

Titanic

Lucas Zanolla

6 de abril de 2018

Com a Tabela do Titanic, é possível comparar alguns dados para se obter informações relevantes. Utilizando a ferramenta R (junto com o R Markdown) foi feito cálculos de probabilidade e estatística para demonstrar essas informações. Como método de Data Science, é importante notar fatores(variáveis) que serão influentes na pesquisa. Neste caso do Titanic, as variáveis mais importantes para serem analisadas são: Sexo, idade e classe dos passageiros em relação a quantidade de sobreviventes. Para começar após a leitura dos dados do Titanic, foram criados novos Datasets para serem usados como comparação ao original e ver quais tem uma porcentagem de sobrevivência maior. Foram criados datasets só com as mulheres, só com os homens, idosos, crianças e adultos.

```
#titanic = read.table("titanic.txt")
female <- titanic[titanic$Sex == "female", ]
male <- titanic[titanic$Sex == "male", ]

criancas <- titanic[titanic$Age < 5, ]
idosos <- titanic[titanic$Age > 65, ]
adultos <- titanic[titanic$Age < 65 && titanic$Age > 5, ]

prop.table(table(titanic$Survived)) *100
```

```
##
##      0      1
## 61.61616 38.38384
```

```
prop.table(table(titanic$Sex)) *100
```

```
##
## female   male
## 35.2413 64.7587
```

```
prop.table(table(male$Survived)) *100
```

```
##
##      0      1
## 81.10919 18.89081
```

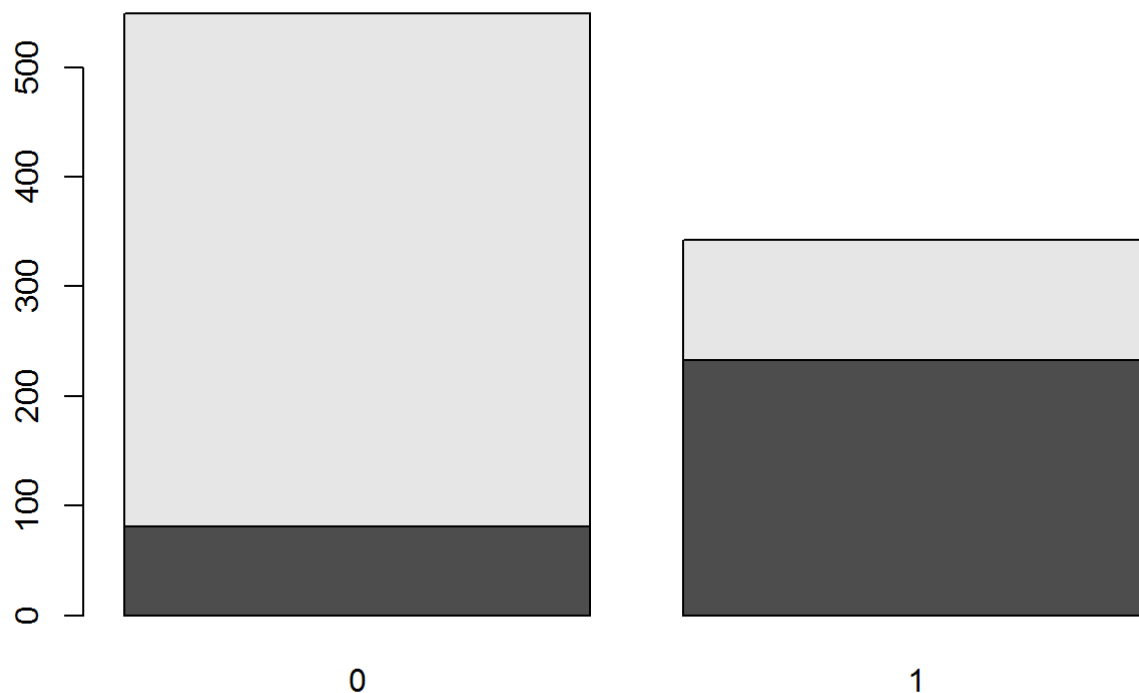
```
prop.table(table(female$Survived)) *100
```

```
##
##      0      1
## 25.79618 74.20382
```

```
prop.table(table(titanic$Sex, titanic$Survived)) *100
```

```
##
##           0           1
##  female  9.090909 26.150393
##   male   52.525253 12.233446
```

```
barplot(table(titanic$Sex, titanic$Survived))
```



Os resultados são muito interessantes, primeiramente o fato de somente 38.38% das pessoas sobreviveram ao desastre, e entre essas 38.38% 26.15% eram mulheres e somente 12.23% eram homens. Para analisar os dados de forma mais individual, nota-se que 64.7% da população do Titanic era Masculina e somente 35.3% era feminina.

Entre os Homens, somente 18.89% deles sobreviveram, diferentemente do dado coletado no qual entre as Mulheres 74.2% sobreviveram! Algo muito interessante pelo fato de ter uma diferença tão grande.

O Gráfico demonstra a diferença, mostrando em tom claro os Homens e em tom escuro as Mulheres. (dados tendo em vista que 1 é verdadeiro com o termo “sobreviventes”)

Tendo em vista esta diferença, é interessante fazer o mesmo estudo em relação à idade das pessoas embarcadas.

```
prop.table(table(criancas$Survived))*100
```

```
##
##    0    1
## 32.5 67.5
```

```
prop.table(table(criancas$Sex, criancas$Survived)) *100
```

```
##  
##           0      1  
##  female 12.5 30.0  
##   male   20.0 37.5
```

```
prop.table(table(adultos$Survived))*100
```

```
##  
##           0      1  
## 61.61616 38.38384
```

```
prop.table(table(adultos$Sex, adultos$Survived)) *100
```

```
##  
##           0      1  
##  female  9.090909 26.150393  
##   male   52.525253 12.233446
```

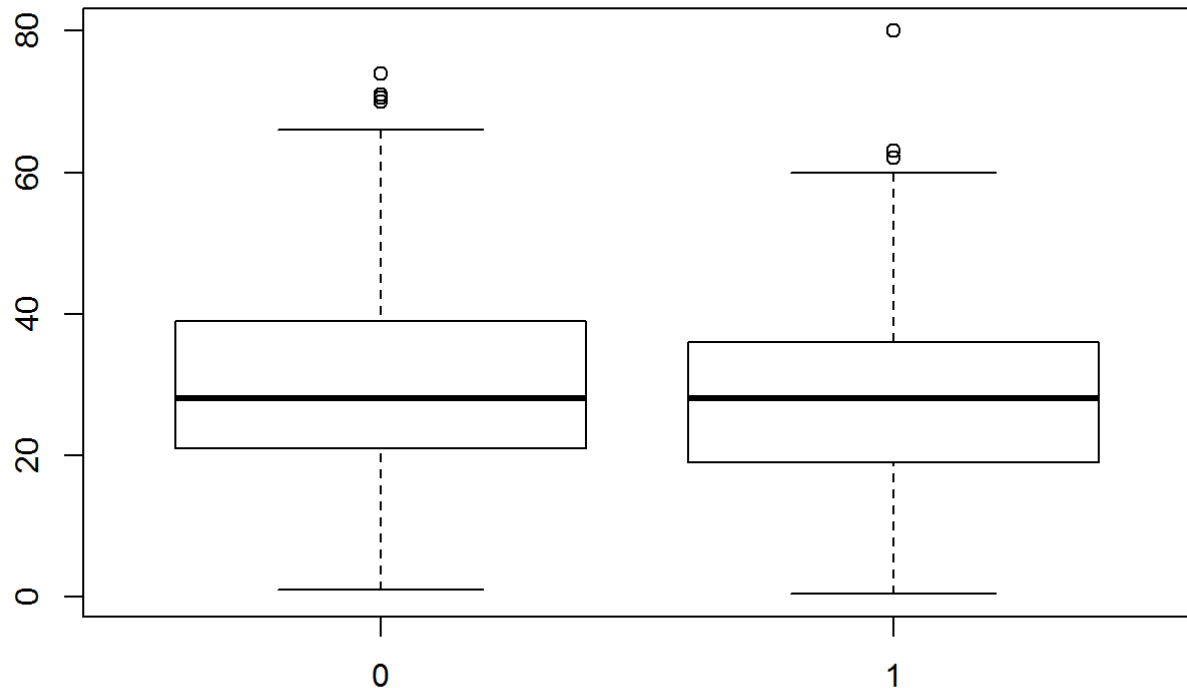
```
prop.table(table(idosos$Survived))*100
```

```
##  
##           0      1  
## 87.5 12.5
```

```
prop.table(table(idosos$Sex, idosos$Survived)) *100
```

```
##  
##           0      1  
##  female  0.0  0.0  
##   male   87.5 12.5
```

```
boxplot(titanic$Age ~ titanic$Survived)
```



Este resultado demonstrou que 67.5% das pessoas com menos de 5 anos na época, foram salvas, e curiosamente, diferente do total o fato de ser do sexo masculino ou feminino não importou tanto para esse estudo. sendo que dos 67.5% dos sobreviventes 30% era do sexo feminino e 37% do sexo masculino. Uma diferença basicamente irrelevante ao comparar que há mais homens com menos de 5 anos do que mulheres (embarcadas).

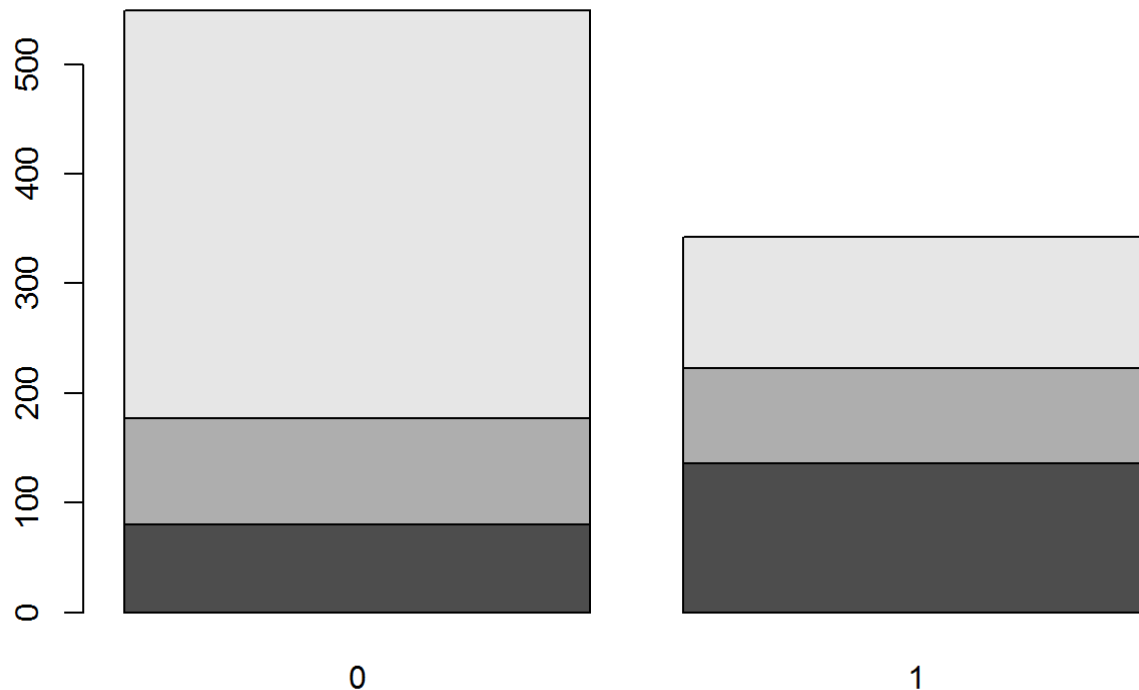
A grande diferença foi entre os “adultos” (maior que 5 anos e menor que 65) onde somente 38% sobreviveram, e a grande maioria mulher (com 26%) sendo que a maioria era homem (74% do total de pessoas entre essa idade). Isso ajuda a entender os dados prévios.

Curiosamente o grupo de “idosos” (maior que 65 anos) só havia homens e somente 12.5% deles sobreviveram.

```
prop.table(table(titanic$Pclass, titanic$Survived)) *100
```

```
##
##           0           1
##  1  8.978676 15.263749
##  2 10.886644  9.764310
##  3 41.750842 13.355780
```

```
barplot(table(titanic$Pclass, titanic$Survived))
```



Já obtendo vários dados importantes, por último é curioso o fato de que a Classe dos passageiros também influencia e muito na sobrevivência, entre os 24% que estavam em primeira classe 15% deles sobreviveram, e o valor vai diminuindo em relação a segunda classe, onde basicamente metade sobreviveu, e não a grande maioria como a primeira classe. A terceira classe foi a que mais sofreu com a tragédia, tendo 41% do total de passageiros que morreram, eram da terceira classe.

Com esse Data Science, é possível fazer uma conclusão, na qual a maior porcentagem de sobreviventes são mulheres, crianças e quem estava na primeira classe!