

临床意义 - 预后 Cox 回归分析[云]

Characteristics	Total(N)	HR(95% CI) Univariate analysis	P value Univariate analysis	HR(95% CI) Multivariate analysis
Pathologic T stage	79		0.961	
T1	8	Reference		
T2	27	1.180 (0.303 - 4.588)	0.811	
T3&T4	44	1.058 (0.296 - 3.785)	0.931	
Pathologic N stage	78		0.063	
N0	46	Reference		Reference
N1	26	1.960 (0.789 - 4.869)	0.147	1.960 (0.789 - 4.869)
N2&N3	6	4.310 (1.301 - 14.279)	0.017	4.310 (1.301 - 14.279)
ERBB2	82		0.407	
Low	41	Reference		
High	41	0.699 (0.300 - 1.625)	0.405	
ERBB3	82		0.229	

网址: <https://www.xiantao.love>

更新时间: 2023.03.06

目录

基本概念	3
应用场景	4
分析流程	4
主要结果/结果解读	6
数据格式	8
参数说明	9
特殊参数	9
预后参数	11
阈值控制	11
数据处理	12
结果说明	13
主要结果:	13
补充结果: 变量情况统计表	14
补充结果: 中位生存时间表	15
补充结果: 单因素 cox 回归分析表	16
补充结果: 多因素 cox 回归分析表	17
补充结果: PH 比例风险假设检验表	18
补充结果: 方差膨胀因子表	19
方法学	20
如何引用	21
常见问题	22

基本概念

- Cox 回归模型：又称为比例风险回归模型，是一种半参数回归模型。Cox 模型以生存结局和生存时间为因变量，分析众多自变量因素对生存期的影响

- 数据要求

- ◆ 结局建议用数字编码（0/1，1/2），其中最好用 0 代表删失或者未发生事件，1 表发生事件

- ◆ 自变量（协变量）可以是数值或者分类变量。分类变量如果是含有等级的含义，则需要以等级资料纳入，需要设置参考组，其他组 and 这个参考组作对比；如果分类变量是无等级含义，一般是需要经过哑变量编码，但是经过哑变量编码后结果有可能不好解读，故无等级关系的分类变量也可以通过组合的方式形成二分类变量纳入。二分类的分类变量以等级或者非等级纳入的结果都是一致的（二分类分不分等级都一样）。数值变量可以直接以数值变量的形式纳入，亦可转换为等级资料或者二分类资料纳入

- 条件假设：观测值独立，风险比不随时间改变（比例风险假设）。（模块内默认是满足此条件）

- 对于回归模型的假设检验通常采用似然比检验、Wald 检验和记分检验

- PH 假设：比例风险（Proportional hazards）假定。Cox 模型应用的前提条件。基本假设为：协变量对生存率的影响不随时间的改变而改变，即风险比值 $h(t)/h_0(t)$ 为固定值。而在实际进行生存分析的过程中，有些自变量对风险函数（事件发生概率）的影响会随时间的变化而变化，因此在构建 Cox 回归模型之前，必须对 PH 假定进行判定，只有 PH 假定得到满足时，Cox 回归模型的结果才有意义

- 中位生存时间（半数生存期）：即当累积生存率为 50%时所对应的生存时间，表示有且只有 50%的生病个体可以活过这个时间。只有当分组内最终累积生存率低于 50%才会有中位生存时间

应用场景

Cox 回归模块主要用于评估变量对于预后的影响，或者判断某个变量是否是独立预后因素（多因素中还有统计学意义）。一般在进行多因素 Cox 回归前，会先进行单因素 Cox 回归对每个变量逐个进行分析，将单因素有意义（ $p < 0.1$ ，这个一般不会设置 0.05）纳入到多因素中进行分析

分析流程：

云端数据 ➡ 单因素 Cox 分析 ➡ (单因素分析 p 值是否满足设定的阈值)
➡ 多因素 Cox 分析

- 云端数据：提供预清洗好的云端数据，不同平台的云端数据集的分子和临床变量可能会有不同
 - 通过特殊参数[变量]选择临床变量或者分子，可以进行输入搜索
 - 通过特殊参数[分组]将选择的临床变量或分子进行分组，得到最终需要进行分析的数据（具体参数操作可看后面参数部分）
- 单因素 Cox 回归分析：

- 构建预后 cox 回归模型：选择好的数据进行 cox 模型构建
- 通过模型得到模型所有变量的分析结果
- 多因素 Cox 回归分析：
 - **筛选变量**：将单因素结果得到的 p 值，没有达到参数“阈值控制”p 值大小的变量筛选出来，不进行后续多因素 Cox 回归
 - 多因素 Cox 回归分析



主要结果/结果解读

Characteristics	Total(N)	HR(95% CI) Univariate analysis	P value Univariate analysis	HR(95% CI) Multivariate analysis
Pathologic T stage	79		0.961	
T1	8	Reference		
T2	27	1.180 (0.303 - 4.588)	0.811	
T3&T4	44	1.058 (0.296 - 3.785)	0.931	
Pathologic N stage	78		0.063	
N0	46	Reference		Reference
N1	26	1.960 (0.789 - 4.869)	0.147	1.960 (0.789 - 4.869)
N2&N3	6	4.310 (1.301 - 14.279)	0.017	4.310 (1.301 - 14.279)
ERBB2	82		0.407	
Low	41	Reference		
High	41	0.699 (0.300 - 1.625)	0.405	
ERBB3	82		0.229	

	A	B	C	D	E	F
1	Characteristics	Total(N)	HR(95% CI) Univariate analysis	P value Univariate analysis	HR(95% CI) Multivariate analysis	P value Multivariate analysis
2	Pathologic T stage	79		0.961		
3	T1	8	Reference			
4	T2	27	1.180 (0.303 - 4.588)	0.811		
5	T3&T4	44	1.058 (0.296 - 3.785)	0.931		
6	Pathologic N stage	78		0.063		
7	N0	46	Reference		Reference	
8	N1	26	1.960 (0.789 - 4.869)	0.147	1.960 (0.789 - 4.869)	0.147
9	N2&N3	6	4.310 (1.301 - 14.279)	0.017	4.310 (1.301 - 14.279)	0.017
10	ERBB2	82		0.407		
11	Low	41	Reference			
12	High	41	0.699 (0.300 - 1.625)	0.405		
13	ERBB3	82		0.229		
14	Low	41	Reference			
15	High	41	1.622 (0.730 - 3.602)	0.235		

➤ Characteristic: 变量以及分组

■ 如果变量是等级/分类变量，则紧接其后的变量的分组，其中第一个分组为参考组

■ 如果变量是数值类型，则该变量在表格中只有有一行结果

➤ Total(N): 数量情况。对应变量总所选分组总的数量以及各组的数量，此样本数量为进行单因素分析时变量和对应分组的数目

■ 由于可能包含缺失信息，所以不同的变量之间的总数可能是不同。（可在参数中选择“进行单因素前先过滤缺失样本”，就能保证变量的总数是一致的）

■ 这一列在文章中并不是一定需要的，可以不提供

- HR(95% CI) Univariate analysis: 单因素分析得到的 HR (Hazard ratio, 风险比) 以及对应的置信区间。一般 $HR > 1$ 说明变量是危险因素, $HR < 1$ 为保护因素
 - 如果是等级/分类变量的参考组, 则此分组 HR 为 Reference
 - 如果该行对应的是等级/分类变量的变量名 (非具体分组), 则不会有 HR 值
- P value Univariate analysis: 单因素分析得到的自变量对应的 p 值 (一般是满足 < 0.1 就会纳入到模型中)
 - 如果是等级/分类变量的参考组, 则此分组单因素 p 值为空
 - 如果该行对应的是等级/分类变量的变量名 (非具体分组), 则单因素 p 值为整个变量整体性检验的 p 值, 这个 p 值影响该变量是否纳入到多因素。(该 p 值在文章中不需要报告)
- HR(95% CI) Multivariate analysis: (只有变量满足进入多因素 Cox 模型的 p 值阈值才会有值) 多因素分析得到的 HR (Hazard ratio, 风险比) 以及对应的置信区间。一般 $HR > 1$ 说明变量是危险因素, $HR < 1$ 为保护因素
- P value Multivariate analysis: (只有变量满足进入多因素 Cox 模型的 p 值阈值才会有值) 多因素分析得到的自变量对应的 p 值。当纳入一定的变量时, 多因素仍然有意义则说明该变量可能是独立预后因素。

数据格式

提供预清洗好的云端数据, 不同平台的云端数据集的分子和临床变量可能会有不同。如果有一些想要的临床变量不存在, 则可能是对应的数据集没有提供或者信息较少。



参数说明

(说明：标注了颜色的为常用参数。)

特殊参数

特殊参数
重置参数

变量 ①

(临床)Pathologic_T_stage

(临床)Pathologic_N_stage

(分子-中位数分组)ERBB2[ENSG000000]

(分子-中位数分组)ERBB3[ENSG000000]

分组(一个框内的分类组合成一个组)

T1 [8] ×

T2 [32] ×

T3 [50] ×

T4 [4] ×

N0 [55] ×

N1 [29] ×

N2 [6] ×

N3 [3] ×

Low [中位数] ×

High [中位数] ×

Low [中位数] ×

High [中位数] ×

➤ 变量：第一个框为自变量变量，可以键盘输入进行搜索，下拉选择，可以搜索分子。

- 输入“临床”关键字，可以搜到对应云端数据录入的所有临床数据
 - ◆ 临床变量有分类类型 和 数值类型
- 直接输入分子名，也可以搜索分子

变量 ①

临床
可以直接输入
-

(临床)Columnar_metaplasia

(临床)Radiation therapy

(临床)Age_数值

(临床)Height_数值

分类类型

数值类型

- 第一个框后面的 + 和 -，代表增加或者删去 一行临床变量。
- 第二个框以及以后的框，为选择对应的变量的分组组合。

- 如果是分类类型的变量，则第二个框为参考组，后面的框对应的分组组合和这个参考组进行对比。分类变量的分组中后的中括号为对应临床资料中的分组的数量(未匹配对应的平台, 仅仅只是临床数据中的, 有可能会和最终的结果不符 (因为存在有临床资料没有对应平台的检测结果))
- 如果是数值类型，则框内只有数值可以选择

变量 ①

(临床)Pathologic_T_stage -

分组(一个框内的分类组合成一个组)

T1 [8] × - T2 [32] × -

T3 [50] × T4 [4] × - +

(临床)Pathologic_N_stage -

N0 [55] × - N1 [29] × -

N2 [6] × N3 [3] × - +

(分子-中位数分组)ERBB2[ENSG00000] -

Low [中位数] × - High [中位数] × - +

(分子-中位数分组)ERBB3[ENSG00000] - +

Low [中位数] × - High [中位数] × - +

- 设置某个变量的分组组合时: 对应的分组的第 1 个框除了有 T1 外, 还有一个变量, 这两个变量组合成一个整体作为参考组。点击 × 可以删除这个变量。下拉选项框内显示无数据, 说明这个临床变量的所有分组都已经选上了
- 当一个临床变量还有分组没有选上时, 先机下拉框会显示这个还没有选上的分组

变量 ①

(临床)Pathologic_stage -

分组(一个框内的分类组合成一个组)

Stage I [7] × - +

Stage II [56] -

Stage III [27] - +

Stage IV [4] -

(临床)Pathologic_N_stage -

N1 [29] × -

(分子-中位数分组)ERBB2[ENSG00000] -

Low [中位数] × - High [中位数] × - +

(分子-中位数分组)ERBB3[ENSG00000] - +

Low [中位数] × - High [中位数] × - +

所有变量的选择 以及 对应的分组组合、参考组设置都是可以自定义选择的, 请尽量保存所有的分组组合的数量比较平均和合适。 请注意查看每个变量对应的

分组的数量，如果某个变量含有的分组对应的数量很少，则说明这个变量很可能信息缺失严重，建议是不纳入

预后参数

- 预后类型：可选不同的预后类型。不同的数据集之间的预后类型可能不一样。可以选择：

- OS[Overall Survival]：总体生存期
- DSS[Disease Specific Survival]：无病生存期
- PFI[Progress Free Interval]：无进展间隔

阈值控制

- p 值（单因素进入多因素）：可以控制变量是否进入到多因素 Cox 模型中，不同的数据集之间的预后类型可能不一样，常规阈值可选 0.1, 0.2。如果输入的是 1，则代表所有变量都纳入到多因素中
- 纳入多因素分析的 p 值阈值：默认为 0.1

数据处理



- 缺失值处理：可以选择对数据中缺失值进行处理
 - 默认为 单因素后多因素前处理变量缺失，表示在经过单因素分析之后，通过变量缺失处理在进行多因素分析
 - 还可以选择 单因素前统一处理缺失，则是在进行分析之前对全部的缺失值进行处理



结果说明

主要结果:

Cox分析-云

· Cox回归: 比例回归风险模型, 用于预后资料的分析。条件假设: 观测值独立, 风险比不随时间改变 (比例风险假设)

· 预后类型: OS[Overall Survival]

· P值阈值: 0.1

Characteristics	Total(N)	HR(95% CI) Univariate analysis	P value Univariate analysis	HR(95% CI) Multivariate analysis	P
Pathologic T stage	79		0.961		
T1	8	Reference			
T2	27	1.180 (0.303 - 4.588)	0.811		
T3&T4	44	1.058 (0.296 - 3.785)	0.931		
Pathologic N stage	78		0.063		
N0	46	Reference		Reference	
N1	26	1.960 (0.789 - 4.869)	0.147	1.960 (0.789 - 4.869)	
N2&N3	6	4.310 (1.301 - 14.279)	0.017	4.310 (1.301 - 14.279)	
ERBB2	82		0.407		
Low	41	Reference			
High	41	0.699 (0.300 - 1.625)	0.405		
ERBB3	82		0.229		

[Cox回归结果.xlsx](#) [Cox回归结果.docx](#)

· Characteristics: 变量以及分组

- Characteristics: 变量以及分组
- Total(N): 数量情况。对应变量总所选分组总的数量以及各组的数量, 此样本数量为进行单因素分析时变量和对应分组的数目
- HR(95%CI) Univariate analysis: 单因素分析得到的 HR (Hazard ratio, 风险比) 以及对应的置信区间。其中 Reference 代表等级资料的参考组, 变量中其他组和这个组进行比较
- P value Univariate analysis: 单因素分析得到的自变量对应的 p 值 (一般是满足设定的 p 值阈值就会纳入到多因素模型中)
- HR(95% CI) Multivariate analysis: (只有变量满足进入多因素 Cox 模型的 p 值阈值才会有值) 多因素分析得到的 HR (Hazard ratio, 风险比) 以及对应的置信区间
- P value Multivariate analysis: (只有变量满足进入多因素 Cox 模型的 p 值阈值才会有值) 多因素分析得到的 p 值

补充结果：变量情况统计表

变量情况

各个变量识别出来的类型 以及 是否纳入 进行分析

变量	类型	分类数量	缺失数量	是否纳入分析	补充说明
event	数值变量	-	0	纳入	
time	数值变量	-	0	纳入	
Pathologic T stage	分类变量	3	3	纳入	
Pathologic N stage	分类变量	3	4	纳入	
ERBB2	分类变量	2	0	纳入	
ERBB3	分类变量	2	0	纳入	

总样本数: 82

· 如果某个分类变量的分类 > 10, 将无法识别为分类变量/等级变量

· 如果变量的分组是以 0 1 2 此类进行编码, 如果分类数量 < 5, 会被识别为分类变量; 如果 > 5, 会被识别为数值变量

· 如果数据中含有无穷值, 无穷值会被当做缺失处理

补充说明: 单因素分析前, 会先去掉 结局和时间列 中的缺失的样本(时间或者结局缺失的样本是无法纳入进行分析的)

缺失处理策略: 单因素后多因素前处理变量缺失

这里提供变量情况统计表:

- 如果某个分类变量的分类 > 10, 将无法识别为分类变量/等级变量
- 如果变量的分组是以 0 1 2 此类进行编码, 如果分类数量 < 5, 会被识别为分类变量; 如果 > 5, 会被识别为数值变量
- 如果数据中含有无穷值, 无穷值会被当做缺失处理

补充说明:

- 单因素分析前, 会先去掉 结局和时间列 中的缺失的样本(时间或者结局缺失的样本是无法纳入进行分析的)
- 缺失处理策略: 单因素后多因素前处理变量缺失

补充结果：中位生存时间表

中位生存时间

中位生存时间只针对分类变量进行，数值变量无法统计中位生存时间

Pathologic T stage						
分组	数目	总事件数	总删失数	总删失比例	中位生存时间	中位生存时间置信区间
T1	8	3	5	62.5%	855	557-?
T2	27	8	19	70.4%	763	650-?
T3&T4	44	12	32	72.7%	1263	553-?
Pathologic N stage						
分组	数目	总事件数	总删失数	总删失比例	中位生存时间	中位生存时间置信区间
N0	46	9	37	80.4%	1458	764-?
N1	26	10	16	61.5%	763	557-?
N2&N3	6	4	2	33.3%	471.5	247-?
ERBB2						
分组	数目	总事件数	总删失数	总删失比例	中位生存时间	中位生存时间置信区间
Low	41	11	30	73.2%	764	553-?
High	41	15	26	63.4%	1263	681-?
ERBB3						
分组	数目	总事件数	总删失数	总删失比例	中位生存时间	中位生存时间置信区间
Low	41	10	31	75.6%	1263	557-?

这里提供分类变量中位生存时间表：

- 中位生存时间只针对分类变量进行，数值变量无法统计中位生存时间

补充结果：单因素 cox 回归分析表

单因素Cox					
变量	类型	数量	HR	置信区间	p值
Pathologic T stage	等级变量	79			0.9615
T1		8	Reference		
T2		27	1.180	0.303 - 4.588	0.8113
T3&T4		44	1.058	0.296 - 3.785	0.9308
Pathologic N stage	等级变量	78			0.0634
N0		46	Reference		
N1		26	1.960	0.789 - 4.869	0.1470
N2&N3		6	4.310	1.301 - 14.279	0.0168
ERBB2	等级变量	82			0.4066
Low		41	Reference		
High		41	0.699	0.300 - 1.625	0.4049
ERBB3	等级变量	82			0.2289
Low		41	Reference		

单因素中满足 $p < 0.1$ 就会纳入到多因素Cox回归中

这里提供单因素 cox 回归分析表：

- 表中所有变量都会纳入到多因素中



补充结果：多因素 cox 回归分析表

多因素Cox				
变量	系数β	HR	置信区间	p值
Pathologic N stage				
N0		Reference		
N1	0.67311	1.960	0.789 - 4.869	0.1470
N2&N3	1.461	4.310	1.301 - 14.279	0.0168

模型常数/截距(Intercept): 0

原始数据一共有82个, 变量信息缺失的样本有4个, 最终纳入的样本数: 78

备注: 如果出现纳入了多因素但是对应的统计量为空的情况, 说明(1)这个变量在去除变量信息缺失后某个分类数目过少(只有1个或者0个)或者是(2)存在严重共线性导致这个变量导致没办法计算。

备注: 当如果多因素中出现HR异常大或者异常小时, 说明这个变量的这个分类数量过少或者是存在共线性问题导致

(分类/等级)变量(非分组)对应的单因素p值为对应变量单因素模型全局性检验的p值, 该变量是否纳入取决于此p值

= 模型全局性统计检验情况:

-- 一致性(Concordance, C-index): 0.624(0.555-0.693)

-- Likelihood ratio test= 5.52 on 2 df, p=0.0634

-- Wald test = 6.06 on 2 df, p=0.0483

-- Score (logrank) test = 6.81 on 2 df, p=0.0333

这里提供多因素 cox 回归分析表：

- 如果出现纳入了多因素但是对应的统计量为空的情况, 说明(1)这个变量在去除变量信息缺失后某个分类数目过少(只有 1 个或者 0 个)或者是(2)存在严重共线性导致这个变量导致没办法计算
- 当多因素中出现 HR 异常大或者异常小时, 说明这个变量的这个分类数量过少或者是存在共线性问题导致
- (分类/等级)变量(非分组)对应的单因素 p 值为对应变量单因素模型全局性检验的 p 值, 该变量是否纳入取决于此 p 值

补充结果：PH 比例风险假设检验表

比例风险假设(PH)

Cox回归应用的前提是要求自变量满足等比例风险假设($P > 0.05$)，即自变量的风险不会随着时间改变而改变，若不满足，则不适合用Cox回归进行检验。

这里只对多因素模型以及纳入的变量进行ph假设检验

备注: (1)单个变量直接PH假设和在模型里面这个变量的PH假设的结果是不一样的; (2)同一份数据不同Cox模型中同一个变量的PH假设的结果也是不一样的

变量	统计量(卡方值)	自由度(df)	p值
Pathologic N stage	0.26234	2	0.8771
GLOBAL	0.26234	2	0.8771

如果全局(GLOBAL)满足 $p > 0.05$ ，可以认为多因素模型满足比例风险假设

这里提供 PH 假设检验表：

- Cox 回归应用的前提是要求自变量满足等比例风险假设($P > 0.05$)，即自变量的风险不会随着时间改变而改变，若不满足，则不适合用 Cox 回归进行检验
- 这里只对多因素模型以及纳入的变量进行 ph 假设检验
 - 单个变量直接 PH 假设和在模型里面这个变量的 PH 假设的结果是不一样的
 - 同一份数据不同 Cox 模型中同一个变量的 PH 假设的结果也是不一样的

补充结果：方差膨胀因子表

方差膨胀因子(VIF)

方差膨胀因子可用于分析模型中的变量是否存在多重共线性问题

变量	类型	VIF
Pathologic N stage	等级变量	
N0		Reference
N1		1.1681
N2&N3		1.1681

一般认为，当 $0 < VIF < 10$ ，不存在多重共线性(补充: 也有认为 $VIF > 4$ 就存在多重共线性); 当 $10 \leq VIF < 100$ ，存在较强的多重共线性; 当 $VIF \geq 100$ 或者是出现 NaN，多重共线性非常严重

这里提供方差膨胀因子表：

➤ 方差膨胀因子可用于分析模型中的变量是否存在多重共线性问题

■ 当 $1 < VIF < 10$ ，不存在或存在较轻的多重共线性

■ 当 $10 \leq VIF < 100$ ，存在较强的多重共线性

■ 当 $VIF \geq 100$ 或者是出现 NaN，多重共线性非常严重

方法学

统计分析和可视化均在 R 4.2.1 版本中进行

涉及的 R 包: survival[3.4.0]

处理过程:

- (1) 使用 survival 包进行比例风险假设检验 并进行 Cox 回归分析
- (2) 变量筛选策略: 单因素中样本满足设定的 p 值阈值就会进入到多因素 Cox 中构建模型



如何引用

生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 (www.xiantao.love)。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。



常见问题

1. 为什么说明文本处只有 计数变量和等级变量？为什么没有分类变量？如何设置分类变量（二分类/多分类变量）？

答：所有的分类变量只有 2 个设置方法：设置哑变量或者设置等级变量。模块是默认 都设置为等级，如果要设置要哑变量，请看“数据格式”部分的说明。二分类的分类变量，设置为哑变量或者等级变量或者数值变量，结果都是一样的，因为只有 2 个分类

2. 为什么不同变量的数量不同？

Characteristics	Total(N)	HR(95% CI) Univariate analysis	P value Univariate analysis	HR(95% CI) Multivariate analysis
Pathologic T stage	79		0.961	
T1	8	Reference		
T2	27	1.180 (0.303 - 4.588)	0.811	
T3&T4	44	1.058 (0.296 - 3.785)	0.931	
Pathologic N stage	78		0.063	
N0	46	Reference		Reference
N1	26	1.960 (0.789 - 4.869)	0.147	1.960 (0.789 - 4.869)
N2&N3	6	4.310 (1.301 - 14.279)	0.017	4.310 (1.301 - 14.279)
ERBB2	82		0.407	
Low	41	Reference		
High	41	0.699 (0.300 - 1.625)	0.405	
ERBB3	82		0.229	

答：结果中的这个数量为进行单因素时的数量（如果是分组（而不是变量），则为对应分组的数量）。由于可能包含缺失信息，所以不同的变量之间的总数可能是不同。（可在参数中选择“进行单因素前先过滤缺失样本”，就能保证变量的总数是一致的）。这一列在文章中并不是一定需要的，可以不提供

3. 为什么有一些变量在多因素中统计学数值为 NA?

答：有可能是这么几种情况：

(1) 变量存在共线性

(2) 去除任一变量信息缺失后，某个变量的某个分组变成了 0

4. 为什么多因素模型中某些变量值很大？或者为 Inf?

答：出现值很大的原因可能是因为纳入了缺失严重的变量，导致了某个变量的某些分组的数量变少了，最终导致数值异常了。需要检查是否纳入了信息缺失的变量。（可看说明文本中各个变量的数量）

5. 为什么文章里面的多因素 Cox 模型没有常数?

答：由于多因素 Cox 模型是广义线性模型的一种，也是存在有常数项的。如果在文章中是有可能看到没有写这个常数项的情况，这种情况有这么几种可能：

在 R 里面(print)输出 cox 模型的表是不带常数项的，只有变量和系数，所以就忽略了这个常数项的情况

常数项本身在模型里面也不是关键的，因为是常数，不会对结果有什么影响

6. 为什么文章里面的模型是放的多因素 p 值有意义的变量，而工具却给的是多因素纳入的变量？

答：首先可以明确的是，多因素模型中的自变量就是纳入到多因素模型的所有变量，包括进入多因素模型后 p 值没有意义的变量，肯定不是只要多因素 p 值有意义的变量才是多因素模型的变量

(1) 多因素模型里面正是这些所有变量在模型里面经过变量之间和混杂因素分析后 才得到的每个变量的校正后的情况

(2) 如果是提取了多因素 p 值有意义的变量再构建一个新的多因素模型，那么这些变量的系数肯定不是用的之前的那个模型，肯定是来自一个这个新的模型，而且这些在上一个多因素中 p 值有意义的变量在新的多因素模型中未必还都是 p 值有意义的

