

## 生存分析：比例风险模型（Cox 模型）

Cox 模型又称比例风险模型，分析发生某种结局（如死亡）的时间与危险因素的联系。Cox 模型可理解为两个部分：基础风险（发生结局事件）函数，常用  $\Lambda_0(t)$  表示，表示随时间变化的发生结局事件的风险；效应参数，表示当某种暴露（危险）因素存在时，基础风险函数的变化。

**比例风险：**指的是暴露（危险）因素对基础风险函数的改变是相乘性的。不管基础死亡风险是多少，某种药物治疗可能会减少病人在某一时点  $t$  的死亡风险 20%。

暴露（危险）因素不一定是两分类的，也可以是连续性的。如是连续性的，则分析的是每单位  $X$  的变化，导致发生结局事件的风险增加（或减少）多少倍。

**结局事件变量：**应编码成：1 表示事件发生，0 表示未发生。

**时间变量：**如结局事件发生，记录发生结局事件的时间；如到观察截止时结局事件尚未发生，记录观察时间。

**重合时间：**当观察的数据中结局事件发生的时间有很多重合时，可有几种方法处理。

Breslow 's 方法使用未修改的方法。另一种被认为能给出较好结果的方法是 Efron 's 方法。

**随时间变化的自变量：**如果预测变量（暴露）随时间变化，将原观察对象记录按时间段分开，每个时间段有开始时间、结束时间及结束状态。如下表，要分析被动吸烟对受孕率的影响，但妇女在不同月经周期的被动吸烟暴露情况不同。

Subject ID	开始周期	结束周期	被动吸烟	是否怀孕
...				
100	1	2	0	0
100	3	5	1	0
100	6	7	0	1

上例中，ID= 100 妇女花了 7 个月经周期怀孕，前 2 个周期没有被动吸烟，周期 3-5 有被动吸烟，周期 6-7 没有被动吸烟。如果有随时间变化的暴露变量，分析时应注明开始时间与结束时间。

模型中的 STRATA 变量：用来表示基础风险存在组间（STRATA）差异。

Landmark 分析：如果随访的不同时间段间基础风险发生变化，需要分时间段分别分析。如手术治疗术后一段时间内病人死亡风险大，但如活过某时间点，死亡风险大大降低。本模块可通过设置分层变量为“时间变量|时间节点”指定软件进行 landmark 分析。

本模块还可以用抽样调查 (Survey) 数据分析，详见易侬抽样调查数据分析 (<http://www.empowerstats.com/cn/manuals/articles/surveyAnalysis.pdf>)

例：练习项目 preg，cox 模型分析怀孕影响因素，并指定按时间 cycle 3 为节点进行 landmark 分析，输入界面如下：

**Cox 模型** ?

标题:

选择分析对象:

结果变量(1=表示事件发生)  处理重复事件方法

自变量

变量	选择
AGE	.
BMI	.
EDU	.
HSMK	.

选择时间变量

开始时间(如有)

模型中的分层(strata)变量

分层(分析)变量

自动检验与选择的自变量(S)的交互作用

‘输出结果:

结局变量: PREG

时间变量: CYCLE

CYCLE segment : <= 3

model: Surv( CYCLE , PREG )~ AGE+BMI+factor(EDU)+HSMK

观察数(人数)	199
事件发生数	132

比例灾害模型

	回归系数 (coef)	回归系数标准 误差	z	危险比 (HR=exp(coef))	1/危险 比	95%区间下 限	95%区间下 限	P 值
AGE	-0.0047	0.0296	- 0.1574	0.9954	1.0047	0.9393	1.0548	0.8749
BMI	-0.0266	0.0620	- 0.4285	0.9738	1.0269	0.8624	1.0996	0.6683
factor(EDU) 1	-0.1426	0.2141	- 0.6660	0.8671	1.1532	0.5700	1.3192	0.5054
factor(EDU) 2	0.2391	0.2896	0.8255	1.2701	0.7874	0.7200	2.2405	0.4091

HSMK	0.0886	0.1881	0.4712	1.0927	0.9152	0.7557	1.5799	0.6375
------	--------	--------	--------	--------	--------	--------	--------	--------

rsq = 0.0137, maxrsq = 0.9984

Log Likelihood (对数似然值) = -636.195613884066

	test	df	pvalue
Likelihood ratio test	2.75390	5.00000	0.73786
Wald test	2.89000	5.00000	0.71681
Score(log rank)test	2.91399	5.00000	0.71324

#### 比例灾害模型诊断

	rho	chisq	p
AGE	-0.0278	0.0955	0.7573
BMI	-0.0445	0.2846	0.5937
factor(EDU)1	0.0683	0.6202	0.4310
factor(EDU)2	-0.0288	0.1105	0.7395
HSMK	0.0122	0.0205	0.8862
GLOBAL		1.8740	0.8663

#### Somers' Rank Correlation for Censored Data

	C	Dxy	aDxy	SD	Z	P	n
Model	0.5454	0.0907	0.0907	0.0653	1.3892	0.1648	199

C: Harrell's C index; Dxy = 2(C - 0.5), Dxy is a U-statistic.

SD: Standard error of Dxy, so, SE. for C = SD/2

#### Step wise model selection

	NA	NA	194	1272.391	1282.391
- AGE	1	0.02477269	195	1272.416	1280.416
- BMI	1	0.16846238	196	1272.584	1278.584
- factor(EDU)	2	2.26430952	198	1274.849	1276.849
- HSMK	1	0.29635462	199	1275.145	1275.145

CYCLE segment : > 3

model: Surv( CYCLE , PREG ) ~ AGE+BMI+factor(EDU)+HSMK

观察数(人数)	62
事件发生数	53

#### 比例灾害模型

	回归系数 (coef)	回归系数标 准误	z	危险比 (HR=exp(coef))	1/危险 比	95%区间下 限	95%区间下 限	P 值
AGE	-0.0180	0.0454	- 0.3957	0.9822	1.0181	0.8985	1.0736	0.6923

BMI	0.0537	0.1110	0.4836	1.0552	0.9477	0.8488	1.3117	0.6287
factor(EDU) 1	0.6271	0.3495	1.7945	1.8722	0.5341	0.9438	3.7140	0.0727
factor(EDU) 2	0.1802	0.5530	0.3259	1.1974	0.8351	0.4051	3.5395	0.7445
HSMK	0.3522	0.3206	1.0988	1.4222	0.7031	0.7588	2.6658	0.2719

rsq = 0.0766, maxrsq = 0.9968

Log Likelihood (对数似然值) = -175.344072940097

	test	df	pvalue
Likelihood ratio test	4.93841	5.00000	0.42344
Wald test	4.72000	5.00000	0.45136
Score(log rank)test	4.81037	5.00000	0.43946

#### 比例灾害模型诊断

	rho	chisq	p
AGE	-0.0774	0.3564	0.5505
BMI	0.0897	0.5054	0.4771
factor(EDU) 1	-0.2001	1.9497	0.1626
factor(EDU) 2	-0.0224	0.0285	0.8659
HSMK	0.1005	0.5826	0.4453
GLOBAL		4.0437	0.5431

#### Somers' Rank Correlation for Censored Data

	C	Dxy	aDxy	SD	Z	P	n
Model	0.6153	0.2306	0.2306	0.1072	2.1501	0.0315	62

C: Harrell's C index; Dxy = 2(C - 0.5), Dxy is a U-statistic.

SD: Standard error of Dxy, so, SE. for C = SD/2

#### Step wise model selection

	NA	NA	57	350.6881	360.6881
- AGE	1	0.1564482	58	350.8446	358.8446
- BMI	1	0.3240969	59	351.1687	357.1687
- HSMK	1	0.9331098	60	352.1018	356.1018
- factor(EDU)	2	3.5247522	62	355.6266	355.6266

Total:

model: Surv( CYCLE , PREG ) ~ AGE+BMI+factor(EDU)+HSMK +factor( TIME.PHASE.4LMCOX )

观察数 (人数)	261
事件发生数	185

#### 比例灾害模型

	回归系数 (coef)	回归系数标准 误差	z	危险比 (HR=exp(coef))	1/危险 比	95%区 间下限	95%区 间下限	P 值
AGE	-0.0087	0.0247	-0.3508	0.9914	1.0087	0.9444	1.0406	0.7257
BMI	-0.0123	0.0539	-0.2275	0.9878	1.0123	0.8888	1.0978	0.8200
factor(EDU) 1	0.0863	0.1835	0.4703	1.0901	0.9173	0.7609	1.5619	0.6381
factor(EDU) 2	0.2690	0.2567	1.0480	1.3086	0.7642	0.7913	2.1641	0.2947
HSMK	0.1356	0.1608	0.8430	1.1452	0.8732	0.8356	1.5696	0.3992
factor(TIME.PHASE.4LMCOX) 2	-19.716	2310.4353	-0.0085	0.0000	inf.	0.0000	inf.	0.9932

rsq = 0.3475, maxrsq = 0.9987

Log Likelihood (对数似然值) = -814.29724233625

	test	df	pvalue
Likelihood ratio test	111.42252	6.00000	0.00000
Wald test	2.21000	6.00000	0.89952
Score(log rank) test	70.19346	6.00000	0.00000

#### 比例灾害模型诊断

	rho	chisq	p
AGE	-0.0389	0.2773	0.5985
BMI	0.0279	0.1629	0.6865
factor(EDU) 1	0.1242	2.8926	0.0890
factor(EDU) 2	-0.0078	0.0115	0.9145
HSMK	0.0530	0.5543	0.4566
factor(TIME.PHASE.4LMCOX) 2	0.5111	0.0000	0.9999
GLOBAL		5.0033	0.5434

#### Somers' Rank Correlation for Censored Data

	C	Dxy	aDxy	SD	Z	P	n
Model	0.6883	0.3766	0.3766	0.0468	8.0491	0	261

C: Harrell's C index; Dxy = 2(C - 0.5), Dxy is a U-statistic.

SD: Standard error of Dxy, so, SE. for C = SD/2

#### Step wise model selection

	NA	NA	255	1628.594	1640.594
- factor(EDU)	2	1.07571992	257	1629.670	1637.670
- BMI	1	0.04899817	258	1629.719	1635.719
- AGE	1	0.19573175	259	1629.915	1633.915
- HSMK	1	0.85674567	260	1630.772	1632.772

Final model:

	回归系数 (coef)	回归系数标 准误	z	危险比 (HR=exp(coef))	1/危 险比	95%区间 下限	95%区 间下 限	P 值
factor(TIME.PHASE.4LMCOX) 2	-19.7243	2310.8627	-0.0085	0.0000	inf.	0.0000	inf.	0.9932

