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

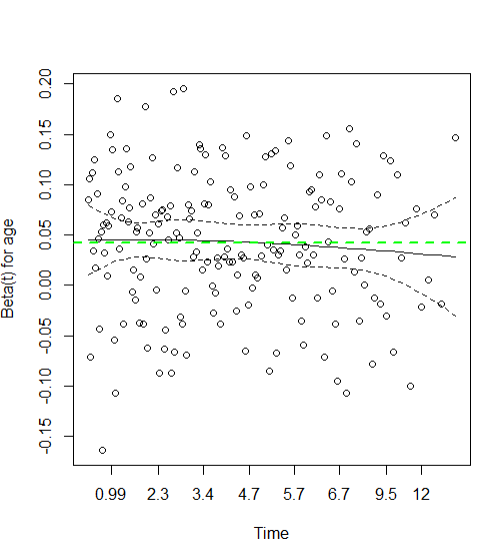
sex 3.443 1 0.064

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

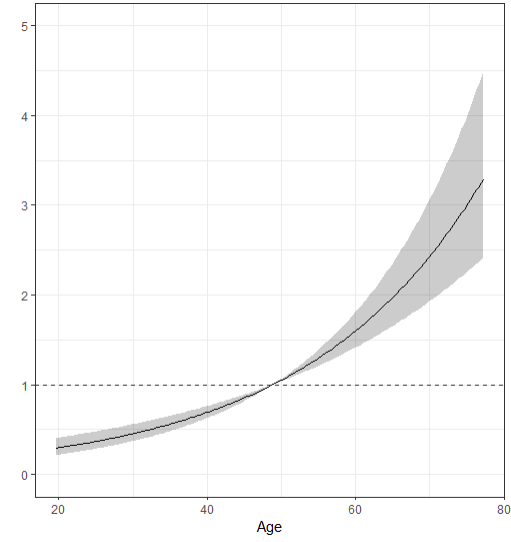
HR~age的绘制

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()

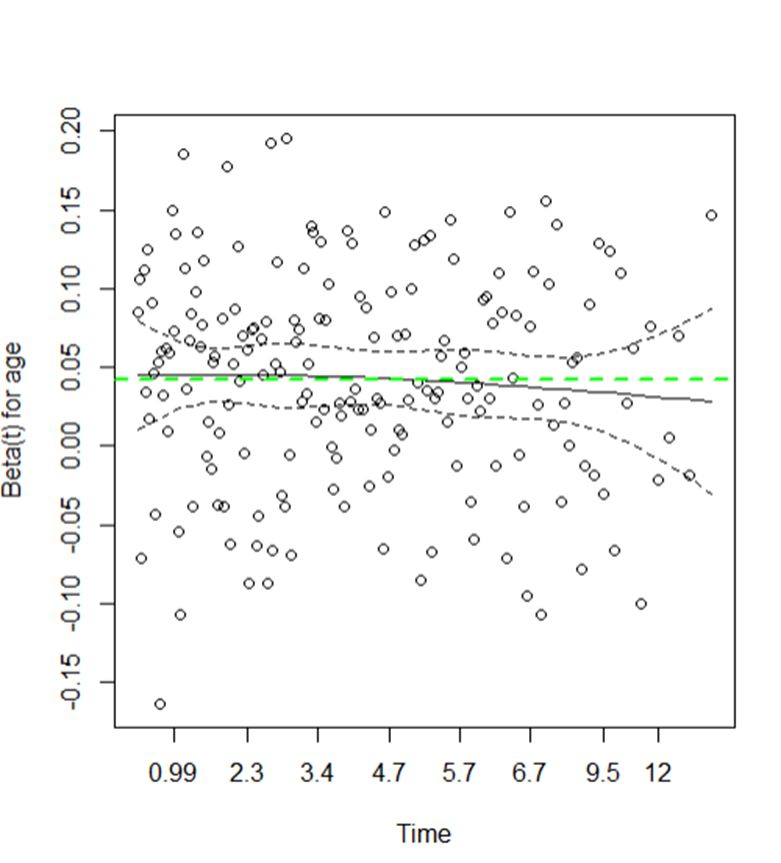


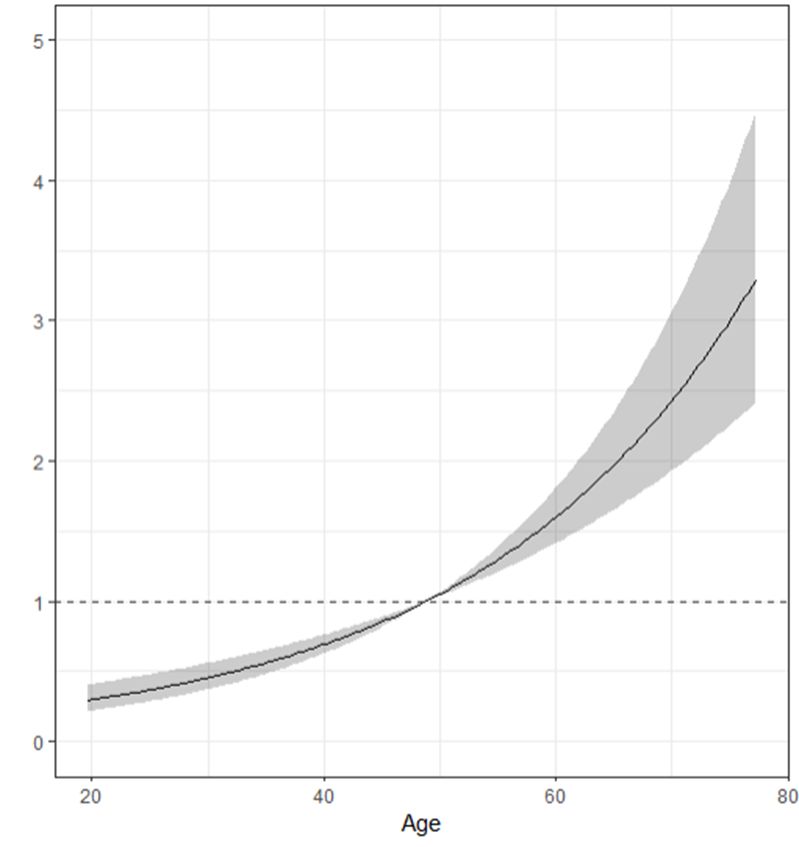
我们可以发现随着age自身的变化，其HR值是在变化的

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

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 也就是随着生存时间的变化，age的HR一直是不变的 但是 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

**用户:**





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

也就是随着生存时间的变化，age的HR一直是不变的

但是

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

**元宝:**

### 1. 核心概念澄清

#### (1) PH假设下的HR（HR ~ Time）加一岁和前一岁的λ(t)的比值

* **PH假设检验**（cox.zph()）验证的是：**“年龄每增加1岁，其HR是否随时间变化”**（即HR是否恒定）。
  + 你的结果中，age的p=0.440（不显著），说明 **HR不随时间变化**（满足PH假设）。
  + 图中Beta(t) for age围绕绿色水平线波动（无时间趋势），印证这一点。

#### (2) HR随年龄的变化（HR ~ Age）现在年龄和参考年龄的λ(t)的比值

* **HR ~ Age图** 展示的是：**“不同年龄的基线风险比（相对于参考年龄）”**。
  + 即使满足PH假设，不同年龄的HR值仍可能不同，因为：
    - **HR的计算基于线性模型**：HR = exp(β \* Δage)。
    - **参考点不同**：图中HR是相对于ref.zero（如年龄=0岁的理论值）的比值，实际年龄越大，Δage越大，HR自然变化。

### 2. 为什么看似矛盾？

#### (1) 两种HR的含义不同

|  |  |  |
| --- | --- | --- |
| **HR类型** | HR ~ Time（PH假设） | HR ~ Age（模型预测） |
| **定义** | 年龄每增加1岁的HR是否随时间变化 | 不同年龄相对于参考年龄的HR |
| **模型要求** | 要求HR不随时间变化（PH假设） | 允许HR随年龄值变化（线性或非线性） |
| **数学表达** | `h(t | age) = h₀(t) \* exp(β\*age)` |

#### (2) 举例说明

假设模型系数 β\_age = 0.05（PH假设成立）：

* **HR ~ Time**：
  + 在任何时间点，年龄每增加1岁，HR均为 exp(0.05) ≈ 1.051（恒定）。
  + 例如：50岁 vs. 49岁的HR在t=1年和t=5年均为1.051。
* **HR ~ Age**：
  + 若参考年龄=40岁，则：
    - 50岁的HR = exp(0.05\*(50-40)) ≈ 1.65
    - 60岁的HR = exp(0.05\*(60-40)) ≈ 2.72
  + **HR随年龄绝对值变化，但每1岁的增量HR仍为1.051**。

### 3. 如何统一理解？

#### (1) 连续变量的HR是“单位增量效应”

* **PH假设** 仅要求 **单位增量（如1岁）的HR不随时间变化**，但允许 **不同年龄的累积HR不同**。
  + 类似“利息复利”：每年利率（单位HR）固定，但本金（年龄基数）越大，总收益（累积HR）越高。

#### (2) 可视化对比

* **HR ~ Time图**（PH检验）：

plot(zp[1])

abline(h = fit\_linear$coef["age"], col="green") # 水平线表示HR恒定

* + 横轴是时间，纵轴是β(t)（应接近水平线）。
* **HR ~ Age图**（模型预测）：

ggplot(Predict(fit\_linear, age, fun=exp, ref.zero=TRUE)) +

geom\_hline(yintercept=1, linetype=2) +

labs(y="HR (vs. age=0)")

* + 横轴是年龄，纵轴是HR相对于年龄=0的比值（呈指数上升）。

#### (3) 数学关系

* 若 h(t|age) = h₀(t) \* exp(β\*age)，则：
  + **单位增量HR**：HR\_per\_1year = exp(β)（恒定，PH假设要求）。
  + **绝对年龄HR**：HR\_age = exp(β\*age)（随年龄变化，非矛盾）。

### 4. 总结

* **PH假设成立**：年龄每增加1岁的HR（如1.05）不随时间变化。
* **HR随年龄变化**：不同年龄的累积HR（如50岁 vs. 40岁）因基数不同而不同，这是模型预期的线性效应。
* **无矛盾**：两者分别描述 **单位增量效应** 和 **累积效应**，本质是同一模型的不同视角。

**建议**：在论文中明确区分“单位增量HR”和“绝对年龄HR”的解释，避免读者混淆。

**用户:**

但是我通过这个代码我怎么知道我的HR ~ Age图中的参考年龄是多少岁呢

**元宝:**

在 **HR ~ Age 图（第2张图）** 中，参考年龄（即 **HR=1** 对应的年龄）可以通过以下方法确定：

### 1. 如何判断参考年龄？

#### (1) 默认参考年龄

* **Predict() 函数的默认行为**：当使用 ref.zero = TRUE 时，参考年龄（HR=1的点）是 **连续变量 age 的中位数**（或模型拟合时的中心化值）。
  + 例如：若数据中 age 的中位数是 50 岁，则 HR=1 对应的年龄为 50 岁。

#### (2) 代码验证参考年龄

通过以下代码可显式查看参考年龄：

# 查看模型中 age 的中心化值（参考点）

ref\_age <- mean(data$age, na.rm = TRUE) # 或 median(data$age)

print(ref\_age)

* **输出示例**：

[1] 50.2

表示参考年龄约为 50 岁（HR=1 的点）。

#### (3) 从图中直接读取

* **第2张图的解释**：
  + 图中 **水平虚线（HR=1）** 与曲线的交点为参考年龄。
  + 若曲线在 age=50 时穿过 HR=1，则参考年龄为 50 岁。
  + 若曲线整体在 HR=1 上方（如你的图中），说明参考年龄可能低于数据范围（如 age=0）。

### 2. 为什么你的图中 HR=1 不在可见范围内？

* **可能原因**：

1. **ref.zero = TRUE 但参考年龄为 0 岁**：

* 若未中心化 age，模型默认 age=0 为参考点（此时 HR=1 对应 age=0，但图中未显示 0 岁）。
* **解决方法**：手动指定参考年龄：

Predict(fit\_linear, age = seq(20, 80, by=5), ref.zero = TRUE, refvalue = 50) # 以50岁为参考

1. **数据中 age 的最小值 > 0**：

* 若年龄范围是 20~80 岁，HR=1 的点可能在 age=20 左侧（不可见）。

### 3. 如何修改参考年龄？

#### (1) 方法1：显式指定参考值

# 以60岁为参考年龄（HR=1）

ggplot(Predict(fit\_linear, age, fun = exp, ref.zero = TRUE, refvalue = 60)) +

geom\_hline(yintercept = 1, linetype = 2) +

labs(y = "HR (vs. age=60)")

#### (2) 方法2：中心化变量

在建模前将 age 中心化，使参考年龄有意义（如以均值或中位数为基准）：

data$age\_centered <- data$age - median(data$age)

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

# 此时 HR=1 对应中位年龄

### 4. 结合你的图片解释

* **第2张图（HR ~ Age）**：
  + 曲线从 HR≈0.5（20岁）上升到 HR≈4（80岁），说明：
    - 若未指定 refvalue，参考年龄可能是 age=0（HR=1 在 0 岁，图中未显示）。
    - **实际意义**：80岁 vs. 20岁的 HR≈4/0.5=8（风险高8倍），但每增加1岁的HR仍恒定（由PH假设保证）。

### 5. 总结

* **默认参考年龄**：由 mean(data$age) 或 median(data$age) 决定，可通过 refvalue 参数修改。
* **图中 HR=1 的位置**：若不在可见范围，需检查是否未中心化或参考年龄设置不合理。
* **关键代码**：

# 显式设置参考年龄为60岁

Predict(fit\_linear, age, ref.zero = TRUE, refvalue = 60)

那么到底什么时候适用样条样的cox回归呢

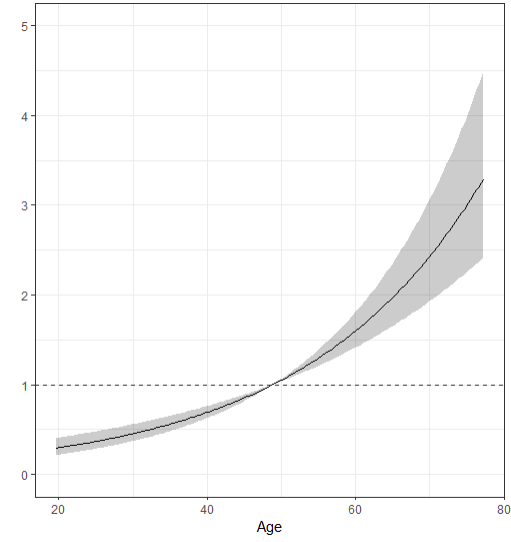
同样举例

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

  
如果在我的简单的cox模型中，我的所有的连续变量的PH假设检验也通过，绘制出来的HR~连续变量图也没有呈现明显的U型的那种明显非线性只是有一点小小的弯曲，这时候我怎们知道对某个变量是否要进行样条化，后续对age进行样条处理

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

并对后续的模型就行检验

anova(fit\_spline )

Wald Statistics Response: Surv(time, death)

Factor Chi-Square d.f. P

age 57.75 3 <.0001

Nonlinear 8.17 2 0.0168

sex 18.75 1 <.0001

TOTAL 75.63 4 <.0001

结果显示age的非线性项显著

且和原本的简单模型进行比较

lrtest(fit\_linear, fit\_spline)

显示

L.R. Chisq d.f. P

8.33954813 2.00000000 0.01545575

显示

若P<0.05提示我们的样条样的cox模型拟合的更好

所以我想知道的是，在一开始我们的简单模型就在我的简单的cox模型中，我的所有的连续变量的PH假设检验也通过，绘制出来的HR~连续变量图也没有呈现明显的U型的那种明显非线性只是有一点小小的弯曲，这时候我怎们知道对某个变量是否要进行样条化

**1. 初步检查：何时考虑样条化？​​**

即使满足以下条件，仍可能需要样条化：

1. ​**PH 假设通过**​（cox.zph p > 0.05）：仅说明 HR 不随时间变化，但 ​**不保证变量与log(HR)的关系是线性的**。
2. ​**HR~变量图有轻微弯曲**​：提示可能存在非线性，但需进一步量化。
3. ​**临床原有知识点支持非线性**​：如年龄对死亡风险的影响可能呈J型或阈值效应。——比如65岁以上房颤卒中风险极具增加

值得注意的是在PH假设检验通过的时候且曲线有点弯曲需要考虑样条样cox

但是PH假设检验不通过的话，就必须对时间分层或者引入时序变量或者时序系数

思维导图如下

