

MAE0514 - Prova 2 - Parte 2

Rubens Santos Andrade Filho¹

Julho de 2021

Sumário

Atividade 2	2
(a) Análise descritiva	2
(b) Modelo inicial	5
(c) Seleção de variáveis	6
(d) Interpretação	7
Código	8

¹Número USP: 10370336

Atividade 2

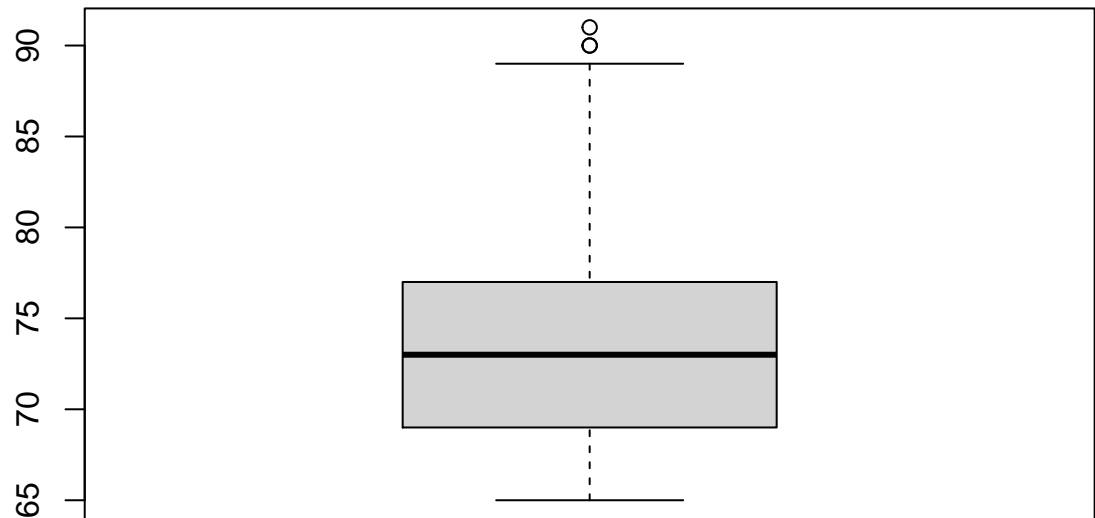
O arquivo FRTCS. dat contém dados sobre um estudo realizado na França (French Three Cities Study), com o objetivo de estudar o efeito da pressão e uso de drogas para hipertensão na sobrevivência de pessoas de uma determinada população. Os indivíduos na amostra foram acompanhados até óbito ou perda de acompanhamento, e os valores da pressão diastólica, sistólica e o uso de drogas anti-hipertensivas foram avaliados no momento de entrada no estudo e em mais dois outros instantes.

(a) Análise descritiva

Inicialmente, nota-se que existem 3 pacientes com observações faltantes de pressão diastólica e pressão sistólica, ambas no segundo acompanhamento. Com isso, vamos desconsiderar esses 3 pacientes e considerar apenas n=694 pacientes.

d	age	sex	date0	sbp0	dbp0	antihyp0	date1	sbp1	dbp1	antihyp1	date2	sbp2	dbp2	antihyp2	date_event	censor
3	75	2	1999-03-29	145.5	86.5	1	2001-07-04	158.5	96.5	1	2003-07-15	NA	NA	1	2004-03-28	0
237	80	2	1999-11-26	210.0	106.0	1	2001-10-16	155.5	103.0	1	2003-06-06	NA	NA	1	2004-11-25	0
263	74	2	2000-01-19	134.5	87.0	0	2001-10-24	174.5	100.5	0	2003-06-10	NA	NA	0	2005-01-18	0

A idade dos pacientes tem uma distribuição levemente assimétrica à direita com uma idade média de 73 anos e desvio padrão de 5,27 anos.



```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      65.00  69.00   73.00   73.71   77.00   91.00
```

```
## [1] 5.277194
```

Além disso, nota-se que que 66% dos pacientes são do sexo feminino. E cerca de 10% das observações foram censuradas.

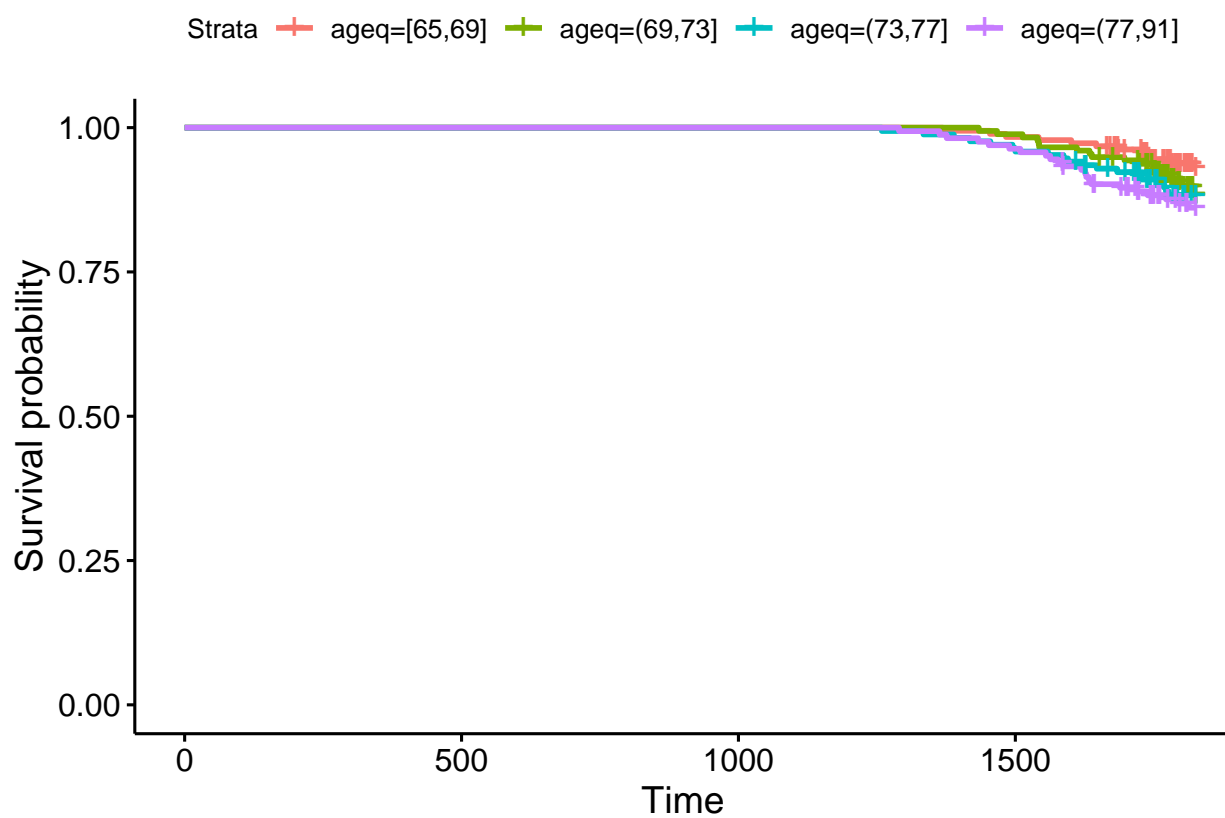
```
## dados$sex    n percent
##      male 233   33.6%
##      female 461  66.4%
```

```
## dados$censor  n percent
##              0 622   89.6%
##              1  72   10.4%
```

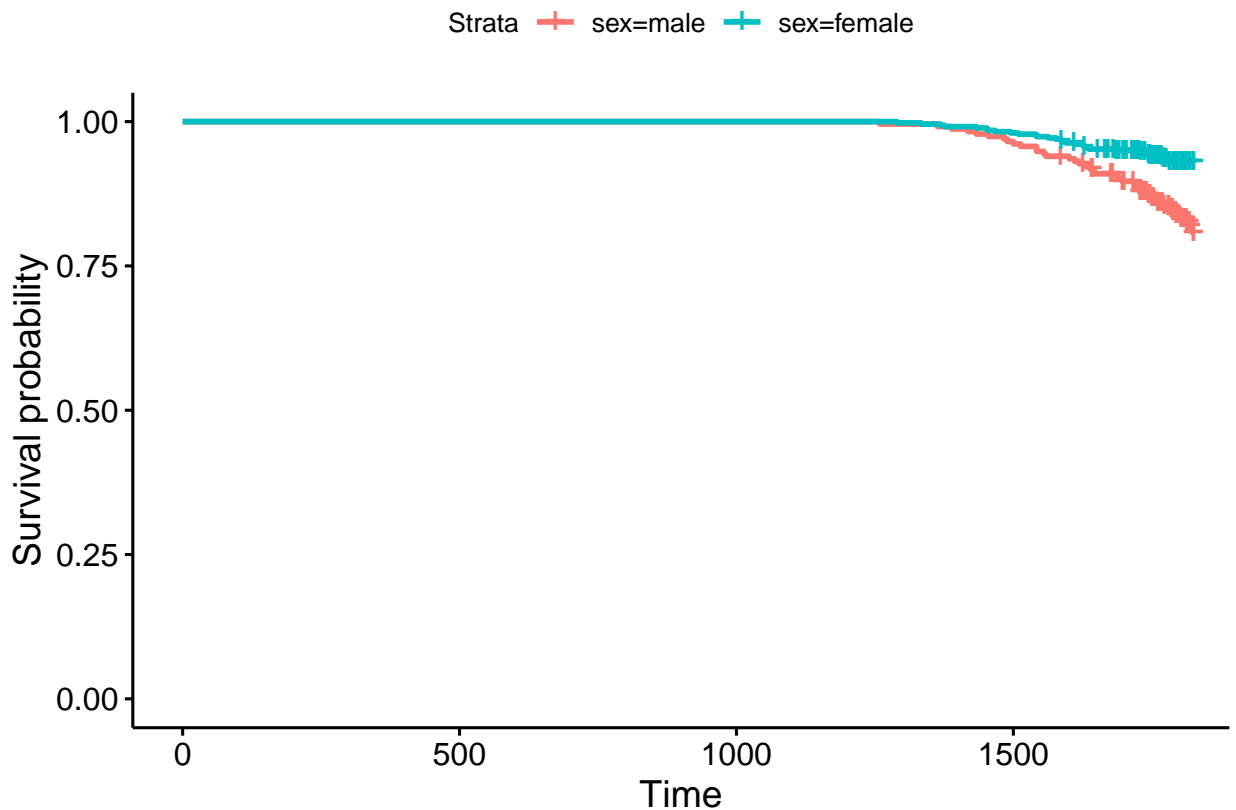
```
## Call: survfit(formula = Surv(tevent, censor) ~ ageq, data = dados)
```

```
##
```

```
##              n events median 0.95LCL 0.95UCL
## ageq=[65,69] 185     12    NA      NA     NA
## ageq=(69,73] 176     19    NA      NA     NA
## ageq=(73,77] 169     19    NA      NA     NA
## ageq=(77,91] 164     22    NA      NA     NA
```



```
## Call: survfit(formula = Surv(tevent, censor) ~ sex, data = dados)
##
##           n events median 0.95LCL 0.95UCL
## sex=male   233     41    NA      NA     NA
## sex=female 461     31    NA      NA     NA
```



Observa-se indícios de diferenças nas curvas de sobrevivência entre os sexos. As curvas de sobrevivência estratificadas por idade (a idade foi categorizada pelos quartis) mostra que as maiores idades apresentam uma menor estimativa de sobrevivência, entretanto graficamente é difícil avaliar se essa diferença é significativa.

(b) Modelo inicial

Organizamos os dados de maneira apropriada para ajustar o modelo semiparamétrico de Cox (incluimos os códigos utilizados no final da prova) com variáveis dependentes do tempo. Ajustamos o modelo semiparamétrico de Cox, incluindo como variáveis explicativas: sexo, idade, pressão diastólica, sistólica e o uso de drogas anti-hipertensivas.

```
## Call:
## coxph(formula = Surv(tini, tfim, censor) ~ age + sex + sbp +
##       dbp + antihyp, data = dados_long)
##
##               coef exp(coef)  se(coef)      z        p
## age           0.032882  1.033428  0.023466  1.401    0.1611
## sexfemale     -1.246693  0.287454  0.255405 -4.881 1.05e-06
## sbp            0.006349  1.006370  0.007058  0.900    0.3683
## dbp           -0.032007  0.968500  0.014457 -2.214    0.0268
## antihypyes    1.368619  3.929921  0.275214  4.973 6.60e-07
##
## Likelihood ratio test=59.16  on 5 df, p=1.813e-11
```

```

## n= 694, number of events= 72
## (1388 observations deleted due to missingness)

## Call:
## coxph(formula = Surv(tini, tfim, censor) ~ age + sex + sbp +
##       dbp + antihyp, data = dados_long)
##
## n= 694, number of events= 72
## (1388 observations deleted due to missingness)
##
##               coef exp(coef)  se(coef)      z Pr(>|z|)
## age           0.032882  1.033428  0.023466  1.401  0.1611
## sexfemale    -1.246693  0.287454  0.255405 -4.881 1.05e-06 ***
## sbp           0.006349  1.006370  0.007058  0.900  0.3683
## dbp          -0.032007  0.968500  0.014457 -2.214  0.0268 *
## antihypyes   1.368619  3.929921  0.275214  4.973 6.60e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## age           1.0334      0.9677      0.9870      1.0821
## sexfemale     0.2875      3.4788      0.1742      0.4742
## sbp           1.0064      0.9937      0.9925      1.0204
## dbp           0.9685      1.0325      0.9414      0.9963
## antihypyes    3.9299      0.2545      2.2915      6.7398
##
## Concordance= 0.736 (se = 0.028 )
## Likelihood ratio test= 59.16 on 5 df,  p=2e-11
## Wald test              = 54.77 on 5 df,  p=1e-10
## Score (logrank) test = 60.91 on 5 df,  p=8e-12

```

O resumo mostra que pelo tanto o teste de RV, Wald e Socré deram significativos a um nível de 5%, indicando que pelo menos uma das variáveis explicativas foi significativa e o modelo é diferente do modelo apenas com intercepto.

(c) Seleção de variáveis

Selecionamos as variáveis explicativas significativas e apresentamos o modelo final ajustado. Começamos com o modelo com todas as variáveis e uma a uma retiramos a variável do modelo e realizamos o teste de razão de verossimilhanças entre este e o anterior, não incluindo a variável no modelo caso o teste não seja significativo a 5% (não rejeitando a hipótese nula do menor número de parâmetros).

```

## Analysis of Deviance Table
## Cox model: response is Surv(tini, tfim, censor)
## Model 1: ~ sex + sbp + dbp + antihyp
## Model 2: ~ age + sex + sbp + dbp + antihyp
##      loglik  Chisq Df P(>|Chi|)
## 1 -388.87
## 2 -387.91 1.9127 1 0.1667

```

```

## Analysis of Deviance Table

```

```
## Cox model: response is Surv(tini, tfim, censor)
## Model 1: ~ age + sbp + dbp + antihyp
## Model 2: ~ age + sex + sbp + dbp + antihyp
##      loglik  Chisq Df P(>|Chi|)
## 1 -399.90
## 2 -387.91 23.978  1 9.746e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(tini, tfim, censor)
## Model 1: ~ age + sex + sbp + antihyp
## Model 2: ~ age + sex + sbp + dbp + antihyp
##      loglik  Chisq Df P(>|Chi|)
## 1 -390.47
## 2 -387.91 5.1315  1  0.02349 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(tini, tfim, censor)
## Model 1: ~ age + sex + sbp + dbp
## Model 2: ~ age + sex + sbp + dbp + antihyp
##      loglik Chisq Df P(>|Chi|)
## 1 -402.27
## 2 -387.91 28.72  1 8.364e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

O modelo selecionado incluiu **sex**, **dbp** e **antihyp**. Entretanto, **dbp** não é significativa a 5% no modelo apenas com essas três covariáveis. Com isso, o modelo final inclui apenas **sex** e **antihyp**.

```
## Call:
## coxph(formula = Surv(tini, tfim, censor) ~ sex + antihyp, data = dados_long)
##
##              coef exp(coef) se(coef)      z      p
## sexfemale  -1.1606   0.3133   0.2402 -4.832 1.35e-06
## antihypyes  1.3503   3.8587   0.2691  5.017 5.25e-07
##
## Likelihood ratio test=50 on 2 df, p=1.387e-11
## n= 694, number of events= 72
## (1388 observations deleted due to missingness)
```

(d) Interpretação

Interpretamos a seguir o modelo ajustado no item anterior e discutimos os resultados.

```
## Call:
## coxph(formula = Surv(tini, tfim, censor) ~ sex + antihyp, data = dados_long)
##
```

```
## n= 694, number of events= 72
## (1388 observations deleted due to missingness)
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## sexfemale -1.1606    0.3133   0.2402 -4.832 1.35e-06 ***
## antihypyes  1.3503    3.8587   0.2691  5.017 5.25e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## sexfemale    0.3133    3.1919    0.1957    0.5017
## antihypyes    3.8587    0.2592    2.2769    6.5393
##
## Concordance= 0.705 (se = 0.029 )
## Likelihood ratio test= 50 on 2 df,  p=1e-11
## Wald test              = 45.93 on 2 df,  p=1e-10
## Score (logrank) test = 51.5 on 2 df,  p=7e-12
```

Utilizando a suposição de as taxas de falha serem proporcionais, interpretamos que pacientes do sexo feminino tem uma chance de óbito $(1 - \exp \hat{\beta}_1)100\% = 69\%$ menor do que pacientes do sexo masculino. Enquanto que para pacientes que fizeram o uso de drogas anti-hipertensivas, a chance de óbito é $(\exp \hat{\beta}_2 - 1)100\% = 286\%$ maior que em pacientes que não fizeram o uso.

Código

```
library(knitr)
library(tidyverse)
library(dplyr)
library(readr)
library(ggplot2)
library(survival)
library(survminer)
library(gtsummary)

knitr::opts_chunk$set(warning=FALSE,
                       # fig.dim = c(5,5),
                       # out.height = '40%',
                       # fig.align = 'center',
                       message=FALSE,
                       echo=FALSE
)

# helper com padroes predefinidos
ggsurv <- function(
  fit,
  conf.int = T,
  surv.median.line = "hv",
  ggtheme = theme_bw(),
```



```

xlim=NULL,break.time.by = NULL,
risk.table = T, tables.height = 0.32,
legend='top', ...) {

ggsurvplot(
  fit,conf.int = conf.int,
  surv.median.line = surv.median.line,
  ggtheme = ggtheme,
  xlim=xlim,
  risk.table = risk.table,
  tables.height = tables.height,
  break.time.by = break.time.by,
  legend=legend,
  ...
)
}

library(readr)
library(dplyr)

# install.packages("timereg")
library(timereg)

col_names <- c(
  "d",
  "age",
  "sex",
  "date0",
  "sbp0",
  "dbp0",
  "antihyp0",
  "date1",
  "sbp1",
  "dbp1",
  "antihyp1",
  "date2",
  "sbp2",
  "dbp2",
  "antihyp2",
  "date_event",
  "censor")

# c("0"="Sim", "1"="Não")
col_types <- cols(
  date0 = col_datetime("%d%b%y"),
  date1 = col_datetime("%d%b%y"),
  date2 = col_datetime("%d%b%y"),
  date_event = col_datetime("%d%b%y"),
  .default = col_double()
)

dados_raw <- readr::read_table(

```

```

"prova2/FRTCS.dat",
col_names = col_names,
col_types = col_types, na = c("."))
)

dados <- dados_raw %>%
  filter(!is.na(sbp2)) %>%
  mutate(
    sex = factor(sex, 1:2, c("male", "female")),
    antihyp0 = factor(antihyp0, 0:1, c("no", "yes")),
    antihyp1 = factor(antihyp1, 0:1, c("no", "yes")),
    antihyp2 = factor(antihyp2, 0:1, c("no", "yes")),
    tini0 = lubridate::ddays(0), # tempo entre follow ups
    tini1 = date1 - date0,
    tini2 = date2 - date0,
    tevent = date_event - date0,
    tfim0 = tini1,
    tfim1 = tini2,
    tfim2 = tevent,
    ageq = qcut(age, cuts = 4)
  )

dados_raw %>% filter(is.na(sbp2)) %>% knitr::kable()
dados$age %>% boxplot()
dados$age %>% summary()
dados$age %>% sd()
janitor::tabyl(dados$sex) %>% janitor::adorn_pct_formatting()
janitor::tabyl(dados$censor) %>% janitor::adorn_pct_formatting()
library(survival)
library(survminer)

fit <- survfit(Surv(tevent, censor)~ageq, dados)
print(fit)
survminer::ggsurvplot(fit)

fit <- survfit(Surv(tevent, censor)~sex, dados)
print(fit)
survminer::ggsurvplot(fit)

dados_long <- bind_rows(
  dados %>%
    dplyr::select(-dplyr::matches("\\d"),
                  dplyr::ends_with("0")) %>%
    rename_with(~gsub("\\d", "", .x), dplyr::ends_with("0")),
  dados %>% dplyr::select(-dplyr::matches("\\d"),
                          dplyr::ends_with("1")) %>%
    rename_with(~gsub("\\d", "", .x), dplyr::ends_with("1")),
  dados %>% dplyr::select(-dplyr::matches("\\d"),

```

```

        dplyr::ends_with("2")) %>%
  rename_with(~gsub("\\d", "", .x), dplyr::ends_with("2")),
  .id = "fup"
)

fit = coxph(Surv(tini, tfim, censor)~
            age + sex + sbp +dbp + antihyp,
            data = dados_long)

fit

summary(fit)
#retirando age
fit1 = coxph(Surv(tini, tfim, censor)~
            sex + sbp +dbp + antihyp,
            data = dados_long)
anova(fit1, fit)
#retirando sex
fit1 = coxph(Surv(tini, tfim, censor)~
            age + sbp +dbp + antihyp,
            data = dados_long)
anova(fit1, fit)
#retirando sbp
fit1 = coxph(Surv(tini, tfim, censor)~
            age+sex +dbp + antihyp,
            data = dados_long)
#retirando dbp
fit1 = coxph(Surv(tini, tfim, censor)~
            age+sex +sbp + antihyp,
            data = dados_long)
anova(fit1, fit)
#retirando antihyp
fit1 = coxph(Surv(tini, tfim, censor)~
            age+sex + sbp + dbp,
            data = dados_long)
anova(fit1, fit)

fitfinal = coxph(Surv(tini, tfim, censor)~
                sex + antihyp,
                data = dados_long)
fitfinal
fitfinal %>% summary()

```