第5章

生存分析

f 究者在临床研究中,有时不仅仅关心结局如何,还需要观察出现这种结局所持续的 时间。例如在对肿瘤的治疗与观察研究中,对结局都是死亡的患者,生存的时间长 短不同,说明患者的预后是有差别的。此时收集的数据资料应包含结局和时间两方面的信 息,称为生存时间资料。对这类资料的研究应采用与生存分析相关的数据分析方法。

5.1 生存分析的基本概念

生存分析(Survival Analysis)是将事件(如死亡、生存等)的结果(发生死亡、存活) 与出现这个结果的时间长短结合在一起进行分析的方法。常用方法是生存概率(生存率) 的计算,其原理是概率乘积原理。

生存分析主要包含两个方面的研究:第一,描述生存过程,即研究生存时间的分布特点,计算某个时间点的生存率,生存曲线的变动趋势;第二,生存过程的影响因素分析(Cox 回归)。生存分析中主要包括参数法(威布尔回归模型)、非参数法和半参数法(Cox 比例 风险模型)。

5.2 生存分析的资料特点

生存分析中随访数据包括两类数据:一类是完全数据(Complete Data),即从起点至死 亡所经历的时间;另一类是截尾数据(Censored Data),即生存时间被人为地截止,其原因 包括失访、退出和终止。

生存分析的主要研究方法如下。

- (1) 生存率的估计与生存曲线。
- (2) 生存曲线的比较。
- (3) 生存过程的影响因素分析(Cox回归)。

生存时间资料有如下3个明显的特点。

(1) 随访数据包括最基本的两个变量: 生存时间和结局。

(2) 存在截尾数据(常用符号+表示)。

(3) 分布类型复杂,分布常呈偏态。

表 5.1 给出的是一个典型的生存时间资料。

また	协会	变量		观	察记录		生存时间
灶石	性别	手术	开始	终止	结局	原因	/ d
张	1	0	00-07-08	00-11-27	0	失访	142+
王	0	1	00-07-10	00-12-15	1	复发死亡	158
李	1	1	00-07-16	00-12-31	0	生存 (研究截止)	168+
赵	0	0	00-08-18	00-11-22	1	复发死亡	96
孙	1	1	00-10-10	00-11-12	1	转移死亡	33

表 5.1 生存时间资料描述

其中,性别变量中的1表示男性,0表示女性;手术变量中的1表示实施了手术,0表 示未实施手术;+表示截尾数据。生存时间为终止时间点与起始时间点之间的天数。该资料 包含生存时间和结局,存在截尾数据,符合生存时间资料的特点。

5.3 生存资料的分析方法

生存资料的分析方法主要包括两种:一种为非参数方法,其中包括适合于大样本生存分析的寿命表法(Life Table)和适合于小样本生存分析的 Kaplan-Meier 法。另一种是半参数法的 Cox 回归,常用于进行生存过程的影响因素分析。

5.4 应用实例1:累积生存率的计算

累积生存率的计算过程主要有如下步骤。

- (1) 计算校正人口数:校正人口数=年初人口数-(截尾例数/2)。
- (2) 计算死亡概率: 死亡概率=年内死亡数/校正人口数。
- (3) 计算生存概率: 生存概率=1-死亡概率。

(4) 计算生存率: 生存率(累积生存概率)=截止到t时刻各时段生存概率的乘积。 累积生存率的计算过程如表 5.2 所示。

下面采用 SPSS 软件对上述数据进行分析,将该表中的数据进行录入,如图 5.1 所示。

术后年数	期初人数	死亡人数	截尾人数	校正人数	死亡概率	生存概率	生存率
0	233	68	8	229.0	0.2969	0.7031	0.7031
1	157	61	7	153.5	0.3974	0.6026	0.4237
2	89	38	3	87.5	0.4343	0.5657	0.2397
3	48	16	1	47.5	0.3368	0.6632	0.1589
4 ~ 5	31	8	23	19.5	0.4103	0.5897	0.0937

表 5.2 累积生存率的计算过程

时间	人数	类型
0.00	68.00	1.00
1.00	61.00	1.00
2.00	38.00	1.00
3.00	16.00	1.00
4.00	8.00	1.00
0.00	8.00	0.00
1.00	7.00	0.00
2.00	3.00	0.00
3.00	1.00	0.00
4.00	23.00	0.00

图 5.1 累积生存率计算的数据录入界面

对人数进行加权后,选择菜单 Analyze→Survival→Life Tables,如图 5.2 所示。

<u>V</u> iew	<u>D</u> ata	∏ransform	Analyze	<u>G</u> raphs	Utilities	<u>W</u> indow	Help	
<u> </u>	5	<u> </u>	Re <u>p</u> i D <u>e</u> so	orts riptive Stat	istics 🔸	0		
0			Com	pare <u>M</u> eans	s 🔸			
H	计间	人数	<u>G</u> ene	eral Linear N	1odel 🕨	-	var	var
	0.00	68	<u>C</u> orr	elate	•			
-	1.00	61	Reg	ession	•			
-	2.00	38	Logi	near	*			
	3.00	16	Clas:	SITY Reduction				
-	4.00	8	Scale	i Neducuon i				
-	0.00	8	Non	barametric`	Tests 🔸			
	1.00	7	Sur∖	ival	•	Life Tab	oles	
-	2.00) 3	Muļti	ple Respon	se 🔸	Kaplan-	Meier	
	3.00	1	.00	0.00		<u>C</u> ox Re	gression	•
	4.00	23	.00	0.00		C <u>o</u> x w/	' lime-Dep	
24		5						

图 5.2 累积生存率分析菜单

在弹出的如图 5.3 所示的对话框"Life Tables"中,单击"Status"框下面【Define Event...】 按钮,打开"Life Tables: Define Event Status"子对话框,并在"Single value"中输入"1", 单击【Continue】按钮返回"Life Tables"主对话框。

Life Tables		×	
●人数	Lime: ④ 时间 Display Time Intervals 0 through 4 by 1 Status: 受型fine Event Eactor: Define Range By Factor: Define Range	OK Paste Beset Cancel Help Life Tables: Define Event Value(s) Indicating Event Has Occurred © Single values 1 Range of yalues: phions	for Statu 🗙 Continue Cancel Help

图 5.3 累积生存率分析主对话框及其子对话框

单击【OK】按钮后的输出结果如图 5.4 所示。

This su	bfile cc	ntains:	10	observat	ions				
Life T Surv	able ival Var	iable H	小问						
	Number	Number	Number	Number			Cumul		
Intrvl	Entrng	Wdrawn	Exposd	of	Propn	Propn	Propn	Proba-	
Start	this	During	to	Termnl	Termi-	Sur-	Surv	bility	Hazard
Time	Intrvl	Intrvl	Risk	Events	nating	viving	at End	Densty	Rate
0.0	233.0	8.0	229.0	68.0	0.2969	0.7031	0.7031	0.2969	0.3487
1.0	157.0	7.0	153.5	61.0	0.3974	0.6026	0.4237	0.2794	0.4959
2.0	89.0	3.0	87.5	38.0	0.4343	0.5657	0.2397	0.1840	0.5547
3.0	48.0	1.0	47.5	16.0	0.3368	0.6632	0.1589	0.0807	0.4051
4.0+	31.0	23.0	19.5	8.0	0.4103	0.5897	0.0937	**	**

图 5.4 输出结果

在如图 5.4 所示的输出结果中,"Cumul Propn Surv at End"表示的就是累积生存率。输出结果中还有一句"The median survival time is 1.7268",该句表明中位生存时间是 1.7268,即病人死亡人数达到一半的时间为 1.73 年。

在如图 5.3 所示的主对话框 "Life Tables"中单击【Options】按钮, 弹出如图 5.5 所示的对话框 "Life Tables: Options", 在 "Plot"复选框中勾选 "Survival", 单击【Continue】 按钮可以绘制出生存曲线。

绘制出的累积生存曲线如图 5.6 所示。



5.5 应用实例 2: 小样本生存率的 Kaplan-Meier 估计

Kaplan-Meier 方法是由 Kaplan-Meier 提出的,是直接用概率乘法原理估计生存率,又称为乘积极限法 Product-Limit (PL),是最为基本的一种生存分析方法。

假设有 23 例肿瘤患者的生存时间(月)分别为 1、3、5(3)、6(3)、7、8、10(2)、 14+、17、19+、20+、22+、26+、31+、34、34+、44、59。其中括号内表示重复死亡人数, "+"表示截尾数据。现采用 Kaplan-Meier 法估计生存率。这里仍使用 SPSS 软件,数据录 入界面如图 5.7 所示。

<u>F</u> ile <u>E</u> d	it <u>V</u> iew <u>D</u> a	ta <u>T</u> ransfor	m <u>A</u> nalyze	Graphs
2				ř I
15 :				
	time	die	frequency	N N
1	1.00	1.00	1.00	
2	3.00	1.00	1.00	
3	5.00	1.00	3.00	
4	6.00	1.00	3.00	
5	7.00	1.00	1.00	
6	8.00	1.00	1.00	
7	10.00	1.00	2.00	
8	14.00	.00	1.00	
9	17.00	1.00	1.00	
10	19.00	.00	1.00	
11	20.00	.00	1.00	
12	22.00	.00	1.00	
13	26.00	.00	1.00	
14	31.00	.00	1.00	
15	34.00	1.00	1.00	
16	34.00	.00	1.00	
17	44.00	1.00	1.00	
18	59.00	1.00	1.00	

图 5.7 Kaplan-Meier 法数据录入界面

其中, time 变量表示生存时间; die 表示类型(1表示死亡,0表示截尾); frequency 表示对应的人数。首先对人数进行加权,然后选择菜单 Analyze→Survival→Kaplan-Meier, 如图 5.8 所示。

<u>V</u> iew <u>D</u> ate	a <u>T</u> ransform	Analyze Graphs Utilities	<u>¥</u> ir	ndow	Help	
	No En Reports 00 1.00 General Linear Mode 00 1.00 General Linear Mode 00 1.00 Miged Models 00 1.00 Egression 00 1.00 Regression 00 1.00 Loglinear 00 1.00 Data Reduction 00 1.00 Scale 00 1.00 Scale 00 0.00 Multiple Response 00 .00 Multiple Response	Reports Descriptive Statistics	*	٥		
		Tables				
time	die f	Compare Means	•		var	var
1.00	1.00	General Linear Model	٠			
3.00	1.00	Mixed Models	٠			
5.00	1.00	Correlate	۲	T		
6.00	1.00	Regression	•	IT		
7.00	1.00	Loglinear	•	IT		
8.00	1.00	Classify	•	H		
10.00	1.00	Data Reduction	•	H		
14.00	.00	Sc <u>a</u> le	•	H		
17.00	1.00	<u>N</u> onparametric Tests		H		
19.00	.00	T <u>i</u> me Series	-	H.		
20.00	.00	Survival		1	ife Tables	
22.00	.00	Multiple Kesponse		1	aplan-Mere	r
26.00	.00	Missing Value Analysis	1	2	ox Kegress	Ion
31.00	.00	Lompiex Samples			ox w/ lime	-Dep CoV

图 5.8 Kaplan-Meier 法分析菜单

在弹出的"Kaplan-Meier"主对话框中的"Time"输入生存时间"time",在"Status"中输入"die",并将"1"设为结局,单击【Save...】按钮,在弹出的子对话框"Kaplan-Meier: Save New ..."中选择"Survival",如图 5.9 所示。单击【Continue】按钮返回"Kaplan-Meier" 主对话框,单击【OK】按钮运行分析。

Kaplan-Meier		×
* frequency	Lime: time Status: die(1) Define Event	OK Paste Beset Cancel Help
	Factor. Stigta: Label Cases by:	Kaplan-Teier: Save Kev X V Survival Continue Standard gror of survival Hazard
Compare Factor	Save Option	IS Cumulative events Help

图 5.9 Kaplan-Meier 法分析的主对话框和 Save 子对话框

输出结果如图 5.10 和图 5.11 所示。

			Cumulative Surviving a	Proportion t the Time	N of	N of Remaining
	Time	Status	Estimate	Std. Error	Events	Cases
1	1.000	1.00	0.957	0.043	1	22
2	3.000	1.00	0.913	0.059	2	21
3	0.000	0.00	0.000	0.000	0	0
4	0.000	0.00	0.000	0.000	0	0
5	5.000	1.00	0.783	0.086	5	18
6	0.000	0.00	0.000	0.000	0	0
7	0.000	0.00	0.000	0.000	0	0
8	6.000	1.00	0.652	0.099	8	15
9	7.000	1.00	0.609	0.102	9	14
10	8.000	1.00	0.565	0.103	10	13
11	0.000	0.00	0.000	0.000	0	0
12	10.000	1.00	0.478	0.104	12	11
13	14.000	0.00			12	10
14	17.000	1.00	0.430	0.104	13	9
15	19.000	0.00			13	8
16	20.000	0.00			13	7
17	22.000	0.00			13	6
18	26.000	0.00			13	5
19	31.000	0.00			13	4
20	34.000	1.00	0.323	0.122	14	3
21	34.000	0.00		a megada ang di sadi hele 2	14	2
22	44.000	1.00	0.161	0.129	15	1
23	59.000	1.00	0.000	0.000	16	0

Survival Table

图 5.10 输出结果 (Survival Table)

图 5.10 所示的输出结果为生存分析表。该输出表输出的结果是用乘积极限法估计的各时间点的生存率及生存率的标准误差。例如,3 个月的累计生存率为 0.913,标准误差为 0.059;5 个月的累计生存率为 0.783,标准误差为 0.086。

图 5.11 所示的输出结果表示的是平均和中位生存时间及它们的 95% 置信区间。输出表中的 Mean 表示平均生存时间(本例为 24.228 个月), Median 表示中位生存时间(本例为 10 个月)。

		Mean ^a			١	Median	
		95% Confide	ence Interval			95% Confide	ence Interval
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
24.228	4.991	14.444	34.011	10.000	6.955	0.000	23.632

Means and Medians for Survival Time

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

图 5.11 输出结果(Means and Medians for Suvival Time)

5.6 应用实例 3: 生存曲线比较的 Log-rank 检验

在实践中,有时常常需要进行两组生存曲线的比较。此时可以采用 Log-rank 检验来比较两组生存曲线。

假设有两组肿瘤患者的生存情况:

对照组: 2, 2, 3, 3, 5+, 9, 9+, 10, 12+, 19+, 25, 32+, 36+, 42+

处理组: 2, 5+, 10, 13, 13+, 14+, 14+, 19+, 19+, 25+, 26+, 30+, 39+, 41+, 51+, 57+, 61+

其中, "+"表示截尾数据, 患者仍生存或失访。生存时间的单位为月。要比较两组的 生存率有无差别, 这里使用 SPSS 软件实现。

首先录入数据,如图 5.12 所示。其中,组别变量中的 1 表示对照组,2 表示处理组; 类型变量中的 1 表示完全数据,0 表示截尾数据。

选择菜单 Analyze→Survival→Kaplan-Meier 命令,如图 5.13 所示。

	组别	生存时间	类型	1
1	1.00	2.00	1.00	
2	1.00	2.00	1.00	
3	1.00	3.00	1.00	
4	1.00	3.00	1.00	
5	1.00	5.00	.00	
6	1.00	9.00	1.00	
7	1.00	9.00	.00	
8	1.00	10.00	1.00	
9	1.00	12.00	.00	
10	1.00	19.00	.00	
11	1.00	25.00	1.00	
12	1.00	32.00	.00	
13	1.00	36.00	.00	
14	1.00	42.00	.00	
15	2.00	2.00	1.00	
16	2.00	5.00	.00	
17	2.00	10.00	1.00	
18	2.00	13.00	1.00	
19	2.00	13.00	.00	
20	2.00	14.00	.00	

7 – SI	PSS Data Editor	
asform	Analyze Graphs Utilities	Window Help
<u>- 67</u>	Reports D <u>e</u> scriptive Statistics <u>I</u> ables	<u>ا</u>
生在	Compare <u>M</u> eans Compare Lincor Model	ar var
	Mixed Models	•
	<u>C</u> orrelate	•
	Regression	•
	Loglinear	P
	Classify	P
	Data Reduction	•
	Sc <u>a</u> le	F
	<u>N</u> onparametric Tests	F
	T <u>i</u> me Series	•
	Survival	Life Tables
	Multiple Response	▶ <u>K</u> aplan-Meier
	Missing <u>V</u> alue Analysis	. <u>C</u> ox Regression
	Comp <u>l</u> ex Samples	Eox v/ Time=Tep Cov
	19.00 .00	

图 5.12 两组生存曲线数据录入界面

图 5.13 两组生存曲线比较菜单

在弹出的对话框 "Kaplan-Meier"中的"Time"中输入"生存时间",在 "Status"中输入"类型",其结局事件输入"1",在 "Factor"中输入 "组别",如图 5.14 所示。

单击【Compare Factor...】按钮,弹出对话框 "Kaplan-Meier: Compare Factor Levels",

勾选"Log rank"选项,如图 5.15 所示。单击【Continue】按钮返回"Kaplan-Meier"主对话框,单击【OK】按钮运行分析。

Kaplan-Meier		Þ
		OK
		Paste
		<u>R</u> eset
	Define Eve	nt Cancel
	Eactor:	Help
	Strata:	
	Label Cases by	r
Compare Factor	Saye	Options

图 5.14 两组生存曲线比较主对话框

	Continue
I <u>v Log ranki</u> I <u>B</u> reslow IIarone- <u>W</u> are	Cancel
Linear trend for factor levels Pooled over strata, C Pairwise over strata	Help
C For each stratum C Pairwise for each stratum	

图 5.15 Log rank 检验子对话框

主要输出结果如图 5.16~图 5.18 所示。

图 5.16 描述了对照组和处理组的完全事件数及截尾事件数。

Case Processing Summary

			Cens	ored
组别	Total N	N of Events	N	Percent
1.00	14	7	7	50.0%
2.00	17	3	14	82.4%
Overall	31	10	21	67.7%

图 5.16 输出结果 (Case Processing Summary)

图 5.17 描述了对照组和处理组的平均及中位生存时间和它们的 95% 置信区间。对照组 的平均及中位生存时间分别为 22.880 月和 25.000 月。处理组仅输出了平均生存时间为 51.318 月。

Means	and	Medians	for	Survival	Time

			Mean ^a			1	Median	
			95% Confide	ence Interval			95% Confide	ence Interval
组别	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
1.00	22.880	4.975	13.128	32.631	25.000	11.139	3.167	46.833
2.00	51.318	5.068	41.385	61.250				
Overall	42.082	4.871	32.535	51.629				

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

图 5.17 输出结果(Means and Medians for Suvival Time)

扫码看彩图

图 5.18 输出的是 Log-rank 检验结果 (P=0.033<0.05),说明两组生存曲线的差异有统 计学意义。

在图 5.19 所示的两组生存曲线比较对话框中单击【Option】按钮,并在"Plots"组中 勾选"Survival"选项,单击【Continue】按钮返回主对话框。最后单击【OK】按钮运行 分析。

rall Compariso	ns	
Chi-Square	df	Sig.
4 538	1	0.022
	rall Compariso	rall Comparisons Chi-Square df

Survival table(s)	Continue
✓ Mean and median survival	Cancel
🗖 Quartiles	Help
Plots	
✓ Survival	
🔲 🖸 ne minus survival	
🗆 <u>H</u> azard	
Log survival	

Test of equality of survival distributions for the different levels o 组别

图 5.18 输出结果 (Overall Comparisons)

图 5.19 绘制两组生存曲线 Options 子对话框

此时就绘制出了两组生存曲线比较图,如图 5.20 所示。从图 5.20 中看出,两条生存曲 线的分辨率很好,说明处理组和对照组的生存曲线有差异,而处理组的累积生存率明显高 于对照组的累积生存率。



