
SPSS：生存分析的Cox回归模型（比例风险模型）

作者：张倩 来源：爱科学

本文原地址：<https://www.iikx.com/news/statistics/5969.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

SPSS：生存分析的Cox回归模型(比例风险模型)

一、生存分析基本概念

1、事件(Event)

指研究中规定的生存研究的终点，在研究开始之前就已经制定好。根据研究性质的不同，事件可以是患者的死亡、疾病的复发、仪器的故障，也可以是下岗工人的再就业等等。

2、生存时间(Survival time)

指从某一起点到事件发生所经过的时间。生存是一个广义的概念，不仅仅指医学中的存活，也可以是机器出故障前的正常运行时间，或者下岗工人再就业前的待业时间等等。有的时候甚至不是通用意义上的时间，比如汽车在出故障前的行驶里程，也可以作为生存时间来考虑。

3、删失(Sensoring)

指由于所关心的事件没有被观测到或者无法观测到，以至于生存时间无法记录的情况。常由两种情况导致：(1)失访;(2)在研究终止时，所关心的事件还未发生。

4、生存函数(Survival distribution function)

又叫累积生存率，表达式为 $S(t)=P(T>t)$ ，其中 T 为生存时间，该函数的意义是生存时间大于时间点 t 的概率。 $t=0$ 时 $S(t)=1$ ，随着 t 的增加 $S(t)$ 递减(严格的说是不增)， $1-S(t)$ 为累积分布函数，表示生存时间 T 不超过 t 的概率。

二、生存分析的方法

1、生存分析的主要目的是估计生存函数，常用的方法有Kaplan-Meier法和寿命表法。对于分组数据，在不考虑其他混杂因素的情况下，可以用这两种方法对生存函数进行组间比较。

2、如果考虑其他影响生存时间分布的因素，可以使用Cox回归模型(也叫比例风险模型)，利用数学模型拟合生存分布与影响因子之间的关系，评价影响因子对生存函数分布的影响程度。这里的前体是影响因素的作用不随时间改变，如果不满足这个条件，则应使用含有时间依存协变量的C

ox回归模型。

下面用一个例子来说明SPSS中Cox回归模型的操作方法。

例题

要研究胰腺癌术中放疗对患者生存时间的影响，收集了下面所示的数据：

表 19.3 胰脏癌术中放疗效果研究数据的说明

变量名	变量说明	变量类型	分类变量的编码
caseno	患者编号		
time	生存时间（月）	连续	
censor	删失	2 分类	0: 死亡、1: 删失
age	手术时的年龄	连续	
trt	处理组别（有无术中放疗）	2 分类	0: 无术中放疗、1: 有术中放疗
sex	性别	2 分类	0: 男、1: 女
bui	占位处	2 分类	0: 胰脏头部、1: 头部以外
ch	胰胆管浸润程度	有序多分类	1: ch0、2: ch1、3: ch2、4: ch3
p	有无腹膜转移	2 分类	0: 无、1: 有
stage	TNM 分类	2 分类	3: III 期、4: IV 期

操作步骤：

SPSS变量视图

名称	类型	宽度	小数	标签	值	...	列	对齐	度量标准
caseno	数值(N)	8	0		无	无	8	右	度量(S)
time	数值(N)	8	1	生存时间（月）	无	无	8	右	度量(S)
censor	数值(N)	8	0	删失	{0, 死亡}...	无	8	右	名义(N)
age	数值(N)	8	0	手术时的年龄	无	无	8	右	度量(S)
sex	数值(N)	8	0	性别	{0, 男}...	无	8	右	名义(N)
trt	数值(N)	8	0	有无术中放疗	{0, 无术中放...	无	8	右	名义(N)
bui	数值(N)	8	0	占位处	{0, 胰脏头部...	无	8	右	名义(N)
ch	数值(N)	8	0	胰胆管浸润程度	{1, CH0}...	无	8	右	序号(O)
p	数值(N)	8	0	有无腹膜转移	{0, 无}...	无	8	右	名义(N)
stage	数值(N)	8	0	TNM分期	{3, III期}...	无	8	右	名义(N)

菜单选择：

文件(F) 编辑(E) 视图(V) 数据(D) 转换(T) 分析(A) 直销(M) 图形(G) 实用程序(U) 窗口(W) 帮助

2: stage 4

	caseno	time	ce
1	1	2.4	
2	2	1.7	
3	3	.1	
4	4	1.0	
5	5	4.8	
6	6	6.4	
7	7	10.8	
8	9	5.1	
9	10	1.1	
10	11	.5	
11	12	.8	
12	14	4.0	
13	15	4.0	
14	16	4.0	
15	17	8.5	
16	18	3.6	
17	19	6.9	
18	20	6.2	

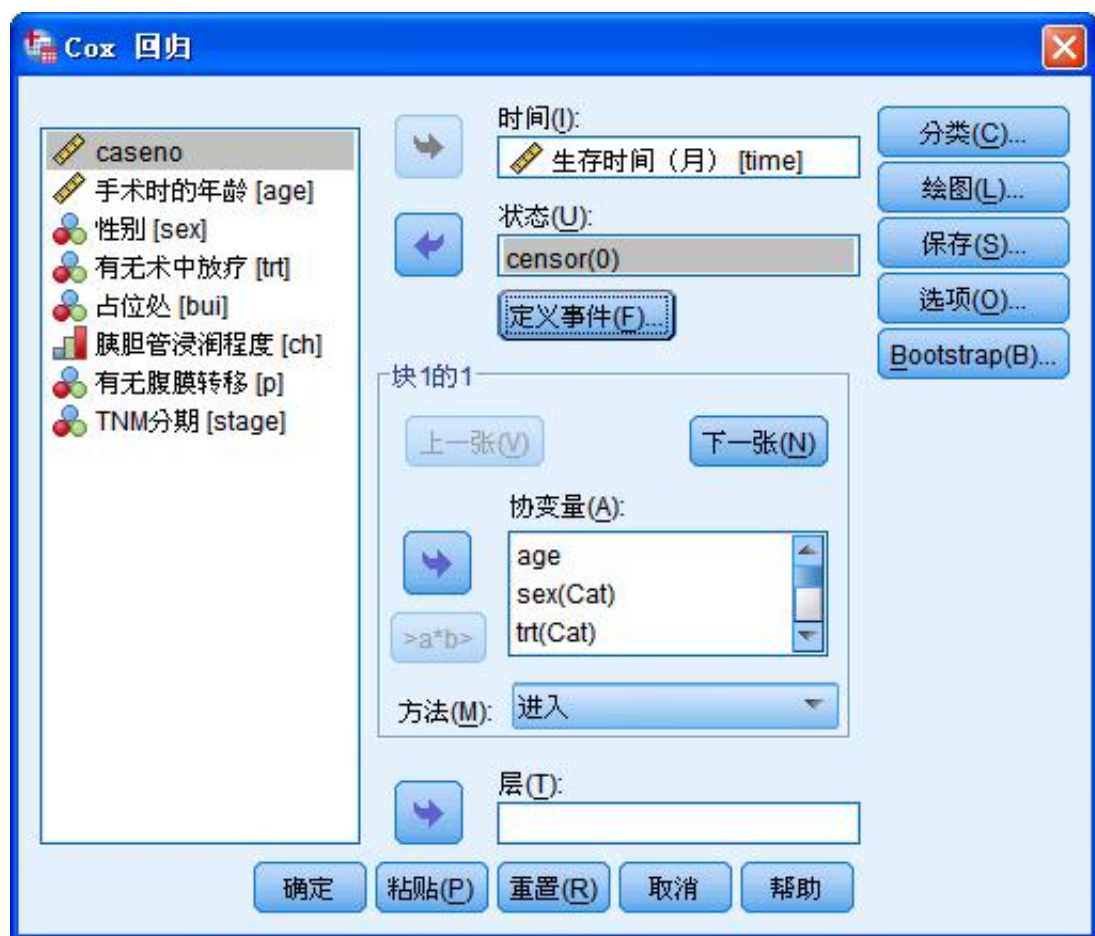
报告
描述统计
表(T)
比较均值(M)
一般线性模型(G)
广义线性模型
混合模型(X)
相关(C)
回归(R)
对数线性模型(O)
神经网络
分类(F)
降维
度量(S)
非参数检验(N)
预测(T)
生存函数(S)
多重响应(U)
缺失值分析(Y)...
多重归因(T)
复杂抽样(L)

sex trt bu

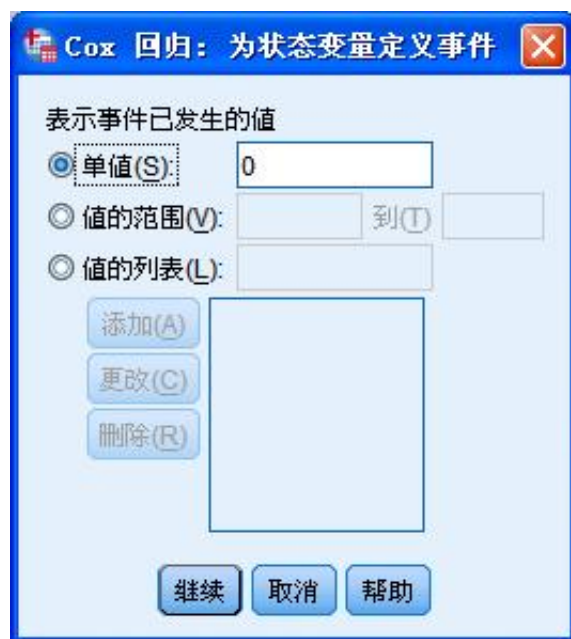
0 0
0 0
0 0
0 0
0 0
0 0
1 0
1 0
0 0
0 0
0 0
1 0
0 0

寿命表(L)...
Kaplan-Meier...
Cox 回归(C)...
Cox 依时协变量(O)...

点击进入Cox主对话框，如下，将time选入“时间”框，将代表删失的censor变量选入“状态”框，其余分析变量选入“协变量”框。“方法”下拉菜单是指变量筛选的方法，可以选择“前向”、“后项”、“进入”等，这里选择“进入”为例，即所有变量同时进入。



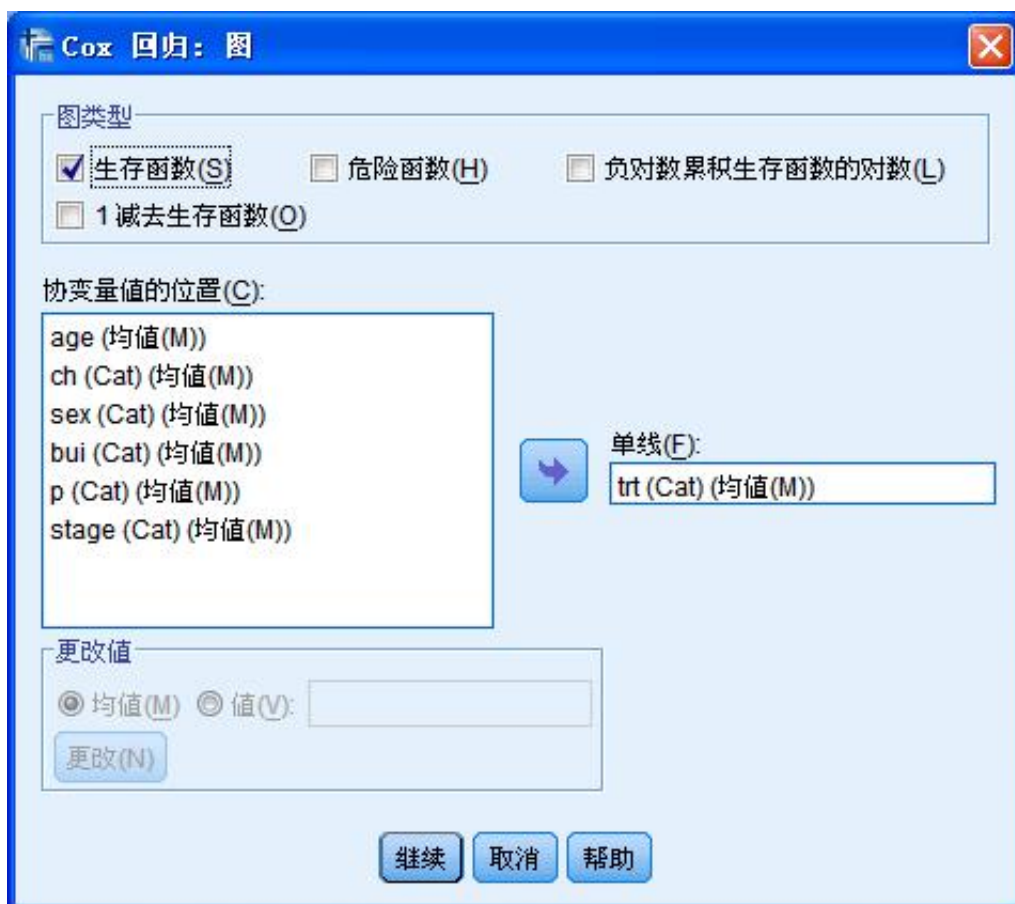
点击“状态”框下方的“定义事件”，将事件发生的标志设为值0，即0代表事件发生。



在主对话框中点击“分类”按钮，进入如下的对话框，将所有分类变量选入右边框中。



在主对话框中点击“绘图”按钮，进入如下的对话框，选择绘图的类型，这里只选择“生存函数”。由于我们关心的主要变量是trt(是否放疗)，所以将trt选入“单线”框中，绘制生存曲线。



在主对话框中点击“选项”按钮，进入如下的对话框，设置如下，输出RR的95%置信区间。回到主界面，点击“确定”输出结果。



结果输出

案例处理摘要		N	百分比
分析中可用的案例	事件 ^a	82	98.8%
	删失	1	1.2%
	合计	83	100.0%
删除的案例	带有缺失值的案例	0	0.0%
	带有负时间的案例	0	0.0%
	层中的最早事件之前删失的案例	0	0.0%
	合计	0	0.0%
	合计	83	100.0%

a. 因变量: 生存时间 (月)

这是案例处理摘要，有一个删失数据。

分类变量编码^{a,d,e,f,g,h}

		频率	(1) ^c	(2)	(3)
sex^b	0=男	48	1		
	1=女	35	0		
trt^b	0=无术中放疗	22	1		
	1=有术中放疗	61	0		
bui^b	0=胰脏头部	18	1		
	1=头部以外	65	0		
ch^b	1=CH0	32	1	0	0
	2=CH1	8	0	1	0
	3=CH2	17	0	0	1
	4=CH3	26	0	0	0
p^b	0=无	60	1		
	1=有	23	0		
stage^b	3=III期	44	1		
	4=IV期	39	0		

- a. 分类变量: **sex** (性别)
- b. 示性参数编码
- c. 已经记录了 (0, 1) 变量, 所以其系数不会与指示符 (0, 1) 编码相同。
- d. 分类变量: **trt** (有无术中放疗)
- e. 分类变量: **bui** (占位处)
- f. 分类变量: **ch** (胰胆管浸润程度)
- g. 分类变量: **p** (有无腹膜转移)
- h. 分类变量: **stage** (TNM分期)

这是分类变量的编码方式。

模型系数的综合测试^a

-2 倍对数似然值	整体 (得分)			从上一步骤开始更改			从上一块开始更改		
	卡方	df	Sig.	卡方	df	Sig.	卡方	df	Sig.
552.313	18.280	9	.032	18.116	9	.034	18.116	9	.034

- a. 起始块编号 1. 方法 = 输入

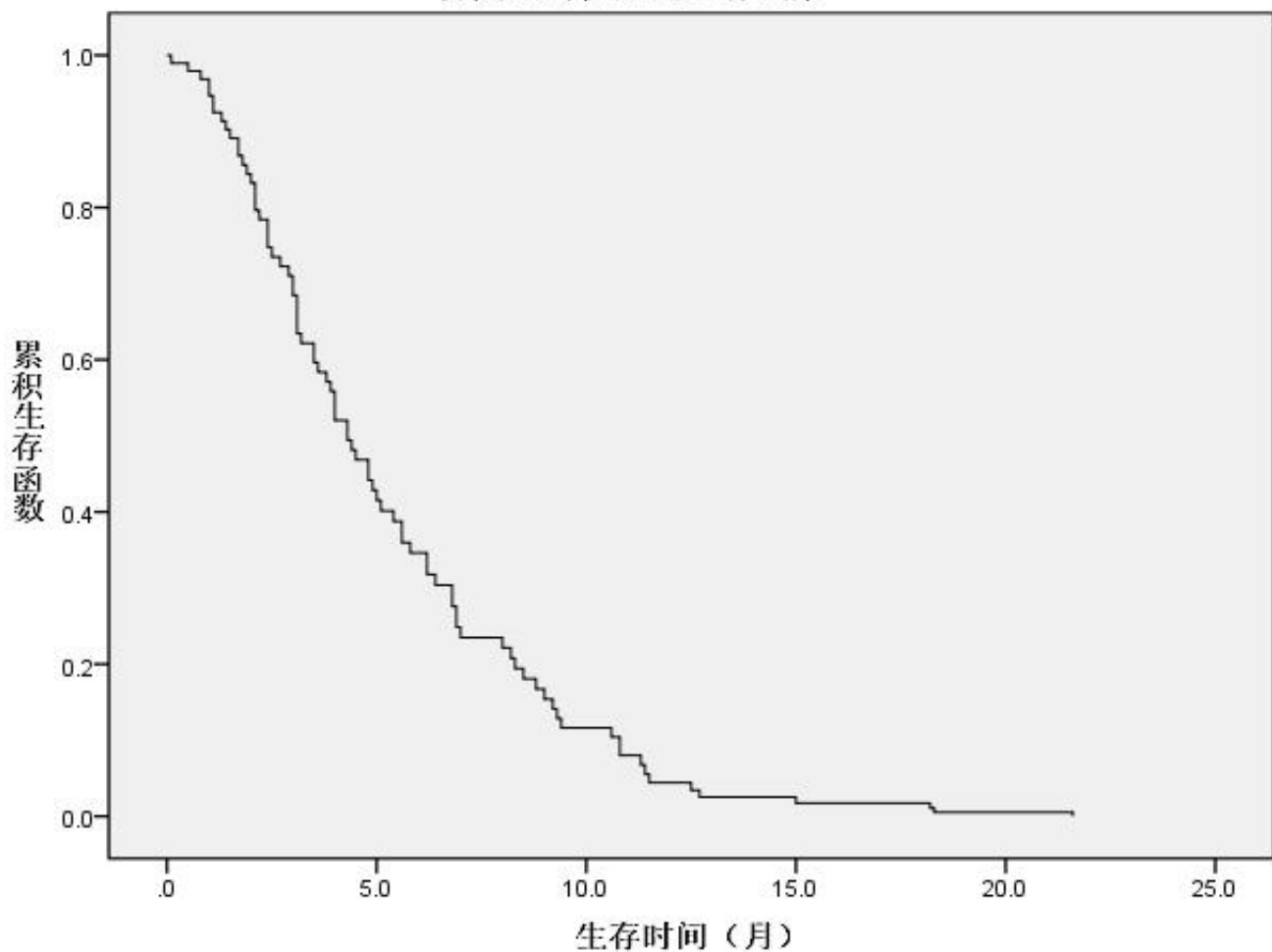
这是对拟合模型的检验, 原假设是“所有影响因素的偏回归系数均为0”, 这里可以看出 $P=0.032 < 0.05$ 拒绝原假设, 认为有偏回归系数不为零的因素, 值得进一步分析。

方程中的变量

	B	SE	Wald	df	Sig.	Exp (B)	95.0% CI 用于 Exp (B)	
							下部	上部
age	.018	.013	1.863	1	.172	1.018	.992	1.045
sex	.034	.245	.019	1	.891	1.034	.640	1.673
trt	.818	.326	6.298	1	.012	2.265	1.196	4.289
bui	-.599	.350	2.939	1	.086	.549	.277	1.090
ch			4.679	3	.197			
ch (1)	-.135	.359	.142	1	.707	.874	.433	1.765
ch (2)	.697	.463	2.260	1	.133	2.007	.809	4.976
ch (3)	.454	.388	1.372	1	.241	1.575	.737	3.366
p	-.360	.278	1.666	1	.197	.698	.404	1.205
stage	-.453	.281	2.596	1	.107	.636	.367	1.103

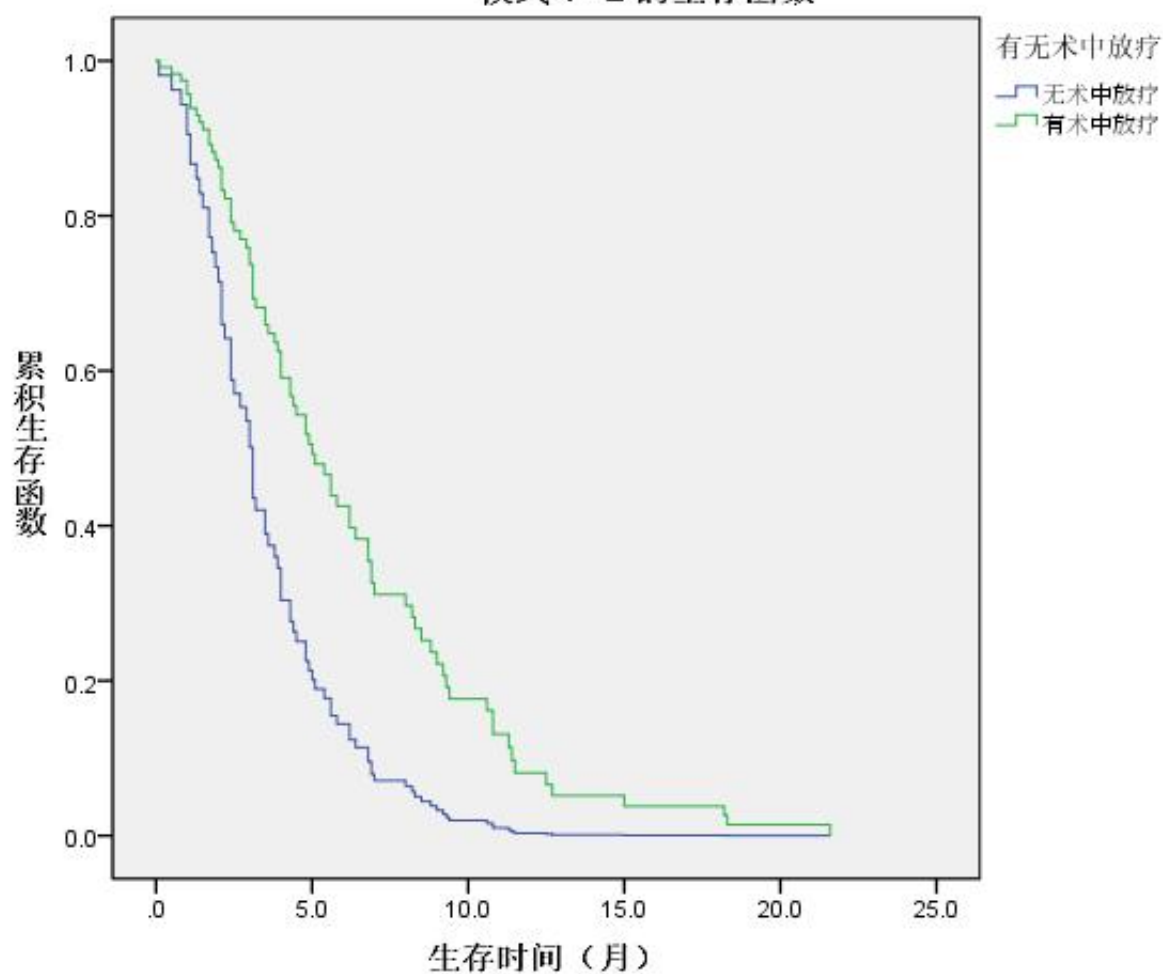
这是多元回归结果，第二列B为偏回归系数，最后三列为OR值及其置信区间。

协变量均值处的生存函数



这是总体的生存函数。

模式 1 - 2 的生存函数



这是在控制了其他变量后，有无放疗组的生存函数对比，可以直观看出，术中放疗患者的生存情况优于不放疗的患者。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发