

# EPI10 - Análise de Sobrevida

## Estrutura dos dados em análise de sobrevida

Rodrigo Citton P. dos Reis  
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

Porto Alegre, 2022



# A história da análise de sobrevivência e seu progresso

## Um breve histórico

- ▶ Originalmente, a análise de sobrevivência era usada exclusivamente para investigações de mortalidade e morbidade nos registros de estatísticas vitais.
- ▶ A primeira análise aritmética dos processos de sobrevivência humana remonta ao **século XVII**, quando o estatístico inglês **John Graunt** publicou a primeira **tábua de vida** em **1662**.

# Um breve histórico



Natural and Political  
OBSERVATIONS  
Mentioned in a following INDEX,  
and made upon the  
Bills of Mortality.

By JOHN GRAUNT,  
Citizen of  
LONDON.

With reference to the Government, Religion, Trade,  
Growth, Age, Diseases, and the several Change of the  
said CITY.

— Nil, nisi ut miratur Turba, labens.  
Contentus parvis Letificibus —

LONDON,

Printed by Tho: Knaplock, for John Martin, James Allestry,  
and Tho: Dineen, at the Sign of the Bell in St. Paul's  
Church-yard, MDCLXII.

1606.

A TABLE of the  
CHRISTENINGS and MORTALITY  
For the Year 1605 and 1606.\*

Weeks.	Days of the Month.	Child.	Bur.	Pla.	Per. Index.	Weeks.	Days of the Month.	Child.	Bur.	Pla.	Per. Index.
1	Dec. 26.	100	116	5	5	28	July 3.	109	110	25	12
2	January 8.	117	151	6	5	29	10.	111	136	33	18
3	5.	120	158	4	4	30	12.	112	146	50	23
4	16.	124	158	2	2	31	24.	96	140	46	26
5	23.	123	181	6	4	31	31.	131	158	66	29
6	30.	124	181	3	3	31	Aug. 7.	131	151	67	29
7	Feb. 6.	121	105	5	5	34	14.	141	197	75	31
8	13.	131	115	7	6	35	21.	133	189	52	28
9	20.	116	109	12	6	36	28.	132	169	82	29
10	27.	102	137	9	8	37	Sept. 4.	123	241	110	31
11	March 6.	110	98	7	4	38	11.	134	216	105	28
12	13.	186	137	9	7	39	18.	131	214	91	26
13	20.	182	133	14	11	40	25.	133	204	87	28
14	27.	134	133	17	8	41	Oct. 2.	121	256	141	40
15	April 5.	127	114	13	9	42	9.	124	218	106	38
16	10.	123	145	17	11	43	16.	141	227	117	37
17	17.	129	219	17	8	44	23.	135	244	129	38
18	24.	113	110	11	7	45	30.	124	216	101	34
19	May 1.	93	126	17	10	46	Nov. 6.	126	153	66	27
20	8.	116	103	13	11	47	13.	125	165	41	20
21	15.	128	94	13	8	48	20.	131	145	28	11
22	22.	113	128	14	9	49	27.	142	153	23	13
23	29.	94	98	9	7	50	Dec. 4.	150	180	46	17
24	June 5.	129	113	10	8	51	11.	135	137	38	20
25	12.	127	112	19	14	52	18.	136	131	28	15
26	19.	117	119	15	10	53	25.	134	135	38	19
27	26.	121	120	24	16						

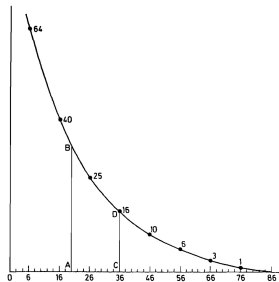
The Totals { Christened — — 6614  
Buried — — 7920  
Whereof of the Plague 1124

\* See London's Remembrance.

A TABLE  
Original from  
THE OHIO STATE UNIVERSITY

# Um breve histórico

Christian Huygens (1629-1695)



## Um breve histórico

- ▶ Por um longo período de tempo, a análise de sobrevivência foi considerada um instrumento analítico, principalmente nos estudos biomédicos e demográficos.
- ▶ Numa fase posterior, expandiu-se gradualmente para o domínio da engenharia para descrever/avaliar o curso dos produtos industriais.

## Um breve histórico

- ▶ Nos últimos cinquenta anos, o escopo da análise de sobrevivência cresceu tremendamente como consequência do rápido **desenvolvimento da ciência da computação**, particularmente o avanço de poderosos pacotes de software estatístico.
- ▶ A conveniência de usar software de computador para criar e utilizar modelos estatísticos complexos levou cientistas de muitas disciplinas a começar a usar modelos de sobrevivência.

## Um breve histórico

- ▶ Como as aplicações da análise de sobrevivência cresceram rapidamente, a inovação metodológica se acelerou em um ritmo sem precedentes nas últimas décadas.
- ▶ O advento do **modelo de Cox** e a perspectiva da **verossimilhança parcial** em **1972** desencadeou o avanço de um grande número de métodos e técnicas estatísticas caracterizadas por modelagem de regressão na análise de dados de sobrevivência.

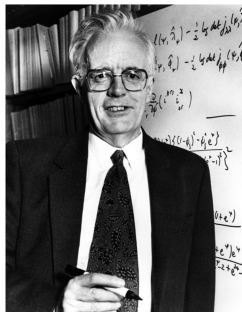


# Um breve histórico

$$\lambda(t) = \lambda_0(t)e^{(x^T \beta)}$$

---


$$\begin{aligned} \ell(\beta|X) &= \sum_j \ln \frac{\exp x_j^T \beta}{\sum_{i:t_i \geq t_j} \exp x_i^T \beta} \\ &= \sum_j \left[ x_j^T \beta - \ln \sum_{i:t_i \geq t_j} \exp x_i^T \beta \right] \end{aligned}$$



## Um breve histórico

- ▶ A principal contribuição do modelo de Cox, dada sua capacidade de gerar procedimentos de estimativa simplificados na análise de dados de sobrevivência, é o fornecimento de uma **abordagem estatística flexível** para modelar os processos de sobrevivência complexos associados a covariáveis mensuráveis.

## Um breve histórico

- ▶ Mais recentemente, o surgimento da **teoria dos processos de contagem** destaca o início de uma nova era na análise de sobrevivência devido ao seu tremendo poder inferencial e alta flexibilidade para modelar **eventos repetidos** para a mesma observação e alguns outros processos de sobrevivência complexos.
  - ▶ Em particular, esta perspectiva moderna combina elementos da **teoria de grandes amostras**, a **teoria de martingales** e a **teoria de integração estocástica**, fornecendo um novo conjunto de procedimentos estatísticos e regras na modelagem de dados de sobrevivência.

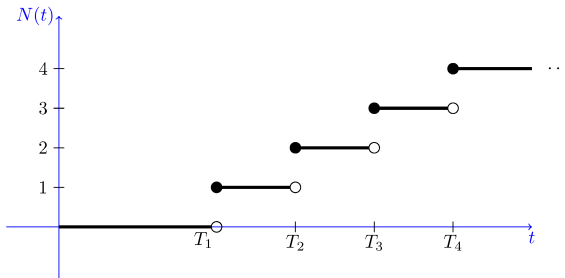
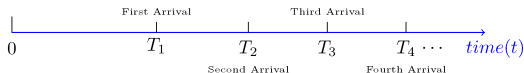
# Características gerais da estrutura de dados de sobrevivência

## Estrutura dos dados

- ▶ Em essência, um processo de sobrevivência descreve um período de vida de um tempo de início especificado até a ocorrência de um evento particular.
- ▶ Portanto, a **principal característica** dos dados de sobrevivência é a descrição de uma **mudança no status** como medida de resultado subjacente.
- ▶ Mais formalmente, uma mudança de status é a ocorrência de um **evento** que designa o fim de um período de vida ou o término de um processo de sobrevivência.

## Estrutura dos dados

- Por exemplo, uma mudança de status ocorre quando uma pessoa morre, se casa ou quando um automóvel quebra.



## Estrutura dos dados

- ▶ Esta característica de um “salto” de status torna a análise de sobrevivência um tanto semelhante a algumas perspectivas estatísticas mais convencionais sobre dados de desfechos qualitativos, como o **modelo logístico**.
- ▶ Esses modelos tradicionais também podem ser usados para examinar uma mudança de status ou a ocorrência de um determinado evento, comparando o status no início e o status no final de um intervalo de observação.
- ▶ Entretanto, essas abordagens **ignoram o tempo de ocorrência** desse evento e, portanto, não possuem a capacidade de descrever um processo de tempo até o evento.
  - ▶ A falta dessa capacidade pode ser prejudicial à qualidade dos resultados analíticos, gerando conclusões enganosas.

# Estrutura dos dados

## Exemplo

- ▶ Suponha que dois grupos da população têm a mesma taxa de ocorrência de um determinado evento até o final de um período de observação, mas os membros de um grupo costumam experimentar o evento significativamente mais tarde do que aqueles no outro.
- ▶ O primeiro grupo populacional tem um padrão de sobrevivência vantajoso porque sua vida média é estendida.
- ▶ Obviamente, a regressão logística ignora esse fator de tempo, não fornecendo informações precisas.



## Estrutura dos dados

- ▶ A maioria dos modelos de sobrevivência leva em consideração o **fator tempo** no salto de status.
- ▶ Dada essa capacidade, a **segunda característica** dos dados de sobrevivência é a **descrição de um processo de tempo até o evento**.
- ▶ Na literatura de análise de sobrevivência, **o tempo de ocorrência de um** determinado **evento** é considerado uma **variável aleatória**, referida como **tempo do evento**, **tempo de falha** ou **tempo de sobrevivência**.

## Estrutura dos dados

- ▶ A **terceira característica** principal da estrutura de dados de sobrevivência é a **censura**.
- ▶ Os dados de sobrevivência são geralmente coletados por um intervalo de tempo no qual as ocorrências de um determinado evento são observadas.
- ▶ Como resultado, os pesquisadores só podem observar os eventos que ocorrem dentro de uma janela de monitoramento entre dois limites de tempo<sup>1</sup>.

---

<sup>1</sup>Tempo de seguimento (*follow up time*).

## Estrutura dos dados

- ▶ Consequentemente, os tempos de sobrevivência completos para muitas unidades sob exame não são observados, com perda de informações ocorrendo antes do início ou após o final do intervalo do estudo.
- ▶ Algumas unidades podem ser perdidas para observação no meio de uma investigação devido a vários motivos.
- ▶ Na análise de sobrevivência, esse status ausente em tempos de evento é chamado de censura, que pode ser dividido em uma variedade de tipos.

## Estrutura dos dados

- ▶ Para a maioria dos tipos de censura, uma porção do tempo de sobrevivência para observações censuradas é observável e pode ser utilizada no cálculo do risco de experimentar um determinado evento.
  - ▶ Na análise de sobrevivência, esta porção dos tempos observados é referida como **tempos de sobrevivência censurados**.
- ▶ Como a censura ocorre com frequência, a maioria das análises de sobrevivência lida literalmente com **dados de sobrevivência incompletos** e, conseqüentemente, os cientistas encontraram maneiras de usar essas informações limitadas para analisar corretamente os dados de sobrevivência incompletos com base em algumas suposições restritivas sobre a distribuição dos tempos de sobrevivência censurados.

# Estrutura dos dados

- ▶ Como os processos de sobrevivência variam essencialmente com base nas características básicas das observações e condições ambientais, um campo considerável de análise de sobrevivência é conduzido por meio de **modelagem de regressão de dados censurados** envolvendo uma ou mais variáveis preditoras.

# Estrutura dos dados

- ▶ Dada a adição de **covariáveis**, a **estrutura de dados de sobrevivência** pode ser vista como consistindo em informações sobre três fatores principais:
  - ▶ tempos de sobrevivência;
  - ▶ status de censura;
  - ▶ e covariáveis.
- ▶ Dada uma **amostra aleatória** de  $n$  unidades, a estrutura de dados para análise de sobrevivência na verdade contém  $n$  destas triplas.
- ▶ A maioria dos modelos de sobrevivência é construída sobre essa estrutura de dados.

# Estrutura dos dados

- Dadas as diferentes ênfases na variedade de recursos, a análise de sobrevivência também é conhecida como **análise de duração**, **análise de tempo até o evento**, **análise de histórico de eventos** ou **análise de dados de confiabilidade**.

# Por hoje é só!

Bons estudos!

