

R 軟體 – Kaplan-Meier 存活曲線的繪製

蔡靜雯 副統計分析師

存活分析中很常使用存活曲線呈現追蹤時間(或存活時間)和感興趣事件的發生情況，或者藉由不同群體之間的存活曲線圖形比較存活差異，本期 eNews 內容將介紹如何使用 `survminer` 套件繪製基本存活曲線，以及透過調整 `ggsurvplot()` 函數中的指令，進一步客製化存活曲線。

1. 存活曲線介紹

常見估計存活函數的方法有生命量表、Kaplan-Meier法、Nelson-Aalen法等，這些都是不需要有存活時間分配(distribution)的假設條件，可從存活曲線圖形中方便觀察出存活時間中位數和存活時間對應的存活百分比。

1. 生命量表方法(life table method)或精算法(actuarial method)

為流行病學和保險學中估計存活函數的典型方法，以一群人在某時段的存活人數估計，藉由將觀測時間分割成一連串小的時間區間(不互相重疊)，區間不須相同時間長度(但大多數都分割成相同長度)，只需知道每一區間的總人數和死亡人數，不需要個人的資訊。

2. Kaplan-Meier方法(product-limit method)

為 Kaplan 和 Meier 在1958年提出的 Product-Limit(PL) 估計式，類似精算法的估計，差別在生命量表的精算估計是一群人的匯整資料，而 Kaplan-Meier 方法是保留每一個時間點的所有資訊，以個人死亡時間估計，因此，需知道每一個人確切的死亡時間。

2. Kaplan-Meier 存活曲線圖形所需的基本變項

繪製存活曲線圖形有兩個基本變項是必備的，分別是追蹤時間(例如：追蹤天數、追蹤月、追蹤年的連續變項)和感興趣事件是否發生(例如：有無死亡、有無復

發的二元變項)，若要比較不同群體或類別之間的存活曲線，就需再加上一個類別變項(例如：性別、治療方式、種族等)。

3. 範例資料檔

在這期 eNews 使用的資料是由美國4個研究肺癌的醫學中心，所整合出來的樣本資料。變數說明如下：

變數名稱	變數說明
SITE	4個研究中心
AGE	年齡(歲)
GENDER	性別(Female；Male)
gender	性別(0, Female；1, Male)
CHEMO	化療與否(0, No；1, Yes)
RT	放射治療(Yes；No；NA:未知)
rt	放射治療(0, 沒有；1, 有；NA, 未知)
Vital.Status	研究結束時存活狀態(Alive；Dead)
vital_status	研究結束時存活狀態(0, 存活；1, 死亡)
FIRST_PROGRESSION_OR_RELAPSE	術後第一次復發與否(Yes；No)
MONTHS_TO_FIRST_PROGRESSION	手術後到第一次復發時間(單位:月)
SMOKING	抽菸狀態(Never smoked, Smoked in the past, Currently smoking)
N_STAGE	癌症N分期(N0, 第0期；N1, 第一期；N1 or N2, 第一或二期)
T_STAGE	癌症T分期(T1, 第一期；T2 or T3, 第二或三期；T2, T3 or T4, 第二、三或四期)
SURVIVAL_MONTHS	存活時間(單位:月)

資料來源可在數據處健康資料加值暨統計中心的檔案下載網頁下載，請參考以下路徑，臺北醫學大學數據處首頁 > 二級單位 健康資料加值暨統計中心 > 檔案下載-資料檔 > 點選肺癌研究檔案下載或點選以下網址

https://ods.tmu.edu.tw/upload_file/tmudc/526/15880545261.csv

肺癌研究檔案下載的格式為csv檔，另存命名為 lung cancer data，檔案內容呈現如下圖：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SITE	GENDER	gender	AGE	CHEMO	RT	rt	Vital.Status	vital_status	FIRST_PROGRESSION_OR_RELAPSGRESSIO	MONTHS_TO_FIRST_PRO	SMOKING	N_STAGE	T_STAGE	SURVIVAL_MONTHS
2	DFCI	Female	0	55	No	No	0	Alive	0	No	NA	Smoked in the past	N0	T2 or T3	110
3	DFCI	Female	0	41	No	No	0	Alive	0	Yes	2	Smoked in the past	N0	T2 or T3	98
4	DFCI	Male	1	47	Yes	No	0	Alive	0	No	NA	Smoked in the past	N0	T2 or T3	110
5	DFCI	Male	1	73	NA	NA	NA	Alive	0	NA	NA	Never smoked	N0	T2 or T3	66
6	DFCI	Female	0	63	NA	NA	NA	Dead	1	Yes	17	Currently smoking	N1	T2 or T3	29

4. R語法說明與範例演練

✧ 讀入資料檔 lung cancer data 並 檢查讀入的資料內容和結構

> 使用 read.csv() 指令讀入檔名為lung cancer data的檔案，並且命名為lung_cancer

【語法】 指定在程式中欲使用的檔名= read.csv("檔案儲存路徑\\檔案名稱.csv")

其中，檔案儲存路徑中的斜線，需改為雙斜線或反斜線

[R程式碼]-讀入資料

```
lung_cancer=read.csv("J:\\eNews 第四十六期\\lung cancer data.csv")
lung_cancer=read.csv("J:/eNews 第四十六期/lung cancer data.csv")
```

[R程式碼]- 檢查是否成功讀入資料以及確認變數資料

```
View(lung_cancer) #以表格形式的視窗呈現，可查看整體資料
head(lung_cancer) #查看部份筆資料(預設是前6筆)
str(lung_cancer) #查看資料結構
```

> View(lung_cancer)

	SITE	GENDER	gender	AGE	CHEMO	RT	rt	Vital.Status	vital_status	FIRST_PROGRESSION
1	DFCI	Female	0	55	No	No	0	Alive	0	No
2	DFCI	Female	0	41	No	No	0	Alive	0	Yes
3	DFCI	Male	1	47	Yes	No	0	Alive	0	No
4	DFCI	Male	1	73	NA	NA	NA	Alive	0	NA
5	DFCI	Female	0	63	NA	NA	NA	Dead	1	Yes
6	DFCI	Male	1	72	NA	NA	NA	Dead	1	Yes
7	DFCI	Female	0	57	NA	NA	NA	Alive	0	NA
8	DFCI	Female	0	55	NA	NA	NA	Alive	0	NA

```
> head(lung_cancer)
  SITE GENDER gender AGE CHEMO   RT rt Vital.Status vital_status
1 DFCI Female     0  55   No   No  0         Alive           0
2 DFCI Female     0  41   No   No  0         Alive           0
3 DFCI  Male     1  47  Yes   No  0         Alive           0
4 DFCI  Male     1  73 <NA> <NA> NA         Alive           0
5 DFCI Female     0  63 <NA> <NA> NA         Dead            1
6 DFCI  Male     1  72 <NA> <NA> NA         Dead            1
  FIRST_PROGRESSION_OR_RELAPSE MONTHS_TO_FIRST_PROGRESSION
1                               No                          NA
2                               Yes                          2
3                               No                          NA
4                               <NA>                       NA
5                               Yes                         17
6                               Yes                          5
  SMOKING N_STAGE  T_STAGE SURVIVAL_MONTHS
1 Smoked in the past    N0 T2 or T3          110
2 Smoked in the past    N0 T2 or T3           98
3 Smoked in the past    N0 T2 or T3          110
4      Never smoked     N0 T2 or T3           66
5  Currently smoking    N1 T2 or T3           29
6      Never smoked     N0 T2 or T3           7
```

[R程式碼] - 查看資料結構

```
#查看資料結構(資料筆數、變數數量、變數名稱和變數類型)
```

```
str(lung_cancer)
```

```
> str(lung_cancer)
'data.frame':  478 obs. of  15 variables:
 $ SITE           : chr  "DFCI" "DFCI" "DFCI" "DFCI" ...
 $ GENDER         : chr  "Female" "Female" "Male" "Male" .
 $ gender         : int  0 0 1 1 0 1 0 0 1 0 ...
 $ AGE            : int  55 41 47 73 63 72 57 55 64 40 ...
 $ CHEMO          : chr  "No" "No" "Yes" NA ...
 $ RT            : chr  "No" "No" "No" NA ...
 $ rt            : int  0 0 0 NA NA NA NA NA NA NA ...
 $ Vital.Status   : chr  "Alive" "Alive" "Alive" "Alive" .
 $ vital_status   : int  0 0 0 0 1 1 0 0 0 0 ...
 $ FIRST_PROGRESSION_OR_RELAPSE: chr  "No" "Yes" "No" NA ...
 $ MONTHS_TO_FIRST_PROGRESSION : num  NA 2 NA NA 17 5 NA NA NA NA ...
 $ SMOKING        : chr  "Smoked in the past" "Smoked in t
 $ N_STAGE        : chr  "N0" "N0" "N0" "N0" ...
 $ T_STAGE        : chr  "T2 or T3" "T2 or T3" "T2 or T3"
 $ SURVIVAL_MONTHS : num  110 98 110 66 29 7 53 63 23 62 ..
```

◇ 安裝繪製Kaplan-Meier存活曲線所需套件，並在開始使用前叫出宣告使用

survival 套件：執行存活分析，估計 Kaplan-Meier 存活函數和檢定

survminer 套件：繪製Kaplan-Meier存活曲線圖

[R程式碼]- 安裝套件和宣告使用套件

```
# 安裝套件 - 同一台電腦安裝成功, 之後使用 library()叫出, 即可使用, 不用再重新
    安裝一次
install.packages("survival")
install.packages("survminer")

# 使用 library()叫出宣告要使用的套件
library(survival)
library(survminer)
```

確定讀入的資料和宣告使用的套件都沒問題後，就可以開始進行下一步，繪製 Kaplan-Meier 存活曲線圖形。

範例4-1、整體資料中所有人的存活曲線

如前面提到，繪製存活曲線圖形有兩個基本變項是必備的，追蹤時間和感興趣事件是否發生，對應範例檔肺癌研究檔案中的變項，即為 SURVIVAL_MONTHS(存活時間) 和 vital_status(研究結束時存活狀態，0：存活；1：死亡)，其中要注意的是，感興趣事件是否發生的資料記錄方式要為 0、1，未發生為 0，有發生為 1。

1. 首先使用 survival 套件計算存活函數

> 使用 survival 套件中的 survfit() 函數計算存活函數

【語法】 survfit(Surv(time, event)~1, data)

其中，

time 為追蹤時間

event 為感興趣事是否發生 (1：有發生，0：未發生)

data 欲分析的檔案

> 使用 survminer 套件中的 ggsurvplot() 函數，繪製 Kaplan-Meier 存活曲線圖

【語法】 ggsurvplot(fit)

其中， fit 為 survfit() 函數估計的存活函數

[R程式碼]-

```
fit <- survfit(Surv(SURVIVAL_MONTHS, vital_status) ~ 1, data = lung_cancer) # 使用 survfit 函數計算存活函數並存檔命名為 fit
fit # 整筆檔案計算存活函數的資訊說明
summary(fit) # 列出每一筆感興趣事件發生時的存活函數計算結果
ggsurvplot(fit) # Kaplan-Meier 存活曲線圖
```

[output 解讀]

有9筆資料在存活函數計算過程被刪除，n=469為實際有使用到的資料筆數，event=253為感興趣事件有發生的筆數，在此範例檔即為死亡人數。median=68.6、0.95LCL=57.3、0.95UCL=77.6 分別為追蹤時間(或存活時間)中位數、95%信賴區間下界和上界。

```
> fit <- survfit(Surv(SURVIVAL_MONTHS, vital_status) ~ 1, data = lung_cancer)
> fit
Call: survfit(formula = Surv(SURVIVAL_MONTHS, vital_status) ~ 1, data = lung_cancer)

因為不存在，9 個觀察量被刪除了
      n events median 0.95LCL 0.95UCL
[1,] 469     253   68.6    57.3   77.6
```

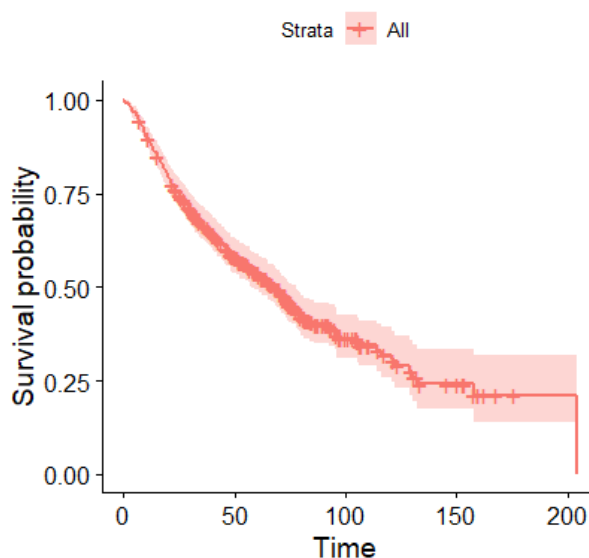
[output 解讀]

使用summary()函數，同樣的也有說明9筆資料被刪除，但會詳細列出每一筆感興趣事件發生時的存活函數計算結果，包含事件發生時的追蹤時間(存活時間)、在對應的追蹤時間之前，事件還未發生的人數、在對應的追蹤時間，事件發生的人數、存活率(或未發生率)…等

```
> summary(fit)
Call: survfit(formula = Surv(SURVIVAL_MONTHS, vital_status) ~ 1, data = lung_cancer)

因為不存在，9 個觀察量被刪除了
      time n.risk n.event survival std.err lower 95% CI upper 95% CI
0.03     469      1    0.998 0.00213    0.994    1.000
0.43     468      1    0.996 0.00301    0.990    1.000
0.53     467      1    0.994 0.00368    0.986    1.000
2.00     466      1    0.991 0.00425    0.983    1.000
2.24     465      1    0.989 0.00474    0.980    0.999
2.27     464      1    0.987 0.00519    0.977    0.997
2.40     463      1    0.985 0.00560    0.974    0.996
2.70     462      1    0.983 0.00598    0.971    0.995
3.25     461      1    0.981 0.00633    0.968    0.993
3.30     460      1    0.979 0.00667    0.966    0.992
```

[output] Kaplan-Meier 存活曲線圖



範例4-2、比較不同群體之間的存活曲線

這邊以範例檔中的癌症N分期作為分類，比較不同癌症N分期之間的存活曲線。

同樣的要先使用 **survival** 套件計算存活函數，再使用 **survminer** 套件繪製 **Kaplan-Meier 存活曲線圖**

> 使用survival套件中的 `survfit()` 函數計算存活函數，要注意的是，整體資料的存活函數計算在`survfit()`函數中~右邊是放1，若是比較不同群體之間的存活函數時，`survfit()`函數中~右邊要改放成分組依據的類別變項。

【語法】 `survfit(Surv(time, event) ~ var, data)`

其中 var 分組依據

[R程式碼]-

```
fit2 <- survfit(Surv(SURVIVAL_MONTHS, vital_status) ~ N_STAGE,
data = lung_cancer) # 使用 survfit 函數計算存活函數並存檔命名為 fit2
fit2                # 整筆檔案計算存活函數的資訊說明
summary(fit2)       # 列出每一筆感興趣事件發生時的存活函數計算結果
ggsurvplot(fit2)    # Kaplan-Meier 存活曲線圖
```


[output 解讀]

有15筆資料在存活函數計算過程被刪除，癌症N分期分成3個類別N0、N1和N1 or N2，分別呈現3個類別實際有使用到的資料筆數、感興趣事件發生的筆數和追蹤時間(或存活時間)中位數以及95%信賴區間下界和上界。

```
> fit2 <- survfit(Surv(SURVIVAL_MONTHS, vital_status) ~ N_STAGE, data = lung_cancer)
> fit2
Call: survfit(formula = Surv(SURVIVAL_MONTHS, vital_status) ~ N_STAGE,
  data = lung_cancer)
```

因為不存在，15 個觀察量被刪除了

	n	events	median	0.95LCL	0.95UCL
N_STAGE=N0	316	136	85.7	74.2	128.8
N_STAGE=N1	24	18	36.5	13.0	NA
N_STAGE=N1 or N2	123	95	28.2	22.3	40.2

[output 解讀]

使用summary() 函數，同樣的也有說明15筆資料被刪除，並詳細列出3種癌症N分期類別，每一筆感興趣事件發生時的存活函數計算結果。

```
> summary(fit2)
Call: survfit(formula = Surv(SURVIVAL_MONTHS, vital_status) ~ N_STAGE,
  data = lung_cancer)
```

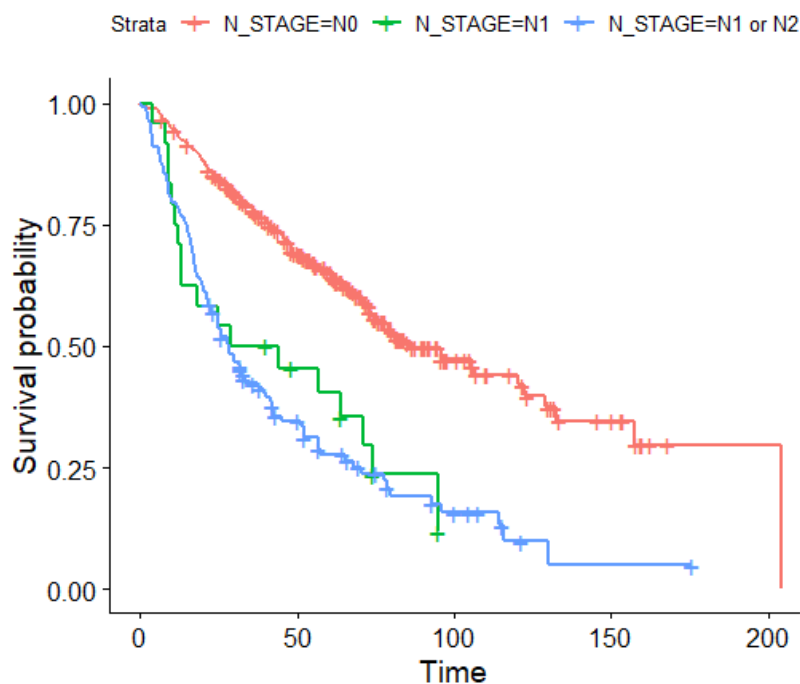
因為不存在，15 個觀察量被刪除了

N_STAGE=N0								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0.03	316	1	0.997	0.00316		0.991		1.000
0.43	315	1	0.994	0.00446		0.985		1.000
2.24	314	1	0.991	0.00546		0.980		1.000
4.00	313	1	0.987	0.00629		0.975		1.000

N_STAGE=N1								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
4	24	1	0.958	0.0408		0.8816		1.000
8	23	1	0.917	0.0564		0.8125		1.000
9	22	2	0.833	0.0761		0.6968		0.997
10	20	1	0.792	0.0829		0.6448		0.972
11	19	1	0.750	0.0884		0.5953		0.945
12	18	1	0.708	0.0928		0.5480		0.916
13	17	2	0.625	0.0988		0.4585		0.852

N_STAGE=N1 or N2								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0.53	123	1	0.9919	0.0081		0.97613		1.000
2.00	122	1	0.9837	0.0114		0.96164		1.000
2.27	121	1	0.9756	0.0139		0.94873		1.000
2.40	120	1	0.9675	0.0160		0.93664		0.999
2.70	119	1	0.9593	0.0178		0.92508		0.995
3.25	118	1	0.9512	0.0194		0.91390		0.990

[output] Kaplan-Meier 存活曲線圖



[補充說明]

在執行過程中，若有資料被刪除的情況，建議回頭檢查被刪除的原因，以確認是資料遺失值問題，還是資料整理過程中造成的資料錯誤。在範例4-1和範例4-2中，分別有9筆和15筆資料被刪除，使用到的變數為 SURVIVAL_MONTHS、vital_status 和 N_STAGE，檢查這3個變數原始資料的分佈情況。

[R程式碼] 檢查被刪除的資料

```
# ----- 步驟 1. 確認讀入的資料總筆數 -----
str(lung_cancer)

# ----- 步驟 2. 檢查原始資料的分佈 -----
# 連續型變數使用 summary() 函數檢查統計量
summary(lung_cancer$SURVIVAL_MONTHS)

# 類別變數使用 table() 函數檢查分佈
table(lung_cancer$vital_status)
table(lung_cancer$N_STAGE)
```

```
# 使用 is.na() 函數檢查遺失值情況分佈
sum(is.na(lung_cancer$vital_status))
sum(is.na(lung_cancer$N_STAGE))

# vital_status 和 N_STAGE 兩個變數一起檢查，呈現所有遺失值，其中 c(9,13,
  15) 為指定只呈現第 9、13、15 個變數，分別為 vital_status、N_STAGE 和 SU
  RVIVAL_MONTHS。
lung_cancer[is.na(lung_cancer$SURVIVAL_MONTH)*1==1 |
is.na(lung_cancer$N_STAGE)*1==1 ,c(9,13,15)]
```

[output] 使用 str() 檢查資料結構，可以看到讀的資料有 478 筆，15 個變數

```
> str(lung_cancer)
'data.frame': 478 obs. of 15 variables:
 $ SITE           : chr "DFCI" "DFCI" "DFCI" "DFCI" ...
 $ GENDER         : chr "Female" "Female" "Male" "Male" .
 $ gender         : int 0 0 1 1 0 1 0 0 1 0 ...
 $ AGE           : int 55 41 47 73 63 72 57 55 64 40 ...
 $ CHEMO         : chr "No" "No" "Yes" NA ...
 $ RT            : chr "No" "No" "No" NA ...
 $ rt            : int 0 0 0 NA NA NA NA NA NA NA ...
 $ Vital.Status   : chr "Alive" "Alive" "Alive" "Alive" .
 $ vital_status   : int 0 0 0 0 1 1 0 0 0 0 ...
 $ FIRST_PROGRESSION_OR_RELAPSE: chr "No" "Yes" "No" NA ...
 $ MONTHS_TO_FIRST_PROGRESSION : num NA 2 NA NA 17 5 NA NA NA NA ...
 $ SMOKING        : chr "Smoked in the past" "Smoked in t
 $ N_STAGE        : chr "NO" "NO" "NO" "NO" ...
 $ T_STAGE        : chr "T2 or T3" "T2 or T3" "T2 or T3"
 $ SURVIVAL_MONTHS : num 110 98 110 66 29 7 53 63 23 62 ..
```

[output] vital_status 存活狀態，這個變數類別為 0 和 1 (存活和死亡)，無遺失值。

```
> table(lung_cancer$vital_status)

 0    1
221 257
> sum(is.na(lung_cancer$vital_status))
[1] 0
```

[output] SURVIVAL_MONTHS 存活時間這個變數的統計量數值都合理，其中有 9 筆遺失值，此為範例 4-1 刪除 9 筆資料的原因。

```
> summary(lung_cancer$SURVIVAL_MONTHS)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.   NA's
  0.03  23.20   45.80   51.78  73.20  204.00     9
```

[output] N_STAGE 癌症N分期，這個變數有3個類別，其中有11筆遺失值

```
> table(lung_cancer$N_STAGE)
      N0      N1 N1 or N2
      317      24      126
> sum(is.na(lung_cancer$N_STAGE))
[1] 11
```

[output] SURVIVAL_MONTHS 存活時間和 N_STAGE 癌症N分期兩個變數一起檢查，列出有遺失值的筆數共15筆，此為範例4-2刪除15筆資料的原因。

```
> lung_cancer[is.na(lung_cancer$SURVIVAL_MONTH)*1==1 |
+ is.na(lung_cancer$N_STAGE)*1==1 ,c(9,13,15)]
  vital_status N_STAGE SURVIVAL_MONTHS
87           1    <NA>          22.74
98           0    <NA>          106.58
101          1    <NA>           30.72
104          1    <NA>           13.33
113          0    <NA>             NA
120          1 N1 or N2             NA
134          0    <NA>             NA
138          0    <NA>           84.44
163          1    <NA>           38.77
166          0    <NA>             NA
167          1 N1 or N2             NA
168          1    <NA>             NA
176          1      N0             NA
180          0    <NA>             NA
210          0 N1 or N2             NA
```

經過以上檢定過程，確認都是遺失值造成的刪除，若是整理過程中造成的資料錯誤，例如：時間長度為負數，表示資料有問題或整理過程的失誤，此時就要修正錯誤後再重新執行。

範例4-3、客製化存活曲線

> ggsurvplot() 函數中包含許多可以調整圖形呈現的指令，在沒有另外指定的情況下，都會以預設值的設定，只呈基本的 Kaplan-Meier 存活曲線圖形，接下來透過調整 ggsurvplot() 函數中的指令，美化或在圖形中添加更方便判讀的相關資訊。

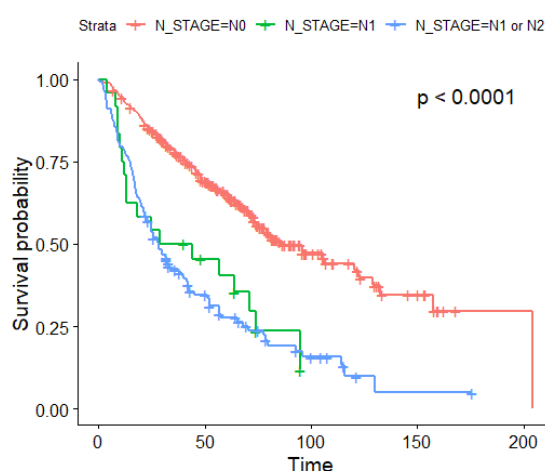
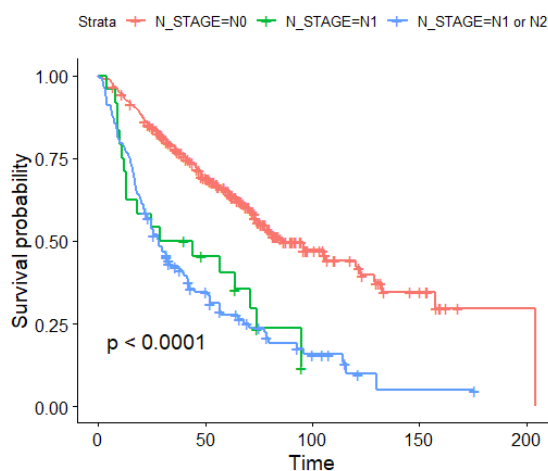
[R程式碼] 圖形中添加相關資訊

```
#(a) 增加 p value (log-rank test)
ggsurvplot(fit2, pval = TRUE, data = lung_cancer)

#(b) 增加 p value 且指定 p value 呈現位置, 預設呈現位置在左下方
ggsurvplot(fit2, pval = TRUE, pval.coord = c(150, 0.95), data = lung_cancer)
```

(a)

(b)

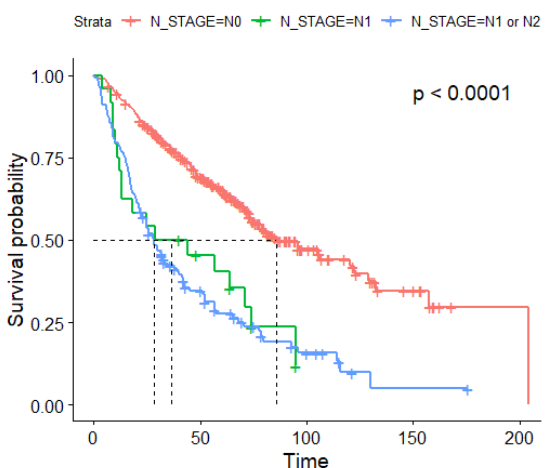
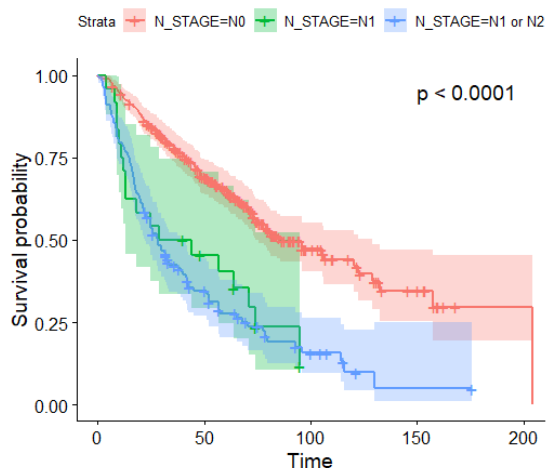


```
#(c) 增加 95%信賴區間
ggsurvplot(fit2, pval = TRUE, pval.coord = c(150, 0.95),
            conf.int = TRUE, data = lung_cancer)
```

```
#(d) 標註存活時間中位數
ggsurvplot(fit2, pval = TRUE, pval.coord=c(150, 0.95),
            surv.median.line = "hv", data = lung_cancer)
```

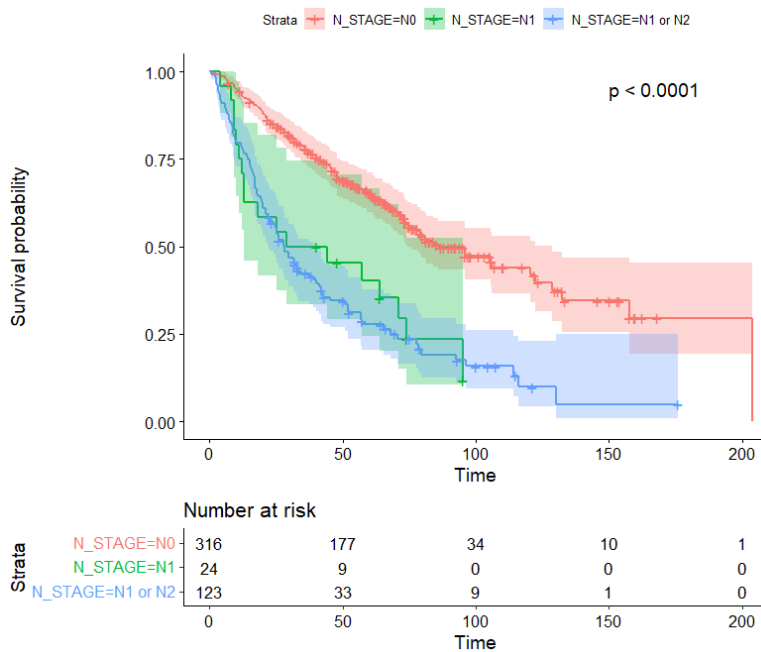
(c)

(d)



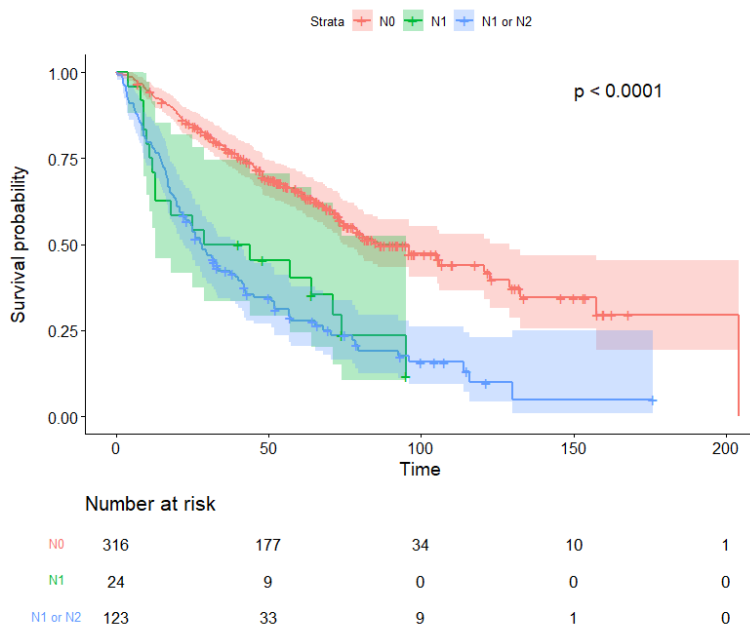
(e) 增加風險人數表格

```
ggsurvplot(fit2, pval = TRUE, pval.coord = c(150, 0.95),
            conf.int = TRUE, risk.table = TRUE,
            data = lung_cancer)
```



(f) 風險人數表格調整，指定圖例名稱、風險表格去除框線和 y 軸名稱

```
ggsurvplot(fit2, pval = TRUE, pval.coord = c(150, 0.95),
            conf.int = TRUE, risk.table = TRUE,
            legend.labs = c("N0", "N1", "N1 or N2"),
            tables.theme = clean_theme(), data = lung_cancer)
```



[R程式碼] 格式調整

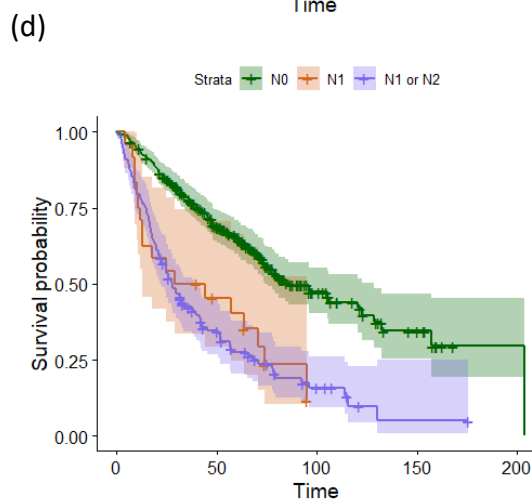
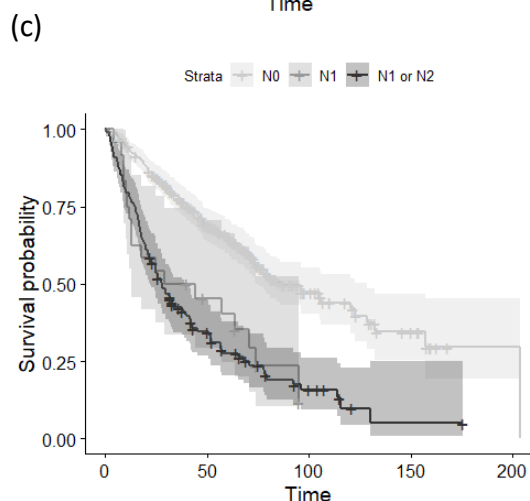
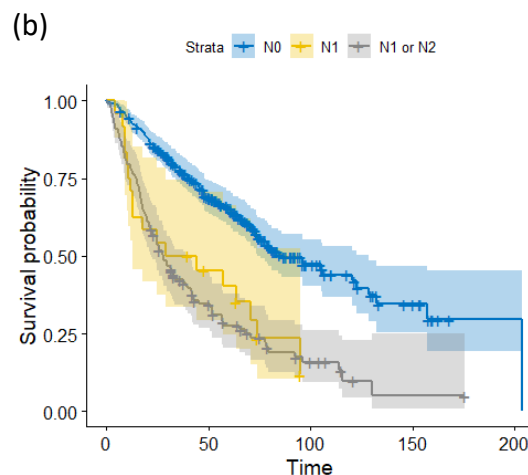
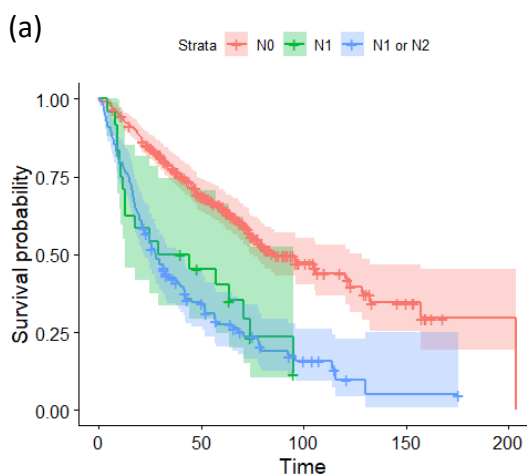
◇ 顏色設定

- `palettes = c(color1, color2, ..)` 依分組數給定顏色，顏色代碼請參考色碼表，或指定搭配好顏色的調色盤代碼，預設的調色盤顏色為"hue"。提供選擇的調色盤顏色有"grey", "npg", "aaas", "lancet", "jco", "ucscgb", "uchicago", "simpsons"和"rickandmorty"

```
# (a) 調色盤預設顏色
ggsurvplot(fit2, conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"))

# (b) 調色盤顏色"jco"、(c) 調色盤顏色"grey"
ggsurvplot(fit2, conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"), palette = "jco")
ggsurvplot(fit2, conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"), palette = "grey")

# (d) 使用色碼表自行指定顏色
ggsurvplot(fit2, conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            palette = c("#006400", "#D2691E", "#7B68EE"))
```



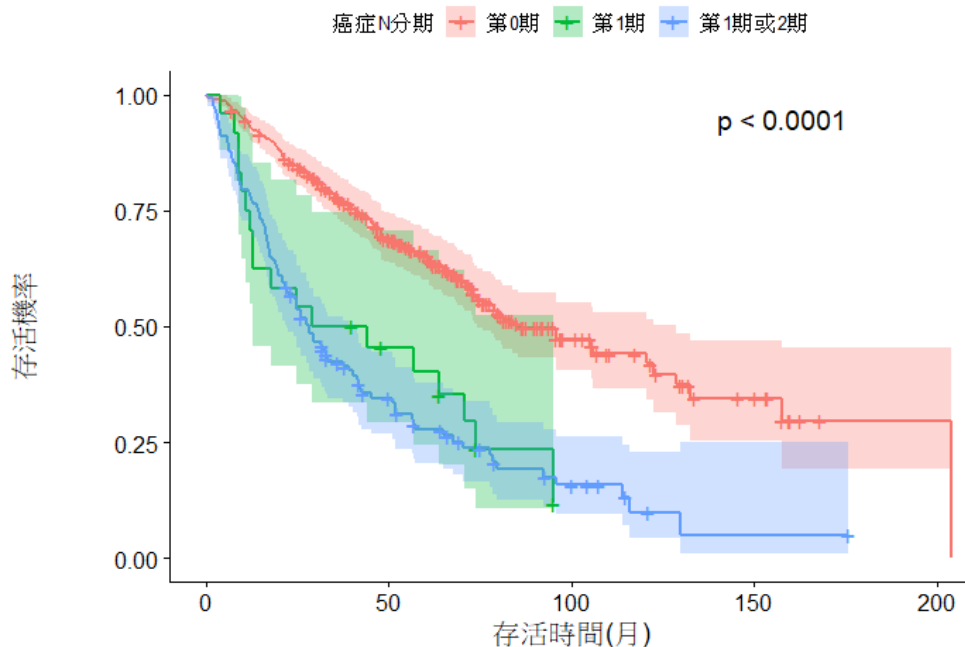
◇ 標題設定

- title = "", 指定圖形標題
- xlab = "", 指定X軸標題
- ylab = "", 指定Y軸標題設定
- legend.title = "", 指定圖例標題
- legend.labs = c("lab1", "lab2", ..) , 指定圖例分組名稱

```

ggsurvplot(fit2, conf.int = TRUE, data = lung_cancer,
            pval=TRUE, pval.coord=c(140, 0.95),
            risk.table = TRUE,
            title = "癌症 N 分期的存活曲線比較",
            xlab = "存活時間(月)",
            ylab = "存活機率",
            legend.title = "癌症 N 分期",
            legend.labs = c("第 0 期", "第 1 期", "第 1 期或 2 期"))
    
```

癌症N分期的存活曲線比較



Number at risk

癌症N分期	0	50	100	150	200
第0期	316	177	34	10	1
第1期	24	9	0	0	0
第1期或2期	123	33	9	1	0

存活時間(月)

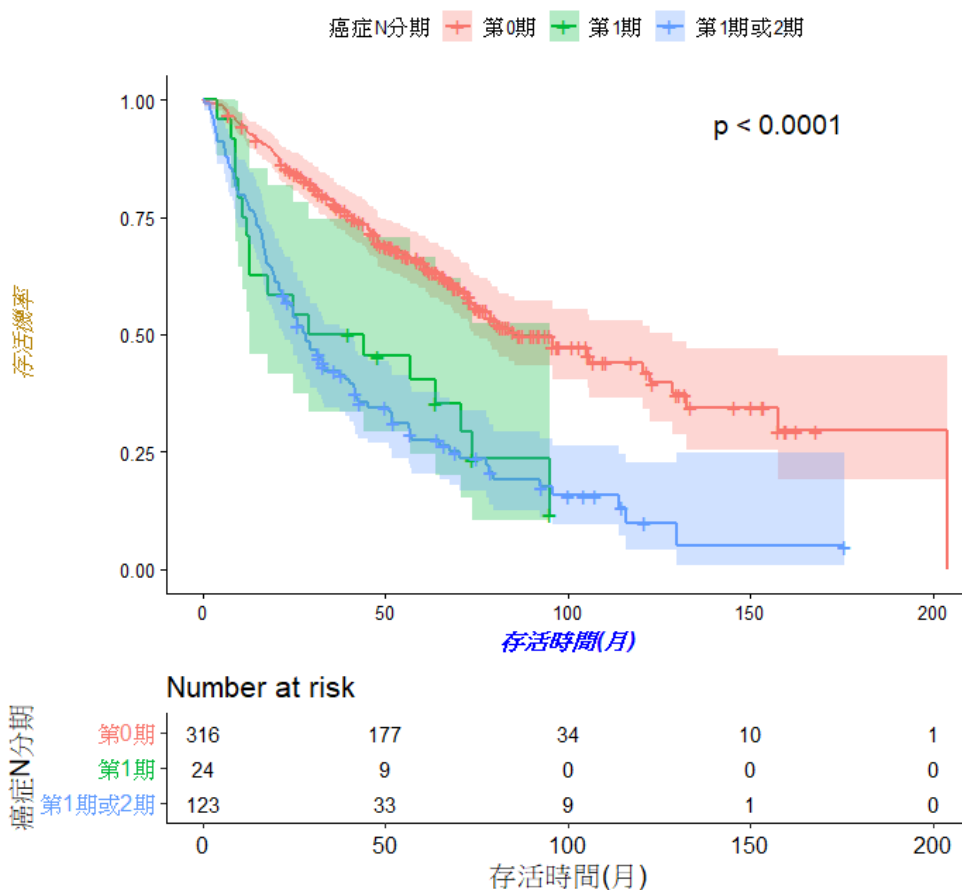
- ◇ 字型設定，依序指定大小、類型和顏色，不指定則顯示預設值
 - font.main = c(16,"bold","red")，指定圖形標題字體樣式
 - font.x = c(14,"italic","blue")，指定X軸標題字體樣式
 - font.y = c(14,"bold.italic","gray")，指定Y軸標題字體樣式
 - font.tickslab = c(12, "plain","green")，指定刻度標籤字體樣式
 - fontsize = 數值 指定風險表和累積事件表的字體大小

也可以只設置其中一項，例如：

- font.main = 16，只設定字體大小；
- font.main = "italic"，只設定字體類型

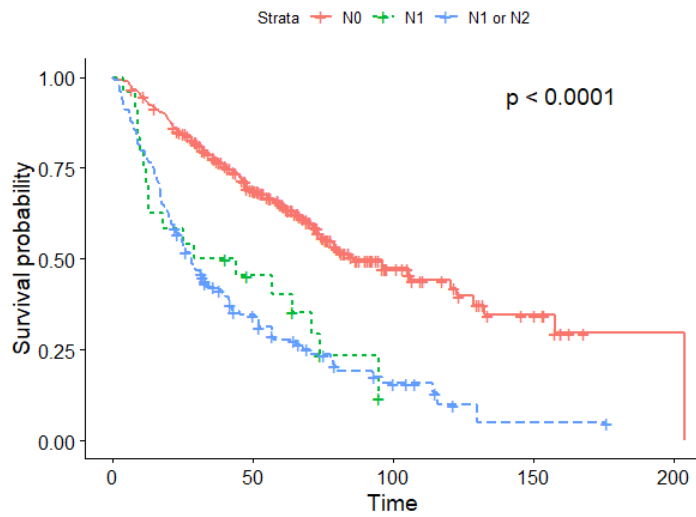
```
ggsurvplot(fit2, conf.int = TRUE, data = lung_cancer,
            pval=TRUE, pval.coord=c(140, 0.95), risk.table = TRUE,
            title = "癌症 N 分期的存活曲線比較",
            xlab = "存活時間(月)", ylab = "存活機率",
            legend.title = "癌症 N 分期",
            legend.labs = c("第 0 期", "第 1 期", "第 1 期或 2 期"),
            font.main = c(14,"bold","red"),
            font.x = c(12,"bold.italic","blue"),
            font.y = c(12,"italic","darkgoldenrod"),
            font.tickslab = 10, fontsize = 4 )
```

癌症 N 分期的存活曲線比較

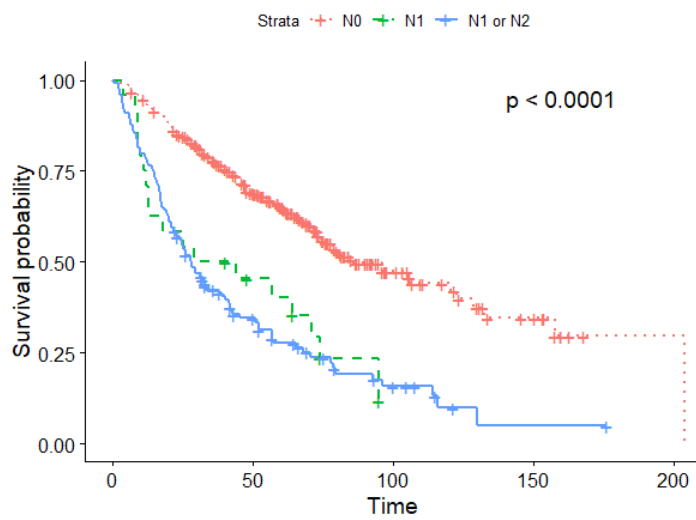



```
legend.labs = c("N0", "N1", "N1 or N2"),  
linetype = "solid")
```

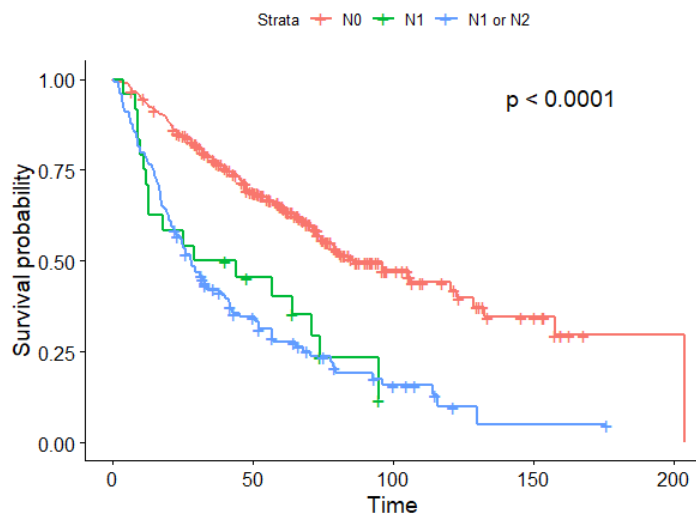
(a)



(b)



(c)



◇ 坐標軸範圍、刻度間距設定

- `xlim = c(數值1, 數值2)` 指定X軸起始和結束範圍
- `ylim = c(數值1, 數值2)` 指定Y軸起始和結束範圍
- `break.x.by = 數值` 指定X軸該度間距
- `break.y.by = 數值` 指定Y軸該度間距

(a) 預設不調整的圖形

```
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month)")
```

(b) 指定 X 軸 Y 軸範圍

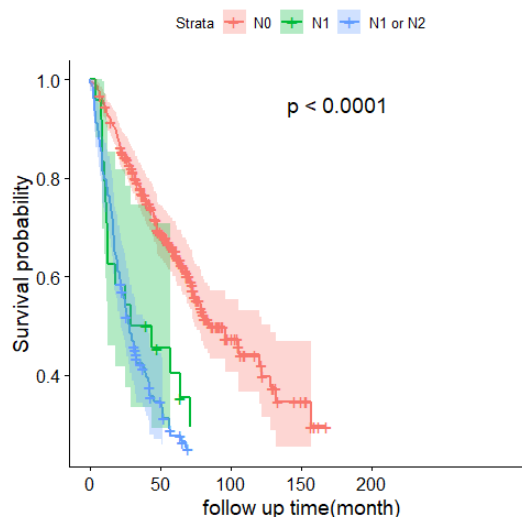
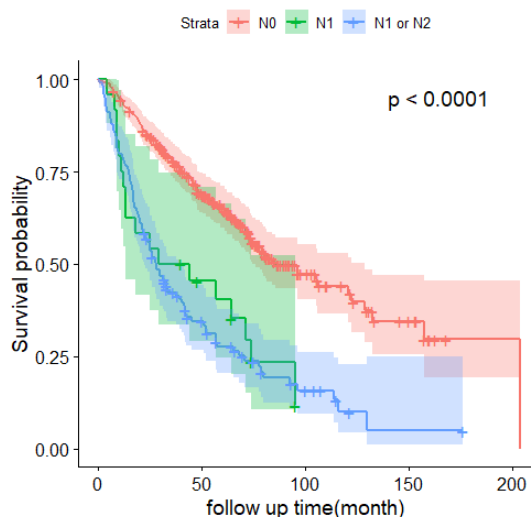
```
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month)",
            xlim = c(0, 300), ylim = c(0.25, 1))
```

(c) 指定 X 軸 Y 軸該度間距

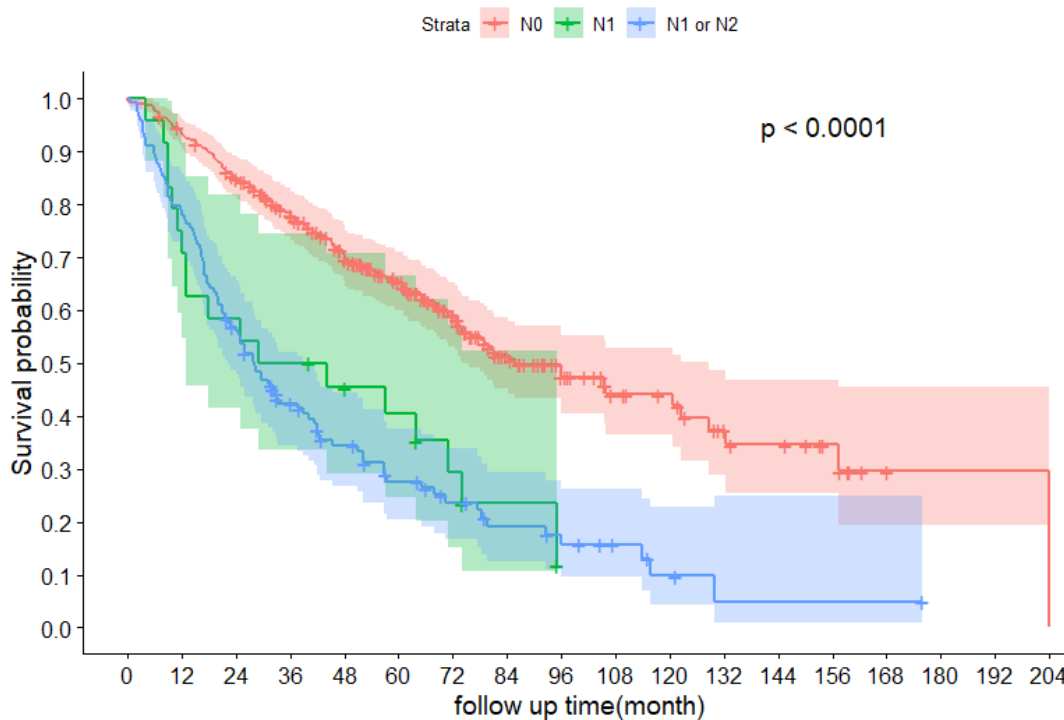
```
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month)",
            break.x.by = 12, break.y.by = 0.10 )
```

(a)

(b)



(c)



◇ 圖例位置設定

- legend = "", 圖例位置設置，提供選擇有"top", "bottom", "left", "right" 和 "none"等，其中"none"為不顯示圖例，預設值為"top"。

#(a) 圖例位置在上方。預設值，可寫可不寫

```
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month) ")
```

#(b) 不顯示圖例

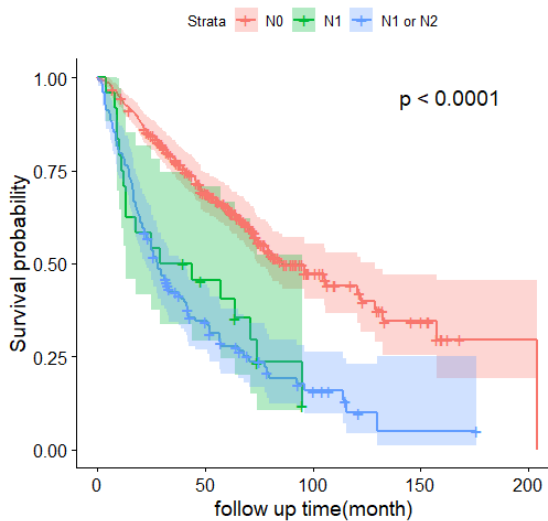
```
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month) ",
            legend = "none")
```

#(c) 圖例位置在右方

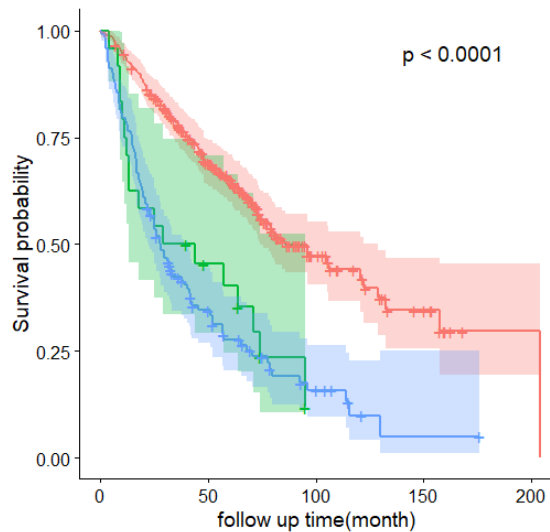
```
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month) ",
            legend = "right" )
```

```
#(d) 圖例位置在下方
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            legend = "bottom")
```

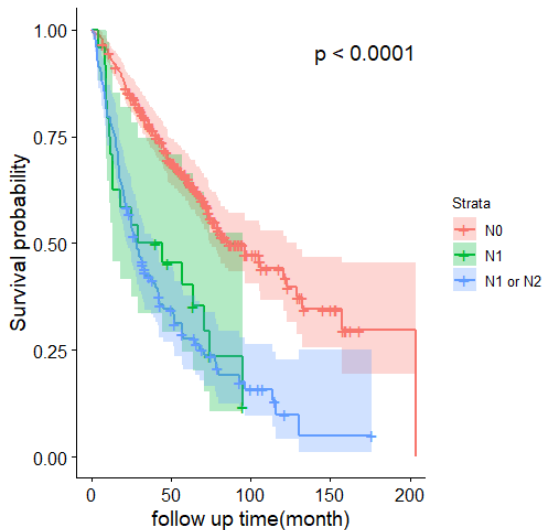
(a)



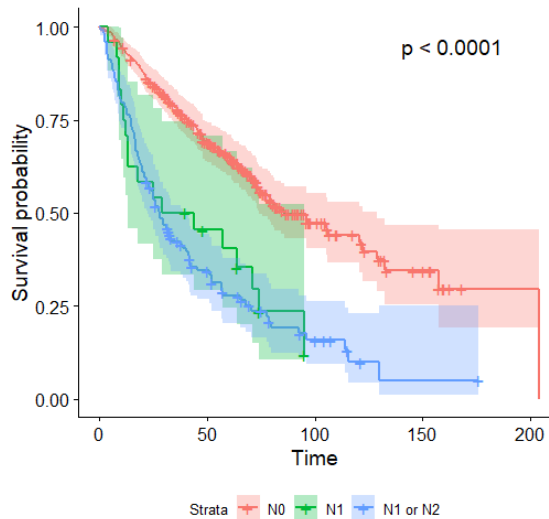
(b)



(c)



(d)

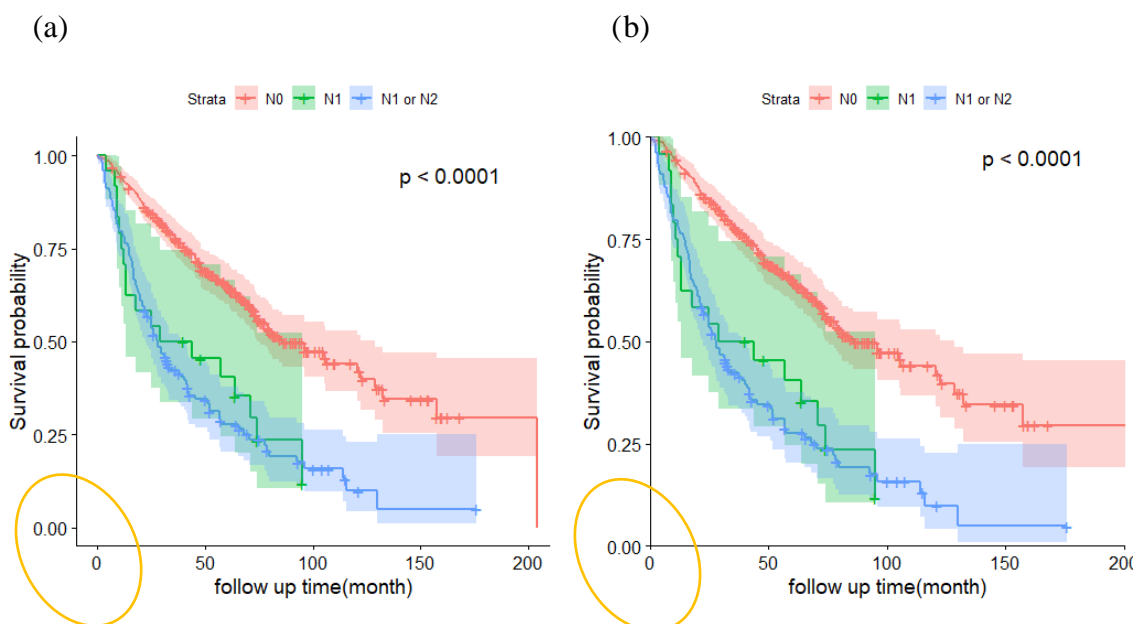


◇ 其他設定

- axes.offse，指定坐標軸從原點開始 (FALSE)，預設為 TURE 不從原點開始。

```
# (a) 坐標軸從原點開始
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month)")

# (b) 坐標軸不從原點開始
ggsurvplot(fit2, pval=TRUE, pval.coord=c(140, 0.95),
            conf.int = TRUE, data = lung_cancer,
            legend.labs = c("N0", "N1", "N1 or N2"),
            xlab = "follow up time(month)",
            axes.offse = FALSE )
```



以上是使用 `survminer` 套件中的 `ggsurvplot()` 函數所繪製的Kaplan-Meier存活曲線，除了基本圖形的呈現，還可以透過指令美化調整成符合需求的圖形，除了上述的範例外，還其它更多的指令可進行調整，此篇僅介紹比較常使用到的相關調整指，提供初學者入門參考。

參考資料

1. 存活分析, 林建甫著, 初版, 2008, 雙葉書局。
2. <https://rpkgs.datanovia.com/survminer/index.html>.
3. survminer R package: Survival Data Analysis and Visualization