



Universidade Federal de Rondônia
Departamento de Matemática e Estatística
Bacharelado em Estatística

Edimar
Jossivana Macedo
Douglas Vinícius

Relatório do Trabalho de Análise de Sobrevivência

Ji-Paraná
2019

Edimar
Jossivana Macedo
Douglas Vinícius

Relatório do Trabalho de Análise de Sobrevivência

Relatório apresentado à Disciplina de Análise
de Sobrevivência do Curso de Bacharelado
em Estatística, da Universidade Federal de
Rondônia - UNIR, para obtenção de aprovação.

Orientador:
Prof. Dr.

Ji-Paraná
2019

Sumário

Lista de Figuras	2
	2
1 Introdução	3
2 Descrição do Banco de Dados	4
3 Desenvolvimento	6
3.1 Carregar os dados	6
3.2 Estimador de Kaplan-Meier	6
3.3 Modelo de Riscos Proporcionais de Cox	9
4 Resultados e Discussões	12
5 Considerações Finais	13

Lista de Figuras

1. Introdução

2. Descrição do Banco de Dados

Os dados são provenientes de coortes hospitalares de pacientes portadores de HIV. A primeira coorte é constituída dos pacientes portadores de HIV atendidos entre 1986 e 2000 no Instituto de Pesquisa Clínica Evandro Chagas (Ipec/Fiocruz). Dessa coorte, obteve-se uma amostra de 193 indivíduos que foram diagnosticados como portadores de Aids (critério CDC 1993) durante o período de acompanhamento.

As variáveis registradas para cada paciente estão listadas na tabela a seguir. Elas foram obtidas a partir dos prontuários clínicos, como descrito em Campos (2005). Nesse artigo também se encontra uma análise exploratória completa desses dados, assim como a análise de sobrevivência em Aids, utilizando modelos não paramétricos e modelos de Cox clássicos.

Esses dados estão disponíveis no arquivo `ipecc.csv`, que está organizado para análise de sobrevivência usando os métodos não estendidos, isto é, com uma linha para cada paciente e sem covariáveis tempo-dependentes.

Tabela 2.1: Descrição dos dados

Variáveis	Descrição
id	identificação do paciente
ini	data do diagnóstico da Aids(em dias)
fim	data do óbito (ou perda do paciente)
tempo	dias de sobrevivência do diagnóstico até o óbito
status	0 = censura, 1 = óbito
sexo	F = feminino, M = masculino
escola	0 = sem escolaridade, 1 = ensino fundamental, 2 = ensino médio, 3 = ensino superior
idade	idade na data do diagnóstico de Aids (20 a 68 anos)
risco	0 = homossexual masculino, 1 = usuário de drogas injetáveis, 2 = transfusão, 3 = contato sexual com HIV+, 5 = hétero c/múltiplos parceiros, 6 = dois fatores de risco
acompan	acompanhamento: 0 = ambulatorial/hospital-dia, 1 = internação posterior, 2 = internação imediata
obito	S = óbito, N = não óbito, I = ignorado
anotrat	ano do início do tratamento (1990 a 2000), 9 = sem tratamento
tratam	terapia antirretroviral: 0 = nenhum, 1 = mono, 2 = combinada, 3 = potente
doença	de apresentação: 1 = pcp, 2 = pcp pulmonar, 3 = pcp disseminada, 4 = toxoplasmose, 5 = sarcoma, 7 = outra doença, 8 = candidíase, 9 = duas doenças, 10 = herpes, 99 = definido por cd4
propcp	profilaxia para pneumocistis: 0 = sem profilaxia, 2 = primária, 3 = secundária, 4 = ambas

3. Desenvolvimento

3.1 Carregar os dados

Esse primeiro bloco de código carrega os pacotes necessários, juntamente com o veteran conjunto de dados do survivalpacote que contém dados de um estudo randomizado de dois tratamentos para câncer de pulmão.

```
library(survival)
library(ranger)
library(ggplot2)
library(dplyr)
library(ggfortify)
#-----
data(veteran)
head(veteran)
```

##	trt	celltype	time	status	karno	diagtime	age	prior
## 1	1	squamous	72	1	60	7	69	0
## 2	1	squamous	411	1	70	5	64	10
## 3	1	squamous	228	1	60	3	38	0
## 4	1	squamous	126	1	60	9	63	10
## 5	1	squamous	118	1	70	11	65	10
## 6	1	squamous	10	1	20	5	49	0

3.2 Estimador de Kaplan-Meier

```
# Kaplan Meier Survival Curve
km <- with(veteran, Surv(time, status))
head(km, 80)
```

## [1]	72	411	228	126	118	10	82	110	314	100+	42	8	144	25+	11
## [16]	30	384	4	54	13	123+	97+	153	59	117	16	151	22	56	21
## [31]	18	139	20	31	52	287	18	51	122	27	54	7	63	392	10
## [46]	8	92	35	117	132	12	162	3	95	177	162	216	553	278	12
## [61]	260	200	156	182+	143	105	103	250	100	999	112	87+	231+	242	991
## [76]	111	1	587	389	33										

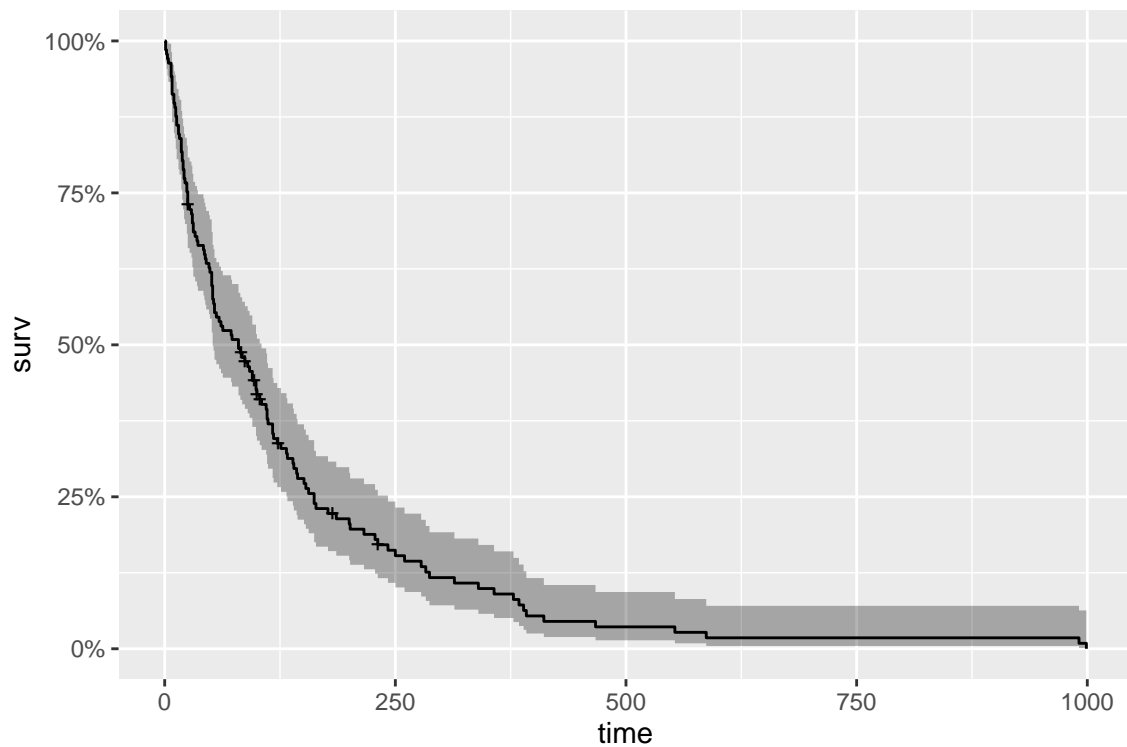
```
km_fit <- survfit(Surv(time, status) ~ 1, data=veteran)
summary(km_fit, times = c(1,30,60,90*(1:10)))
```

```
## Call: survfit(formula = Surv(time, status) ~ 1, data = veteran)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    1     137      2   0.985  0.0102   0.96552    1.0000
##   30      97     39   0.700  0.0392   0.62774    0.7816
```

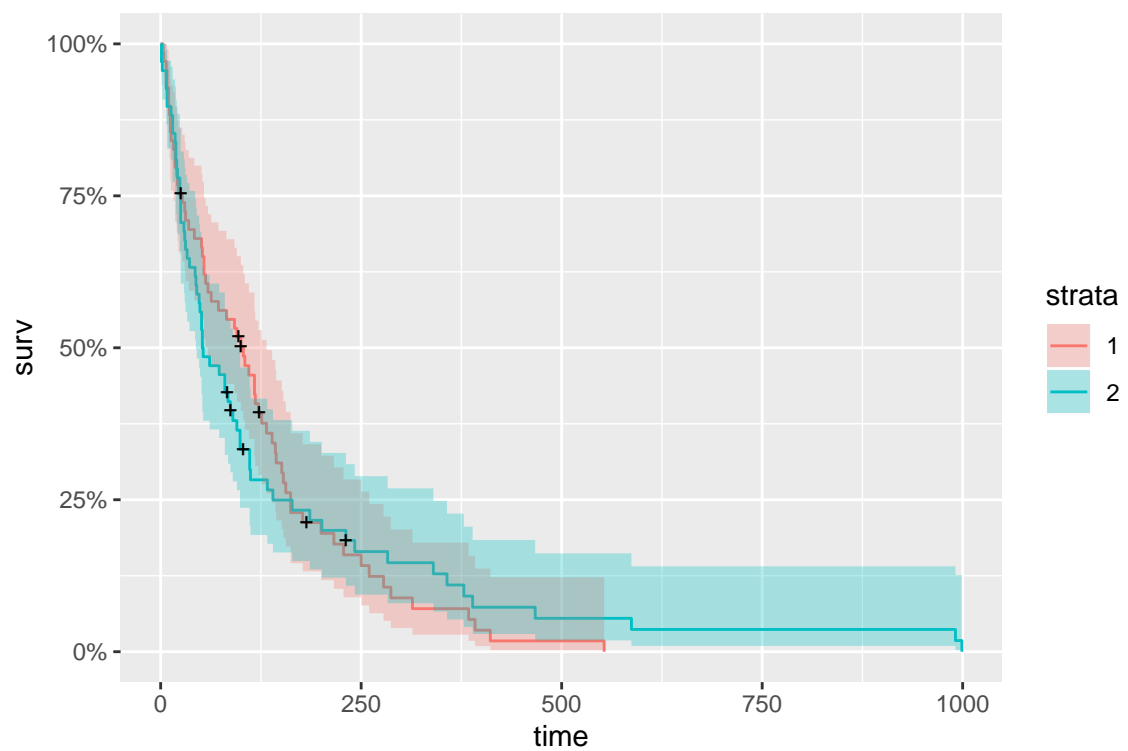


```
##      60      73      22      0.538 0.0427      0.46070      0.6288
##      90      62      10      0.464 0.0428      0.38731      0.5560
##     180      27      30      0.222 0.0369      0.16066      0.3079
##     270      16       9      0.144 0.0319      0.09338      0.2223
##     360      10       6      0.090 0.0265      0.05061      0.1602
##     450       5       5      0.045 0.0194      0.01931      0.1049
##     540       4       1      0.036 0.0175      0.01389      0.0934
##     630       2       2      0.018 0.0126      0.00459      0.0707
##     720       2       0      0.018 0.0126      0.00459      0.0707
##     810       2       0      0.018 0.0126      0.00459      0.0707
##     900       2       0      0.018 0.0126      0.00459      0.0707
```

```
autoplot(km_fit)
```

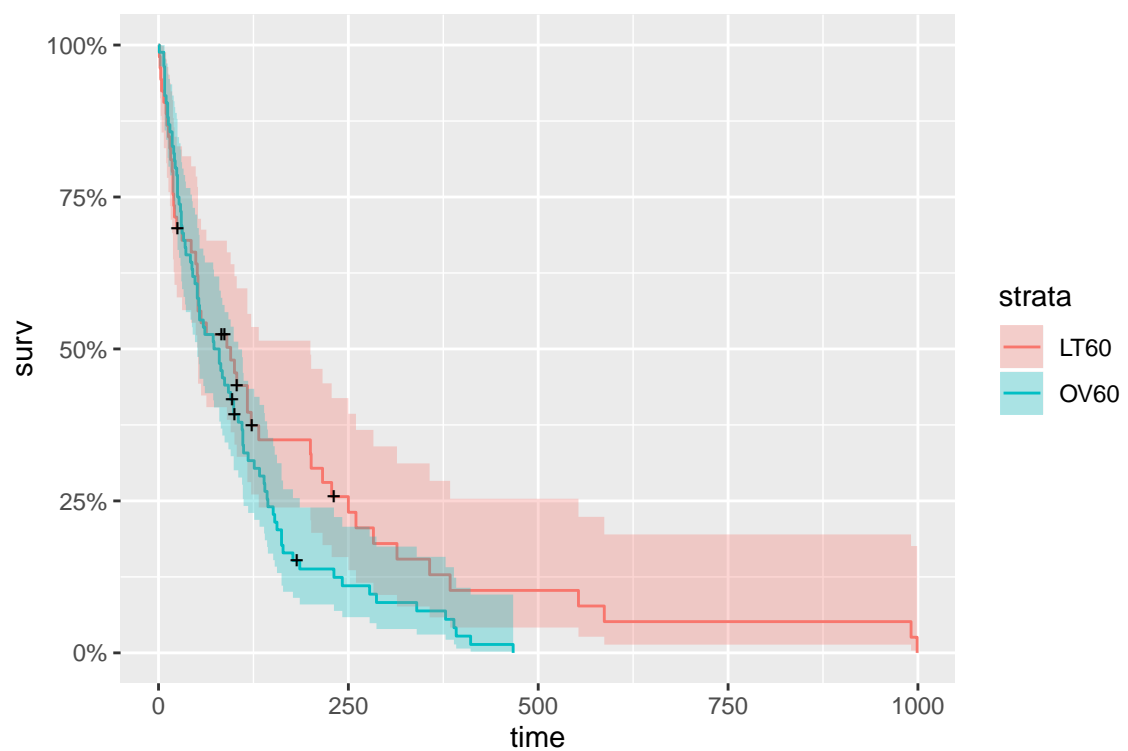


```
km_trt_fit <- survfit(Surv(time, status) ~ trt, data=veteran)
autoplot(km_trt_fit)
```



```
vet <- mutate(veteran, AG = ifelse((age < 60), "LT60", "OV60"),
              AG = factor(AG),
              trt = factor(trt, labels=c("standard", "test")),
              prior = factor(prior, labels=c("NO", "Yes")))

km_AG_fit <- survfit(Surv(time, status) ~ AG, data=vet)
autoplot(km_AG_fit)
```

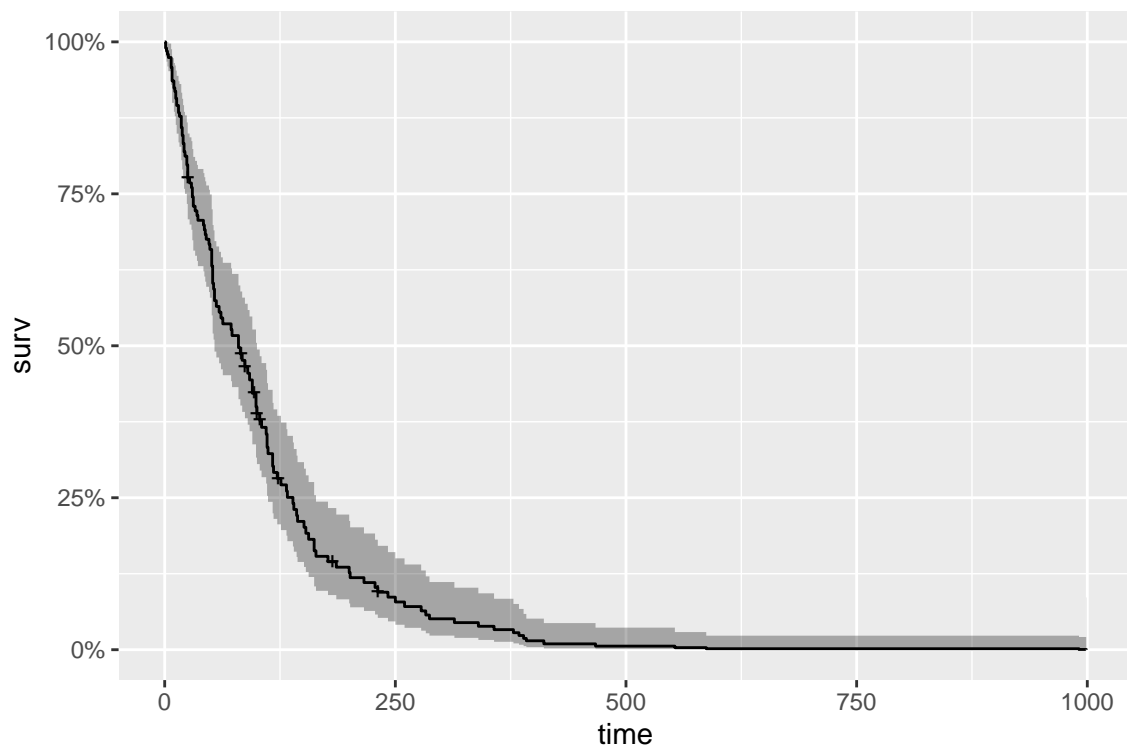


3.3 Modelo de Riscos Proporcionais de Cox

```
# Fit Cox Model
cox <- coxph(Surv(time, status) ~ trt + celltype + karno + diagtime + age + priorYes, data = vet)
summary(cox)

## Call:
## coxph(formula = Surv(time, status) ~ trt + celltype + karno +
##       diagtime + age + prior, data = vet)
##
## n= 137, number of events= 128
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## trttest          2.946e-01  1.343e+00  2.075e-01  1.419  0.15577
## celltypesmallcell 8.616e-01  2.367e+00  2.753e-01  3.130  0.00175 **
## celltypeadeno    1.196e+00  3.307e+00  3.009e-01  3.975  7.05e-05 ***
## celltypelarge    4.013e-01  1.494e+00  2.827e-01  1.420  0.15574
## karno            -3.282e-02  9.677e-01  5.508e-03 -5.958  2.55e-09 ***
## diagtime         8.132e-05  1.000e+00  9.136e-03  0.009  0.99290
## age              -8.706e-03  9.913e-01  9.300e-03 -0.936  0.34920
## priorYes         7.159e-02  1.074e+00  2.323e-01  0.308  0.75794
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## trttest          1.3426      0.7448    0.8939    2.0166
## celltypesmallcell 2.3669      0.4225    1.3799    4.0597
## celltypeadeno    3.3071      0.3024    1.8336    5.9647
## celltypelarge    1.4938      0.6695    0.8583    2.5996
## karno            0.9677      1.0334    0.9573    0.9782
## diagtime         1.0001      0.9999    0.9823    1.0182
## age              0.9913      1.0087    0.9734    1.0096
## priorYes         1.0742      0.9309    0.6813    1.6937
##
## Concordance= 0.736 (se = 0.021 )
## Likelihood ratio test= 62.1  on 8 df,  p=2e-10
## Wald test              = 62.37  on 8 df,  p=2e-10
## Score (logrank) test = 66.74  on 8 df,  p=2e-11

cox_fit <- survfit(cox)
#plot(cox_fit, main = "cph model", xlab="Days")
autoplot(cox_fit)
```



```
aa_fit <-aareg(Surv(time, status) ~ trt + celltype +
               karno + diagtime + age + prior ,
               data = vet)

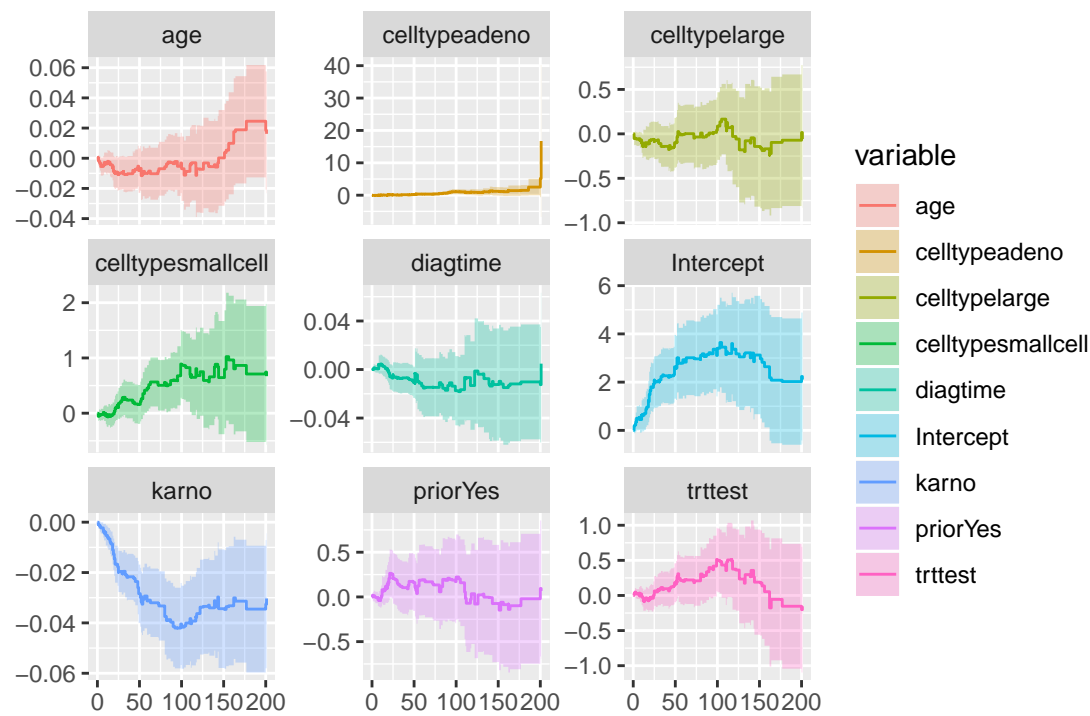
aa_fit

## Call:
## aareg(formula = Surv(time, status) ~ trt + celltype + karno +
##       diagtime + age + prior, data = vet)
##
## n= 137
## 75 out of 97 unique event times used
##
##
```

	slope	coef	se(coef)	z	p
## Intercept	0.083400	3.81e-02	1.09e-02	3.490	4.79e-04
## trttest	0.006730	2.49e-03	2.58e-03	0.967	3.34e-01
## celltypesmallcell	0.015000	7.30e-03	3.38e-03	2.160	3.09e-02
## celltypeadeno	0.018400	1.03e-02	4.20e-03	2.450	1.42e-02
## celltypelarge	-0.001090	-6.21e-04	2.71e-03	-0.229	8.19e-01
## karno	-0.001180	-4.37e-04	8.77e-05	-4.980	6.28e-07
## diagtime	-0.000243	-4.92e-05	1.64e-04	-0.300	7.65e-01
## age	-0.000246	-6.27e-05	1.28e-04	-0.491	6.23e-01
## priorYes	0.003300	1.54e-03	2.86e-03	0.539	5.90e-01

```
##
## Chisq=41.62 on 8 df, p=1.6e-06; test weights=aalen

#summary(aa_fit) # provides a more complete summary of results
autoplot(aa_fit)
```



4. Resultados e Discussões

5. Considerações Finais