

Estatística aplicada à epidemiologia II

Modelos para desfecho binário

Leo Bastos – leonardo.bastos@fiocruz.br

PROCC – Fundação Oswaldo Cruz

<https://github.com/lsbastos/eae2>

- Vamos explorar o exemplo do famoso naufrágio de Titanic em 1912.
- O dado original consiste em um array com 4 dimensões dos 2201 passageiros contendo as seguintes informações:
 - Survived: {Yes, No}
 - Sex: {Male, Female}
 - Age: {Child, Adult}
 - Class: {1st, 2nd, 3rd, Crew}
- “Mulheres e crianças primeiro!”, será que isso aconteceu no naufrágio do Titanic?
- Como podemos verificar isso?

- Podemos explicar a probabilidade de sobreviver segundo sexo e idade
- Podemos comparar as probabilidades (ou avaliar as ORs) associadas a sobrevivência entre os sexos, e entre os dois grupos de idade.
- Então, basta calcularmos as ORs

$$OR_{Sex} = \frac{Odds(Female)}{Odds(Male)}$$

e

$$OR_{Sex} = \frac{Odds(Child)}{Odds(Adult)}$$

Lendo os dados no R

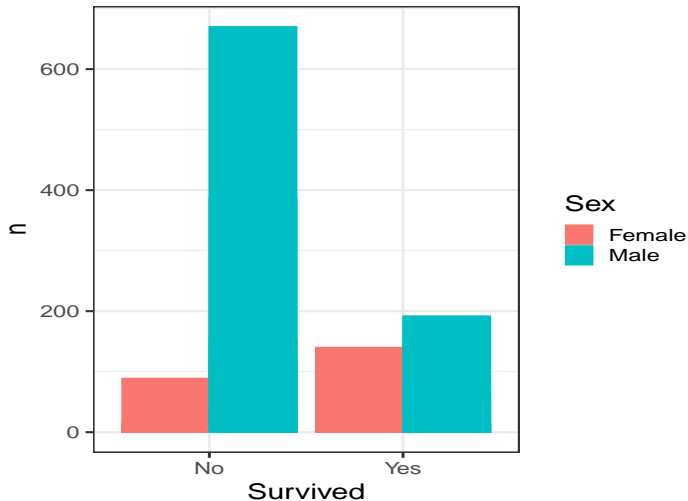
- Chamando o banco
`> data(Titanic)`
- O dado é um data.frame de 4 dimensões, podemos transformá-lo em uma tabela (tibble)

```
> library(tidyverse)
> Titanic2 <- as_tibble(Titanic) %>%
+   spread(Survived, n)
> head(Titanic2)
```

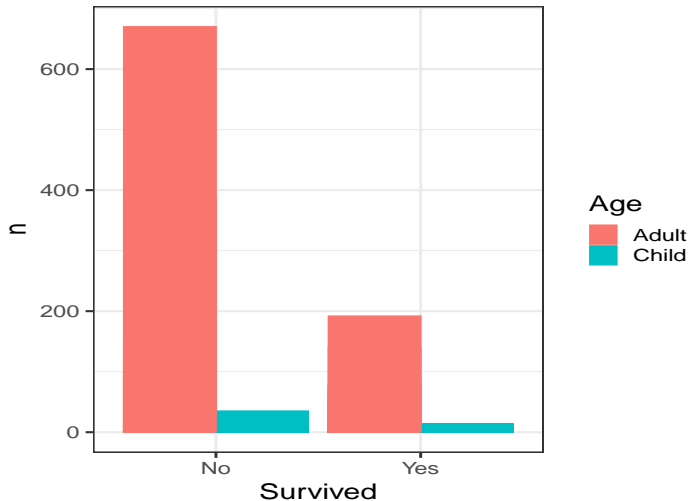
A tibble: 6 × 5

	Class	Sex	Age	No	Yes
	<chr>	<chr>	<chr>	<dbl>	<dbl>
1	1st	Female	Adult	4	140
2	1st	Female	Child	0	1
3	1st	Male	Adult	118	57
4	1st	Male	Child	0	5
5	2nd	Female	Adult	13	80
6	2nd	Female	Child	0	13

Visualizando: Sobrevivência por sexo



Visualizando: Sobrevivência por idade



Tabelas para Sexo e Idade

- Tabela para Sexo

	No	Yes
Male	1364	367
Female	126	344

- Tabela para idade

	No	Yes
Child	52	57
Adult	1438	654

- ORs? Existe associação significativa?

- Ambas associações são significativas, via teste chi-quadrado ($p\text{-value} < 0.001$)
- As chances de uma mulher sobreviver ao naufrágio são 10.14 vezes as chances de um homem sobreviver.
- As chances de uma criança sobreviver ao naufrágio são 2.41 vezes as chances de um adulto sobreviver.

- O modelo logístico tem a forma

$$Y_i \sim \text{Ber}(\theta_i)$$

onde $Y_i = 1$ se o indivíduo sobreviveu, $g(\theta_i) = \alpha + \beta X_i$

- Se o dado estiver agregado, cada linha representa um conjunto de categorias, podemos escrever o modelo da seguinte forma
- O modelo logístico tem a forma

$$Z_k \sim \text{Bin}(n_k, \theta_k)$$

onde Z_k é o número de sobreviventes na linha k , n_k o número de passageiros na linha k , e $g(\theta_k) = \alpha + \beta X_k$

Olhando o dado agregado

- O dado em sua forma agregada

```
> head(Titanic2)
```

```
# A tibble: 6 × 5
```

	Class	Sex	Age	No	Yes
	<chr>	<chr>	<chr>	<dbl>	<dbl>
1	1st	Female	Adult	4	140
2	1st	Female	Child	0	1
3	1st	Male	Adult	118	57
4	1st	Male	Child	0	5
5	2nd	Female	Adult	13	80
6	2nd	Female	Child	0	13

- $Y \sim \text{Sex}$

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.3128	0.0588	-22.33	0.0000
SexFemale	2.3172	0.1196	19.38	0.0000

- $Y \sim \text{Age}$

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.7879	0.0472	-16.71	0.0000
AgeChild	0.8797	0.1975	4.45	0.0000

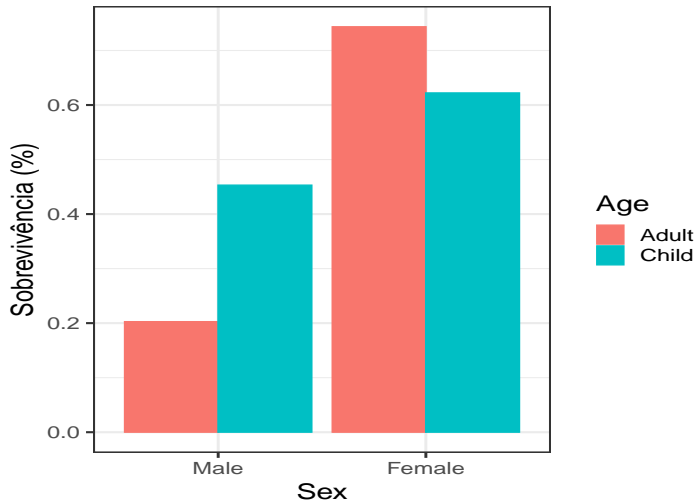
- $OR_{\text{Sex}} = 10.15$ e $OR_{\text{Age}} = 2.41$

- Podemos encontrar as ORs ajustadas, i.e. a OR do sexo ajustada por idade, e a OR da idade ajustada pelo sexo.
- $Y \sim \text{Age} + \text{Sex}$

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.3365	0.0599	-22.32	0.0000
SexFemale	2.2940	0.1199	19.13	0.0000
AgeChild	0.5564	0.2276	2.44	0.0145

- $OR_{\text{Sex}} = 9.91$ e $OR_{\text{Age}} = 1.74$
- O efeito da idade mudou pouco, já o efeito do sexo sofreu uma mudança considerável.

Visualizando: Sobrevivência por idade e sexo



- $Y \sim \text{Age}$ (para os homens)

	Estimate	Std. Error	z value	$\text{Pr}(> z)$
(Intercept)	-1.3691	0.0609	-22.47	0.0000
AgeChild	1.1811	0.2584	4.57	0.0000

- $Y \sim \text{Age}$ (para as mulheres)

	Estimate	Std. Error	z value	$\text{Pr}(> z)$
(Intercept)	1.0644	0.1111	9.58	0.0000
AgeChild	-0.5654	0.3269	-1.73	0.0837

Criando uma categoria única Sexo e idade

- Agora vamos criar uma categoria para sexo e idade, vamos chamala de SexAge.
- $Y \sim \text{AgeSex}$

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.3691	0.0609	-22.47	0.0000
AgeSexChild:Male	1.1811	0.2584	4.57	0.0000
AgeSexChild:Female	1.8681	0.3134	5.96	0.0000
AgeSexAdult:Female	2.4335	0.1267	19.21	0.0000

- Lembrando que podemos recuperar as probabilidades aplicando a função logit inversa,

$$\theta = \text{logit}^{-1}\left(\alpha + \sum_i \text{AgeSex}_i \beta_i\right)$$

- Resultando na tabela:

	Levels	OR	SurvProb	LI	LS
1	Adult:Male	1.000	0.203	0.184	0.223
2	Child:Male	3.258	0.453	0.336	0.575
3	Child:Female	6.476	0.622	0.474	0.751
4	Adult:Female	11.399	0.744	0.700	0.783

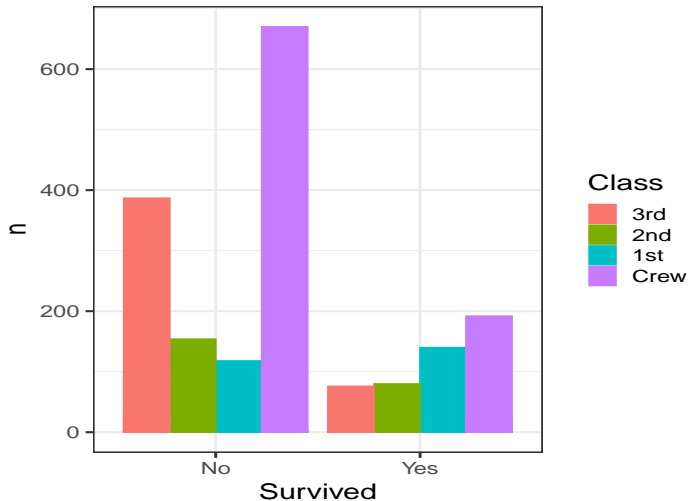
- Comparando modelos via AIC

	Modelos	AIC
1	1 + Age	703.02
2	1 + Sex	288.11
3	1 + Age + Sex	284.22
4	1 + AgeSex	269.90

Conclusão?

- E aí, Mulheres e crianças primeiro?
- Mulheres adultas primeiro, depois crianças!

Visualizando: Sobrevivência por classe



Modelo adicionando a Classe

- $Y \sim \text{AgeSex} + \text{Class}$

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.2461	0.1320	-17.01	0.0000
AgeSexChild:Male	1.7921	0.2829	6.33	0.0000
AgeSexChild:Female	2.5072	0.3332	7.52	0.0000
AgeSexAdult:Female	2.6172	0.1510	17.33	0.0000
Class2nd	0.7767	0.1777	4.37	0.0000
Class1st	1.8105	0.1759	10.29	0.0000
ClassCrew	1.0072	0.1530	6.58	0.0000

- ORs por classe

	OR
Class2nd	2.17
Class1st	6.11
ClassCrew	2.74

- ORs (Idade e sexo)

	OR.crude	OR.adj
AgeSexChild:Male	3.26	6.00
AgeSexChild:Female	6.48	12.27
AgeSexAdult:Female	11.40	13.70

- $Y \sim \text{AgeSex}$ (AIC = 269.9)
- $Y \sim \text{AgeSex} + \text{Class}$ (AIC = 155.2)

Incluindo a interação com a Classe

- $Y \sim \text{AgeSex} + \text{AgeSex} : \text{Class}$ (AIC= 74.6)

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.6409	0.1262	-13.01	0.0000
AgeSexChild:Male	0.6505	0.3484	1.87	0.0619
AgeSexChild:Female	1.4468	0.3823	3.78	0.0002
AgeSexAdult:Female	1.4830	0.2008	7.39	0.0000
AgeSexAdult:Male:Class2nd	-0.7570	0.3063	-2.47	0.0135
AgeSexChild:Male:Class2nd	27.1934	89551.9055	0.00	0.9998
AgeSexChild:Female:Class2nd	26.5477	88813.3546	0.00	0.9998
AgeSexAdult:Female:Class2nd	1.9750	0.3374	5.85	0.0000
AgeSexAdult:Male:Class1st	0.9133	0.2048	4.46	0.0000
AgeSexChild:Male:Class1st	26.5269	95180.5436	0.00	0.9998
AgeSexChild:Female:Class1st	24.7602	131010.7755	0.00	0.9998
AgeSexAdult:Female:Class1st	3.7133	0.5306	7.00	0.0000
AgeSexAdult:Male:ClassCrew	0.3912	0.1504	2.60	0.0093
AgeSexAdult:Female:ClassCrew	2.0550	0.6385	3.22	0.0013

Probabilidades de sobrevivência

