所以样条回归的cox回归针对的是cox回归中的连续型变量进行操作，而引入时许变量的cox回归其实是针对分组变量或者是分类变量导致的等比例风险假设不满足的情况下进行的操作，从这个角度是否可以对区分这两种不同的回归模型

**用户:**

所以样条回归的cox回归针对的是cox回归中的连续型变量进行操作，而引入时许变量的cox回归其实是针对分组变量或者是分类变量导致的等比例风险假设不满足的情况下进行的操作，从这个角度是否可以对区分这两种不同的回归模型

**元宝:**

你的理解**部分正确**，但需要进一步澄清两者的核心目标和适用变量类型。以下是更准确的区分和总结：

### 1. 样条Cox回归（Spline Cox Regression）

* **核心目标**：解决**连续型变量**与风险比（HR）之间的**非线性关系**，与是否违反PH假设无关。
* **适用变量**：连续型变量（如年龄、血压、BMI等）。
* **操作方式**：通过样条函数（如rcs()或ns()）将连续变量转换为非线性项，允许HR随变量值变化（如U型或J型效应）。
* **与PH假设的关系**：
  + 不直接解决PH假设问题，但能改善模型对非线性效应的拟合。
  + 如果变量效应非线性且同时违反PH假设，可能需要结合时依变量方法。

#### 示例：

* 研究年龄对死亡风险的影响，发现风险在中年最低，年轻和老年较高（U型效应）。
* **R代码**：

fit\_spline <- cph(Surv(time, death) ~ rcs(age, 4) + sex, data = data)

### 2. 时依变量Cox回归（Time-Dependent Cox Regression）

* **核心目标**：解决**任何类型变量**（连续或分类）导致的**PH假设违反**，即HR随时间变化。
* **适用变量**：所有类型变量（连续型、分类变量或分组变量）。
* **操作方式**：
  + 引入时间交互项（如variable \* time）或时依函数（如tt()函数）。
  + 或对违反PH假设的变量进行分层（strata()）。
* **与PH假设的关系**：
  + 直接解决PH假设违反问题，无论变量类型如何。

#### 示例：

* 研究化疗药物（二分类变量：使用/未使用）的效果，发现药物保护效应随时间衰减。
* **R代码**：

# 方法1：时间交互项

fit\_time <- coxph(Surv(time, death) ~ drug + drug:log(time), data = data)

# 方法2：分层模型（针对分类变量）

fit\_strat <- coxph(Surv(time, death) ~ age + strata(drug), data = data)

### 关键区分总结

|  |  |  |
| --- | --- | --- |
| **维度** | **样条Cox回归** | **时依变量Cox回归** |
| **核心问题** | 连续变量的非线性效应 | 任何变量导致的PH假设违反 |
| **变量类型** | 仅连续变量 | 连续变量、分类变量、分组变量均可 |
| **PH假设处理** | 不直接解决PH假设 | 直接解决PH假设 |
| **典型场景** | 年龄与风险呈U型关系 | 药物效果随时间衰减或分组变量HR变化 |

### 注意事项

1. **两类方法可结合使用**：

* 若一个连续变量**同时存在非线性效应和PH假设违反**，可同时使用样条和时依项：

fit\_combined <- coxph(Surv(time, death) ~ ns(age, 3) + age:log(time) + sex, data = data)

1. **分类变量的PH假设处理**：

* 对分类变量（如性别、治疗方案分组），若PH假设违反，优先使用**分层模型**（strata()），而非样条回归：

# 正确：分层处理分类变量

fit\_strat <- coxph(Surv(time, death) ~ age + strata(sex), data = data)

### 操作流程图

graph TD

A[变量类型] -->|连续型| B{效应是否非线性?}

A -->|分类/分组| C{是否违反PH假设?}

B -->|是| D[样条Cox回归]

B -->|否| E[标准Cox回归]

C -->|是| F[时依变量/分层Cox]

C -->|否| E

通过以上分析，可明确两者的区别：

* **样条Cox回归**专攻**连续变量的非线性效应**，与PH假设无关。
* **时依变量Cox回归**专攻**任何变量的PH假设违反**，与变量类型无关。

风险率函数*λ(t)*表示在病人存活至时间*t*后,在随后的趋近于0的*δ*时间段内发生死亡事件的速率。

风险比（Hazard ratio,HR）是指同一时点两组的风险函数之比，这个比即为相对危险度

注意的是我们可能认为对于年龄这个和生存时间高度共线的连续变量，在其违反PH检验的时候，其说明age肯定也是可以用样条回归处理，因为对于和生存时间不高度共线的连续变量，比如血脂值，这个就可以出现time变化时，血脂值得HR一直不变，不违反PH假设，但是会有可能血脂值自身变化时，其HR变化，也就是血脂值得HR随着血脂值的数值变化 对于年龄来说也是可以出现简单模型不违法PH检验但是其会有年龄自身变化时其HR值变化的，即age的样条cox回归模型显著有效，原因是age的时间跨度（一般就是十几年的年龄差距了）一般 远大于生存随访时间（一般3-5年随访时间）

首先绘制普通模型连续变量age的HR~TIME图还有HR~age图

fit\_linear <- cph(Surv(time, death) ~ age + sex, data = data)

zp <- cox.zph(fit\_linear)

chisq df p

age 0.597 1 0.440

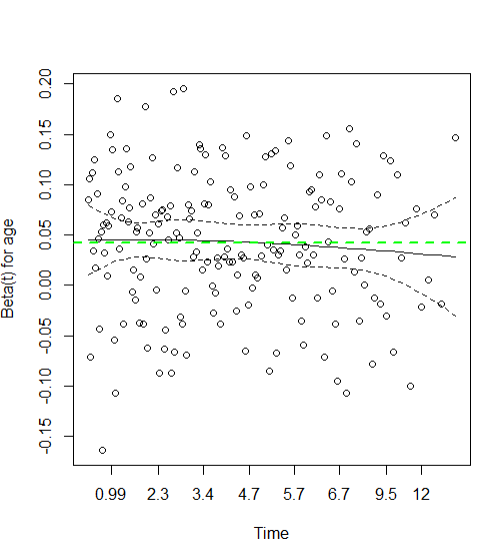
sex 3.443 1 0.064

GLOBAL 3.904 2 0.142

plot(zp[1])

abline(0,0, col="red") # 0水平线

abline(h= fit\_linear $coef[1], col="green", lwd=2, lty=2) # 整体估计



图像进一步说明age满足PH假设检验

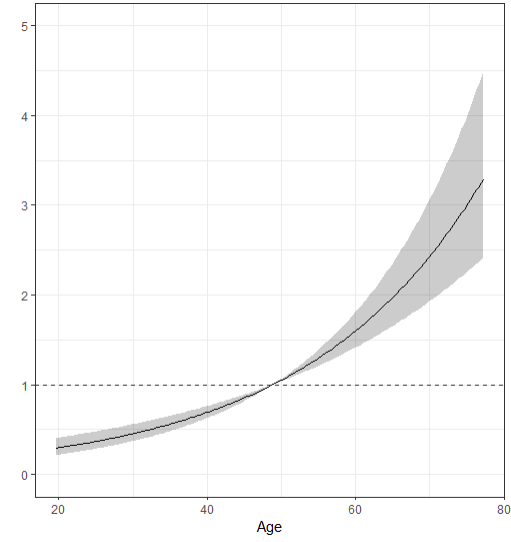
HR~age的绘制

library(ggplot2)

ggplot(Predict(fit\_linear, age,fun = exp, ref.zero = T))+ # fun = exp，计算HR值，不加这句计算概率

geom\_hline(yintercept = 1, color="grey20",linetype=2)+ # HR=1的横线

theme\_bw()



我们可以发现随着age自身的变化，其HR值是在变化的

这时候针对连续变量的疑问点就来了，对于连续变量age在其满足PH假设检验的时候，其=每增加一岁时，其死亡增加的风险都是一样的，也就是age的HR是一定的，也就是管你多少岁HR应该都是一样的，但是在HR~age图中，明显是随着age的变化，HR在发生变化，这应该怎么解释这种看似前后矛盾的结果呢