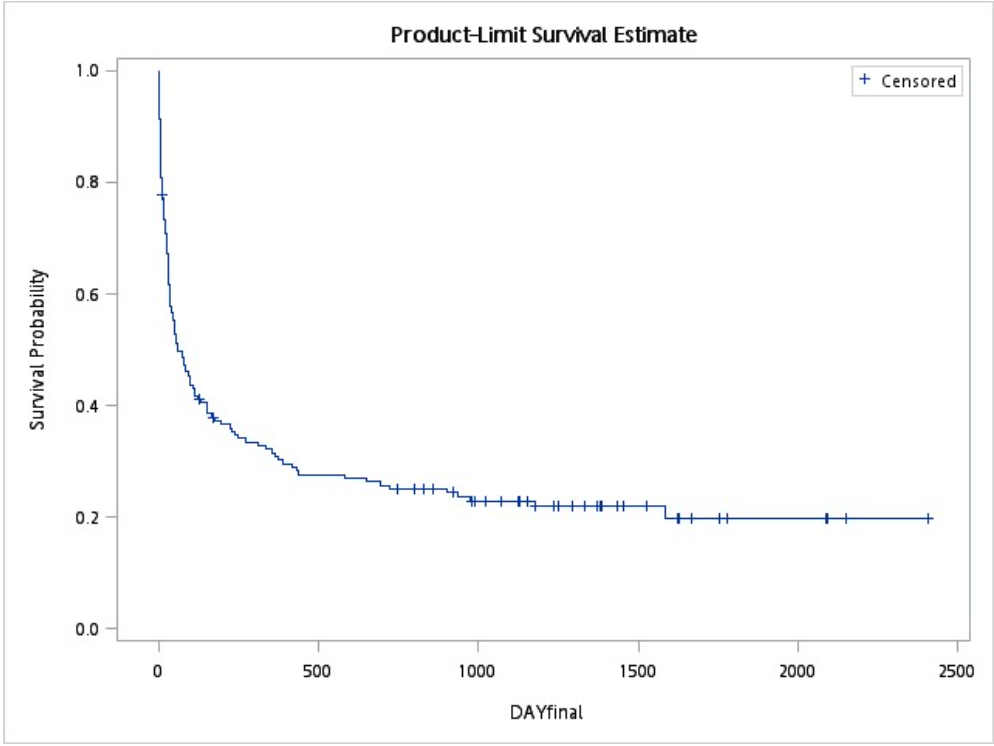
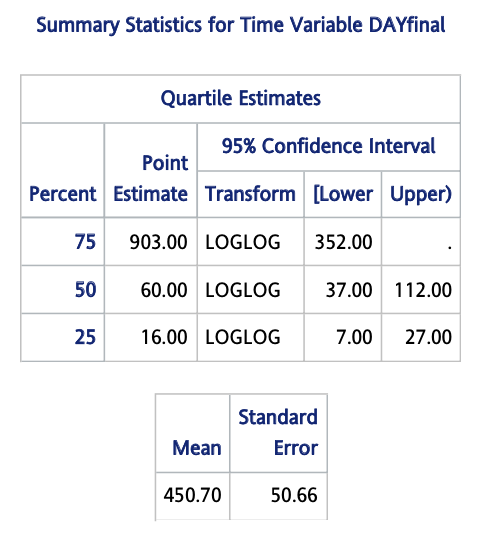
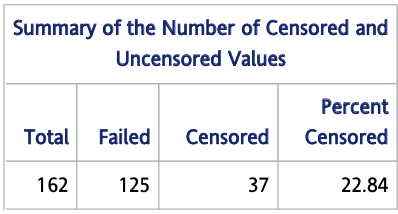
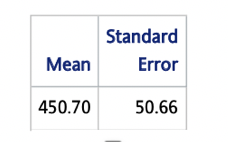
select sum(dead0\_30) / sum(day0\_30) as IR0\_30,

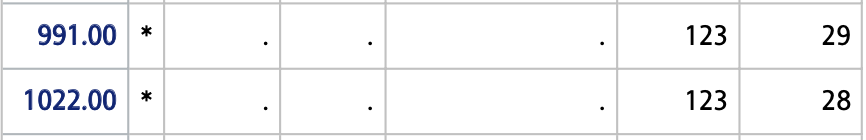
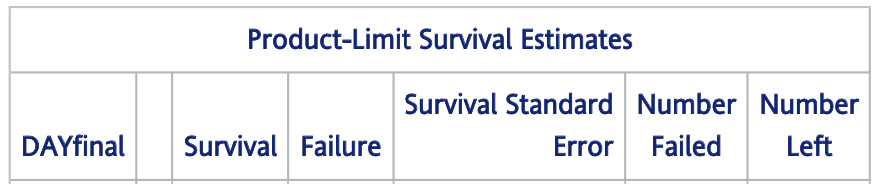
sum(dead30\_60) / sum(day30\_60) as IR30\_60,

sum(dead60\_90) / sum(day60\_90) as IR60\_90

from hwk1.ir0\_30\_60\_90;

quit;

* 1. Time-to-event curve：
     1. 請以 Kaplan-Meier 法，計算 Dataset 3 MRSA 菌血症病人的存活曲線 (Survival curve) 。
        1. Result：



**0.1593**

**100**

**0.4356**

**0.2287**

**0.2763**

**0.0476**

**0.5644**

**Mean : 450.70 (day)**

**Medium : 60.00 (day)**

* + - 1. Descriptions：
         1. 上圖為Kaplan-Meier survival curve，橫軸為第一次血液培養陽性後的天數，縱軸為MRSA bacteremia病人的存活機率。
         2. 觀察上圖（如附圖標線）：

MRSA bacteremia病人的存活機率存活機率在0到100天內快速下降至0.4356；

100到500天內存活機率下降速度趨緩但仍然快速，自0.4356下降至0.2763，在400天內下降0.1593；

500天後下降速度趨緩，自500到1000天僅自0.2763下降至0.2287，在500天內下降0.0476。

1000 天之後存活機率則幾乎持平。

或可推測對於MRSA bacteremia病人而言100天內仍屬危險期，病情控制不易，死亡風險高，故存活機率下降快；100至500天情況稍佳，仍有一定的死亡風險但稍稍降低，存活機率下降稍緩；500天後病況穩定，存活機率下降趨緩。

然以上僅為推測，實際情況仍須分析個別死因，如Q3-2, 3。

* + - * 1. 觀察平均值（mean）與中位數（medium）：如上圖及上表所列，存活天數的medium為60.00，mean為450.70，在不考慮誤差的情況下（見(4)）兩者差異顯著，medium遠小於mean，類似右偏分佈。可見多數MRSA bacteremia病人的存活天數短，少數病人的存活天數極長。
        2. Mean的誤差：MRSA bacteremia病人平均存活天數為450.70日。然而，此數值可能被嚴重低估。由於研究有follow-up-time的限制，存活時間只會記錄至follow-up-time的截止日期（2005/12/07），卻有相當比例的病人存活超過該日期而未被記錄，故實際存活時間被嚴重低估。
      1. SAS code：

title hwk3;

/\* hwk3 \*/

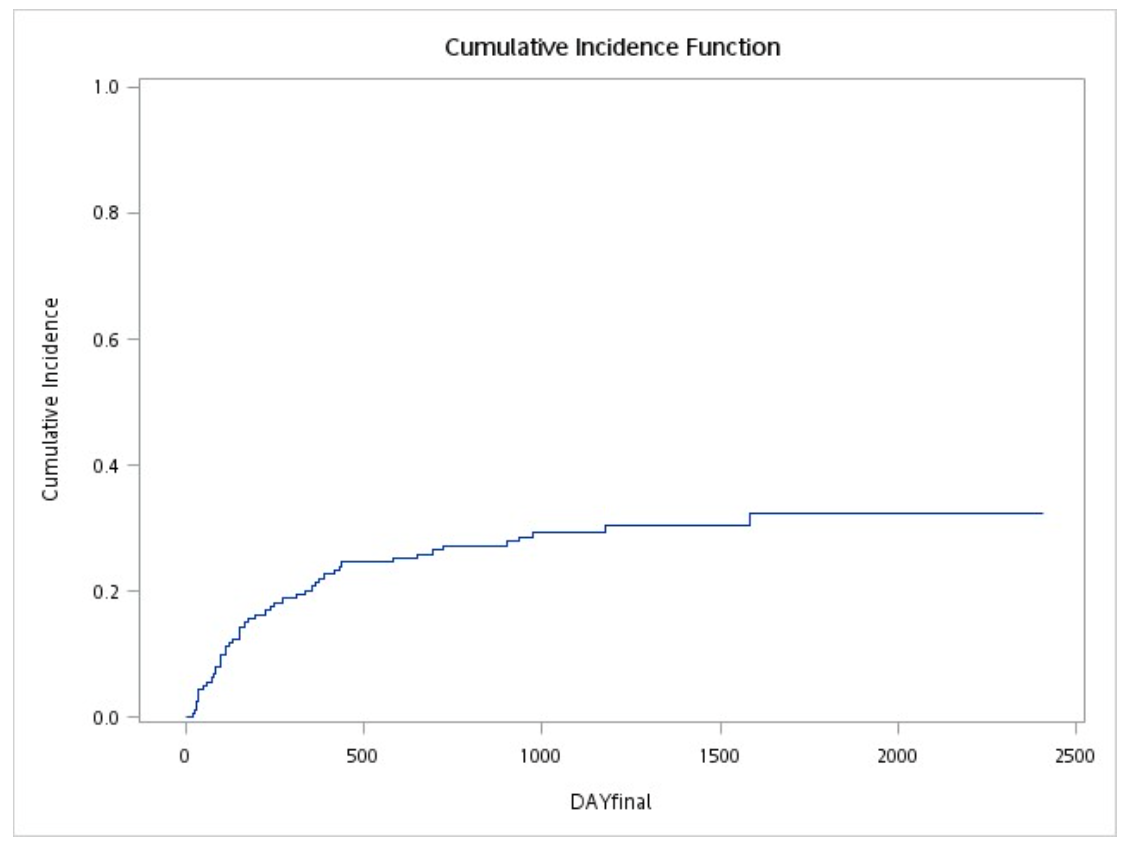
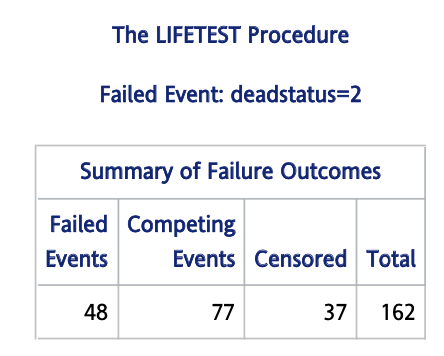
/\* hwk3-1 KM survival curve of MRSA patients \*/

proc lifetest data = hwk1.dataset3 plots = (s) method = KM;

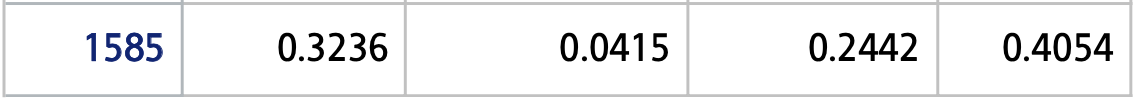
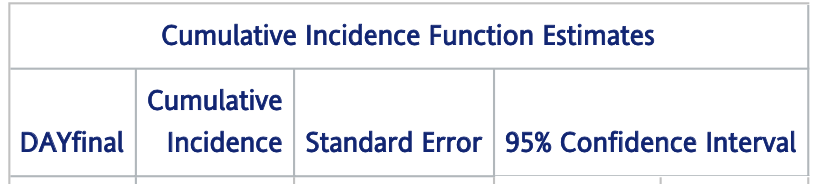
time dayfinal \* deadfinal(0) ;

title "survival curve";

run;

* + 1. 請以 Competing risk 方法，計算 Dataset 3 MRSA 菌血症病人因其他原因死亡 (不是死於 MRSA 菌血症死亡，而是死於其他原因) 的累積發生率曲線。
       1. Results：

**Day1585**



**Death before day500 :**

**37 / 48**

**Death after day500 :**

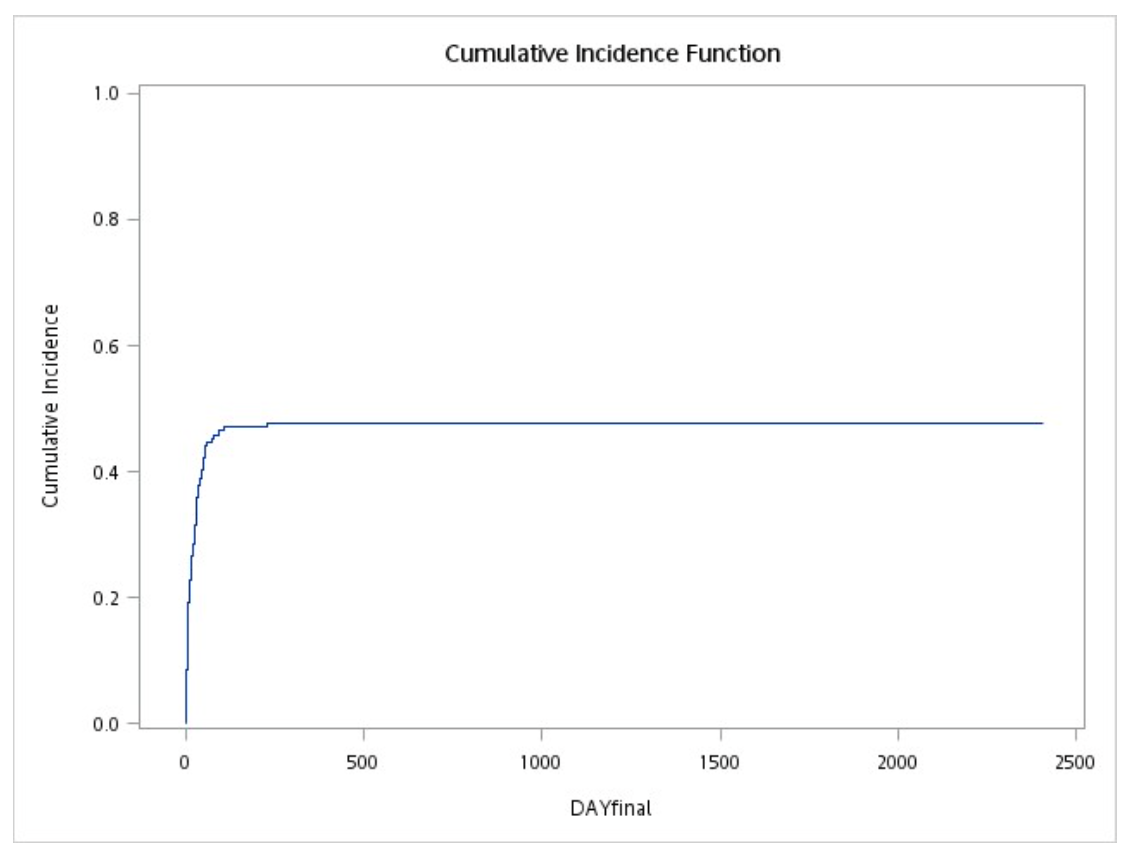
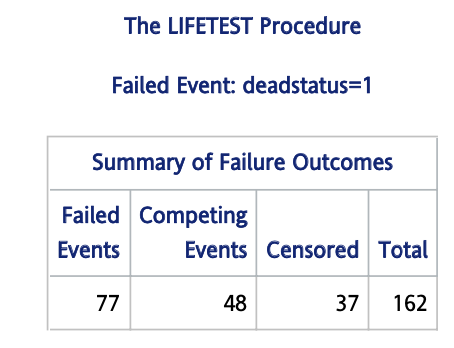
**9 / 48**

* + - * 1. 如上表所示，總共有162位MRSA bacteremia病人，其中，在觀察時間內死亡者有125人，而不是死於 MRSA bacteremia，而是死於其他原因者有48位。
        2. 觀察累積發生率曲線：

可發現曲線在500天以前上升較快，500天以後略呈平緩。

多數非死於MRSA bacteremia者，死於第一次血液培養陽性後500天以內（37 / 48），少數死於500天以後（9 / 48）。

最後一名在觀察時間內非死於MRSA bacteremia如圖上所標記，在第1585天，累計發生率為0.3236。

* + 1. 請以 Competing risk 方法，計算 Dataset 3 MRSA 菌血症病人因 MRSA 菌血症 死亡的累積發生率曲線
       1. Result：

**Death after day100 :**

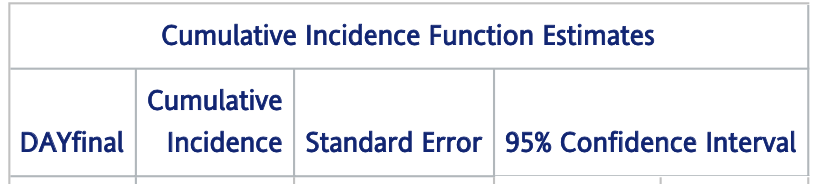
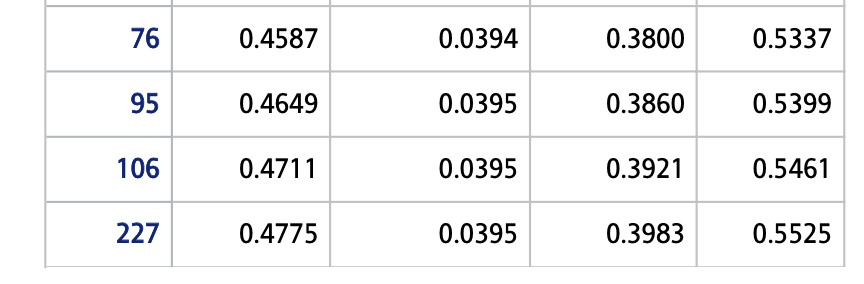
**2 / 77**

**Death before day500 :**

**75 / 77**

**Day106**

**Day227**



* + - 1. Description：
         1. 如上表所示，總共有162位MRSA bacteremia病人，其中，在觀察時間內死亡者有125人，而死於 MRSA bacteremia者有77位
         2. 觀察累積發生率曲線：

可發現曲線在100天以內上升極快，100天後幾乎持平。

絕大多數死於MRSA bacteremia者，死於第一次血液培養陽性後100天以內（75 / 77），少數死於500天以後（2 / 75）。唯二死於100天後者如圖上箭頭標示

最後一名在觀察時間內死於MRSA bacteremia如圖上所標記，在第227天，累計發生率為0.4775。

* + - * 1. 此累計發生率曲線可與Q3-1的survival curve相互對應。Q3-1的survival curve顯示MRSA bacteremia病人的存活機率在100天內快速下降。又比較Q3-2兩累計發生率曲線，可知死於MRSA bacteremia的病人集中於100天內；非死於MRSA bacteremia者，死亡發生率在500天以內都略為相似，並沒有集中於100天以前的現象。由此可知，造成Q3-1 survival curve現象的原因，是MRSA bacteremia，而非其他死因。
      1. SAS code

/\* hwk3-2 cumulative risk of competing risk \*/

/\* grouping \*/

data hwk1.hwk3\_2;

set dataset3;

if deadfinal = 1 then

if MRSA\_relateddeath = 1 then deadstatus = 1;

else deadstatus = 2;

else deadstatus = 0;

run;

/\* just to check \*/

proc print data = hwk1.hwk3\_2;

var deadfinal MRSA\_relateddeath deadstatus;

run;

/\* plotting cumulative risk curve \*/

proc lifetest data = hwk1.hwk3\_2 plots = CIF outcif = MRSA\_nonMRSA;

time dayfinal \* deadstatus(0) / eventcode = 1;

time dayfinal \* deadstatus(0) / eventcode = 2;

run;