

第三十讲 R语言-Cox比例风险模型的假设检验条件

T 投必得论... 已认证的官方帐号

10 人赞同了该文章

在介绍完 Cox比例风险模型的详细理论和R实现以后（插入链接!!!），我们已经知道Cox比例风险模型可以较好的同时矫正多个混杂因素对结果的影响，同时我们遗留下来了一个问题：**Cox比例风险模型的假设检验条件是什么？这一讲，我们将对此进行解答。**

请注意，如果使用不当，统计模型可能会产生误导性的结论。因此，检查给定的模型是否适合数据类型是非常重要的！

1. Cox比例风险模型的假设条件

Cox比例风险模型的成立是建立在数据满足一定假设条件的基础上的。因此，评估拟合的Cox回归模型是否满足这些假设条件非常重要。

我们需要测试Cox模型成立的三个基本条件：

- 测试比例风险假设。
- 检查影响观察结果的值（或异常值）。
- 检测对数风险值与协变量之间关系的非线性情况。



Cox模型的常见残差包括：

- Schoenfeld残差：用以检查比例风险假设
- Martingale残差：应以评估非线性情况
- Deviance残差（Martingale残差的对称变换）：用以检查异常值

2. 评估 Cox比例风险模型的有效性

2.1 安装和加载所需的R包

我们将使用两个R包：

- **survival**用于计算生存分析
- **survminer**用于可视化生存分析的结果
- 安装软件包

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

- 加载包

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

2.2 计算Cox模型

我们将使用在survival包中的自带lung部数据集和coxph () 函数。

计算Cox模型，请输入以下内容：

```
library("survival")
res.cox <- coxph(Surv(time, status) ~ age + sex + wt.loss, data = lung)
res.cox
Call:
coxph(formula = Surv(time, status) ~ age + sex + wt.loss, data = lung)
      coef exp(coef) se(coef)      z      p
age      0.02009   1.02029  0.00966   2.08 0.0377
sex     -0.52103   0.59391  0.17435  -2.99 0.0028
wt.loss  0.00076   1.00076  0.00619   0.12 0.9024
Likelihood ratio test=14.7 on 3 df, p=0.00212
n= 214, number of events= 152
(14 observations deleted due to missingness)
```

2.3 测试比例风险假设

我们可以使用一定的统计学方法和图形来描绘**标准化 (scaled) 后的Schoenfeld残差**，从而检查比例风险假设。

原则上，**Schoenfeld残差与时间无关**。如果结果显示该残差与时间有非随机性关系，则说明表明违反了比例风险假设。

函数cox.zph () [在survival软件包中]提供了一种简便的解决方案，可以对Cox模型拟合中包含的每个协变量进行比例风险假设检验。



对于每个协变量，函数`cox.zph()`将标准化的Schoenfeld残差的相应集合与时间相关联，以测试残差和时间之间的独立性。此外，它还会对整个模型进行全局测试。

当比例风险假设中，残差和时间之间的线性关系不显著时，可以说，模型符合比例风险假设。

针对上一步骤得到的比例风险模型，我们进行比例风险假设检验时，请键入以下内容：

```
test.ph <- cox.zph(res.cox)
test.ph
```

	rho	chisq	p
age	-0.0483	0.378	0.538
sex	0.1265	2.349	0.125
wt.loss	0.0126	0.024	0.877
GLOBAL	NA	2.846	0.416

从上面的输出中，每个协变量都不具有统计显著性 ($p > 0.05$)，全局检验也不具有统计学显著性。因此，我们可以认为该Cox模型符合比例风险假设。

可以使用`ggcoxzph()` [在survminer软件包中] 函数进行图形诊断，该函数针对每个协变量生成标准化的Schoenfeld残差相对于时间的相关性图像。

```
ggcoxzph(test.ph)
```



在上图中，实线是曲线的平滑样条拟合，虚线代表拟合周围的+/- 2倍标准误差范围。

请注意，如图所示，当线条基本平行于X轴时，我们可以认为符合风险比例假设，因为我们假设的是，系数 β_1 ， β_2 ， β_3 不 随时间变化。

从图形检查来看，没有协变量是随时间规律变化的。所有的协变量都满足比例风险的假设。其中，由于性别是一个双数值变量，有男女两个数值，所以图中可以看到有两个条带，分别代表男女。

另一种图形方法来检查比例风险的是绘制 $\log(-\log(S(t)))$ 对 t 或 $\log(t)$ 的相关图。这只能用于分类协变量。其原理类似于上面的方法，这里不做赘述。



那么，如果数据不符合比例风险假设，我们可以通过以下方法解决：

- 添加协变量*时间交互
- 分层

分层对于随时间变化的混杂因素很有用，因为这可以有效的避免了这个混杂因素对结果的影响。但是这也使得你没有办法研究这个因素本身对结果的影响。

2.4 检查异常值

为了测试对结果产生影响的异常观察值或离群值，我们可以用下面两个方法可视化：

- deviance残差或
- dfbeta值

ggcoxdiagnostics () 函数（在survminer软件包中）为检查异常值提供了一种方便的解决方案。简化格式如下：

```
ggcoxdiagnostics(fit, type = , linear.predictions = TRUE)
```

- fit: 类coxph.object的对象
- type: 要在Y轴上显示的残差类型。允许的值包括c (" martingale" , " deviance" , " score" , " schoenfeld" , " dfbeta" , " dfbetas" , " scaledsch" , " partial") 。
- linear.predictions: 一个逻辑值，指示是显示观测值的线性预测 (TRUE) 还是仅在X轴上显示观测值的索引 (FALSE) 。

指定参数类型 = " dfbeta" 时，将依次删除每个观察值后绘制回归系数的估计变化；同样，type = " dfbetas" 会产生系数的估计变化除以其标准误。

例如：

```
ggcoxdiagnostics(res.cox, type = "dfbeta",  
                  linear.predictions = FALSE, ggtheme = theme_bw())
```





(dfbeta的指数图，显示了在Cox回归中，用于死亡时间对年龄，性别和体重损失的影响)

上面的指数图表明，将最大dfbeta值的幅度与回归系数进行比较表明，即使年龄和体重损失的某些dfbeta值相对于其他值而言很大，但这些观察结果与每一个排它结果都没有很大的区别，都是均匀分布在 $x = 0$ 的参考线两边。

也可以通过可视化deviance残差来检查离群值。deviance残差是martingale残差的正态分布变换形式。当这些残差应大致对称地分布在零附近，标准差应该为1。

- 正值对应于：与预期生存时间相比“过早死亡”的个体。
- 负值对应于：“寿命过长”的个体。
- 非常大或小的值是离群值，模型无法很好地预测它们。

deviance残差示例：

```
ggcoxdiagnostics(res.cox, type = "deviance",  
  linear.predictions = FALSE, ggtheme = theme_bw())
```

该图形显示各点均匀分布在0附近，并且相对对称。



2.5 检测对数风险值与协变量之间关系的非线性情况

通常，我们假设连续协变量是线性形式。但是，这个假设需要被检验。

用连续协变量绘制Martingale残差是一种用于检测非线性情况，或换句话说，来评估协变量的函数形式，的常用方法。对于给定的连续协变量，图中的模式可以表明变量是否正确拟合了比例风险模型所需要的线性关系。

非线性不是分类变量的问题，因此我们仅针对连续变量绘制martingale残差图和部分残差图。

Martingale残差可以表示为在 $(-\infty, +1)$ 范围内的任何值：

- Martingale残差接近1表示“过早死亡”的个人，
- 较大的负值对应于：“寿命过长”的个体。

为了评估Cox比例风险模型中连续变量的函数形式，我们将使用功能`ggcoxfunctional()` [在 `survminer` 程序包中]。

`ggcoxfunctional()` 函数可以显示连续协变量图，用以显示连续性的协变量相对于cox比例风险模型的原假设的martingale残差。这有助于正确判断Cox模型中连续变量的函数形式。如果拟合线类似为线性，表示满足线性假设。

例如，要评估年龄的函数形式，请输入以下内容：

```
ggcoxfunctional(Surv(time, status) ~ age + log(age) + sqrt(age), data = lung)
```




从以上图来看，年龄的非线性影响在该Cox模型中是比较少的。

参考内容：sthda.com/

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 07-31

R语言绘图 风险模型 生存分析

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. 期刊学术润色服务
2. 润色学术润色服务
3. 润色学术润色服务

4. 期刊学术润色服务
5. 润色学术润色服务
6. 润色学术润色服务

7. 期刊学术润色服务
8. 润色学术润色服务
9. 润色学术润色服务

10. 期刊学术润色服务
11. 润色学术润色服务
12. 润色学术润色服务

第二十九讲 R语言-Cox比例风险模型2

投必得论文... 发表于投必得统计...

投必得七大服务类型

1. 期刊学术润色服务
2. 润色学术润色服务
3. 润色学术润色服务

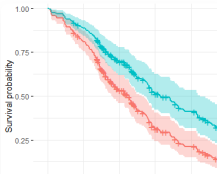
4. 期刊学术润色服务
5. 润色学术润色服务
6. 润色学术润色服务

7. 期刊学术润色服务
8. 润色学术润色服务
9. 润色学术润色服务

10. 期刊学术润色服务
11. 润色学术润色服务
12. 润色学术润色服务

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...



COX比例风险模型

经济小海狸

还没有评论

写下你的评论...



赞同 10 添加评论 分享 喜欢 收藏 ...