

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESPECIALIZAÇÃO EM ESTATÍSTICA



BIOESTATÍSTICA
ANÁLISE DE SOBREVIVÊNCIA

Professora: Arminda Lúcia Siqueira
Aluno: Zilma Reis, Ricardo Veiga e Augusto Filho
Belo Horizonte - MG

Exercício 1. *Apresente exemplos de pesquisa em que as técnicas de análise de sobrevivência podem ser empregadas.*

Resolução:

Existem vários exemplos de aplicação dos modelos de análise de sobrevivência. Na área médica, eles são muito utilizados na identificação de fatores de prognóstico para uma doença, bem como na comparação de tratamentos. Entretanto condições similares ocorrem em outras áreas em que se usam as mesmas técnicas de análise de dados.

- 1. **Em oncologia** qualquer nova terapêutica ou droga para o combate ao câncer requer um estudo, em que a resposta de interesse é geralmente o tempo de sobrevida dos pacientes, que é chamada de sobrevida global pelos oncologistas.*
- 2. **Em engenharia**, são comuns os estudos onde produtos ou componentes são colocados sob teste para se estimar características relacionadas aos seus tempos de vida, tais como o tempo médio ou a probabilidade de um certo produto durar mais do que 5 anos.*
- 3. **Em ciências sociais**, onde várias situações de interesse têm a resposta o tempo entre eventos. Criminalistas estudam o tempo entre a liberação de presos e a ocorrência, desempregos, promoções e aposentadorias; demógrafos com nascimentos, mortes, casamentos, divórcios e migrações. O crescimento observado no número de aplicações em medicina também pode ser observado nestas outras áreas.*

Exercício 2. Compare os métodos de tabela de vida e de Kaplan-Meier para estimar as curvas de sobrevida, destacando quando cada um deles deve ser utilizado. Compare o método de Kaplan-Meier e o modelo de Cox, destacando quando cada um deles deve ser utilizado.

Resolução:

- **O Método da Tábua de Vida:** É um método clássico para estimar uma curva de sobrevivência. Para iniciar, os tempos de sobrevivência são agrupados dentro de intervalos de comprimentos fixos. As primeiras três colunas da tabela enumeram o intervalo de tempo de t a $t + n$, a proporção de indivíduos vivos no tempo t . Se l_0 é o número de indivíduos vivos no tempo 0 e l_t é o número deles vivos no tempo t , a proporção de indivíduos que ainda não morreram no tempo t pode ser calculada como $S(t) = \frac{l_t}{l_0}$.
- **Método de Kaplan-Meier:** Este é um estimador não paramétrico para a função de sobrevivência. Ele foi proposto por Kaplan e Meier (1958) e é também chamado de estimador limite-produto. Ele é uma adaptação da função de sobrevivência empírica que, na ausência de censuras, é definida como:

$$\hat{S}(t) = \frac{\text{número de observações que não falharam até o tempo } t}{\text{número total de observações no estudo}}$$

O estimador de Kaplan-Meier, na sua construção, considera tantos intervalos de tempo quantos forem o número de falhas distintas. Os limites dos intervalos de tempo são os tempos de falhas da amostra.

A comparação entre os métodos de Tabela de Vida e Kaplan-Meier consiste no fato de que quando usamos o método da tábua de vida, a função de sobrevivência estimada modifica-se somente durante os intervalos de tempo em que ocorre pelo menos uma morte. Para conjuntos de dados menores, tal como a amostra de 12 pacientes hemofílicos diagnosticados com Aids, pode haver muitos intervalos sem uma única morte. Nesses casos, pode não fazer sentido apresentar a função de sobrevivência desse modo. O método de Kaplan-Meier, é uma técnica que usa o tempo de sobrevivência exato para cada indivíduo na amostra, em vez de agrupar os tempos em intervalos.

- **Modelo de Cox:** A metodologia para comparação de curvas de sobrevida, é adequada somente se os pacientes dos diferentes grupos diferem apenas na variável que os divide em grupos. Na prática, entretanto, o tempo de sobrevida depende, com

frequência, de outros fatores. Cox generalizando resultados anteriores, introduziu uma metodologia que resolve o problema de comparação de curvas de sobrevida na presença de fatores de confusão. Para isto adaptou os modelos de regressão para esta nova situação permitindo a análise de dados provenientes de estudos de tempo de vida em que a resposta é o tempo até a ocorrência de um evento de interesse, ajustando por covariáveis.

Exercício 3. *Miller (1981) descreve um ensaio clínico para avaliar a eficácia da manutenção da quimioterapia no tratamento de pacientes em leucemia aguda. Após entrar em remissão através do tratamento de quimioterapia, os pacientes que participaram desse estudo foram aleatoriamente alocados a um dos dois grupos. O primeiro grupo recebeu manutenção de quimioterapia, enquanto que o segundo não (grupo controle). Os dados preliminares desse estudo conduzido por Embury et al. na Universidade de Stanford referentes ao tempo de remissão em meses estão apresentados a seguir.*

Grupo 1	($n_1 = 11$)	9	13	13*	18	23	28*	31	34	45*	48	161*	
Grupo 2	($n_1 = 12$)	5	5	8	8	12	16*	23	27	30	33	43	45

(a) *A tabela a seguir apresenta os cálculos do método de Kaplan-Meier para a obtenção da curva de remissão para o grupo 1. Complete as contas para o grupo 2 e desenhe as curvas de remissão.*

Resolução:

Utilizando o Minitab, obteve-se os seguintes valores:

Grupo 1	Grupo 2
$\hat{S}(0) = 1$	$\hat{S}(0) = 1$
$\hat{S}(9) = 0,909091$	$\hat{S}(5) = 0,833333$
$\hat{S}(13) = 0,818182$	$\hat{S}(8) = 0,666667$
$\hat{S}(18) = 0,715909$	$\hat{S}(12) = 0,583333$
$\hat{S}(23) = 0,613636$	$\hat{S}(23) = 0,486111$
$\hat{S}(31) = 0,490909$	$\hat{S}(27) = 0,388889$
$\hat{S}(34) = 0,368182$	$\hat{S}(30) = 0,291667$
$\hat{S}(48) = 0,184091$	$\hat{S}(33) = 0,194444$
	$\hat{S}(43) = 0,097222$
	$\hat{S}(45) = 0,000000$

A seguir, veremos a Curva de remissão para os grupos 1 e 2.

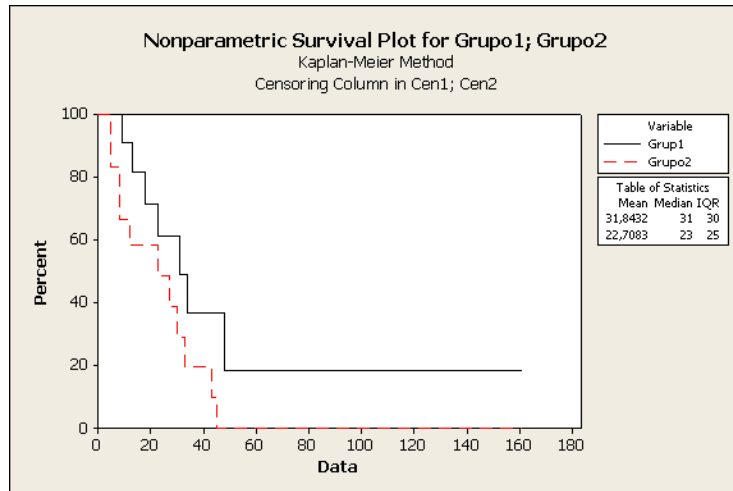


Figura 1: Curva de Remissão

- (b) *Avalie se a manutenção da quimioterapia prolonga o tempo de remissão.*

Resolução:

De acordo com a análise do gráfico acima a manutenção da quimioterapia prolonga o tempo de remissão no grupo 1.

- (c) *Estime os quartis para os dois grupos.*

Resolução:

Utilizando o Minitab, obtemos os seguintes valores para o grupo 1:

Characteristics of Variable

	Standard	95,0% Normal CI	
Mean(MTTF)	Error	Lower	Upper
31,8432	4,89067	22,2576	41,4287
Median = 31			
IQR = 30 Q1 = 18 Q3 = 48			

Logo, os quartis para o grupo 1 são: $Q_1 = 18$; $Q_2 = 31$ e $Q_3 = 48$.

Para o grupo 2, temos:

Characteristics of Variable

	Standard	95,0% Normal CI	
Mean(MTTF)	Error	Lower	Upper
22,7083	4,38501	14,1139	31,3028
Median = 23			
IQR = 25 Q1 = 8 Q3 = 33			

Portanto, os quartis para o grupo 2, são: $Q_1 = 8$; $Q_2 = 23$ e $Q_3 = 33$.

Exercício 4. Um estudo tinha como objetivo avaliar fatores prognósticos no tempo livre de doença entre 235 mulheres que foram tratadas de câncer de mama. Analise os dados do arquivo *brestca.xls* cujas variáveis consideradas foram:

<i>Variável</i>	<i>Descrição</i>
<i>ID</i>	<i>Número de identificação (1 a 235)</i>
<i>AGE</i>	<i>Idade em anos</i>
<i>RACE</i>	<i>Raça: 0=branca, 1=outra</i>
<i>ER</i>	<i>Concentração de esgrógeneo (fMol/mg)</i>
<i>PR</i>	<i>Concentração de progesterona (fMol/mg)</i>
<i>SIZE</i>	<i>Tamanho do tumor (cm)</i>
<i>NODSTAR</i>	<i>Status nodal: 1=negativo, 2=1 a 3, 3=4 a 8, 4=9 ou mais nódulos</i>
<i>HGRADE</i>	<i>Grau histológico: 1=I, 2=II, 3=III</i>
<i>MITO</i>	<i>Índice mitótico: 0=1 a 10/10 HPF, 1=mais que 10/10 HPF</i>
<i>NGRADE</i>	<i>Grau nuclear: 1=I, 2=II, 3=III</i>
<i>LUMPH</i>	<i>Infiltração linfocítica: 1=baixa, 2=moderada, 3=intensa</i>
<i>VASINV</i>	<i>Invasão vascular: 0=ausente, 1=presente</i>
<i>NEC</i>	<i>Necrose do tumor: 0=ausente, 1=presente</i>
<i>TYPE</i>	<i>Tipo histológico: 1=ductal, 2=mucoide, 3=papilar, 4=outro</i>
<i>DFI</i>	<i>Intervalo livre da doença (meses)</i>
<i>CENS</i>	<i>Status da observação: 0=não censurada, 1=censurada</i>

Resolução:

Abaixo, temos uma análise descritiva dos dados no Minitab:

Descriptive Statistics: AGE; RACE; ER; PR; SIZE; NODSTAT; HGRADE; MITO; ...

Variable	Mean	StDev	CoefVar	Minimum	Q1	Median	Q3	Maximum
AGE	56,804	12,123	21,34	28,000	48,000	56,000	66,000	85,000
RACE	0,2468	0,4321	175,06	0,0000	0,0000	0,0000	0,0000	1,0000
ER	66,59	105,61	158,60	0,00	6,60	28,00	82,00	880,00
PR	95,3	186,6	195,90	0,0	7,8	28,4	102,0	1312,0
SIZE	4,965	3,094	62,32	1,000	3,000	4,000	6,000	23,000
NODSTAT	2,4809	1,1851	47,77	1,0000	1,0000	2,0000	4,0000	4,0000
HGRADE	2,1319	0,5347	25,08	1,0000	2,0000	2,0000	2,0000	3,0000
MITO	0,3277	0,4704	143,55	0,0000	0,0000	0,0000	1,0000	1,0000
NGRADE	2,1915	0,5245	23,94	1,0000	2,0000	2,0000	3,0000	3,0000
LYMPH	2,2851	0,6599	28,88	1,0000	2,0000	2,0000	3,0000	3,0000
VASINV	0,8979	0,3035	33,80	0,0000	1,0000	1,0000	1,0000	1,0000
NEC	0,4596	0,4994	108,67	0,0000	0,0000	0,0000	1,0000	1,0000
TYPE	1,4638	0,9967	68,09	1,0000	1,0000	1,0000	1,0000	4,0000
DFI	28,95	17,43	60,23	0,10	13,10	29,30	43,50	59,40
CENS	0,5957	0,4918	82,55	0,0000	0,0000	1,0000	1,0000	1,0000

A seguir, temos a curva estimada pelo método de Kaplan-Meier.

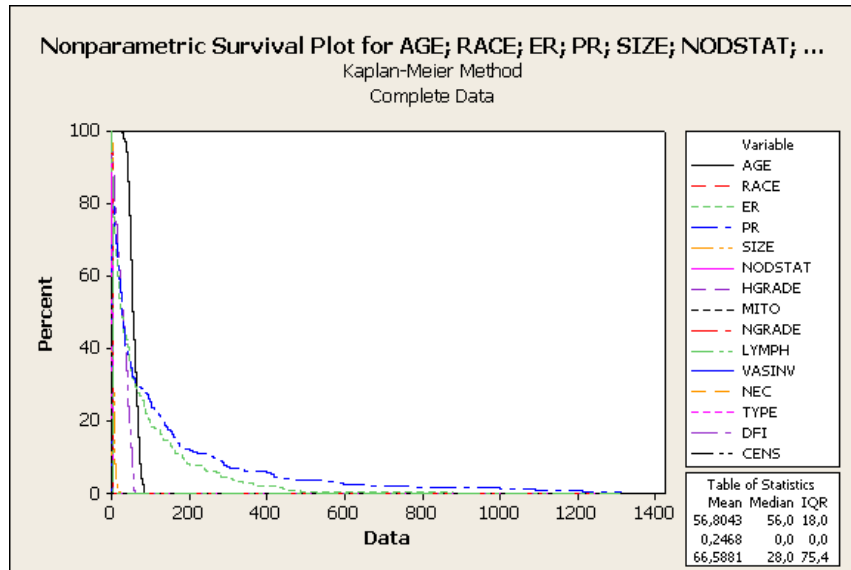


Figura 2: Curvas estimadas

Segue os quartis para algumas variáveis:

Variable: AGE

Censoring Information Count
 Uncensored value 235

Nonparametric Estimates

Characteristics of Variable

	Standard Error	95,0% Normal CI	
Mean(MTTF)		Lower	Upper
56,8043	0,790802	55,2543	58,3542

Median = 56

IQR = 18 Q1 = 48 Q3 = 66

Para a variável idade, pode-se observar que 25% das pacientes tinha idade abaixo de 48 anos e metade das pacientes tinha idade de 56 anos.

Variable: ER

Censoring Information Count
 Uncensored value 235

Nonparametric Estimates

Characteristics of Variable

	Standard	95,0% Normal CI	
Mean(MTTF)	Error	Lower	Upper
66,5881	6,88925	53,0854	80,0908

Median = 28

IQR = 75,4 Q1 = 6,6 Q3 = 82

O valor mediano para a variável Concentração de estrogênio (ER) é 28. Pode-se observar que 75% das informações estão abaixo do valor de ER igual a 82 (fMol/mg).

Para a variável PR, temos:

Variable: PR

Censoring Information Count
 Uncensored value 235

Nonparametric Estimates

Characteristics of Variable

	Standard	95,0% Normal CI	
Mean(MTTF)	Error	Lower	Upper
95,2519	12,1721	71,3951	119,109

Median = 28,4

IQR = 94,2 Q1 = 7,8 Q3 = 102

Logo, a variável PR - (Concentração de progesterona), possui 50% das informações abaixo e acima do valor 28,4 (fMol/mg).

Exercício 5. Analise a saída do SPSS relativa ao Exemplo 4.5

Kaplan-Meier
Survival Analysis for TEMPO
Factor GROUP = C

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
1.00	0			1	20
1.00	0	.9048	.0641	2	19
2.00	0			3	18
2.00	0	.8095	.0857	4	17
3.00	0	.7619	.0929	5	16
4.00	0			6	15
4.00	0	.6667	.1029	7	14
5.00	0			8	13
5.00	0	.5714	.1080	9	12
8.00	0			10	11
8.00	0			11	10
8.00	0			12	9
8.00	0	.3810	.1060	13	8
11.00	0			14	7
11.00	0	.2857	.0986	15	6
12.00	0			16	5
12.00	0	.1905	.0857	17	4
15.00	0	.1429	.0764	18	3
17.00	0	.0952	.0641	19	2
22.00	0	.0476	.0465	20	1
23.00	0	.0000	.0000	21	0

Number of Cases: 21 Censored: 0 (.00%) Events: 21

	Survival Time	Standard Error	95% Confidence Interval
Mean:	8.67	1.41	(5.90, 11.43)
Median:	8.00	1.67	(4.73, 11.27)

A partir da saída do SPSS é possível observar que a mediana do tempo de sobrevivência é de apenas 8 semanas para o grupo controle (placebo).

É possível compararmos os dois grupos em estudo. Logo é possível observar que no tempo $t = 23$ ocorreram duas mortes, uma em cada grupo. Esperava-se $2 \cdot \frac{6}{7} = 1,71$ mortes no grupo tratamento e $2 \cdot \frac{1}{7} = 0,29$ no grupo controle. (log-rank)

Observando os p - valores, com grande certeza existe diferença entre o tempo em remissão entre os dois grupos.

Survival Analysis for TEMPO

Factor GROUP = T

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
6.00	0			1	20
6.00	0			2	19
6.00	0	.8571	.0764	3	18
6.00	1			3	17
7.00	0	.8067	.0869	4	16
9.00	1			4	15
10.00	0	.7529	.0963	5	14
10.00	1			5	13
11.00	1			5	12
13.00	0	.6902	.1068	6	11
16.00	0	.6275	.1141	7	10
17.00	1			7	9
19.00	1			7	8
20.00	1			7	7
22.00	0	.5378	.1282	8	6
23.00	0	.4482	.1346	9	5
25.00	1			9	4
32.00	1			9	3
32.00	1			9	2
34.00	1			9	1
35.00	1			9	0

Number of Cases: 21 Censored: 12 (57.14%) Events: 9

	Survival Time	Standard Error	95% Confidence Interval
Mean: (Limited to 35.00)	23.29	2.83	(17.75, 28.83)
Median:	23.00	5.26	(12.70, 33.30)

Survival Analysis for TEMPO

		Total	Number Events	Number Censored	Percent Censored
GROUP	C	21	21	0	.00
GROUP	T	21	9	12	57.14
Overall		42	30	12	28.57

Test Statistics for Equality of Survival Distributions for GROUP

	Statistic	df	Significance
Log Rank	16.79	1	.0000
Breslow	13.46	1	.0002
Tarone-Ware	15.12	1	.0001

Enquanto que a mediana é de 23 semanas para o grupo tratamento (6-MP).

Pode-se comentar que este teste não compara a \bar{x} ou \tilde{x} e sim a curva de sobrevivência.

*Os valores p são usados de houve riscos proporcionais, 1 gl (porque são dois grupos).
Pode-se observar que há diferença entre os dois grupos controles.*