第十八單元　存活分析(一)--基本概念

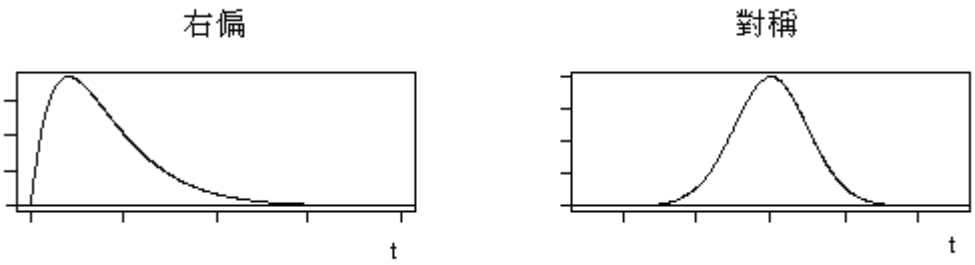
梁文敏 教授 葉懿諄 統計分析師

18.1 存活分析基本概念

存活分析資料的特質

存活分析是在醫學統計中常用的一種分析，它的主要結果變數為時間(time)，例如：癌症患者手術後的存活時間、參與戒菸計畫者的戒菸持續時間。這類存活時間(time to event)不同於一般連續變數，主要特性有二：

(1)首先，這類資料通常並不會服從常態分布，例如若要探討手術後的存活時間，通常此時間的分布並不會呈對稱分布，這些個案的存活時間均較短，因為手術後數月內病人處於高危險期，死亡機率最高，因此這類資料通常會呈右偏的分布，所以平均值通常並不能適切地代表資料的中央趨勢，而以中位數(median)取而代之。



(2)其次，這類時間資料經常有些是無法獲得的，例如，要探討患者術後的存活時間，若研究時間為5年則研究結束時，那些存活的患者我們無法確知其實際的存活時間，只能說他們至少活了5年。因此，我們常將這類資料分為兩類：**(a)第一類針對事件已發生的個案，其資料稱之為完整資料 (complete data)**：其存活時間的計算是從起始點(initial point)到事件(event)發生點的時間。一般事件起始點常是手術日、疾病診斷日或是納入研究的時間，而事件在醫學上常被定義為死亡、復發等。**(b)第二類係針對事件未發生的個案數，其資料稱之為設限資料 (censored data)**：其時間的計算是從起始點(initial point)到可以追蹤到的時間(follow-up time)，在這時間內事件(event)都沒有發生，我們無法

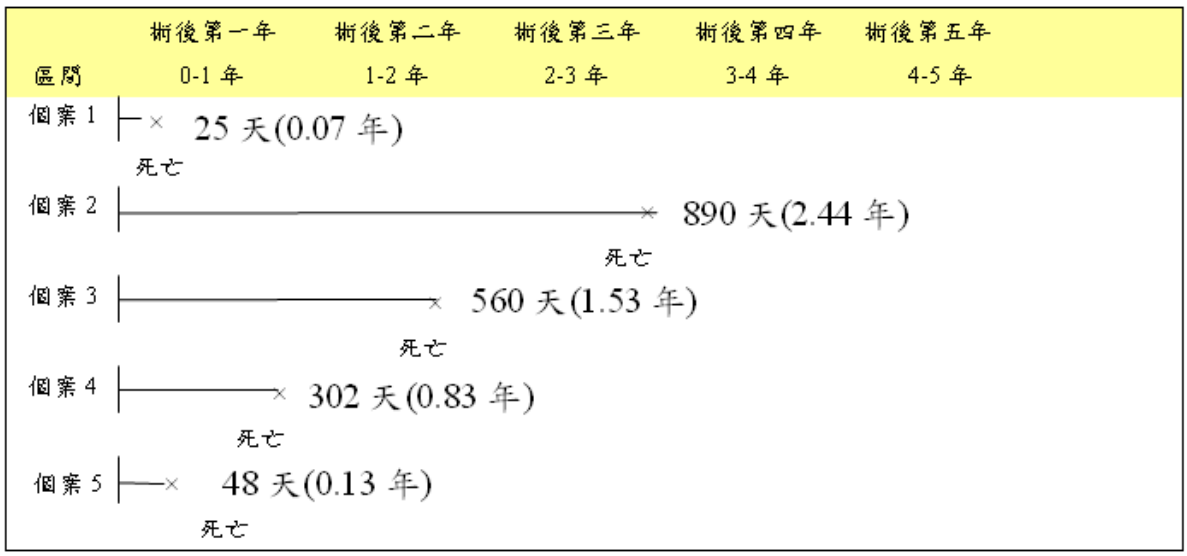
知道他們從起始點到事件(死亡)發生的時間，我們僅知道他們至少存活多少時間。設限資料的個案常見為失去追蹤者或是一直到研究結束仍存活的患者，故無法確知其實際的存活時間。

因此，若用平均值來代表存活時間的集中趨勢常是有偏差的，因為其分布大都是偏斜的並經常有設限資料，而中位數較適合用來描述其中央趨勢。因此，此類資料有其特色，在真實的世界中，我們很難都獲得每個個案的存活時間，故需要利用存活分析來處理這類偏斜資料及不完整觀察值的問題。

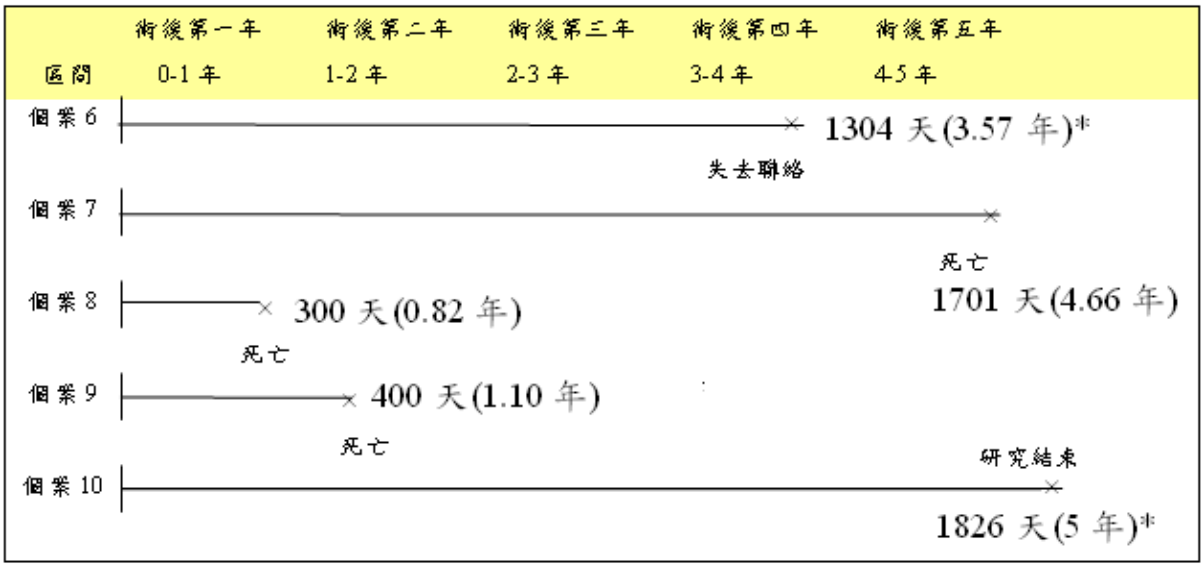
範例資料

在本範例中我們主要探討A與B兩種手術療法對於某癌症治療存活時間的比較，研究期間共5年。假設A手術為目前主要的治療方式，B手術為新的治療方式，總收案數為10人，被隨機分派至兩組，每組5人，其存活情況如下。

(a) **A療法組(對照組)**：編號為1至5，個案1手術後共存活了25天，個案2存活了890天，個案3存活了560天，個案4存活了302天，個案5存活了48天，這五名個案都在研究期間內死亡，故知其存活時間，此5筆資料皆為完整資料。



(b) **B療法組(實驗組)**：編號為6到10，個案6於手術後追蹤了1304天，後即失去聯絡，個案7存活了1701天，個案8存活了300天，個案9存活了400天，個案10至5年研究結束後仍存活，亦即個案10至少術後存活了1826天，故此組5名個案中有3名個案(個案7、個案8及個案9)的存活資料為完整資料、有2名(個案6及個案10)為設限資料。



【範例】計算 A 及 B 兩種療法存活時間的中央趨勢。

範例資料：

茲將兩組的資料整理如下，其中，id 表示個案編號，sex 表示性別(0: 女生, 1: 男生)，group 表示治療組(0: A 療法, 1: B 療法)，s_day 表示存活天數(單位：天)，s_year 表示存活年數(單位：年)，s_censor 表示資料類型(1: 完整, 0:設限)。

A 療法組(對照組)

id	sex	group	s_day	s_year	s_censor
1	1	0	25	0.07	1
2	0	0	890	2.44	1
3	0	0	560	1.53	1
4	1	0	302	0.83	1
5	1	0	48	0.13	1

B 療法組(實驗組)

id	sex	group	s_day	s_year	s_censor
6	0	1	1304	3.57	0
7	0	1	1701	4.66	1
8	0	1	300	0.82	1
9	1	1	400	1.1	1
10	0	1	1826	5	0

使用方法：

茲分別計算兩組存活時間的平均值及中位數。

(一) 不考慮設限情形：

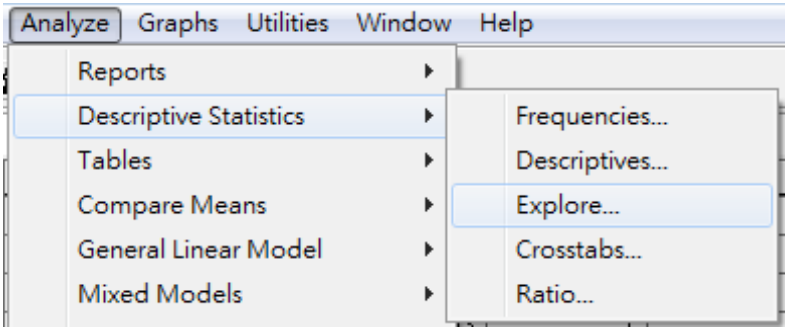
軟體操作步驟：

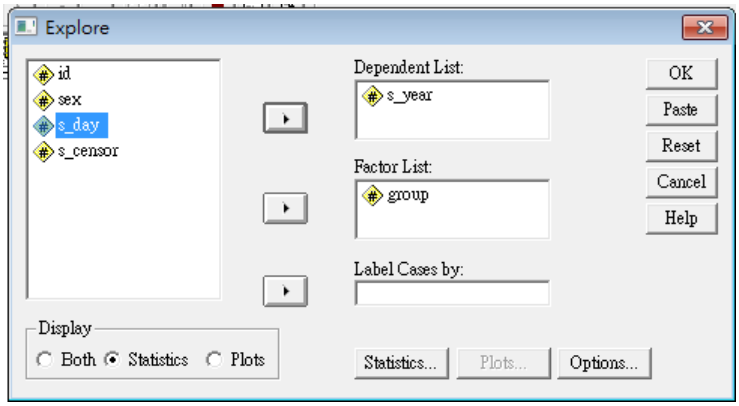
【SPSS步驟】

步驟一：選擇功能【Analysis分析】→進入【Descriptive Statistics敘述統計】→進入【Explore預檢資料】。

步驟二：在Dependent List依變數清單(選入變數：s_year)→在Factor List因子清單(選入變數：group)。

步驟三：在框格左下方的Display中，點選：Statistics統計量→ 選擇OK確定。





輸出結果：

Descriptives

	group			Statistic	Std. Error
s_year	0	Mean		1.0000	.44750
		95% Confidence Interval for Mean	Lower Bound	-.2425	
			Upper Bound	2.2425	
		5% Trimmed Mean		.9717	
		Median		.8300	
		Variance		1.001	
		Std. Deviation		1.00065	
		Minimum		.07	
		Maximum		2.44	
		Range		2.37	
		Interquartile Range		1.89	
		Skewness		.693	.913
		Kurtosis		-.893	2.000
	1	Mean		3.0300	.87859
		95% Confidence Interval for Mean	Lower Bound	.5906	
			Upper Bound	5.4694	
		5% Trimmed Mean		3.0433	
		Median		3.5700	
		Variance		3.860	
		Std. Deviation		1.96459	
		Minimum		.82	
		Maximum		5.00	
		Range		4.18	
		Interquartile Range		3.87	
		Skewness		-.321	.913
		Kurtosis		-2.971	2.000

上表分別計算了 A 療法(group=0)及 B 療法(group=1)的描述性統計。A 療法：平均值為 1 年，中位數為 0.83 年；B 療法：平均值為 3.03 年，中位數為 3.57 年。用此方法所計算的結果，若有設限資料存在則平均值及中位數都會有誤差，尤其是平均值可能產生的誤差更大。例如：假設 B 療法中個案 6 及個案 15 都在手術後存活了 15 年，則上述所計算出來的平均值顯然低估了許多。

(二) 考慮設限情形：

以上是不考慮資料是否有設限情形來計算，下列我們針對存活時間資料(s_year)，再加入設限情形(s_censor)。

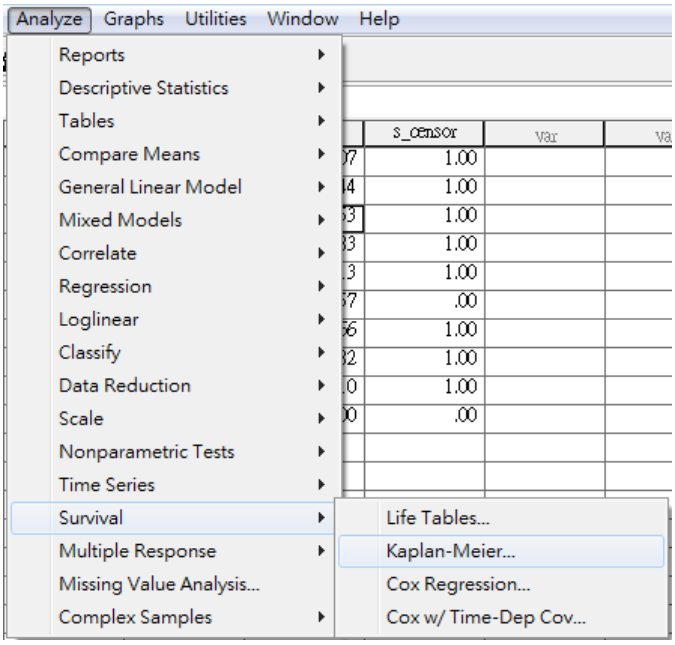
軟體操作步驟：

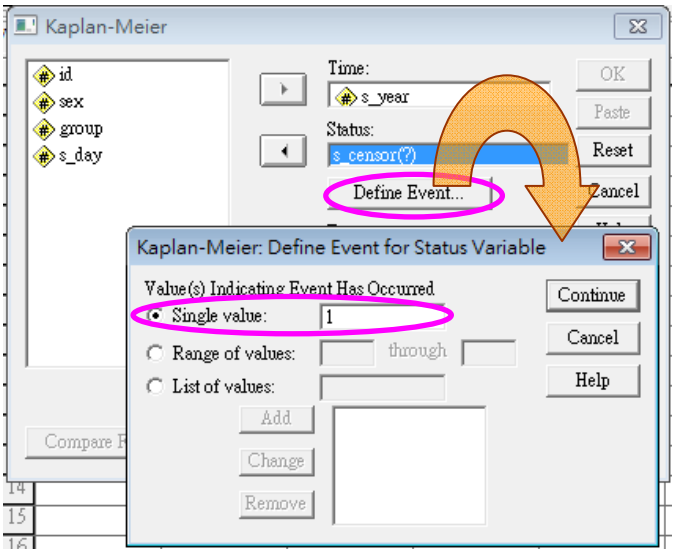
【SPSS步驟】

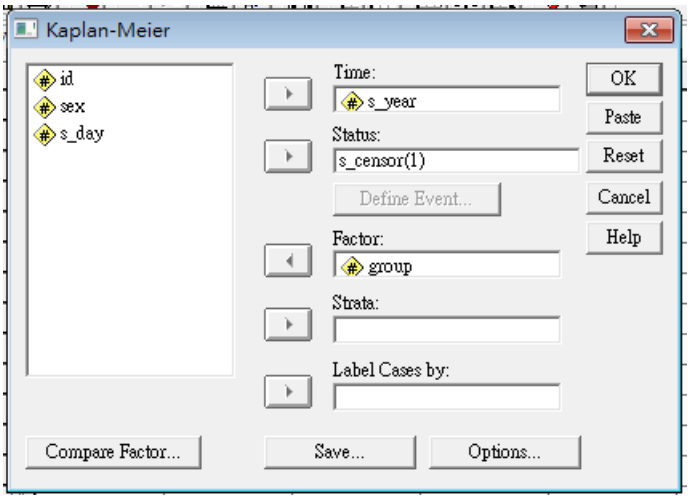
步驟一：選擇功能【Analyze分析】→進入【Survival存活分析】→進入【Kaplan-Meier統計】。

步驟二：在Time時間框格中(選入變數：s_year)→在Status狀態框格中(選入變數：s_censer)→點選下方的Event Defined定義事件（選擇：Single value單一數值：1）→Continue繼續。

步驟三：在Factor因子框格中(選入變數：group) → Continue繼續→OK確定。


Analyze
Graphs
Utilities
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Life Tables...
Kaplan-Meier...
Cox Regression...
Cox w/ Time-Dep Cov...
s_censor
var


Kaplan-Meier
id
sex
group
s_day
Time:
s_year
Status:
s_censor(?)
Define Event...
OK
Paste
Reset
Cancel
Kaplan-Meier: Define Event for Status Variable
Value(s) Indicating Event Has Occurred
Single value:
1
Range of values:
through
List of values:
Add
Change
Remove
Continue
Cancel
Help


Kaplan-Meier
id
sex
s_day
Time:
s_year
Status:
s_censor(1)
Define Event...
Factor:
group
Strata:
Label Cases by:
OK
Paste
Reset
Cancel
Help
Compare Factor...
Save...
Options...

輸出結果：

Means and Medians for Survival Time

	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
group	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
.00	1.000	.448	.123	1.877	.830	.767	.000	2.333
1.00	3.282	.852	1.612	4.952	4.660	2.833	.000	10.213
Overall	2.141	.593	.979	3.303	1.100	.553	.015	2.185

a. Estimation is limited to the largest survival time if it is censored.

從上表中可以知道A療法(group=0)的存活時間的平均值為1.00年、中位數為0.83年，B療法(group=1)的平均值為3.28年、中位數為4.66年，與不考慮設限資料在B療法中所得到的結果是不同的(不考慮設限B療法的平均值為3.03年，中位數為3.57年)。根據存活分析模組其平均值及中位數均是根據存活模型所估算出來的，整體而言，中位數較平均值來的穩健(robust)，更能代表實際資料得中央趨勢。