

Prova 2 - Parte 1

SME0821 - Análise de Sobrevivência e Confiabilidade

Sidnei Gazola Junior – N^o USP: 9378888

Exercício 1

a)

Leitura dos dados

```
InsufCardiaca <- read_delim("C:/Users/junio/Desktop/p2/InsufCardiaca.csv",
  ";", escape_double = FALSE, col_types = cols(Idade = col_integer(), sexo = col_factor(levels = c("1",
    "2")), diag = col_factor(levels = c("1",
    "2", "3", "4", "5")), frac = col_factor(levels = c("1",
    "2", "3")), ritmo = col_factor(levels = c("1",
    "2", "3", "4"))),
  trim_ws = TRUE)
```

Limpeza dos dados

Foi verificado que a variável 'diag' conta com dois valores iguais a 6, o que não confere com a descrição, então é necessário a limpeza dos dados retirando essas linhas.

```
InsufCardiaca<- subset(InsufCardiaca, InsufCardiaca[,"diag"]!=6)
```

Análise descritiva

```
st_options(lang = 'pt',footnote = NA,style = "rmarkdown",plain.ascii = FALSE )
dfSummary(InsufCardiaca,na.col = F,varnumbers = F,valid.col = F)
```

Resumo dos dados

InsufCardiaca Dimensões: 945 x 15

Duplicatas: 0

Variável	Estatísticas / Valores	Freqs (% de Válidos)	Grafo
tempo	Média (dp) : 31.1 (14.4)	691 valores distintos	:
[numeric]	mín < mediana < máx:		. . .
	0.1 < 33.3 < 67.2	
	IQE (CV) : 18.3 (0.5)	
		
censura	Mín : 0	0 : 777 (82.2%)	IIIIIIIIIIIIIIIIIIII
[numeric]	Média : 0.2	1 : 168 (17.8%)	III
	Máx : 1		

Variável	Estatísticas / Valores	Freqs (% de Válidos)	Grafo
Idade	Média (dp) : 2.7 (1.2)	0 : 52 (5.5%)	I
[integer]	mín < mediana < máx:	1 : 95 (10.1%)	II
	0 < 3 < 4	2 : 251 (26.6%)	IIII
	IQE (CV) : 2 (0.4)	3 : 262 (27.7%)	IIII
		4 : 285 (30.2%)	IIIII
sexo	1. 1	542 (57.4%)	IIIIIIIIII
[factor]	2. 2	403 (42.6%)	IIIIIII
imc	Média (dp) : 27.6 (6)	724 valores distintos	:
[numeric]	mín < mediana < máx:		::
	10.7 < 26.7 < 63.6		::
	IQE (CV) : 7.5 (0.2)		. : : :
			: : : : .
pas	Média (dp) : 143.4 (29.3)	55 valores distintos	:
[numeric]	mín < mediana < máx:		. : :
	80 < 140 < 250		: : : .
	IQE (CV) : 40 (0.2)		. : : : :
			: : : : : .
diag	1. 1	129 (13.6%)	II
[factor]	2. 2	192 (20.3%)	IIII
	3. 3	449 (47.5%)	IIIIIIII
	4. 4	104 (11.0%)	II
	5. 5	71 (7.5%)	I
class	Média (dp) : 2.7 (1.5)	1 : 140 (14.8%)	II
[numeric]	mín < mediana < máx:	2 : 372 (39.4%)	IIIIII
	1 < 2 < 9	3 : 247 (26.1%)	IIII
	IQE (CV) : 1 (0.6)	4 : 152 (16.1%)	III
		9 : 34 (3.6%)	
hb	Média (dp) : 14 (1.8)	99 valores distintos	:
[numeric]	mín < mediana < máx:		. :
	5.9 < 14.1 < 19.2		: : .
	IQE (CV) : 2.1 (0.1)		: : : :
			. : : : : :
trigli	Média (dp) : 137.8 (90.8)	268 valores distintos	:
[numeric]	mín < mediana < máx:		:
	28 < 110 < 931		: .
	IQE (CV) : 90 (0.7)		: :
			: : : .
frac	1. 1	429 (45.4%)	IIIIIIII
[factor]	2. 2	273 (28.9%)	IIII
	3. 3	243 (25.7%)	IIII

Variável	Estatísticas / Valores	Freqs (% de Válidos)	Grafo
ritmo [factor]	1. 1	581 (61.5%)	IIIIIIIIII
	2. 2	141 (14.9%)	II
	3. 3	162 (17.1%)	III
	4. 4	61 (6.5%)	I
droga1	Mín : 0	0 : 257 (27.2%)	IIIII
[numeric]	Média : 0.7	1 : 688 (72.8%)	IIIIIIIIII
	Máx : 1		
droga2	Mín : 0	0 : 516 (54.6%)	IIIIIIII
[numeric]	Média : 0.5	1 : 429 (45.4%)	IIIIIIII
	Máx : 1		
droga3	Mín : 0	0 : 626 (66.2%)	IIIIIIIIII
[numeric]	Média : 0.3	1 : 319 (33.8%)	IIIII
	Máx : 1		

```
#style = "grid", tmp.img.dir = "/imgt",graph.magnif = 0.5,split.cells = Inf,split.tables = Inf
```

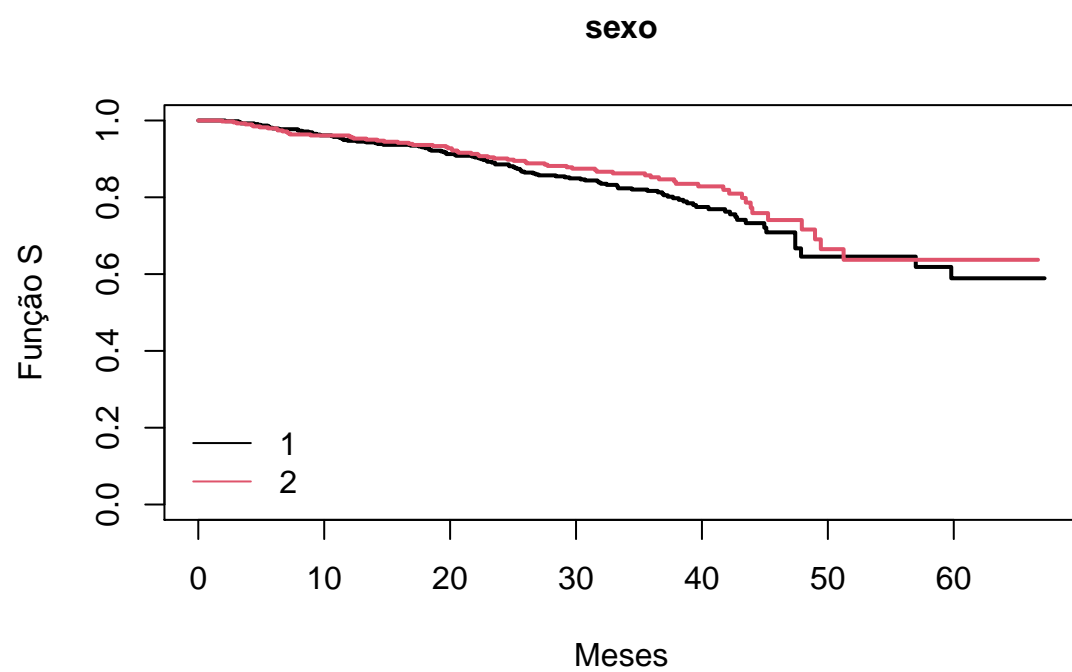
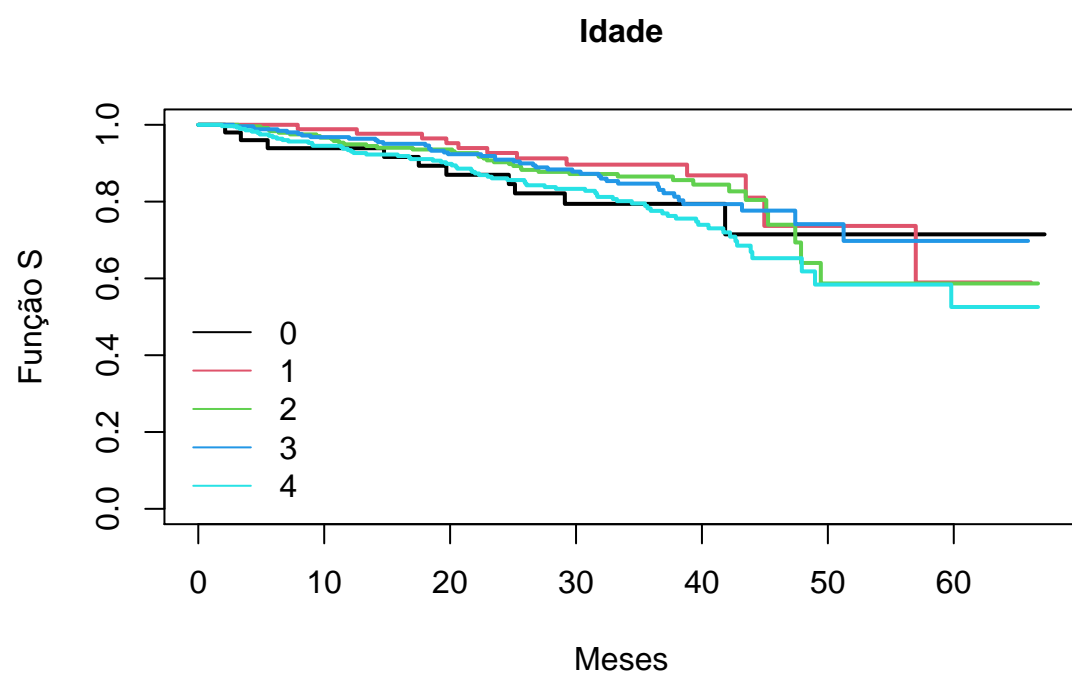
b)

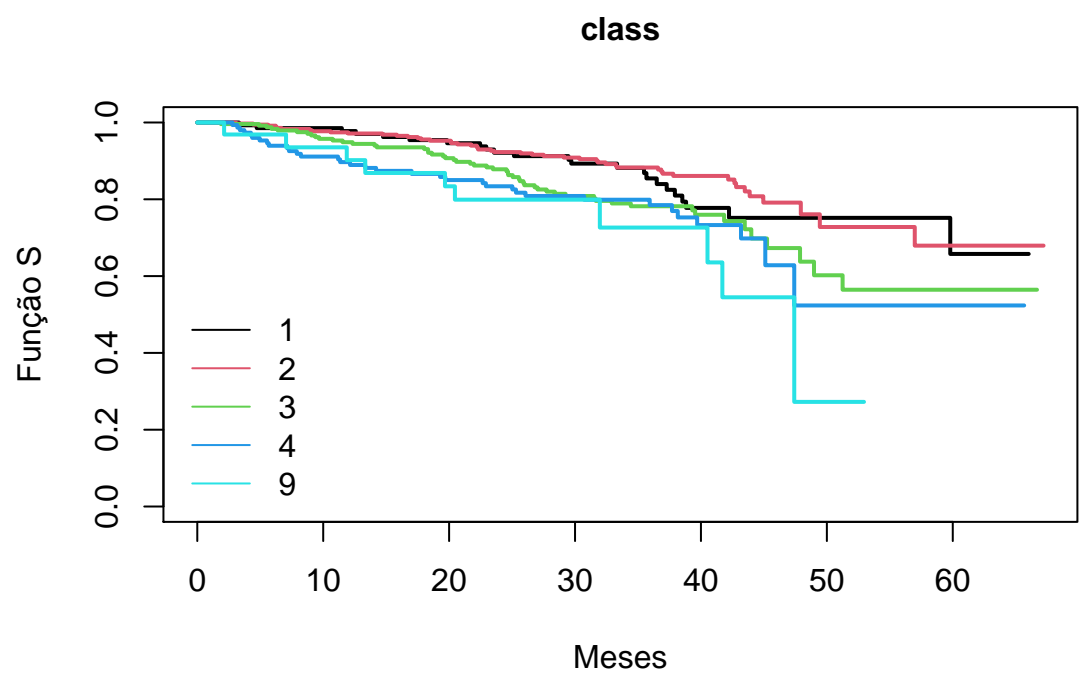
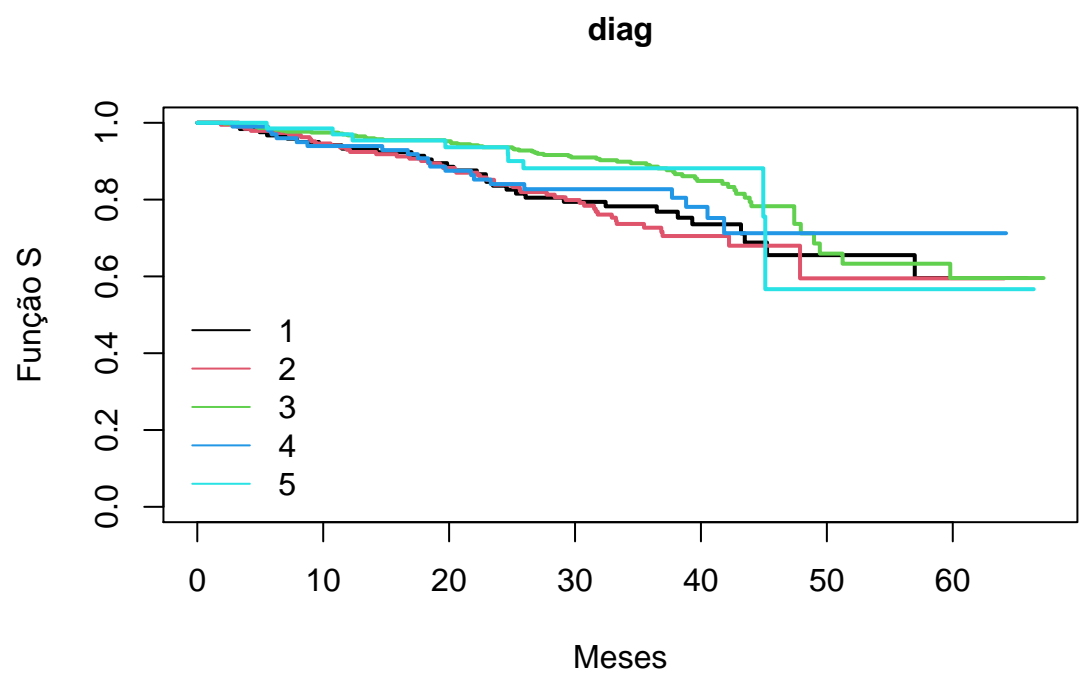
Gráficos de sobrevivência

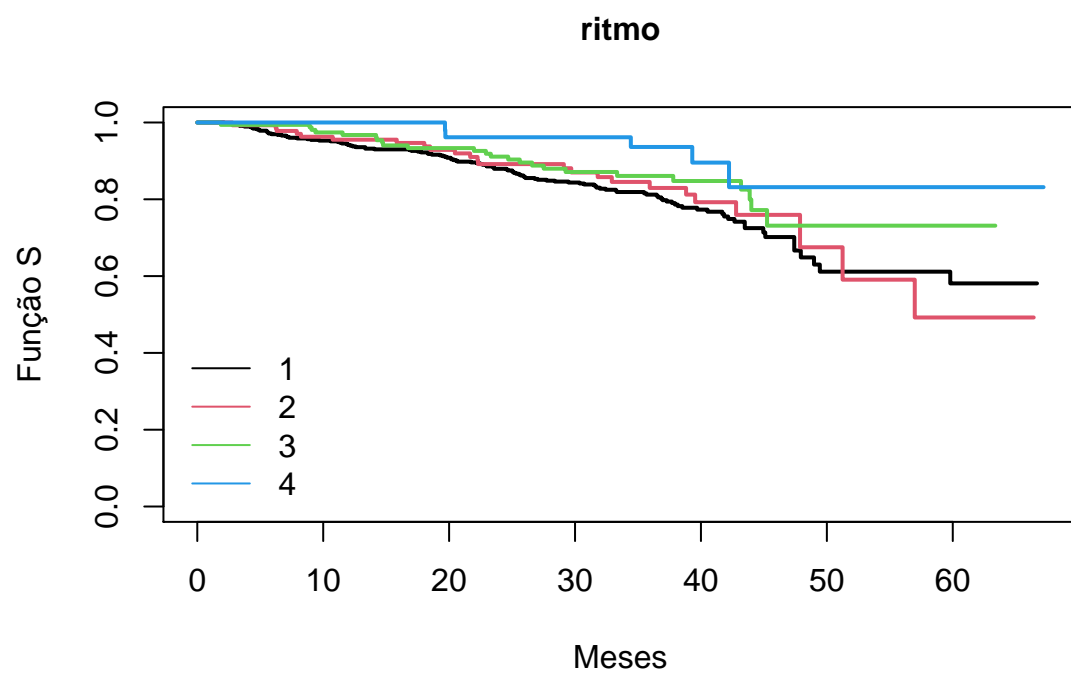
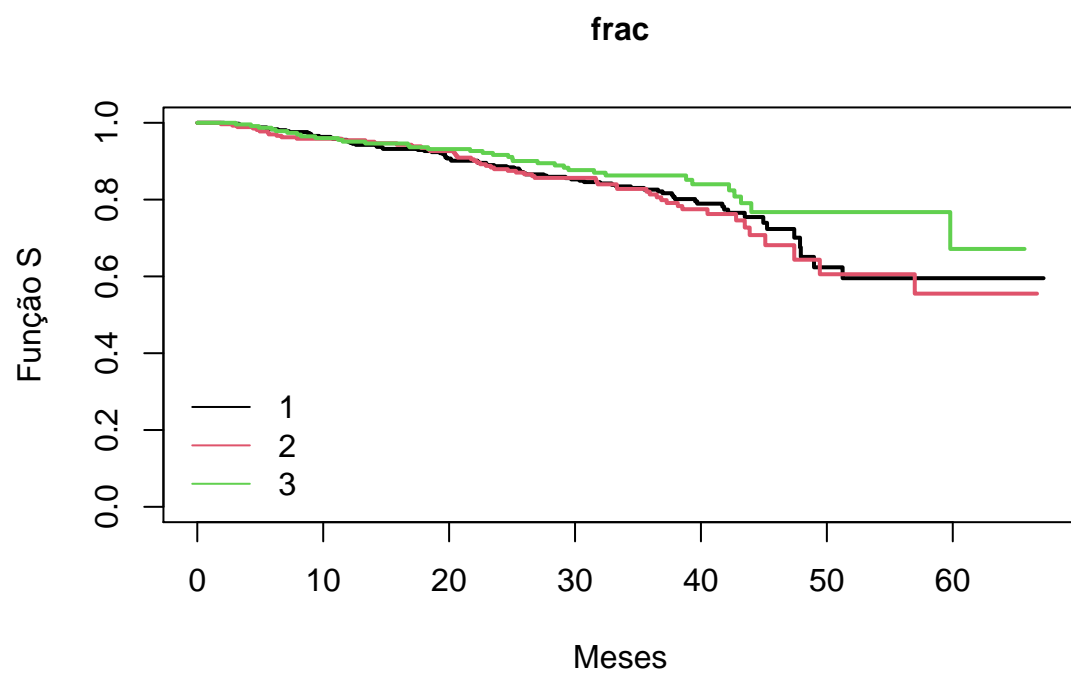
Para confecção dos gráficos de sobrevivência é necessário categorizar as variáveis contínuas, esse processo foi realizado por meio da mediana, valores menor que a mediana receberam 1, e maiores que a mediana receberam 2.

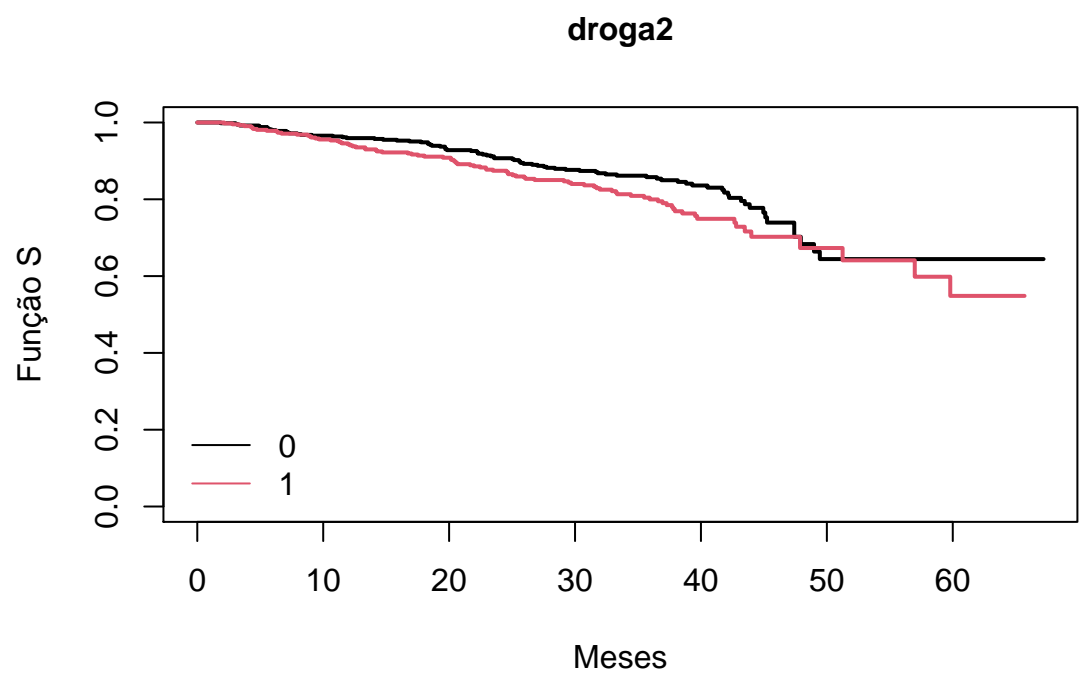
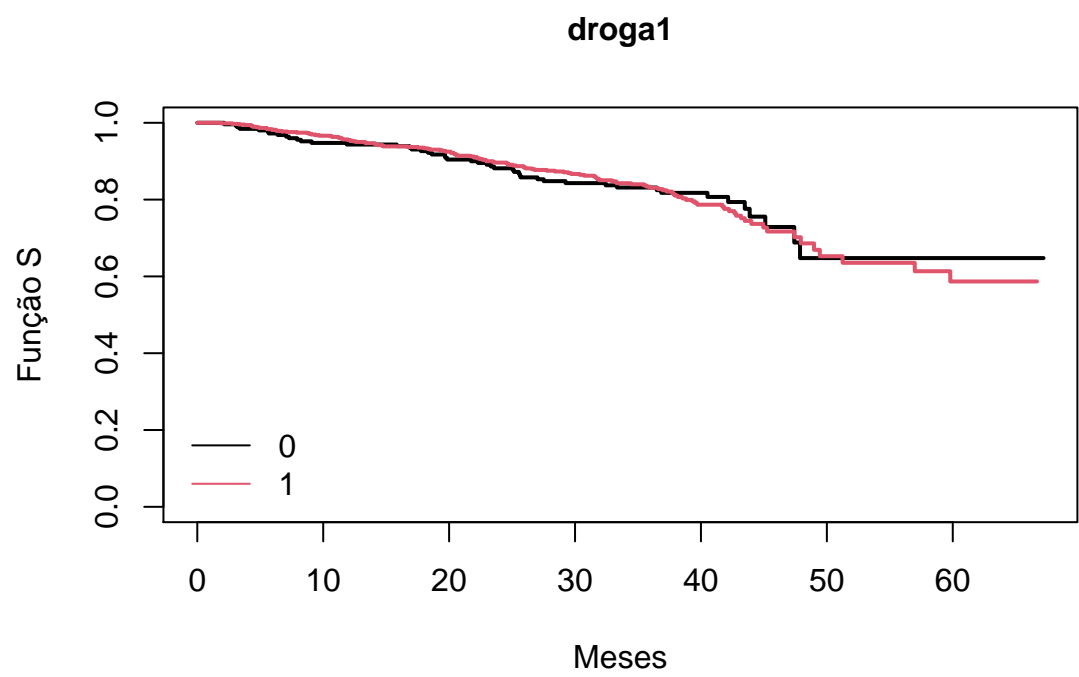
```
InsufCardiacaT = InsufCardiaca %>%
  transmute( imc = ifelse(imc > 25, 2, 1),
             pas = ifelse(pas > median(pas), 2, 1),
             hb = ifelse(hb > median(hb), 2, 1),
             trigli = ifelse(trigli>200, 2, 1))%>%
  mutate(across(.cols = everything(), factor))
var = names(InsufCardiaca)
covn = c("imc", "pas", "hb", "trigli")
covc = var[!var %in% c("tempo", "censura", covn)]
InsufCardiaca[,covc]<- lapply(InsufCardiaca[,covc], factor)
dadoscor = cbind(InsufCardiaca[, var[!var %in% covn]], InsufCardiacaT)
varc = var[-c(1,2,5,6,9,10)]
for ( i in 1:length(varc)){
  ncol = length(levels(dadoscor[,varc[i]]))
  km = survfit(Surv(tempo, censura)~ get(varc[i]), data = dadoscor, se.fit = F)
  plot(km, xlab = "Meses", ylab = "Função S", main = varc[i], col = 1:ncol, lwd = 2, cex.lab = 1, ce
  legend("bottomleft",levels(as.factor(dadoscor[,varc[i]])), bty="n", lty = 1, col = 1:ncol)
}

plot(survfit(Surv(tempo, censura)~ pas, data = dadoscor, se.fit = F), xlab = "Meses", ylab = "Função S",
legend("bottomleft", levels(as.factor(dadoscor[, "pas"])), bty="n", lwd = 2, col = 1:2)
plot(survfit(Surv(tempo, censura)~ imc, data = dadoscor, se.fit = F), xlab = "Meses", ylab = "Função S",
legend("bottomleft", levels(as.factor(dadoscor[, "imc"])), bty="n", lwd = 2, col = 1:2)
plot(survfit(Surv(tempo, censura)~ trigli, data = dadoscor, se.fit = F), xlab = "Meses", ylab = "Função S",
legend("bottomleft", levels(as.factor(dadoscor[, "trigli"])), bty="n", lwd = 2, col = 1:2)
plot(survfit(Surv(tempo, censura)~ hb, data = dadoscor, se.fit = F), xlab = "Meses", ylab = "Função S",
legend("bottomleft", levels(as.factor(dadoscor[, "hb"])), bty="n", lwd = 2, col = 1:2)
```

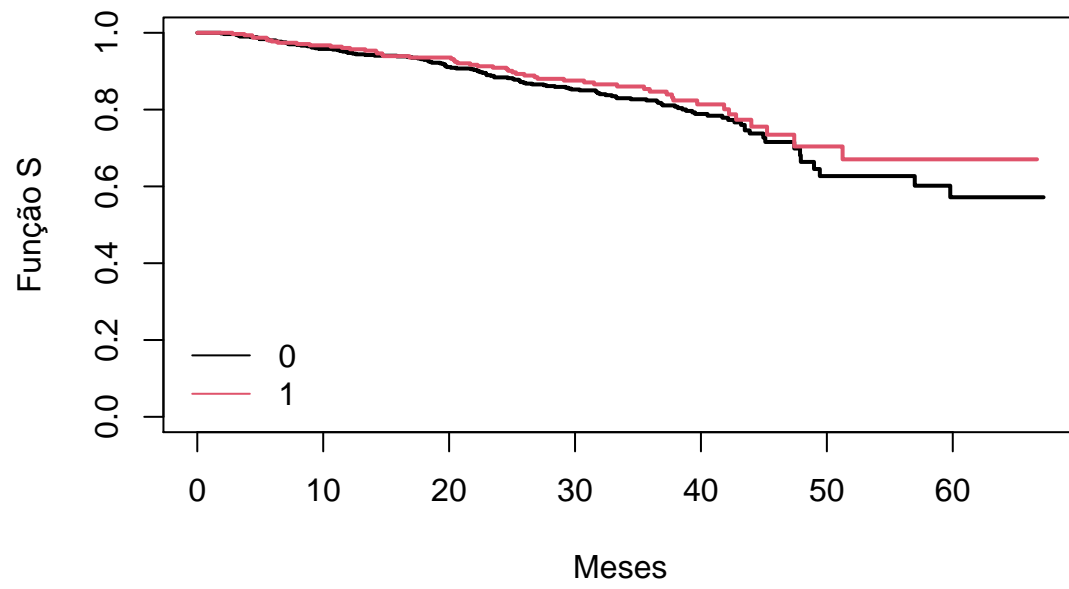




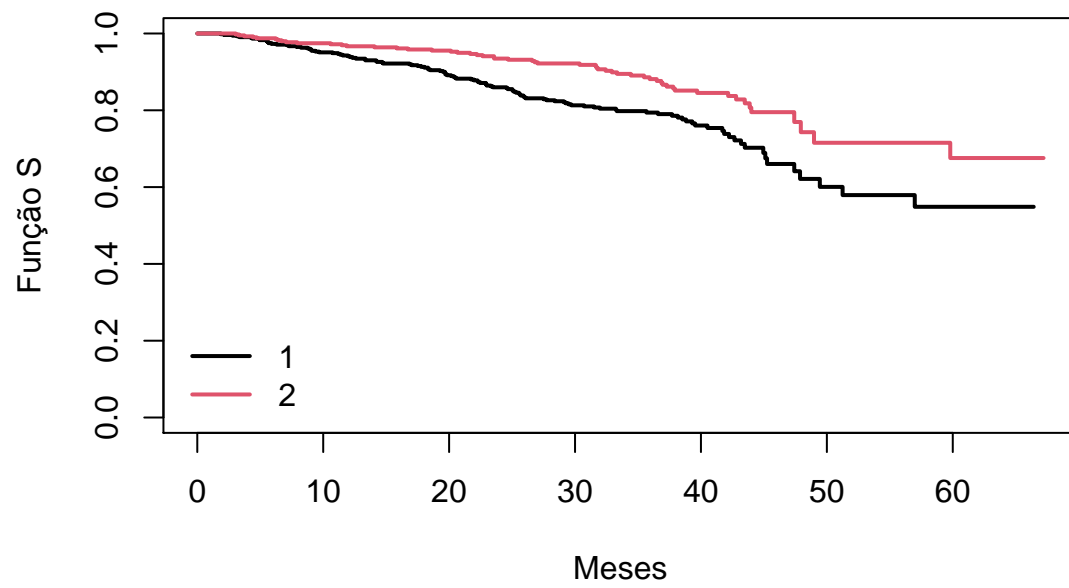




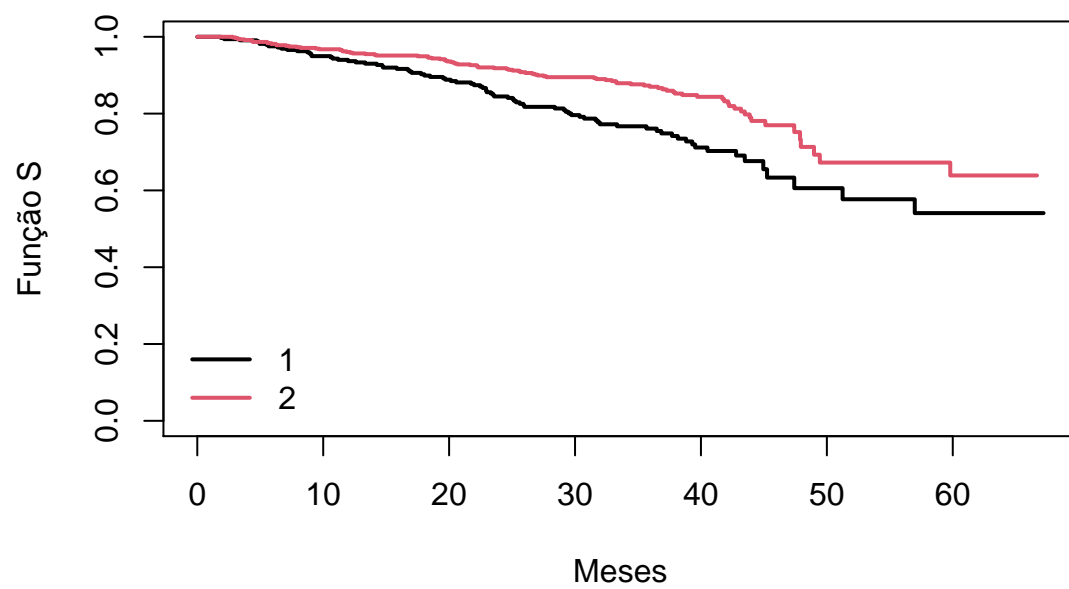
droga3



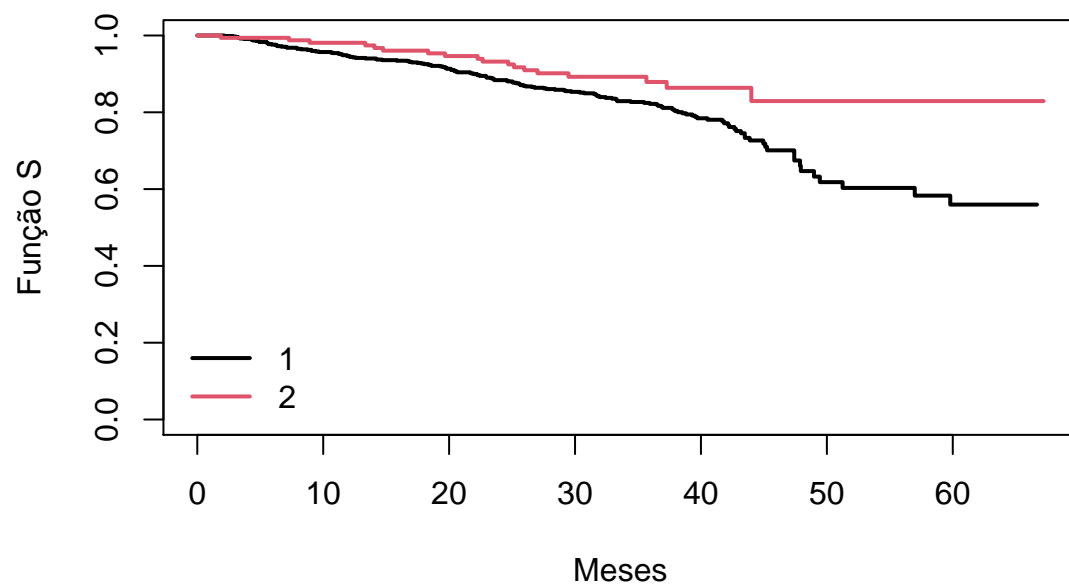
pas

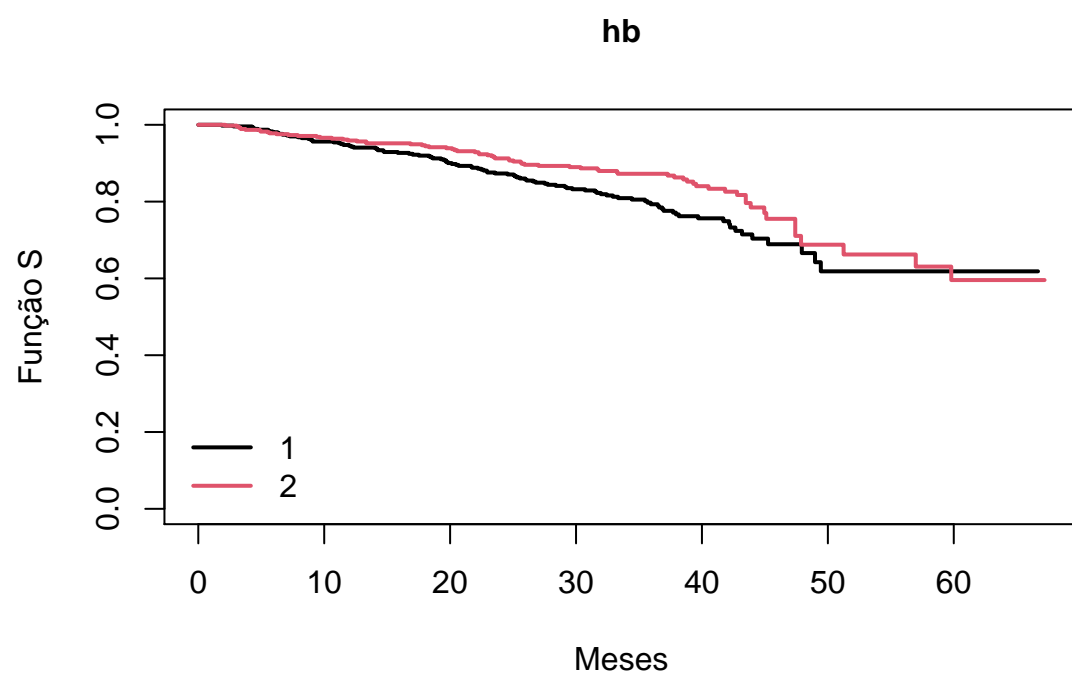


imc



trigli





Testes logrank

```
ptestl =c()
estatc = c()
testelog = data.frame()
for ( i in 1:length(var)){
  tlr = survdiff(Surv(tempo, censura)~ get(var[i]), data = dadoscor, rho = 0)
  estatc[i] = tlr$chisq
  ptestl[i] = 1 - pchisq(tlr$chisq, ncol-1)
}
testelog <- data.frame(Variavel = var, Estatistica = estatc, pvalor = ptestl)
xtable(testelog,digits = 2,caption = "Teste logrank")
```

	Variavel	Estatistica	pvalor
1	tempo	4131.39	0.00
2	censura	1164.04	0.00
3	Idade	8.47	0.00
4	sexo	1.71	0.19
5	imc	14.60	0.00
6	pas	13.27	0.00
7	diag	15.68	0.00
8	class	18.74	0.00
9	hb	5.20	0.02
10	trigli	6.00	0.01
11	frac	2.63	0.10
12	ritmo	7.41	0.01
13	droga1	0.03	0.86
14	droga2	4.24	0.04
15	droga3	1.04	0.31

Tabela 2: Teste logrank

Para complementar os gráficos de sobrevivência é realizado testes logrank. Nesses testes é comumente adotado nível de de significância de 25%. Com este nível de significância as variáveis ‘droga3’, ‘frac’ e ‘droga1’ tiveram p-valor abaixo do esperado, e portanto, não serão inclusas no modelo.

Ajuste do modelo Weibull

Para este exercício primeiramente será escolhido o modelo Weibull para ajuste dos dados, caso o modelo não se adeque satisfatoriamente aos dados, serão ajustados outros modelos, como por exemplo Log-Normal e Log-Logística.

```
m_wei <- survreg(Surv(tempo, censura) ~ Idade + sexo + imc + pas + diag + class
                + hb + trigli + ritmo + droga2, data = dadoscor, dist = "weibull")
a = 1/m_wei$scale
exp = (dadoscor$tempo * exp(-m_wei$linear.predictors))^a
mkap = survfit( Surv(exp, dadoscor$censura) ~ 1)
time = mkap$time
fs = mkap$surv
fse = exp(-mkap$time)
par(mfrow = c(1,2))
plot(fs, fse, xlab = "S: K-M", ylab = "S: Exp", pch = 16, cex = 1.5)
plot(mkap, conf.int = F, xlab = "Res Cox-Snell", ylab = "Sei est", cex = 1.5)
lines(time, fse, lty = 4)
summary(m_wei)
```

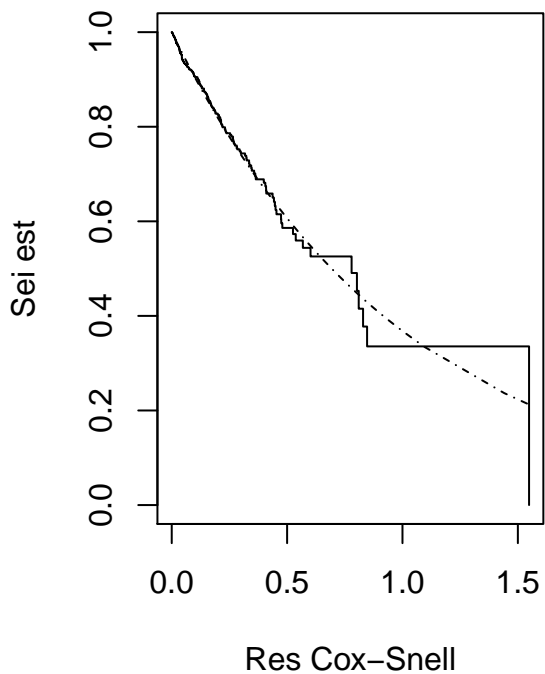
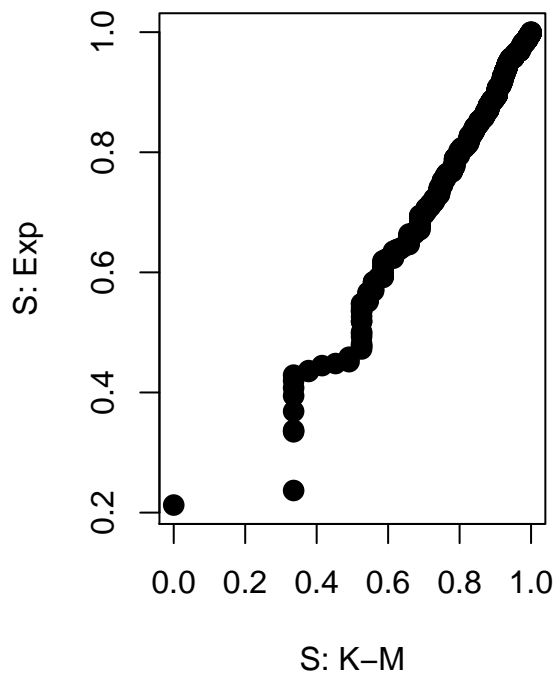
```
##
## Call:
## survreg(formula = Surv(tempo, censura) ~ Idade + sexo + imc +
##      pas + diag + class + hb + trigli + ritmo + droga2, data = dadoscor,
##      dist = "weibull")
##
```

	Value	Std. Error	z	p
## (Intercept)	2.833	0.414	6.84	7.7e-12
## Idade1	0.266	0.287	0.93	0.3541
## Idade2	0.193	0.241	0.80	0.4248
## Idade3	0.154	0.243	0.63	0.5270
## Idade4	-0.118	0.234	-0.50	0.6137
## sexo2	0.260	0.117	2.23	0.0255
## imc	0.301	0.109	2.76	0.0058
## pas	0.295	0.121	2.44	0.0146
## diag2	-0.080	0.162	-0.50	0.6206
## diag3	0.273	0.156	1.75	0.0806
## diag4	0.264	0.197	1.34	0.1794
## diag5	0.452	0.262	1.73	0.0839
## class2	0.103	0.170	0.61	0.5427
## class3	-0.316	0.169	-1.87	0.0614
## class4	-0.401	0.182	-2.21	0.0274
## class9	-0.576	0.260	-2.22	0.0265
## hb	0.278	0.112	2.48	0.0133
## trigli	0.291	0.167	1.75	0.0808

```
## ritmo2      0.245      0.154  1.59  0.1121
## ritmo3      0.206      0.150  1.37  0.1718
## ritmo4      0.735      0.304  2.41  0.0158
## droga21     -0.261      0.107 -2.45  0.0143
## Log(scale)  -0.422      0.068 -6.21 5.3e-10
##
## Scale= 0.656
##
## Weibull distribution
## Loglik(model)= -978.2   Loglik(intercept only)= -1023.3
## Chisq= 90.05 on 21 degrees of freedom, p= 1.6e-10
## Number of Newton-Raphson Iterations: 8
## n= 945
```

```
AIC(m_wei)
```

```
## [1] 2002.455
```



Com o modelo ajustado podemos notar que o AIC do modelo é de 2002.455. Como valor-p para o teste de máxima verossimilhança é 1.6×10^{-10} rejeitamos a hipótese de todos os parâmetros serem nulos.

Fator de aceleração das Covariáveis

```
xtable(data.frame(FatorGamma = exp(m_wei$coefficients)),digits = 2,  
        caption = "Fator de aceleração das Covariáveis")
```

	FatorGamma
(Intercept)	17.00
Idade1	1.30
Idade2	1.21
Idade3	1.17
Idade4	0.89
sexo2	1.30
imc	1.35
pas	1.34
diag2	0.92
diag3	1.31
diag4	1.30
diag5	1.57
class2	1.11
class3	0.73
class4	0.67
class9	0.56
hb	1.32
trigli	1.34
ritmo2	1.28
ritmo3	1.23
ritmo4	2.08
droga21	0.77

Tabela 3: Fator de aceleração das Covariáveis

Com o fator de aceleração $\gamma = e^{\beta_i}$ calculado para cada coeficiente do modelo, pode-se verificar se tal coeficiente influencia no tempo de sobrevida do paciente. Para valores maiores que 1 podemos concluir que tal coeficiente influencia diminuindo o tempo de sobrevida, para valores iguais a 1 podemos concluir que tal coeficiente não afeta o tempo de sobrevida e para valores maiores que 1 podemos concluir que tal coeficiente influencia aumentando o tempo de sobrevida.

Portanto de acordo com os resultados da tabela a acima podemos concluir que as únicas características benéficas ao tempo de sobrevida do paciente foram: Idade4 (ter 65 anos ou mais), diag2 (ter doença cardíaca hipertensiva), droga (usar Digoxina). Em relação à variável classes, a única que aumentou o risco de morte foi class2 (limitado a grandes esforços).