

## 投必得七大服务类型



投必得学术

## 第二十七讲 R-生存分析：生存函数的假设检验



投必得论...

已认证的官方帐号

6 人赞同了该文章

在第二十六讲中，我们讲到了KM生存曲线的绘制（插入链接XXXXXXXXXX），并且提到，对于不同组（层），如男人与女人，可以在同一张图里头绘制多条曲线。那么，我们如何确认在两组中的生存曲线是否有统计学差异呢？这一讲，我们将向大家介绍如何对多个生存函数进行假设检验。

### 1. 常见检验方法

对KM生存函数有三种常见检验方法：log-rank、breslow、tarone。这三种都属于卡方检验的方法。KM法根据观察时点的顺序，把生存资料从小到大排列，依次计算实际死亡数和预期死亡数。

▲ 赞同 6 ▼    ● 添加评论    ➤ 分享    ❤ 喜欢    ★ 收藏    ...

- Breslow检验则在Log Rank检验的基础上增加了权重，并设置权重为各时点开始时存活的人数。也就是开始存活人数多的时点死亡情况的变化对整个模型的贡献较大，而开始存活人数少的时点死亡情况的变化对整个模型的贡献较小。
- Tarone-Ware检验是权重的取值方法介于以上两种方法之间，设置权重为各时点开始时存活的人数的平方根。同样是开始存活人数多的时点死亡情况的变化对整个模型的贡献较大，而开始存活人数少的时点死亡情况的变化对整个模型的贡献较小。只是开始存活人数多的时点对整个模型的贡献不如Breslow检验大。

简而言之，log-rank法侧重于远期差别，breslow法侧重于近期差别，tarone法介于两者之间。

现实生活中，我们常用的是log-rank检验。

## 2. 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

```
install.packages(c("survival", "survminer"))
```

- 加载软件包

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

## 3. 示例数据集

我们将使用survival包中提供的肺癌数据。

```
data("lung")  
head(lung)
```

▲ 赞同 6 ▼    添加评论    分享    喜欢    收藏    ...

知乎

首发于  
投必得统计分析大讲堂

|   |    |      |   |    |   |   |     |    |      |    |
|---|----|------|---|----|---|---|-----|----|------|----|
| 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. 比较生存曲线的对数秩 (Log-Rank) 检验: survdiff()

对数秩检验是比较两条或更多条生存曲线是否有差异的最常用方法。原假设是两组之间的生存率没有差异。对数秩检验是一种非参数检验, 它不对生存分布做出任何假设。本质上, 对数秩检验将每个组中观察到的事件数与原假设为真 (即, 生存曲线相同) 下的预期事件进行比较。对数秩统计量近似作为卡方检验统计量分布。

函数survdiff()[在survival包中]可用于比较两个或多个生存曲线的对数秩检验。

survdiff()可以如下使用:

```
surv_diff <- survdiff(Surv(time, status) ~ sex, data = lung)
surv_diff
Call:
survdiff(formula = Surv(time, status) ~ sex, data = lung)
      N Observed Expected (O-E)^2/E (O-E)^2/V
sex=1 138      112    91.6      4.55     10.3
sex=2  90       53    73.4      5.68     10.3
Chisq= 10.3 on 1 degrees of freedom, p= 0.00131
```

▲ 赞同 6 ▼    ● 添加评论    ➤ 分享    ♥ 喜欢    ★ 收藏    ...

- n: 每组的受试者总人数。
- obs: 每组中事件观测到的数量。
- exp: 每个组中预期事件数。
- chisq: 用于卡方检验统计量。
- strata: (可选) 每个层中包含的主题数。

生存差异的对数秩检验得出p值为 $p = 0.0013$ ，表明性别群体的生存差异显著。

## 4. 拟合复杂的生存曲线

当我们的生存曲线受多个因素的影响时，需要将多个因素相结合，来完成`ggsurvplot()`的输出

### 4.1 使用结肠数据集拟合（复杂）生存曲线（多个因素：性别、rx、粘附力）

```
library("survival")
library("survminer")
require("survival")
data("colon")
head(colon)
```

|   | id | study | rx      | sex | age | obstruct | perfor | adhere | nodes | status | differ | extent | surg | node |
|---|----|-------|---------|-----|-----|----------|--------|--------|-------|--------|--------|--------|------|------|
| 1 | 1  | 1     | Lev+5FU | 1   | 43  | 0        | 0      | 0      | 5     | 1      | 2      | 3      | 0    |      |
| 2 | 1  | 1     | Lev+5FU | 1   | 43  | 0        | 0      | 0      | 5     | 1      | 2      | 3      | 0    |      |
| 3 | 2  | 1     | Lev+5FU | 1   | 63  | 0        | 0      | 0      | 1     | 0      | 2      | 3      | 0    |      |
| 4 | 2  | 1     | Lev+5FU | 1   | 63  | 0        | 0      | 0      | 1     | 0      | 2      | 3      | 0    |      |
| 5 | 3  | 1     | Obs     | 0   | 71  | 0        | 0      | 1      | 7     | 1      | 2      | 2      | 0    |      |
| 6 | 3  | 1     | Obs     | 0   | 71  | 0        | 0      | 1      | 7     | 1      | 2      | 2      | 0    |      |

```
#拟合生存曲线
fit2 <- survfit( Surv(time, status) ~ sex + rx + adhere,
                data = colon )
```

### 4.2 使用survminer可视化输出。下图显示了根据rx&粘附力值按性别变量划分的生存曲线。

```
ggsurv <- ggsurvplot(fit2, fun = "event", conf.int = TRUE,
                    ggtheme = theme_bw())
```

图中第一列图为sex = 0的人群中的生存曲线，第二列为sex = 1的人群中的生存曲线，每一行的图分别为rx = Obs, rx = Lev, rx = Lev+5FU的人群中的生存曲线。图中曲线代表了粘附力为0或者1的生存曲线情况。

参考内容：[sthda.com](https://sthda.com)

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 07-26

[生存分析](#)   [R（编程语言）](#)   [数据统计](#)

文章被以下专栏收录

T

投必得统计分析大讲堂

进入专栏

推荐阅读

▲ 赞同 6 ▼

● 添加评论

➦ 分享

♥ 喜欢

★ 收藏

...

- 期刊编辑组
  - 1. 拟定刊例
  - 2. 润色稿件
  - 3. 润色稿件
- 审稿和编辑组
  - 1. 英文润色稿件
  - 2. 中文润色稿件
- 学术评审组
  - 1. 编辑组评审和编辑
  - 2. 专业学术评审组
- 最终出版

**第二十六讲 R-生存分析：绘制KM生存曲线**  
发表于投必得统计...

- 期刊编辑组
  - 1. 拟定刊例
  - 2. 润色稿件
  - 3. 润色稿件
- 审稿和编辑组
  - 1. 英文润色稿件
  - 2. 中文润色稿件
- 学术评审组
  - 1. 编辑组评审和编辑
  - 2. 专业学术评审组
- 最终出版

**第二十五讲 R语言基础概念**  
发表于投必得统计...

还没有评论

写下你的评论...