

Análise de Sobrevivência e Confiabilidade

Atividade final

Abner Ignacio Melin Leite - 10415165
Brayan Gomes Rafael - 10294814
Eduardo Almeida Merussi Neiva - 10817027
Mariana Soares Alves de Souza - 10294706
Murilo Sernaglia Fernandes - 10415186
Natally F. Souza Cunha Redondo - 10388155

26/07/2021

Sumário

1	Introdução	2
1.1	Sobre o câncer	2
1.2	Objetivos do estudo	3
2	Materiais e métodos	4
2.1	Censura	4
2.2	Estimador de Kaplan-Meier	4
2.3	Teste Logrank	5
2.4	Modelagem	5
2.4.1	Modelo Weibull	6
2.4.2	Modelo Log-normal	7
2.4.3	Modelo Log-logística	7
2.4.4	Teste da Razão de Verossimilhança	7
3	Análise de dados	8
3.1	Leitura e pré processamento	8
3.2	Análise exploratória	8
3.2.1	Breve comparação dos dados	10
3.3	Comparando as Curvas por Teste de Hipótese	11
3.4	Seleção de variáveis e ajuste do modelo	11
3.5	Adequação do modelo	13
4	Conclusões e observações	15
4.1	Conclusões	15
4.2	Observações	15
5	Fontes	16

1 Introdução

1.1 Sobre o câncer

O banco de dados designado ao nosso grupo tem o nome de "BD_AVR_URIN.rda", este que contém dados sobre o **câncer de bexiga**.

O câncer de bexiga está em 12^a colocação no ranking de ocorrência e em 15^a na causa de morte ao redor do mundo. Em todas as etnias essa enfermidade ocorre em uma taxa maior em homens do que mulheres, tendo a proporção de 3 á 4 vezes maior e aumenta ao decorrer da idade.

Assim, no ocidente, é a 4^a com maior incidência entre homens e 8^a entre mulheres. A ocorrência deste câncer é o 8^o maior na América do Norte, Leste do Mediterrâneo, Sul e Oeste da Europa e alguns países do Norte da África. A maior taxa de mortalidade desta doença foi registrada no Egito.

Os riscos conhecidos para câncer de bexiga incluem: fumo, má alimentação, sedentarismo, exposições prolongadas a aminas aromáticas, esquistossomose, tinturas para cabelo e ingestão de água contaminada.

Os riscos em potencial que podem vir a causar câncer de bexiga são efeitos colaterais de remédios e localização geográfica na questão de condições ruins econômicas e sociais. Mais de 420 mil novos casos e 165 mil mortes são registrados mundialmente por causa de câncer de bexiga.

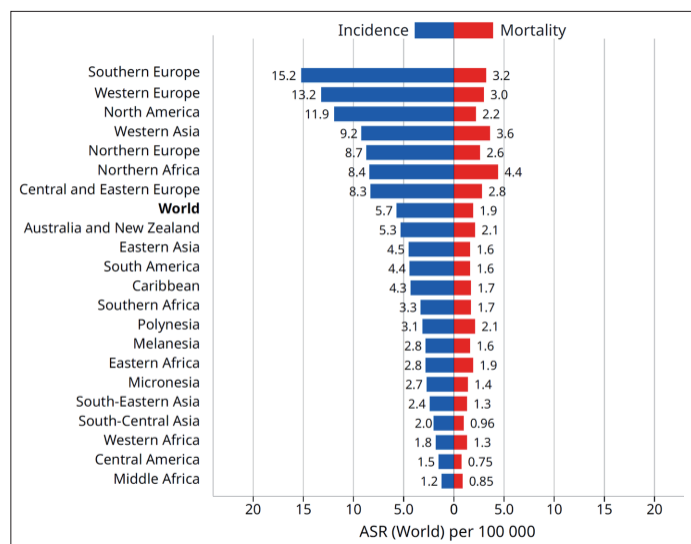


FIGURA 1

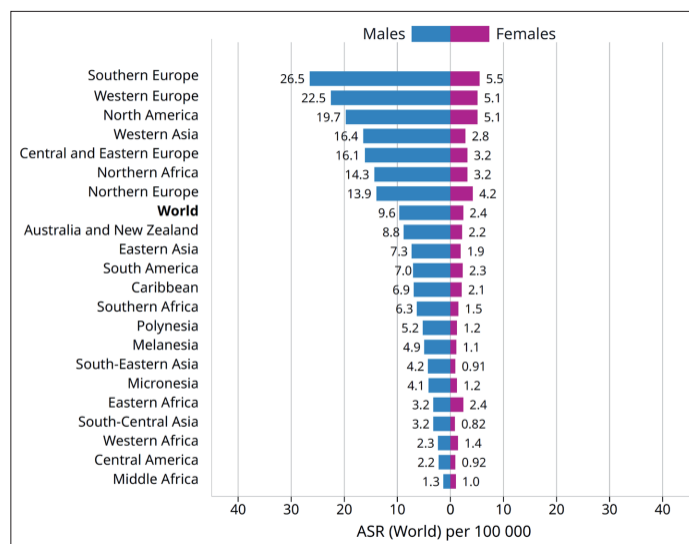


FIGURA 2

Na figura 1, retirada de *Recent patterns of bladder cancer incidence and mortality: a global overview*, temos a taxa de incidência padronizada por idade e a taxa de mortalidade padronizada por idade de câncer de bexiga no mundo em 2018.

Na figura 2, também retirada do link referido, a taxa de incidência padronizada por idade de câncer de bexiga por gênero no mundo em 2018

1.2 Objetivos do estudo

1. Comparar os dados nacionais com os mundiais
2. Aplicar o conteúdo aprendido na disciplina
3. Obter covariáveis importantes para análise

2 Materiais e métodos

2.1 Censura

Censura se trata de uma particularidade da análise de sobrevivência, se trata de casos em que na unidade experimental por algum motivo não ocorre o evento de interesse. Censuras podem ser:

- Tipo I - Unidades experimentais retiradas pois o evento de interesse não ocorre até o fim do tempo estipulado para o experimento.
- Tipo II - Dado que foi determinado que o experimento termina depois que ocorre o evento de interesse em n unidades experimentais, as $N-n$ unidades experimentais que sobraram são censuradas (N =Número total de observações).
- Aleatória - É assumido que o tempo de censura L_i $i = 1, \dots, n$, para a i unidade é uma variável aleatória contínua com funções de sobrevivência e de densidade $G_i(t)$ e $g_i(t)$, respectivamente, e assume-se que L_1, \dots, L_n são estocasticamente independentes uns dos outros e dos tempos de falha T_1, \dots, T_n .
- Intervalar - Temos um intervalo $[L, U]$, sabe-se apenas que o evento ocorreu no intervalo porém não podemos dizer o exato tempo que ocorreu.

No caso dos nossos dados temos censura caso o paciente não tenha tido reincidência da doença ou tenha morrido por algum motivo não relacionado com câncer, nestes casos temos censura tipo I. Como só é possível observar que o paciente teve reincidência da doença quando este vai para o hospital, os nossos dados também tiveram censura intervalar.

2.2 Estimador de Kaplan-Meier

Estimador de Kaplan-Meier pode ser utilizado para estimar $S(t)$, ou seja a função de sobrevivência, a vantagem deste é ser um estimador consistente sob certas condições, ser estimador de máxima verossimilhança e possuir distribuição assintótica normal.

Sua definição é:

$$\widehat{S}(t) = \prod_{j:t_j < t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j < t} \left(1 - \frac{d_j}{n_j} \right)$$

Em que $t_1 < t_2 < \dots < t_k$ são os k ($k < n$) tempos distintos e ordenados de falha, d_j é o número de falhas em t_j e n_j o número de indivíduos sob risco em t_j .

A variância é estimada por ((F.Greenwood):

$$\widehat{Var}(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{j:t_j < t} \frac{d_j}{n_j(n_j - d_j)}$$

Intervalo de $100(1 - \alpha)\%$ de confiança para $S(t)$ é dado por:

$$\hat{S}(t) \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{S}(t))}$$

Vale mencionar que também existe o Intervalo de confiança de Kalbfleish & Prentice que utiliza a transformação $\log[-\log(\hat{S}(t))]$, porém não iremos nos aprofundar neste assunto pois o Software R não utiliza este método como padrão e o grupo manteve o padrão do comando.

O estimador de Kaplan-Meier foi utilizado pelo grupo para verificar a curva de algumas variáveis no início da análise e também para observar o ajuste dos modelos selecionados.

2.3 Teste Logrank

Podemos utilizar este método para comparar duas curvas de sobrevivência inteiras até o tempo t máximo observados, podemos considerar as hipóteses do teste como:

$$H_0 : S_0(t) = S_1(t) \text{ vs } H_1 : S_0(t) = [S_1(t)]^\phi$$

Ou:

$$\begin{cases} H_0 : h_0(t) = h_1(t) \text{ vs } & \rightarrow H_0 : \frac{h_0(t)}{h_1(t)} = 1 \\ H_1 : h_0(t) = \phi h_1(t) & \rightarrow H_1 : \frac{h_0(t)}{h_1(t)} = \phi \end{cases}$$

Para todo $t > 0$ e $\phi > 0$.

Para aplicar o teste Logrank, vamos supor que temos duas amostras para cada grupo de interesse (Ex: Grupo de controle vs Grupo de teste) com $t_1 < t_2 < \dots < t_k$ tempos distintos de falha. Sendo n_{ij} o número de indivíduos em risco do grupo i no tempo t_j , d_{ij} o número de falhas do grupo i no tempo t_j e n_j a quantidade total de indivíduos em risco conseguimos deduzir o seguinte D_{ij} que sob H_0 têm distribuição hipergeométrica:

$$D_{ij} \sim \text{Hipergeométrica}(n_j, n_{ij}, d_j)$$

Utilizando a propriedade da qui-quadrada $\chi^2 \equiv \sum_{j=1}^n \frac{(O_j - E_j)^2}{E_j}$ podemos obter a estatística do teste. Seja:

$$O_{ij} = d_{ij} \text{ observado}; \quad E_{ij} = E(D_{ij}) = n_{ij} \frac{d_j}{n_j} \text{ observado};$$

Se definirmos $U_j = O_{ij} - E_{ij}$, então sob H_0 :

$$\begin{aligned} E(U_j) &= 0 \\ \text{Var}(U_j) &= n_{ij} \left(\frac{d_j}{n_j} \right) \left(1 - \frac{d_j}{n_j} \right) \left(\frac{n_j - n_{ij}}{n_j - 1} \right) = V_j \end{aligned}$$

Logo a estatística do teste é dada por:

$$\frac{\left[\sum_{j=1}^k (O_{ij} - E_{ij}) \right]^2}{\sum_{j=1}^k V_j} = \frac{\left[\sum_{j=1}^k U_j \right]^2}{\sum_{j=1}^k V_j}$$

2.4 Modelagem

Nas análises feitas foram utilizados três modelos diferentes: o modelo Weibull, modelo Log-normal e modelo Log-logístico, sendo aplicado a estimativa K-M da função de sobrevivência em todos. Abaixo entraremos em mais detalhes sobre cada modelo.

2.4.1 Modelo Weibull

O modelo de Weibull vem da distribuição de Weibull cuja f.d.p é dada por:

$$f(t; \alpha, \beta) = \alpha \lambda t^{\alpha-1} e^{-\lambda t^\alpha}, \quad t > 0$$

E a função de risco e função de sobrevivência são:

$$h(t; \alpha, \beta) = \alpha \lambda t^{\alpha-1}, \quad t > 0$$

$$S(t; \alpha, \beta) = e^{-\lambda t^\alpha}, \quad t > 0$$

A avaliação adequação do modelo é feita gráficamente, considerando que $\log(-\log(S(t))) = \log(H(t))$ é linear em relação a $\log(t)$ podemos plotar $\log(-\log(\widehat{S}(t)))$ vs $\log(t)$. Vale lembrar que neste caso $\log(-\log(S(t))) = \log(\lambda) + \alpha \log(t)$

Como foi mencionado no tópico 2.1 temos censura nos dados, a função de verossimilhança do para dados censurados $\theta = (\alpha, \lambda)$ é dada por:

$$L(\theta) = \alpha^r \lambda^r \exp - \lambda \sum_{i=1}^n t_i^\alpha + (\alpha - 1) \sum_{i=1}^n \delta_i \log(t_i)$$

Onde δ_i é o indicador de censura e $t_i = \min(T_i, C_i)$. Portanto o log-verossimilhança é dado por:

$$l(\theta) = r \log(\alpha) + r \log(\lambda) - \lambda \sum_{i=1}^n t_i^\alpha + (\alpha - 1) \sum_{i=1}^n \delta_i \log(t_i)$$

Assim, os componentes do escore são:

$$\frac{\partial l(\theta)}{\partial \alpha} = \frac{r}{\alpha} - \lambda \sum_{i=1}^n \log(t_i) t_i^\alpha + \sum_{i=1}^n \delta_i \log(t_i)$$

$$\frac{\partial l(\theta)}{\partial \lambda} = \frac{r}{\lambda} - \sum_{i=1}^n t_i^\alpha$$

E os componentes da matriz de informação observada é dado por:

$$\frac{\partial^2 l(\theta)}{\partial \alpha^2} = \frac{r}{\alpha^2} - \lambda \sum_{i=1}^n \log(t_i)^2 t_i^\alpha + \sum_{i=1}^n \delta_i \log(t_i)$$

$$\frac{\partial^2 l(\theta)}{\partial \alpha \partial \lambda} = - \sum_{i=1}^n \log(t_i) t_i^\alpha$$

$$\frac{\partial^2 l(\theta)}{\partial \lambda^2} = \frac{r}{\lambda^2}$$

2.4.2 Modelo Log-normal

Para o modelo Log-normal vem da seguinte f.d.p:

$$f(t) = \frac{1}{\sqrt{2\pi t\sigma}} \exp \left\{ -\frac{1}{2} \left(\frac{\log t - \mu}{\sigma} \right)^2 \right\}$$

Onde $T \sim N(\mu, \sigma^2)$

E a função de risco e função de sobrevivência são representadas por:

$$h(t) = \frac{\exp \left\{ \frac{1}{2} \left(\frac{\log t - \mu}{\sigma} \right)^2 \right\}}{\sqrt{2\pi t\sigma} \phi \left(-\frac{\log t - \mu}{\sigma} \right)}$$

$$S(t) = \phi \left(-\frac{\log t - \mu}{\sigma} \right)$$

Quanto a adequabilidade do modelo, $\phi^{-1}(S(t))$ é linear em relação a $\log(t)$ logo podemos fazer o gráfico de $\phi^{-1}(\widehat{S(t)})$ vs $\log(t)$. Neste modelo temos que $\phi^{-1}(S(t)) = \frac{\mu}{\sigma} - \frac{\log(t)}{\sigma}$.

2.4.3 Modelo Log-logística

Para o modelo Log-logística vem da seguinte f.d.p:

$$f(t) = \frac{\alpha \lambda t^{\alpha-1}}{(1 + \lambda t^\alpha)^2}$$

E a função de risco e função de sobrevivência são:

$$h(t) = \frac{\alpha \lambda t^{\alpha-1}}{1 + \lambda t^\alpha}$$

$$S(t) = (1 + \lambda t^\alpha)^{-1}$$

Dado que $\log \left(\frac{1-S(t)}{S(t)} \right)$ tem relação linear com $\log(t)$ então podemos avaliar a adequabilidade do modelo plotando $\log \left(\frac{1-S(t)}{S(t)} \right)$ vs $\log(t)$. Podemos considerar: $\log \left(\frac{1-S(t)}{S(t)} \right) = \log + \alpha \log t$.

2.4.4 Teste da Razão de Verossimilhança

Para um modelo com um vetor $\theta = (\theta_1, \dots, \theta_p)'$ de parâmetros, muitas vezes há o interesse em testar hipóteses relacionadas a este vetor ou a um subconjunto dele.

O teste da razão de verossimilhanças é baseado na comparação dos valores dos logaritimos das funções de verossimilhança maximizada sem restrição e sob H_0 , ou seja, compara-se $\log L(\hat{\theta})$ e $\log L(\hat{\theta}_0)$. A estatística para este teste é dada por:

$$TRV = -2 \log \left[\frac{L(\hat{\theta}_0)}{\log L(\hat{\theta})} \right] = 2[\log L(\hat{\theta}) - \log L(\hat{\theta}_0)],$$

que, sob $H_0 : \theta = \theta_0$, segue aproximadamente uma distribuição qui-quadrado com p graus de liberdade. Para amostras grandes, H_0 é rejeitada à um nível $100\alpha\%$ de significância, se $TRV > \chi_{p,1-\alpha}$

3 Análise de dados

3.1 Leitura e pré processamento

```
1 load("BD_AVR_URIN.rda")
2 df = BD_AVR_URIN
3 rm(BD_AVR_URIN)
```

Há muitas linhas com dados faltando. Como a ausência da informação de tempo ou de censura inviabilizam a análise, iremos remover todas as linhas que não possuem alguma dessas variáveis.

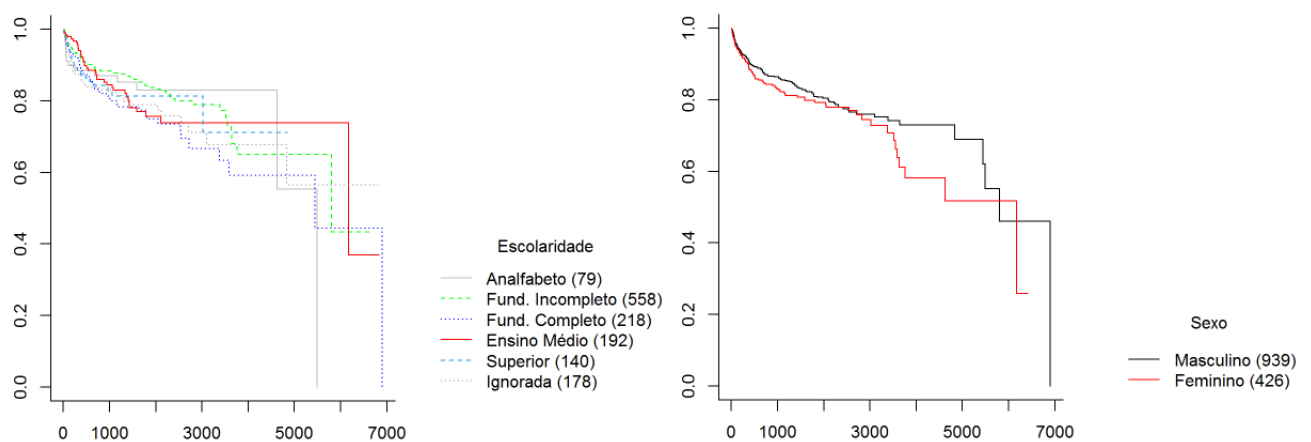
```
1 library(tidyverse)
2
3 df = drop_na(data=df, ULTDIAG)
4 df = drop_na(data=df, VCAN)
5
6 dim(df)
```

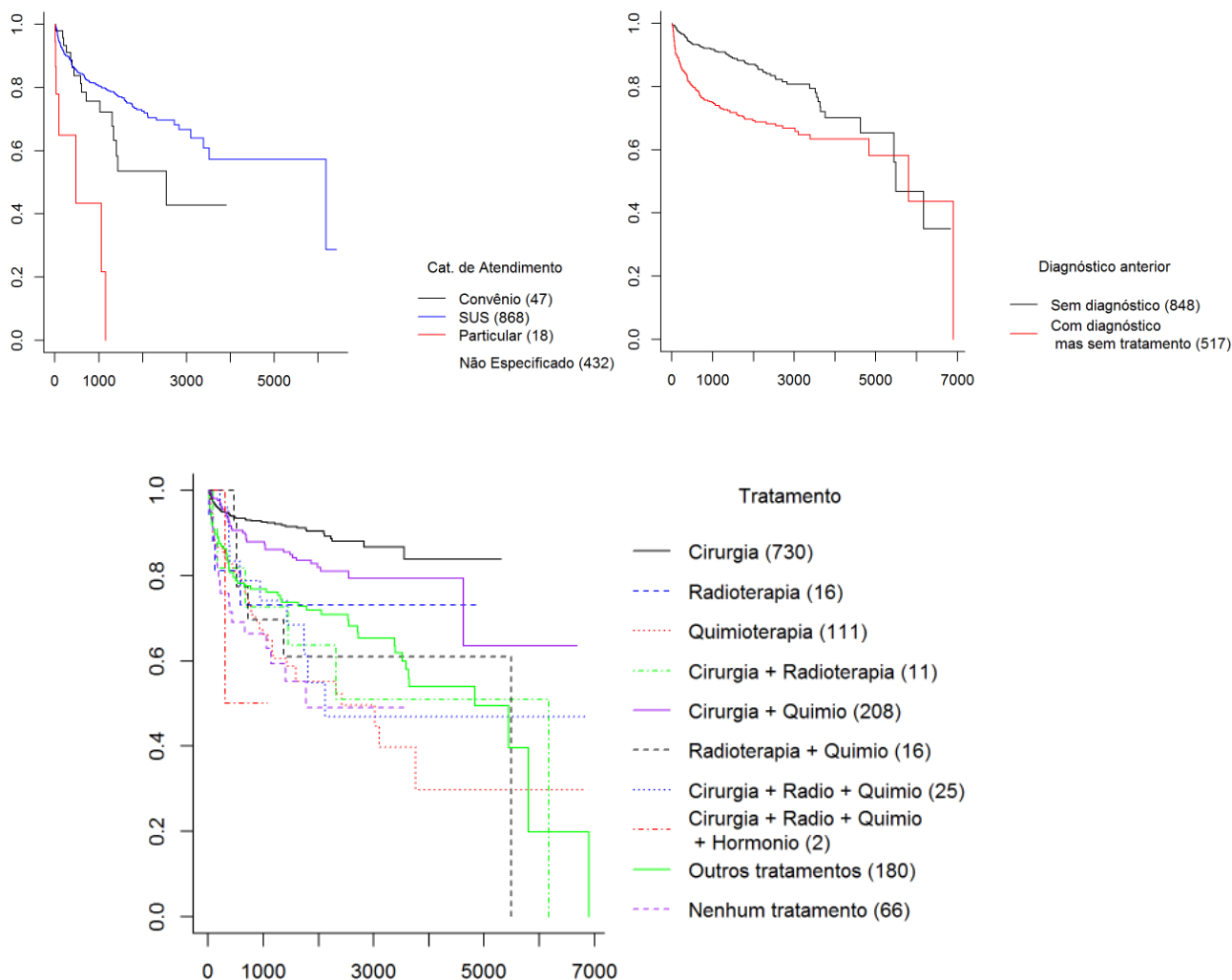
```
## [1] 1365    99
```

Das 5496 linhas, ficamos com 1365 que contem, no mínimo, o tempo e a indicação de censura.

3.2 Análise exploratória

Iremos começar obtendo estimativas de Kaplan-Meier para algumas das variáveis





Podemos notar que muitas das curvas são impróprias, o que indica uma alta proporção de censuras e limita a análise dos gráficos.

Parece não haver diferença significativa entre os níveis de escolaridade ou de sexo. A curva para pacientes que não tinham diagnóstico anterior se distanciava da curva dos pacientes já diagnosticados inicialmente, mas depois voltaram a se aproximar. Esse último gráfico sugere que pacientes previamente diagnosticados tem prognóstico inicial pior.

Já para a categoria de atendimento, vemos que as curvas apresentaram diferenças significativas, no entanto é difícil fazer qualquer afirmação apenas com base no gráfico, considerando o desbalançamento extremo entre as categorias.

Para os tratamentos, vemos que a cirurgia como método único tem a curva mais alta, seguido do método cirurgia + quimioterapia. A quimioterapia como método único teve as mais baixas estimativas para $S(t)$ no geral.

A análise gráfica indica que as variáveis categoria do atendimento e tratamento serão relevantes na regressão.

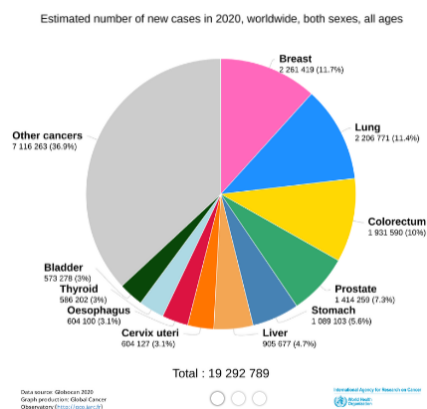
Os códigos para encontrar as contagens de cada nível das variáveis podem ser encontrados no arquivo Rmd.

3.2.1 Breve comparação dos dados



Nos dados mundiais a incidência de câncer de bexiga em homens é de 9.8 para 2.4 de mulheres cerca de 4 vezes maior a relação entre homens e mulheres. Comparando com a nossa base de dados temos 939 homens e 426 mulheres com câncer de bexiga, como nossa base é pequena comparada ao mundo seguimos a ideia de uma proporção a cada 100 pessoas logo essa proporção fica 9.39 para homens e 4.26 para mulheres sendo assim 2.2 vezes maior e comparando a América do Sul 7.0 para homens e 2.3 para mulheres 3 vezes maior, formando a seguinte tabela.

	Homens	Mulheres
Mundo	9,8	2.4
América do Sul	7.0	2.3
Brasil	9.39	4.26

Também temos o comparativo de incidência mundial de câncer de bexiga por meio de um gráfico de setores retirados da fonte *GloboCAN 2020:Bladder cancer 10th most commonly diagnosed worldwide*



Onde fica notório que a incidência do câncer de bexiga entre todos os outros cânceres tem uma taxa de 3% mundialmente, fazendo um comparativo com o Brasil, pela fonte *Inca - Estimativas sobre Câncer* temos a seguinte tabela de incidência por sexo

Localização primária	Casos	%	Homens	Mulheres	Localização primária	Casos	%
Próstata	65.840	29,2%			Mama feminina	66.280	29,7%
Cólon e Reto	20.540	9,1%			Cólon e Reto	20.470	9,2%
Traqueia, Brônquio e Pulmão	17.760	7,9%			Colo do útero	16.710	7,5%
Estômago	13.360	5,9%			Traqueia, Brônquio e Pulmão	12.440	5,6%
Cavidade Oral	11.200	5,0%			Glândula Tireoide	11.950	5,4%
Esôfago	8.690	3,9%			Estômago	7.870	3,5%
Bexiga	7.590	3,4%			Ovário	6.650	3,0%
Linfoma não Hodgkin	6.580	2,9%			Corpo do útero	6.540	2,9%
Laringe	6.470	2,9%			Linfoma não Hodgkin	5.450	2,4%
Leucemias	5.920	2,6%			Sistema Nervoso Central	5.230	2,3%

Logo nota-se que a incidência deste câncer no Brasil embora siga próximo da proporção mundial de 3% é algo mais recorrente em homens como comentado nos tópicos anteriores,

— **Observação:** Utilizamos essas duas fontes, pois não tínhamos no banco de dados alguma forma para comparar o câncer de bexiga com os demais cânceres

3.3 Comparando as Curvas por Teste de Hipótese

As curvas de sobrevivência para diversos grupos serão comparadas pelo método logrank para evitar que padrões de censura diferentes influenciem na análise, bem como considerar um peso igual para todo o eixo do tempo.

A hipótese nula do teste é que todas as curvas são iguais. Iremos considerar um valor descritivo de 0,25. A proposta de escolher um nível relativamente alto de significância é baseada em recomendações de Bendel e Afifi (1977) para regressão linear, Costanza e Afifi (1979) para análise de discriminante e Mickey E Greenland (1989) para mudanças nos coeficientes do modelo de regressão logística.

Fizemos uma avaliação das variáveis categóricas significativas pelo teste do logrank (códigos encontrados no arquivo rmd)

Algumas das covariáveis que apresentaram diferença entre seus níveis foram: faixa etária, metástase (1 e 4), Classificação TNM (T, N e M), Estadiamento pós cirúrgico, combinação de tratamentos realizados no hospital, tratamentos realizados após admissão fora do hospital, local de recidiva/-metástase (1).

Obs: utilizar $\alpha = 0.25$ retorna 4 variáveis a mais do que utilizando $\alpha = 0.10$, sendo três delas escolaridade, sexo e morfologia.

3.4 Seleção de variáveis e ajuste do modelo

Iremos inicialmente ajustar todos os modelos de uma única covariável e continuar a análise apenas com aqueles que são significativos à 5%. As variáveis utilizadas são algumas daquelas consideradas

relevantes pelo teste logrank e mais algumas numéricas que não foram testadas anteriormente: IDADE E DIAGTRAT (Diferença em dias entra as datas de tratamento e diagnóstico).

Não iremos considerar todas as variáveis que apresentaram diferenças pelo teste logrank por conta de seus significados: algumas informações só podem ser obtidas após certo tempo de acompanhamento ou podem ser desconhecidas. A prioridade é pelo modelo mais simples, que utiliza informações de fácil acesso.

O resultado foi que quase todas as variáveis geraram bons modelos, por isso decidiu-se por fazer a escolha do modelo baseado no teste da razão de verossimilhança, a partir de um modelo Weibull.

Modelos univariados e significância

Modelos	Covariáveis	Log Verossimilhança	valor p
1	nenhuma	-2474	
2	Tratamento	-2415,5	$< 2e - 16$
3	Sexo	-2473,06	0,16
4	DiagPrev	-2450,27	$5,58e - 12$
5	Idade	-2461,52	$5,85e - 7$
6	M	-2454,94	$1,7e - 6$
7	N	-2456,9	0,0051
8	Cirurgia	-2444,3	$1,29e - 14$
9	DiagTrat	-2275,3	0,197
10	UFResid	-2450	0,00933

As variáveis Sexo e DiagTrat não foram significativas, porém como sexo se manteve relevante após a inclusão de tratamento e, considerando sua significância clínica, continuaremos utilizando-a. A variável numérica DiagTrat será descartada.

Tabela TRV

Efeito	Hipótese nula	TRV	valor p
Sexo Tratamento	$H_0 : \beta_2 = 0$	4,14	0,0419
Diagprev Tratamento+Sexo	$H_0 : \beta_3 = 0$	4,732	0,0296
Idade Tratamento+Sexo+DiagPrev	$H_0 : \beta_4 = 0$	15,728	7,313354e-05
M Tratamento+Sexo+DiagPrev+Idade	$H_0 : \beta_5 = 0$	20,2	6,975294e-06
N Tratamento+Sexo+DiagPrev+Idade+M	$H_0 : \beta_6 = 0$	4,4	0,0359
Cirurgia Tratamento+Sexo+DiagPrev+Idade+M+N	$H_0 : \beta_7 = 0$	2	0,1380
UFResid Tratamento+Sexo+DiagPrev+Idade+M+N	$H_0 : \beta_8 = 0$	34	7,074463e-10

Podemos ver que a variável cirurgia não é significativa à 5%, considerando as anteriores. O Estado de nascimento foi altamente significativo mesmo após a inclusão de 7 outras variáveis, o que faz sentido pois como foi visto, o local é um fator de influência no câncer de bexiga.

```

1 fit_weibull = survreg(Surv(ULTDIAG,VCAN)~TRATAMENTO+SEXO+
2   DIAGPREV+IDADE+M+N+UFRESID, dist='weibull', data=df)
3
4 fit_lognormal=survreg(Surv(ULTDIAG,VCAN)~TRATAMENTO+SEXO+
5   DIAGPREV+IDADE+M+N+UFRESID, dist='lognormal', data=df)
6
7 fit_loglog=survreg(Surv(ULTDIAG,VCAN)~TRATAMENTO+SEXO+
8   DIAGPREV+IDADE+M+N+UFRESID, dist='loglogistic', data=df)

```

Assim foram ajustados três modelos paramétricos com distribuições de base Weibull, lognormal e log-logística e 7 variáveis: tratamento, sexo, diagnóstico anterior, idade, classificação TNM (metástase ou linfonodos) e UF de residência.

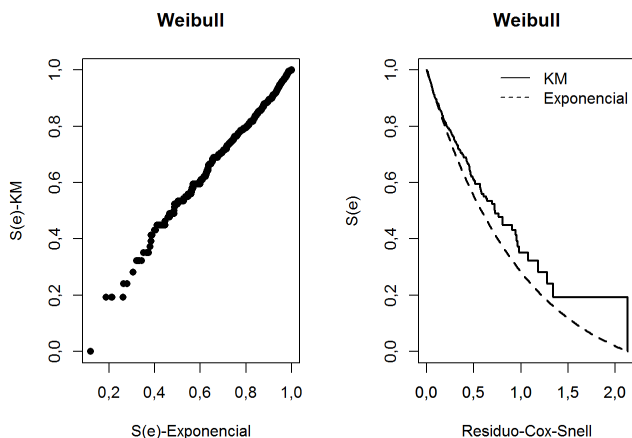
3.5 Adequação do modelo

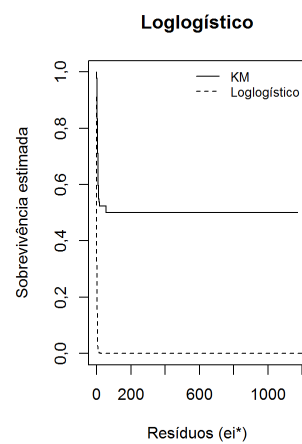
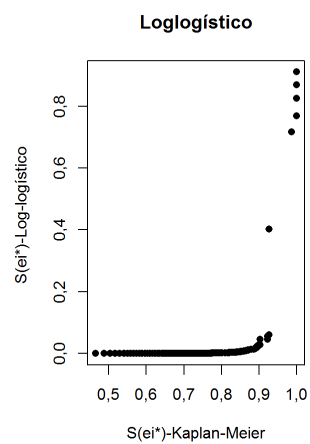
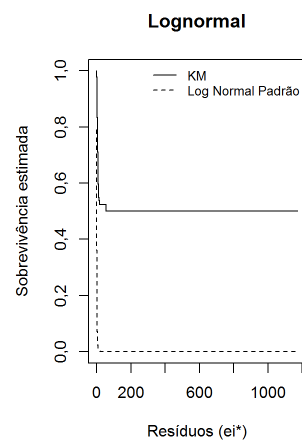
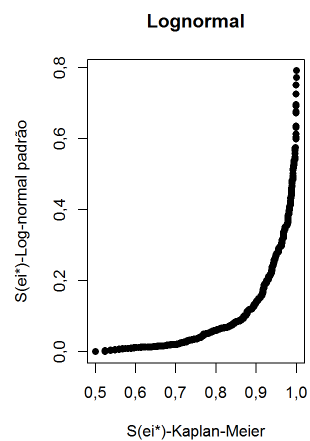
Podemos padronizar os resíduos de acordo com a seguinte relação:

$$\hat{v}_i = \frac{y_i - x_i' \hat{\beta}}{\hat{\sigma}}; \quad \text{com } y_i = \log t_i$$

Dessa forma, se os modelos log-normal e log-logístico forem adequados, as probabilidades de sobrevivência $\hat{S}(\hat{v}_i)$ estimadas por Kaplan-Meier VS os respectivos valores obtidos pelo modelo devem se aproximar de uma reta.

Já os resíduos do modelo Weibull serão analisados pelo método padrão, o código pode ser encontrado no arquivo Rmd.





4 Conclusões e observações

4.1 Conclusões

Finalizando o trabalho concluímos que na questão comparativa de dados mundiais e nacionais, a incidência do câncer de bexiga comparado aos outros cânceres tem proporção por cerca 3% sendo que no Brasil ele é um dos 10 mais recorrentes entre homens. Sendo que dentro da América Latina o Brasil possui uma das maiores taxas de incidência do câncer de bexiga.

Na modelagem (cujo os modelos estão no arquivo Rmd) concluímos que tanto os modelos Log-normal e Log-logístico possuem violações brutas as suposições de adequabilidade já o modelo Weibull não apresenta violações preocupantes.

Pelo que foi analisado de toda a pesquisa temos entre as 2 melhores curvas de sobrevivência entre os pacientes cujo atendimento foi feito dentro de hospitais para os seguintes atendimentos: Cirurgia (método único) seguido de Cirurgia + Quimioterapia, uma observação sobre a Quimioterapia, é que ela sendo tratamento único possui a curva de sobrevivência mais baixa entre todos os tratamentos.

A variável UF de residência, embora tenha 28 níveis, se mostrou muito relevante para o modelo, confirmando as hipóteses iniciais de que as condições socioeconômicas e geográficas são um fator de influência para o câncer. Outras variáveis categóricas geraram aumento dos resíduos e um modelo mal ajustado.

4.2 Observações

Utilizamos de diversas fontes para fazer análises breves interligando-as com nossa base de dados. Tendo um foco maior para incidência do câncer e comparações gerais.

Colocamos vários gráficos de resíduos no tópico *3.5:adequação do modelo* com o objetivo de mostrar a necessidade de fazer uma modelagem de longa duração.

Foi tentado o ajuste de modelos de sobrevivência de longa duração, o que é mais apropriado a este caso pois existe taxa de cura e muitas censuras. No entanto, devido à limitações computacionais, não foi possível obter a verossimilhança dos modelos: diversos valores apresentaram ordem de grandeza 10^{-200} ou ainda mais baixa, inviabilizando os cálculos. Ainda sim, o ajuste paramétrico feito apresentou bons resultados.

5 Fontes

- RECENT PATTERNS OF BLADDER CANCER INCIDENCE AND MORTALITY: A GLOBAL OVERVIEW
- Maio Vermelho alerta para o diagnóstico precoce do câncer de bexiga
- Inca - Estimativas sobre Câncer
- GloboCAN 2020:Bladder cancer 10th most commonly diagnosed worldwide
- Análise de Sobrevivência Aplicada, Enrico Antônio Colosimo.