

# Você sobreviveria no Titanic? Um exemplo de Regressão Logística

(<http://www.abgconsultoria.com.br/blog/sobreviveria-no-titanic-um-exemplo-de-regressao-logistica/>)

28 de ago de 2017

A tragédia do Titanic, em 1912, é um dos desastres marítimos mais conhecidos da história. Eternizado no cinema em 1997, o navio construído em Belfast (Irlanda) naufragou quatro dias após sua viagem inaugural, que tinha como destino a cidade de Nova Iorque. Quando construído, o navio prometia ser o mais luxuoso e seguro de sua época. Entretanto, estudos posteriores indicaram falhas no sistema de segurança e evacuação. A estimativa é de 1514 mortes entre os 2224 passageiros, ou seja, aproximadamente 68% da tripulação.

**Caso você estivesse no navio, qual seria a sua chance de sobrevivência?**

A tragédia do Titanic é um clássico exemplo de aplicação da **Regressão Logística**. A Regressão Logística faz parte de uma família de modelos chamada Modelos Lineares Generalizados (GLM) (<http://www.abgconsultoria.com.br/blog/modelos-lineares-generalizados/>) e é adequada quando a variável de interesse (resposta) é binária, isto é, “sim” ou “não”. Através da Regressão Logística é possível avaliar os fatores de influenciam a ocorrência de determinado evento.

Leia também: O que são Modelos Lineares Generalizados? (<http://www.abgconsultoria.com.br/blog/modelos-lineares-generalizados/>)



No Titanic, a variável resposta é a morte ou a sobrevivência do passageiro. Disponível na internet, há um banco de dados referente a 891 indivíduos que estavam na tripulação do navio. O banco está disponível no Kaggle (<https://www.kaggle.com/datasets>) e é necessário o cadastro no site para o download gratuito da base. No banco de dados em questão estão disponíveis algumas informações sobre os tripulantes, mas neste exemplo consideraremos suas respectivas idades, sexo, classe (1ª a 3ª) e desfecho (morte ou sobrevivência). Dessa forma, é possível calcular a probabilidade de sobrevivência do indivíduo a partir destas características.

Na base de dados faltam registros sobre a idade de 177 indivíduos e, como o objetivo é uma análise didática, os mesmos foram excluídos das análises. Dessa forma, as análises seguintes são referentes a 714 passageiros do navio (cerca de

32% do total de tripulantes). Vale ressaltar que existem técnicas mais sofisticadas para o tratamento de dados perdidos, que podem ser discutidas posteriormente em nossos artigos.

## Função de Ligação da Regressão Logística e a Razão de Chances

Como uma característica do GLM, a **Regressão Logística utiliza a função de ligação “logit”, o que possibilita a interpretação dos resultados em função da Razão de Chances (Odds Ratio).**

A função de ligação conecta os fatores de influência à variável resposta. Por exemplo, na Regressão Linear Simples, que não possui nenhuma função de ligação, a variável de interesse (resposta) pode assumir qualquer valor nos números reais e, dessa forma, a função de ligação “logit” limita os valores previstos pelo modelo ao intervalo 0 e 1, como uma característica da resposta. A Razão de Chances pode ser definida como a razão de um evento ocorrer em um grupo A em função de um grupo B.

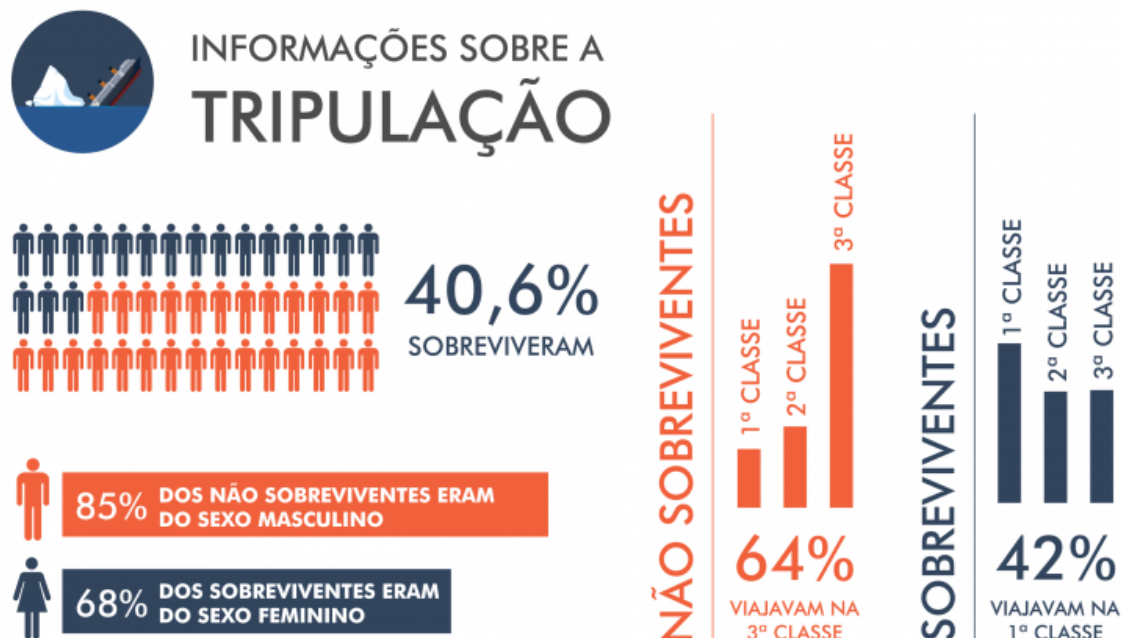


Como mostrado nos momentos finais do filme estrelado por Leonardo DiCaprio e Kate Winslet, mulheres e crianças eram prioridade na evacuação do navio. Dessa forma, é esperado que a chance de sobrevivência dos indivíduos do sexo Feminino (grupo A) seja superior a chance de sobrevivência do sexo Masculino (grupo B). Neste caso, por exemplo, a Razão de Chances da variável ‘Sexo’, em relação a variável resposta, seria a razão entre a chance de sobrevivência de indivíduos do sexo Feminino e Masculino.

## Características da tripulação

Na amostra, 424 indivíduos (59,38%) não sobreviveram e 290 (40,62%) sobreviveram. Cerca de 68% dos sobreviventes eram do sexo feminino e 85% dos não sobreviventes eram do sexo masculino. Em relação a classe, 42% dos sobreviventes eram da 1ª Classe e 64% dos não sobreviventes eram da 3ª Classe.

A idade média dos indivíduos que não sobreviveram foi de 30,62 anos, sendo a idade mínima de 1 ano e a máxima de 74 anos. A Idade média dos indivíduos que sobreviveram foi de 28,24 anos, sendo a idade mínima menor que 1 ano e a máxima de 80 anos.



## Ajuste do modelo logístico

O ajuste do modelo é mostrado abaixo e é possível observar que todas as variáveis foram significativas ao nível de 5% de significância (valor-p < 0,050). Logo, a chance de sobrevivência para indivíduos do sexo Feminino foi 12,46 [8,3; 18,71] vezes maior que a chance de sobrevivência para os indivíduos do sexo Masculino.

Em relação a chance de sobrevivência dos indivíduos da 3ª a Classe, a Chance de sobrevivência para os indivíduos da 2ª Classe foi 3,36 [2,21; 5,75] vezes maior e para os indivíduos da 1ª Classe a chance de sobrevivência foi 13,21 [7,61; 22,93] vezes maior. Sobre a idade, a cada ano acrescido na idade do indivíduo a chance de sobrevivência foi 0,96 [0,95; 0,98] vezes menor.

### Ajuste do Modelo de Regressão Logística

Variáveis	$\beta$	E.P.( $\beta$ )	O.R.	I.C. – 95%	Valor P
Sexo = Masculino	–	–	1,00	–	–

Variáveis	$\beta$	E.P.( $\beta$ )	O.R.	I.C. – 95%	Valor P
Sexo = Feminino	2,52	0,21	12,46	[8,3 ; 18,71]	0,000
Classe = 3ª	–	–	1,00	–	–
Classe = 2ª	1,27	0,24	3,56	[2,21 ; 5,75]	0,000
Classe = 1ª	2,58	0,28	13,21	[7,61 ; 22,93]	0,000
Idade	-0,04	0,01	0,96	[0,95 ; 0,98]	0,000



## Equação do Modelo de Regressão Logística

Sendo  $g(X)$  a função de ligação, a equação do modelo é dada por:

$$g(X) = -1,33 + 2,55 * (I_{Feminino}) + 1,27 * (I_{2ª\ classe}) + 2,58 * (I_{1ª\ classe}) - 0,04 * (Idade)$$

" $I_{Feminino}$ " recebe 1 se o indivíduo é do sexo feminino e 0 se é do sexo masculino, " $I_{2ª\ Classe}$ " recebe 1 se o indivíduo navegava na 2ª Classe e 0 caso contrário e " $I_{1ª\ Classe}$ " recebe 1 se o indivíduo navegava na 1ª Classe e 0 caso contrário.

Para calcular a probabilidade de sobrevivência de um indivíduo,  $E(Y)$ , é necessário aplicar uma função inversa a função de ligação, sendo representada por:

$$E(Y) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Dessa forma, uma Mulher de 42 anos que estava na 1ª Classe tem probabilidade de sobrevivência de 0,90, enquanto que uma mulher de mesma idade na 3ª Classe tem probabilidade de sobrevivência de 0,41. Um homem de 30 anos de idade da 1ª Classe tem probabilidade de sobrevivência de 0,42, enquanto que um homem de mesma idade na 2ª Classe tem probabilidade de sobrevivência de 0,16.

Rose DeWitt Bukater, interpretada por Kate Winslet, era uma garota de 17 anos que estava na primeira classe. Jack Dawson, interpretado por Leonardo DiCaprio, tinha 20 anos e estava na terceira classe. Segundo o modelo logístico, a probabilidade de sobrevivência de Rose é de aproximadamente 0,96, enquanto que a probabilidade de sobrevivência de Jack é de aproximadamente 0,11.

## Qual seria a sua probabilidade de sobrevivência?

Elaboramos uma calculadora para que você saiba qual seria a sua probabilidade de sobrevivência caso estivesse viajando no Titanic. Clique aqui e faça a simulação!

Calcule sua probabilidade de sobrevivência no Titanic! (<http://www.abgconsultor>

A aplicação ao evento de morte no Titanic é apenas um exemplo ilustrativo. Caso queira saber mais sobre o **modelo de regressão logística**, entre em contato (<http://www.abgconsultoria.com.br/contato>) com nossos consultores. E não deixe de se inscrever em nosso blog para conferir nossos artigos!

Artigo desenvolvido com a colaboração de Leonardo Gonçalves.

## Gostou deste conteúdo?

Inscreva-se para receber artigos, eBooks, planilhas, dicas e guias exclusivos.

Cadastrar



---

A ABG Consultoria Estatística nasceu em 2008 para transformar dados em informações confiáveis para a tomada de decisão. Desde então, sua atuação no mercado acadêmico a mantém atualizada nas metodologias de análise mais recentes, permitindo-a oferecer para as empresas soluções com o que há de mais avançado no mercado. Registrada no Conselho Regional de Estatística da 6ª Região sob o número 099.

---

Receba notícias, dicas e novidades sobre **Estatística**



Quero receber

---

☎ +55 31 2516-0068 (tel:+55-31-2516-0068)

✉ contato@abgconsultoria.com.br (mailto:contato@abgconsultoria.com.br)

🏠 Edifício Vivaldi - Av. Bias Fortes, 349, Sala 302

Bairro Lourdes, Belo Horizonte - MG, CEP: 30170-011

Principal (<http://www.abgconsultoria.com.br>)

/ Para Você (<http://www.abgconsultoria.com.br/estatistica-para-voce>)

/ Para Sua Empresa (<http://www.abgconsultoria.com.br/estatistica-para-empresas>)

/ Cursos (<http://www.abgconsultoria.com.br/cursos-e-treinamentos>)

/ Blog (<http://www.abgconsultoria.com.br/blog>) / Sobre (<http://www.abgconsultoria.com.br/sobre>)

/ Contato (<http://www.abgconsultoria.com.br/contato>)

Copyrights 2016 © ABG Consultoria Estatística

