Devoir 5

Mathieu Lemire

2025-11-15

Exercice 1 a)

```r
data <- read.table("boneMarrowTransplant.txt", header = TRUE)

data$t1_cens <- pmin(data$t1, 365)
data$d1_cens <- ifelse(data$t1 > 365, 0, data$d1)

ALL <- data[data$group == 1,]
AML_low_risk <- data[data$group == 2,]
AML_high_risk <- data[data$group == 3,]
```

```r
library(survival)

surv_ALL <- Surv(ALL$t1_cens, ALL$d1_cens)
surv_AML_low <- Surv(AML_low_risk$t1_cens, AML_low_risk$d1_cens)
surv_AML_high <- Surv(AML_high_risk$t1_cens, AML_high_risk$d1_cens)



fit_ALL <- survfit(surv_ALL ~ 1)
fit_AML_low <- survfit(surv_AML_low ~ 1)
fit_AML_high <- survfit(surv_AML_high ~ 1)
```

```r
paste("Fonction de survie pour ALL :",
      round(summary(fit_ALL, times = 365)$surv, 4))
```

```
## [1] "Fonction de survie pour ALL : 0.5996"
```

```r
paste("Fonction de survie pour AML low risk :",
      round(summary(fit_AML_low, times = 365)$surv, 4))
```

```
## [1] "Fonction de survie pour AML low risk : 0.8333"
```

```r
paste("Fonction de survie pour AML high risk :",
      round(summary(fit_AML_high, times = 365)$surv, 4))
```

```
## [1] "Fonction de survie pour AML high risk : 0.4222"
```

Exercice 1 b)

On utilise le test du log-rang ajusté

```
alpha_ajusté <- 0.05 / 3
alpha_ajusté
```

```
## [1] 0.01666667
```

```
test1 <- survdiff(Surv(t1, d1) ~ group, data=data[data$group %in% c(1,2),])
test1$pvalue
```

```
## [1] 0.01519682
```

```
alpha_ajusté > test1$pvalue
```

```
## [1] TRUE
```

On rejette H_0 car $0.01667 > 0.015197$, les fonctions de survie de ALL et AML low risk sont significativement différentes au niveau 0.05 avec la correction de Bonferonni

```
test2 <- survdiff(Surv(t1, d1) ~ group, data=data[data$group %in% c(1,3),])
test2$pvalue
```

```
## [1] 0.1380662
```

```
alpha_ajusté > test2$pvalue
```

```
## [1] FALSE
```

On ne peut pas rejetter H_0 car $0.01667 < 0.13807$, les fonctions de survie de ALL et AML high risk ne sont pas significativement différente.

```
test3 <- survdiff(Surv(t1, d1) ~ group, data=data[data$group %in% c(2,3),])
test3$pvalue
```

```
## [1] 0.0001474087
```

```
alpha_ajusté > test3$pvalue
```

```
## [1] TRUE
```

On rejette H_0 car $0.01667 > 0.000147$ les fonctions de survie de AML low risk et AML high risk sont significativement différente

Exercice 2 a)

$$\begin{aligned}
S(t \mid x_1) &= \int_{-\infty}^{\infty} S(t \mid x_1, x_2) g(x_2 \mid x_1) dx_2 \\
&= \int_{-\infty}^{\infty} \exp[-H(t \mid x_1, x_2)] g(x_2 \mid x_1) dx_2 \\
&= \int_{-\infty}^{\infty} \exp\left[-\int_0^t h(u \mid x_1, x_2) du\right] g(x_2 \mid x_1) dx_2 \\
&= \int_{-\infty}^{\infty} \exp\left[-\int_0^t h_o(u) \exp(\beta_1 x_1 + \beta_2 x_2) du\right] g(x_2 \mid x_1) dx_2 \\
&= \int_{-\infty}^{\infty} \exp\left[-\exp(\beta_1 x_1 + \beta_2 x_2) \int_0^t h_o(u) du\right] g(x_2 \mid x_1) dx_2 \\
&= \int_{-\infty}^{\infty} \exp\left[-e^{\beta_1 x_1 + \beta_2 x_2} H_o(t)\right] g(x_2 \mid x_1) dx_2
\end{aligned}$$

$$\begin{aligned}
h(t \mid x_1) &= -\frac{d}{dt} \log(S(t \mid x_1)) \\
&= -\frac{d}{dt} \log\left(\int_{-\infty}^{\infty} \exp\left[-e^{\beta_1 x_1 + \beta_2 x_2} H_o(t)\right] g(x_2 \mid x_1) dx_2\right)
\end{aligned}$$

On ne peut pas écrire $h(t \mid x_1) = h_o(t \mid x_1) c(\beta^\top Z)$, donc ce n'est pas un modèle de risque proportionnel.

Exercice 2 b)

Si X_1 est indépendant de X_2, on aurait

$$h(t \mid x_1) = -\frac{d}{dt} \log\left(\int_{-\infty}^{\infty} \exp\left[-e^{\beta_1 x_1 + \beta_2 x_2} H_o(t)\right] g(x_2) dx_2\right)$$

On ne peut toujours pas l'écrire sous la forme $h(t \mid x_1) = h_o(t \mid x_1) c(\beta^\top Z)$. Il n'est toujours pas un modèle de risque proportionnel.

Si X_1 est indépendant de X_2 et X_1 est binaire, on aurait

$$h(t \mid X_1 = 0) = -\frac{d}{dt} \log\left(\int_{-\infty}^{\infty} \exp\left[-e^{\beta_2 x_2} H_o(t)\right] g(x_2) dx_2\right)$$

Et

$$h(t \mid X_1 = 1) = -\frac{d}{dt} \log\left(\int_{-\infty}^{\infty} \exp\left[-e^{\beta_1 + \beta_2 x_2} H_o(t)\right] g(x_2) dx_2\right)$$

Leur risque relatif étant

$$\frac{h(t \mid X_1 = 1)}{h(t \mid X_1 = 0)} = \frac{-\frac{d}{dt}\log\left(\int_{-\infty}^{\infty} \exp\left[-e^{\beta_1+\beta_2 x_2} H_o(t)\right] g(x_2) dx_2\right)}{-\frac{d}{dt}\log\left(\int_{-\infty}^{\infty} \exp\left[-e^{\beta_2 x_2} H_o(t)\right] g(x_2) dx_2\right)} \neq \exp(\beta_1)$$

Donc leur risque relatif n'est pas constant par rapport à t, le modèle n'est toujours pas un modèle de risque proportionnel.