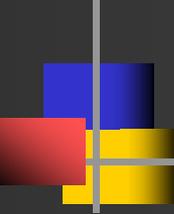


# **Análise de Sobrevivência**

**Gisela Tunes da Silva  
Julio da Motta Singer**

**SP, 14/10/2004  
Revisto em 06/06/2016**



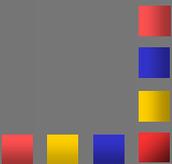
# Análise de Sobrevivência

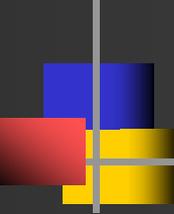
---

- Procedimentos para a análise de dados cuja resposta é

**TEMPO ATÉ OCORRÊNCIA DE UM EVENTO!**

- TEMPO: anos, meses, semanas, dias, horas, etc.
- EVENTO: morte, cura, reincidência da doença, etc.
- Terminologia:
  - TEMPO: tempo de sobrevivência.
  - EVENTO: falha.

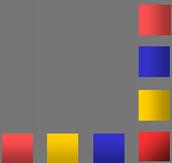


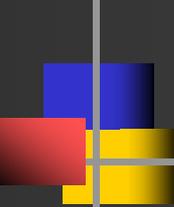


# EXEMPLO: Dados de Câncer Bucal

---

- **OBJETIVO:** comparar a experiência de sobrevivência de pacientes de câncer na cavidade oral submetidos a diferentes tratamentos.
- **TRATAMENTOS:** braquiterapia, radioterapia e externo, quimioterapia e externo além de radioterapia, externo e braquiterapia.
- **DADOS:** 1014 pacientes com câncer na cavidade oral (assoalho de boca, língua, bochecha, gengiva inferior, gengiva superior), acompanhados por um período de 6 anos no Japão.
- **ORIGEM** (do tempo): Início do tratamento.
- **EVENTO:** Óbito.

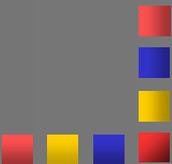


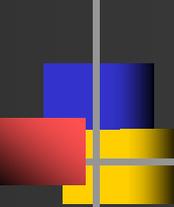


# EXEMPLO: Memória de ratos

---

- Experimento realizado no ICB-USP
- **OBJETIVO:** verificar se lesões em áreas do sistema nervoso de ratos influenciam o padrão de memória.
- 3 grupos de ratos submetidos a cirurgias:
  - **GRUPO 1:** submetidos a lesões pequenas no giro denteado dorsal (região supostamente envolvida com memória espacial).
  - **GRUPO 2:** submetidos a lesões pequenas no giro denteado ventral.
  - **GRUPO 3:** controle (submetidos a cirurgia, sem lesões).

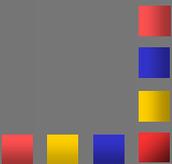


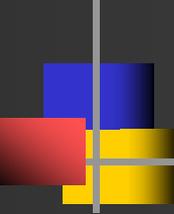


# EXEMPLO: Memória de ratos



- Após recuperação da cirurgia, ratos submetidos a treinamento:
  - Ratos colocados em uma piscina de água turva contendo uma plataforma fixa.
  - Ratos deixados na piscina por um tempo máximo de 2 minutos. Se não encontrasse a plataforma, rato conduzido até ela.
- Depois de 1 semana, mediu-se o tempo até o rato encontrar a plataforma.
- **VARIÁVEL RESPOSTA**: tempo até o encontro da plataforma.





# Definições Importantes

---

## ■ ORIGEM (do tempo)

- Bem definida.
- Comparável.

## EXEMPLOS

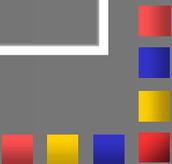
- Início do tratamento
- Aleatorização
- Diagnóstico

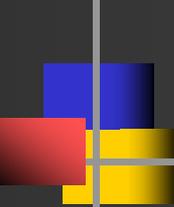
## ■ EVENTO

- Bem definido.

## EXEMPLOS

- Óbito
- Transplante
- Cura

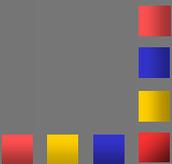




# Censuras

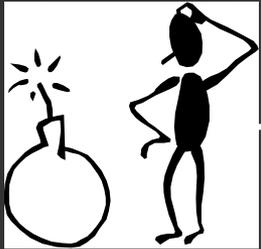
---

- Característica comum em dados de sobrevivência:
  - Tempo exato de sobrevivência não é conhecido.
  - Sabe-se que o tempo é maior que um determinado valor (**tempo de censura**).
- **Possíveis razões para censura:**
  - a unidade amostral (pessoa, rato) não apresenta o evento antes do fim do estudo.
  - Perda de contato com o paciente durante o estudo.
  - O paciente sai do estudo por outros motivos (morte por outra razão/fim do tratamento devido a efeitos colaterais/etc.).
- **Exemplo:** ratos que não encontravam a plataforma em 2 minutos.



# Censuras

- Esquemáticamente:

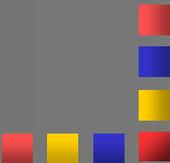


Início  
do  
estudo

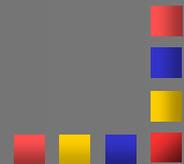
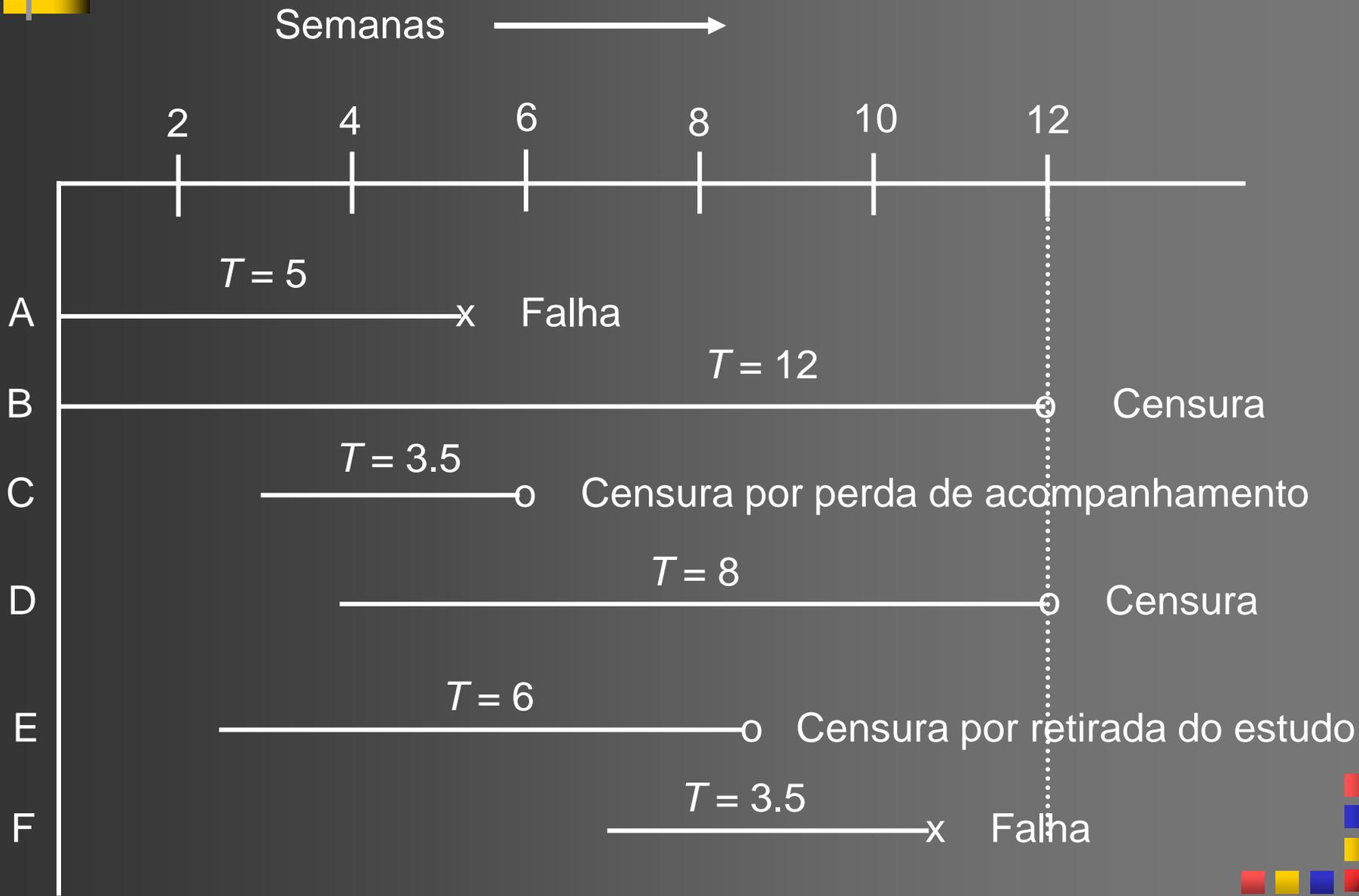
Final  
do  
estudo

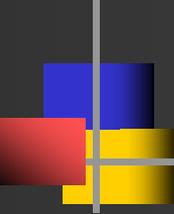
$T$

→ Censura (à direita)



# Censuras

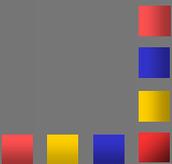


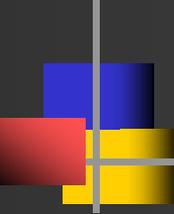


# Organização dos dados

---

Paciente	Tempo de sobrevivência	Falha (1); Censura (0)
A	5	1
B	12	0
C	3.5	0
D	8	0
E	6	0
F	3.5	1



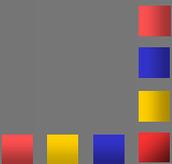


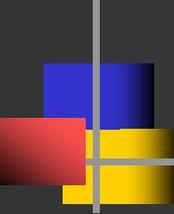
# Notação

---

- $T$ : Tempo de Sobrevivência (ou falha).
  - $T$  é variável aleatória positiva.
- $\delta$ : variável aleatória indicadora de falha

$$\delta = \begin{cases} 1, & \text{para falha} \\ 0, & \text{para censura} \end{cases}$$





# Caracterização de T

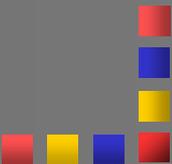
---

- FUNÇÃO DE SOBREVIVÊNCIA:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t)$$

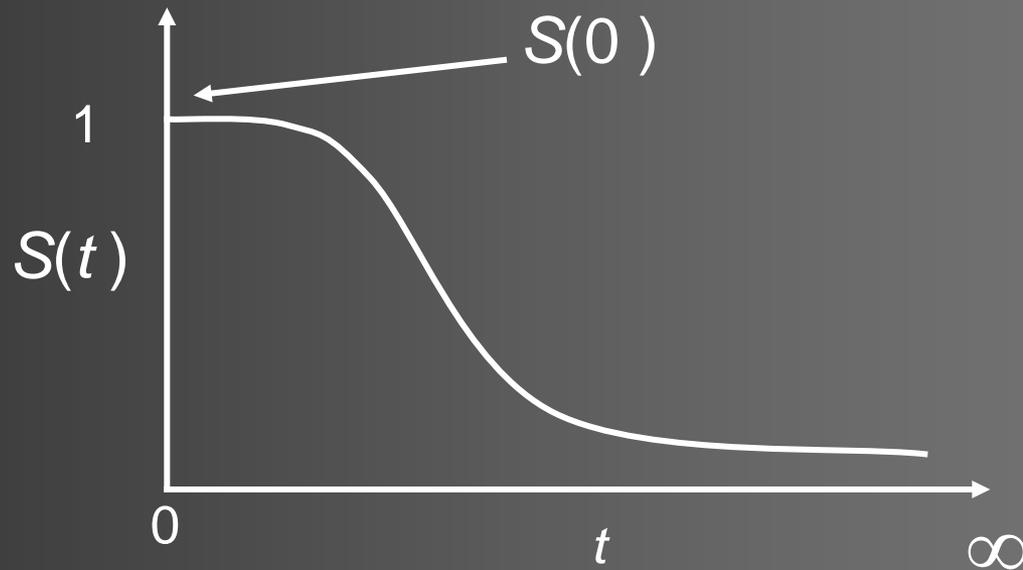
<b>t</b>	<b>S(t)</b>
1	$S(1) = P(T > 1)$
2	$S(2) = P(T > 2)$
3	$S(3) = P(T > 3)$
•	•
•	•
•	•

---

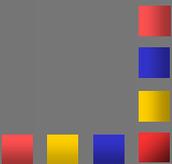


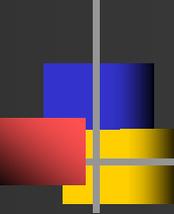
# Função de Sobrevivência

## ■ CURVA DE SOBREVIVÊNCIA TEÓRICA:



- Como estimar  $S(t)$ ?
- Como incorporar observações censuradas?

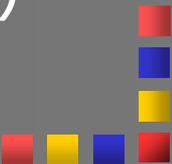


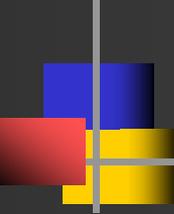


# Caracterização de T

---

- **FUNÇÃO DE RISCO:**  $h(t)$ 
  - “Potencial instantâneo de ocorrência do evento de interesse por unidade de tempo, dado que o paciente sobreviveu até  $t$ ”;
  - “Potencial instantâneo, no instante  $t$ , de ocorrência do evento de interesse, dada a sobrevivência até aquele momento”;
  - “ Fornece o *risco* de ocorrência do evento de interesse no instante  $t$  para um paciente ainda não sujeito ao evento”.
  - $h(t) \geq 0$ ;
  - $h(t)$  não tem um valor máximo (pode ser infinito...)

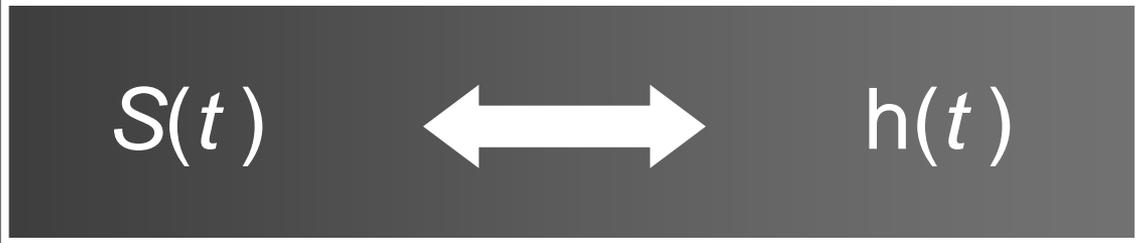




# Usos para a Função de Risco

---

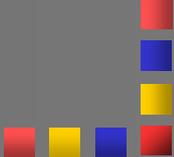
- Fornece uma idéia do comportamento da taxa condicional de falha.
- Ajuda a selecionar modelos probabilísticos adequados ao fenômeno estudado.
- Ajuda a comunicação entre o estatístico e o cliente,
- A forma da função  $h(t)$  determina a forma de  $S(t)$ .



$S(t)$

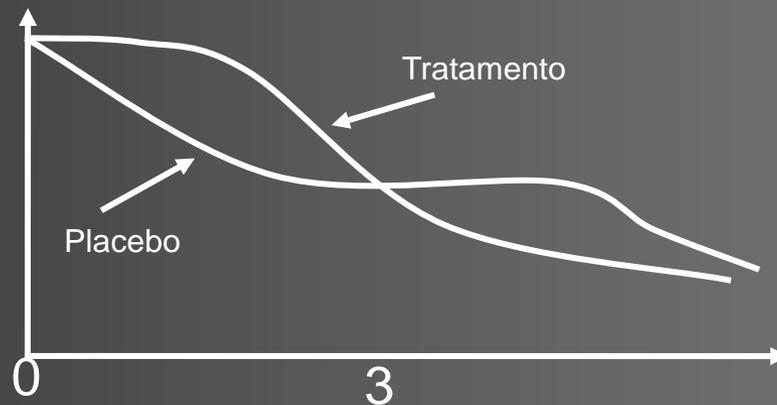


$h(t)$

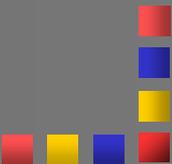


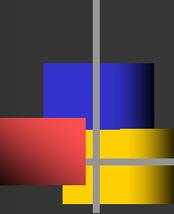
# Objetivos da Análise de Sobrevivência

- Estimar e interpretar a função de sobrevivência.
- Interpretar funções de risco.
- Comparar funções de sobrevivência (ou funções de risco).



- Averiguar a contribuição de fatores de interesse (variáveis explicativas - idade, sexo, etc.) para a sobrevivência de pacientes.





# Estimação da Função de Sobrevivência

---

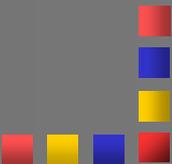
- Para dados não censurados:
  - Função distribuição empírica

$$\bar{F}(x) = \frac{\# obs. \leq x}{n}$$

- Função de sobrevivência pode ser estimada:

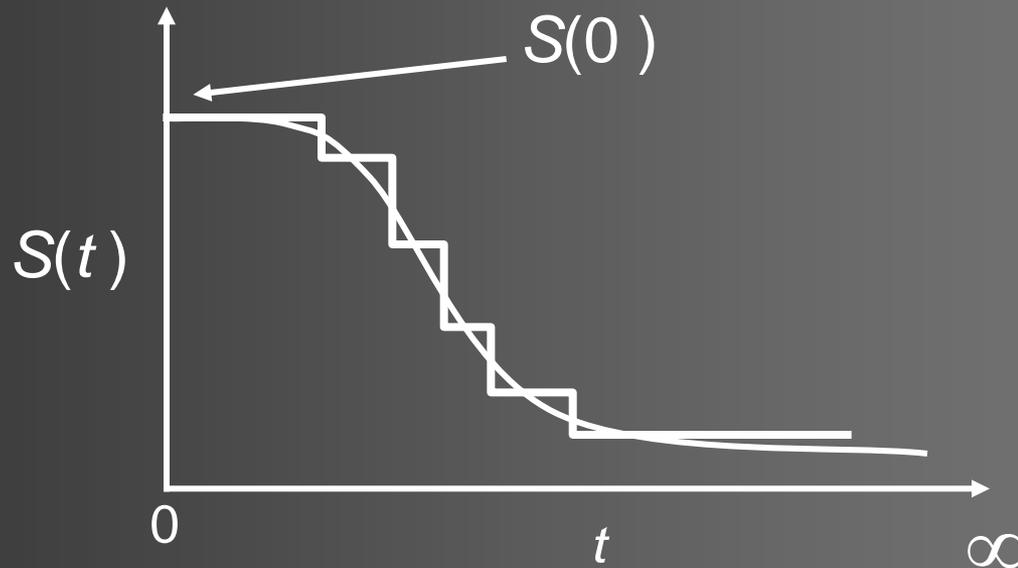
$$\bar{S}(x) = 1 - \bar{F}(x)$$

- Dados censurados??

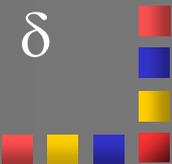


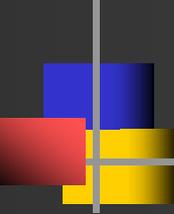
# Estimação da Função de Sobrevivência

- Estimador de Kaplan-Meier:



- Função Escada;
- $S(0) = 1$ ;
- Instantes de falha (quando  $\delta = 1$ ) ou censura (quando  $\delta = 0$ ) precisam ser conhecidos.



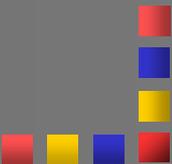


# Estimador de Kaplan-Meier

---

Organização dos dados:

Observ.	Tempo	$\delta$
1	$t_1$	$\delta_1$
2	$t_2$	$\delta_2$
3	$t_3$	$\delta_3$
.	.	.
.	.	.
.	.	.
n	$t_n$	$\delta_n$

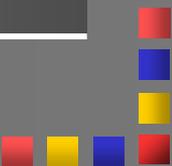


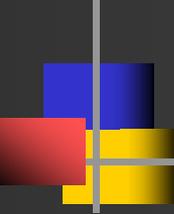
# Estimador de Kaplan-Meier

Outra maneira de organização dos dados:

Tempos de falha ordenados	Número de falhas	Número de censuras em $[t_{(j)}; t_{(j+1)})$	Conjunto em risco
$t_{(0)}=0$	$m_0=0$	$q_0$	$R(t_{(0)})$
$t_{(1)}$	$m_1$	$q_1$	$R(t_{(1)})$
$t_{(2)}$	$m_2$	$q_2$	$R(t_{(2)})$
$t_{(3)}$	$m_3$	$q_3$	$R(t_{(3)})$
·	·	·	·
·	·	·	·
·	·	·	·
$t_{(k)}$	$m_k$	$q_k$	$R(t_{(k)})$

**Conjunto em risco:** número de observações que não falharam nem foram censuradas (sobreviveram) até  $t^-$ .





# Estimador de Kaplan-Meier

---

- Por definição:  $S(t_{(j)}) = P(T > t_{(j)})$

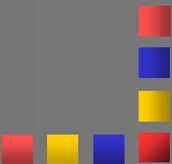
- Lembrando que:  $P(A \cap B) = P(A)P(B | A)$

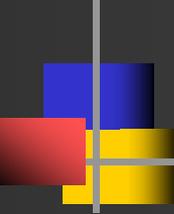
- Pode-se concluir que

$$\begin{aligned} S(t_{(j)}) &= P\left((T > t_{(j-1)}) \cap (T > t_{(j)})\right) \\ &= P\left((T > t_{(j)}) \mid (T > t_{(j-1)})\right)P(T > t_{(j-1)}) \\ &= P\left((T > t_{(j)}) \mid (T > t_{(j-1)})\right)P(T > t_{(j-1)}) \end{aligned}$$

- Aplicando esse resultado várias vezes:

$$S(t_{(j)}) = \prod_{i=1}^j P\left((T > t_{(i)}) \mid (T > t_{(i-1)})\right)$$





# Estimador de Kaplan-Meier

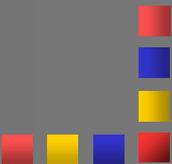
---

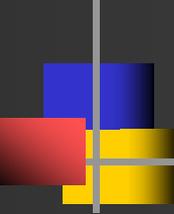
- Denotando

$$p_i = P\left(T > t_{(i)} \mid T > t_{(i-1)}\right)$$

- $p_i$ : probabilidade de sobreviver a  $t_{(i)}$  dado que sobreviveu a  $t_{(i-1)}$ .
- Pode-se estimar

$$\begin{aligned}\hat{p}_i &= \frac{\#(\text{sobreviventes a } t_{(i)})}{\#(\text{sobreviventes a } t_{(i)}^-)} \\ &= \frac{R(t_{(i)}) - m_{(i)}}{R(t_{(i)})}\end{aligned}$$





# Estimador de Kaplan-Meier

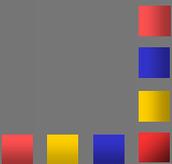
---

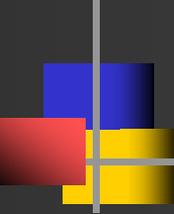
- Estimador Produto-Limite:

$$\hat{S}(t) = \begin{cases} 1 & \text{se } t < t_{(1)}, \\ \prod_{t_{(i)} \leq t} \left[ 1 - \frac{m_{(i)}}{R(t_{(i)})} \right] & \text{se } t_{(i)} \leq t. \end{cases}$$

- Variância do estimador pode ser estimada através da fórmula de Greenwood:

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{t_{(i)} \leq t} \frac{m_{(i)}}{R(t_{(i)}) (R(t_{(i)}) - m_{(i)})}$$

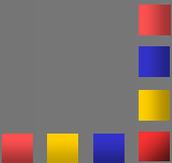




# Kaplan- Meier: Exemplo

---

- Dados (21 observações):  
6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+,  
17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+.
- 12 observações censuradas
- 9 falhas observadas



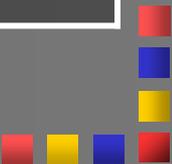
# Kaplan- Meier: Exemplo

6, 6, 6, 7, 10, 13, 16, 22, 23, 6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+.

Tempos de falha	Número de falhas	Núm. censuras $[t_{(j)}; t_{(j+1)})$	Conjunto em risco
0	0	0	21
6	3	1	21
7	1	1	17
10	1	2	15
13	1	0	12
16	1	3	11
22	1	0	7
23	1	5	6
> 23	-	-	0

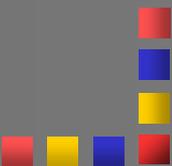
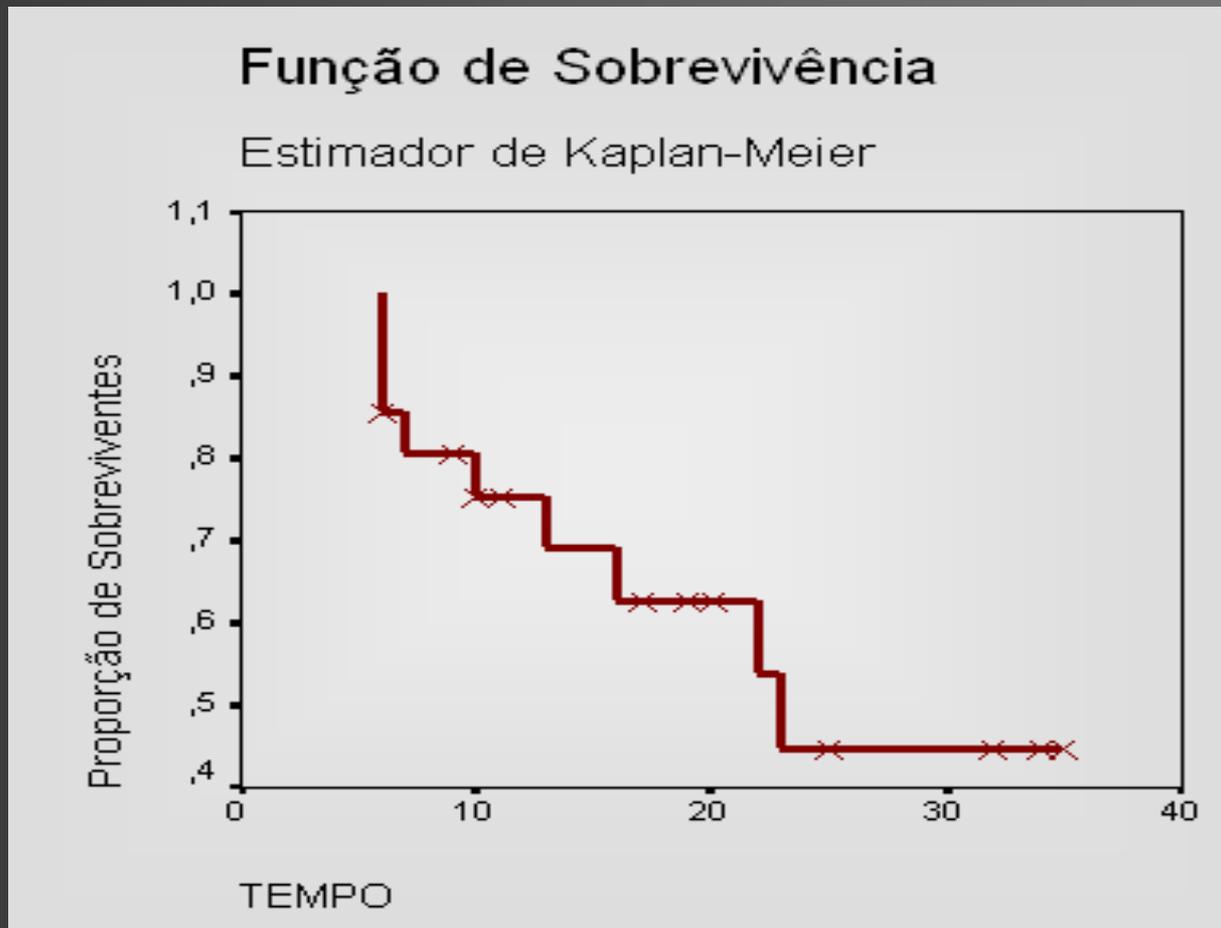
# Kaplan- Meier: Exemplo

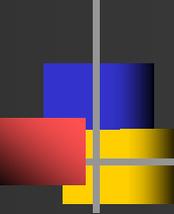
Tempos de falha	Número falhas	Núm. censuras	Conj. risco	$\hat{S}(t_{(j)})$
0	0	0	21	1
6	3	1	21	$1 \times 18/21 = 0,8571$
7	1	1	17	$0,8571 \times 16/17 = 0,8067$
10	1	2	15	$0,8067 \times 14/15 = 0,7529$
13	1	0	12	$0,7529 \times 11/12 = 0,6902$
16	1	3	11	$0,6902 \times 10/11 = 0,6275$
22	1	0	7	$0,6275 \times 6/7 = 0,5378$
23	1	5	6	$0,5378 \times 5/6 = 0,4482$



# Kaplan- Meier: Exemplo

- Graficamente (SPSS):





# Estimador de Kaplan-Meier

---

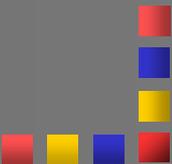
- Através das estimativas Kaplan-Meier, algumas medidas descritivas podem ser calculadas, por exemplo:

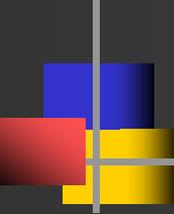
- **Tempo médio** de acompanhamento:

- Variáveis aleatórias positivas:

$$\mu = \int_0^{\infty} S(t)dt$$

- Estimador: área sob a curva!
- OBS.: estimador restrito a  $\tau$  (maior tempo observado/ definido pelo investigador).





# Estimador de Kaplan-Meier

---

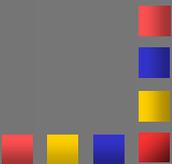
- Através das estimativas Kaplan-Meier, algumas medidas descritivas podem ser calculadas, por exemplo:

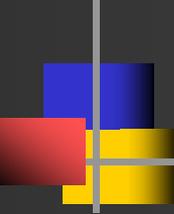
- **Tempo médio** de acompanhamento:

- Variáveis aleatórias positivas:

$$\mu = \int_0^{\infty} S(t)dt$$

- Estimador: área sob a curva!
- OBS.: estimador restrito a  $\tau$  (maior tempo observado/ definido pelo investigador).



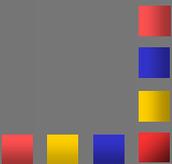


# Estimador de Kaplan-Meier

---

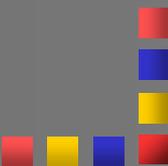
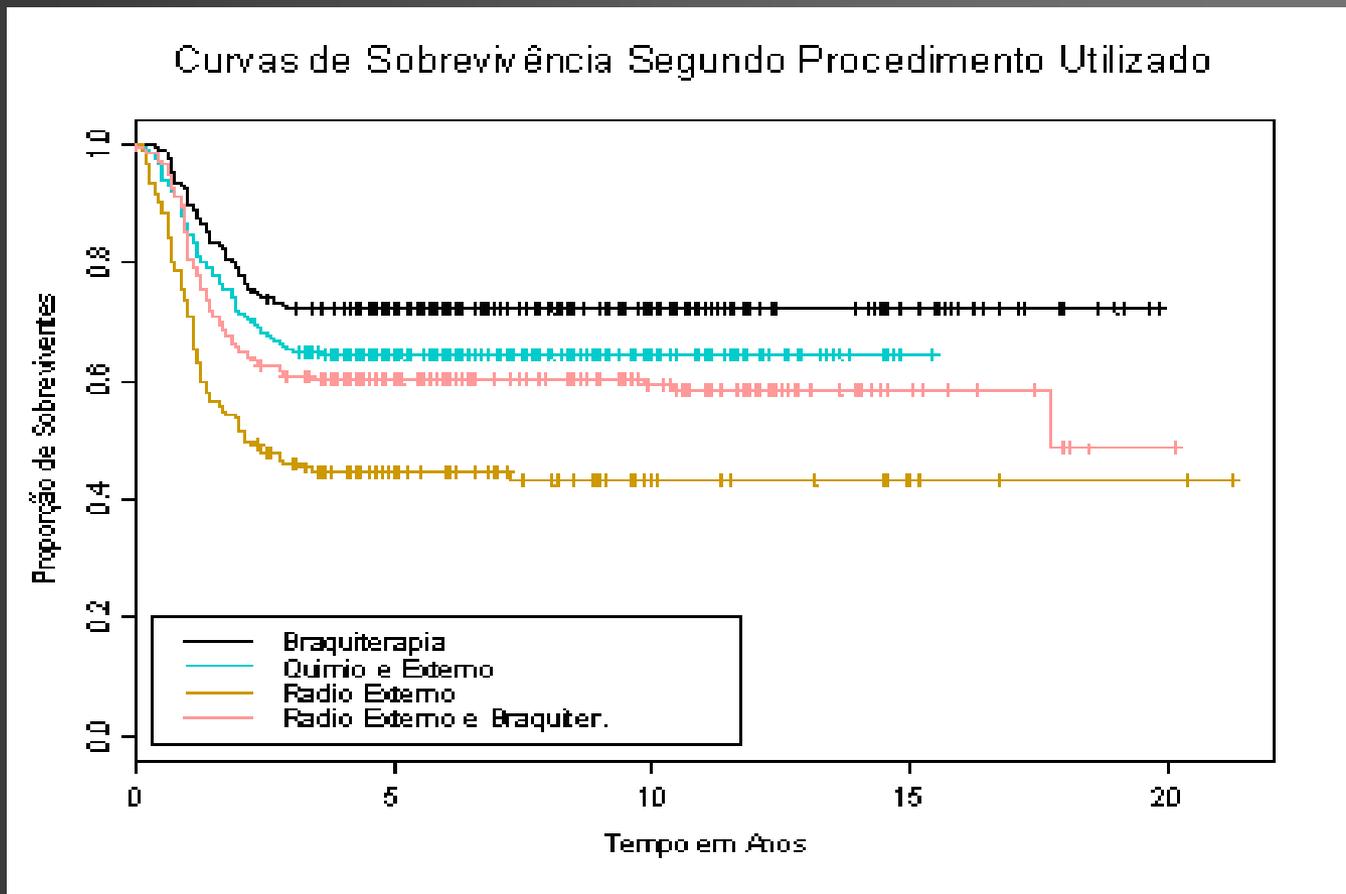
- Tempo mediano de sobrevivência:
  - $T_{med} = \{\inf t : S(t) \leq 1/2\}$ .
- Quantil de ordem  $p$  :
  - $T_p = \{\inf t : S(t) \leq 1-p\}$ .
- No exemplo, temos

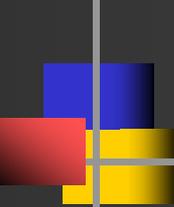
	Estimativa	Erro padrão
<b>Tempo Médio</b>	23,29*	2,83
<b>Tempo Mediano</b>	23,00	5,26



# EXEMPLO: Dados de Câncer Bucal

- Relembrando: 1014 pacientes com câncer na cavidade oral, submetidos a 4 diferentes tratamentos.



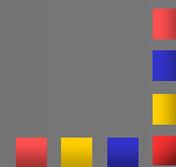


# EXEMPLO: Dados de Câncer Bucal

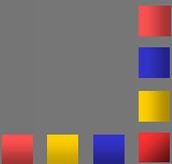
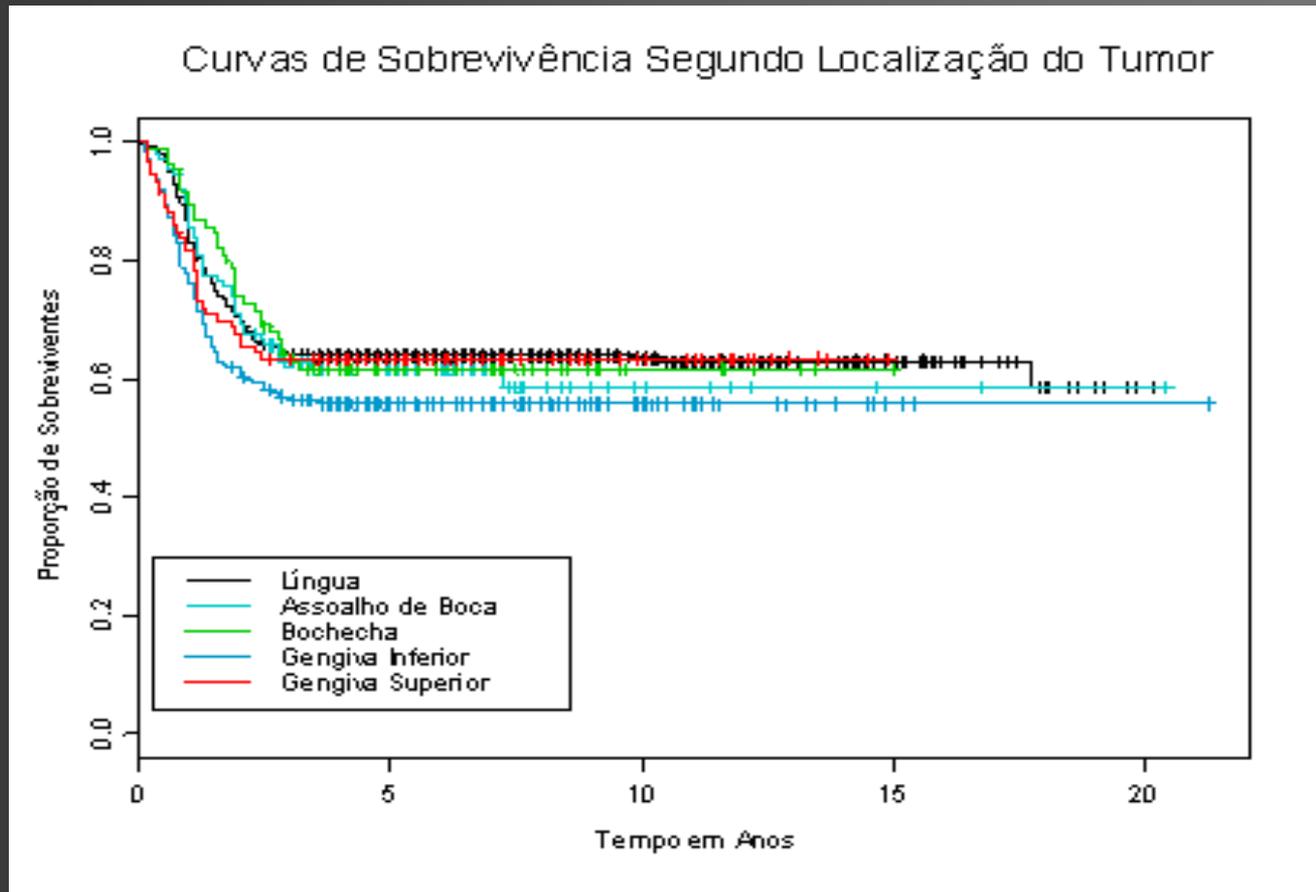
## MEDIDAS DESCRITIVAS

Tratamento	Tempo Médio (meses)	1º. Quartil	Tempo Mediano	3º. Quartil
Braquiterapia	176,94* (6,35)	.	.	27,00 (.)
Químico+ externo	125,08** (4,30)	.	.	22,00 (2,28)
Radio+ externo	118,85*** (9,38)	.	25,00 (7,80)	11,00 (1,15)
Radio, externo, braquiterapia	146,83**** (7,20)	.	213,00 (.)	16,00 (1,74)

Restrito a: \*238; \*\*185; \*\*\*255;\*\*\*\*242

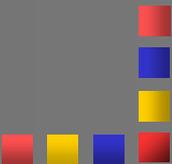
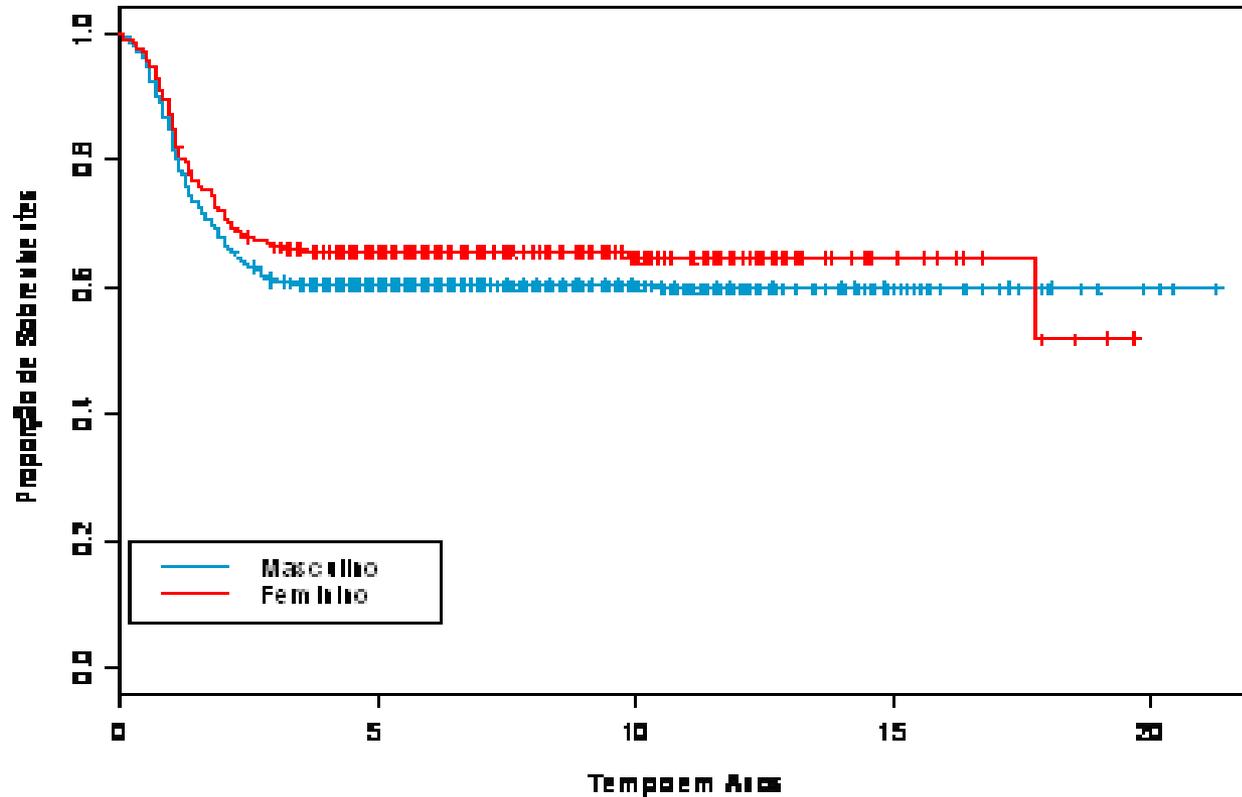


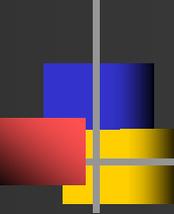
# EXEMPLO: Dados de Câncer Bucal



# EXEMPLO: Dados de Câncer Bucal

Curvas de Sobrevivência Segundo Sexo





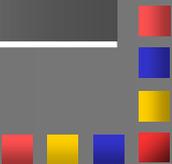
# EXEMPLO: Memória de ratos

---

- 32 ratos, divididos em 3 grupos:
  - **Grupo 1:** lesão em área relacionada com memória espacial.
  - **Grupo 2:** lesão em área não relacionada com memória espacial.
  - **Grupo 3:** controle.

## Porcentagem de censuras:

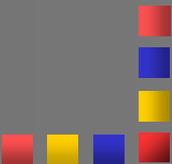
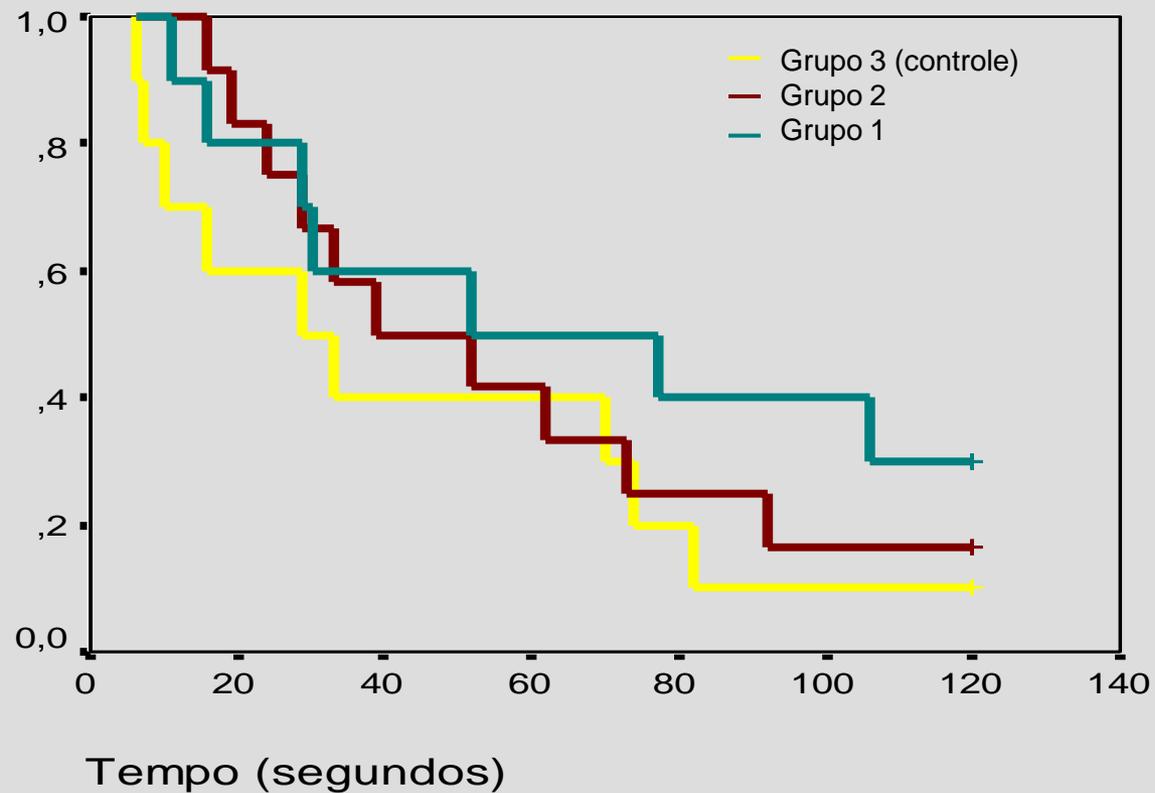
Grupo	Número de observações	Número de censuras	Porcentagem de censuras
1	10	3	30,00%
2	12	2	16,67%
3 (cont.)	10	1	10,00%

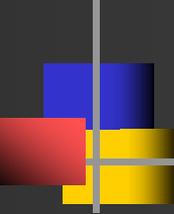


# EXEMPLO: Memória de ratos

## Função de Sobrevivência

Estimador de Kaplan-Meier



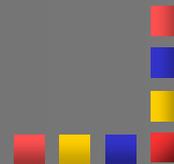


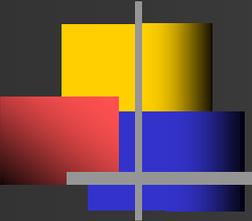
# EXEMPLO: Memória de ratos

## MEDIDAS DESCRITIVAS

Tratamento	Tempo Médio (meses)	1º. Quartil	Tempo Mediano	3º. Quartil
Grupo 1	68,10* (13,72)	.	52,00 (37,16)	29,00 (18,84)
Grupo 2	56,58* (10,33)	73,00 (22,50).	39,00 (16,45)	24,00 (7,5)
Grupo 3 (cont.)	44,70* (11,79)	74,00 (5,06)	29,00 (13,44)	10,00 (4,35)

\*Restrito a 120.





# Bibliografia

---

Colosimo, E.A. e Giolo, S.R. (2006). **Análise de Sobrevivência Aplicada**. São Paulo: Blücher

Kleinbaum, D.G. (1996). **Survival Analysis: a Self-Learning Text**. New York: Springer-Verlag

Lee, E.T. and Wang, J.W. (2013). **Statistical Methods for Survival Data Analysis, 4th Edition**. New York: Wiley

Miller, R.G. and Klein, M. (2011). **Survival Analysis: a self learning text, 3rd Edition**. New York: Springer