

第二十九讲 R语言-Cox比例风险模型2

T 投必得论... 

已认证的官方帐号

5 人赞同了该文章

在第二十八讲中，我们介绍了Cox比例风险模型的基本概念、计算公式，以及基本的Cox模型的函数coxph()。

在进行Cox比例风险建模时，我们首先需要进行单变量Cox模型，选取有意义的协变量值，然后将选取出来的协变量值用于多元Cox模型中。

这一讲，我们将为大家介绍，如何进行多元Cox回归分析，如何解释回归分析的结果，以及如何通过作图来展示结果。

1安装并加载所需的R包

我们将使用两个R包：

- survival用于计算生存分析
- survminer用于总结和可视化生存分析结果
- 安装软件包

- 加载软件包



```
library("survival")
library("survminer")
```

2 示例数据集

我们将使用survival包中提供的肺癌数据。

```
data("lung")
head(lung)
  inst time status age sex ph.ecog ph.karno pat.karno meal.cal wt.loss
1    3  306     2  74  1      1      90      100     1175      NA
2    3  455     2  68  1      0      90      90      1225      15
3    3 1010     1  56  1      0      90      90        NA      15
4    5  210     2  57  1      1      90      60      1150      11
5    1  883     2  60  1      0     100      90        NA       0
6   12 1022     1  74  1      1      50      80      513       0
```

- inst: 机构代码
- time: 以天为单位的生存时间
- status: 删失状态1 = 删失, 2 = 出现失效事件
- age: 岁
- sex: 性别, 男= 1女= 2
- ph.ecog: ECOG评分 (0 =好, 5 =死)
- ph.karno: 医师进行的Karnofsky评分 (0 = 差, 100 = 好)
- pat.karno: 患者自行进行的Karnofsky评分 (0 = 差, 100 = 好)
- meal.cal: 用餐时消耗的卡路里
- wt.loss: 最近六个月的体重减轻

3 多元Cox回归分析

多元cox回归分析的R代码如下, 其中3个因素(性别, 年龄和ph.ecog)为上一讲中选择出来的变量, 纳入多元模型:

```
res.cox <- coxph(Surv(time, status) ~ age + sex + ph.ecog, data = lung)
summary(res.cox)
Call:
coxph(formula = Surv(time, status) ~ age + sex + ph.ecog, data = lung)
n= 227, number of events= 164
(1 observation deleted due to missingness)
      coef exp(coef) se(coef)      z Pr(>|z|)
age      0.011067  1.011128  0.009267  1.194 0.232416
sex     -0.552612  0.575445  0.167739 -3.294 0.000986 ***
ph.ecog  0.463728  1.589991  0.113577  4.083 4.45e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
age      1.0111      0.9890      0.9929      1.0297
sex      0.5754      1.7378      0.4142      0.7994
ph.ecog  1.5900      0.6289      1.2727      1.9864
```

赞同 5 添加评论 分享 喜欢 收藏 ...



```
Likelihood ratio test= 30.5 on 3 df, p=1.083e-06
Wald test              = 29.93 on 3 df, p=1.428e-06
Score (logrank) test = 30.5 on 3 df, p=1.083e-06
```

所有三个总体检验（似然性，Wald和得分）的p值均显著，表明该模型具有显著性意义。这些检验评估了总体beta的综合原假设（ β ）为0。

在上面的示例中，检验统计数据非常一致，并且完全拒绝了综合原假设。即由3个因素（性别，年龄和ph.ecog）组成的模型对风险比的影响系数不为0。

在元Cox分析中，协变量性别和ph.ecog保持显著性（ $p < 0.05$ ）。但是，协变量年龄不显著（ $p = 0.23$ ，大于0.05）。

性别的p值为0.000986，风险比 $HR = \exp(\text{coef}) = 0.58$ ，表明患者的性别与死亡风险降低之间有很强的关系。协变量的风险比可解释为对危险的倍增效应。例如，保持其他协变量不变的前提下，女性（性别 = 2）相比于男性，死亡风险低0.58或42%。我们得出的结论是，成为女性与良好的预后相关。

同样，ph.ecog的p值为4.45e-05，风险比 $HR = 1.59$ ，表明ph.ecog值与死亡风险增加之间有很强的关系。如果保持其他协变量不变，则ph.ecog的值越高，存活率越低，即在其他协变量都一致的人群中，ph.ecog每增高一个单位，死亡风险增高59%。

相比之下，年龄的p值现在为 $p = 0.23$ 。风险比 $HR = \exp(\text{coef}) = 1.01$ ，95%置信区间为0.99至1.03。由于HR的置信区间为1，因此该结果表明，在调整了ph.ecog值和患者的性别后，年龄对HR差异的贡献较小，并且不显著。例如，在其他协变量保持不变的情况下，每老一岁会引起每日死亡危险，1%，但这并没有统计学意义，也不是重大贡献。

4. 可视化生存时间的估计分布情况

将Cox模型拟合到数据后，就可以可视化特定风险组在任何给定时间点的预测的生存率。函数`survfit()`估计生存比例，默认情况下输出的为协变量的平均值。

与KM生存曲线不同的是，Cox模型拟合曲线输出的是在矫正了其他协变量因素以后的预测的生存率，而不是实际观察到的生存率情况。

```
# 绘出基线的生存函数
ggsurvplot(survfit(res.cox), color = "#2E9FDF",
            ggtheme = theme_minimal())
```



我们可能希望显示估算的生存率如何被特定协变量影响的。

考虑到这一点，我们想评估性别对估计生存率的影响。在这种情况下，我们先创建一个有两行的新数据表，每一行代表一种性别。其他协变量则固定为其平均值（如果是连续变量）或最低水平（如果它们是离散变量）。对于协变量为哑变量的，平均值为数据集中编码为1的比例。该数据表通过newdata参数传递给survfit（）：

```
# 创建新的数据表
sex_df <- with(lung,
  data.frame(sex = c(1, 2),
             age = rep(mean(age, na.rm = TRUE), 2),
             ph.ecog = c(1, 1)
            )
)

sex_df
  sex    age ph.ecog
1   1 62.44737      1
2   2 62.44737      1
# 绘制曲线图
fit <- survfit(res.cox, newdata = sex_df)
ggsurvplot(fit, conf.int = TRUE, legend.labs=c("Sex=1", "Sex=2"),
  ggtheme = theme_minimal())
```



参考内容：sthda.com

如果您觉得我说的对您有帮助，请点赞让我感到您的支持，您的支持是我写作最大的动力~

[ijournal](#)：高颜值的期刊检索网站，助您快速找到理想目标期刊（weixin小程序也上线了哦）

[投必得](#)：全专业中英文论文润色编辑助力您的论文快速发表，点击了解业务详情

专栏传送门：

[投必得科研软件安装使用手册](#)；[投必得：SCI期刊介绍与选择](#)；[投必得，教你写论文](#)；[投必得统计分析大讲堂](#)；[投必得科研生活解忧杂货店](#)

发布于 07-28

[风险模型](#) [生存分析](#) [R（编程语言）](#)

文章被以下专栏收录



投必得统计分析大讲堂

进入专栏

推荐阅读

投必得七大服务类型

1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

第二十八讲 R语言-Cox比例风险模型1

投必得论文... 发表于投必得统计...

投必得七大服务类型

1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

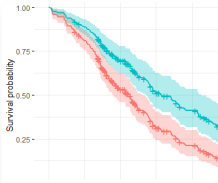
1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

1. 项目早期诊断服务
2. 项目中期诊断服务
3. 项目后期诊断服务

第三十讲 R语言-Cox比例风险模型的假设检验条件

投必得论文... 发表于投必得统计...



COX比例风险模型

经济小海狸

还没有评论

写下你的评论...

😊