

Gabarito da 3ª Lista de Exercícios - MAE514

Professora: GISELA TUNES

Monitor: RODRIGO PASSOS MARTINS

Exercício 1

Em um estudo com pacientes com mieloma múltiplo, os pacientes foram tratados com tratamento padrão e os tempos de sobrevivência (em meses) dos 25 pacientes, contados a partir do início do tratamento, estão disponíveis abaixo (tempos censurados à direita são denotados por um sinal “+”):

- 0,3; 5,9; 20,8; 1,7; 7,2; 2,1; 6,4; 2,5; 2,3; 0,3; 0,4; 0,6; 23,0; 6,9; 2,1; 42,6; 0,8
- 28,0+; 65,4+; 64,9+; 73,6+; 42,6+; 48,0+; 43,6+; 12,0+

Suponha que deseja-se fazer um estudo para comparar o tratamento padrão com um novo tratamento. Esse novo estudo terá um período de acompanhamento total igual a 4 anos e meio e espera-se que a mortalidade inicial seja reduzida e, após um ano e meio de tratamento, ainda se tenha em torno de 65% de pacientes vivos.

$$S_N(1,5 \text{ anos}) = S_N(18 \text{ meses}) = 65\%$$

a)

Vamos obter o tamanho da amostra necessário se o recrutamento for feito por um período de 2 anos ($a = 2$ anos) e, nos próximos dois anos e meio ($f = 2,5$ anos), for feito apenas o acompanhamento desses pacientes.

Primeiro, vamos calcular o estimador de KM para esses pacientes:

```
# CALCULANDO O ESTIMADOR DE KM (NO R)

Tempo <- c(0.3, 5.9, 20.8, 1.7, 7.2, 2.1, 6.4, 2.5, 2.3,
           0.3, 0.4, 0.6, 23.0, 6.9, 2.1, 42.6, 0.8,
           28.0, 65.4, 64.9, 73.6, 42.6, 48.0, 43.6, 12.0)
Evento <- c(rep(1, 17), rep(0, 8))

dados <- data.frame(Tempo, Evento)

library(survival)
library(survminer)

## Loading required package: ggplot2
## Loading required package: ggpubr
## Loading required package: magrittr
```

```
S_KM <- survfit(Surv(Tempo, Evento) ~ 1, data = dados)
summary(S_KM)
```

```
## Call: survfit(formula = Surv(Tempo, Evento) ~ 1, data = dados)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   0.3    25      2    0.920  0.0543    0.820    1.000
##   0.4    23      1    0.880  0.0650    0.761    1.000
##   0.6    22      1    0.840  0.0733    0.708    0.997
##   0.8    21      1    0.800  0.0800    0.658    0.973
##   1.7    20      1    0.760  0.0854    0.610    0.947
##   2.1    19      2    0.680  0.0933    0.520    0.890
##   2.3    17      1    0.640  0.0960    0.477    0.859
##   2.5    16      1    0.600  0.0980    0.436    0.826
##   5.9    15      1    0.560  0.0993    0.396    0.793
##   6.4    14      1    0.520  0.0999    0.357    0.758
##   6.9    13      1    0.480  0.0999    0.319    0.722
##   7.2    12      1    0.440  0.0993    0.283    0.685
##  20.8    10      1    0.396  0.0986    0.243    0.645
##  23.0     9      1    0.352  0.0970    0.205    0.604
##  42.6     7      1    0.302  0.0953    0.162    0.560
```

O comando **summary()** nos dá os valores referentes a estimativa de Kaplan-Meier. Por ele, temos que a estimativa da sobrevida após 18 meses com o tratamento padrão é $\hat{S}_p(18 \text{ meses}) = 0,440$ (poderíamos usar interpolação para o cálculo exato, mas 7,2 meses é o maior tempo, entre os registrados no estudo piloto, menor ou igual a 18 meses).

Os pesquisadores esperam que a mortalidade inicial seja reduzida para 65% após 18 meses. Assim, supondo riscos proporcionais:

$$S_N(18 \text{ meses}) = [S_P(18 \text{ meses})]^{\psi_R} \Leftrightarrow \psi_R \cdot \log\{S_N(18 \text{ meses})\} = \log\{S_P(18 \text{ meses})\} \Leftrightarrow \psi_R = \frac{\log\{S_P(18 \text{ meses})\}}{\log\{S_N(18 \text{ meses})\}}$$

Assim,

$$\psi_R = \frac{\log\{S_P(18 \text{ meses})\}}{\log\{S_N(18 \text{ meses})\}} = \frac{\log\{0,65\}}{\log\{0,44\}} \cong 0,5247$$

Além disso,

$$\theta_R = \log(\psi_R) = \log(0,5247) \cong 0,6449$$

Agora, vamos calcular o números de falhas considerando 5% e 8% de nível de significância do teste e poder igual a 80%, 85% e 90%, com:

$$d = \frac{4}{\theta_R^2} (z_{\alpha/2} + z_\beta)^2$$

Os valores dessas falhas estão na tabela abaixo:

Poder	Nível de significância de 5%	Nível de significância de 8%
80%	$d = \frac{4}{0,6449^2} (1,96 + 0,8416)^2 \cong 75,49$	$d = \frac{4}{0,6449^2} (1,7507 + 0,8416)^2 \cong 64,63$
85%	$d = \frac{4}{0,6449^2} (1,96 + 1,0364)^2 \cong 86,35$	$d = \frac{4}{0,6449^2} (1,7507 + 1,0364)^2 \cong 74,71$
90%	$d = \frac{4}{0,6449^2} (1,96 + 1,2816)^2 \cong 101,06$	$d = \frac{4}{0,6449^2} (1,7507 + 1,2816)^2 \cong 88,43$

Como destacado anteriormente, $a = 24$ e $f = 30$. Com isso, vamos calcular:

$$\hat{p}_{F|P} = \hat{P}(\text{falha} \mid \text{padrão}) \approx 1 - \frac{1}{6} \left[\hat{S}_P(f) + 4\hat{S}_P\left(f + \frac{a}{2}\right) + \hat{S}_P(f + a) \right] = 1 - \frac{1}{6} \left[\hat{S}_P(30) + 4\hat{S}_P(42) + \hat{S}_P(54) \right]$$

Assim,

$$\hat{p}_{F|P} \approx 1 - \frac{1}{6} \left[\hat{S}_P(30) + 4\hat{S}_P(42) + \hat{S}_P(54) \right] = 1 - \frac{1}{6} [0,352 + 4 \cdot 0,352 + 0,302] \cong 0,6563$$

Para as falhas do tratamento novo, podemos estimar usando o coeficiente de referência ψ_R com:

$$\hat{S}_N(t) = [\hat{S}_P(t)]^2$$

Portanto,

- $\hat{S}_N(30) = (0,352)^{0,5247} = 0,5782$
- $\hat{S}_N(42) = (0,352)^{0,5247} = 0,5782$
- $\hat{S}_N(54) = (0,302)^{0,5247} = 0,5335$

Analogamente ao tratamento padrão, temos para o novo que:

$$\hat{p}_{F|N} \approx 1 - \frac{1}{6} \left[\hat{S}_N(30) + 4\hat{S}_N(42) + \hat{S}_N(54) \right] = 1 - \frac{1}{6} [0,352 + 4 \cdot 0,352 + 0,302] \cong 0,6563$$

Logo, podemos calcular,

$$\hat{p}_F = \frac{\hat{p}_{F|P} + \hat{p}_{F|N}}{2} = \frac{0,6563 + 0,4293}{2} = 0,5428$$

Tendo os valores de \hat{p}_F , vamos calcular o tamanho amostral para cada cenário com:

$$n = \frac{d}{\hat{p}_F}$$

Logo, os valores amostrais estão na tabela seguinte:

Poder	Nível de significância de 5%	Nível de significância de 8%
80%	$n = \frac{75,49}{0,5428} = 140$	$n = \frac{64,63}{0,5428} = 120$
85%	$n = \frac{86,35}{0,5428} = 160$	$n = \frac{74,71}{0,5428} = 138$
90%	$n = \frac{101,06}{0,5428} = 187$	$n = \frac{88,43}{0,5428} = 163$

b)

Agora, vamos assumir que o recrutamento poderá ser feito ao longo dos 4 anos e meio ($a = 54$ meses) e ($f = 0$ meses). Assim, temos que as falhas são iguais as calculadas no item a) e precisamos recalculiar $\hat{p}_{F|P}$, $\hat{p}_{F|N}$ e \hat{p}_F .

Para tal, façamos:

$$\hat{p}_{F|P} \approx 1 - \frac{1}{6} \left[\hat{S}_P(0) + 4\hat{S}_P(27) + \hat{S}_P(54) \right] = 1 - \frac{1}{6} [1 + 4.0,352 + 0,302] \cong 0,5483$$

Com isso,

- $\hat{S}_N(0) \cong (1)^{0,5247} = 1,0000$
- $\hat{S}_N(27) \cong (0,352)^{0,5247} = 0,5782$
- $\hat{S}_N(54) \cong (0,302)^{0,5247} = 0,5335$

Analogamente ao tratamento padrão, temos para o novo que:

$$\hat{p}_{F|N} \approx 1 - \frac{1}{6} \left[\hat{S}_N(0) + 4\hat{S}_N(27) + \hat{S}_N(54) \right] = 1 - \frac{1}{6} [1 + 4.0,5782 + 0,5335] \cong 0,3590$$

Logo, podemos calcular,

$$\hat{p}_F = \frac{\hat{p}_{F|P} + \hat{p}_{F|N}}{2} = \frac{0,5483 + 0,3590}{2} = 0,4537$$

Tendo os valores de \hat{p}_F , vamos calcular os tamanhos amostrais:

Poder	Nível de significância de 5%	Nível de significância de 8%
80%	$n = \frac{75,49}{0,4537} = 167$	$n = \frac{64,63}{0,4537} = 143$
85%	$n = \frac{86,35}{0,4537} = 191$	$n = \frac{74,71}{0,4537} = 165$
90%	$n = \frac{101,06}{0,4537} = 223$	$n = \frac{88,43}{0,4537} = 195$

Podemos notar que com o tempo de recrutamento maior, os tamanhos amostrais ideais também aumentaram.

Exercício 2

Considerando $a_1 < \dots < a_J$ uma partição do eixo do tempo. A função de sobrevivência da distribuição Exponencial Segmentada (*Piecewise Exponential*) é dada por

$$S(t) = \exp \left\{ - \sum_{j=1}^J \lambda_j \nabla_j(t) \right\}, \quad t > 0,$$

em que,

$$\nabla_j(t) = \begin{cases} 0, & \text{se } t < a_{j-1} \\ t - a_{j-1}, & \text{se } a_{j-1} \leq t < a_j \\ a_j - a_{j-1}, & \text{se } t \geq a_j \end{cases}$$

para $j = 1, \dots, J$. Além disso, temos que $a_0 = 0$ e $a_{J+1} = \infty$.

a)

Vamos mostrar que a função de taxa de falha desta distribuição é constante no intervalo (a_j, a_{j+1}) , $\forall j = 0, 1, \dots, J$.

Lembremos que,

$$\alpha(t) = \frac{-d(\log S(t))}{dt}$$

Então,

$$\alpha(t) = \frac{-d \left(\log \left(\exp \left\{ - \sum_{j=1}^J \lambda_j \nabla_j(t) \right\} \right) \right)}{dt} = \frac{-d \left(- \sum_{j=1}^J \lambda_j \nabla_j(t) \right)}{dt} = \sum_{j=1}^J \lambda_j \nabla_j'(t)$$

Analisando $\nabla_j(t)$, a sua derivada $\nabla_j'(t)$ é dada por:

$$\nabla_j'(t) = \begin{cases} 0, & \text{se } t < a_{j-1} \\ 1, & \text{se } a_{j-1} \leq t < a_j \\ 0, & \text{se } t \geq a_j \end{cases}$$

Portanto,

$$\alpha(t) = \sum_{j=1}^J \lambda_j \nabla_j'(t) = \sum_{j=1}^J \lambda_j \cdot [0 \cdot \mathbb{1}(t < a_{j-1}) + 1 \cdot \mathbb{1}(a_{j-1} \leq t < a_j) + 0 \cdot \mathbb{1}(t \geq a_j)] = \sum_{j=1}^J \lambda_j \cdot \mathbb{1}(a_{j-1} \leq t < a_j)$$

Logo, no intervalo $(a_{j-1} \leq t < a_j)$, a função de taxa de falha é constante $\forall j = 0, 1, \dots, J$.

$$\alpha(t) = \sum_{j=1}^J \lambda_j \cdot \mathbb{1}(a_{j-1} \leq t < a_j) = \sum_{j=1}^J \lambda_j, \text{ no intervalo } (a_{j-1} \leq t < a_j) \blacksquare$$

b)

Para a amostra (censurada) de tamanho n , o estimador de máxima verossimilhança do vetor $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_J)$ é dada por $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_J)$, com:

$$\hat{\lambda}_j = \frac{d_j}{\sum_{i=1}^n \nabla_j(t_i)}, \text{ com } j = 1, \dots, J$$

Para essa amostra, temos que:

$$L(\boldsymbol{\lambda}) \propto \prod_{i=1}^n [\alpha(t_i|\boldsymbol{\lambda})]^{\delta_i} \cdot S(t_i|\boldsymbol{\lambda}), \text{ com } \delta_i = \mathbb{1}(\text{falha em } t_i)$$

Aplicando o \log , temos:

$$l(\boldsymbol{\lambda}) \propto \sum_{i=1}^n \delta_i \cdot \log(\alpha(t_i|\boldsymbol{\lambda})) + \sum_{i=1}^n S(t_i|\boldsymbol{\lambda}) = \sum_{i=1}^n \delta_i \cdot \log \left(\sum_{j=1}^J \lambda_j \cdot \mathbb{1}(a_{j-1} \leq t < a_j) \right) + \sum_{i=1}^n \left(- \sum_{j=1}^J \lambda_j \nabla_j(t_i) \right)$$

Derivando em λ_j , temos que:

$$\frac{dl(\boldsymbol{\lambda})}{d\lambda_j} = \frac{\sum_{i=1}^n \delta_i \cdot \mathbb{1}(a_{j-1} \leq t < a_j)}{\sum_{j=1}^J \lambda_j \cdot \mathbb{1}(a_{j-1} \leq t < a_j)} + \sum_{i=1}^n \left(- \sum_{j=1}^J \lambda_j \nabla_j(t_i) \right)$$

Como visto na letra **a)**, no intervalo $[a_{j-1}, a_j)$, temos a constante. Assim,

$$\frac{dl(\boldsymbol{\lambda})}{d\lambda_j} = 0 \Leftrightarrow \frac{\sum_{i: t_i \in [a_{j-1}, a_j)} d_i}{\hat{\lambda}} - \sum_{i=1}^n \nabla_j(t_i) = 0 \Leftrightarrow \frac{d_j}{\hat{\lambda}} = \sum_{i=1}^n \nabla_j(t_i) \Leftrightarrow \hat{\lambda}_j = \frac{d_j}{\sum_{i=1}^n \nabla_j(t_i)}, \text{ com } j = 1, \dots, J \blacksquare$$

c)

Vamos mostrar que essa distribuição pertence à classe de modelos de riscos proporcionais em um modelo de regressão com $\lambda_j = \exp\{\beta_{0j} + x^T \beta\}$.

Para tal, vamos provar que:

$$\frac{\alpha(t|X_j)}{\alpha(t|X_i)} = k, \text{ sendo } k \text{ constante}$$

Pelo o que já vimos no item **a)**, a função $\alpha(t)$ é diferente de zero apenas no intervalo $[a_{j-1}, a_j)$:

$$\alpha(t) = \sum_{j=1}^J \lambda_j$$

Assim, vamos fazer a razão:

$$\frac{\alpha(t|X_j)}{\alpha(t|X_i)} = \frac{\sum_{j=1}^J \exp(\beta_{0j} + x^T \cdot \beta)}{\sum_{i=1}^J \exp(\beta_{0i} + x^T \cdot \beta)} = \frac{\sum_{j=1}^J \exp(\beta_{0j}) \cdot \exp(x^T \cdot \beta)}{\sum_{i=1}^J \exp(\beta_{0i}) \cdot \exp(x^T \cdot \beta)} = \frac{\exp(x^T \cdot \beta) \cdot \sum_{j=1}^J \exp(\beta_{0j})}{\exp(x^T \cdot \beta) \cdot \sum_{i=1}^J \exp(\beta_{0i})} = \frac{\sum_{j=1}^J \exp(\beta_{0j})}{\sum_{i=1}^J \exp(\beta_{0i})}$$

Como a expressão final não depende de t , esta distribuição pertence à classe de modelos de riscos proporcionais em um modelo de regressão com $\lambda_j = \exp\{\beta_{0j} + x^T \beta\}$. ■

Exercício 3

Os dados no arquivo HOD_NHL.csv são referentes a 43 pacientes com linfoma de Hodgking ou linfoma não Hodgking, submetidos a transplante de medula óssea. Alguns pacientes receberam transplante de doador aparentado compatível (transplante alogênico) e outros receberam transplante autólogo (ou seja, a própria medula óssea do paciente é coletada e posteriormente infundida). Nos dados também há informação sobre o escore de Karnofsky, que é uma medida de performance que classifica os pacientes segundo o bem-estar dos pacientes.

O objetivo principal do estudo é a comparação dos tipos de transplante, considerando-se o tempo (dias) livre de doença (i.e., tempo antes de ocorrência da recorrência ou óbito).

Primeiro, vamos importar os dados do exercício:

```
library(readr)
dados <- read_delim("HOD_NHL.csv", ";", escape_double = FALSE, trim_ws = TRUE)

## Parsed with column specification:
## cols(
##   Graft = col_double(),
##   Time = col_double(),
##   D_R = col_double(),
##   Karnofsky = col_double()
## )

names(dados) <- c("Transplante", "Tempo", "D_R", "Karnofsky")

library(magrittr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

dados %<>% mutate(Transplante = ifelse(Transplante == 1, "Alogênico", "Autólogo"))
```

a)

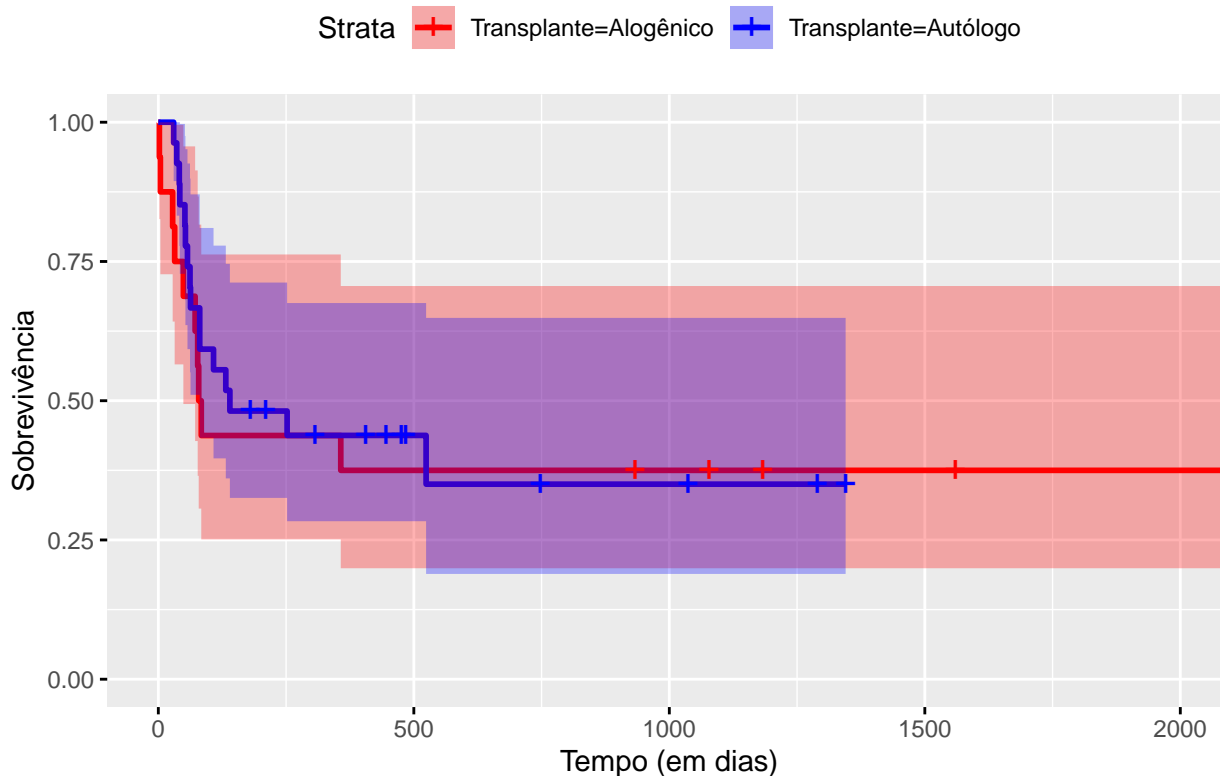
Vamos construir as curvas de Kaplan-Meier para o tempo de sobrevivência dos pacientes com linfoma para cada um dos grupos de tratamento:

```
# FAZENDO O GRÁFICO DE KM (EM R)
S_KM_Trans <- survfit(Surv(Tempo, D_R) ~ Transplante, data = dados)
```



```
ggsurvplot(S_KM_Trans, data = dados, palette = c('red', 'blue'),
  conf.int = T, ggtheme=theme_gray()) +
  labs(x = "Tempo (em dias)", y = "Sobrevivência",
  title = "Estimativas de Kaplan-Meier (Transplante)")
```

Estimativas de Kaplan-Meier (Transplante)



Analisando o gráfico, vemos que as curvas de sobrevivência são parecidas entre os transplantes, indicando que os dois tipos de procedimento levam aos mesmos resultados.

Agora, vamos categorizar a variável score de Karnofsky, criando uma variável com duas categorias (score < 80 ou score ≥ 80):

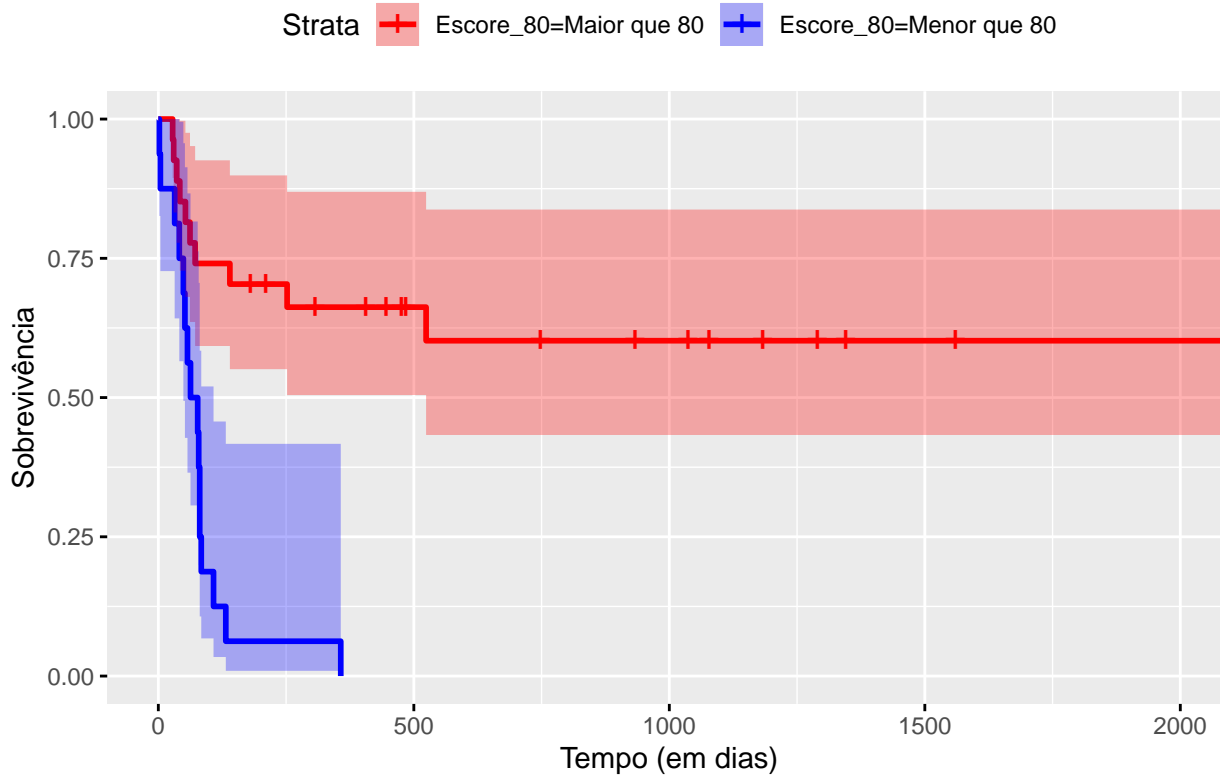
```
dados %<>% mutate(Escore_80 = ifelse(Karnofsky >= 80, "Maior que 80", "Menor que 80"))
```

Assim, podemos construir os gráficos de Kaplan-Meier para essa nova variável categorizada:

```
# FAZENDO O GRÁFICO DE KM (EM R)
S_KM_Escore <- survfit(Surv(Tempo, D_R) ~ Escore_80, data = dados)

ggsurvplot(S_KM_Escore, data = dados, palette = c('red', 'blue'),
  conf.int = T, ggtheme=theme_gray()) +
  labs(x = "Tempo (em dias)", y = "Sobrevivência",
  title = "Estimativas de Kaplan-Meier (Escore < 80)")
```

Estimativas de Kaplan–Meier (Escore < 80)



A sobrevivência dos pacientes com o escore ≥ 80 é muito maior do que a dos pacientes com escore < 80 . Isso é um indicativo de que o alto bem-estar dos pacientes está associado com um alto tempo de sobrevivência.

b)

Vamos ajustar um modelo de regressão Weibull com a variável tratamento e a variável escore de Karnofsky categorizada no item a):

```
# AJUSTANDO O MODELO WEIBULL (EM R)
modelo_weibull <- survreg(Surv(Tempo, D_R) ~ Transplante + Escore_80,
                          data = dados, dist = "weibull")
modelo_weibull
```

```
## Call:
## survreg(formula = Surv(Tempo, D_R) ~ Transplante + Escore_80,
##         data = dados, dist = "weibull")
##
## Coefficients:
##      (Intercept)  TransplanteAutólogo Escore_80Menor que 80
##           7.9442140          -0.5500871          -3.3250980
##
## Scale= 1.300982
##
```

```
## Loglik(model)= -168.4   Loglik(intercept only)= -183.3
##  Chisq= 29.72 on 2 degrees of freedom, p= 3.51e-07
## n= 43
```

Considerando a distribuição do valor extremo, temos as seguintes estimativas:

- $\hat{\gamma}_0$ (Intercept): 8,494301
- $\hat{\gamma}_1$ (Transplante): -0,550087
- $\hat{\gamma}_2$ (Escore80 (Menor que 80)): -3,325098
- $\hat{\sigma}$ (Scale): 1,300982

Temos que reparametrizar a saída para os parâmetros de máxima verossimilhança do modelo Weibull como um modelo de localização-escala:

- $\hat{\rho} = \frac{1}{\hat{\sigma}} = \frac{1}{1,300982} = 0,768650$
- $\hat{\beta}_0 = \exp\left(\frac{-\hat{\mu}}{\hat{\sigma}}\right) = \exp\left(\frac{-8,494301}{1,300982}\right) = \exp(-6,529146) = 0,001460$
- $\hat{\beta}_1 = \frac{-\hat{\gamma}_1}{\hat{\sigma}} = \frac{-(-0,550087)}{1,300982} = 0,422825$
- $\hat{\beta}_2 = \frac{-\hat{\gamma}_2}{\hat{\sigma}} = \frac{-(-3,325098)}{1,300982} = 2,555837$

Para que não seja necessária essa reparametrização, o R disponibiliza uma biblioteca chamada **SurvRegCensCov**, que transforma os resultados do comando **survreg** da distribuição valor extremo para a parametrização do modelo Weibull:

```
# AJUSTANDO O MODELO WEIBULL (EM R)
library(SurvRegCensCov)

## Registered S3 method overwritten by 'SurvRegCensCov':
##   method      from
##   print.src    dplyr

objeto_modelo <- Surv(time = dados$Tempo, event = dados$D_R)

modelo_weibull_2 <- WeibullReg(objeto_modelo ~ dados$Transplante + dados$Escore_80)

modelo_weibull_2

## $formula
## objeto_modelo ~ dados$Transplante + dados$Escore_80
##
## $coef
##                               Estimate          SE
```

```
## lambda                0.002228736 0.002256216
## gamma                 0.768650076 0.131172353
## dados$TransplanteAutólogo 0.422824485 0.411928401
## dados$Escore_80Menor que 80 2.555836817 0.490132578
##
## $HR
##                HR                LB                UB
## dados$TransplanteAutólogo 1.526266 0.6807635 3.421877
## dados$Escore_80Menor que 80 12.882075 4.9292883 33.665683
##
## $ETR
##                ETR                LB                UB
## dados$TransplanteAutólogo 0.57689956 0.20626795 1.6134989
## dados$Escore_80Menor que 80 0.03596899 0.01192999 0.1084467
##
## $summary
##
## Call:
## survival::survreg(formula = formula, data = data, dist = "weibull")
##                Value Std. Error      z      p
## (Intercept)      7.944      0.546 14.54 < 2e-16
## dados$TransplanteAutólogo -0.550      0.525 -1.05 0.29
## dados$Escore_80Menor que 80 -3.325      0.563 -5.91 3.5e-09
## Log(scale)         0.263      0.171 1.54 0.12
##
## Scale= 1.3
##
## Weibull distribution
## Loglik(model)= -168.4 Loglik(intercept only)= -183.3
## Chisq= 29.72 on 2 degrees of freedom, p= 3.5e-07
## Number of Newton-Raphson Iterations: 5
## n= 43
```

Note que os parâmetros do modelo estão numericamente próximos aos reparametrizados anteriormente.

Agora, vamos ajustar o modelo Weibull como um modelo de riscos proporcionais.

Considerando duas unidades experimentais com covariáveis \mathbf{X}_1 e \mathbf{X}_2 , temos que:

$$\frac{\alpha_T(t|\mathbf{X}_1)}{\alpha_T(t|\mathbf{X}_2)} = \frac{\theta_0 \cdot \rho \cdot t^{\rho-1} \cdot \exp(\mathbf{X}_1^T \cdot \boldsymbol{\beta})}{\theta_0 \cdot \rho \cdot t^{\rho-1} \cdot \exp(\mathbf{X}_2^T \cdot \boldsymbol{\beta})} = \exp[(\mathbf{X}_1^T - \mathbf{X}_2^T) \cdot \boldsymbol{\beta}] = K(\mathbf{X}_1, \mathbf{X}_2)$$

Note que $K(\mathbf{X}_1, \mathbf{X}_2)$ é uma constante que não depende de t e que representa a proporcionalidade das taxas de falha.

Nesse caso, uma estimativa para o parâmetro $\boldsymbol{\beta}$ é a relação direta de $\hat{\beta}_1$ e $\hat{\beta}_2$ que vimos anteriormente:

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \text{Intercepto} \\ \text{Coef. Transplante} \\ \text{Coef. Escore 80} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0,001460 \\ 0,422825 \\ 2,555837 \end{pmatrix}$$

c)

Vamos encontrar uma estimativa pontual para a razão de taxas de falha de pacientes que receberam transplante autólogo e alogênico. Para tal, temos a seguinte razão:

$$\frac{\alpha_T(t|\mathbf{X}_{aut})}{\alpha_T(t|\mathbf{X}_{alo})} = \exp[(\mathbf{X}_{aut}^T - \mathbf{X}_{alo}^T) \cdot \beta] = K(\mathbf{X}_1, \mathbf{X}_2)$$

Vamos considerar que \mathbf{X}_{aut}^T e \mathbf{X}_{alo}^T são vetores que indicam que os pacientes receberam transplante autólogo e alogênico, respectivamente. Além disso, vamos considerar que eles possuem o mesmo escore de Karnofsky, ou seja:

$$\mathbf{X}_{aut} = \begin{pmatrix} 1 \\ \mathbb{1}(\text{indicador de Autólogo}) \\ \mathbb{1}(\text{indicador do escore} < 80) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ k \end{pmatrix} \text{ e } \mathbf{X}_{alo} = \begin{pmatrix} 1 \\ \mathbb{1}(\text{indicador de Autólogo}) \\ \mathbb{1}(\text{indicador do escore} < 80) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ k \end{pmatrix}$$

Sendo assim,

$$\exp[(\mathbf{X}_{aut}^T - \mathbf{X}_{alo}^T) \cdot \beta] = \exp[(0 \ 1 \ 0) \cdot (0,001460 \ 0,422825 \ 2,555837)^T] = \exp(0,422825) = 1,526267$$

Então, a taxa de falha dos pacientes que receberam o transplante autólogo é 52,6% maior que os pacientes que receberam transplante alogênico.

Agora, vamos encontrar uma estimativa do fator de aceleração. Tendo em vista os resultados do item b), podemos escrever que:

$$T_i = \exp\{\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \sigma \omega_i\}$$

Vamos definir como $\mu_i = \gamma_0 + \sigma \omega_i$, então:

$$T_i = \exp\{\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \sigma \omega_i\} = \exp\{\mu_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2}\}$$

Para esse indivíduo i , a função de sobrevivência é dada por:

$$S_i(t) = P(T > t | x_i) = P(T_i > t) = P(\exp\{\mu_i + \gamma_1 x_{i1} + \gamma_2 x_{i2}\} > t) = P(\mu_i > \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\} \cdot t)$$

Vamos reescrever como a função de sobrevivência:

$$S_i(t) = P(\mu_i > \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\} \cdot t) = S_0(\exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\} \cdot t)$$

Portanto,

$$\psi_{x_i} = \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\}$$

A estimativa de ψ é dada por:

$$\hat{\psi}_{x_i} = \exp\{-\hat{\gamma}_1 x_{i1} - \hat{\gamma}_2 x_{i2}\} = \exp\{-(-0,550087)x_{i1} - (-3,325098)x_{i2}\} = \exp\{0,550087x_{i1} + 3,325098x_{i2}\}$$

Como x_{i1} é $\mathbb{1}(\text{indicador de Autólogo})$ e x_{i2} é $\mathbb{1}(\text{indicador do escore} < 80)$, temos que:

- Fator de aceleração para pacientes com transplante autólogo: $\exp\{0,550087+3,325098.1(\text{indicador do escore} < 80)\}$
- Fator de aceleração para pacientes com transplante alogênico: $\exp\{3,325098.1(\text{indicador do escore} < 80)\}$

Assumindo o escore de Karnofsky ≥ 80 , temos que:

- Fator de aceleração para pacientes com transplante autólogo: $\exp\{0,550087\} = 1,733404$
- Fator de aceleração para pacientes com transplante alogênico: $\exp\{0\} = 1$

Como o valor é superior a 1, a ocorrência de eventos é acelerada, o que diminui o tempo de sobrevivência daqueles pacientes que fizeram o transplante autólogo.

A seguir, temos os mesmos cálculos para a variável (escore de Karnofsky):

Primeiro, calculemos a estimativa pontual para a razão de taxas de entre os pacientes com escore de Karnofsky < 80 e escore de Karnofsky ≥ 80 :

$$\frac{\alpha_T(t|\mathbf{X}_{\text{escore}<80})}{\alpha_T(t|\mathbf{X}_{\text{escore}\geq 80})} = \exp[(\mathbf{X}_{\text{escore}<80}^T - \mathbf{X}_{\text{escore}\geq 80}^T) \cdot \boldsymbol{\beta}] = K(\mathbf{X}_1, \mathbf{X}_2)$$

Vamos considerar que $\mathbf{X}_{\text{escore}<80}^T$ e $\mathbf{X}_{\text{escore}\geq 80}^T$ são vetores que indicam que os pacientes que possuem escore < 80 e escore ≥ 80 , respectivamente. Além disso, vamos assumir que ambos os pacientes fizeram o mesmo tipo de transplante:

$$\mathbf{X}_{\text{escore}<80} = \begin{pmatrix} 1 \\ \mathbb{1}(\text{indicador de Autólogo}) \\ \mathbb{1}(\text{indicador do escore} < 80) \end{pmatrix} = \begin{pmatrix} 1 \\ k \\ 1 \end{pmatrix} \text{ e } \mathbf{X}_{\text{escore}\geq 80} = \begin{pmatrix} 1 \\ \mathbb{1}(\text{indicador de Autólogo}) \\ \mathbb{1}(\text{indicador do escore} < 80) \end{pmatrix} = \begin{pmatrix} 1 \\ k \\ 0 \end{pmatrix}$$

Sendo assim,

$$\exp[(\mathbf{X}_{\text{escore}<80}^T - \mathbf{X}_{\text{escore}\geq 80}^T) \cdot \boldsymbol{\beta}] = \exp[(0 \ 0 \ 1) \cdot (0,001460 \ 0,422825 \ 2,555837)^T] = \exp(2,555837) = 12.88208$$

Então, a taxa de falha dos pacientes que obtiveram um escore de Karnofsky < 80 é 1188,2% maior que os pacientes que receberam obtiveram escore de Karnofsky ≥ 80 .

Agora, vamos encontrar uma estimativa do fator de aceleração análoga à anterior:

$$S_i(t) = P(T > t | \mathbf{x}_i) = P(T_i > t) = P(\exp\{\mu_i + \gamma_1 x_{i1} + \gamma_2 x_{i2}\} > t) = P(\mu_i > \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\} \cdot t)$$

Vamos reescrever como a função de sobrevivência:

$$S_i(t) = P(\mu_i > \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\} \cdot t) = S_0(\exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\} \cdot t)$$

Portanto,

$$\psi_{x_i} = \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\}$$

A estimativa de ψ é dada por:

$$\hat{\psi}_{x_i} = \exp\{-\hat{\gamma}_1 x_{i1} - \hat{\gamma}_2 x_{i2}\} = \exp\{-(-0,550087)x_{i1} - (-3,325098)x_{i2}\} = \exp\{0,550087x_{i1} + 3,325098x_{i2}\}$$

Como x_{i1} é $\mathbb{1}$ (indicador de Autólogo) e x_{i2} é $\mathbb{1}$ (indicador do escore < 80), temos que:

- Fator de aceleração para pacientes com escore < 80 : $\exp\{0,550087 \cdot \mathbb{1}(\text{indicador de Autólogo}) + 3,325098\}$
- Fator de aceleração para pacientes com escore ≥ 80 : $\exp\{0,550087 \cdot \mathbb{1}(\text{indicador de Autólogo})\}$

Assumindo que o tratamento foi alogênico, temos que:

- Fator de aceleração para pacientes com escore < 80 : $\exp\{3,325098\} = 27,80172$
- Fator de aceleração para pacientes com escore ≥ 80 : $\exp\{0\}$

Como o valor é superior a 1, a ocorrência de eventos é acelerada, o que diminui (e muito) o tempo de sobrevivência daqueles pacientes que obtiveram um escore < 80 .

d)

Usaremos a estatística de Wald, com um nível de significância de 10%, para testar a hipótese de igualdade dos tipos de transplante e a igualdade das categorias do escore de Karnofsky. O R já solta os valores-p dessa estatística para cada um dos parâmetros:

```
# CONSULTANDO OS RESULTADOS DO AJUSTE DO MODELO (EM R)
summary(modelo_weibull)

##
## Call:
## survreg(formula = Surv(Tempo, D_R) ~ Transplante + Escore_80,
##   data = dados, dist = "weibull")
##               Value Std. Error      z      p
## (Intercept)    7.944     0.546 14.54 < 2e-16
## TransplanteAutólogo -0.550     0.525 -1.05  0.29
## Escore_80Menor que 80 -3.325     0.563 -5.91 3.5e-09
## Log(scale)      0.263     0.171  1.54  0.12
##
## Scale= 1.3
##
```

```
## Weibull distribution
## Loglik(model)= -168.4   Loglik(intercept only)= -183.3
##  Chisq= 29.72 on 2 degrees of freedom, p= 3.5e-07
## Number of Newton-Raphson Iterations: 5
## n= 43
```

As colunas p de cada um dos parâmetros indicam o valor-p de cada teste de hipótese do tipo:

$$\begin{cases} H_0 : \gamma_j = 0 \\ H_1 : \gamma_j \neq 0 \end{cases}$$

Nesse teste, sob H_0 ,

$$\frac{\hat{\gamma}_j^2}{EP^2(\hat{\gamma}_j)} \approx \chi_{(k-1)}^2,$$

em que EP é o erro-padrão associado a estimativa $\hat{\gamma}$ e k é a quantidade de níveis que a variável pode assumir.

No caso de testarmos γ_1 , estamos testando se os tratamentos são iguais (já que o parâmetro é igual a zero para ambos). Assim,

$$\frac{\hat{\gamma}_j^2}{EP^2(\hat{\gamma}_j)} = \frac{(-0,55)^2}{(0,525)^2} = 0,705186 \Rightarrow \text{valor-p: } 1 - \frac{\hat{\gamma}_j^2}{EP^2(\hat{\gamma}_j)} = 1 - 0,705186 = 0,294814$$

Portanto, há um nível de significância de 10%, não rejeitamos a hipótese $H_0 : \gamma_1 = 0$, ou seja, os tratamentos não estatisticamente diferentes em questão de sobrevida.

Analogamente, temos que o valor-p do teste $H_0 : \gamma_2 = 0$ é igual a $3,5 \times 10^{-9}$, ou seja, a 10% de significância, a sobrevivência dos pacientes com escore de Karnofsky < 80 e ≥ 80 são estatisticamente diferentes (a sobrevivência do grupo com escore ≥ 80 é maior que a do outro grupo).

Exercício 4

Nesse exercício, ainda estamos considerando os dados do **Exercício 3**.

a)

Vamos refazer o ajuste do item **b)** utilizando agora a distribuição log-logística:

```
# AJUSTANDO O MODELO LOG-LOGÍSTICO (EM R)
modelo_log_logistico <- survreg(Surv(Tempo, D_R) ~ Transplante + Escore_80,
                                data = dados, dist = "loglogistic")
modelo_log_logistico

## Call:
## survreg(formula = Surv(Tempo, D_R) ~ Transplante + Escore_80,
##       data = dados, dist = "loglogistic")
##
## Coefficients:
##      (Intercept)  TransplanteAutólogo Escore_80Menor que 80
##      7.1013155      -0.3546855      -2.8753331
##
## Scale= 1.096619
##
## Loglik(model)= -170.9   Loglik(intercept only)= -180.2
##  Chisq= 18.56 on 2 degrees of freedom, p= 9.32e-05
## n= 43
```

O modelo de localização-escala é da forma:

$$Z_i = \ln(T_i) = \mathbf{X}_i^T \boldsymbol{\gamma} + \sigma \omega_i$$

No caso, temos que:

$$\mathbf{X}_1 = \begin{pmatrix} 1 \\ 1(\text{indicador de Autólogo}) \\ 1(\text{indicador do escore} < 80) \end{pmatrix} \text{ e } \hat{\boldsymbol{\gamma}} = \begin{pmatrix} \text{Intercepto} \\ \text{Coef. Transplante} \\ \text{Coef. Escore 80} \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{pmatrix} = \begin{pmatrix} 7,101315 \\ -0,354686 \\ -2,875333 \end{pmatrix}$$

$$\sigma = 1,096619 \text{ e } \omega_i \sim \text{Logística}$$

b)

Para encontrarmos uma estimativa pontual para o fator de aceleração, temos um algebrismo análogo ao do **Exercício 3**:

$$T_i = \exp\{\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \sigma \omega_i\}$$

Vamos definir como $\mu_i = \gamma_0 + \sigma \omega_i$, então:

$$T_i = \exp\{\gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \sigma \omega_i\} = \exp\{\mu_i + \gamma_1 x_{i1} + \gamma_2 x_{i2}\}$$

Para esse indivíduo i , a função de sobrevivência é dada por:

$$S_i(t) = P(T > t | \mathbf{x}_i) = P(T_i > t) = P(\exp\{\mu_i + \gamma_1 x_{i1} + \gamma_2 x_{i2}\} > t) = P(\mu_i > \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\}.t)$$

Vamos reescrever como a função de sobrevivência:

$$S_i(t) = P(\mu_i > \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\}.t) = S_0(\exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\}.t)$$

Portanto,

$$\psi_{x_i} = \exp\{-\gamma_1 x_{i1} - \gamma_2 x_{i2}\}$$

A estimativa de ψ é dada por:

$$\hat{\psi}_{x_i} = \exp\{-\hat{\gamma}_1 x_{i1} - \hat{\gamma}_2 x_{i2}\} = \exp\{-(-0,354686)x_{i1} - (-2,875333)x_{i2}\} = \exp\{0,354686x_{i1} + 2,875333x_{i2}\}$$

Como x_{i1} é $\mathbb{1}$ (indicador de Autólogo) e x_{i2} é $\mathbb{1}$ (indicador do escore < 80), temos que:

- Fator de aceleração para pacientes com transplante autólogo: $\exp\{0,354686 + 2,875333.\mathbb{1}(\text{indicador do escore} < 80)\}$
- Fator de aceleração para pacientes com transplante alogênico: $\exp\{2,875333.\mathbb{1}(\text{indicador do escore} < 80)\}$

Assumindo o escore de Karnofsky ≥ 80 , temos que:

- Fator de aceleração para pacientes com transplante autólogo: $\exp\{0,354686\} = 1,425733$
- Fator de aceleração para pacientes com transplante alogênico: $\exp\{0\} = 1$

Como o valor é superior a 1, a ocorrência de eventos é acelerada, o que diminui o tempo de sobrevivência daqueles pacientes que fizeram o transplante autólogo.

A seguir, temos os mesmos cálculos para a variável (escore de Karnofsky):

- Fator de aceleração para pacientes com escore < 80 : $\exp\{0,354686.1(\text{indicador de Autólogo}) + 2,875333\}$
- Fator de aceleração para pacientes com escore ≥ 80 : $\exp\{2,875333.1(\text{indicador de Autólogo})\}$

Assumindo que o tratamento foi alogênico, temos que:

- Fator de aceleração para pacientes com escore < 80 : $\exp\{2,875333\} = 17,73133$
- Fator de aceleração para pacientes com escore ≥ 80 : $\exp\{0\}$

Como o valor é superior a 1, a ocorrência de eventos é acelerada, o que diminui (e muito) o tempo de sobrevivência daqueles pacientes que obtiveram um escore < 80 .

c)

A chance de sobrevivência após t é definida como

$$\frac{S(t|\mathbf{x})}{1 - S(t|\mathbf{x})}$$

Vamos mostrar que no modelo logístico

$$\frac{S(t|\mathbf{x})}{1 - S(t|\mathbf{x})} = \exp[-\mathbf{x}^T \beta] \frac{S(t|\mathbf{x} = \mathbf{0})}{1 - S(t|\mathbf{x} = \mathbf{0})}$$

Seja $t = -\mathbf{x}^T \beta + \omega$, com $\omega \sim \text{Logística Padrão}$.

$$\omega = t + \mathbf{x}^T \beta$$

Por definição, temos que:

$$f_{\omega}(\omega) = \frac{e^{\omega}}{(1 + e^{\omega})^2} \text{ e } S(\omega) = \frac{1}{1 + e^{\omega}}$$

$$S(t|x) = \frac{1}{1 + \exp(t + \mathbf{x}^T \beta)} \Leftrightarrow S(t|\mathbf{x} = \mathbf{0}) = \frac{1}{1 + \exp(t)}$$

$$\frac{S(t|\mathbf{x} = \mathbf{0})}{1 - S(t|\mathbf{x} = \mathbf{0})} = \frac{\frac{1}{1 + \exp(t)}}{1 - \frac{1}{1 + \exp(t)}} = \frac{\frac{1}{1 + \exp(t)}}{\frac{\exp(t)}{1 + \exp(t)}} = \frac{1}{1 + \exp(t)} = \exp(-t)$$

$$\frac{S(t|\mathbf{x})}{1 - S(t|\mathbf{x})} = \frac{\frac{1}{1 + \exp(t + \mathbf{x}^T \beta)}}{1 - \frac{1}{1 + \exp(t + \mathbf{x}^T \beta)}} = \frac{\frac{1}{1 + \exp(t + \mathbf{x}^T \beta)}}{\frac{\exp(t + \mathbf{x}^T \beta)}{1 + \exp(t + \mathbf{x}^T \beta)}} = \frac{1}{1 + \exp(t + \mathbf{x}^T \beta)} = \exp(-t) \cdot \exp(\mathbf{x}^T \beta)$$

Sendo assim, temos que:

$$\frac{S(t|\mathbf{x})}{1 - S(t|\mathbf{x})} = \exp(-\mathbf{x}^T \beta) \cdot \exp(-t) = \exp(-\mathbf{x}^T \beta) \cdot \frac{S(t|\mathbf{x} = \mathbf{0})}{1 - S(t|\mathbf{x} = \mathbf{0})} \blacksquare$$

d)

Vamos obter uma estimativa da razão de chances de sobrevivência após um tempo t dos pacientes que realizaram os transplantes autólogo e alogênico, respectivamente.

Para tal, vamos definir \mathbf{x}_{Aut} e \mathbf{x}_{Alo} como os dados dos pacientes que realizaram os transplantes autólogo e alogênico, respectivamente. Além disso, vamos assumir que o escore de Karnofsky é igual a k para ambos. Então a razão de chances de sobrevivência é calculada por:

$$\frac{\frac{S(t|\mathbf{x}_{Aut})}{1 - S(t|\mathbf{x}_{Aut})}}{\frac{S(t|\mathbf{x}_{Alo})}{1 - S(t|\mathbf{x}_{Alo})}} = \frac{\exp(-\mathbf{x}_{Aut}^T \beta) \cdot \frac{S(t|\mathbf{x} = \mathbf{0})}{1 - S(t|\mathbf{x} = \mathbf{0})}}{\exp(-\mathbf{x}_{Alo}^T \beta) \cdot \frac{S(t|\mathbf{x} = \mathbf{0})}{1 - S(t|\mathbf{x} = \mathbf{0})}} = \frac{\exp(-\mathbf{x}_{Aut}^T \beta)}{\exp(-\mathbf{x}_{Alo}^T \beta)} = \exp[(\mathbf{x}_{Alo}^T - \mathbf{x}_{Aut}^T) \cdot \beta]$$

Sendo que,

$$\mathbf{x}_{aut} = \begin{pmatrix} 1 \\ \mathbb{1}(\text{indicador de Autólogo}) \\ \mathbb{1}(\text{indicador do escore} < 80) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ k \end{pmatrix} \text{ e } \mathbf{x}_{alo} = \begin{pmatrix} 1 \\ \mathbb{1}(\text{indicador de Autólogo}) \\ \mathbb{1}(\text{indicador do escore} < 80) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ k \end{pmatrix}$$

Portanto,

$$\exp[(\mathbf{x}_{Alo}^T - \mathbf{x}_{Aut}^T) \cdot \beta] = \exp[(0 \ 1 \ 0) \cdot \gamma / \sigma] = \exp(-0,354686/1,096619) = 0,723658$$

Então, a chance de um indivíduo com células normais sobreviver é maior so que um indivíduo que recebeu as células anormais.

e)

Analogamente a letra **d)** do **Exercício 3**, temos que os valores-p da estatística de Wald são disponibilizados pelo seguinte comando:

```
# CONSULTANDO OS RESULTADOS DO MODELO (EM R)
summary(modelo_log_logistico)
```

```
##
## Call:
## survreg(formula = Surv(Tempo, D_R) ~ Transplante + Escore_80,
## data = dados, dist = "loglogistic")
##
##              Value Std. Error      z      p
## (Intercept)    7.1013    0.6637 10.70 < 2e-16
## TransplanteAutólogo -0.3547    0.6519 -0.54  0.59
## Escore_80Menor que 80 -2.8753    0.6480 -4.44 9.1e-06
```

```
## Log(scale)          0.0922    0.1688  0.55    0.58
##
## Scale= 1.1
##
## Log logistic distribution
## Loglik(model)= -170.9    Loglik(intercept only)= -180.2
##  Chisq= 18.56 on 2 degrees of freedom, p= 9.3e-05
## Number of Newton-Raphson Iterations: 4
## n= 43
```

Vale lembrar que o cálculo da estatística de Wald é análogo ao que já calculamos:

$$W = \frac{\hat{\gamma}^2}{[DP(\hat{\gamma})]^2}$$

Logo, há um nível de significância de 10%, não rejeitamos a hipótese $H_0 : \gamma_1 = 0$ (igualdade de tratamentos) e rejeitamos a hipótese $H_0 : \gamma_2 = 0$.

Esses resultados vão ao encontro dos resultados vistos com o ajuste pela distribuição Weibull.

Exercício 5

Nesse exercício, ainda estamos considerando os dados do **Exercício 3**.

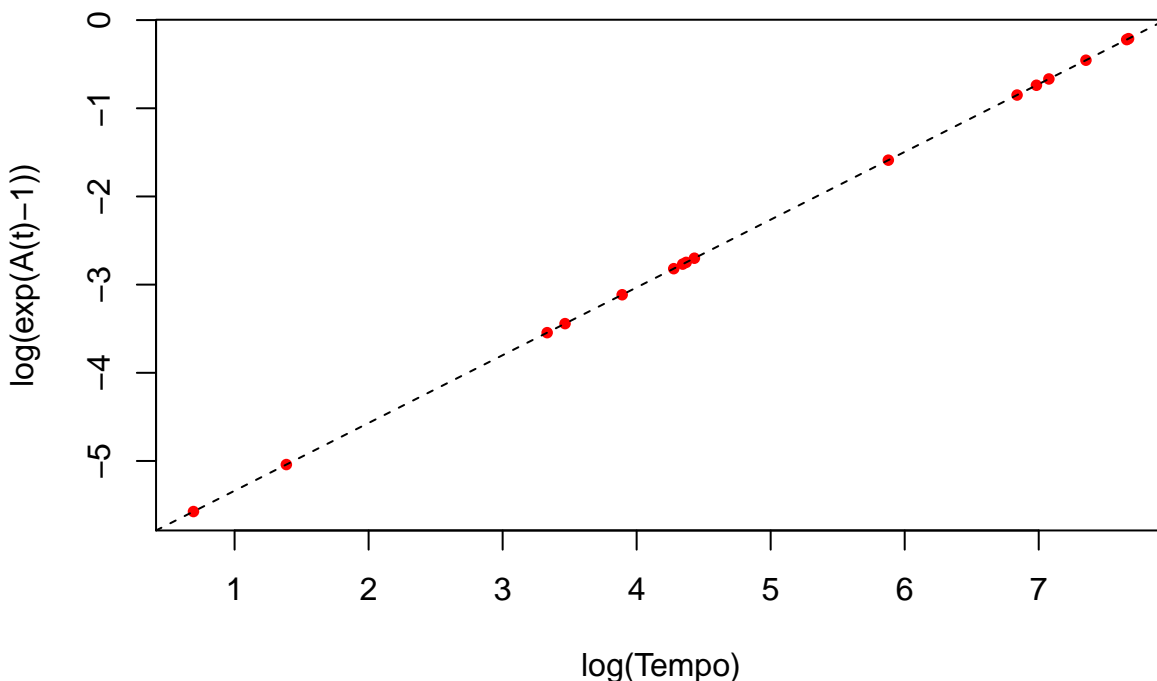
Vamos fazer gráficos da taxa de falha acumulada para verificar a adequabilidade dos modelos Weibull e Log-logístico. Para isso, vamos utilizar o estimador de Nelson-Aalen da função de taxa de falha acumulada considerando cada grupo separadamente (ou seja, com estimativas da função de taxa de falha acumulada para cada grupo).

Primeiro, façamos os gráficos para o **Modelo Weibull**:

```
# CALCULANDO OS ESTIMADORES DE NELSON AALEN POR GRUPO (EM R)
gamma = modelo_weibull$coefficients
rho = 1/modelo_weibull$scale
lambda = exp(-gamma[1]*rho)

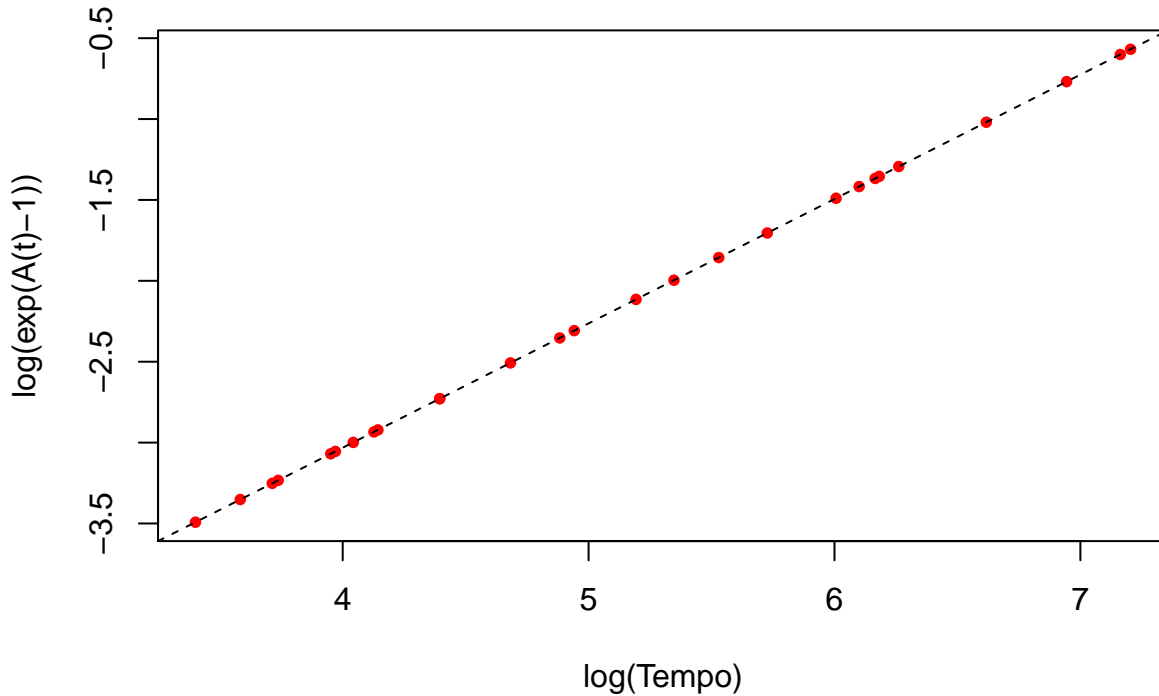
A = log(1+lambda*dados$Tempo[1:16]^rho)
la = log(exp(A)-1)
plot(log(dados$Tempo[1:16]), la, col = "red", pch = 16,
     main = "Grupo 1: Transplante tipo Alogênico - Weibull",
     xlab = "log(Tempo)", ylab="log(exp(A(t)-1))", cex = 0.8)
abline(log(lambda), rho, lty = 2)
```

Grupo 1: Transplante tipo Alogênico – Weibull



```
A = log(1+lambda*dados$Tempo[17:43]^rho)
la = log(exp(A)-1)
plot(log(dados$Tempo[17:43]), la, col = "red", pch = 16,
     main = "Grupo 1: Transplante tipo Autólogo - Weibull",
     xlab = "log(Tempo)", ylab="log(exp(A(t)-1))", cex = 0.8)
abline(log(lambda), rho, lty = 2)
```

Grupo 1: Transplante tipo Autólogo – Weibull



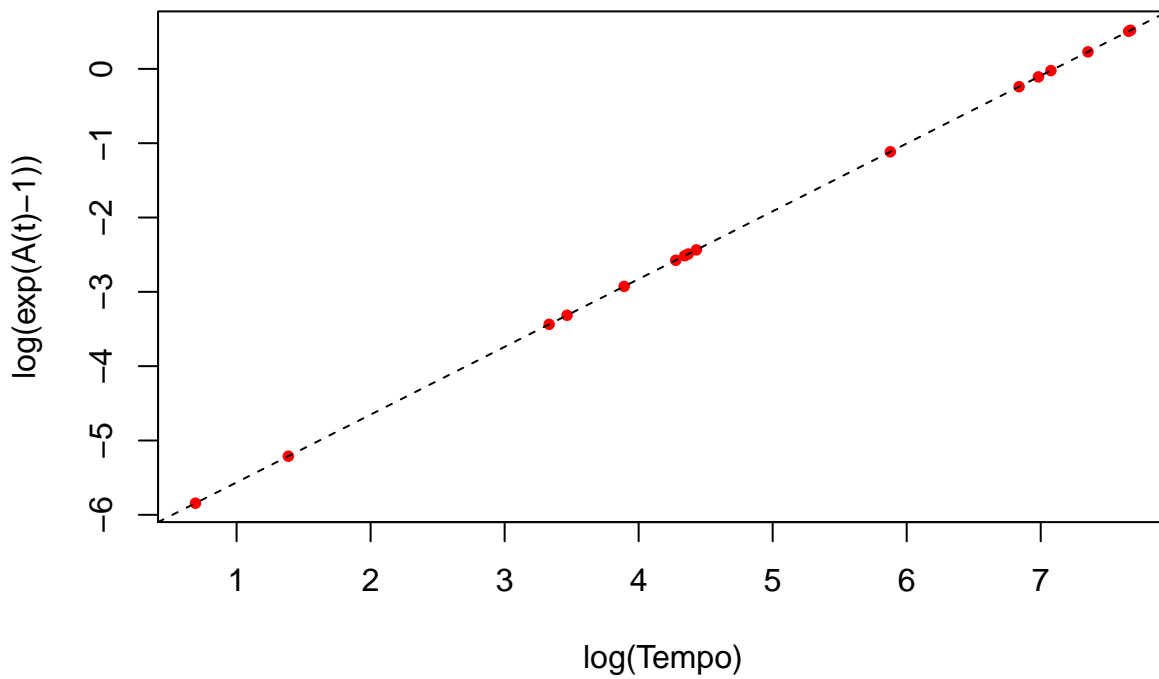
Os gráficos mostram uma boa adequabilidade do modelo Weibull para esses dois grupos.

Agora, façamos os gráficos para o **Modelo Log-logístico**:

```
# CALCULANDO OS ESTIMADORES DE NELSON AALEN POR GRUPO (EM R)
gamma = modelo_log_logistico$coefficients
rho = 1/modelo_log_logistico$scale
lambda = exp(-gamma[1]*rho)

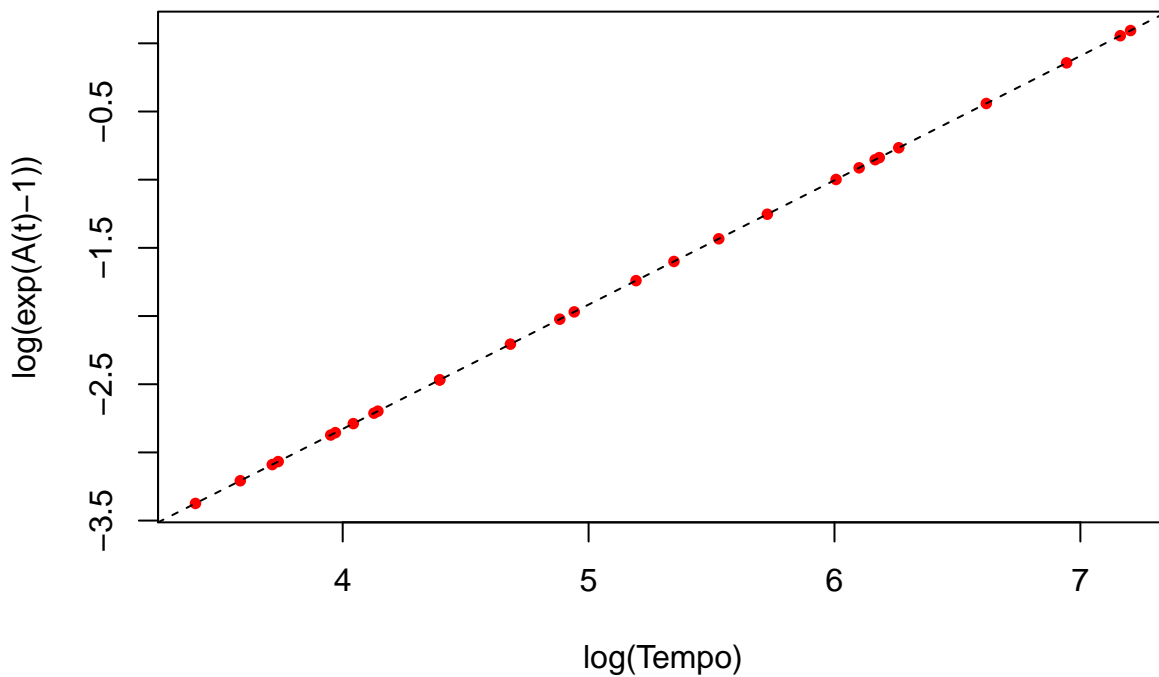
A = log(1+lambda*dados$Tempo[1:16]^rho)
la = log(exp(A)-1)
plot(log(dados$Tempo[1:16]), la, col = "red", pch = 16,
     main = "Grupo 1: Transplante tipo Alogênico - Log-logístico",
     xlab = "log(Tempo)", ylab="log(exp(A(t))-1)", cex = 0.8)
abline(log(lambda), rho, lty = 2)
```

Grupo 1: Transplante tipo Alogênico – Log-logístico



```
A = log(1+lambda*dados$Tempo[17:43]^rho)
la = log(exp(A)-1)
plot(log(dados$Tempo[17:43]), la, col = "red", pch = 16,
     main = "Grupo 1: Transplante tipo Autólogo - Log-logístico",
     xlab = "log(Tempo)", ylab="log(exp(A(t))-1)", cex = 0.8)
abline(log(lambda), rho, lty = 2)
```

Grupo 1: Transplante tipo Autólogo – Log-logístico



Os gráficos também mostram uma boa adequabilidade do modelo Log-logístico para esses dois grupos.

Exercício 6

Nesse exercício, ainda estamos considerando os dados do **Exercício 3**.

Vamos calcular os resíduos de Cox-Snell para o modelo Weibull:

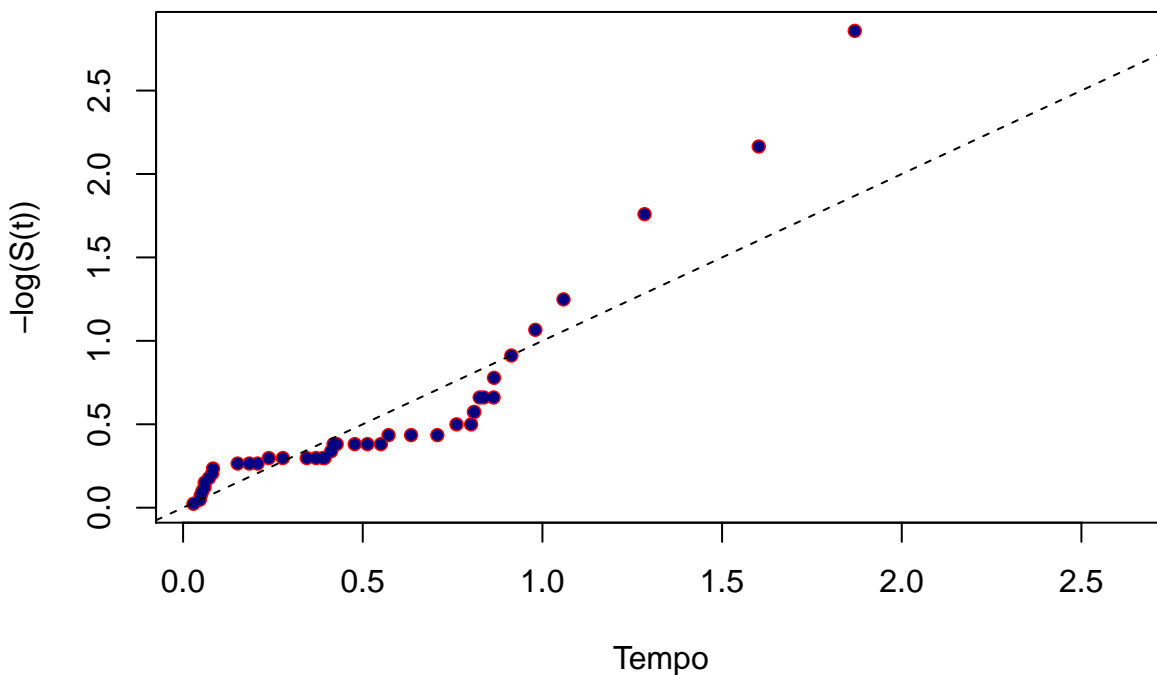
```
# CALCULANDO OS RESÍDUOS DE COX-SNELL PARA WEIBULL (EM R)
dados %<>% mutate(ind_aut = ifelse(Transplante == "Autólogo", 1, 0),
                  ind_menor_80 = ifelse(Escore_80 == "Menor que 80", 1, 0))

xb_wei <- modelo_weibull$coef[1]+modelo_weibull$coef[2]*dados$ind_aut+modelo_weibull$coef[3]*dados$ind_menor_80
res_wei <- (log(dados$Tempo)-xb_wei)/modelo_weibull$scale
resid_wei <-exp(res_wei)
coxsnell_wei <- (dados$Tempo^(1/modelo_weibull$scale))*exp(-xb_wei/modelo_weibull$scale)

KM_wei <- survfit(Surv(coxsnell_wei, dados$D_R) ~ 1)
TFacum_KM_wei <- -log(KM_wei$surv)

Surv_Aa_wei <- survfit(coxph(Surv(coxsnell_wei, dados$D_R) ~ 1))
TFacum_Aa_wei <- -log(Surv_Aa_wei$surv)
plot(KM_wei$time, TFacum_KM_wei, col = "red", pch = 16,
     main = "Gráfico da função de risco acumulada (cox-snell) - Weibull",
     ylab = "-log(S(t))", xlab = "Tempo")
points(Surv_Aa_wei$time, TFacum_Aa_wei, col = "navy blue", pch = 16, cex = 0.8)
abline(0, 1, lty = 2)
```

Gráfico da função de risco acumulada (cox-snell) – Weibull

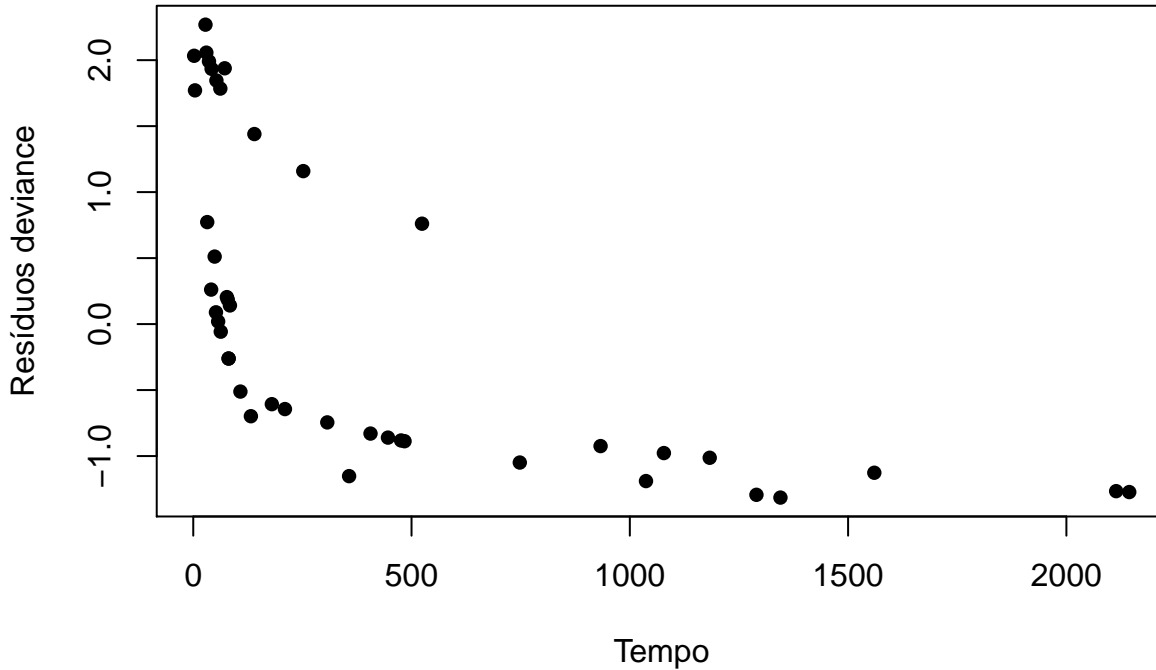


Os pontos apresentam aumentam seu valor ao longo do tempo, o que pode ser considerada uma falta de qualidade do ajuste.

Agora, façamos os resíduos *deviance* para o modelo Weibull:

```
# CALCULANDO OS RESÍDUOS DEVIANCE PARA WEIBULL (EM R)
m_wei <- dados$D_R - coxsnell_wei
deviance_wei <- sqrt(-2*(m_wei + dados$D_R*log(dados$D_R - m_wei)))*ifelse(m_wei < 0, -1, 1)
plot(dados$Tempo, deviance_wei, pch = 16, main = "Resíduos deviance - Weibull",
     ylab = "Resíduos deviance", xlab = "Tempo")
```

Resíduos deviance – Weibull



O gráfico apresenta alguns pontos que distam muito dos demais, contudo, a maioria está entre uma faixa parecida de valores e bem pequena (entre 2 e -1).

Vamos calcular os resíduos de Cox-Snell para o modelo Log-logístico:

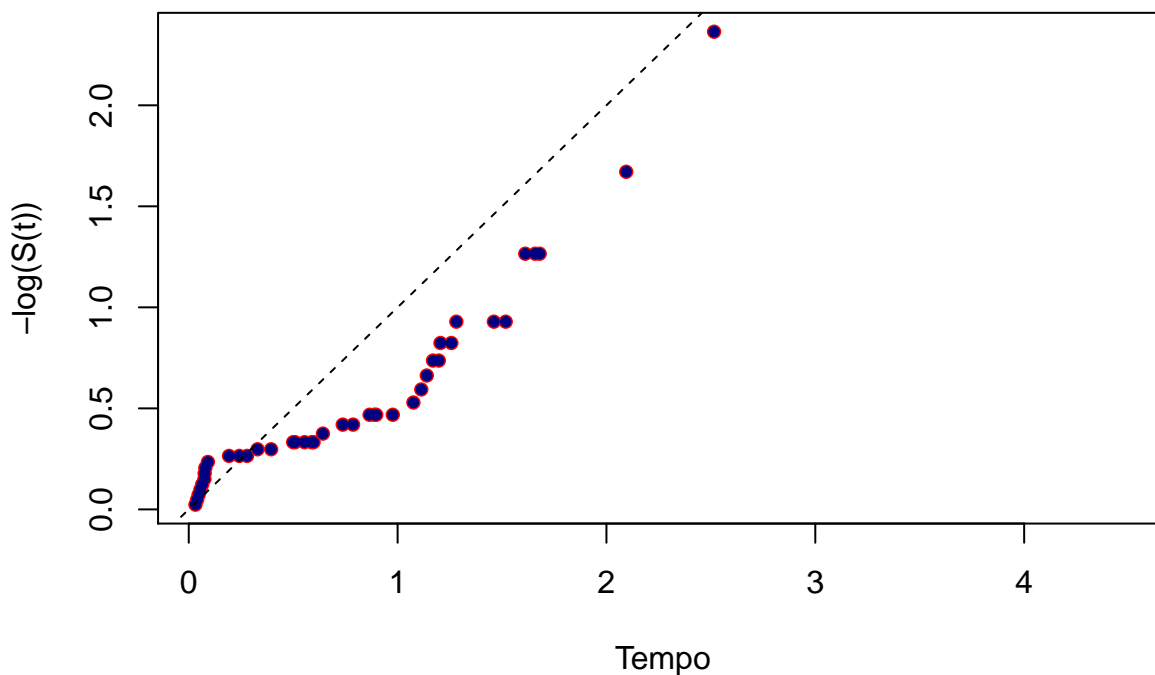
```
# CALCULANDO OS RESÍDUOS DE COX-SNELL PARA LOG-LOGÍSTICO (EM R)
xb_ll <- modelo_log_logistico$coef[1]+modelo_log_logistico$coef[2]*dados$ind_aut+modelo_log_logistico$coef[3]*dados$Tempo
res_ll <- (log(dados$Tempo)-xb_ll)/modelo_log_logistico$scale
resid_ll <- exp(res_ll)
coxsnell_ll <- (dados$Tempo^(1/modelo_log_logistico$scale))*exp(-xb_ll/modelo_log_logistico$scale)
coxsnell_ll
```

```
## [1] 0.03216009 0.49994154 0.73732952 1.20537078 4.50966885 0.78682749
## [7] 0.89761297 0.97700910 1.25734194 1.65884383 1.68029721 0.03989235
## [13] 0.07505801 0.07609442 1.11342637 1.13976859 0.06432207 0.07952168
## [19] 1.16955309 1.28131537 1.61132894 0.19282793 1.61132894 0.32957279
## [25] 0.64249602 0.27909141 0.58860274 1.19728920 0.04732676 0.05588711
## [31] 0.86603638 1.07562615 0.09174870 2.09466739 2.51528256 0.24249238
## [37] 0.39458008 0.50912898 0.55467824 0.59761700 0.88883707 1.46102144
## [43] 1.51771970
```

```
KM_11 <- survfit(Surv(coxsnell_11, dados$D_R) ~ 1)
TFacum_KM_11 <- -log(KM_11$surv)

Surv_Aa_11 <- survfit(coxph(Surv(coxsnell_11, dados$D_R) ~ 1))
TFacum_Aa_11 <- -log(Surv_Aa_11$surv)
plot(KM_11$time, TFacum_KM_11, col = "red", pch = 16,
     main = "Gráfico da função de risco acumulada (cox-snell) - Log-logístico",
     ylab = "-log(S(t))", xlab = "Tempo")
points(Surv_Aa_11$time, TFacum_KM_11, col = "navy blue", pch = 16, cex = 0.8)
abline(0, 1, lty = 2)
```

Gráfico da função de risco acumulada (cox-snell) – Log-logístico

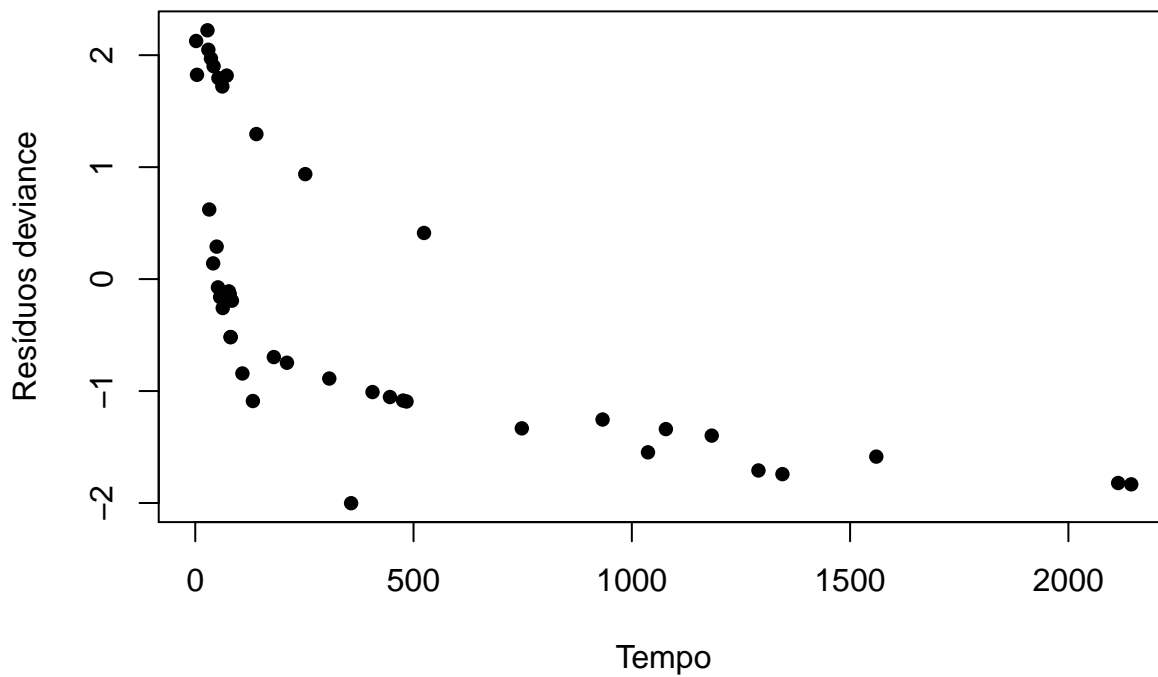


Os pontos se deslocam da reta logo no começo do gráfico, o que pode ser prejudicial ao ajuste.

Agora, façamos os resíduos *deviance* para o modelo Log-logístico:

```
# CALCULANDO OS RESÍDUOS DEVIANCE PARA LOG-LOGÍSTICO (EM R)
m_11 <- dados$D_R - coxsnell_11
deviance_11 <- sqrt(-2*(m_11 + dados$D_R*log(dados$D_R - m_11)))*ifelse(m_11 < 0, -1, 1)
plot(dados$Tempo, deviance_11, pch = 16, main = "Resíduos deviance - Log-logístico",
     ylab = "Resíduos deviance", xlab = "Tempo")
```

Resíduos deviance – Log-logístico



A partir do tempo 500, os resíduos ficam sempre abaixo de zero. Podemos desconsiderar esse fato sabendo que a variação está entre -2 e 2.

A qualidade geral do ajuste do modelo é avaliada nos gráficos dos resíduos de Cox-Snell. Nos dois ajustes, os gráficos de quantis que comparam com a distribuição padrão dos modelos não formam uma reta de formato $x = y$ (reta de 45 graus) indicando que nenhum dos modelos está bem ajustado. Alguma revisão pode ser feita nos dois ajustes.