

# **Análise de sobrevivência com dados de diálise - Segunda Avaliação**

**Universidade Federal da Paraíba - CCEN**

Gabriel de Jesus Pereira

13 de outubro de 2024

```

library(flexsurv)
library(survminer)
library(discSurv)
library(survival)
library(tidyverse)
library(vroom)
library(ggsurvfit)
library(mice)

df <- read_delim(
  "sobrevivencia/segunda_avaliacao/includes/dialcompete.txt",
  delim = " ") |>
mutate(
  intervalo = cut(
    tempo,
    breaks = 1:44,
    labels = paste0("[", 1:43, ",", 2:44, ")"),
    right = FALSE
  ),
  doenca = factor(doenca),
) |>
select(-intervalo)

```

# Questão 1

Com seu banco de dados utilizado na primeira prova, escolhido em: <http://sobrevida.fiocruz.br/dados.html>. A partir do banco de dados escolhido por você, faça o que se pede a seguir.

## (a) Ajuste o modelo de Cox aos dados. Interprete os coeficientes do modelo escolhido. Quais critérios você considerou para obter os modelos?

O modelo foi construído considerando os diferentes grupos de doença de base dos pacientes em diálise e também a idade dos pacientes. A partir dos dados da tabela a seguir, é possível ver que há indícios de que, ao nível de 5% de significância, apenas a variável idade é significativa, mas os grupos de doença não são significativos.

```
fit_cox <- coxph(  
  Surv(tempo, status) ~ doenca + idade,  
  data = df,  
  method = "breslow")  
summary(fit_cox)
```

Call:

```
coxph(formula = Surv(tempo, status) ~ doenca + idade, data = df,  
      method = "breslow")
```

n= 2453, number of events= 686

	coef	exp(coef)	se(coef)	z	Pr(> z )
doencadiab	0.344346	1.411066	0.249773	1.379	0.168
doencahiper	-0.162741	0.849812	0.237105	-0.686	0.492
doencaoutr	-0.222810	0.800267	0.257236	-0.866	0.386
doencarim	-0.297050	0.743007	0.243107	-1.222	0.222
idade	0.011220	1.011283	0.002845	3.943	8.04e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
doencadiab	1.4111	0.7087	0.8648	2.302
doencahiper	0.8498	1.1767	0.5339	1.353
doencaoutr	0.8003	1.2496	0.4834	1.325
doencarim	0.7430	1.3459	0.4614	1.197
idade	1.0113	0.9888	1.0057	1.017

Concordance= 0.582 (se = 0.012 )

Likelihood ratio test= 53.49 on 5 df, p=3e-10

Wald test = 57.97 on 5 df, p=3e-11

Score (logrank) test = 59.61 on 5 df, p=1e-11

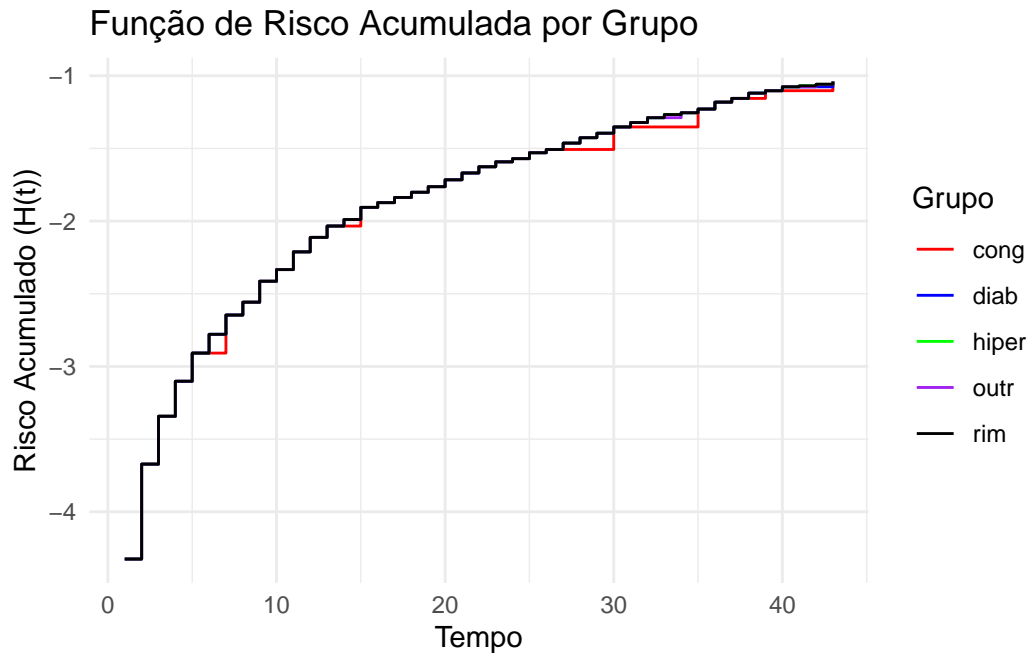
## (b) Verifique a proporcionalidade dos riscos de acordo com o método que você desejar

A proporcionalidade de risco foi verificada a partir de três métodos. O primeiro deles é foi pelo método gráfico, como pode ser visto na figura abaixo:

```
risco_acumulado <- basehaz(fit_cox, centered = FALSE) |>
  mutate(hazard = log(hazard))

#Pareando a base risco_acumulado com
#a base original pela variável chave 'time'
dados <- merge(
  df |>
    rename(time = "tempo"),
  risco_acumulado,
  by="time")

ggplot(dados, aes(x = time, y = hazard, color = doenca)) +
  geom_step() +
  scale_color_manual(values = c("red", "blue", "green", "purple", "black")) +
  labs(title = "Função de Risco Acumulada por Grupo",
       x = "Tempo",
       y = "Risco Acumulado (H(t))",
       color = "Grupo") +
  theme_minimal()
```



Pelo método gráfico, é possível ver que não há indícios sugerindo uma possível falha da suposição de riscos proporcionais. O segundo método foi a partir do teste para suposição de riscos proporcionais. A partir desse teste foi possível observar o mesmo resultado do método gráfico para a suposição de riscos proporcionais, mostrando que, ao nível de 5% de significância, a idade, a doença e o nível global não apresentam nenhum risco a suposição de riscos proporcionais.

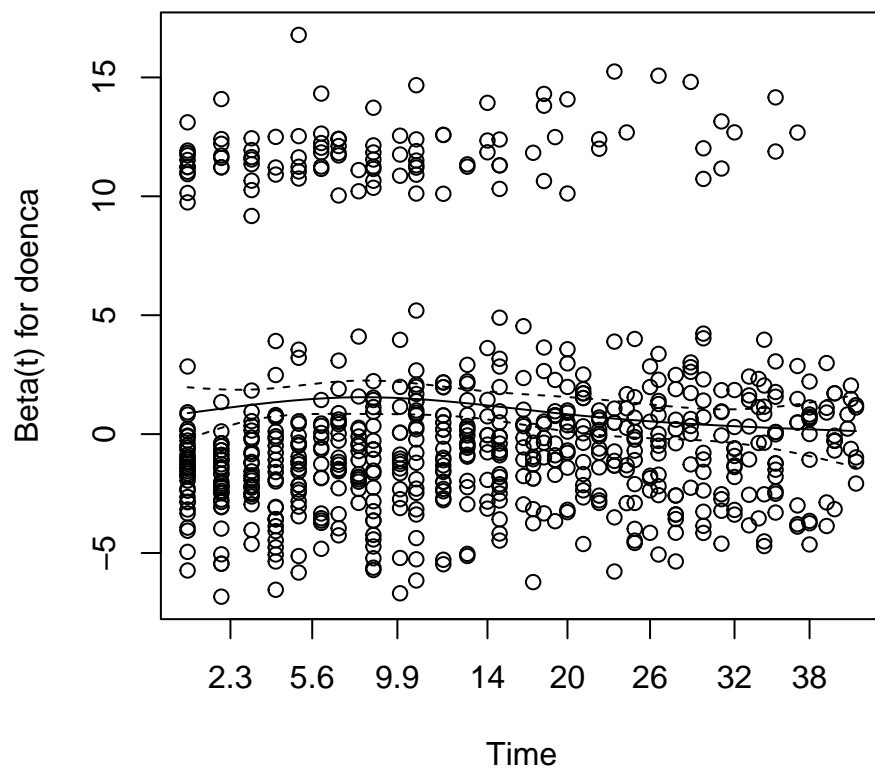
```
cox.zph(fit_cox, transform = "identity")
```

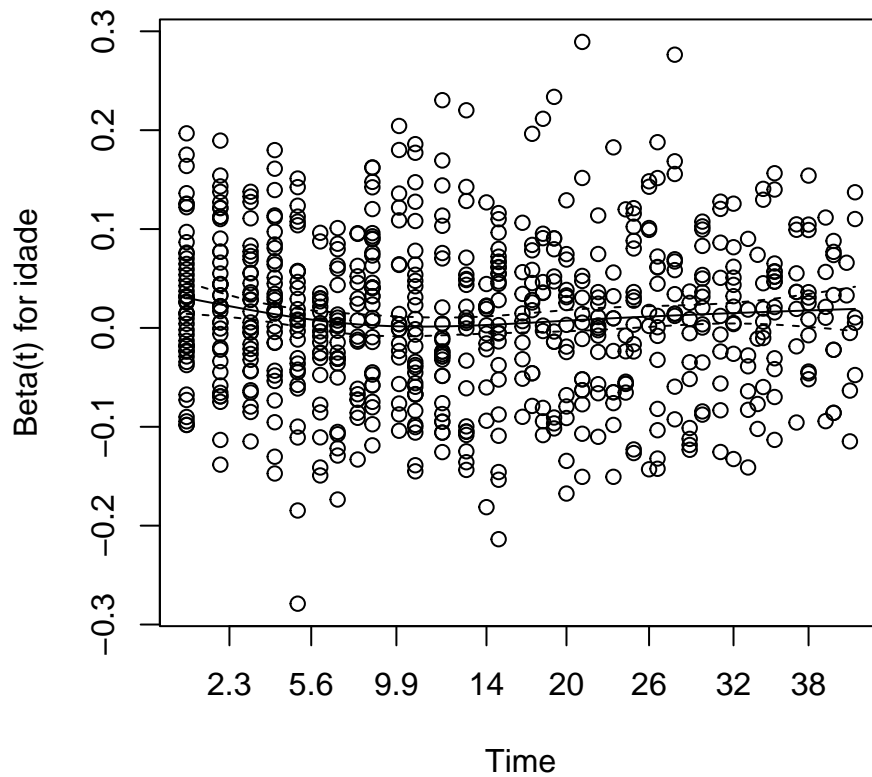
	chisq	df	p
doenca	6.806	4	0.15
idade	0.443	1	0.51
GLOBAL	6.824	5	0.23

Por último, foi utilizado também os resíduos padronizados de Schoenfeld para analisar a suposição de riscos proporcionais, como pode ser visto nos gráficos abaixo:

A partir do gráfico de resíduos padronizados de Schoenfeld para a variável idade utilizada na construção do modelo, é possível ver que não apresenta evidências que não permitem a rejeição da hipótese nula de riscos proporcionais, uma vez que tendências ao longo do tempo não são evidentes. O mesmo pode ser confirmado para a variável doença. Não há tendência evidentes ao longo do tempo.

```
plot(cox.zph(fit_cox))
```





**(c) Faça uma análise de resíduos do modelo que você escolheu. Qual ou quais resíduos você escolheu? Interprete.**

Os resíduos escolhidos foi o resíduo de martingal e de deviance. O resíduo de martingal é uma modificação dos resíduos de Cox-Snell e é visto como uma estimativa do número de falhas em excesso observado na amostra. Ele é definido da seguinte forma:

$$m_i = \delta_i - e_i$$

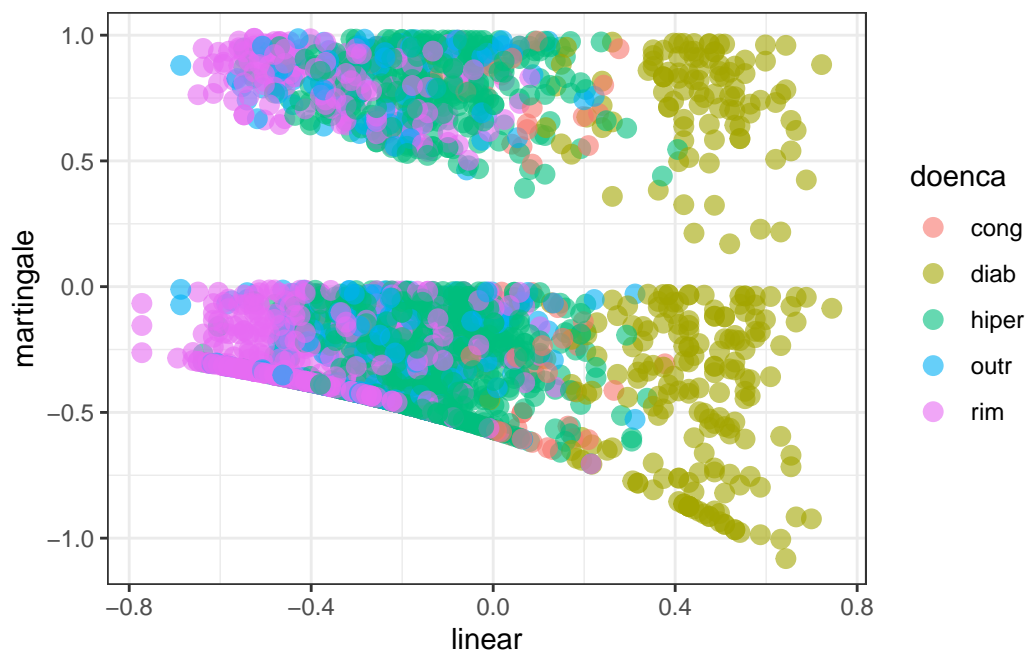
em que  $\delta_i$  é a variável indicadora de falha e  $e_i$  os resíduos de Cox-Snell.

O gráfico abaixo são os resíduos martingal versus o preditor linear do modelo de Cox ajustado. O gráfico não sugere a existência de outliers. É possível verificar o mesmo pelo gráfico dos resíduos de deviance. No entanto, o gráfico dos resíduos apresentam

problemas na adequação do modelo, pois não apresentam um comportamento aleatório dos resíduos em torno de 0.

```
mart = df |>
  mutate(
    linear = fit_cox$linear.predictors,
    martingale = resid(fit_cox, type = "martingale"),
    deviance = resid(fit_cox, type = "deviance")
  ) |>
  ggplot(aes(x = linear, y = martingale, color = doenca)) +
  geom_point(size = 3, alpha = .6) +
  theme_bw()
```

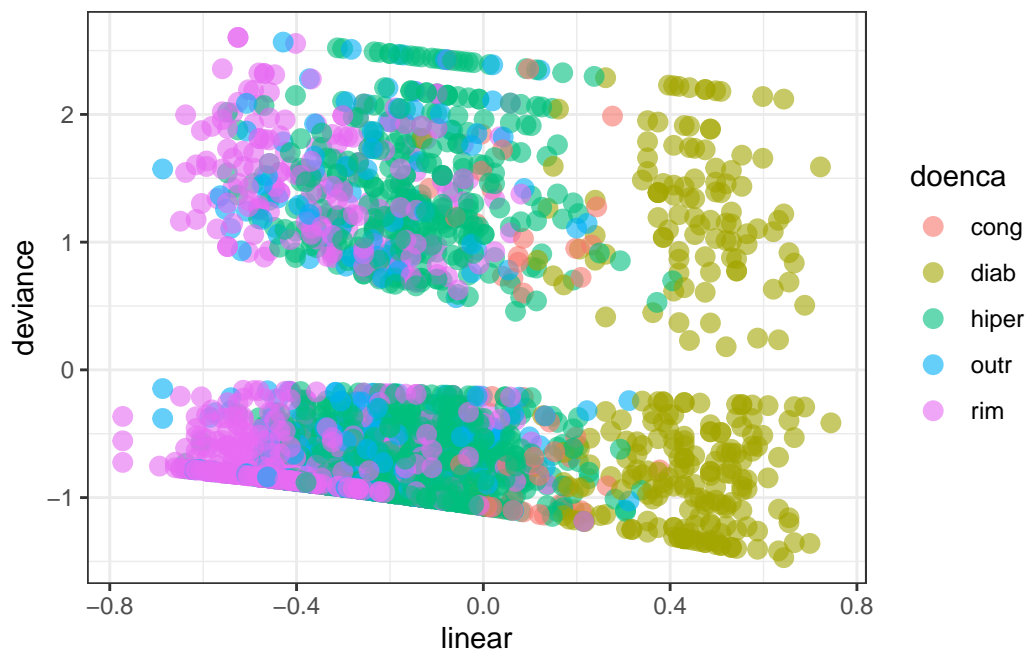
mart



```
dev = df |>
  mutate(
    linear = fit_cox$linear.predictors,
    martingale = resid(fit_cox, type = "martingale"),
    deviance = resid(fit_cox, type = "deviance")
  ) |>
  ggplot(aes(x = linear, y = deviance, color = doenca)) +
  geom_point(size = 3, alpha = .6) +
```



```
theme_bw()
dev
```



**(d)**

Ajuste seu modelo utilizando os modelos paramétricos de regressão Exponencial e Weibull.

```
ekm <- survfit(
  Surv(tempo, status) ~ doenca,
  data = df, type = "kaplan-meier")
```

```
st1<-ekm[1]$surv
time1<-ekm[1]$time
invst1<-qnorm(st1)
```

```
st2<-ekm[2]$surv
time2<-ekm[2]$time
invst2<-qnorm(st2)
```

```
st3<-ekm[3]$surv
time3<-ekm[3]$time
```

```

invst3<-qnorm(st3)

st4<-ekm[4]$surv
time4<-ekm[4]$time
invst4<-qnorm(st4)

st5<-ekm[5]$surv
time5<-ekm[5]$time
invst5<-qnorm(st5)

# st<-ekm$surv
# temp<-ekm$time
# invst<-qnorm(st)

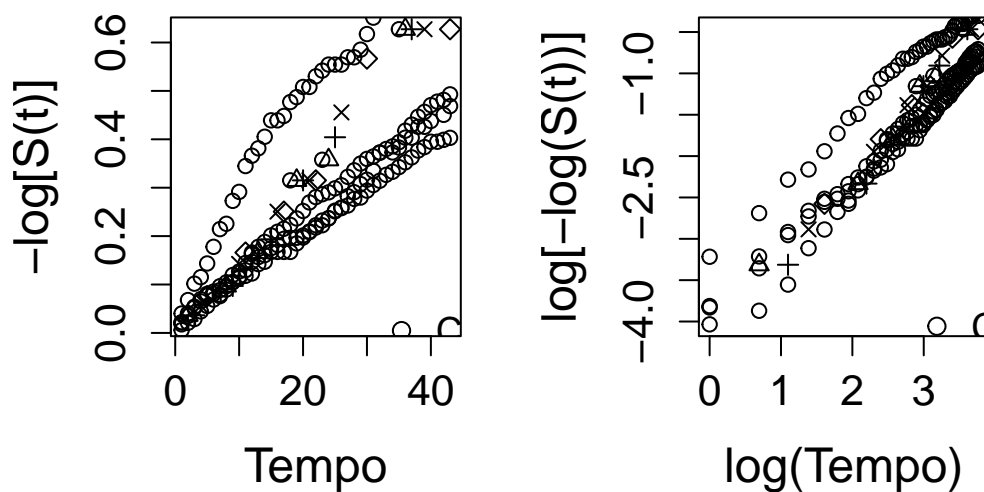
# par(mfrow=c(1, 2))
# plot(temp, -log(st), pch=16, xlab="Tempos", ylab="-log(S(t))")
# plot(log(temp), log(-log(st)), pch=16, xlab="log(Tempos)", ylab="log(-log(S(t)))")

par(mfrow=c(1,2))

#Verificação - Exponencial - Grupos 1 e 2
plot(time1, -log(st1), pch=1:5, xlab="Tempo", ylab="-log[S(t)]",
      cex.lab=1.5, cex.axis=1.3)
points(time2, -log(st2))
points(time3, -log(st3))
points(time4, -log(st4))
points(time5, -log(st5))
legend(30, 0.1, pch=1:5,
      c("Congênito", "Diabetes", "Hipertenso", "Rim", "Outros"),
      bty="n", cex=1.2)

#Verificação - Weibull - Grupos 1 e 2
plot(log(time1), log(-log(st1)), pch=1:5, xlab="log(Tempo)",
      ylab="log[-log(S(t))]", cex.lab=1.5, cex.axis=1.3)
points(log(time2), log(-log(st2)))
points(log(time3), log(-log(st3)))
points(log(time4), log(-log(st4)))
points(log(time5), log(-log(st5)))
legend(2.7, -3.5, pch=1:5,
      c("Congênito", "Diabetes", "Hipertenso", "Rim", "Outros"),
      bty="n", cex=1.2)

```



```
fit_exp <- flexsurvreg(
  Surv(tempo, status) ~ idade + doenca,
  data = df,
  dist='exponential')

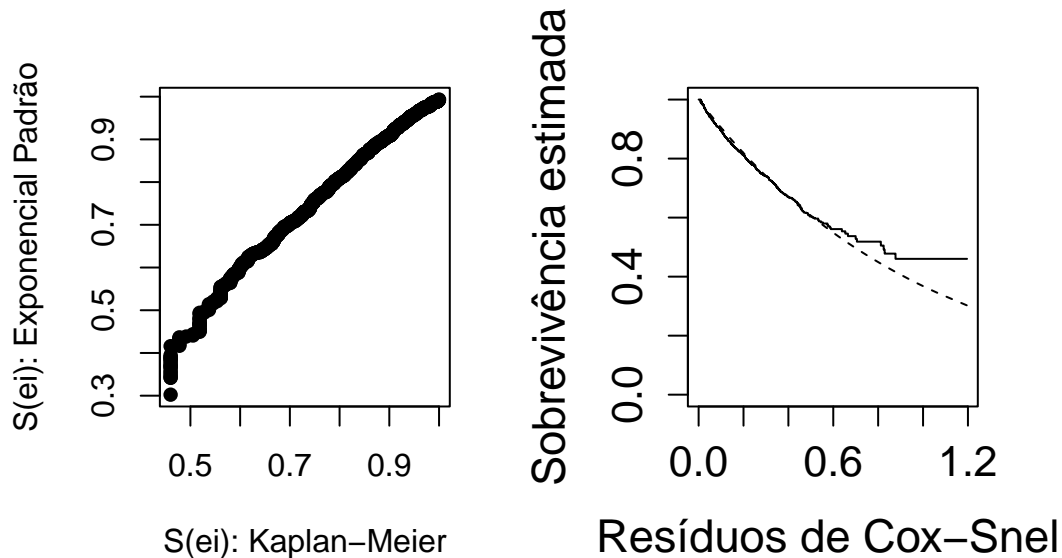
res_exp <- resid(fit_exp, type = "coxsnell")
ekm1 <- survfit(
  Surv(res_exp, status) ~ 1,
  data = df, type = "kaplan-meier")
t <- ekm1$time
st <- ekm1$surv
sexp <- exp(-t)

par(mfrow=c(1,2))

plot(st, sexp, xlab = "S(ei): Kaplan-Meier", ylab = "S(ei): Exponencial Padrão", pch =

plot(ekm1, conf.int=F, lty=c(1,1), xlab="Resíduos de Cox-Snell",
  ylab="Sobrevivência estimada", cex.lab=1.4, cex.axis=1.3)
res_exp<-sort(res_exp)
exp1<-exp(-res_exp)
lines(res_exp, exp1, lty=2)
```

```
legend(1.3, 0.9, lty=c(1,2), c("Kaplan-Meier","Exponencial padrão"),
      lwd=1, bty="n", cex=0.8)
```



```
fit_wei <- flexsurvreg(
  Surv(tempo, status) ~ idade + doenca,
  data = df,
  dist='weibull')
res_wei <- resid(fit_wei, type = "coxsnell")
ekm2 <- survfit(
  Surv(res_wei, status) ~ 1,
  data = df, type = "kaplan-meier")
t_wei <- ekm2$time
st_wei <- ekm2$surv
sexp_wei <- exp(-t_wei)

par(mfrow=c(1,2))

plot(st_wei, sexp_wei, xlab = "S(ei): Kaplan-Meier", ylab = "S(ei): Weibull Padrão", p

plot(ekm2, conf.int=F, lty=c(1,1), xlab="Resíduos de Cox-Snell",
      ylab="Sobrevivência Estimada", cex.lab=1.4, cex.axis=1.3)
res_wei <- sort(res_wei)
exp2 <- exp(-res_wei)
```

```
lines(res_wei, exp2, lty=2)
legend(1.3, 0.9, lty=c(1,2), c("Kaplan-Meier", "Weibull Padrão"),
      lwd=1, bty="n", cex=0.8)
```

