



CE077 - Análise de Sobrevida

Trabalho - Análise de Dados

Luiz Francisco - GRR20213026

Mateus Souza - GRR20207154

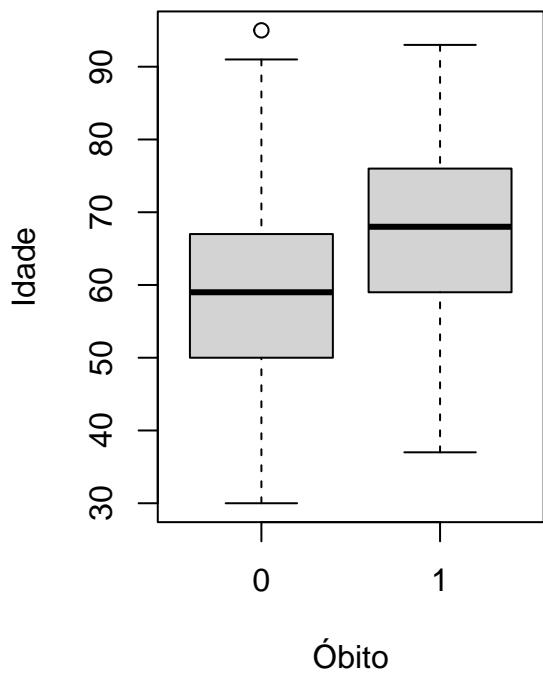
Os dados para este trabalho são um recorte da coorte SamiTrop, um projeto desenvolvido em parceria com pesquisadores de MG e SP e que envolve o acompanhamento e tratamento de mais de dois mil pacientes chagásicos em municípios do norte de MG. Para este estudo, foram realizadas três ondas de avaliação dos pacientes. A primeira ocorreu entre os anos de 2013 e 2014. Os sobreviventes da primeira onda foram avaliados uma segunda vez entre 2015 e 2017. A terceira onda ocorreu entre 2019 e 2021. Em cada onda, os pacientes responderam a questionários, e foram feitos exames laboratoriais, ecocardiogramas e eletrocardiogramas. O objetivo deste trabalho é estudar como algumas covariáveis, medidas na linha de base (2013-2014), estão relacionadas com o desfecho óbito do paciente ao longo de todo o período de acompanhamento. Os dados, contidos no arquivo BancoChagas.RData, contêm informação para 1.000 pacientes aleatoriamente escolhidos e dez covariáveis. A variável Tempo representa o tempo de acompanhamento (em anos) enquanto Obito é a indicadora de óbito do paciente. As demais variáveis são definidas a seguir.

- Sexo: sexo do paciente (masculino ou feminino).
- Idade: idade do paciente em anos.
- Classe_Funcional: classificação de insuficiência cardíaca da New York Heart Association (NYHA), com níveis 1, 2, 3 e 4.
- AIM: infarto (não e sim).
- Rochagan: uso do medicamento Rochagan (não e sim).
- Beta_Bloqueador: uso de medicamento betabloqueador (não e sim).
- Bnp: dosagem de BNP.
- Anticorpos: níveis de anticorpos.
- ECG_Result: alterações do eletro (menores ou maiores).
- FE: fração de ejeção (valor porcentual)

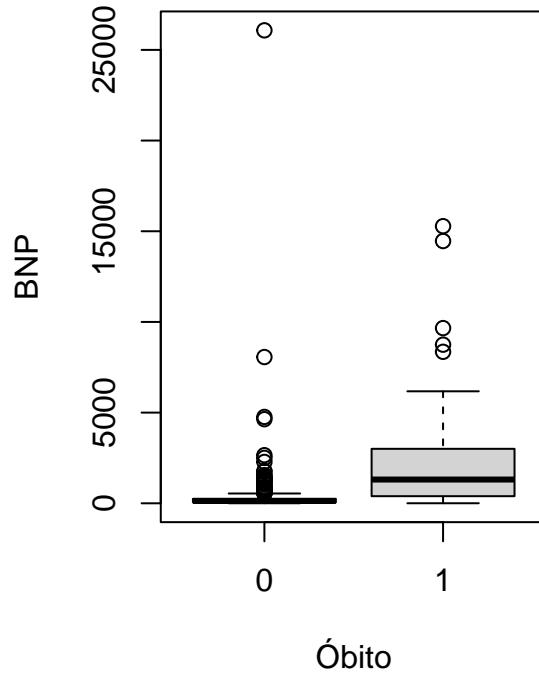
a) Faça uma análise exploratória. Utilize medidas resumo e gráficos de Kaplan-Meier para as curvas de sobrevivência para cada variável.

```
par(mfrow = c(1, 2))
boxplot(Idade ~ Obito, data = dados, main = "Idade do Paciente", ylab = "Idade", xlab = "Óbito")
boxplot(Bnp ~ Obito, data = dados, main = "Dosagem de BNP", ylab = "BNP", xlab = "Óbito")
```

Idade do Paciente

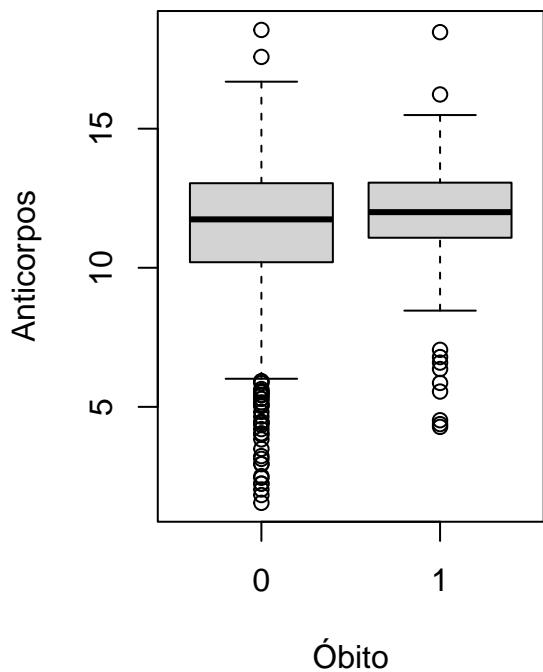


Dosagem de BNP

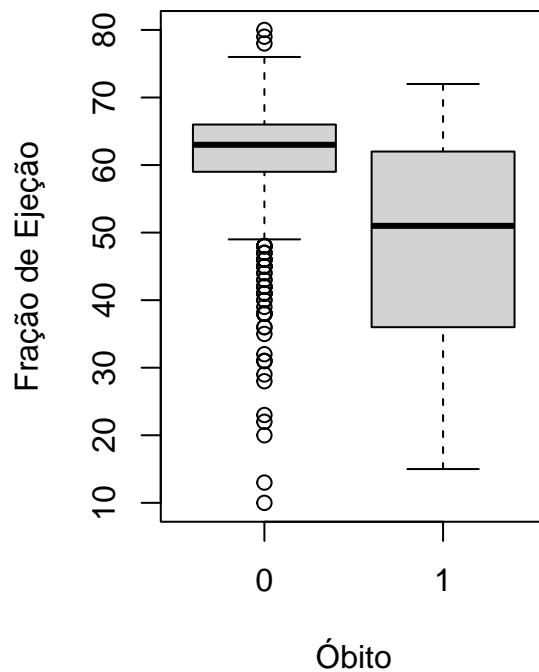


```
par(mfrow = c(1, 2))
boxplot(Anticorpos ~ Óbito, data = dados, main = "Anticorpos", ylab = "Anticorpos", xlab = "Óbito")
boxplot(FE ~ Óbito, data = dados, main = "Fração de Ejeção", ylab = "Fração de Ejeção", xlab = "Óbito")
```

Anticorpos



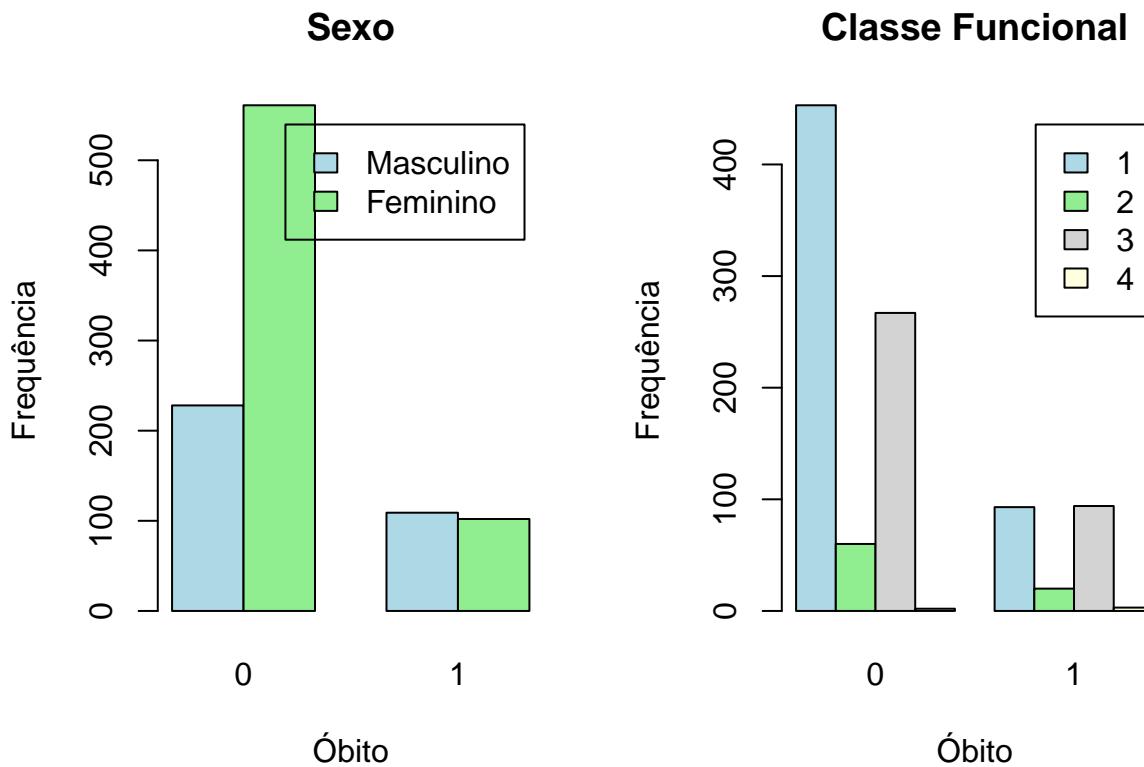
Fração de Ejeção



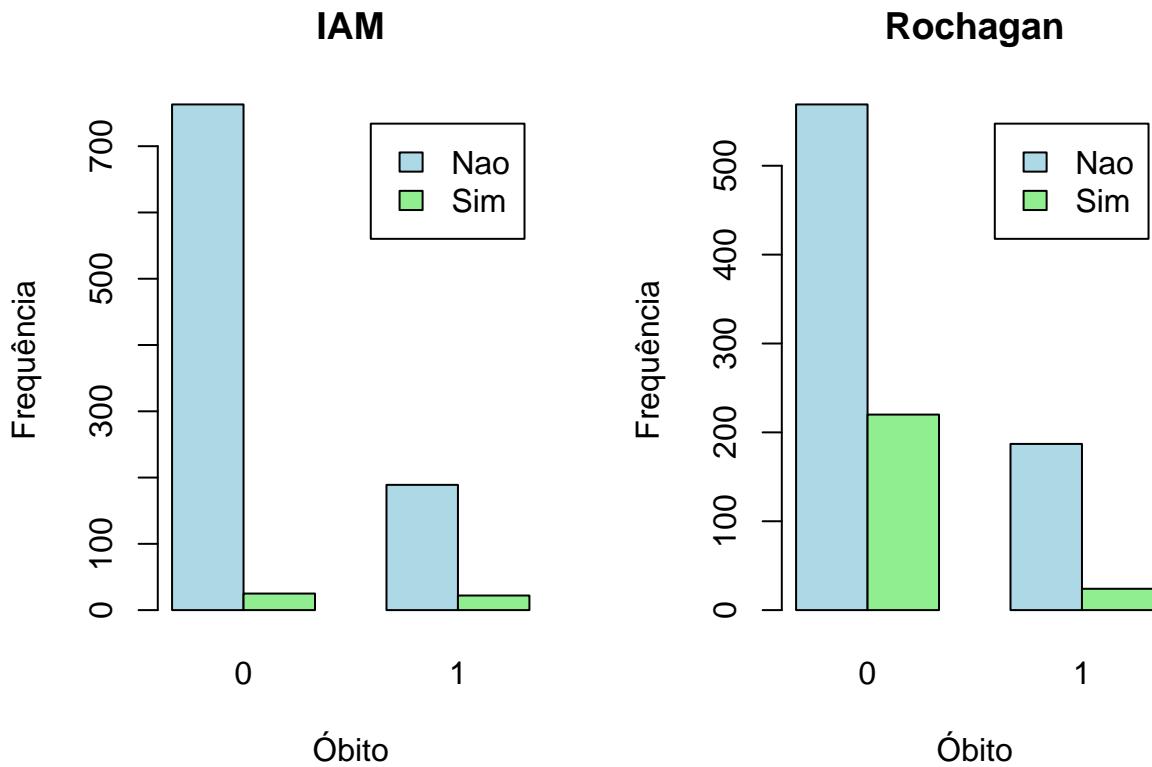
No primeiro gráfico, observamos que a mediana da idade dos pacientes que faleceram (grupo 1) é ligeiramente maior em comparação com os que sobreviveram (grupo 0). Isso indica que a idade pode ser um fator relevante

para o desfecho de óbito. No segundo gráfico, a dosagem de BNP (uma medida relacionada à insuficiência cardíaca) é significativamente maior no grupo de óbito, sugerindo uma associação entre níveis elevados de BNP e maior risco de mortalidade. No terceiro gráfico, os níveis de anticorpos não apresentam uma diferença clara entre os grupos, indicando que essa variável pode não estar fortemente associada ao desfecho. No quarto gráfico, a fração de ejeção, que mede a eficiência do coração em bombear sangue, é menor no grupo de óbito, o que é consistente com a presença de insuficiência cardíaca mais grave nesse grupo. Esses gráficos indicam que variáveis como idade, BNP e fração de ejeção têm uma relação significativa com o desfecho de óbito, enquanto os níveis de anticorpos parecem menos relevantes. É interessante destacar também que os dados apresentam uma grande variação e a visualização destes pelo gráfico boxplot fica prejudicada.

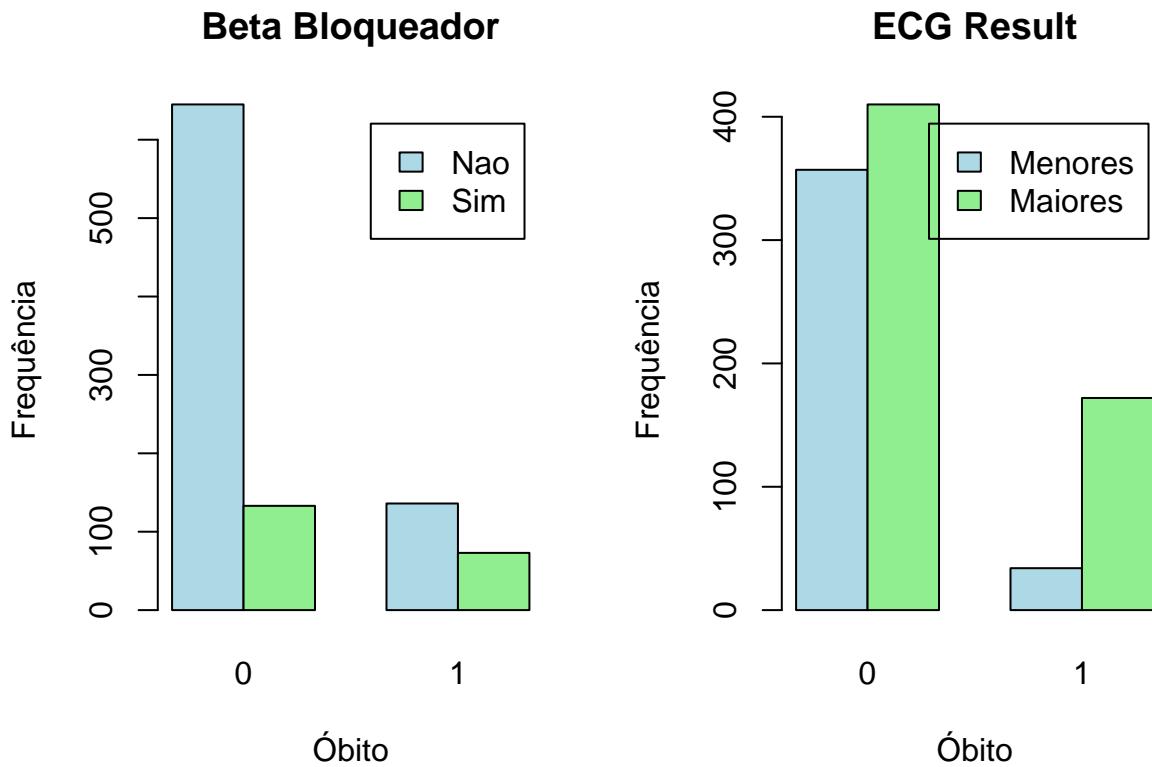
```
par(mfrow = c(1, 2))
barplot(table(dados$Sexo, dados$Obito), beside = TRUE, col = c("lightblue", "lightgreen"),
legend = rownames(table(dados$Sexo)), main = "Sexo", ylab = "Frequência", xlab = "Óbito")
barplot(table(dados$Classe_Funcional, dados$Obito), beside = TRUE, col = c("lightblue",
"lightgreen", "lightgray", "lightyellow"), legend = rownames(table(dados$Classe_Funcional)),
main = "Classe Funcional", ylab = "Frequência", xlab = "Óbito")
```



```
par(mfrow = c(1, 2))
barplot(table(dados$IAM, dados$Obito), beside = TRUE, col = c("lightblue", "lightgreen"),
legend = rownames(table(dados$IAM)), main = "IAM", ylab = "Frequência", xlab = "Óbito")
barplot(table(dados$Rochagan, dados$Obito), beside = TRUE, col = c("lightblue", "lightgreen"),
legend = rownames(table(dados$Rochagan)), main = "Rochagan", ylab = "Frequência", xlab = "Óbito")
```

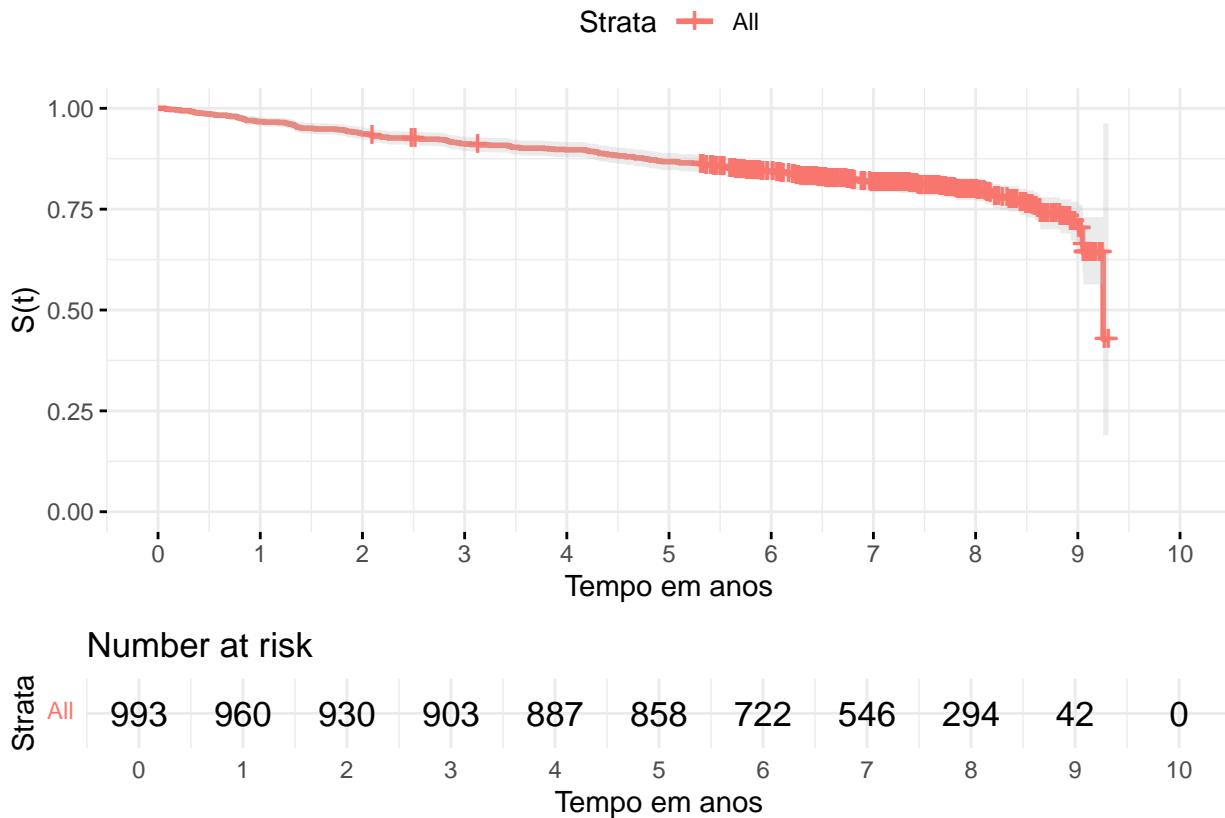


```
par(mfrow = c(1, 2))
barplot(table(dados$Beta_Bloqueador, dados$Obito), beside = TRUE, col = c("lightblue", "lightgreen"),
        legend = rownames(table(dados$Beta_Bloqueador)), main = "Beta Bloqueador",
        ylab = "Frequência", xlab = "Óbito")
barplot(table(dados$ECG_Result, dados$Obito), beside = TRUE, col = c("lightblue", "lightgreen"),
        legend = rownames(table(dados$ECG_Result)), main = "ECG Result", ylab = "Frequência", xlab = "Óbito")
```



Os gráficos de barras fornecem uma visão das variáveis categóricas em relação ao desfecho de óbito (0 para não óbito e 1 para óbito). O primeiro gráfico mostra que a distribuição de sexo sugere maior sobrevida para o sexo feminino. No gráfico da Classe Funcional, a maioria dos pacientes no nível 1 sobreviveram, enquanto aqueles nos níveis mais altos (3 e 4) têm uma maior proporção de óbitos. No gráfico de IAM, a maioria dos pacientes que não sofreram infarto sobreviveram, com poucos óbitos entre os que tiveram infarto. No gráfico de Rochagan, há um número pequeno de óbitos com uso do medicamento, aparentemente indicando sua eficácia na prevenção. No gráfico de Beta Bloqueador, a maioria dos pacientes que não usaram o medicamento sobreviveram, enquanto os que usaram têm uma maior proporção de óbitos. Finalmente, no gráfico de ECG Result, a maioria dos pacientes com alterações menores sobreviveram, enquanto aqueles com alterações maiores têm uma maior proporção de óbitos. Esses gráficos indicam que variáveis como Classe Funcional, AIM, uso de Beta Bloqueador e resultados de ECG podem ter uma relação significativa com o desfecho de óbito, o que será melhor investigado nas curvas de sobrevida por covariável a serem desenvolvidas a seguir.

```
ekm <- survfit(Surv(Tempo, Óbito) ~ 1, data = dados)
ggsurvplot(ekm, risk.table = TRUE, break.time.by = 1, xlab = "Tempo em anos",
           ylab = "S(t)", ggtheme = theme_minimal(), conf.int = TRUE)
```



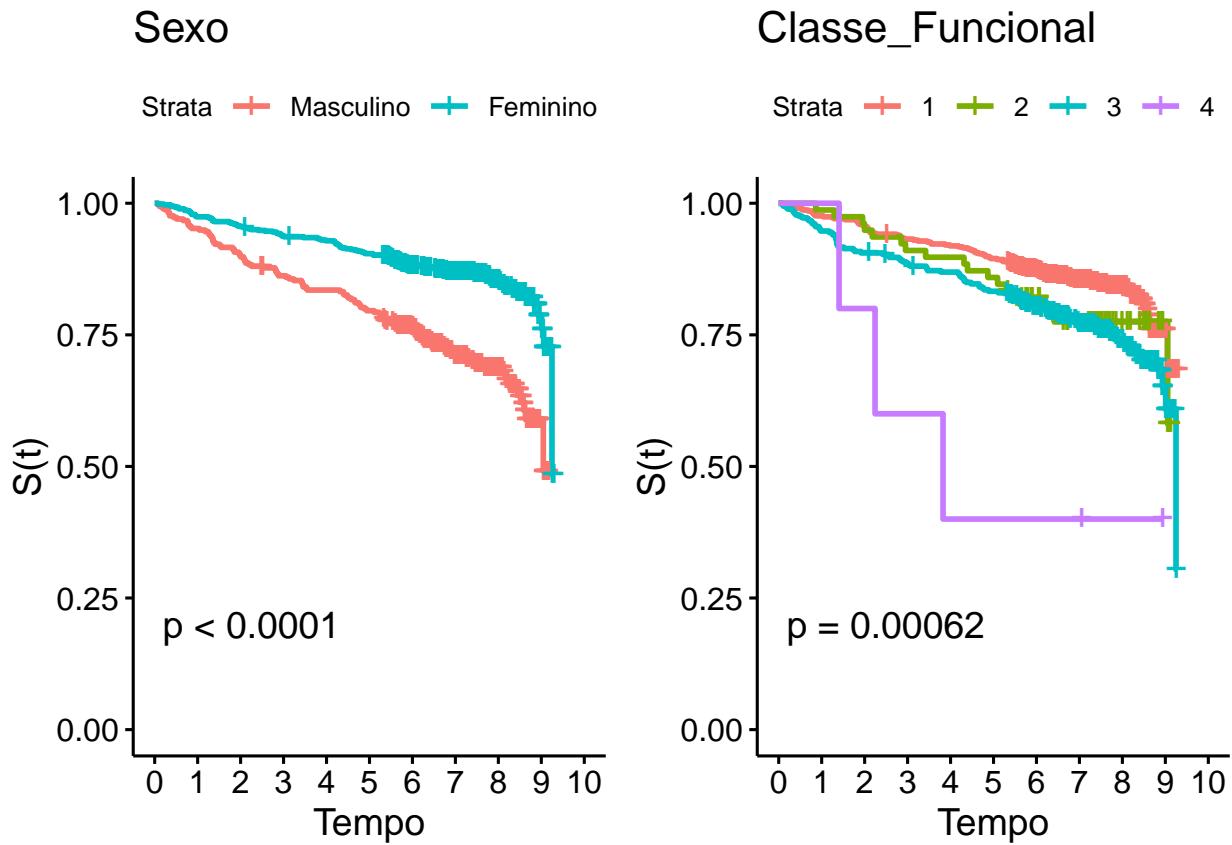
```
print(ekm, print.rmean = TRUE)
```

```
## Call: survfit(formula = Surv(Tempo, Obito) ~ 1, data = dados)
##
##      7 observations deleted due to missingness
##      n events rmean* se(rmean) median 0.95LCL 0.95UCL
## [1,] 993     205    8.09    0.0797   9.25    9.25      NA
##      * restricted mean with upper limit =  9.3
```

Acima, temos as estimativas de tempo médio, acompanhada de seu intervalo com 95% de confiança, e de tempo mediano, que é de 9.25 anos.

```
ekm_V1 <- ggsurvplot(survfit(Surv(Tempo, Obito) ~ Sexo, data = dados), title = "Sexo",
break.time.by = 1, pval = TRUE, xlab = "Tempo", ylab = "S(t)", legend.labs = levels(dados$Sexo))
ekm_V2 <- ggsurvplot(survfit(Surv(Tempo, Obito) ~ Classe_Funcional, data = dados),
title = "Classe_Funcional", break.time.by = 1, pval = TRUE, xlab = "Tempo", ylab = "S(t)",
legend.labs = levels(dados$Classe_Funcional))
ekm_V3 <- ggsurvplot(survfit(Surv(Tempo, Obito) ~ IAM, data = dados), title = "IAM",
break.time.by = 1, pval = TRUE, xlab = "Tempo", ylab = "S(t)", legend.labs = levels(dados$IAM))
ekm_V4 <- ggsurvplot(survfit(Surv(Tempo, Obito) ~ Rochagan, data = dados), title = "Rochagan",
break.time.by = 1, pval = TRUE, xlab = "Tempo", ylab = "S(t)", legend.labs = levels(dados$Rochagan))
ekm_V5 <- ggsurvplot(survfit(Surv(Tempo, Obito) ~ Beta_Bloqueador, data = dados),
title = "Beta_Bloqueador", break.time.by = 1, pval = TRUE, xlab = "Tempo", ylab = "S(t)",
legend.labs = levels(dados$Beta_Bloqueador))
ekm_V6 <- ggsurvplot(survfit(Surv(Tempo, Obito) ~ ECG_Result, data = dados),
title = "ECG_Result", break.time.by = 1, pval = TRUE, xlab = "Tempo", ylab = "S(t)",
legend.labs = levels(dados$ECG_Result))
```

```
arrange_ggsurvplots(list(ekm_V1, ekm_V2), print = TRUE, ncol = 2, nrow = 1, ggtheme = theme_minimal())
```

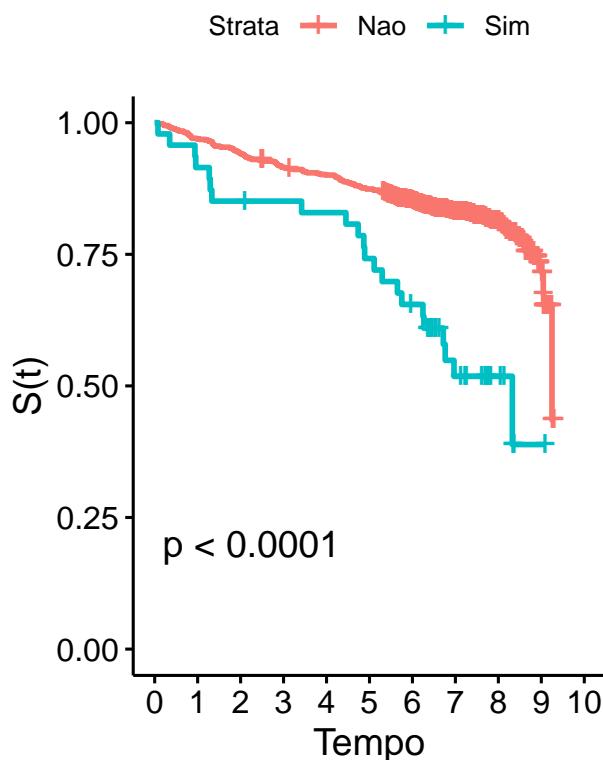


Sexo: A curva de sobrevivência para o sexo masculino apresenta valores consistentemente inferiores em comparação à curva do sexo feminino ao longo de todo o estudo.

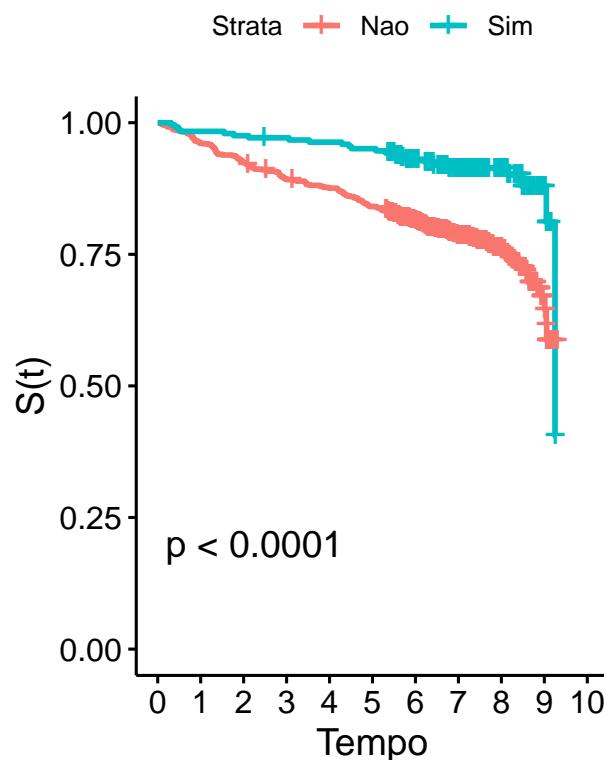
Classe Funcional: A classe funcional 1 mantém-se com a maior taxa de sobrevivência, seguida pela classe 2 e pela classe 3. A classe funcional 4 apresenta a pior performance em termos de sobrevivência durante todo o período do estudo. No entanto, esta classe possui apenas 5 observações no dataset, o que limita a confiança nos resultados.

```
arrange_ggsurvplots(list(ekm_V3, ekm_V4), print = TRUE, ncol = 2, nrow = 1, ggtheme = theme_minimal())
```

IAM



Rochagan

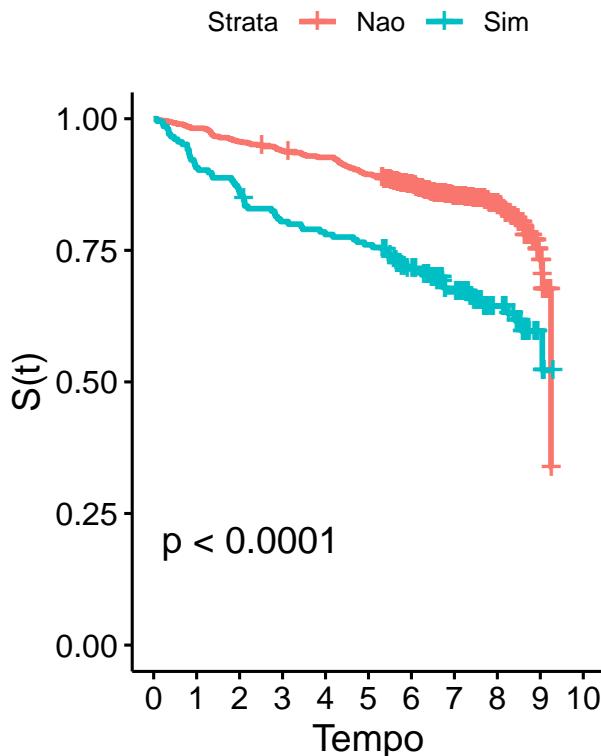


IAM: A curva de sobrevivência para pacientes com IAM (Sim) é inferior à daqueles sem IAM (Não) ao longo de todo o estudo, indicando uma menor sobrevivência para os pacientes com histórico de infarto.

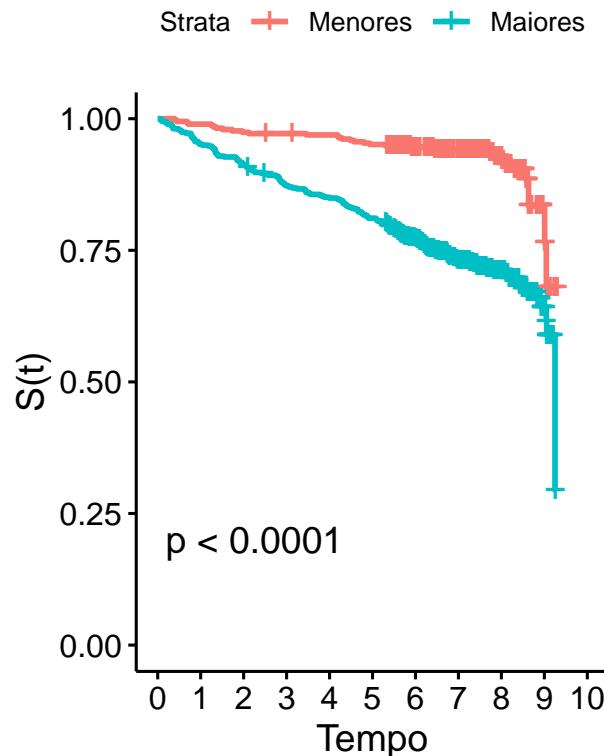
Rochagan: A curva de sobrevivência para pacientes que não utilizaram o medicamento Rochagan é inferior à daqueles que utilizaram, sugerindo uma possível eficácia do medicamento. Contudo, no ponto final do estudo, a curva de sobrevivência dos usuários do medicamento é menor, o que necessita de uma investigação mais aprofundada.

```
arrange_ggsurvplots(list(ekm_V5, ekm_V6), print = TRUE, ncol = 2, nrow = 1, ggtheme = theme_minimal())
```

Beta_Bloqueador



ECG_Result



Beta Bloqueador: A curva de sobrevivência para pacientes que não utilizaram beta bloqueadores é consistentemente superior à daqueles que utilizaram, indicando uma menor eficácia do medicamento para este grupo. Similar ao resultado para Rochagan, no ponto final do estudo, a curva de sobrevivência dos não-usuários do medicamento é menor, o que requer uma análise mais detalhada.

ECG : A curva de sobrevivência para pacientes com resultados de ECG maiores está sempre abaixo da curva daqueles com resultados menores, indicando uma pior sobrevivência para os pacientes com alterações significativas no eletrocardiograma.

Esses resultados sugerem diferenças significativas nas taxas de sobrevivência entre as diferentes categorias das covariáveis analisadas.

b) Ajuste um modelo paramétrico. Faça a seleção das covariáveis. Justifique a escolha pelo modelo

```

ajust1 <- survreg(Surv(Tempo, Obito) ~ 1, dist = 'exponential', data = dados)
ajust2 <- survreg(Surv(Tempo, Obito) ~ 1, dist = 'weibull', data = dados)
ajust3 <- survreg(Surv(Tempo, Obito) ~ 1, dist = 'lognorm', data = dados)

time <- ekm$time
st <- ekm$surv
ste <- exp(-time/exp(ajust1$coefficients[1]))
stw <- exp(-(time/exp(ajust2$coefficients[1]))^(1/ajust2$scale))
stln <- pnorm((-log(time) + ajust3$coefficients[1])/ajust3$scale)

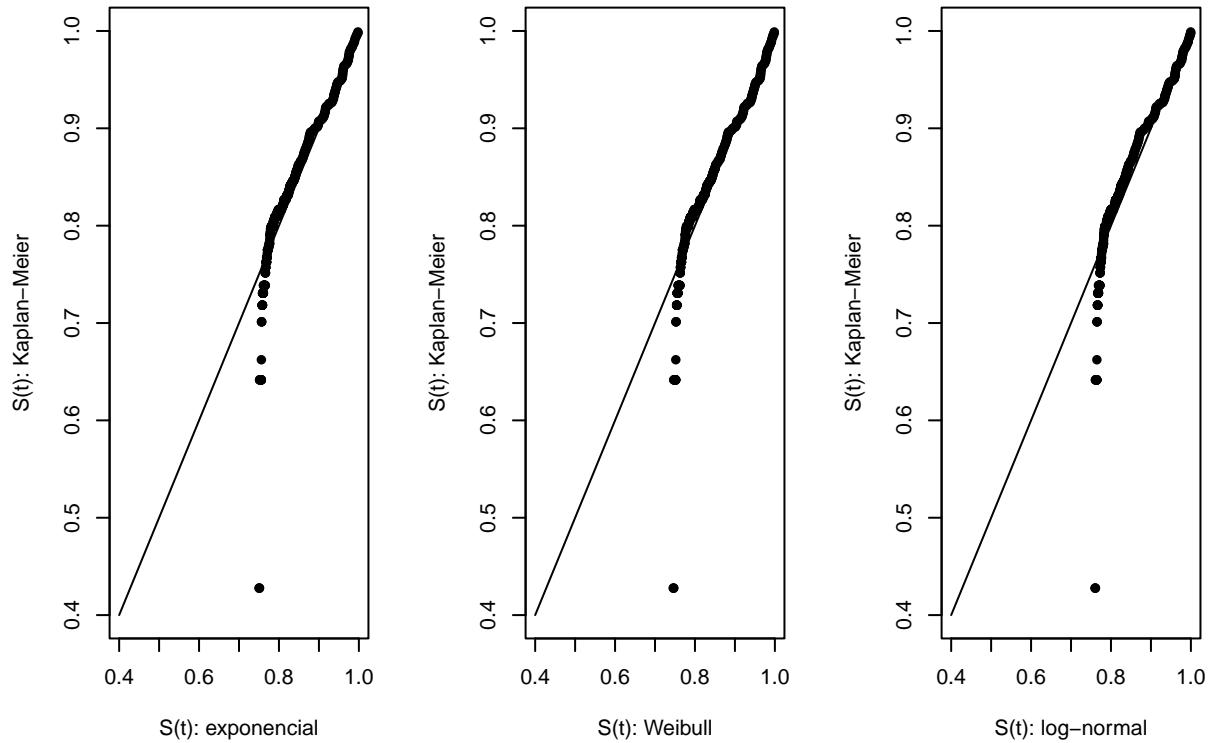
par(mfrow = c(1, 3))
plot(ste, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): exponencial")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stw, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Weibull", xlab = "S(t): exponencial")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stln, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Lognormal", xlab = "S(t): exponencial")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)

```

```

ylab = "S(t): Kaplan-Meier", xlab = "S(t): Weibull")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stln, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): log-normal")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)

```



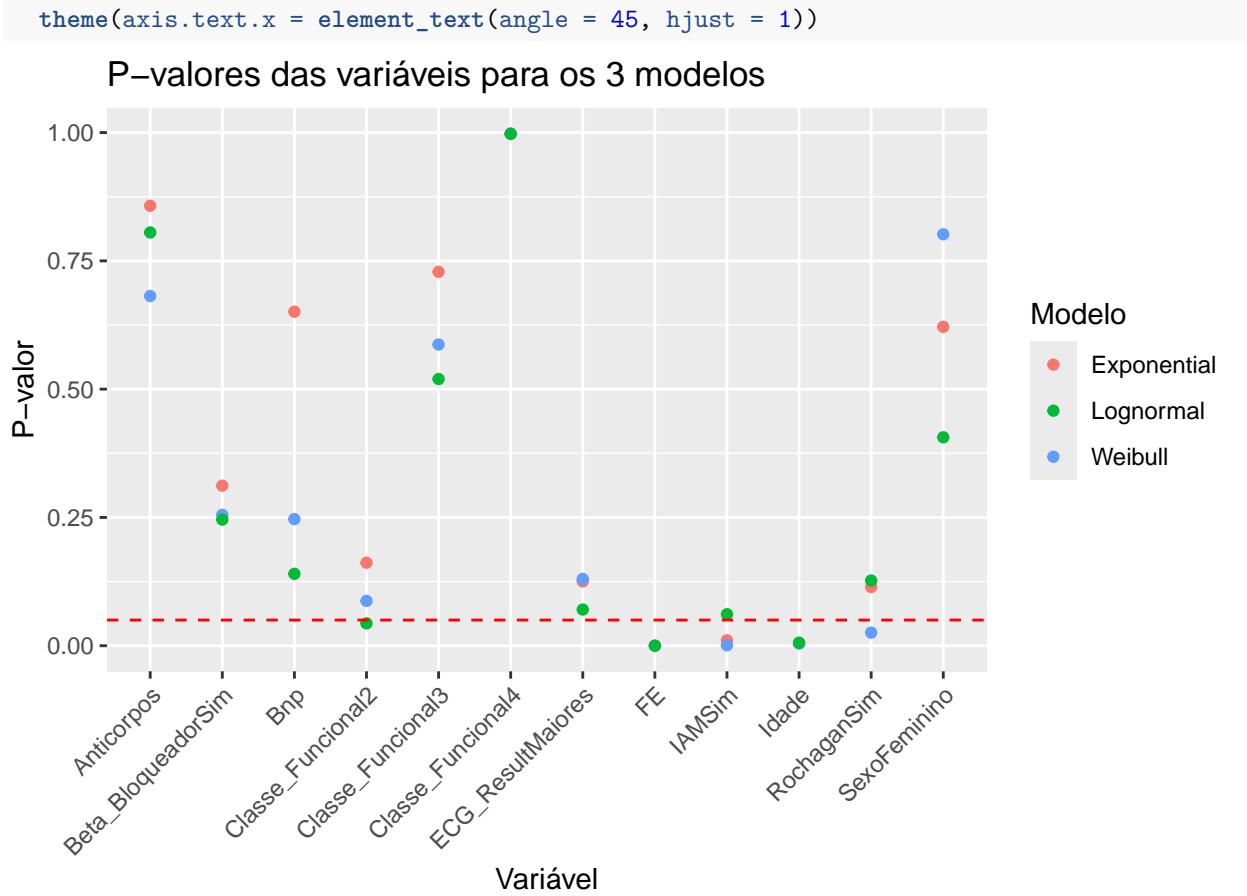
Os modelos ajustados utilizando as distribuições exponencial, Weibull e log-normal não estão representando adequadamente os dados de sobrevida, conforme evidenciado pelos gráficos de ajuste. As curvas Kaplan-Meier estão consideravelmente afastadas da linha de identidade, indicando um ajuste inadequado. Para melhorar o ajuste dos modelos, podemos explorar outras distribuições ou incluir covariáveis relevantes no modelo.

```

ajust1_cov <- survreg(Surv(Tempo, Obito) ~ Sexo + Classe_Funcional + IAM + Rochagan +
Beta_Bloqueador + ECG_Result + Bnp + Idade + FE + Anticorpos, dist = 'exponential', data = dados)
ajust2_cov <- survreg(Surv(Tempo, Obito) ~ Sexo + Classe_Funcional + IAM + Rochagan +
Beta_Bloqueador + ECG_Result + Bnp + Idade + FE + Anticorpos, dist = 'weibull', data = dados)
ajust3_cov <- survreg(Surv(Tempo, Obito) ~ Sexo + Classe_Funcional + IAM + Rochagan +
Beta_Bloqueador + ECG_Result + Bnp + Idade + FE + Anticorpos, dist = 'lognorm', data = dados)

p_values_df <- data.frame(
  Variable = rownames(summary(ajust1_cov)$table)[2:13],
  Exponential = summary(ajust1_cov)$table[, "p"] [2:13],
  Weibull = summary(ajust2_cov)$table[, "p"] [2:13],
  Lognormal = summary(ajust3_cov)$table[, "p"] [2:13])
p_values_long <- gather(p_values_df, Modelo, p_value, -Variable)
ggplot(p_values_long, aes(x = Variable, y = p_value, color = Modelo, group = Modelo)) +
  geom_point() +
  geom_hline(yintercept = 0.05, linetype = "dashed", color = "red") +
  labs(title = "P-valores das variáveis para os 3 modelos",
       x = "Variável", y = "P-valor") +

```



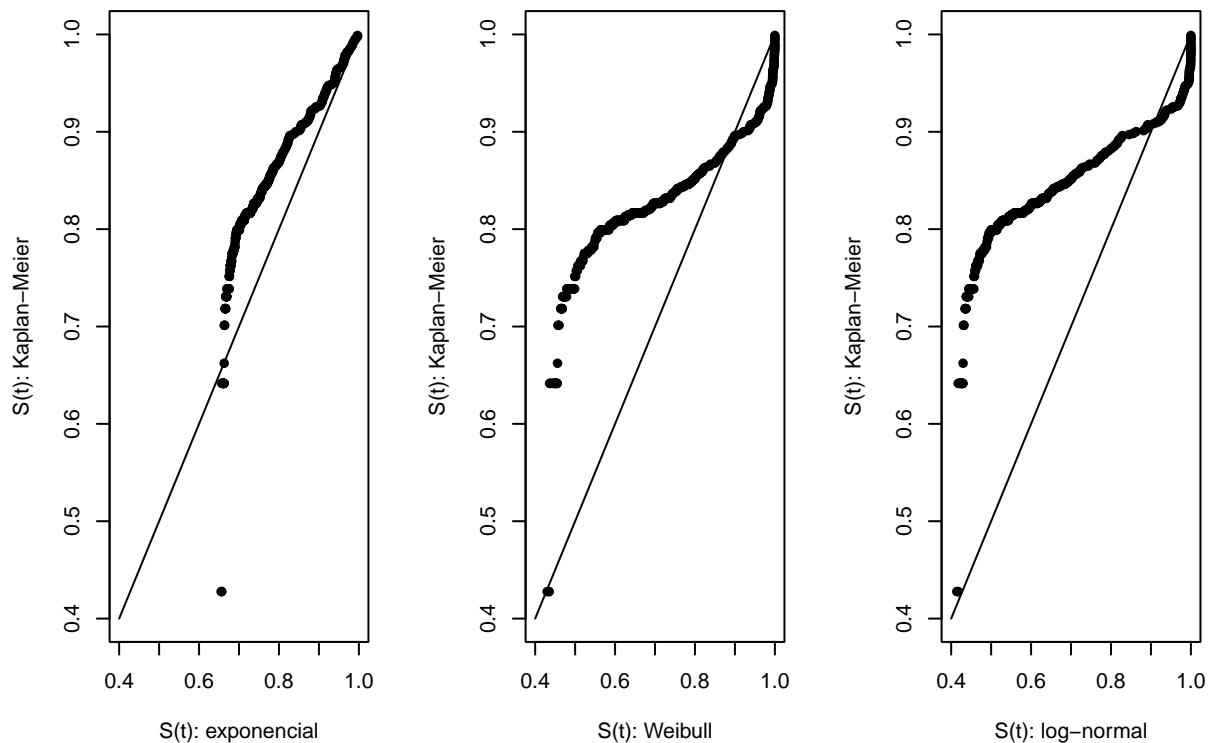
Os próximos modelos a serem ajustados levarão em consideração as seguintes covariáveis: Classe_Funcional, IAM, Rochagan, Idade e FE, pois estas apresentaram valores estatisticamente significativos como evidenciado no gráfico acima. Abaixo, temos nova verificação de ajuste do modelo.

```
ajust4 <- survreg(Surv(Tempo, Obito) ~ Classe_Funcional + IAM + Rochagan + Idade + FE,
                     dist = 'exponential', data = dados)
ajust5 <- survreg(Surv(Tempo, Obito) ~ Classe_Funcional + IAM + Rochagan + Idade + FE,
                     dist = 'weibull', data = dados)
ajust6 <- survreg(Surv(Tempo, Obito) ~ Classe_Funcional + IAM + Rochagan + Idade + FE,
                     dist = 'lognorm', data = dados)

time <- ekm$time
st <- ekm$surv
ste <- exp(-time/exp(ajust4$coefficients[1]))
stw <- exp(-(time/exp(ajust5$coefficients[1]))^(1/ajust5$scale))
stln <- pnorm((-log(time) + ajust6$coefficients[1])/ajust6$scale)

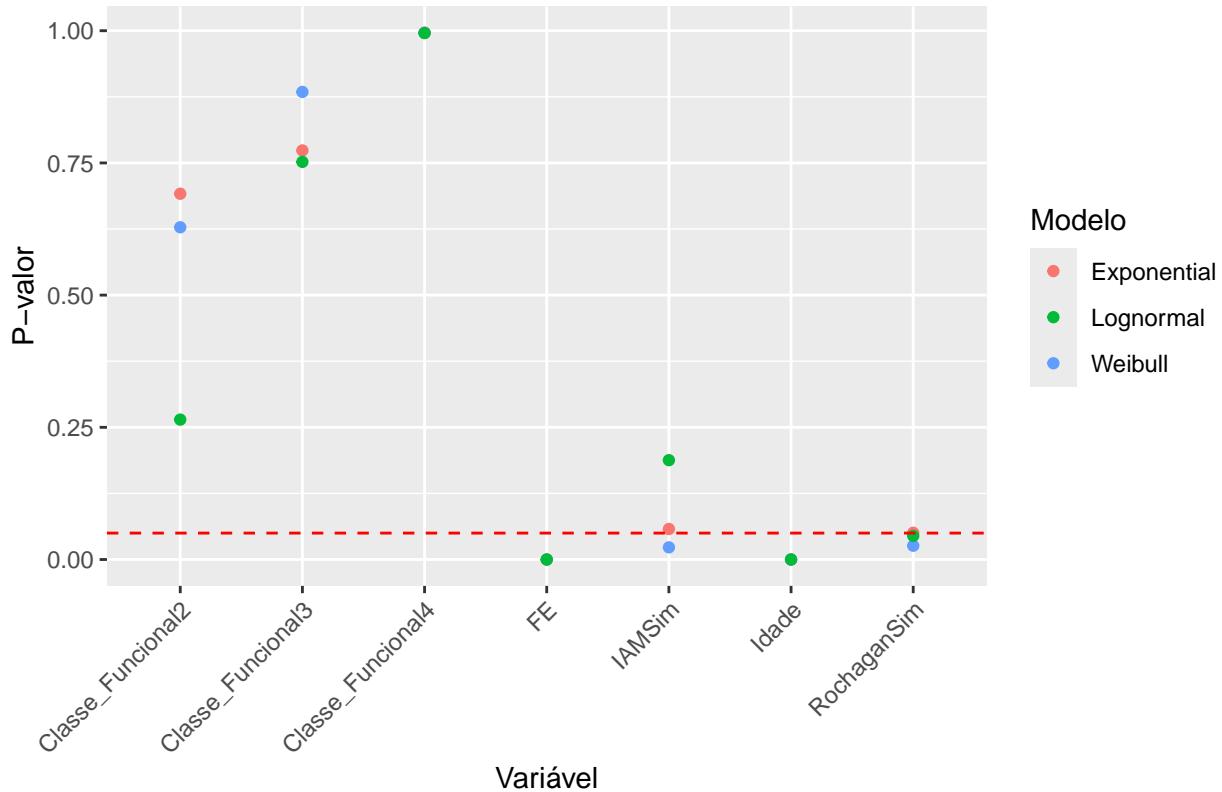
par(mfrow = c(1, 3))
plot(ste, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
     ylab = "S(t): Kaplan-Meier", xlab = "S(t): exponencial")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stw, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
     ylab = "S(t): Kaplan-Meier", xlab = "S(t): Weibull")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stln, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
     ylab = "S(t): Kaplan-Meier", xlab = "S(t): log-normal")
```

```
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
```



```
p_values_df <- data.frame(
  Variable = rownames(summary(ajust4)$table)[2:8],
  Exponential = summary(ajust4)$table[, "p"][2:8],
  Weibull = summary(ajust5)$table[, "p"][2:8],
  Lognormal = summary(ajust6)$table[, "p"][2:8])
p_values_long <- gather(p_values_df, Modelo, p_value, -Variable)
ggplot(p_values_long, aes(x = Variable, y = p_value, color = Modelo, group = Modelo)) +
  geom_point() +
  geom_hline(yintercept = 0.05, linetype = "dashed", color = "red") +
  labs(title = "P-valores das variáveis para os 3 modelos",
       x = "Variável",
       y = "P-valor") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

P–valores das variáveis para os 3 modelos



Como mostrado no gráfico acima a variável Classe Funcional não é importante para o modelo e portanto será removida no decorrer dos próximos ajustes.

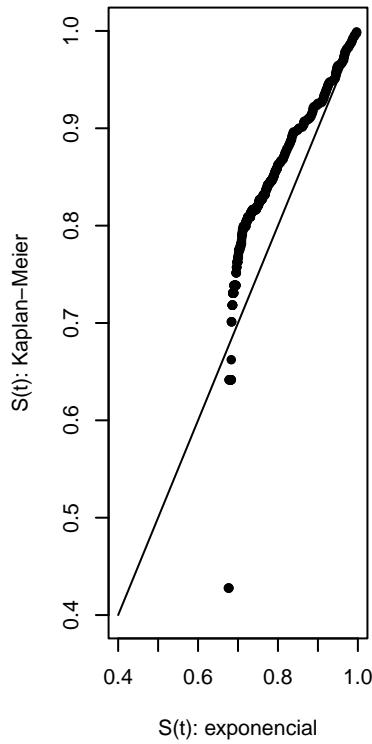
```

ajust7 <- survreg(Surv(Tempo, Obito) ~ IAM + Rochagan + Idade + FE, dist = 'exponential', data = dados)
ajust8 <- survreg(Surv(Tempo, Obito) ~ IAM + Rochagan + Idade + FE, dist = 'weibull', data = dados)
ajust9 <- survreg(Surv(Tempo, Obito) ~ IAM + Rochagan + Idade + FE, dist = 'lognorm', data = dados)

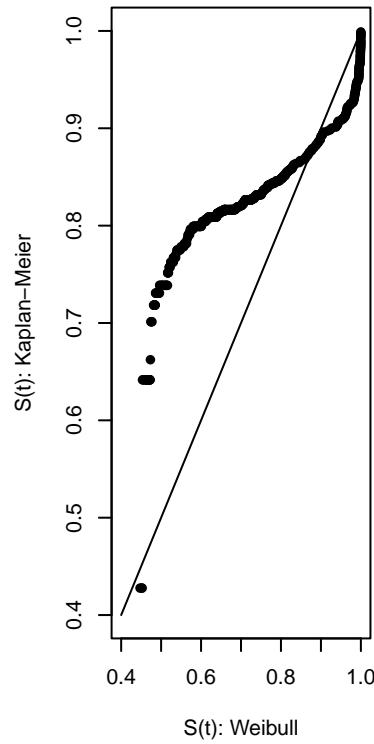
time <- ekm$time
st <- ekm$surv
ste <- exp(-time/exp(ajust7$coefficients[1]))
stw <- exp(-(time/exp(ajust8$coefficients[1]))^(1/ajust8$scale))
stln <- pnorm((-log(time) + ajust9$coefficients[1])/ajust9$scale)

par(mfrow = c(1, 3))
plot(ste, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): exponencial")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stw, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): Weibull")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stln, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): log-normal")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)

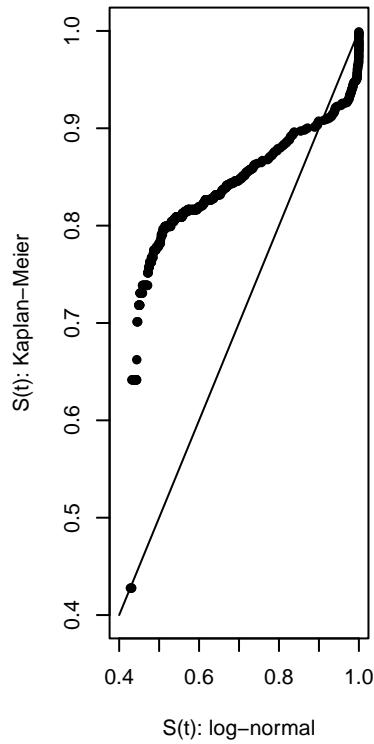
```



S(t): exponential



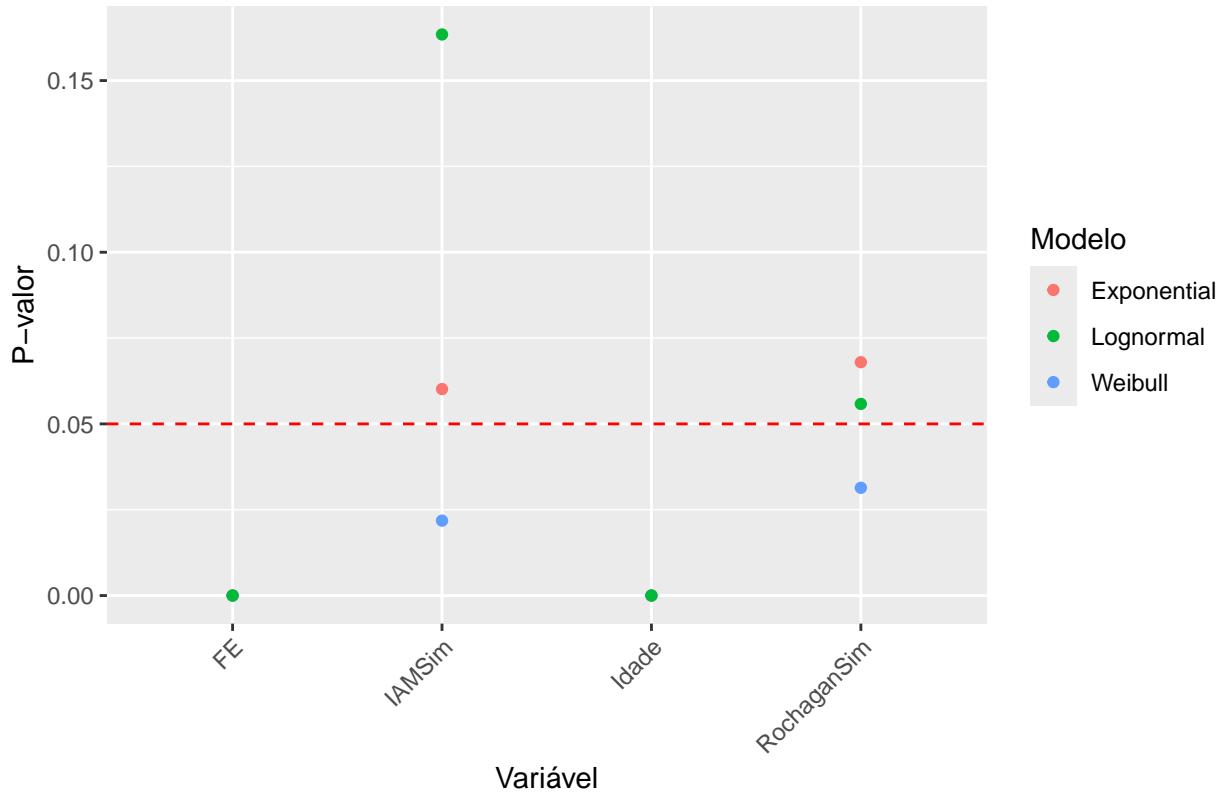
S(t): Weibull



S(t): log-normal

```
p_values_df <- data.frame(
  Variable = rownames(summary(ajust7)$table)[2:5],
  Exponential = summary(ajust7)$table[, "p"][2:5],
  Weibull = summary(ajust8)$table[, "p"][2:5],
  Lognormal = summary(ajust9)$table[, "p"][2:5])
p_values_long <- gather(p_values_df, Modelo, p_value, -Variable)
ggplot(p_values_long, aes(x = Variable, y = p_value, color = Modelo, group = Modelo)) +
  geom_point() +
  geom_hline(yintercept = 0.05, linetype = "dashed", color = "red") +
  labs(title = "P-valores das variáveis para os 3 modelos",
       x = "Variável",
       y = "P-valor") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

P–valores das variáveis para os 3 modelos



Considerando que as variáveis Idade e FE destoam consideravelmente em seus p–valores, iremos realizar um ajuste de modelos levando em consideração apenas as duas.

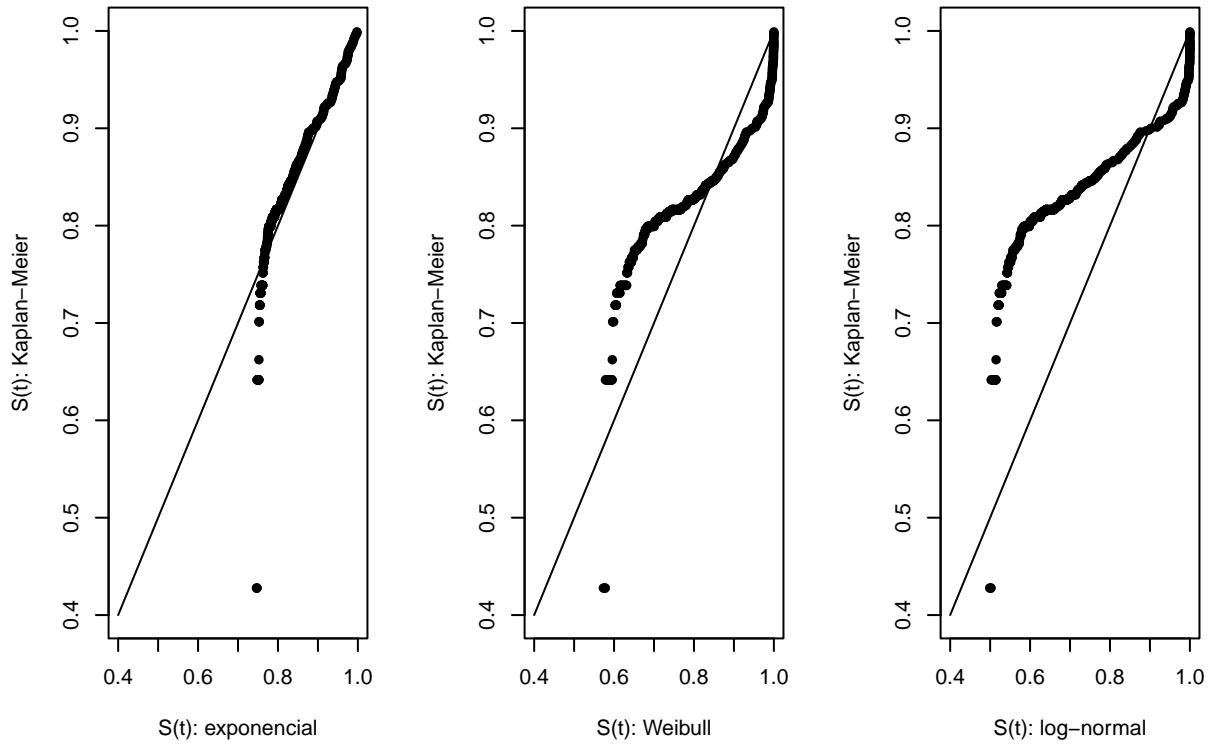
```

ajust10 <- survreg(Surv(Tempo, Obito) ~ Idade + FE, dist = 'exponential', data = dados)
ajust11 <- survreg(Surv(Tempo, Obito) ~ Idade + FE, dist = 'weibull', data = dados)
ajust12 <- survreg(Surv(Tempo, Obito) ~ Idade + FE, dist = 'lognorm', data = dados)

time <- ekm$time
st <- ekm$surv
ste <- exp(-time/exp(ajust10$coefficients[1]))
stw <- exp(-(time/exp(ajust11$coefficients[1]))^(1/ajust11$scale))
stln <- pnorm((-log(time) + ajust12$coefficients[1])/ajust12$scale)

par(mfrow = c(1, 3))
plot(ste, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): exponencial")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stw, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): Weibull")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)
plot(stln, st, pch = 16, ylim = range(c(0.4, 1)), xlim = range(c(0.4, 1)),
ylab = "S(t): Kaplan-Meier", xlab = "S(t): log-normal")
lines(c(0.4, 1), c(0.4, 1), type = "l", lty = 1)

```



```
anova(ajust10,ajust11,ajust12)
```

```
##          Terms Resid. Df -2*LL Test Df Deviance Pr(>Chi)
## 1 Idade + FE     691 905.7723    NA     NA      NA
## 2 Idade + FE     690 823.9037  = 1 81.868527 1.454471e-19
## 3 Idade + FE     690 827.5338  = 0 -3.630022      NA
```

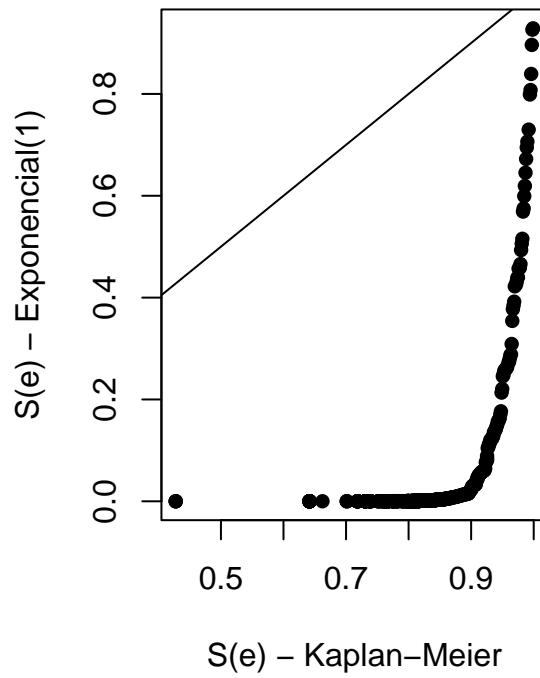
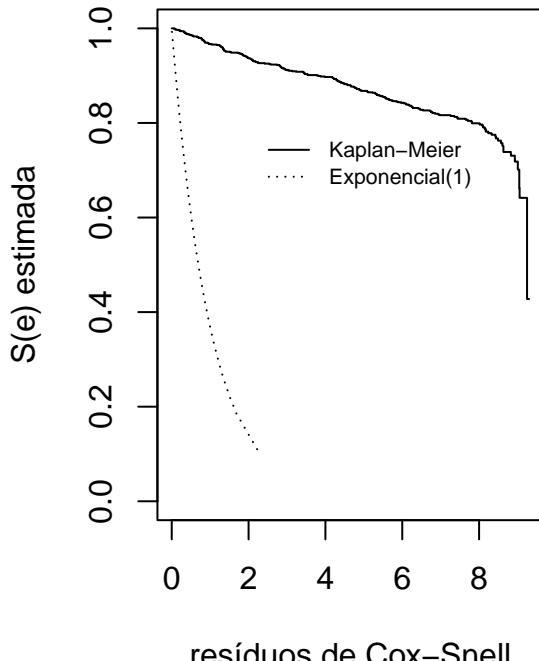
Conforme mostrado no gráfico acima e também pelo resultado da ANOVA, o modelo exponencial foi o que mais se adequou a curva de sobrevivência estimado dos dados, portanto se configura como modelo final paramétrico escolhido, que ficou da forma:

$$\lambda(t|Idade, FE) = \exp(\beta_0 + \beta_{Idade} * Idade + \beta_{FE} * FE)$$

$$S(t|Idade, FE) = \exp(-t/\exp(\beta_0 + \beta_{Idade} * Idade + \beta_{FE} * FE)))$$

```
t <- dados$Tempo
x_idade <- dados$Idade
x_fe <- dados$FE
bo <- ajust10$coefficients[1]
b1 <- ajust10$coefficients[2]
b2 <- ajust10$coefficients[3]
res <- t*exp(-bo-b1*x_idade-b2*x_fe) # resíduos de Cox-Snell
par(mfrow = c(1, 2))
plot(ekm, conf.int = F, lty = c(1, 1), xlab = "resíduos de Cox-Snell",
ylab = "S(e) estimada")
res <- sort(res)
exp1 <- exp(-res)
lines(res, exp1, lty = 3)
legend(2, 0.8, lty = c(1, 3), c("Kaplan-Meier", "Exponencial(1)"), lwd = 1, bty = "n",
cex = 0.7)
st <- ekm$surv
```

```
t <- ekm$time
sexp1 <- exp(-t)
plot(st, sexp1, xlab = "S(e) - Kaplan-Meier", ylab = "S(e) - Exponencial(1)", pch = 16)
abline(coef = c(0, 1))
```



Conforme mostrado nos resíduos acima, o modelo exponencial não parece se adequar bem aos dados, apresentando discrepâncias significativas em várias partes da curva de sobrevivência. Portanto, iremos partir para o modelo de Cox, que é um modelo semi-paramétrico e pode lidar melhor com a complexidade dos dados.

```
print(paste("Número de valores nulos na coluna Idade:", sum(is.na(dados$Idade))))
```

```
## [1] "Número de valores nulos na coluna Idade: 205"
```

```
print(paste("Número de valores nulos na coluna FE:", sum(is.na(dados$FE))))
```

```
## [1] "Número de valores nulos na coluna FE: 302"
```

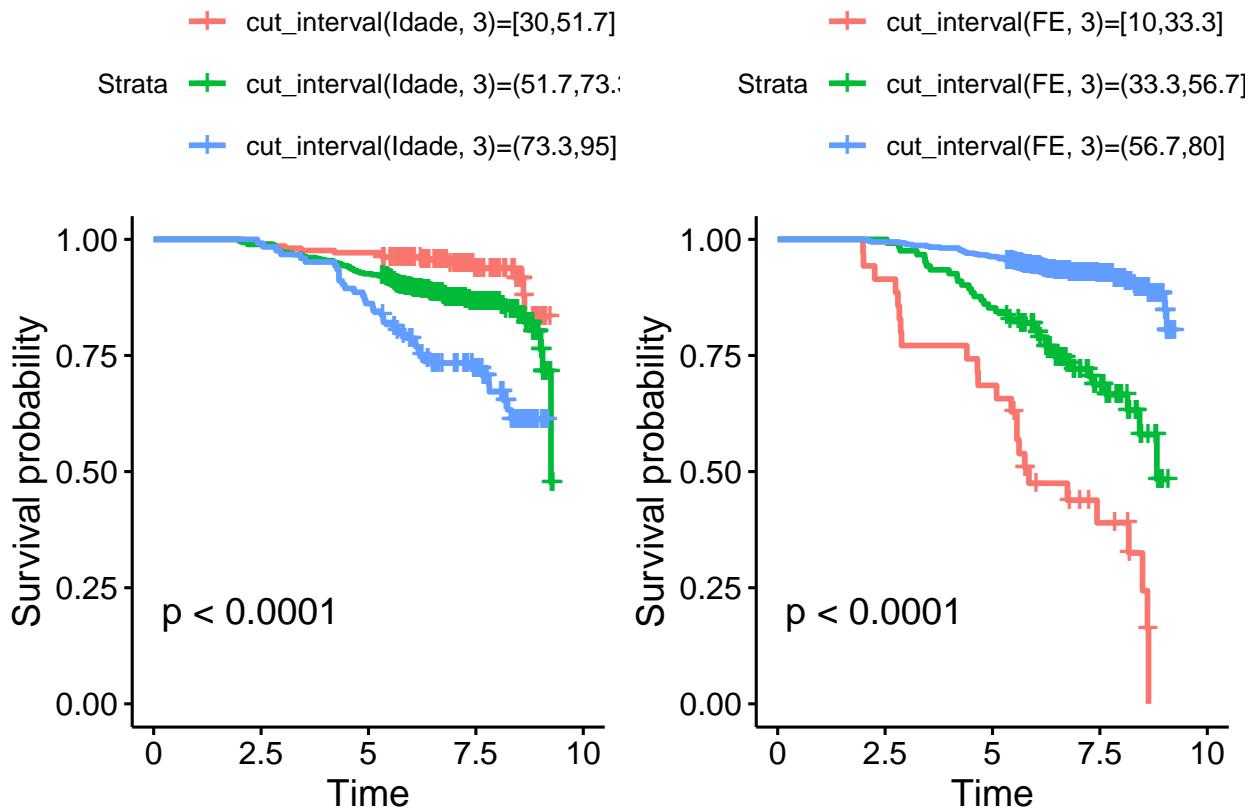
Cabe ressaltar que as duas covariáveis utilizadas apresentam observações com valores nulos em uma quantidade significativa do dataset, o que pode estar contribuindo para o mau ajuste do modelo.

c) Ajuste o modelo de Cox. Faça a seleção das covariáveis

```
ekm <- survfit(Surv(Tempo, Obito) ~ 1, data = dados)
ekm2 <- survfit(Surv(Tempo, Obito) ~ cut_interval(Idade, 3), data = dados)
ekm3 <- survfit(Surv(Tempo, Obito) ~ cut_interval(FE, 3), data = dados)

plot1 <- ggsurvplot(ekm2, pval = TRUE, conf.int = FALSE)
plot1$plot <- plot1$plot + guides(colour = guide_legend(nrow = 3))
plot2 <- ggsurvplot(ekm3, pval = TRUE, conf.int = FALSE)
plot2$plot <- plot2$plot + guides(colour = guide_legend(nrow = 3))

arrange_ggsurvplots(list(plot1, plot2))
```



```
fit <- coxph(Surv(Tempo, Obito) ~ ., data = dados)
summary(fit)
```

```
## Call:
## coxph(formula = Surv(Tempo, Obito) ~ ., data = dados)
##
##      n= 365, number of events= 54
##      (635 observations deleted due to missingness)
##
##              coef  exp(coef)   se(coef)      z Pr(>|z|)    
## SexoFeminino -9.202e-02 9.121e-01 3.340e-01 -0.276 0.78292  
## Idade         3.702e-02 1.038e+00 1.288e-02  2.874 0.00405 ** 
## Classe_Funcional2 8.609e-01 2.365e+00 4.801e-01  1.793 0.07293 .  
## Classe_Funcional3 1.551e-01 1.168e+00 3.246e-01  0.478 0.63266  
## Classe_Funcional4 -1.271e+01 3.017e-06 2.833e+03 -0.004 0.99642  
## IAMSim        1.813e+00 6.126e+00 5.740e-01  3.158 0.00159 ** 
## RochaganSim   -9.276e-01 3.955e-01 4.355e-01 -2.130 0.03317 *  
## Beta_BloqueadorSim -4.706e-01 6.246e-01 3.353e-01 -1.404 0.16039  
## Bnp           3.380e-05 1.000e+00 3.226e-05  1.048 0.29465  
## Anticorpos    -2.638e-02 9.740e-01 6.522e-02 -0.405 0.68583  
## ECG_ResultMaiores 6.826e-01 1.979e+00 4.279e-01  1.595 0.11065  
## FE            -7.791e-02 9.250e-01 1.093e-02 -7.129 1.01e-12 *** 
## ---                                                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
##          exp(coef)  exp(-coef) lower .95 upper .95    
## SexoFeminino 9.121e-01 1.096e+00  0.4740   1.7552  
## Idade        1.038e+00 9.637e-01  1.0118   1.0643
```

```

## Classe_Funcional2 2.365e+00 4.228e-01 0.9231 6.0605
## Classe_Funcional3 1.168e+00 8.563e-01 0.6182 2.2062
## Classe_Funcional4 3.017e-06 3.315e+05 0.0000 Inf
## IAMSim 6.126e+00 1.632e-01 1.9887 18.8699
## RochaganSim 3.955e-01 2.528e+00 0.1685 0.9286
## Beta_BloqueadorSim 6.246e-01 1.601e+00 0.3238 1.2050
## Bnp 1.000e+00 1.000e+00 1.0000 1.0001
## Anticorpos 9.740e-01 1.027e+00 0.8571 1.1068
## ECG_ResultMaiores 1.979e+00 5.053e-01 0.8555 4.5776
## FE 9.250e-01 1.081e+00 0.9054 0.9451
##
## Concordance= 0.845 (se = 0.028 )
## Likelihood ratio test= 111 on 12 df, p=<2e-16
## Wald test = 87.54 on 12 df, p=1e-13
## Score (logrank) test = 178.2 on 12 df, p=<2e-16
fit2 <- coxph(Surv(Tempo, Obito) ~ Idade + IAM + Rochagan + FE, data = dados)
summary(fit2)

```

```

## Call:
## coxph(formula = Surv(Tempo, Obito) ~ Idade + IAM + Rochagan +
##        FE, data = dados)
##
## n= 694, number of events= 106
## (306 observations deleted due to missingness)
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)
## Idade      0.044493  1.045498  0.008547  5.206 1.93e-07 ***
## IAMSim     0.658773  1.932421  0.313676  2.100  0.0357 *
## RochaganSim -0.604738  0.546218  0.290257 -2.083  0.0372 *
## FE         -0.067384  0.934836  0.006019 -11.195 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## Idade       1.0455    0.9565   1.0281   1.0632
## IAMSim      1.9324    0.5175   1.0450   3.5736
## RochaganSim 0.5462    1.8308   0.3092   0.9648
## FE          0.9348    1.0697   0.9239   0.9459
##
## Concordance= 0.806 (se = 0.022 )
## Likelihood ratio test= 140.2 on 4 df, p=<2e-16
## Wald test = 154.8 on 4 df, p=<2e-16
## Score (logrank) test = 190.9 on 4 df, p=<2e-16
fit3 <- coxph(Surv(Tempo, Obito) ~ Idade + FE, data = dados)
summary(fit3)

```

```

## Call:
## coxph(formula = Surv(Tempo, Obito) ~ Idade + FE, data = dados)
##
## n= 694, number of events= 106
## (306 observations deleted due to missingness)
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)

```

```

## Idade  0.049049  1.050272  0.008488   5.779 7.53e-09 ***
## FE     -0.068058  0.934206  0.005893 -11.550 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## Idade     1.0503     0.9521    1.0329    1.0679
## FE        0.9342     1.0704    0.9235    0.9451
##
## Concordance= 0.801 (se = 0.022 )
## Likelihood ratio test= 130.9 on 2 df,  p=<2e-16
## Wald test          = 147.1 on 2 df,  p=<2e-16
## Score (logrank) test = 178.9 on 2 df,  p=<2e-16

```

Os resultados dos modelos de Cox indicam que as covariáveis Idade e FE são as mais estatisticamente significativas, conforme já analisadas anteriormente. Abaixo temos uma ANOVA para testar a diferença entre os modelos fit2 e fit3:

```
anova(fit2,fit3)
```

```

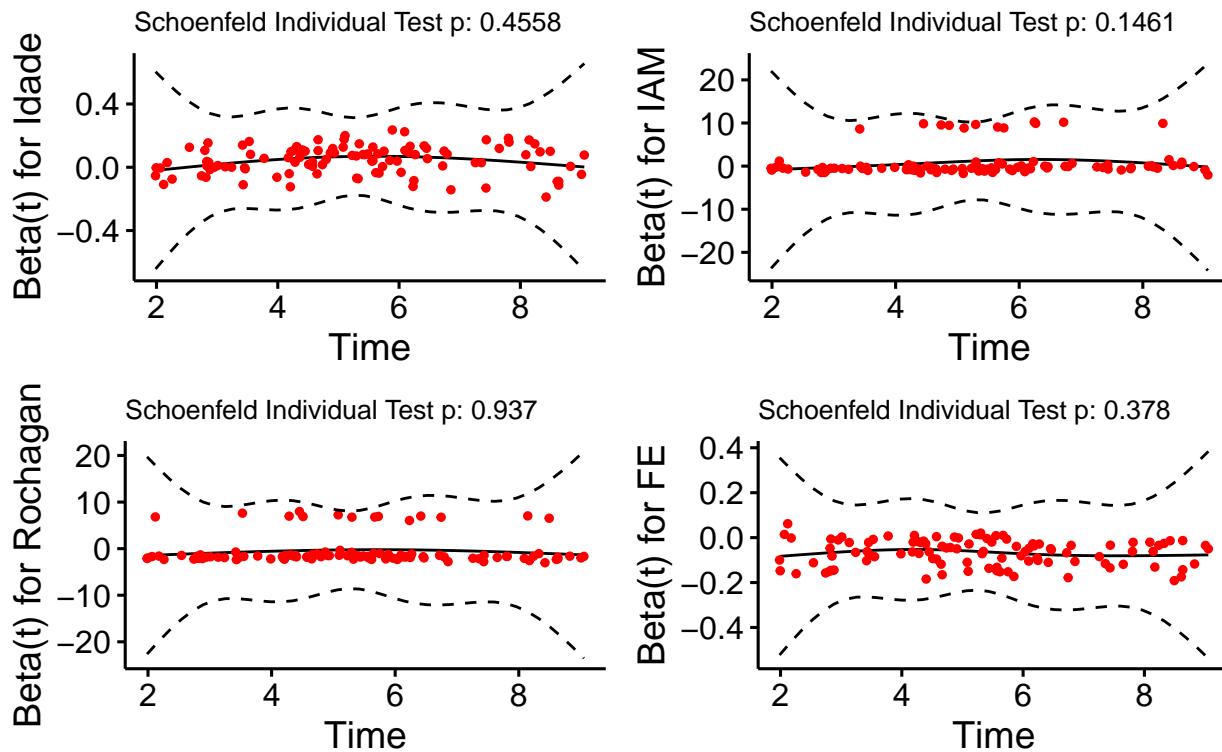
## Analysis of Deviance Table
## Cox model: response is Surv(Tempo, Obito)
## Model 1: ~ Idade + IAM + Rochagan + FE
## Model 2: ~ Idade + FE
##      loglik  Chisq Df Pr(>|Chi|)
## 1 -579.91
## 2 -584.59 9.3645  2  0.009258 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

A análise de deviança entre os modelos fit2 (Idade + IAM + Rochagan + FE) e fit3 (Idade + FE) indica que a inclusão das variáveis IAM e Rochagan no modelo fit2 melhora significativamente o ajuste do modelo. O teste de qui-quadrado (Chisq = 9.3645, Df = 2) resulta em um valor-p de 0.009258, sugerindo que a diferença entre os modelos é estatisticamente significativa ao nível de 1%. Isso significa que IAM e Rochagan contribuem significativamente para o modelo de sobrevivência, melhorando a predição do risco de óbito quando comparado a um modelo que inclui apenas Idade e FE.

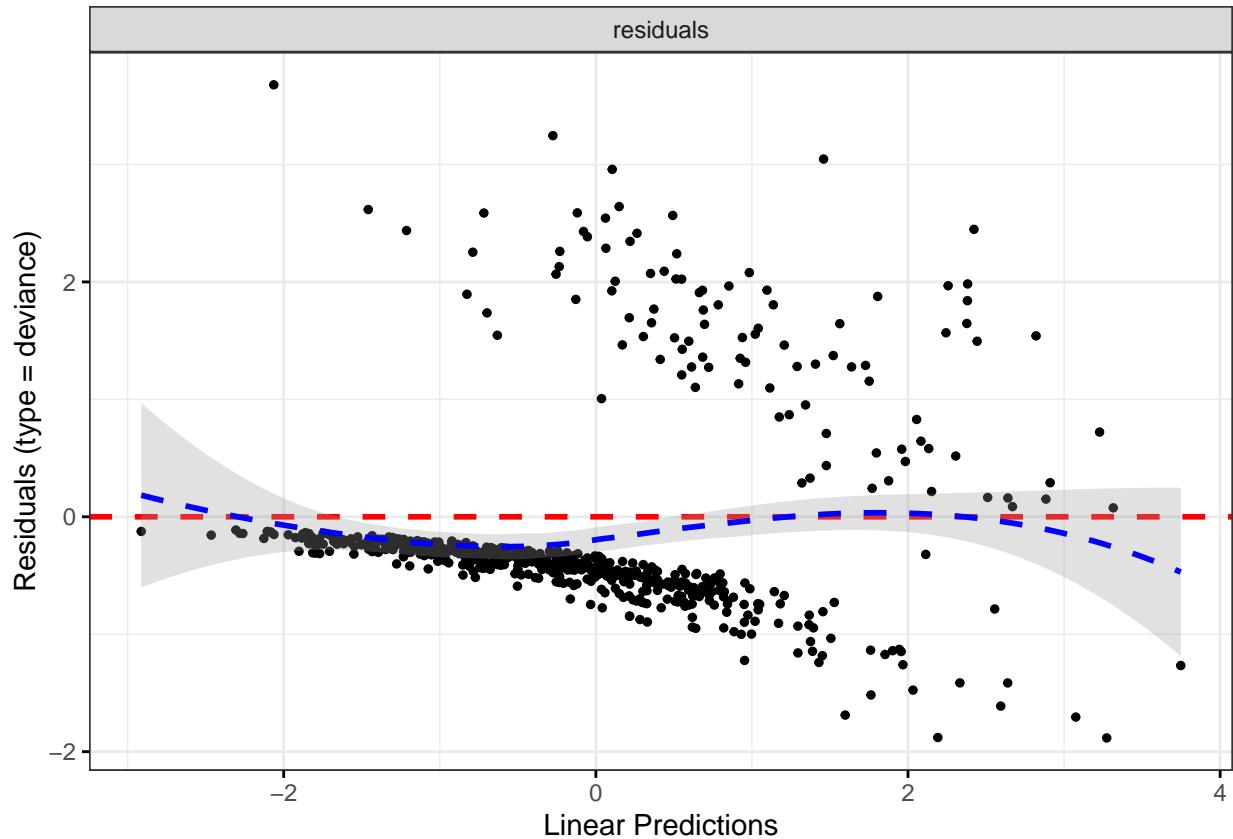
```
zph <- cox.zph(fit2, transform = 'identity')
ggcoxzph(zph, font.main = 10)
```

Global Schoenfeld Test p: 0.4893



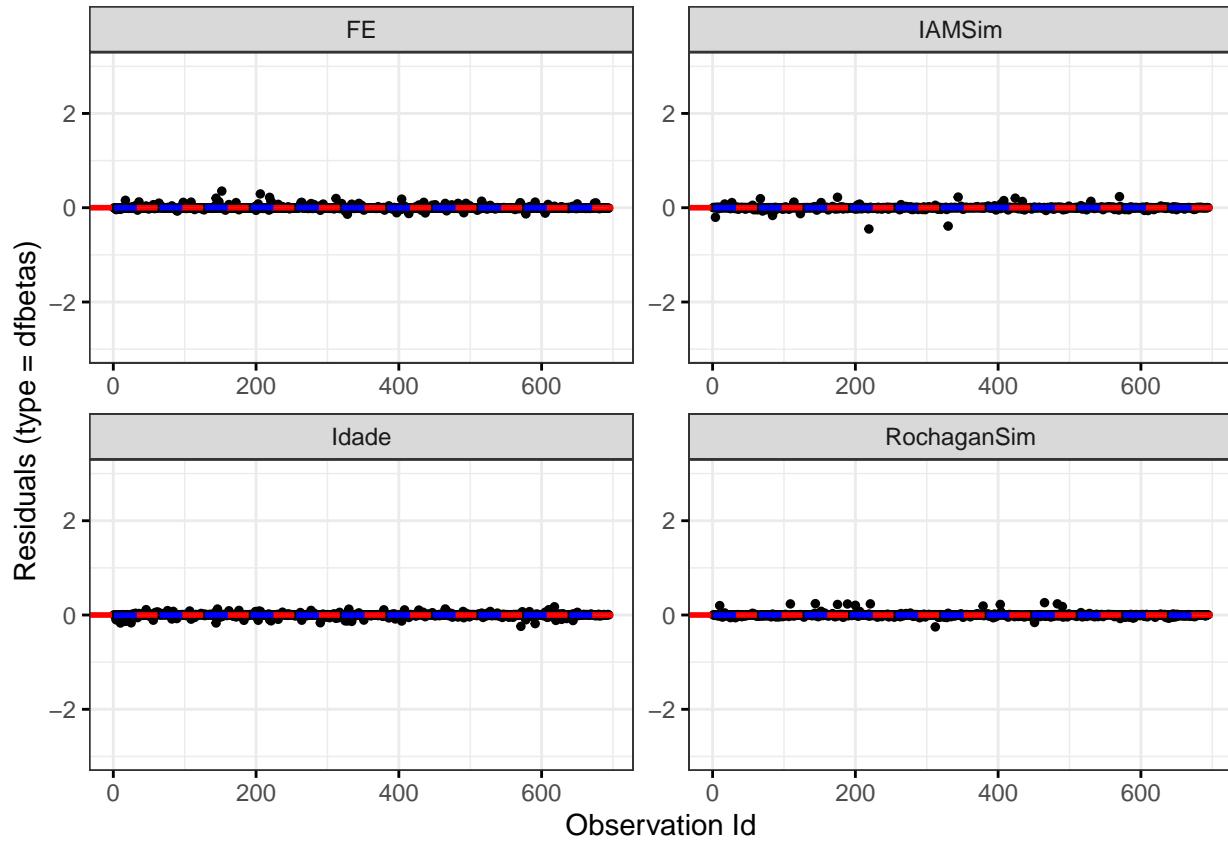
Os resultados do teste de proporcionalidade de Schoenfeld aplicado ao modelo de Cox fit2, que inclui as variáveis Idade, IAM, Rochagan e FE, apresenta um p-valor de 0.4893, indicando que não há evidências significativas de violação da suposição de proporcionalidade dos riscos para o modelo como um todo. Analisando os gráficos individuais, os p-valores para cada variável (Idade: 0.4558, IAM: 0.1461, Rochagan: 0.937, e FE: 0.378) também não são significativos, reforçando que cada uma dessas variáveis atende à suposição de proporcionalidade dos riscos. As linhas horizontais e a distribuição dos resíduos (pontos vermelhos) ao longo do tempo confirmam a adequação do modelo de Cox para essas variáveis.

```
ggcoxdiagnostics(fit2, type = "deviance")
```



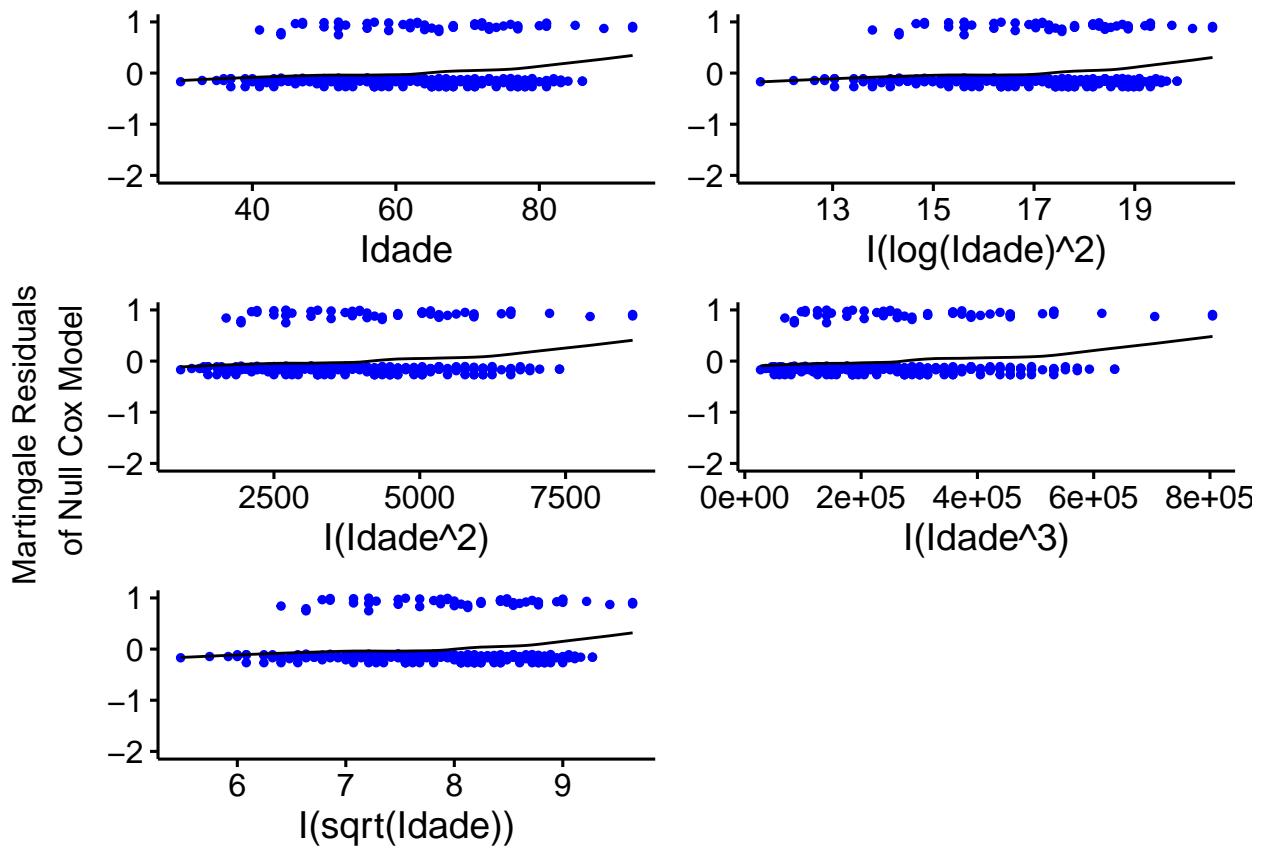
O gráfico de diagnósticos de Cox para os resíduos do desvio do modelo fit2 mostram desvios consideráveis e consistentes para muitas observações.

```
ggcoxdiagnostics(fit2, type = "dfbetas", ox.scale = "observation.id", ylim = c(-3, 3))
```

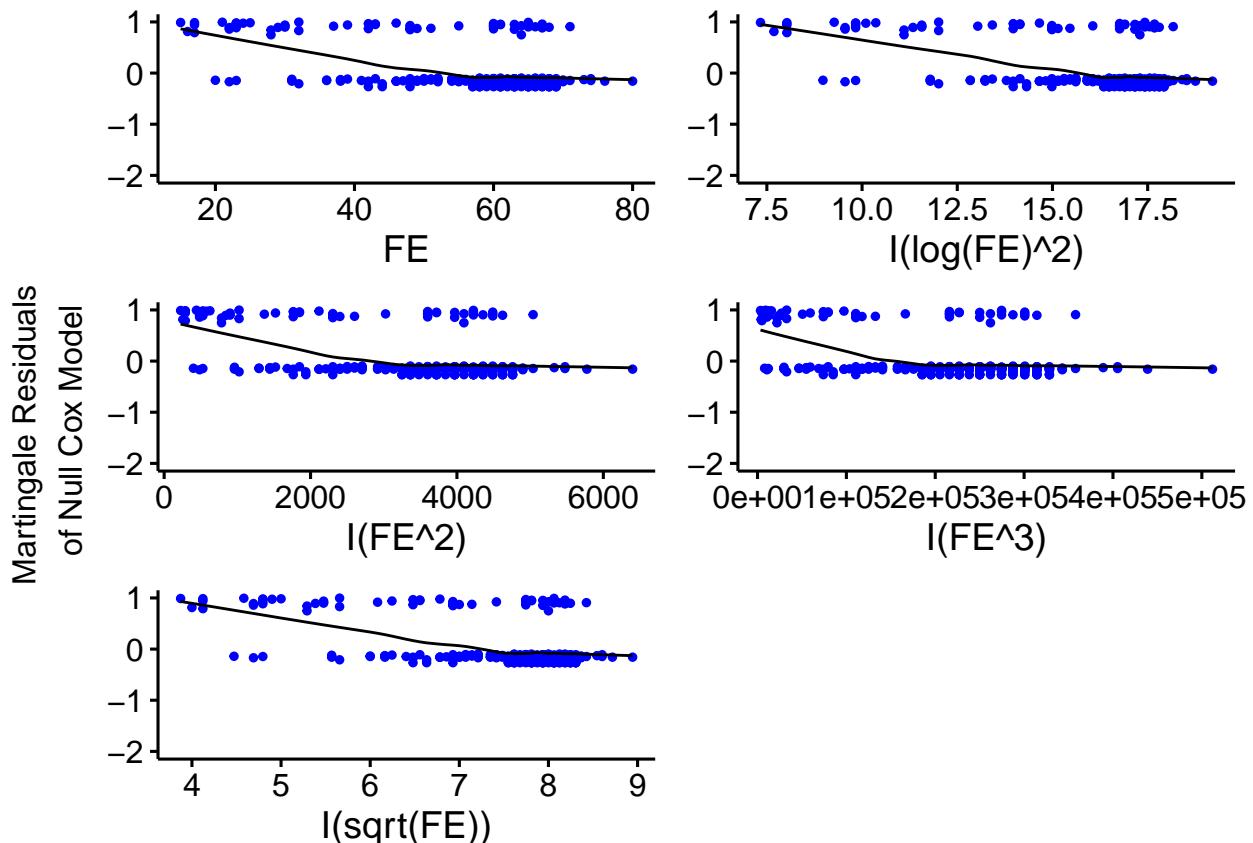


O gráfico de diagnósticos de Cox para os pontos influentes nos indica a influência de cada observação nas estimativas dos coeficientes. A maioria dos resíduos está próxima de zero, sugerindo que poucas observações têm grande influência nos coeficientes. As linhas de suavização são quase retas e próximas de zero, indicando que a assunção de linearidade é razoável. Não há muitos outliers evidentes, sugerindo que o modelo é robusto e não excessivamente influenciado por poucas observações específicas.

```
res.cox <- coxph(Surv(Tempo, Obito) ~ Idade + I(log(Idade)^2) + I(Idade^2) + I(Idade^3)
+ I(sqrt(Idade)), data = dados)
df <- na.omit(dados, cols = "Idade")
ggcoxfunctorial(res.cox, point.col = "blue", data = df, ylim = c(-2,1))
```



```
res.cox <- coxph(Surv(Tempo, Obito) ~ FE + I(log(FE)^2) + I(FE^2) + I(FE^3)
+ I(sqrt(FE)), data = dados)
df <- na.omit(dados, cols = "FE")
ggcoxfunctional(res.cox, point.col = "blue", data = df, ylim = c(-2,1))
```



Os gráficos mostram resíduos de Martingale para diferentes transformações das variáveis Idade e FE em um modelo de Cox. Para Idade, os resíduos não indicam uma relação linear clara, com resíduos levemente inclinados em todas as transformações, sugerindo que nenhuma transformação captura perfeitamente a relação com o tempo de sobrevivência. Para FE, resíduos também mostram padrões não lineares, com tendências de inclinação negativa, indicando que a relação entre FE e o tempo de sobrevivência pode ser mais complexa e talvez precise de uma transformação diferente ou um modelo não linear para melhor ajuste. Esses gráficos ajudam a diagnosticar a adequação das transformações usadas e sugerem que ajustes adicionais podem ser necessários para capturar corretamente as relações dessas variáveis com o desfecho.

d) Avalie a qualidade do ajuste de cada modelo e verifique sua adequação. Investigue a necessidade de transformação nos dados e inclusão de termos de interação.

Os gráficos apresentados indicam que o ajuste do último modelo não parece se adequar bem aos dados. No gráfico dos resíduos de desvio, é possível observar um padrão nos resíduos que não deveria estar presente se o modelo estivesse bem ajustado. A linha azul suavizada mostra uma tendência, indicando que os resíduos não estão distribuídos aleatoriamente em torno de zero, sugerindo que o modelo pode estar omitindo alguma variável ou relação importante.

Além disso, as transformações aplicadas na variável Idade, mostradas nos gráficos de resíduos martingale, não parecem ter melhorado a situação. Os pontos dos resíduos não estão dispostos uniformemente em torno de zero, o que sugere que a relação entre Idade e o risco não é linear ou que outra forma de transformação ou ajuste é necessária. Isso pode indicar que outras variáveis ou interações estão influenciando os resultados, e que o modelo atual não está capturando essas nuances.

Outro ponto que pode estar afetando a análise é a grande quantidade de valores ausentes nas variáveis.

```
vars.cox <- dados %>% select(FE, IAM, Idade, Rochagan)
missings_proporcao <- dados %>%
  summarise_all(~ mean(is.na(.))) %>%
```

```

gather(key = "Variável", value = "Proporção de Dados Faltantes") %>%
arrange(desc(`Proporção de Dados Faltantes`))
print(missings_proporcao)

## # A tibble: 12 x 2
##   Variável      `Proporção de Dados Faltantes`
##   <chr>          <dbl>
## 1 Bnp            0.599
## 2 FE             0.302
## 3 Idade          0.205
## 4 ECG_Result    0.027
## 5 Anticorpos    0.016
## 6 Beta_Bloqueador 0.013
## 7 Classe_Funcional 0.008
## 8 Tempo           0.007
## 9 IAM             0.001
## 10 Óbito          0
## 11 Sexo            0
## 12 Rochagan       0

```

A quantidade de missings pode prejudicar a precisão e a confiabilidade dos resultados. Portanto, a análise continuará agora considerando a possibilidade de imputação desses dados ausentes para melhorar a robustez e a validade dos resultados. Para isso, vamos então conferir a curva de sobrevivência estimada tratando essa ausência de dados como sendo uma estratificação, para verificar se pode trazer alguma informação adicional.

```

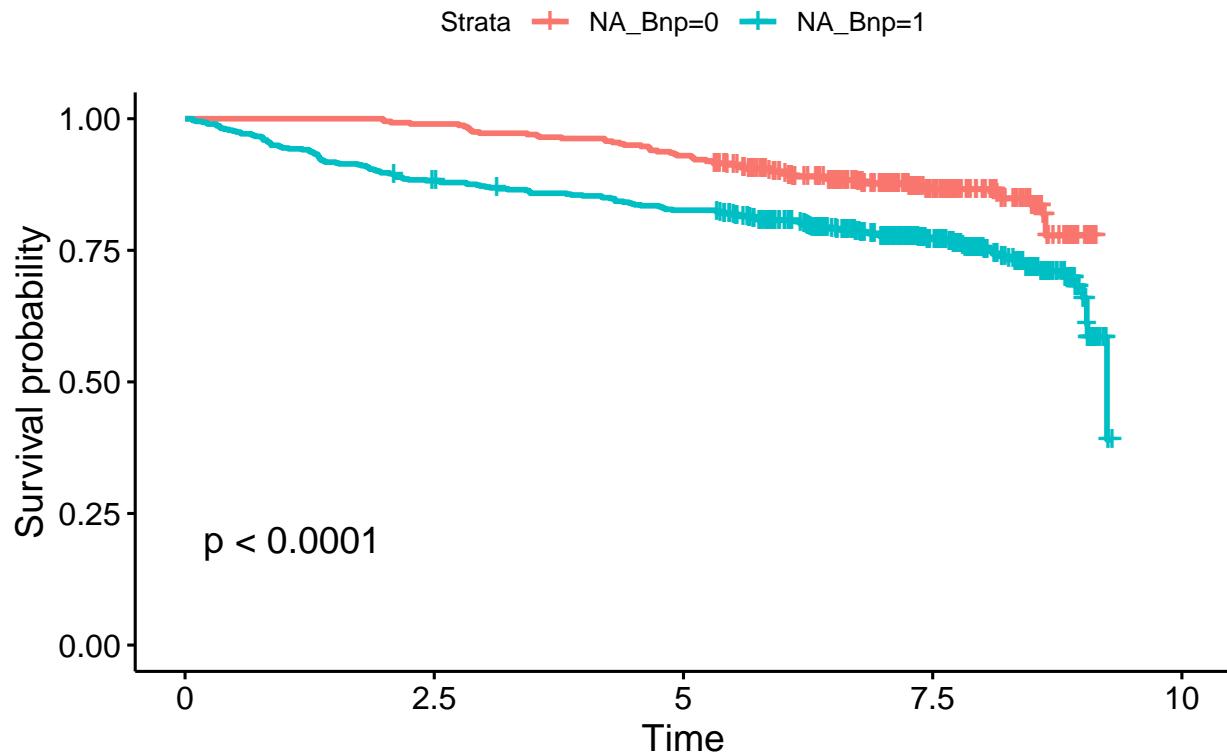
dados <- dados %>%
  mutate(NA_Bnp = ifelse(is.na(Bnp), 1, 0),
        NA_FE = ifelse(is.na(FE), 1, 0),
        NA_Idade = ifelse(is.na(Idade), 1, 0))

surv_object <- Surv(dados$Tempo, dados$Óbito)

fit_Bnp <- survfit(surv_object ~ NA_Bnp, data = dados)
ggsurvplot(fit_Bnp, data = dados, pval = TRUE,
            title = "Curva de Sobrevivência Estratificada por NA_Bnp")

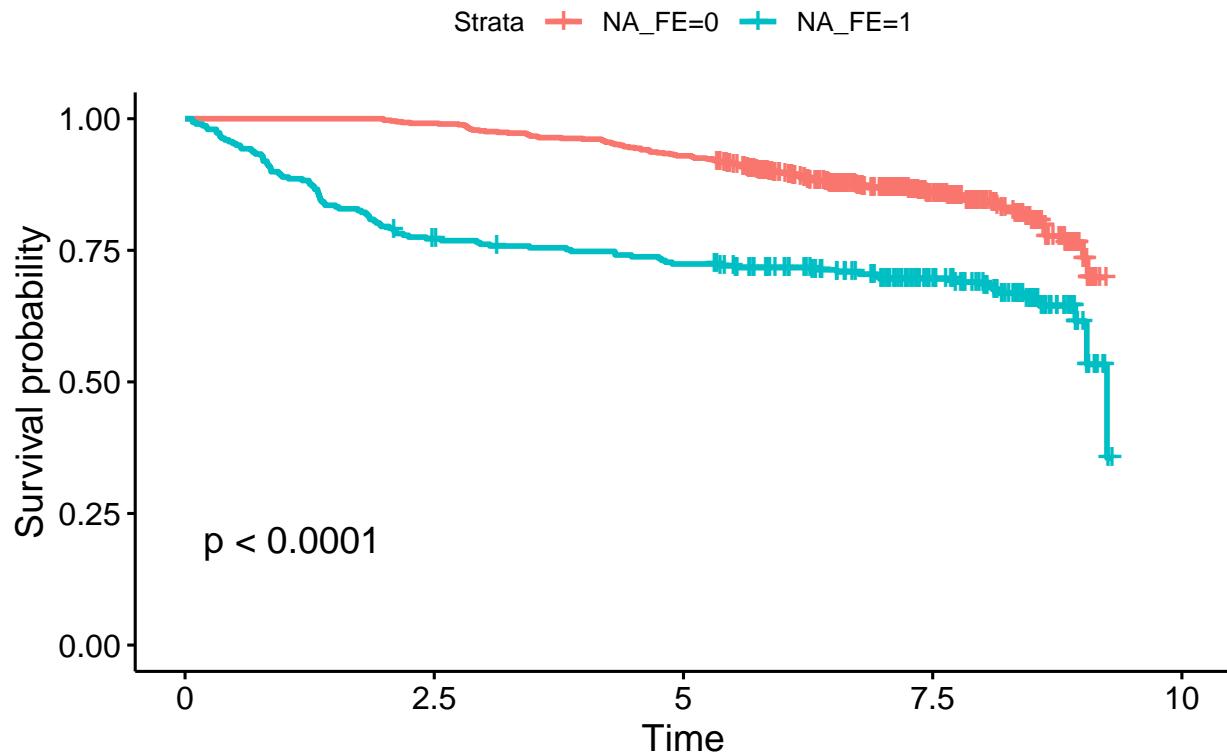
```

Curva de Sobrevida Estratificada por NA_Bnp



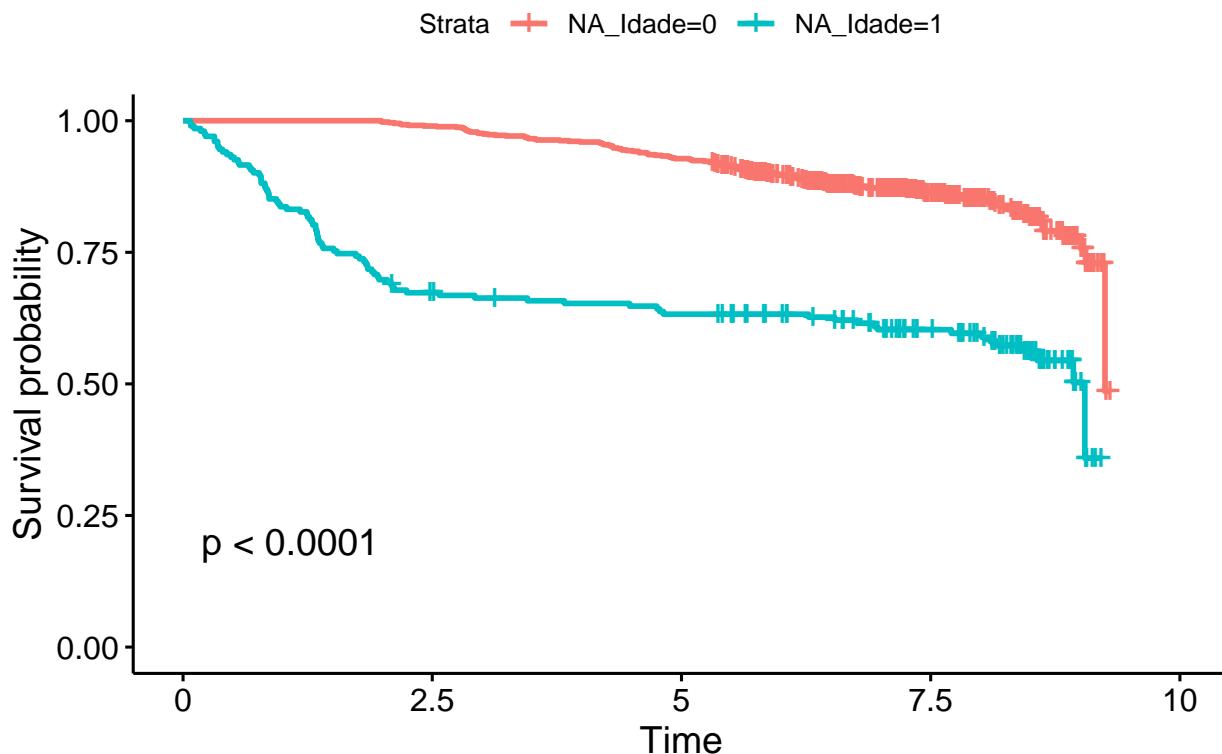
```
fit_FE <- survfit(surv_object ~ NA_FE, data = dados)
ggsurvplot(fit_FE, data = dados, pval = TRUE,
           title = "Curva de Sobrevida Estratificada por NA_FE")
```

Curva de Sobrevida Estratificada por NA_FE



```
fit_Idade <- survfit(surv_object ~ NA_Idade, data = dados)
ggsurvplot(fit_Idade, data = dados, pval = TRUE,
           title = "Curva de Sobrevida Estratificada por NA_Idade")
```

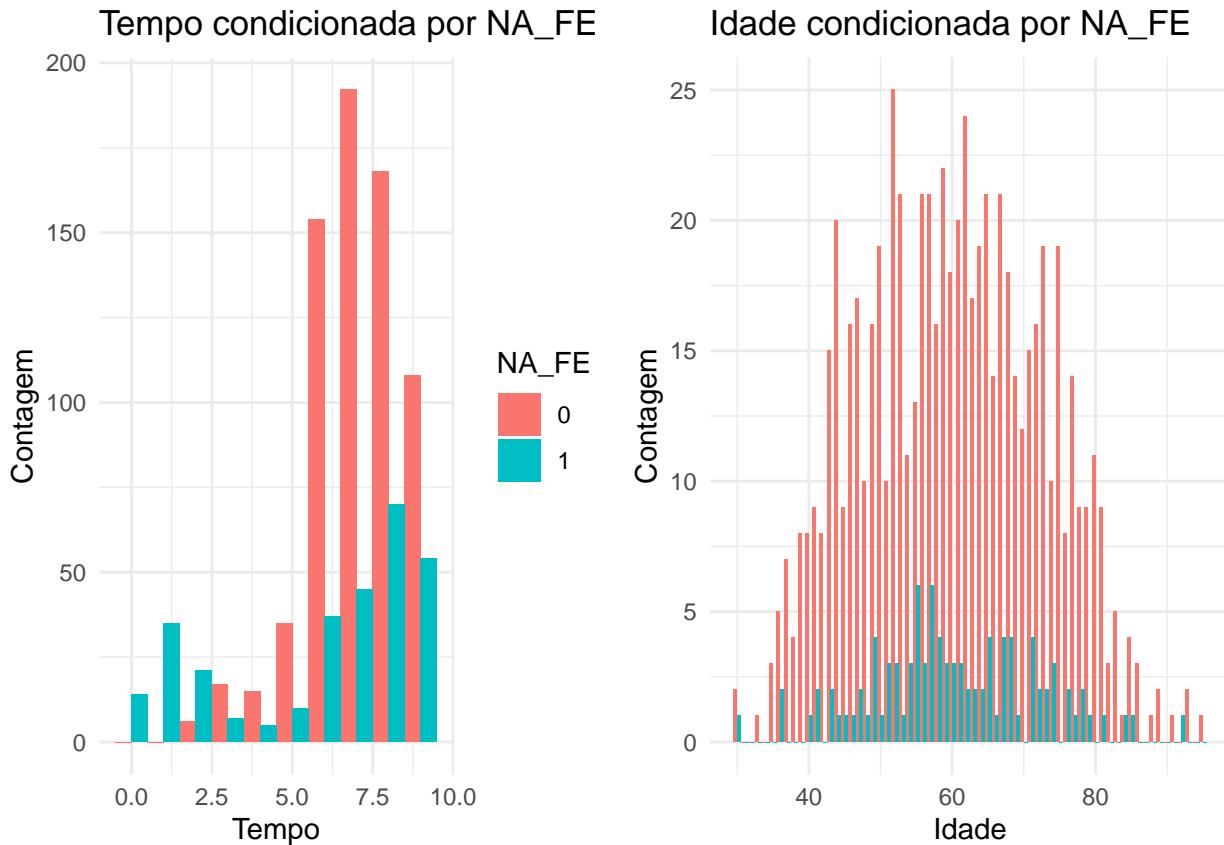
Curva de Sobrevida Estratificada por NA_Idade



Os gráficos das curvas de sobrevida estratificadas pelas variáveis binárias indicadoras de dados faltantes revelam que, consistentemente, as curvas que representam os dados missing estão abaixo das curvas correspondentes a dados não missing. Além disso, o teste log-rank aplicado em cada estratificação indica diferenças significativas nos riscos de sobrevida, sugerindo que a ausência de dados em Bnp, FE, e Idade está associada a um pior prognóstico para os pacientes.

```
variaveis_continuas <- c("Tempo", "Idade")
grafico_lista <- list()

for (variavel in variaveis_continuas) {
  p <- ggplot(dados, aes_string(x = variavel)) +
    geom_histogram(aes(fill = as.factor(NA_FE)), binwidth = 1, position = "dodge") +
    labs(title = paste(variavel, "condicionada por NA_FE"),
         x = variavel, y = "Contagem", fill = "NA_FE") +
    theme_minimal() +
    theme(legend.position = ifelse(variavel == "Idade", "none", "right"))
  grafico_lista[[variavel]] <- p}
do.call(grid.arrange, c(grafico_lista, ncol = 2))
```



Pela análise direta do gráfico acima podemos perceber que a variável Tempo parece se comportar de maneira diferente condicionada à variável indicadora de FE faltante, enquanto Idade não aparece ser diversa. Porém, iremos verificar isso a nível estatístico com o teste abaixo.

```
ks.test(dados %>% filter(NA_FE == 0) %>% pull(Tempo),
        dados %>% filter(NA_FE == 1) %>% pull(Tempo))
```

```
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  dados %>% filter(NA_FE == 0) %>% pull(Tempo) and dados %>% filter(NA_FE == 1) %>% pull(Tempo)
## D = 0.23154, p-value = 3.881e-10
## alternative hypothesis: two-sided

ks.test(dados %>% filter(NA_FE == 0) %>% pull(Idade),
        dados %>% filter(NA_FE == 1) %>% pull(Idade))

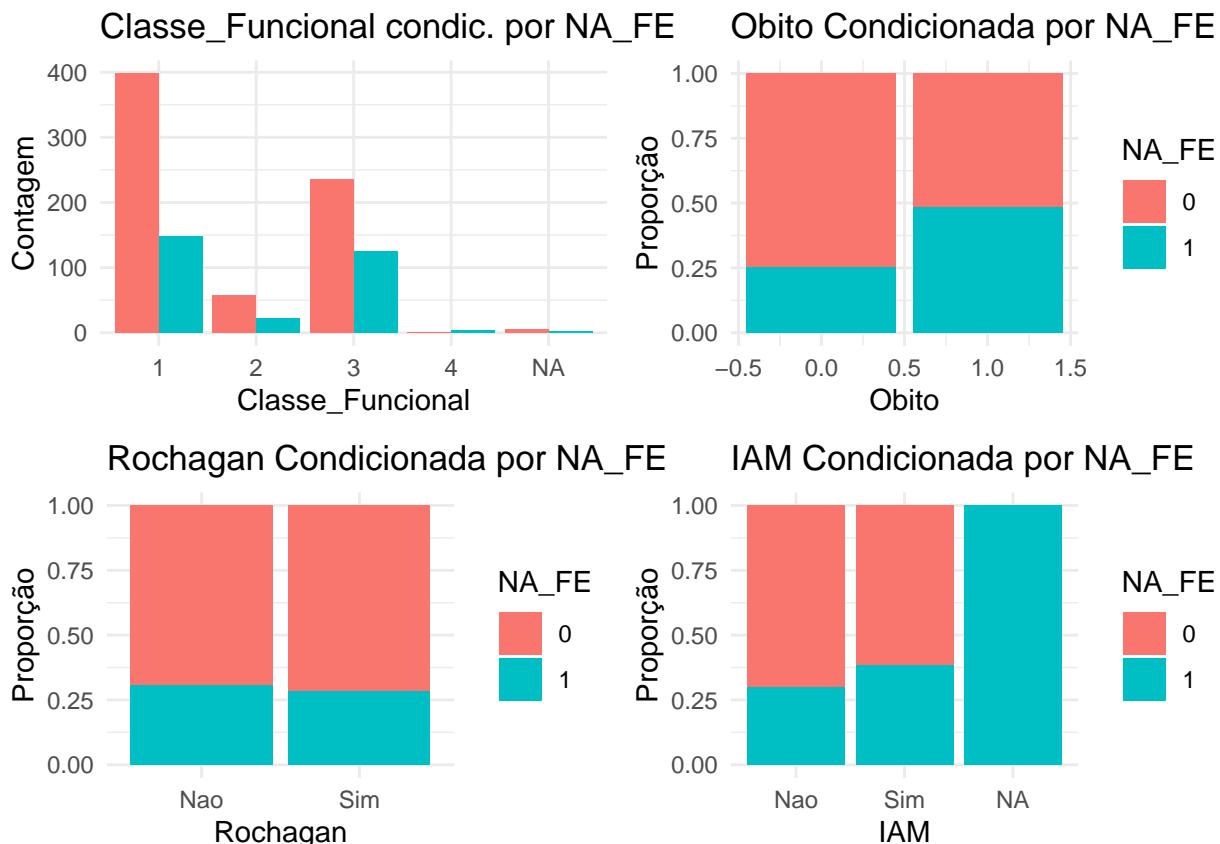
##
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  dados %>% filter(NA_FE == 0) %>% pull(Idade) and dados %>% filter(NA_FE == 1) %>% pull(Idade)
## D = 0.068984, p-value = 0.8082
## alternative hypothesis: two-sided
```

Os resultados dos testes Kolmogorov-Smirnov mostram que o Tempo tem distribuições significativamente diferentes entre dados faltantes ou presentes para FE, com um p-valor muitíssimo baixo, enquanto a distribuição de Idade é semelhante para ambos os grupos, com um p-valor de 0,80. Assim, Tempo parece sensível à variável indicadora de FE faltante, ao contrário de Idade.

```

g1 <- ggplot(dados, aes_string(x = "Classe_Funcional")) +
  geom_bar(aes(fill = as.factor(NA_FE)), position = "dodge") +
  labs(title = "Classe_Funcional condic. por NA_FE",
       x = "Classe_Funcional", y = "Contagem", fill = "NA_FE") +
  theme_minimal() + theme(legend.position = "none")
g2 <- ggplot(dados, aes(x = Obito, fill = as.factor(NA_FE))) +
  geom_bar(position = "fill") +
  labs(title = "Obito Condicionada por NA_FE",
       x = "Obito", y = "Proporção", fill = "NA_FE") +
  theme_minimal() + theme(legend.position = "right")
g3 <- ggplot(dados, aes(x = Rochagan, fill = as.factor(NA_FE))) +
  geom_bar(position = "fill") +
  labs(title = "Rochagan Condicionada por NA_FE",
       x = "Rochagan", y = "Proporção", fill = "NA_FE") +
  theme_minimal() + theme(legend.position = "right")
g4 <- ggplot(dados, aes(x = IAM, fill = as.factor(NA_FE))) +
  geom_bar(position = "fill") +
  labs(title = "IAM Condicionada por NA_FE",
       x = "IAM", y = "Proporção", fill = "NA_FE") +
  theme_minimal() + theme(legend.position = "right")
do.call(grid.arrange, c(list(g1, g2, g3, g4), ncol = 2, nrow = 2))

```



```

fisher_classe_funcional <- fisher.test(table(dados$Classe_Funcional, dados$NA_FE))
fisher_obito <- fisher.test(table(dados$Obito, dados$NA_FE))
fisher_rochagan <- fisher.test(table(dados$Rochagan, dados$NA_FE))
fisher_iam <- fisher.test(table(dados$IAM, dados$NA_FE))

```

```

resultados <- data.frame(
  Variavel = c("Classe Funcional", "Obito", "Rochagan", "IAM"),
  `p-valor` = round(c(fisher_classe_funcional$p.value,
    fisher_obito$p.value,
    fisher_rochagan$p.value,
    fisher_iam$p.value),6))
print(resultados)

##          Variavel p.valor
## 1 Classe Funcional 0.008383
## 2           Obito 0.000000
## 3      Rochagan 0.471419
## 4         IAM 0.253554

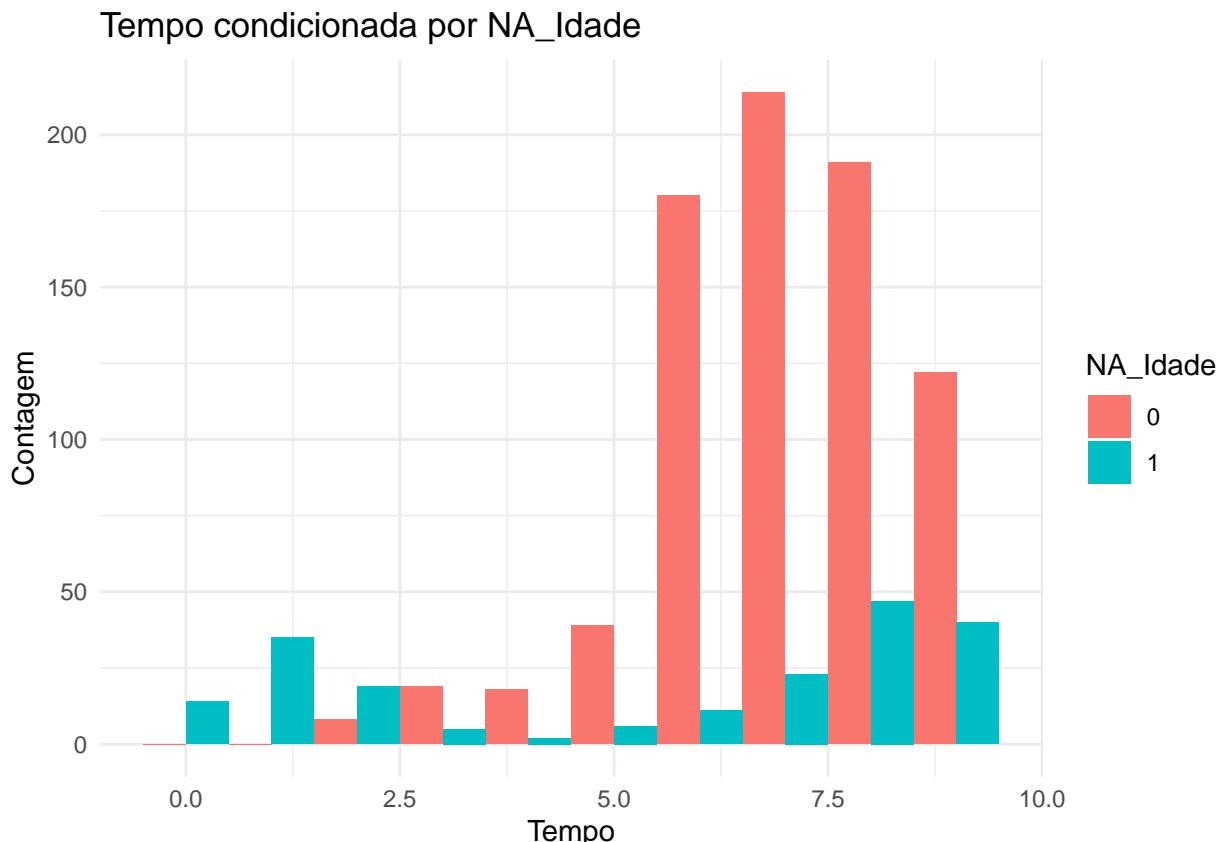
```

Os gráficos mostram a distribuição de Classe Funcional e Obito condicionada por NA_FE, evidenciando variações nas contagens entre os valores. O teste exato de Fisher confirma uma associação estatisticamente significativa tanto entre Classe_Funcional e NA_FE, com p-valor de 0.0084, quanto entre Obito e NA_FE, com p-valor de 4.31e-10, sugerindo uma relação dependente entre essas variáveis. Já para as variáveis de Rochagan e IAM os testes indicam que não há diferença da distribuição condicionada a NA_FE.

```

ggplot(dados, aes_string(x = "Tempo")) +
  geom_histogram(aes(fill = as.factor(NA_Idade)), binwidth = 1, position = "dodge") +
  labs(title = paste("Tempo condicionada por NA_Idade"),
       x = "Tempo", y = "Contagem", fill = "NA_Idade") + theme_minimal()

```



```

ks.test(dados %>% filter(NA_Idade == 0) %>% pull(Tempo),
        dados %>% filter(NA_Idade == 1) %>% pull(Tempo))

```

```

## 
##  Asymptotic two-sample Kolmogorov-Smirnov test
##
## data:  dados %>% filter(NA_Idade == 0) %>% pull(Tempo) and dados %>% filter(NA_Idade == 1) %>% pull()
## D = 0.33516, p-value = 3.994e-16
## alternative hypothesis: two-sided

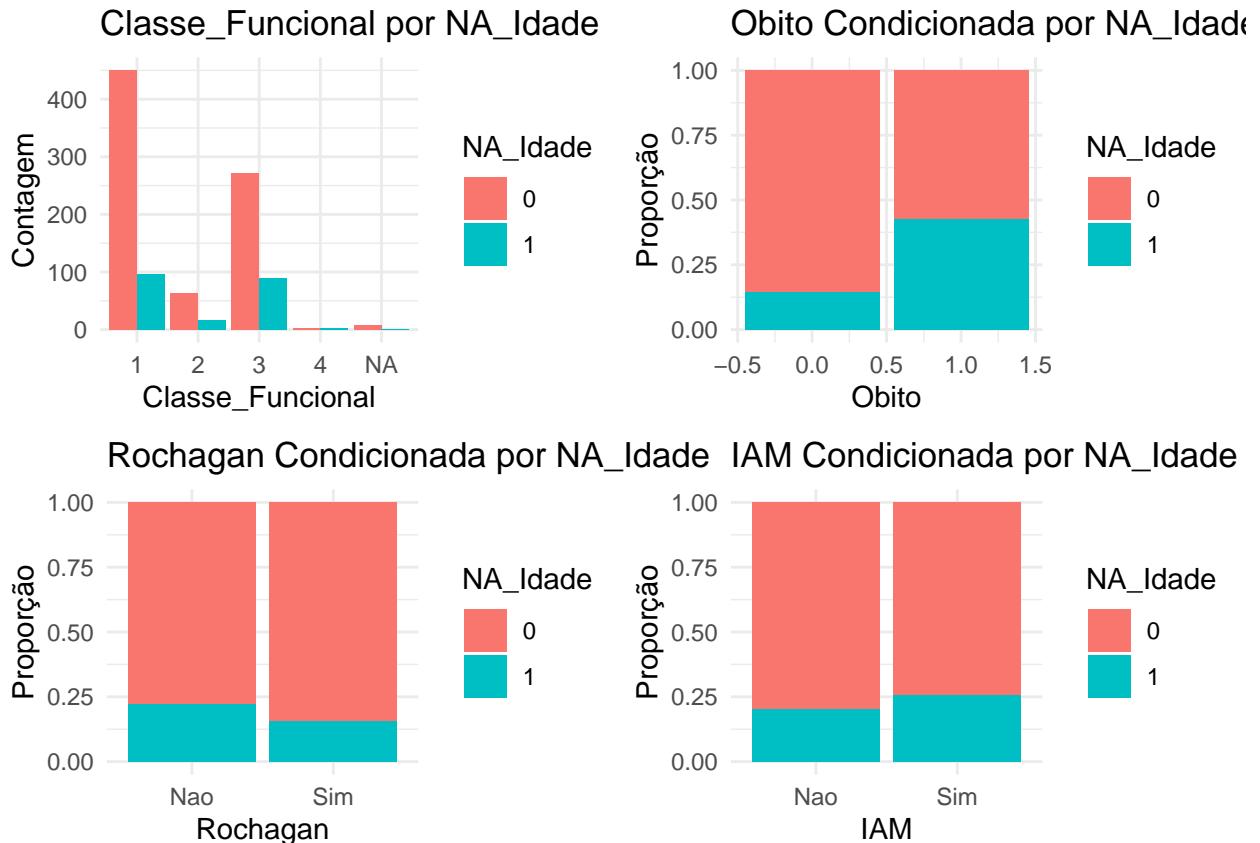
```

O resultados dos testes Kolmogorov-Smirnov mostram que o Tempo tem distribuições significativamente diferentes entre dados faltantes ou presentes para a variável idade, com um p-valor muitíssimo baixo.

```

g5 <- ggplot(dados, aes_string(x = "Classe_Funcional")) +
  geom_bar(aes(fill = as.factor(NA_Idade)), position = "dodge") +
  labs(title = "Classe_Funcional por NA_Idade",
       x = "Classe_Funcional", y = "Contagem", fill = "NA_Idade") + theme_minimal()
g6 <- ggplot(dados, aes(x = Obito, fill = as.factor(NA_Idade))) +
  geom_bar(position = "fill") +
  labs(title = "Obito Condicionada por NA_Idade",
       x = "Obito", y = "Proporção", fill = "NA_Idade") + theme_minimal()
g7 <- ggplot(dados, aes(x = Rochagan, fill = as.factor(NA_Idade))) +
  geom_bar(position = "fill") +
  labs(title = "Rochagan Condicionada por NA_Idade",
       x = "Rochagan", y = "Proporção", fill = "NA_Idade") +
  theme_minimal() + theme(legend.position = "right")
g8 <- ggplot(dados %>% filter(!is.na(IAM)) , aes(x = IAM, fill = as.factor(NA_Idade))) +
  geom_bar(position = "fill") +
  labs(title = "IAM Condicionada por NA_Idade",
       x = "IAM", y = "Proporção", fill = "NA_Idade") +
  theme_minimal() + theme(legend.position = "right")
do.call(grid.arrange, c(list(g5, g6, g7, g8), ncol = 2, nrow = 2))

```



```

fisher_classe_funcional <- fisher.test(table(dados$Classe_Funcional, dados$NA_Idade))
fisher_obito <- fisher.test(table(dados$Obito, dados$NA_Idade))
fisher_rochagan <- fisher.test(table(dados$Rochagan, dados$NA_Idade))
fisher_iam <- fisher.test(table(dados$IAM, dados$NA_Idade))

resultados <- data.frame(
  Variavel = c("Classe Funcional", "Obito", "Rochagan", "IAM"),
  `p-valor` = round(c(fisher_classe_funcional$p.value,
    fisher_obito$p.value,
    fisher_rochagan$p.value,
    fisher_iam$p.value),6))
print(resultados)

##          Variavel p.valor
## 1 Classe Funcional 0.009962
## 2           Obito 0.000000
## 3      Rochagan 0.028727
## 4          IAM 0.357773

```

Os resultados dos testes de Fisher indicam que há uma associação significativa entre a variável indicadora de idade faltante NA_Idade e as variáveis Classe Funcional, Obito, e Rochagan, com p-valores de 0.009962, 0.000000, e 0.028727, respectivamente. Isso sugere que a ausência de informação sobre a idade está relacionada com a distribuição dessas variáveis. Em contraste, não foi encontrada associação significativa com a variável IAM, cujo p-valor foi 0.357773, indicando que a distribuição de “IAM” não parece ser influenciada pela ausência de dados sobre a idade.

Dado esses resultados, concluímos que as covariáveis explicativas do modelo fit2, IAM e Rochagan, não apresentam diferenças de distribuições para os dados faltantes ou presentes das outras covariáveis explicativas,

Idade e FE. Portanto, o modelo não necessita de interações.

```
fit2 <- coxph(Surv(Tempo, Óbito) ~ IAM + Rochagan + Idade + FE , data = dados)
summary(fit2)

## Call:
## coxph(formula = Surv(Tempo, Óbito) ~ IAM + Rochagan + Idade +
##        FE, data = dados)
##
##     n= 694, number of events= 106
##     (306 observations deleted due to missingness)
##
##              coef exp(coef)   se(coef)      z Pr(>|z|)
## IAMSim      0.658773  1.932421  0.313676   2.100  0.0357 *
## RochaganSim -0.604738  0.546218  0.290257  -2.083  0.0372 *
## Idade       0.044493  1.045498  0.008547   5.206 1.93e-07 ***
## FE          -0.067384  0.934836  0.006019 -11.195 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## IAMSim      1.9324    0.5175   1.0450   3.5736
## RochaganSim 0.5462    1.8308   0.3092   0.9648
## Idade       1.0455    0.9565   1.0281   1.0632
## FE          0.9348    1.0697   0.9239   0.9459
##
## Concordance= 0.806  (se = 0.022 )
## Likelihood ratio test= 140.2 on 4 df,  p=<2e-16
## Wald test           = 154.8 on 4 df,  p=<2e-16
## Score (logrank) test = 190.9 on 4 df,  p=<2e-16
```

e) Interprete os resultados e discuta as conclusões obtidas.

Interpretação dos Resultados:

1. IAM: A presença de IAM (“IAMSIM”) está associada a um aumento significativo no risco de óbito, com um coeficiente de 0.6588 e um valor p de 0.0357. O valor de $\exp(\text{coef}) = 1.9324$ sugere que indivíduos com IAM têm aproximadamente 1.93 vezes mais chance de óbito comparado a quem não teve IAM.
2. Rochagan: A variável “RochaganSim” tem um coeficiente negativo (-0.6047) e é estatisticamente significativa ($p = 0.0372$). Isso indica que a presença de Rochagan está associada a uma redução do risco de óbito, com $\exp(\text{coef}) = 0.5462$, o que sugere que indivíduos que tomaram Rochagan têm uma probabilidade de óbito aproximadamente 54.6% menor.
3. Idade: A idade apresenta uma associação positiva com o risco de óbito, com um coeficiente de 0.0445 e um valor p extremamente significativo ($p < 0.0001$). O valor de $\exp(\text{coef}) = 1.0455$ sugere que, para cada aumento de um ano na idade, há um aumento de aproximadamente 4.55% no risco de óbito.
4. FE: A variável “FE” também é altamente significativa ($p < 0.0001$), com um coeficiente negativo (-0.0674), sugerindo que valores maiores de FE estão associados a uma redução no risco de óbito. $\exp(\text{coef}) = 0.9348$ indica que para cada unidade de aumento em FE, há uma redução de cerca de 6.52% no risco de óbito.

Fatores de Risco: A análise destaca que a presença de IAM e a idade avançada são fatores de risco significativos para o óbito. Ambos estão positivamente associados ao risco, implicando que esses fatores devem ser monitorados de perto em pacientes.

Fatores de Proteção: A administração de Rochagan e valores mais altos de FE são fatores de proteção, associados a uma redução no risco de óbito.

O modelo possui uma concordância de 0.806, sugerindo um bom desempenho preditivo. Os testes de razão de verossimilhança, Wald e logrank são todos altamente significativos ($p < 0.0001$), indicando que o modelo é apropriado para os dados. Essas conclusões podem ser úteis para orientar decisões clínicas, focando na monitorização de pacientes com IAM, idades avançadas, e otimizando o tratamento com Rochagan e manejo de FE.