

**CENTRO UNIVERSITA´RIO INSTITUTO DE EDUCAC¸ A˜O SUPERIOR DE BRAS´ILIA**

**Bacharelado em**

**Ciˆencia de Dados e Inteligˆencia Artificial**

***Modelos Lineares Generalizados***

**Fl´avia Guimar˜aes Gaia Paula**

**Regress˜ao Log´ıstica aplicada no dataset titanic**

**Bras´ılia 2022**

# Sum´ario

1. [Introdução](#_bookmark0) 4
2. [Desenvolvimento](#_bookmark1) 5
3. [Conclusão](#_bookmark2) 7

[Referências](#_bookmark3) 8

[Anexo A – Código](#_bookmark5) em linguagem R 10

# 1 Introdução

O naufra´gio do Titanic, aconteceu em abril de 1912 e ´e considerado um dos acidentes mar´ıtimos mais famosos da hist´oria. O navio tinha sido um s´ımbolo das conquistas industriais de Belfast e quando constru´ıdo prometia ser o mais luxuoso e seguro da ´epoca. O tema fez tanto sucesso que foi para o cinema em 1997como uma hist´oria de fic¸c˜ao do naufra´gio real do RMS Titanic, estrelando Leonardo DiCaprio como Jack Dawson, e Kate Winslet como Rose DeWitt Bukater. Estima-se que houveram 1514 mortes entre os 2224 passageiros do navio, ou seja, cerca de 68

Neste trabalho ser´a utilizado o dataset do software R do caso titanic tendo como vari´avel resposta ´e a morte ou a sobrevivˆencia do passageiro no acidente com o navio RMS Titanic. Para isso se aplica a t´ecnica de regress˜ao log´ıstica mu´ltipla, que faz parte da fam´ılia de Modelos Lineares Generalizados (GLM) e ´e adequada quando a varia´vel de interesse (resposta) ´e bin´aria, isto ´e, “sim” ou “n˜ao”. Atrav´es da Regress˜ao Log´ıstica ´e poss´ıvel avaliar os fatores de influenciam a ocorrˆencia de determinado evento.

# Desenvolvimento

Fun¸c˜ao de Liga¸c˜ao da Regress˜ao Log´ıstica e a Raz˜ao de Chances Como uma caracter´ıstica do GLM, a Regress˜ao Log´ıstica utiliza a fun¸c˜ao de liga¸c˜ao “logit”, o que possibilita a interpreta¸c˜ao dos resultados em fun¸c˜ao da Raz˜ao de Chances (Odds Ratio).

A func¸a˜o de ligac¸a˜o conecta os fatores de influˆencia a` varia´vel resposta. Por exemplo, na Regress˜ao Linear Simples, que n˜ao possui nenhuma fun¸c˜ao de liga¸c˜ao, a vari´avel de interesse (resposta) pode assumir qualquer valor nos nu´meros reais e, dessa forma, a func¸a˜o de liga¸c˜ao “logit” limita os valores previstos pelo modelo ao intervalo 0 e 1, como uma caracter´ıstica da resposta. A Raz˜ao de Chances pode ser definida como a raz˜ao de um evento ocorrer em um grupo A em fun¸c˜ao de um grupo B.

Como mostrado nos momentos finais do filme estrelado por Leonardo DiCaprio e Kate Winstley, mulheres e crianc¸as eram prioridade na evacuac¸a˜o do navio. Dessa forma,

´e esperado que a chance de sobrevivˆencia dos indiv´ıduos do sexo Feminino (grupo A) seja superior a chance de sobrevivˆencia do sexo Masculino (grupo B). Neste caso, por exemplo, a Raza˜o de Chances da varia´vel ‘Sexo’, em relac¸a˜o a varia´vel resposta, seria a raza˜o entre a chance de sobrevivˆencia de indiv´ıduos do sexo Feminino e Masculino.

Caracter´ısticas da tripula¸c˜ao Na amostra, 424 indiv´ıduos (59,38

A idade m´edia dos indiv´ıduos que na˜o sobreviveram foi de 30,62 anos, sendo a idade m´ınima de 1 ano e a ma´xima de 74 anos. A Idade m´edia dos indiv´ıduos que sobreviveram foi de 28,24 anos, sendo a idade m´ınima menor que 1 ano e a m´axima de 80 anos.

Ajuste do modelo log´ıstico O ajuste do modelo ´e mostrado abaixo e ´e poss´ıvel observar que todas as vari´aveis foram significativas ao n´ıvel de 5

Em rela

a˜o a chance de sobrevivˆencia dos indiv´ıduos da 3ª a Classe, a Chance de

sobrevivˆencia para os indiv´ıduos da 2ª Classe foi 3,36 [2,21; 5,75] vezes maior e para os

indiv´ıduos da 1ª Classe a chance de sobrevivˆencia foi 13,21 [7,61; 22,93] vezes maior. Sobre

a idade, a cada ano acrescido na idade do indiv´ıduo a chance de sobrevivˆencia foi 0,96 [0,95; 0,98] vezes menor.

Ajuste do Modelo de Regress˜ao Log´ıstica Vari´aveis E.P.() O.R. I.C. – 95Sexo = Masculino – – 1,00 – – 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

Equa¸c˜ao do Modelo de Regress˜ao Log´ıstica Sendo g(X) a fun¸c˜ao de liga¸c˜ao, a equa¸c˜ao do modelo ´e dada por:

“IFeminino“ recebe 1 se o indiv´ıduo ´e do sexo feminino e 0 se ´e do sexo masculino,

“I2ª Classe“ recebe 1 se o indiv´ıduo navegava na 2ª Classe e 0 caso contra´rio e “I1ª Classe“ recebe 1 se o indiv´ıduo navegava na 1ª Classe e 0 caso contr´ario.

Para calcular a probabilidade de sobrevivˆencia de um indiv´ıduo, E(Y) , ´e necessa´rio aplicar uma fun¸c˜ao inversa a fun¸c˜ao de liga¸c˜ao, sendo representada por:

Dessa forma, uma Mulher de 42 anos que estava na 1ª Classe tem probabilidade de sobrevivˆencia de 0,90, enquanto que uma mulher de mesma idade na 3ª Classe tem probabilidade de sobrevivˆencia de 0,41. Um homem de 30 anos de idade da 1ª Classe tem probabilidade de sobrevivˆencia de 0,42, enquanto que um homem de mesma idade na 2ª Classe tem probabilidade de sobrevivˆencia 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 log´ıstico, a probabilidade de sobrevivˆencia de Rose ´e de aproximadamente 0,96, enquanto que a probabilidade de sobrevivˆencia de Jack ´e de aproximadamente 0,11.

Qual seria a sua probabilidade de sobrevivˆencia? Elaboramos uma calculadora para que vocˆe saiba qual seria a sua probabilidade de sobrevivˆencia caso estivesse viajando no Titanic. Clique aqui e fa¸ca a simula¸c˜ao!

A aplicac¸a˜o ao evento de morte no Titanic ´e apenas um exemplo ilustrativo. Caso queira saber mais sobre o modelo de regress˜ao log´ıstica, entre em contato com nossos consultores. E n˜ao deixe de se inscrever em nosso blog para conferir nossos artigos!

# Conclusão

Dessa forma, conclui-se que uma Mulher de 42 anos que estava na 1ª Classe tem probabilidade de sobrevivˆencia de 0,90, enquanto que uma mulher de mesma idade na 3ª Classe tem probabilidade de sobrevivˆencia de 0,41. Um homem de 30 anos de idade da 1ª Classe tem probabilidade de sobrevivˆencia de 0,42, enquanto que um homem de mesma idade na 2ª Classe tem probabilidade de sobrevivˆencia 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 log´ıstico, a probabilidade de sobrevivˆencia de Rose ´e de aproximadamente 0,96, enquanto que a probabilidade de sobrevivˆencia de Jack ´e de aproximadamente 0,11.

# Referências

* <https://smolski.github.io/livroavancado/reglog.html#regressao-logistica-multipla-com-variavel-categorica>;
* [https://www.abgconsultoria.com.br/blog/voce-sobreviveria-no-titanic-um-exemplo-de- regressao-logistica/](https://www.abgconsultoria.com.br/blog/voce-sobreviveria-no-titanic-um-exemplo-de-%20%20%20%20%20%20regressao-logistica/);
* <https://cdr.ibpad.com.br/revishttps://rpubs.com/JulhinhaM/ProjTitanic>;
* <https://datascienceplus.com/perform-logistic-regression-in-r/>;
* <https://rviews.rstudio.com/2019/03/01/some-r-packages-for-roc-curves/>;

# Anexo A – Código em linguagem R

A seguir esta descrito todo o c´odigo em linguagem R utilizados para gerar as an´alises e resultados deste trabalho.