

Lista 4 - MAE0514

Guilherme N^oUSP: 8943160 e Leonardo N^oUSP: 9793436

Exercício 1

Abaixo estão listadas as referências de dois artigos (os arquivos estão disponíveis):

- (i) Gonzalez, A. B. et al. (2010). Body-Mass Index and Mortality among 1.46 Million White Adults, New England Journal of Medicine 363(23): 2211-2219.
- (ii) He, J., McGee, D. , Niu, X. e Choi, W. (2009). Examining the Dynamic Association of BMI and Mortality in the Framingham Heart Study , International Journal of Environmental Research and Public Health 6: 3115-3126.

Leia os dois artigos e faça um breve resumo dos artigos, enfocando principalmente:

- Objetivo do estudo;
- Descrição da forma como os dados utilizados foram coletados;
- Metodologia estatística utilizada na análise dos dados;
- Principais resultados obtidos e conclusões da análise estatística.

Faça uma breve comparação das principais conclusões dos dois artigos e comente.

Resolução

O primeiro artigo Body-Mass Index and Mortality among 1.46 Million White Adults, tem como objetivo estudar a associação do índice de massa corporal e o aumento na mortalidade por doenças cardiovasculares, dessa forma os dados provém de 19 estudos prospectivos envolvendo 1,46 milhão de adultos brancos, de 19 a 84 anos de idade com o seguinte critério de seleção: o estudo que teve mais de 5 anos de acompanhamento, houve mais de 1000 mortes entre não-hispânicos participantes que eram brancos, e o ano de referência foi de 1970 em diante. Os estudos também devem ter apurado altura, peso e tabagismo. A metodologia utilizada foi a regressão de Cox para estimar taxas de risco e intervalos de confiança de 95% para associação entre IMC e mortalidade por todas as causas, com as covariáveis: idade, estudo, atividade, consumo de álcool, educação e estado civil, assim resultando em geral que em adultos brancos, com sobrepeso e obesidade ou que estão abaixo do peso estão associados com aumento da mortalidade por todas as causas. A mortalidade por todas as causas é geralmente mais baixa com um IMC de 20,0 a 24,9.

O segundo artigo examining the Dynamic Association of BMI and Mortality in the Framingham Heart Study, tem como objetivo estudar também estudar a associação do índice de massa corporal e o aumento na mortalidade por todas as causas, porém mostrando que o IMC e a mortalidade podem não ser estáticos, sendo assim os dados provém de múltiplos estudos coortes, porém foi utilizado o estudo da ESF (Framingham Heart Study) com 5209 homens e mulheres com idades entre 30 e 62 anos, de Framingham, Massachusetts. Neste estudo, foi utilizado as informações dos 20 primeiros exames (cerca de 40 anos de acompanhamento) da coorte da ESF para análise, sendo avaliados idade, tabagismo, IMC entre outras variáveis. Uma das técnicas utilizadas foi a regressão logística com o intuito de verificar a mortalidade em relação as outras covariáveis o resultado que a análise mostra é resultado da associação entre o IMC e mortalidade podem diferir quando o desenho de um estudo muda, como usar o IMC medido em um tempo ou alterar a duração do acompanhamento. Também foi utilizado um modelo dinâmico de sobrevivência que é muito parecido

com o modelo de cox, porém não pressupões riscos proporcionais, e chegando a um resultado que o recurso dinâmico é mais forte para homens do que para mulheres.

Fazendo uma breve comparação entre os estudos um mostra que que existe associação entre o IMC e mortes por doenças cardiovasculares, já no outro estudo foi feita uma comparação mais geral entre o IMC e mortes por causas mais gerais, o interessante desse estudo é que foi feita comparação entre duas técnicas sendo que a técnica mais complexa (modelo dinâmico de sobrevivência) obteve um melhor resultado dado que a suposição de riscos proporcionais “na vida real” é constantemente violada, porém temos o contra ponto de que o estudo ficou muito restrito a amostragem do estudo coorte, inviabilizando a reprodutibilidade do mesmo.

Exercício 2

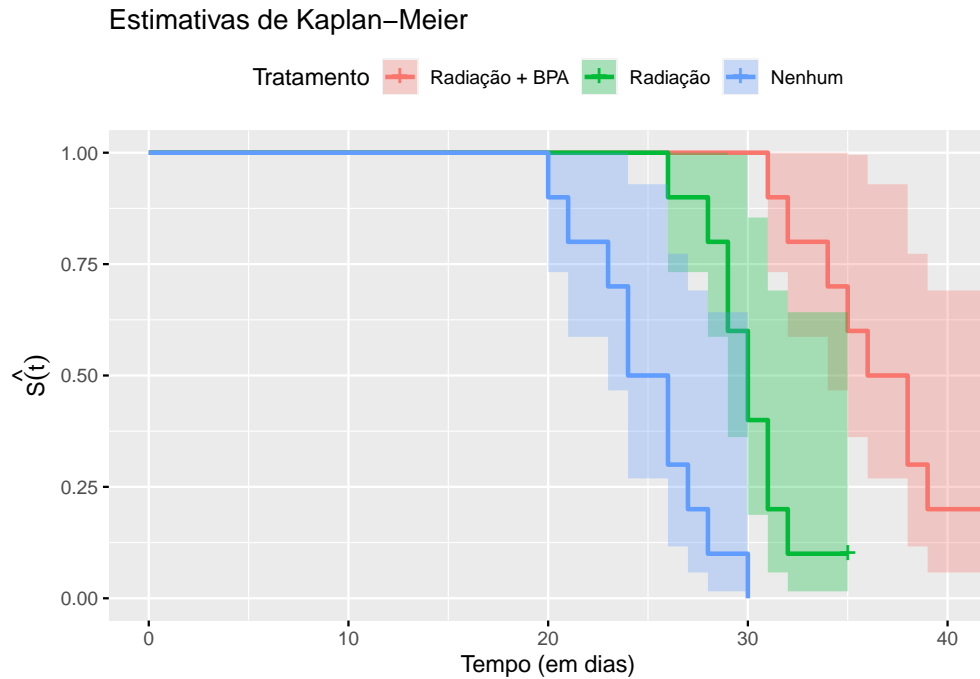
Um estudo foi conduzido para determinar a eficiência de uma terapia conhecida como BNCT (Boron Neutron Capture Therapy), usando BPA (boronophenylalanine) como agente de captura, no tratamento de glioma F98, que é um tipo de tumor no sistema nervoso central, nas células gliais. Células com glioma F98 foram implantadas no cérebro de ratos, que foram divididos em três grupos. O primeiro grupo não recebeu tratamento, o segundo grupo foi tratado apenas com radioterapia e, por fim, o terceiro grupo recebeu radiação e também uma dose apropriada de BPA. Os dados disponíveis são os tempos de vida (em dias) dos ratos em cada um dos três grupos e estão apresentados na tabela abaixo (tempos censurados à direita estão denotados por um sinal “+”).

Sem tratamento	Radiação	Radiação + BPA
20	26	31
21	28	32
23	29	34
24	29	35
24	30	36
26	30	38
26	31	38
27	31	39
28	32	42 ⁺
30	35 ⁺	42 ⁺

- (a) Obtenha e compare as curvas de sobrevivência dos três grupos, utilizando algum teste apropriado. Comente.

Resolução

Para comparar os tratamentos, é feito um gráfico de Kaplan Meier:



Estabelecendo as hipóteses do teste de log-rank, temos:

$$\begin{cases} H_0 : S_1(t) = S_2(t) = S_3(t) \\ H_1 : \text{pelo menos uma curva é diferente} \end{cases}$$

Obtendo o seguinte resultado:

variable	pval	method
trat	1e-07	Log-rank

A partir do gráfico, nota-se que a proporção de sobreviventes dos pacientes que não receberam nenhum tipo de tratamento, aparentemente, decresce mais rapidamente que a proporção dos pacientes que utilizaram a radiação, que por sua vez, decresce mais rapidamente do que os pacientes que utilizaram a radiação e o BPA. O teste Log-rank aponta um p-valor menor que 0.0001, logo rejeita-se a hipótese nula, ao nível de significância de 5%, logo há evidências que há diferença entre os tratamentos.

- (b) Crie duas variáveis binárias Z_1 (igual a 1 se o animal recebeu radiação apenas e igual a 0, caso contrário) e Z_2 (igual a 1 se o animal recebeu radiação e BPA e igual a 0, caso contrário). Considerando o modelo semiparamétrico de riscos proporcionais, obtenha estimativas dos coeficientes associados com cada variável criada e dos respectivos erros padrão utilizando três métodos diferentes para empates. Compare os resultados.

Resolução

Para obter as estimativas dos coeficientes associados com cada variável criada e dos respectivos erros padrão, devido aos empates, utiliza-se as aproximações de Efron, Breslow e Cox/Exato.

Primeiramente, pela aproximação de Efron:

```
## Call:
## coxph(formula = Surv(tempo, delta) ~ Z1 + Z2, data = data, ties = "efron")
##
##      n= 30, number of events= 27
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Z11 -1.92238    0.14626  0.56670 -3.392 0.000693 ***
## Z21 -3.76323    0.02321  0.76612 -4.912 9.01e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## Z11  0.14626      6.837   0.04817   0.4441
## Z21  0.02321     43.087   0.00517   0.1042
##
## Concordance= 0.819 (se = 0.018 )
## Likelihood ratio test= 29.93 on 2 df,  p=3e-07
## Wald test              = 24.52 on 2 df,  p=5e-06
## Score (logrank) test = 35.2 on 2 df,  p=2e-08
```

Pela aproximação de Breslow:

```
## Call:
## coxph(formula = Surv(tempo, delta) ~ Z1 + Z2, ties = "breslow")
##
##      n= 30, number of events= 27
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Z11 -1.81197    0.16333  0.55971 -3.237 0.00121 **
## Z21 -3.55737    0.02851  0.75825 -4.692 2.71e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## Z11  0.16333      6.123   0.054531   0.4892
## Z21  0.02851     35.071   0.006451   0.1260
##
## Concordance= 0.819 (se = 0.018 )
## Likelihood ratio test= 27.37 on 2 df,  p=1e-06
## Wald test              = 22.45 on 2 df,  p=1e-05
## Score (logrank) test = 31.74 on 2 df,  p=1e-07
```

E por fim, pela aproximação de Cox/Exato:

```
## Call:
## coxph(formula = Surv(tempo, delta) ~ Z1 + Z2, ties = "exact")
##
##      n= 30, number of events= 27
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Z11 -2.28322    0.10196  0.71469 -3.195   0.0014 **
## Z21 -4.23235    0.01452  0.90716 -4.665 3.08e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## Z11  0.10196      9.808  0.025123  0.41376
## Z21  0.01452     68.879  0.002453  0.08592
##
## Concordance= 0.819  (se = 0.018 )
## Likelihood ratio test= 30.78  on 2 df,   p=2e-07
## Wald test               = 21.77  on 2 df,   p=2e-05
## Score (logrank) test = 33.38  on 2 df,   p=6e-08
```

Observa-se que as aproximações apontam valores diferentes, entretanto, é possível obter a mesma interpretação para as três aproximações. Os pacientes que utilizam o tratamento de radiação, o risco de óbito diminui em 85%, aproximadamente. Já os pacientes que são tratados com a radiação e o BPA, o risco de óbito diminui em 98%.

- (c) Ainda considerando as variáveis criadas no item (b), teste a hipótese global de que não há efeito de nenhum tratamento (ou seja, $H_0 : \beta_1 = \beta_2 = 0$), utilizando o teste da razão de verossimilhanças, para as três aproximações para empates usadas.

Resolução

Para as três aproximações, observa-se valores da estatística teste entre 27 e 31, resultando em p-valores menores que 0.0001, desse modo as hipóteses nulas são rejeitas, logo não há evidências que os parâmetros são iguais a 0 ou que os tratamentos não possuem diferenças. Nota-se que com a aproximação de Cox/Exato, a hipótese é rejeitada com mais poder de teste e a aproximação de Breslow, com menos poder de teste.

- (d) Repita o item (c), usando o teste de Wald.

Resolução

Observa-se que os valores da estatística teste estão entre 21 e 25, resultando em p-valores menores que 0.0001, desse modo as hipóteses nulas são rejeitas, logo não há evidências que os parâmetros são iguais a 0 ou que os tratamentos não possuem diferenças. Nota-se que com a aproximação de Efron, a hipótese é rejeitada com mais poder de teste e a aproximação de Cox/Exato, com menos poder de teste.

- (e) Teste a hipótese de que o efeito da radiação e da radiação com BPA são iguais (ou seja, $H_0 : \beta_1 = \beta_2$), para as aproximações para empates consideradas, usando o teste da razão de verossimilhanças.

Resolução

Considerando o modelo com a aproximação de Breslow, queremos testar:

$$\begin{cases} H_0 : \beta_1 = \beta_2 \\ H_1 : \beta_1 \neq \beta_2 \end{cases}$$

Utilizando o teste de razão de verossimilhanças, temos a seguinte estatística de teste:

$$\lambda = -2 \ln \left(\frac{L(\hat{\beta})}{L(\hat{\beta}_1, \hat{\beta}_2)} \right) = -2[L(\ln(\hat{\beta})) - L(\ln(\hat{\beta}_1, \hat{\beta}_2))] = -2 \ln(L(\hat{\beta})) + 2 \ln(L(\hat{\beta}_1, \hat{\beta}_2))$$

e como $\lambda \stackrel{a}{\sim} \chi^2_{(2-1)}$, logo:

$$\lambda = -2 * (-64.78) + 2 * (-59.33) = 10.8913$$

Calculando o p-valor é dado por $\mathbb{P}(\chi^2_1 > \lambda) = \mathbb{P}(\chi^2_1 > 10.8913) = 0.001$

Logo, temos evidências para rejeitar H_0 ao nível de significância de 5%.

Exercício 3

Considere um estudo com pacientes com câncer de ovário, em estágio mais avançado. Todos os pacientes foram submetidos a um tratamento padrão e observou-se, para cada paciente, o tempo até sua morte, em dias. No momento do início do tratamento, os pacientes foram separados em dois grupos: pacientes com tumor grande e pacientes com tumor moderado. Deseja-se avaliar o efeito do tamanho do tumor na sobrevivência dos pacientes.

- (a) Crie uma variável binária adequada para indicação do grupo e escreva a forma do modelo semi-paramétrico de Cox.

Resolução

Criando a variável binária a partir do grupo, temos:

```
data3$Tumor <- ifelse(data3$Tumor=="grande",1,0)
```

Ou seja:

$$Grupo = \begin{cases} 1 & \text{se Tumor grande} \\ 0 & \text{se Tumor moderado} \end{cases}$$

E o modelo semiparamétrico de Cox é escrito como:

$$\alpha(t|x_i) = \alpha_0(t)e^{x'_i\beta}$$

E para o problema em específico, com apenas uma covariável binária, pode ser escrito como:

$$\alpha(t) = \begin{cases} \alpha_0(t)e^{\beta} & \text{se } x = 1 \\ \alpha_0(t) & \text{se } x = 0 \end{cases}$$

Em que $\alpha_0(t)$ é a função de taxa de falha de referência e x é variável de grupo.

(b) Assumindo que não há empates, escreva a forma da verossimilhança parcial de Cox para este problema.

Resolução

Como temos apenas uma covariável binária e assumindo que não há empates, a forma da verossimilhança parcial de Cox é dada por:

$$L(\beta) = \prod_{i=1}^{22} \frac{e^{\beta x_i}}{\sum_{j=1}^{n_i} e^{\beta x_j}} = \prod_{i=1}^{35} \left(\frac{e^{\beta x_i}}{\sum_{j=1}^{n_i} e^{\beta x_j}} \right)^{\delta_i}$$

Em que x_i é uma variável binária, assim como definido anteriormente, 1 se tumor grande, 0 caso contrário, $n_i = n_{1i} + n_{2i}$ indivíduos no conjunto de risco ($R(t_i)$) no tempo t_i e n_{1i}, n_{2i} sendo os indivíduos no conjunto de risco dos grupos 1 e 2 respectivamente e por fim δ_i é indicador do evento.

(c) Utilizando a expressão obtida em (b), encontre o escore.

Resolução

O vetor escore, pode ser obtido por:

$$U(\beta) = \frac{\partial \ln(L(\beta))}{\partial \beta}$$

Aplicando o $\ln(L(\beta))$, temos:

$$l(\beta) = \ln(L(\beta)) = \sum_{i=1}^{22} \beta x_i - \sum_{i=1}^{22} \ln \left(\sum_{j=1}^{n_i} e^{\beta x_j} \right)$$

E também quando $x_i = 0$ temos indivíduos do grupo 2 (tumor moderado), logo $\sum_{i=1}^{22} x_i = \sum_{i=1}^{22} d_{1i} = d_1$ que é o total de óbitos no grupo 1 (tumor grande), então neste caso temos:

$$\sum_{j=1}^{n_i} e^{\beta x_j} = n_{1i} e^{\beta} + n_{2i}$$

Resultando em:

$$l(\beta) = \ln(L(\beta)) = \sum_{i=1}^{22} \beta d_{1i} - \sum_{i=1}^{22} \ln(n_{1i} e^{\beta} + n_{2i})$$

Mas como o caso tratado é de apenas uma covariável, o vetor escore se resume em:

$$I(\beta) = \frac{\partial l(\beta)}{\partial \beta}$$

Logo:

$$U(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{22} \left[d_{1i} - \frac{n_{1i} e^{\beta}}{n_{1i} e^{\beta} + n_{2i}} \right] = \sum_{i=1}^{35} \delta_i \left[d_{1i} - \frac{n_{1i} e^{\beta}}{n_{1i} e^{\beta} + n_{2i}} \right]$$

(d) Utilizando a expressão obtida em (b), encontre a expressão para a informação observada.

Resolução

A matriz de informação observada é dada por:

$$I(\beta) = -\frac{\partial^2 \beta}{\partial \beta \partial \beta^T}$$

Mas como o caso tratado é de apenas uma covariável, a informação observada se resume em:

$$I(\beta) = -\frac{\partial^2 l(\beta)}{\partial \beta^2} = -U'(\beta)$$

Assim temos:

$$I(\beta) = -\left(\sum_{i=1}^{22} \left[\frac{-(n_{1i}e^\beta + n_{2i})n_{1i}e^\beta + (n_{1i}e^\beta)^2}{(n_{1i}e^\beta + n_{2i})^2} \right]\right) = \sum_{i=1}^{22} \left[\frac{(n_{1i}e^\beta + n_{2i})n_{1i}e^\beta - (n_{1i}e^\beta)^2}{(n_{1i}e^\beta + n_{2i})^2} \right] = \sum_{i=1}^{22} \frac{n_{1i}n_{2i}e^\beta}{(n_{1i}e^\beta + n_{2i})^2}$$

(e) Suponha que deseja-se testar a hipótese $H_0 : \beta = 0$, em que β é o parâmetro no modelo de Cox associado à covariável binária criada anteriormente. Para realizar este teste, pode-se utilizar a estatística do teste do escore. Escreva a expressão para a estatística do escore seguindo os seguintes passos:

- i. Com base no item (c), obtenha $U(0)$, ou seja, a expressão para o escore avaliado no ponto $\beta = 0$. Simplifique a expressão obtida, escrevendo-a em função do número de indivíduos em risco em cada grupo.

Resolução

Com base no item (c) $U(0)$ pode ser escrito como:

$$U(0) = \sum_{i=1}^{35} \delta_i \left[d_{1i} - \frac{n_{1i}e^0}{n_{1i}e^0 + n_{2i}} \right] = \sum_{i=1}^{35} \delta_i \left[d_{1i} - \frac{n_{1i}}{n_{1i} + n_{2i}} \right] = \sum_{i=1}^{35} \delta_i \left[d_{1i} - \frac{n_{1i}}{n_i} \right]$$

Em que $n_i = n_{1i} + n_{2i}$ indivíduos no conjunto de risco ($R(t_i)$) no tempo t_i e n_{1i}, n_{2i} sendo os indivíduos no conjunto de risco dos grupos 1 e 2 respectivamente.

- ii. Obtenha $I(0)$, ou seja, a expressão para a informação avaliada no ponto $\beta = 0$. Simplifique a expressão obtida, escrevendo-a em função do número de indivíduos em risco em cada grupo.

Resolução

Com base no item (d) $I(0)$ pode ser escrito como:

$$I(0) = \sum_{i=1}^{35} \delta_i \left[\sum_{i=1}^{22} \frac{n_{1i}n_{2i}e^0}{(n_{1i}e^0 + n_{2i})^2} \right] = \sum_{i=1}^{35} \delta_i \left[\sum_{i=1}^{22} \frac{n_{1i}n_{2i}}{(n_{1i} + n_{2i})^2} \right]$$

- iii. Obtenha a estatística do teste do escore, dada por

$$S = \frac{U^2(0)}{I(0)},$$

que, sob $H_0 : \beta = 0$, tem distribuição de qui-quadrado com 1 grau de liberdade.

Resolução

Como:

$$S = \frac{U^2(0)}{I(0)} = \frac{\left[\sum_{i=1}^{35} \delta_i \left(d_{1i} - \frac{n_{1i}}{n_i} \right) \right]^2}{\sum_{i=1}^{35} \delta_i \left[\frac{n_{1i} * n_{2i}}{(n_{1i} + n_{2i})^2} \right]}$$

Em que podemos notar que é a mesma estatística é muito parecida com a do teste do teste log-rank, e na literarura podemos encontrar que o teste de escore do modelo de cox coincide com o teste de log-rank.

Exercício 4

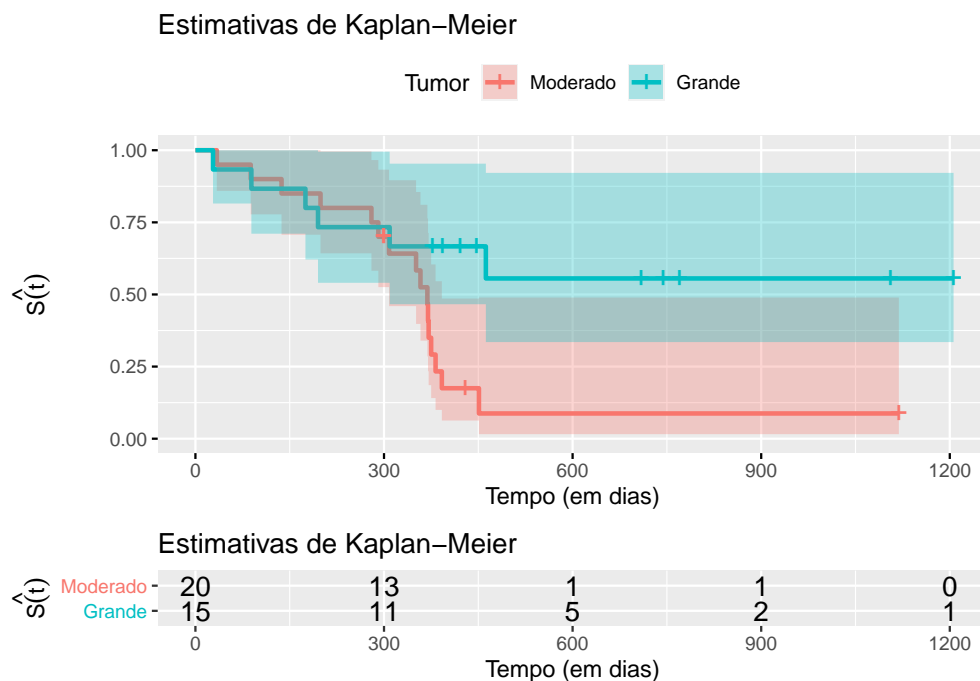
Considere a situação descrita no exercício 3. Esse problema corresponde a um estudo realizado com 35 mulheres, sendo observados os seguintes tempos (tempos censurados à direita estão denotados por um sinal “+”):

Grupo	Tempos
Tumor grande	28, 89, 175, 195, 309, 377 ⁺ , 393 ⁺ , 421 ⁺ , 447 ⁺ , 462, 709 ⁺ , 744 ⁺ , 770 ⁺ , 1106 ⁺ , 1206 ⁺
Tumor moderado	34, 88, 137, 199, 280, 291, 299 ⁺ , 300 ⁺ , 308, 351, 358, 369, 370, 371, 375, 382, 392, 429 ⁺ , 451, 1119 ⁺

- (a) Utilizando um pacote estatístico, obtenha as curvas de Kaplan-Meier de cada um dos grupos. Apresente as duas curvas no mesmo gráfico e compare.

Resolução

Utilizando um R como pacote estatístico as curvas de Kaplan-Meier para cada um dos grupos é:



Em que pode-se notar que até 300 dias de acompanhamento as curvas são bem próximas, com diversos cruzamentos, a partir de então a proporção de sobreviventes de pacientes com o tumor moderado decresce rapidamente. Entretanto, nota-se muitas censuras para pacientes com tumores grandes a partir de 300 dias de acompanhamento.

- (b) Teste a igualdade das curvas utilizando o teste de log-rank. Obtenha também um teste da família de Fleming-Harrington. Conclua.

Resolução

O teste Log-rank obtem um p-valor de 0.018. Logo rejeita-se a hipótese nula de que as curvas são idênticas.

variable	pval	method
Tumor	0.018	Log-rank

O teste da família de Fleming-Harrington obtem um p-valor de 0.0018. Logo rejeita-se a hipótese nula de que as curvas são idênticas.

variable	pval	method
Tumor	0.0018	Fleming-Harrington (p=1, q=1)

- (c) Ajuste um modelo de riscos proporcionais de Cox. Obtenha a estimativa de β , bem como o seu erro padrão. Interprete o coeficiente β .

Resolução

```
## Call:
## coxph(formula = Surv(Tempo, delta) ~ Tumor, data = data3)
##
##      n= 35, number of events= 22
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Tumor -1.1283     0.3236   0.4969 -2.271   0.0232 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Tumor    0.3236      3.091    0.1222    0.8569
##
## Concordance= 0.588 (se = 0.064 )
## Likelihood ratio test= 5.83  on 1 df,  p=0.02
## Wald test            = 5.16  on 1 df,  p=0.02
## Score (logrank) test = 5.62  on 1 df,  p=0.02
```

A partir do modelo ajustado, pode-se dizer que pacientes com tumores grandes, tem um risco de óbito aumentado em 32% do que quem tem tumor moderado.

- (d) Teste a hipótese $H_0 : \beta = 0$ utilizando a estatística do teste de Wald, calculada com base nas estimativas de β e de seu erro-padrão obtidos em (c).

Resolução

$$W = \frac{\hat{\beta}}{se(\hat{\beta})} = \frac{-1.1283}{0.4969} = -2.271$$

Pode-se notar que a partir do teste de Wald, com um p-valor de 0.02, rejeita-se a hipótese nula de que o parâmetro é igual a zero ou que os tempos são idênticos para pacientes com tumores grandes e moderados.

- (e) Teste a hipótese $H_0 : \beta = 0$ utilizando a estatística do teste do escore obtida no exercício 3.e.iii (sem a utilização de software estatístico). Compare com os itens anteriores, comente e conclua.

Resolução

Utilizando a estatística do teste do escore obtida no exercício 3.e.iii, temos:

$$S = \frac{U^2(0)}{I(0)} = \frac{\left[\sum_{i=1}^{35} \delta_i \left(d_{1i} - \frac{n_{1i}}{n_{1i} + n_{2i}} \right) \right]^2}{\sum_{i=1}^{35} \delta_i \left[\frac{n_{1i} * n_{2i}}{(n_{1i} + n_{2i})^2} \right]}$$

Com o auxílio de uma planilha (em anexo):

$$S = \frac{U^2(0)}{I(0)} = \frac{(-5.384)^2}{5.16} = \frac{28.987}{5.16} = 5.6177$$

E com parando com S com a distribuição χ_1^2 , temos um p-valor de: $\mathbb{P}(\chi_1^2 > S) = \mathbb{P}(\chi_1^2 > 5.6177) = 0.0177$

Logo assim como o resultado do teste log-rank, temos evidência estatística para rejeitar a hipótese de que $\beta = 0$.

Exercício 5

Considere o mesmo conjunto de dados da segunda lista de exercícios, disponíveis no arquivo Lista2-hodgkins.xlsx, referentes 60 pacientes com doença de Hodgkins que receberam tratamento padrão para a doença. O tempo de vida (em meses), bem como idade, sexo, histologia e estágio da doença de cada paciente foram observados.

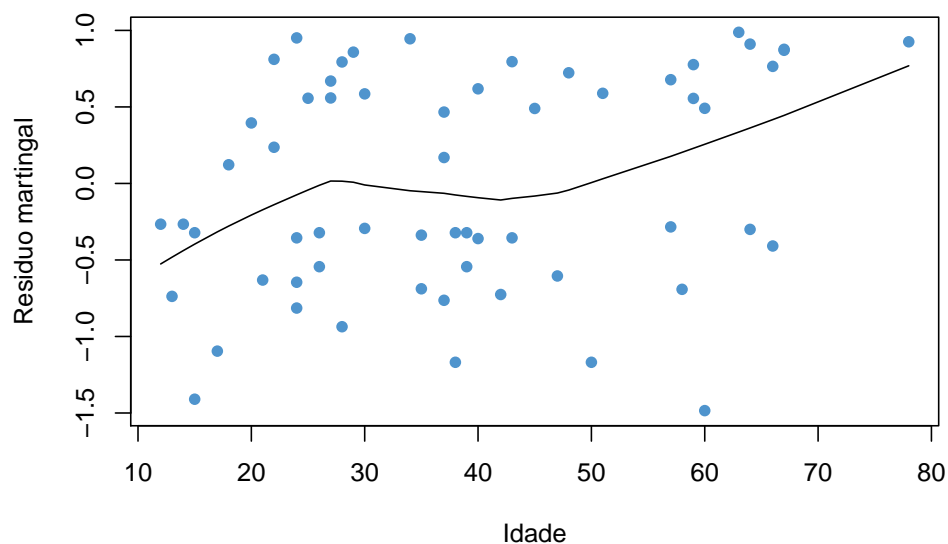
- (a) Ajuste um modelo de Cox sem a variável idade e utilize o resíduo martingal para avaliar a forma funcional dessa variável.

Resolução

Ajustando o modelo de Cox sem a variável idade:

```
## Call:
## coxph(formula = Surv(survivaltime, dead) ~ sex + stage + factor(hist),
##       data = data5, ties = "breslow")
##
##      n= 60, number of events= 30
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## sex           0.19046   1.20981  0.45891  0.415  0.67812
## stage          0.76467   2.14828  0.40826  1.873  0.06107 .
## factor(hist)2  0.06433   1.06644  0.42182  0.153  0.87879
## factor(hist)3  1.66692   5.29586  0.55367  3.011  0.00261 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## sex                1.210     0.8266    0.4921    2.974
## stage              2.148     0.4655    0.9651    4.782
## factor(hist)2      1.066     0.9377    0.4665    2.438
## factor(hist)3      5.296     0.1888    1.7892   15.675
##
## Concordance= 0.667 (se = 0.051 )
## Likelihood ratio test= 12.43 on 4 df,  p=0.01
## Wald test              = 14.51 on 4 df,  p=0.006
## Score (logrank) test = 17.01 on 4 df,  p=0.002
```

Utilizando o resíduo martingal:



Nota-se que a curva se aproxima de uma reta, logo a idade possui uma tendência linear. Portanto, ao utilizar a variável no modelo, não é necessário nenhuma transformação.

(b) Ajuste um modelo de Cox incluindo todas as quatro covariáveis no modelo.

Resolução

Ajustando um modelo de Cox incluindo todas as quatro covariáveis no modelo:

```
## Call:
## coxph(formula = Surv(survivaltime, dead) ~ age + sex + stage +
##       factor(hist), data = data5, ties = "breslow")
##
##      n= 60, number of events= 30
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age              0.03028   1.03075  0.01241  2.441  0.0147 *
## sex              0.24416   1.27654  0.45492  0.537  0.5915
## stage            0.84294   2.32320  0.41313  2.040  0.0413 *
## factor(hist)2    -0.13859   0.87058  0.42942 -0.323  0.7469
## factor(hist)3    1.30686   3.69455  0.59721  2.188  0.0287 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age              1.0307     0.9702     1.0060     1.056
## sex              1.2765     0.7834     0.5234     3.114
## stage            2.3232     0.4304     1.0338     5.221
## factor(hist)2    0.8706     1.1487     0.3752     2.020
## factor(hist)3    3.6946     0.2707     1.1461    11.910
##
## Concordance= 0.74 (se = 0.049 )
## Likelihood ratio test= 18.64 on 5 df,  p=0.002
## Wald test              = 21.75 on 5 df,  p=6e-04
## Score (logrank) test = 25.04 on 5 df,  p=1e-04
```

(c) Selecione as variáveis do modelo em (b) pelo teste da razão de verossimilhanças e apresente um modelo final, apenas com as variáveis significativas. Interprete os parâmetros do modelo final encontrado.

Resolução

Fazendo o teste da razão de verossimilhança temos:

	loglik	Chisq	Df	Pr(> Chi)
NULL	-110.611	NA	NA	NA
age	-107.189	6.844	1	0.009
sex	-107.137	0.103	1	0.748
stage	-104.031	6.213	1	0.013
factor(hist)	-101.292	5.478	2	0.065

Assim, as variáveis mais importantes significativas são: Age e Stage, sendo assim o modelo ajustado:

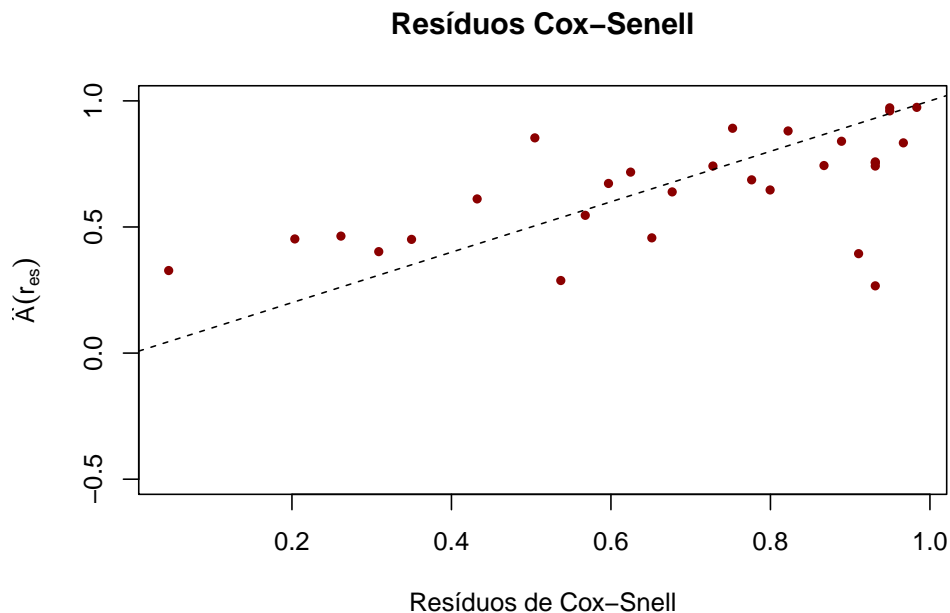
```
## Call:
## coxph(formula = Surv(survivaltime, dead) ~ age + stage, data = data5,
##       ties = "breslow")
##
##      n= 60, number of events= 30
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age   0.03579    1.03644  0.01222  2.929   0.0034 **
## stage 0.97067    2.63970  0.40653  2.388   0.0170 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## age      1.036      0.9648      1.012      1.062
## stage     2.640      0.3788      1.190      5.856
##
## Concordance= 0.711  (se = 0.047 )
## Likelihood ratio test= 13.09  on 2 df,   p=0.001
## Wald test               = 12.1  on 2 df,   p=0.002
## Score (logrank) test = 12.71  on 2 df,   p=0.002
```

A partir do modelo, pode-se dizer que a cada unidade acrescentada na idade do paciente, o risco de óbito aumenta em 3.64% e se o estágio da doença é avançado, o risco de óbito aumenta 163%.

(d) Obtenha os resíduos de Cox-Snell para o modelo ajustado no item (c) e comente.

Resolução

Os resíduos de Cox-Snell são:

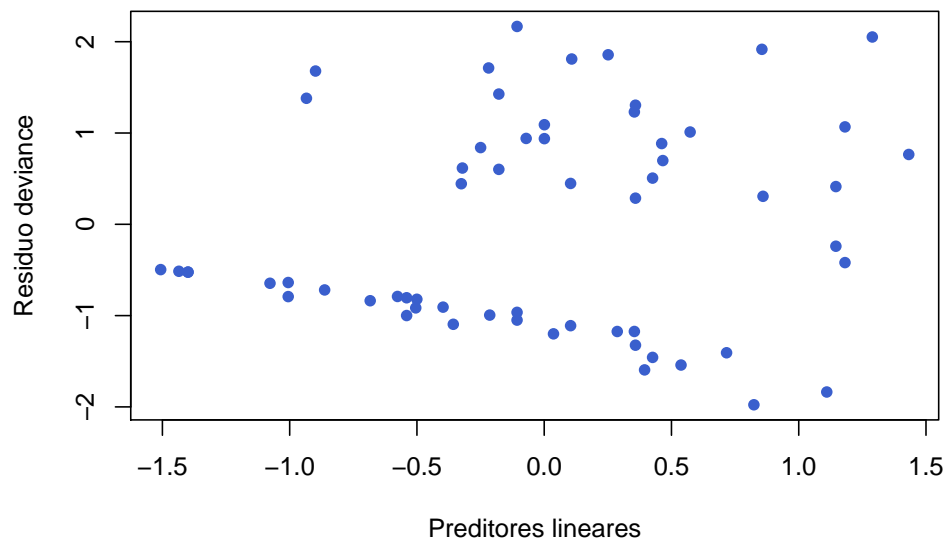


Em que podemos notar uma boa adequação de ajuste global do modelo, pois os pontos se distribuem bem em torno da reta $y=x$.

- (e) Obtenha os resíduos deviance e faça um gráfico apropriado para encontrar pontos atípicos. Identifique o(s) ponto(s), se for o caso, e comente.

Resolução

Os resíduos deviance x preditores lineares:

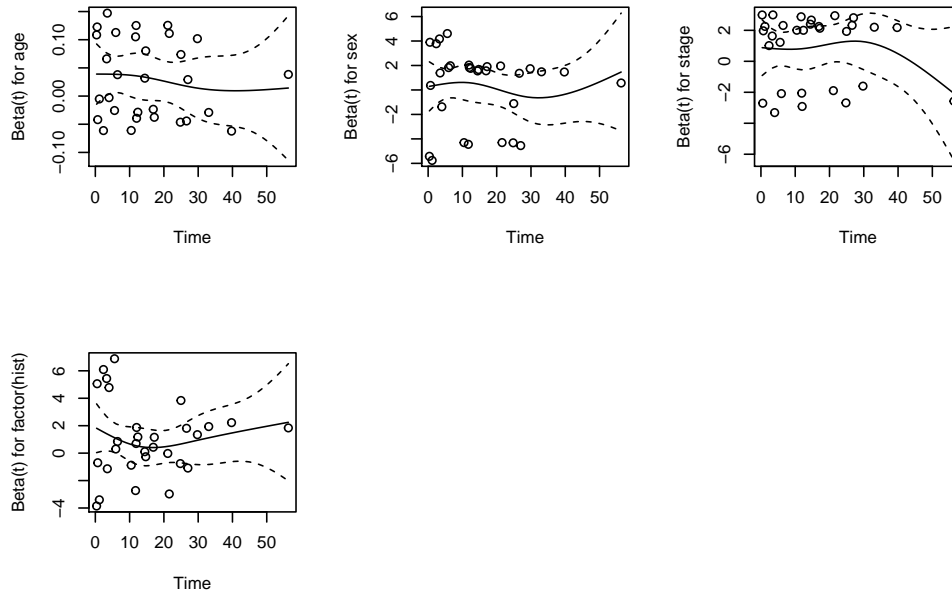


Como pode se notar no gráfico acima, os pontos estão distribuídos em torno do zero de forma aleatória, entre -2 e 2, logo não se percebe pontos atípicos.

- (f) Obtenha os resíduos de Schoenfeld do modelo em (b) e o teste para proporcionalidade dos riscos. Comente os resultados.

Resolução

Resíduos de Schoenfeld do modelo completo:



Em que se pode notar que para as variáveis hist e stage elas não tem um comportamento linear ao longo do tempo.

Fazendo a tabela com o resultado do teste, em que:

$$\begin{cases} H_0 : O \text{ modelo e de riscos proporcionais} \\ H_1 : O \text{ modelo nao e de riscos proporcionais} \end{cases}$$

	chisq	df	p
age	0.618	1	0.432
sex	0.032	1	0.857
stage	0.257	1	0.612
factor(hist)	5.365	2	0.068
GLOBAL	5.527	5	0.355

Queremos testar se o modelo de riscos proporcionais se adequa a esse caso, a ao observar a tabela acima assim como resultado dos gráficos acima as variáveis que rejeita a hipótese nula são hist e stage, ou seja, elas não se adequam no modelo de riscos proporcionais.

Anexo

```
# Lista 3 - MAE0514

# local de trabalho
setwd("~/Área de Trabalho/Lista 4")

# Pacotes
library(ggplot2)
library(survival)
library(survminer)
library(KMsurv)

# Exercício 2

data <- data.frame(trat = c(rep("Sem_trat",10),rep("Radicao",10),
                           rep("Radiacao_BPA",10)),
                  delta = c(rep(1,19),0,rep(1,8),0,0),
                  tempo = c(20,21,23,24,24,26,26,27,28,30,26,28,29,29,30,
                             30,31,31,32,35,31,32,34,35,36,38,38,39,42,42))

# item a

KM2 <- survfit(Surv(tempo, delta)~trat,data = data)

# Grafico Kaplan-Meier
ggsurvplot(KM2, data = data,conf.int = T,legend.title = "Tratamento",
            legend.labs = c("Radiação + BPA", "Radiação","Nenhum"),
            ggtheme=theme_gray() +
            labs(x="Tempo (em dias)",
                 y=expression(hat(S(t))),
                 title = "Estimativas de Kaplan-Meier")

# teste log-rank
knitr::kable(surv_pvalue(KM2, method = c("1"),data=data)[,1:3])

# item b

data$Z1 <- sapply(data$trat,
                  function(x){
                    if (x == 'Radicao') x = '1'
                    else x = '0'
                  })
data$Z1 <- as.factor(data$Z1)

data$Z2 <- sapply(data$trat,
                  function(x){
                    if (x == 'Radiacao_BPA') x = '1'
                    else x = '0'
                  })
data$Z2 <- as.factor(data$Z2)
```

```

attach(data)

# Efron
fit <- coxph(Surv(tempo,delta)~Z1+Z2,ties = "efron",data=data)
summary(fit)

# Breslow
fit2 <- coxph(Surv(tempo,delta)~Z1+Z2,ties = "breslow")
summary(fit2)

# cox/exato
fit3 <- coxph(Surv(tempo,delta)~Z1+Z2,ties = "exact")
summary(fit3)

# item e
l <- -2*logLik(coxph(Surv(tempo,delta)~Z2,ties = "breslow"))+2*logLik(fit2)

1-pchisq(l,1) # p-valor
detach(data)

# Exercício 3

data3 <- data.frame(Tumor = c(rep("grande",15),rep("moderado",20)),
                    delta = c(rep(1,5),rep(0,4),1,rep(0,5),
                                rep(1,6),0,0,rep(1,9),0,1,0),
                    Tempo = c(28,89,175,195,309,377,393,421,447,462,709,744,770,
                                1106,1206,34,88,137,199,280,291,299,300,308,351,358,
                                369,370,371,375,382,392,429,451,1119))

# item a
# dummy tumor
data3$Tumor <- ifelse(data3$Tumor=="grande",1,0)

# Exercício 4

# item a

KM4 <- survfit(Surv(Tempo, delta)~Tumor,data = data3)

# Grafico Kaplan-Meier
ggsurvplot(KM4, data = data3,conf.int = T,legend.title = "Tumor",risk.table = T,
            legend.labs = c("Moderado", "Grande"),
            ggtheme=theme_gray()) +
  labs(x="Tempo (em dias)",
       y=expression(hat(S(t))),
       title = "Estimativas de Kaplan-Meier")

# item b

# teste Log-rank
knitr::kable(surv_pvalue(KM4, method = c("1"))[,1:3],digits = 3)

# teste Familia Fleming-Harrington

```

```

knitr::kable(surv_pvalue(KM4, method = c("FH_p=1_q=1"))[,1:3])

# item c

fit3 <- coxph(Surv(Tempo, delta)~Tumor,data = data3)
summary(fit3)

# Exercício 5

data5 <- read.csv('Lista2_Hodgkins.csv',sep=';',dec=',')

data5$id <- NULL

# item a
fit5 <- coxph(Surv(survivaltime,dead)~sex+stage+factor(hist),
              data = data5, ties="breslow")
summary(fit5)

# martingal
plot(data5$age,resid(fit5), xlab='Idade',
      ylab='Resíduo martingal', pch=16, col="steelblue3" )
smooth <- lowess(data5$age,resid(fit5),iter=0)
lines(smooth)

# item b

fit6 <- coxph(Surv(survivaltime,dead)~age+sex+stage+factor(hist),
              data = data5, ties="breslow")
summary(fit6)

# item c

fit7 <- coxph(Surv(survivaltime,dead)~age+stage,data = data5, ties="breslow")
knitr::kable(anova(fit6, test="Chisq"),digits = 3)

summary(fit7)

# item d

## Cox-Snell residual
coxsnell <- fit6$residuals*data5$dead
x <- coxsnell[coxsnell!=0]
data5 <- data.frame(data5,x)

Surv_Aa <- survfit(coxph(Surv(x, dead)~1,data = data5))

plot(Surv_Aa$surv,x, col="dark red", pch=16, main="Resíduos Cox-Snell",
      xlab="Resíduos de Cox-Snell", ylab=expression(hat(A)(r[es])),
      cex=0.8,ylim=c(-0.5,1) )
abline(0,1,lty=2)

# item e

```

```
plot(fit7$linear.predictors,resid(fit7,type='deviance'),  
      xlab='Preditores lineares', ylab='Residuo deviance',  
      pch=16, col="royalblue3")
```

```
# item f
```

```
scho <- cox.zph(fit6,transform='identity')  
par(mfrow=c(2,3))  
plot(scho)
```

```
# tabela
```

```
knitr::kable(scho$table,digits = 3)
```