



Escola Livre de Inteligência Artificial

Inteligência Artificial ao alcance de todos

Aula 22/06/2021: Regressão Logística Binária

Professor: Eng. Rodolfo Magliari de Paiva



Objetivos da Aula

- Compreender o que é uma Análise de Regressão e seus tipos;
- Aprender a interpretar um Gráfico de Dispersão;
- Compreender o sentido e o objetivo de se efetuar uma Regressão Logística Binária;
- Efetuar uma Regressão Logística Binária.





Análise de Regressão

Parte da Estatística que estuda a relação entre duas ou mais variáveis (dependentes e independentes), de modo que seja possível identificar quais variáveis possuem maior ou menor impacto em um fenômeno de estudo, além de também permitir a explicação de um fenômeno e prever o futuro.

Para isso, utilizamos os chamados **Modelos de Regressão**.



Os Modelos de Regressão podem ser Lineares (ML) ou Não Lineares (MNL):

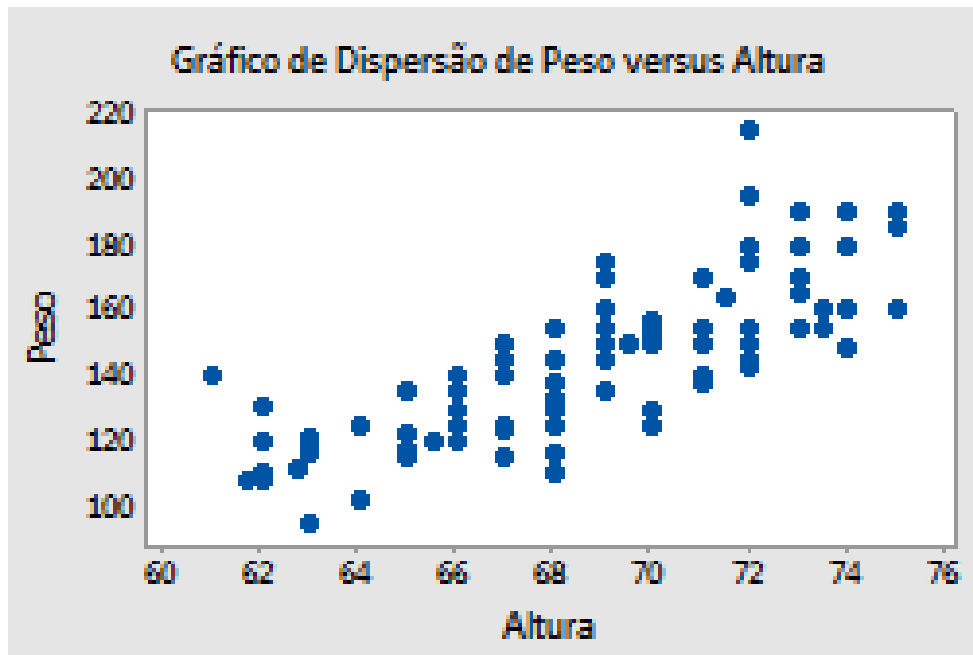
- **Modelos de Regressão Linear:**
 - Regressão Linear Simples;
 - Regressão Linear Múltipla.
- **Modelos de Regressão Não Linear:**
 - Regressão Logística Binária;
 - Regressão Logística Multinomial;
 - Regressão Logística Ordinal;
 - Regressão Exponencial;
 - Regressão Poisson;
 - ...



Gráfico de Dispersão

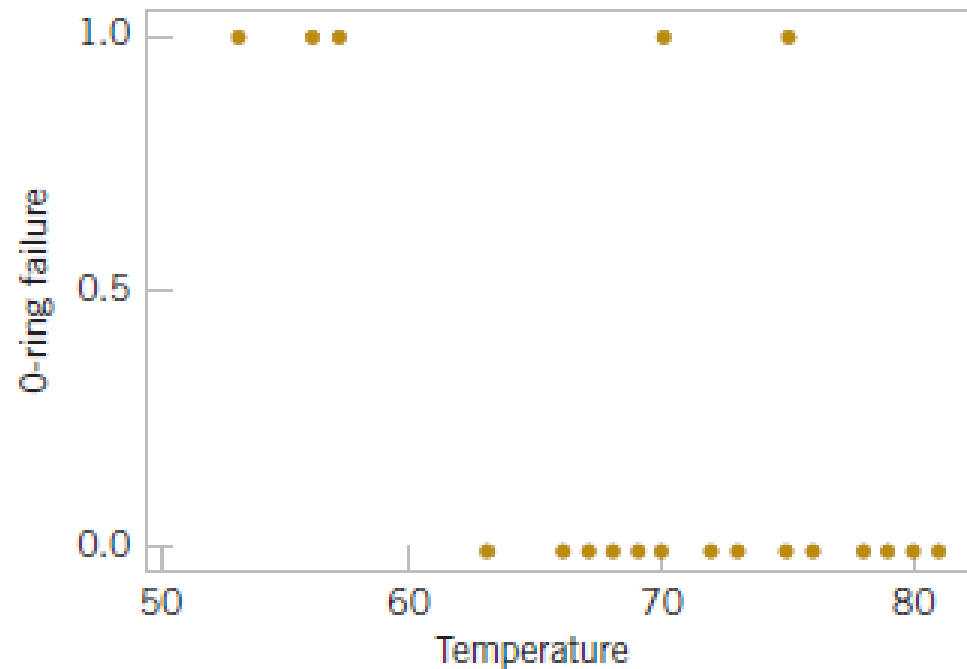
A utilização deste gráfico é muito importante para descobrir se duas variáveis podem estar **associadas**.

Além de auxiliar também a descobrir o comportamento do fenômeno que estamos buscando compreender.



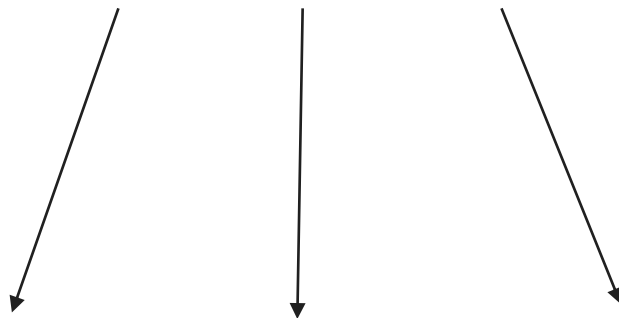


E quando a nuvem de pontos formada resulta em um gráfico assim?





Regressão Logística



Binária

Multinomial

Ordinal





Regressão Logística Binária

Técnica que é considerada um modelo de Classificação, embora receba o nome de Regressão, e sua equação parte da Regressão Linear.

A variável dependente é dicotômica.

A variável resposta (variável dependente), é obrigatoriamente **qualitativa** (categórica).



Para aplicarmos o modelo de Regressão Logística Binária, é necessário cumprir alguns pré-requisitos:

- **Variável Dependente é Categórica e Dicotômica:** Apresenta apenas duas possibilidades de resposta (categorias), podendo assumir os valores 0 e 1 arbitrariamente, além de serem categorias mutuamente exclusivas;
- **Independência das Observações:** Não há medidas repetidas;
- **Ausência de Outliers:** Não podem haver valores discrepantes (pontos influentes ou pontos de alavancagem);
- **Ausência de Multicolinearidade:** As variáveis independentes não podem estar altamente correlacionadas, verificação por meio da Correlação Linear de Pearson ou por meio do VIF (Fatores de Inflação de Variância);
- **Teste de Box-Tidwell:** As variáveis independentes estão linearmente relacionadas ao log das probabilidades.



A equação do modelo de uma Regressão Logística (Função de Resposta Logit) é dada por:

$$E(Y) = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}} + \varepsilon$$

ou

$$E(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} + \varepsilon$$

(ε = Erro, resultado de flutuações aleatórias)

→ **$P(y = 1 | x)$**

→ **$P(y = 0 | x)$**

Onde:

β_0 = Constante do modelo

β_1 = Coeficiente da variável independente

OBS: Podem ter mais β , irá depender do número de variáveis

e = Número de Euler = 2,718281828...



Faz necessário interpretar a **Odds Ratio (OR)**, ou **Razão entre Chances**:

A Razão de Chances pode ser definida como a razão de um evento ocorrer em um grupo A em função de um grupo B, sua fórmula é dada por:

$$e^{(\beta_0 + \beta_1 x)} = \frac{E(Y)}{1 - E(Y)}$$

Onde:

Se $OR > 1$ ➡ Chance aumentou do evento de interesse acontecer;

Se $OR < 1$ ➡ Chance diminuiu do evento de interesse acontecer;

Se $OR = 1$ ➡ Não impacta no evento de interesse.

OBS: Odds Ratio está no intervalo $0 < OR < \infty$



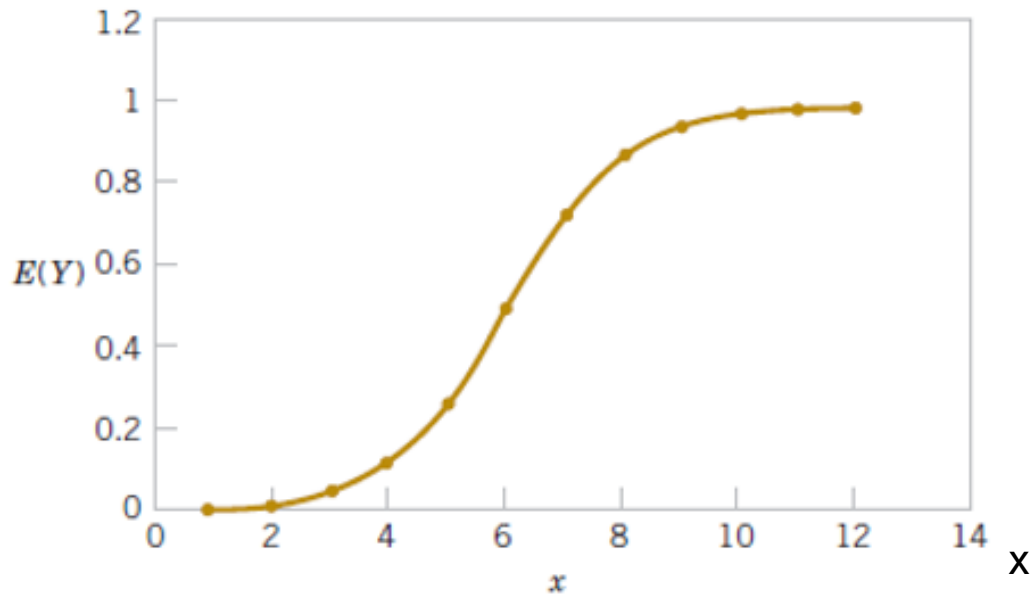
O gráfico da Regressão Logística é um **sigmoide** (forma de S) no **Plano Cartesiano**:

y

Vale lembrar:

Eixo y = Eixo das Ordenadas

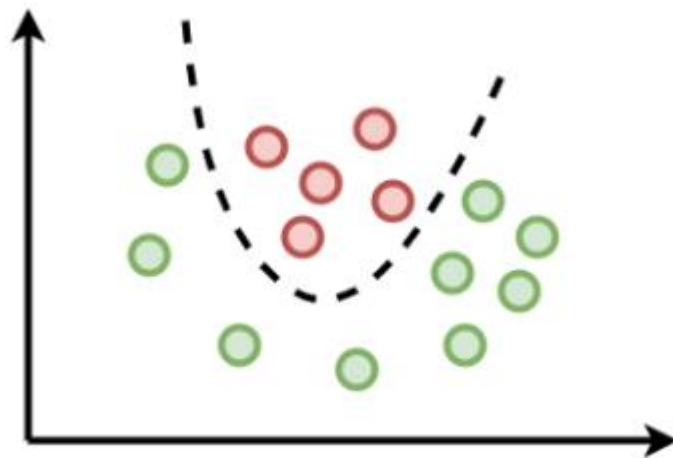
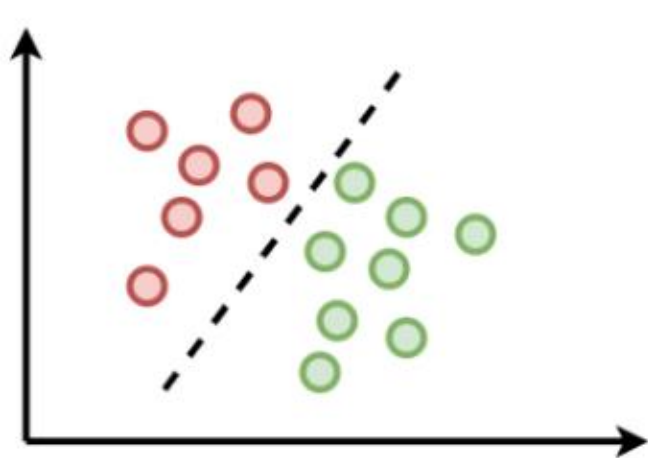
Eixo x = Eixo das Abscissas





Um item que é interessante analisar, é o **Limite de Decisão**.

É uma curva (ou margem), que divide as classes, o próprio modelo realiza um ajuste próprio ao limite de decisão, afim de prever qual classe os próximos dados pertencerão.





Após aplicar o modelo é interessante analisar o **Pseudo R^2 (Coeficiente de Determinação)**.

Nos modelos de Regressão Logística, o R^2 é calculado de forma diferente do que o R^2 nos modelos de Regressão Linear, por isso é chamado de Pseudo R^2 , cujo objetivo é medir o poder preditivo do modelo.

Existem diversas técnicas para a Regressão Logística, sendo a mais utilizada a técnica de McFadden.



Exemplo

Um estudo Estatístico foi realizado para avaliar a chance dos passageiros a bordo do Navio Titanic viverem ou morrerem. O modelo foi ajustado para uma Regressão Logística Binária, tendo como equação:

$$E(Y) = \frac{e^{-1,33 + 2,55.x_1 + 1,27.x_2 + 2,58.x_3 - 0,04.x_4}}{1 + e^{-1,33 + 2,55.x_1 + 1,27.x_2 + 2,58.x_3 - 0,04.x_4}}$$

Onde:

x_1 = Se for do sexo feminino colocar 1, caso contrário 0

x_2 = Se estiver na 2ª Classe colocar 1, caso contrário 0

x_3 = Se estiver na 1ª Classe colocar 1, caso contrário 0

x_4 = Idade da pessoa

Nessas condições, qual a chance que uma mulher com 42 anos da 1ª Classe tinha de sobreviver?

OBS: $E(Y) = 1$ ou próximo de 1 \Rightarrow grande chance de viver

$E(Y) = 0$ ou próximo de 0 \Rightarrow grande chance de morrer



Resolução:

$$E(Y) = e^{\frac{-1,33 + 2,55.x1 + 1,27.x2 + 2,58.x3 - 0,04.x4}{1 + e^{-1,33 + 2,55.x1 + 1,27.x2 + 2,58.x3 - 0,04.x4}}}$$

$$E(Y) = e^{\frac{-1,33 + 2,55.1 + 1,27.0 + 2,58.1 - 0,04.42}{1 + e^{-1,33 + 2,55.1 + 1,27.0 + 2,58.1 - 0,04.42}}}$$

$$E(Y) \approx 0,90$$



Conclusão

Com essas ferramentas da **Análise de Regressão** é possível entender a relação entre duas variáveis e tentar prever o comportamento de uma delas, basta **ter** ou **iniciar** a coleta de dados e na sequência:



APLICAR!





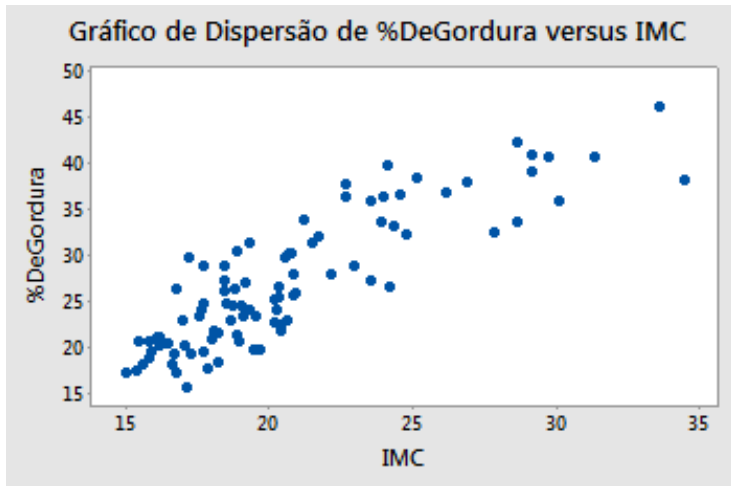
Exercícios

- 1) Em que tipo de situação devemos realizar uma Regressão Logística ao invés de uma Regressão Linear?
- 2) Para que seja possível realizar uma Regressão Linear Múltipla, quais pré-requisitos devemos cumprir?
- 3) Dê dois exemplos de situações onde nossa variável resposta é dicotômica. Escreve a pergunta e as possibilidades de respostas.
- 4) Qual a diferença entre uma variável categórica e uma variável numérica?

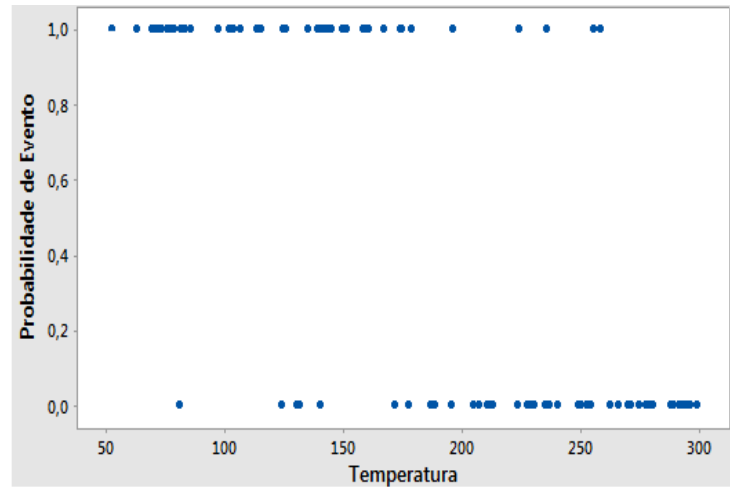


5) Observe os gráficos a seguir:

(I)



(II)



Qual deles pode ser aplicado um modelo de Regressão Logística e qual pode ser aplicado um modelo de Regressão Linear? Justifique sua resposta.

6) Um estudo Estatístico foi realizado para avaliar a chance dos passageiros a bordo do Navio Titanic viverem ou morrerem. O modelo foi ajustado para uma Regressão Logística Binária, tendo como equação:

$$E(Y) = \frac{e^{-1,33 + 2,55.x_1 + 1,27.x_2 + 2,58.x_3 - 0,04.x_4}}{1 + e^{-1,33 + 2,55.x_1 + 1,27.x_2 + 2,58.x_3 - 0,04.x_4}}$$

Onde:

x1 = Se for do sexo feminino colocar 1, caso contrário 0

x2 = Se estiver na 2ª Classe colocar 1, caso contrário 0

x3 = Se estiver na 1ª Classe colocar 1, caso contrário 0

x4 = Idade da pessoa

OBS: E(Y) = 1 ou próximo de 1 => grande chance de viver

E(Y) = 0 ou próximo de 0 => grande chance de morrer

Nessas condições:

a) Qual a chance que uma mulher com 42 anos de idade da 3ª Classe tinha de sobreviver?

b) Qual a chance que um homem com 30 anos de idade da 1ª Classe tinha de sobreviver?

c) Qual a chance que um homem com 30 anos de idade da 2ª Classe tinha de sobreviver?





7) Considerando ainda o exercício anterior, interprete os resultados das OR nas seguintes situações:

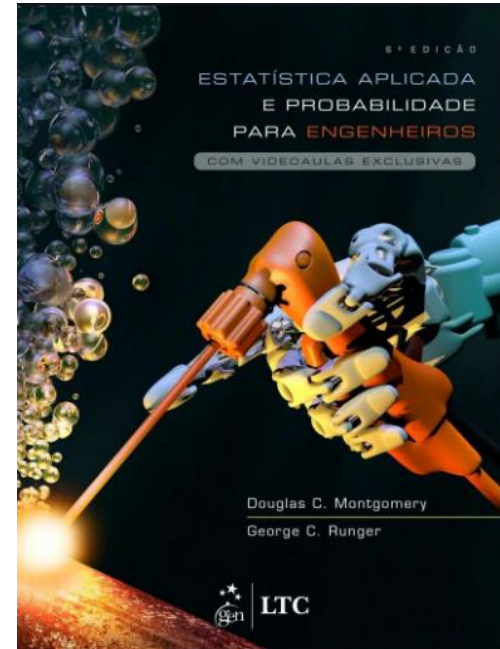
a) $OR = 12,46$ para a variável “Ser do Sexo Feminino”.

b) $OR = 0,96$ para a variável “Idade”.



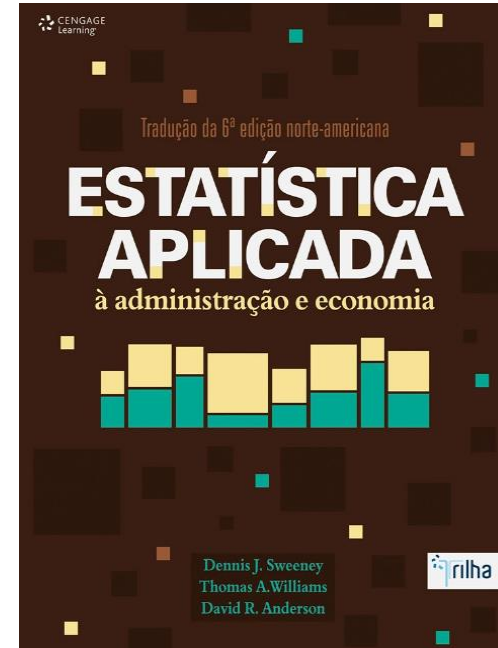
Bibliografia

MONTGOMERY, Douglas C. e RUNGER, George C. ***Estatística Aplicada e Probabilidade para Engenheiros***. 6ª Edição. Rio de Janeiro: Editora GEN|LTC, 2016





SWEENEY, Dennis J; WILLIAMS, Thomas A. e
ANDERSON, David R. ***Estatística Aplicada à
Administração e Economia.***
6ª Edição. São Paulo: Editora Cengage Learning,
2013.





Contatos

Prof. Eng. Rodolfo Magliari de Paiva



Cel.: (11) 9-6866-5501



E-mail: rodolfomagliari@gmail.com



LinkedIn: Rodolfo Magliari de Paiva



Obrigado!