Profª.

Paula Shinozaki

Monitora:

Cristiane Pereira

Regressão linear simples com uso do software R

UNIVASF
UNIVERSIDADE FEDERAL DO VALE DO SÃO FRANCISCO

Proex
PRÓ-REITORIA DE EXTENSÃO

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
CODAI
COLÉGIO AGRÍCOLA DOM AGOSTINHO IKAS DA UFRPE

Regressão Linear

O termo “Regressão” surgiu em 1885 com o antropólogo, matemático e estatístico Francis Galton. As primeiras aplicações do método surgiram na Antropometria, ou seja, estudo das medidas e da matemática dos corpos humanos. Ao estudar as estaturas de pais e filhos, Galton observou que filhos de pais com altura baixa em relação à média tendem a ser mais altos que seus pais, e filhos de pais com estatura alta em relação à média tendem a ser mais baixos que seus pais, ou seja, as alturas dos seres humanos em geral tendem a **regredir** à média.

Regressão Linear

Hoje, conhecemos a análise de regressão como uma técnica que permite estimar o comportamento médio de uma variável resposta em relação a uma ou mais variáveis explicativas. Por exemplo, estimar a altura média dos filhos a partir da altura de seus pais; estimar a produção média de uma lavoura a partir da quantidade de chuva, quantidade de adubo, etc.

Regressão Linear

É importante notar que, apesar de ser uma possibilidade, a análise de regressão não tem como objetivo obter estimativas pontuais de eventos futuros, mas sim de estimar médias condicionais e efeitos.

Regressão Linear

A Análise de Regressão é chamada de **Simples,** quando existe apenas uma variável resposta e uma variável explicativa, e **Múltipla** quando existe uma variável resposta e mais de uma explicativa. Casos em que existe mais de uma variável resposta são analisados pela regressão **Multivariada.**

Regressão Linear

A análise de regressão linear estuda a relação entre a variável dependente ou variável resposta (Y) e uma ou várias variáveis independentes ou regressoras $(X_1, X_2, \dots, X_p)$.

Esta relação representa-se por meio de um modelo matemático, ou seja, por uma equação que associa a variável dependente (Y) com as variáveis independentes $(X_1, X_2, \dots, X_p)$.

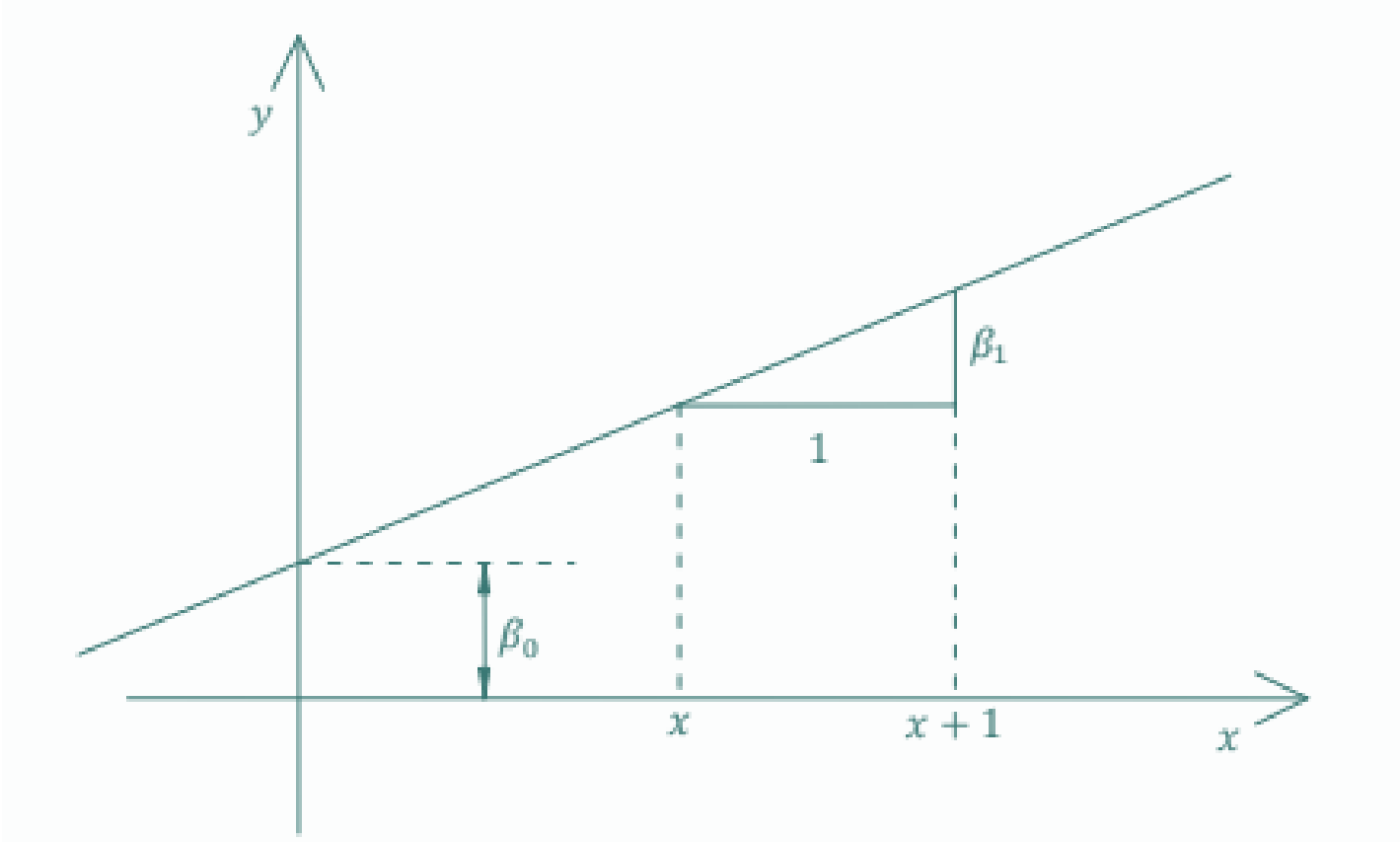
Modelo Teórico

O modelo de Regressão Linear Simples define-se como a relação linear entre a variável dependente (Y) e uma variável independente.

A equação representativa do modelo de regressão linear simples é dado por:

$$y_i = \beta_0 + \beta_1 + \varepsilon_i, \qquad i = 1, \ldots, n$$

Modelo Teórico



Pressupostos do modelo

a) A relação existente entre X e Y é linear;
b) Os erros são independentes com média nula;
c) A variância do erro é constante;
d) Os erros são normalmente distribuídos.

Estimação dos parâmetros

Supondo que existe efetivamente uma relação linear entre X e Y, surge a questão de como estimar os parâmetros β_0 e β_1.

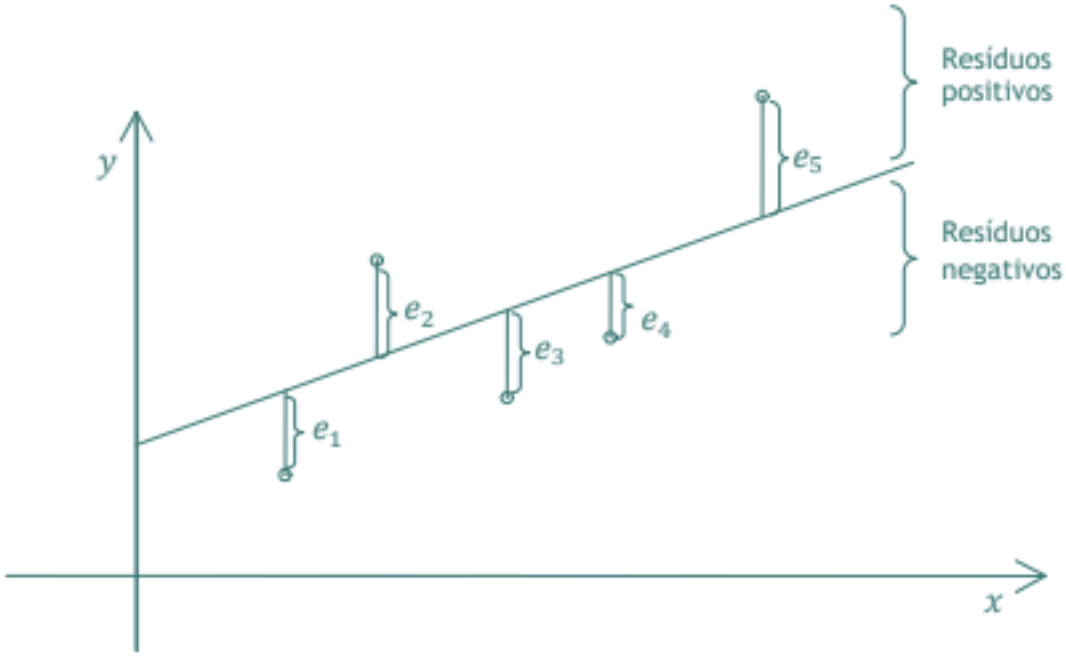
Karl Gauss propôs estimar os parâmetros β_0 e β_1 visando minimizar a soma dos quadrados dos desvios, $e_i, i = 1, \ldots, n$, chamando este processo de método dos mínimos quadrados.

Método dos Mínimos Quadrados

O método dos mínimos quadrados consiste na obtenção dos estimadores dos coeficientes de regressão $\hat{\beta}_0$ e $\hat{\beta}_1$, minimizando os resíduos do modelo de regressão linear, calculados como a diferença entre os valores observados y_i, e os valores estimados, $\hat{y}_i$, isto é:

$$e_i = y_i - \hat{y}_i, \qquad i = 1, \ldots, n$$

Método dos Mínimos Quadrados



Em termos gráficos, os resíduos são representados pelas distâncias verticais entre os valores observados e os valores ajustados

Vamos para o R?

Um psicólogo está investigando a relação entre o tempo que um indivíduo leva para reagir a um estímulo visual e alguns fatores como acuidade visual. Na Tabela 15.1 temos as informações para n=20 indivíduos.

Estatística Básica (Bussab e Morettin, 2017)

O que a teoria nos diz?

Contudo, o que garante que as conclusões anteriores sejam verdadeiras?

A estimativa dos p-valores, do R^2 e dos coeficientes é em sua essência uma função matemática, independente dos valores utilizados, um resultado será obtido. Então, o que garante os resultados tenham valor estatístico?

Vamos para o R?

O cumprimento das considerações do modelo de mínimos quadrados:

1. Resíduos com distribuição normal;
2. Homoscedasticidade dos resíduos (variância constante);
3. Aleatoriedade dos resíduos frente ao valor predito e às variáveis preditoras.