

ANÁLISE DE SOBREVIVÊNCIA

Teoria e aplicações em saúde

Caderno de Respostas

Capítulo 10

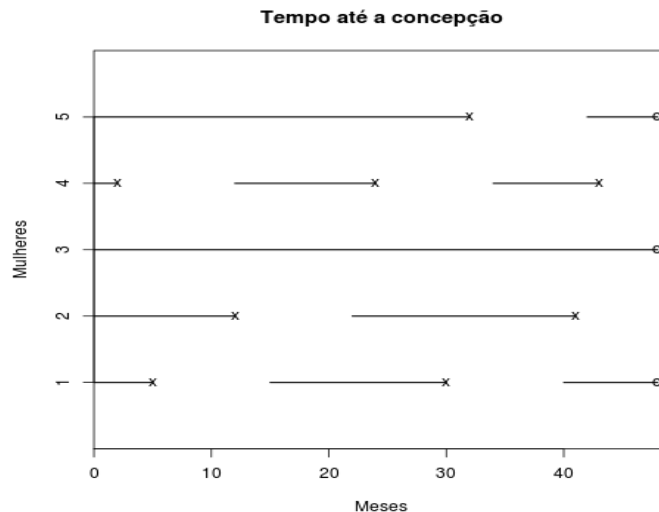
Eventos múltiplos

Exercício 10.1 Em um estudo de fertilidade, acompanhou-se uma coorte de mulheres por 48 meses para medir o tempo desde o parto até a próxima concepção, junto com covariáveis potencialmente associadas ao risco de engravidar. Para 5 mulheres, as seguintes datas de concepção foram registradas (em meses desde o primeiro parto):

| Mulher | Tempo até a próxima concepção (meses) | | | Estado civil | Contraceptivo |
|--------|---------------------------------------|-------------|-------------|--------------|---------------|
| | 2ª Gestação | 3ª Gestação | 4ª Gestação | | |
| 1 | 5 | 30 | - | Casada | Não |
| 2 | 12 | 41 | - | Casada | Não |
| 3 | - | - | - | Solteira | Sim |
| 4 | 2 | 24 | 43 | Casada | Não |
| 5 | 32 | - | - | Solteira | Sim |

Considerando que cada mulher retorna ao grupo sob risco de gravidez exatamente 10 meses após a concepção:

- a) Faça, à mão, um gráfico das trajetórias das cinco mulheres. Use o gráfico referente aos dados de reincidência de diarreia como exemplo.



- b) Concepção é um evento múltiplo ordenado. Neste capítulo, foram apresentados dois modelos marginais para eventos ordenados. Discuta os pressupostos assumidos em cada um deles para estimar o efeito das covariáveis indicativas do estado civil e do uso de contraceptivo no risco de engravidar.

Resposta:

No modelo para os eventos ordenados independentes (modelo AG) o risco é o mesmo para a ocorrência do 1º, 2º, ... ou para o iº evento. Já no modelo para os eventos ordenados estruturados (modelo PWP) o risco varia dependendo da ordem da ocorrência do evento. Assim considerando o exemplo podemos destacar para cada caso:

Eventos ordenados independentes (Modelo AG) – o risco de engravidar seria constante, após cada gravidez se retornaria ao grupo de risco, com o mesmo risco inicial. Este modelo é possível em população de ratas, durante a fase fértil. Em populações humanas, no século XXI e em áreas urbanas, entretanto, o risco de engravidar diminui substancialmente após uma ou duas gestações, em função tanto do uso de contraceptivos como do estado civil. Ou seja, a linha de base do risco tende a variar muito entre as gestações.

Eventos ordenados estruturado (Modelo PWP) – a cada gravidez a linha de base do risco muda. Ou seja, a ordem das gestações é importante na definição da linha de base, e por isso este modelo se chama "condicional". É o mais adequado ao estudo de concepção em populações humanas, seja devido à limitação voluntária da concepção, com o uso de anticoncepcionais, ou à diminuição da fertilidade ao longo do tempo.

- c) Organize uma planilha para esses dados de forma a poder ajustar um modelo de eventos independentes ordenados (modelo AG).

Resposta:

| Mulher | Início | Fim | Status |
|--------|--------|-----|--------|
| 1 | 0 | 5 | 1 |
| 1 | 15 | 30 | 1 |
| 1 | 40 | 48 | 0 |
| 2 | 0 | 12 | 1 |
| 2 | 22 | 41 | 1 |
| 3 | 0 | 48 | 0 |
| 4 | 0 | 2 | 1 |
| 4 | 12 | 24 | 1 |
| 4 | 34 | 43 | 1 |
| 5 | 0 | 32 | 1 |
| 5 | 42 | 48 | 0 |

Observe que entre o final de um período de risco e o início de outro decorrem 10 meses (9 meses da gestação mais um mês de resguardo).

- d) Organize uma planilha para esses dados, como eventos ordenados estruturados, para que se possa ajustar um modelo PWP.

Resposta:

| Mulher | Início | Fim | Status | Enum |
|--------|--------|-----|--------|------|
| 1 | 0 | 5 | 1 | 1 |
| 1 | 15 | 30 | 1 | 2 |
| 1 | 40 | 48 | 0 | 3 |
| 2 | 0 | 12 | 1 | 1 |
| 2 | 22 | 41 | 1 | 2 |
| 3 | 0 | 48 | 0 | 1 |

| | | | | |
|---|----|----|---|---|
| 4 | 0 | 2 | 1 | 1 |
| 4 | 12 | 24 | 1 | 2 |
| 4 | 34 | 43 | 1 | 3 |
| 5 | 0 | 32 | 1 | 1 |
| 5 | 42 | 48 | 0 | 2 |

Neste caso precisamos identificar a ordem de ocorrência de cada gravidez da mulher (evento).

Exercício 10.2 Um exemplo de evento ordenado é a reinternação hospitalar. Se o risco de reinternação for independente do fato de o indivíduo ter sofrido uma internação anterior, então o modelo de incrementos independentes é apropriado. Se esse pressuposto não for aceitável, então o modelo condicional é mais apropriado. O Apêndice 12.4 descreve um banco de dados (reint.csv) de pacientes que tiveram pelo menos uma internação em um hospital público regional.

- a) Ajuste um modelo de incrementos independentes, com as covariáveis sexo e idade.

Resposta:

```
# Exercicio 10.2
```

```
# Item a
```

```
reint <- read.table("reint.csv", header = T, sep = ",")
```

```
names(reint)
```

```
[1] "id"      "sexo"    "diasperm" "idade"    "numint"   "capcid"   "obitpos"
[8] "ini"     "fim"     "ini0"     "fim0"     "enum"     "status"   "tempo"
```

```
dim (reint)
```

```
[1] 21415    14
```

```
reint[1:10, ]
```

```
   id sexo diasperm idade numint   capcid obitpos ini  fim ini0 fim0 enum status tempo
1  1  1     1        26    52     4 CAPVIIIC  0  0   21   0   21   1     1     21
2  1  1     1         7    52     4 CAPVIIIC  0 28   69  28   69   2     1     41
3  1  1     1         5    52     4 CAPVIIIC  0 74  100  74  100   3     1     26
4  1  1     1        51    52     4 CAPVIIIC  0 151 1403 151 1403  4     0    1252
5  2  3     3         1    29     3  CAPVIIB  0  0  128   0  128   1     1     128
6  2  3     3         1    29     3   CAPXI  0 129  326 129  326   2     1     197
7  2  3     3         1    29     3  CAPVIIB  0 327 1428 327 1428  3     0    1101
8  4  3     3         6    56     4  CAPVIA  0  0  344   0  344   1     1     344
9  4  3     3         5    56     4  CAPVIA  0 349  378 349  378   2     1     29
10 4  3     3        13    56     4  CAPVIA  0 391 1313 391 1313  3     1     922
```

O banco está organizado com a repetição dos indivíduos para cada internação. O indivíduo 1, por exemplo, teve 4 períodos de segmentos com 3 internações, a primeira após 21 dias de acompanhamento e que teve duração de 7 dias (diasperm), a segunda ocorreu após 69 dias de acompanhamento (5 dias internado) e a terceira após 100 dias (51 dias internado). No último período de segmento não houve internação (censura).

Observamos que a variável Sexo aparece como numérica. É necessário fazer da variável sexo um fator, pois os valores são 1 e 3, se tratados como numérico estimarão valores errados.

```
# Transformado a variável sexo em fator
reint$sexo <- factor(reint$sexo, labels = c("masc", "fem"))
table (reint$sexo)

masc   fem
10714 10701

require(survival)
# Ajustando um modelo de eventos múltiplos – Incrementos Independentes - AG
# usando cluser para indicar as repetições de informação do indivíduo
reint.ag <- coxph(Surv(ini, fim, status) ~ idade + sexo + cluster(id), data = reint)
summary(reint.ag)
```

```
Call:
coxph(formula = Surv(ini, fim, status) ~ idade + sexo + cluster(id),
      data = reint)
```

```
n= 21415, number of events= 4734
```

```
              coef exp(coef)  se(coef) robust se      z Pr(>|z|)
idade    0.0076869 1.0077165 0.0007261 0.0009592 8.014 1.11e-15 ***
sexofem  0.1946315 1.2148633 0.0294041 0.0418080 4.655 3.23e-06 ***
```

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

```
              exp(coef) exp(-coef) lower .95 upper .95
idade            1.008      0.9923      1.006      1.010
sexofem          1.215      0.8231      1.119      1.319
```

```
Concordance= 0.557 (se = 0.005 )
Rsquare= 0.007 (max possible= 0.98 )
Likelihood ratio test= 152.2 on 2 df, p=0
Wald test = 98.78 on 2 df, p=0
Score (logrank) test = 155.1 on 2 df, p=0, Robust = 92.06 p=0
```

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

As variáveis sexo e idade são significativas. Para cada ano de idade o risco de internação aumenta 0,8% e em relação ao sexo masculino, as mulheres têm risco aumentado em 22%. Considerando que as internações do mesmo indivíduo são independentes.

b) Que pressuposto está sendo assumido ao se utilizar esse modelo?

Resposta:

Estamos assumindo que as reinternações sucessivas têm o mesmo risco de base. Ao sair de uma internação o paciente está de volta ao estado inicial.

c) Compare as estimativas de variância. Como você interpreta esse modelo?

Resposta:

O objetivo de comparar a variância do modelo clássico de Cox e a variância do modelo AG (variância robusta) é avaliar se são próximas (independência dos eventos múltiplos) ou não (dependência).

As estimativas das variâncias foram:

- para idade: $se(coef)=0,000727$ e $robust\ se=0,000891$, ou seja, a variância robusta é apenas 23% maior.
- para sexo: $se(coef)=0,014700$ e $robust\ se=0,019143$, ou seja, a variância robusta é apenas 30% maior.

A diferença entre as variâncias não foram muito grandes, indicando que a independência entre as internações de cada paciente poderia ser aceitável.

Na interpretação do modelo, considerando a independência entre as internações, teríamos:

- O modelo é significativamente diferente do modelo nulo ($p=0$). O teste da razão de verossimilhança (Likelihood ratio test) e Score (logrank) test assumem independência das internações no mesmo indivíduo.
- O modelo explica apenas 0,007 de um máximo de 0,98, ou seja, 0,82% da variabilidade total.
- A variável idade é significativa ($p = 0,000$) e a cada ano a mais de idade o risco de internar aumenta em 1%.
- O risco das mulheres se internarem é 22% maior que dos homens.

d) Você sugeriria outro modelo mais apropriado para descrever o risco de nova reinternação?

Resposta:

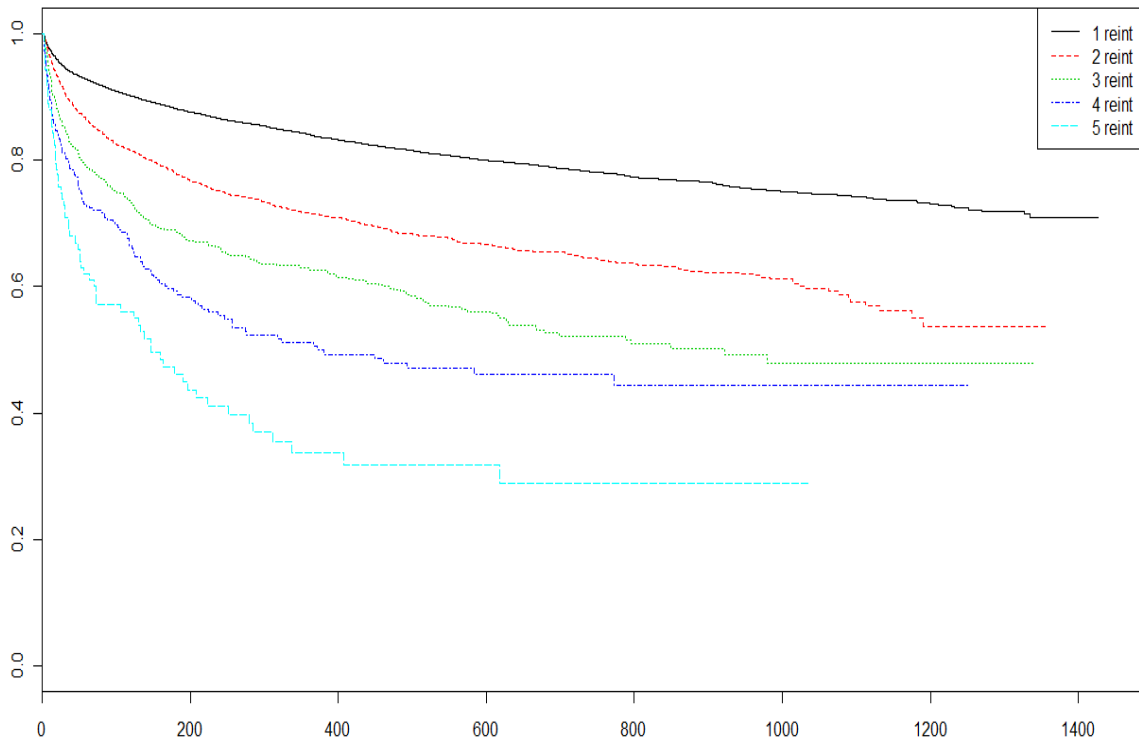
Sim, o modelo condicional, onde a linha de base do risco varia conforme as internações, poderia ser usado neste caso. A melhor forma de explorar esta possibilidade é verificando a estimativa não-paramétrica da sobrevivência por ordem da internação.

item d

Gráfico da sobrevivência por ordem da internação

```
plot(survfit(Surv(tempo, status) ~ enum, data = reint, subset = (enum < 6)), col = 1:5, lty = 1:5, c(100, 0.25), mark.time = F, main = "Tempo até a reinternação")
legend("topright", col = 1:5, lty = 1:5, legend=c("1 reint", "2 reint", "3 reint", "4 reint", "5 reint" ))
```

Tempo até a reinternação



Observe que no gráfico é visível a diferença no tempo até cada reinternação, diminuindo na medida em que aumenta o número de ordem da internação.

e) Ajuste um modelo condicional aos dados e interprete-o.

Resposta:

```
# item e
# Ajustando um modelo condicional – PWP (eventos ordenados)
# Usando strata para indicar a ordenação dos eventos
reint.pwp <- coxph(Surv(ini, fim, status) ~ idade + sexo + cluster(id) +
strata(enum), data = reint)
summary(reint.pwp)
```

```
Call:
coxph(formula = Surv(ini, fim, status) ~ idade + sexo + cluster(id) +
strata(enum), data = reint)
```

n= 21415, number of events= 4734

| | coef | exp(coef) | se(coef) | robust se | z | Pr(> z) |
|---------|-----------|-----------|-----------|-----------|-------|--------------|
| idade | 0.0075175 | 1.0075458 | 0.0007318 | 0.0007894 | 9.523 | < 2e-16 *** |
| sexofem | 0.1356800 | 1.1453153 | 0.0297847 | 0.0341717 | 3.971 | 7.17e-05 *** |

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

```
exp(coef) exp(-coef) lower .95 upper .95
```

```
idade      1.008      0.9925      1.006      1.009
sexofem    1.145      0.8731      1.071      1.225
```

```
Concordance= 0.558 (se = 0.006 )
Rsquare= 0.006 (max possible= 0.966 )
Likelihood ratio test= 119.9 on 2 df, p=0
Wald test = 110.7 on 2 df, p=0
Score (logrank) test = 122.5 on 2 df, p=0, Robust = 101.8 p=0
```

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

O modelo explica 0,06% da variabilidade total da sobrevivência, sendo melhor que o modelo nulo (p-valor do Wald teste = 0). As duas variáveis são significativas: para cada ano a mais de idade o risco de reinternar aumenta em 0,8%, e as mulheres internam 14,5% mais que os homens.

Exercício 10.3 Um pesquisador estuda fatores de risco associados à ocorrência de doenças oportunistas em pacientes com Aids (banco de dados oportunista.csv – seção Aids, doenças oportunistas do Apêndice). Para isso, ele acompanha uma coorte de pacientes na qual registrou a data do diagnóstico da Aids (que define o início do período de acompanhamento - ini) e as datas de ocorrência das doenças de interesse: alterações hematológicas (Hemato), herpes (Herpes), tuberculose (Tuberc), pneumocistose (Pneumocist) e candidíase (Cândida).

- a) Liste as primeiras 10 linhas e descreva o que aconteceu com os pacientes 1 a 4. Observe a variável status, cujo valor é sempre 1. O que isso quer dizer em termos da composição do grupo sob risco?

Resposta:

Exercício 10.3

Item a

```
oport <- read.table("oportunista.csv ", header = T, sep = ",")
oport[1:10, ]
```

```
      reg sex esc idade udi sexual ini      oport fim status tempo
1      1  F  1   45   0      0 3127  Candida 3141      1    14
2      2  M  2   38   1      0 2884   Herpes 3133      1   249
3      2  M  2   38   1      0 2884 Pneumocist 3133      1   249
4      2  M  2   38   1      0 2884  Candida 3343      1   459
5      3  F  4   34   1      0 3330   Hemato 3332      1     2
6      4  M  3   43   0      0 3780   Hemato 3964      1   184
7      5  F  3   32   0      1 2923  Candida 3588      1   665
8      5  F  3   32   0      1 2923 Pneumocist 3987      1  1064
9      5  F  3   32   0      1 2923   Hemato 5243      1  2320
10     6  M  2   33   0      0 2938 Pneumocist 4639      1  1701
```

Observando os dados do banco:

- Paciente 1 – sexo feminino, escolaridade menor do que 4 anos, com 45 anos, não é usuária de droga injetável (UDI) e o comportamento sexual não foi classificado como de risco, 14 dias após o diagnóstico de Aids apresentou candidíase. Não apresentou outras doenças durante o estudo.

- Paciente 2 – sexo masculino, ensino fundamental, 38 anos, UDI, 249 dias após o diagnóstico de Aids apresentou Herpes e pneumocistose, aos 459 dias após, candidíase.
- Paciente 3 – sexo feminino, curso superior, 34 anos, UDI, 2 dias após o diagnóstico de Aids apresentou alterações hematológicas.
- Paciente 4 – sexo masculino, ensino médio, 43 anos, comportamento sexual de risco, 184 dias após o diagnóstico de Aids apresentou alterações hematológicas.

O grupo de risco é composto de todos os pacientes em observação: ninguém sai do risco porque apresentou qualquer doença oportunista. Um paciente só sai do grupo de risco por óbito ou perda de acompanhamento. Neste estudo os eventos poderem ocorrer ao mesmo tempo (paralelos).

- b) Ajuste um modelo separadamente para os desfechos Hemato, Herpes e Cândida. Além de sexo e idade, estime o efeito das covariáveis relacionadas a comportamento de risco (udi e sexual)

Resposta:

```
# Item b
# modelo com doenças Hematológicas
hem.fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex + idade, data = oport,
subset = (oport == "Hemato"))
summary(hem.fit)
```

```
Call:
coxph(formula = Surv(tempo, status) ~ udi + sexual + sex + idade,
      data = oport, subset = (oport == "Hemato"))
```

n= 477, number of events= 477

| | coef | exp(coef) | se(coef) | z | Pr(> z) |
|--------|-----------|-----------|----------|--------|----------|
| udi | -0.120886 | 0.886135 | 0.201968 | -0.599 | 0.5495 |
| sexual | -0.302527 | 0.738948 | 0.171573 | -1.763 | 0.0779 . |
| sexM | 0.191416 | 1.210963 | 0.096727 | 1.979 | 0.0478 * |
| idade | 0.009323 | 1.009367 | 0.005033 | 1.852 | 0.0640 . |

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

| | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|--------|-----------|------------|-----------|-----------|
| udi | 0.8861 | 1.1285 | 0.5965 | 1.316 |
| sexual | 0.7389 | 1.3533 | 0.5279 | 1.034 |
| sexM | 1.2110 | 0.8258 | 1.0018 | 1.464 |
| idade | 1.0094 | 0.9907 | 0.9995 | 1.019 |

```
Concordance= 0.53 (se = 0.015 )
Rsquare= 0.03 (max possible= 1 )
Likelihood ratio test= 14.59 on 4 df, p=0.005626
Wald test = 13.96 on 4 df, p=0.007426
Score (logrank) test = 14.04 on 4 df, p=0.007159
```

Para a ocorrência de complicações hematológicas os únicos fatores de risco significativo é ser do sexo masculino e idade. Os demais comportamentos de risco não são variáveis significativas no risco de complicações hematológicas. O modelo explica apenas 3% mas é significativamente diferente do modelo nulo.

modelo com Herpes

```
her.fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex + idade, data = oport,
subset = (oport == "Herpes"))
summary(her.fit)
```

Call:

```
coxph(formula = Surv(tempo, status) ~ udi + sexual + sex + idade,
data = oport, subset = (oport == "Herpes"))
```

n= 102, number of events= 102

| | coef | exp(coef) | se(coef) | z | Pr(> z) |
|--------|-----------|-----------|----------|--------|----------|
| udi | -0.679818 | 0.506709 | 0.511267 | -1.330 | 0.184 |
| sexual | -0.194769 | 0.823025 | 0.497997 | -0.391 | 0.696 |
| sexM | -0.162708 | 0.849839 | 0.207808 | -0.783 | 0.434 |
| idade | 0.005714 | 1.005730 | 0.012024 | 0.475 | 0.635 |

| | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|--------|-----------|------------|-----------|-----------|
| udi | 0.5067 | 1.9735 | 0.1860 | 1.380 |
| sexual | 0.8230 | 1.2150 | 0.3101 | 2.184 |
| sexM | 0.8498 | 1.1767 | 0.5655 | 1.277 |
| idade | 1.0057 | 0.9943 | 0.9823 | 1.030 |

Concordance= 0.549 (se = 0.034)

Rsquare= 0.041 (max possible= 0.999)

Likelihood ratio test= 4.31 on 4 df, p=0.3662

Wald test = 3.7 on 4 df, p=0.4475

Score (logrank) test = 3.84 on 4 df, p=0.4287

Nenhuma das variáveis é significativa em relação ao Herpes e o modelo não difere do modelo nulo.

modelo com Candidíase

```
can.fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex + idade, data = oport,
subset = (oport == "Candida"))
summary(can.fit)
```

Call:

```
coxph(formula = Surv(tempo, status) ~ udi + sexual + sex + idade,
data = oport, subset = (oport == "Candida"))
```

n= 381, number of events= 381

| | coef | exp(coef) | se(coef) | z | Pr(> z) |
|--------|-----------|-----------|----------|--------|----------|
| udi | -0.001136 | 0.998864 | 0.205791 | -0.006 | 0.9956 |
| sexual | -0.218972 | 0.803344 | 0.179234 | -1.222 | 0.2218 |
| sexM | 0.065779 | 1.067991 | 0.109791 | 0.599 | 0.5491 |
| idade | 0.011931 | 1.012002 | 0.005392 | 2.213 | 0.0269 * |

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

| | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|--------|-----------|------------|-----------|-----------|
| udi | 0.998864 | 1.001136 | 0.998864 | 1.001136 |
| sexual | 0.803344 | 1.244888 | 0.641888 | 1.001136 |
| sexM | 1.067991 | 0.936388 | 0.936388 | 1.067991 |
| idade | 1.012002 | 0.987998 | 0.987998 | 1.012002 |

```

udi          0.9989      1.0011      0.6673      1.495
sexual       0.8033      1.2448      0.5654      1.141
sexM         1.0680      0.9363      0.8612      1.324
idade        1.0120      0.9881      1.0014      1.023

```

```

Concordance= 0.543 (se = 0.017 )
Rsquare= 0.019 (max possible= 1 )
Likelihood ratio test= 7.46 on 4 df, p=0.1136
Wald test = 7.47 on 4 df, p=0.1129
Score (logrank) test = 7.49 on 4 df, p=0.1121

```

Em relação a candidíase, nenhuma das variáveis é significativa, e o modelo não difere do modelo nulo.

- c) Compare o modelo acima com uma abordagem alternativa, na qual o efeito das covariáveis é modelado simultaneamente para os três desfechos, porém permitindo um risco basal diferente para cada doença oportunista.

Resposta:

```

# Item c
# modelo com três desfechos e todas as doenças oportunistas
todas.fit <- coxph(Surv(tempo, status) ~ udi + sexual + sex + idade + cluster(reg)
+ strata(oport), data = oport, )

```

```

Call:
coxph(formula = Surv(tempo, status) ~ udi + sexual + sex + idade +
      cluster(reg) + strata(oport), data = oport)

```

```

n= 1195, number of events= 1195

```

```

      coef exp(coef) se(coef) robust se      z Pr(>|z|)
udi    -0.145726  0.864395  0.119256  0.179444 -0.812  0.417
sexual -0.270660  0.762876  0.110938  0.144493 -1.873  0.061 .
sexM    0.089012  1.093094  0.062204  0.078845  1.129  0.259
idade   0.008501  1.008537  0.003159  0.003712  2.290  0.022 *
---

```

```

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

```

```

      exp(coef) exp(-coef) lower .95 upper .95
udi          0.8644      1.1569      0.6081      1.229
sexual       0.7629      1.3108      0.5747      1.013
sexM         1.0931      0.9148      0.9366      1.276
idade        1.0085      0.9915      1.0012      1.016

```

```

Concordance= 0.527 (se = 0.018 )
Rsquare= 0.017 (max possible= 1 )
Likelihood ratio test= 20.6 on 4 df, p=0.0003802
Wald test = 11.89 on 4 df, p=0.01819
Score (logrank) test = 19.8 on 4 df, p=0.0005465, Robust = 9.64
p=0.04702

```

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

O modelo agora é significativamente diferente do modelo nulo pelo teste de Wald, teste score robusto e no teste da verossimilhança. Apenas a categoria de idade é significativa. O coeficiente de explicação é de apenas 1,7%.