

# Regressão Linear Simples

## Modelos Estatísticos Aplicados

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

- 1 Modelagem
  - Modelos em geral
  - Modelos estatísticos
- 2 Regressão Linear Simples
  - Introdução
  - A regressão
  - Coeficiente de Determinação  $r^2$
  - Exercício
- 3 Resumo
  - Resumo

## 1 Modelagem

- Modelos em geral
- Modelos estatísticos

## 2 Regressão Linear Simples

- Introdução
- A regressão
- Coeficiente de Determinação  $r^2$
- Exercício

## 3 Resumo

- Resumo

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Modelos em geral

Modelos estatísticos

Regressão

Resumo



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Modelos em geral

Modelos estatísticos

Regressão

Resumo

# Modelos animais



Regressão  
Linear  
Simples

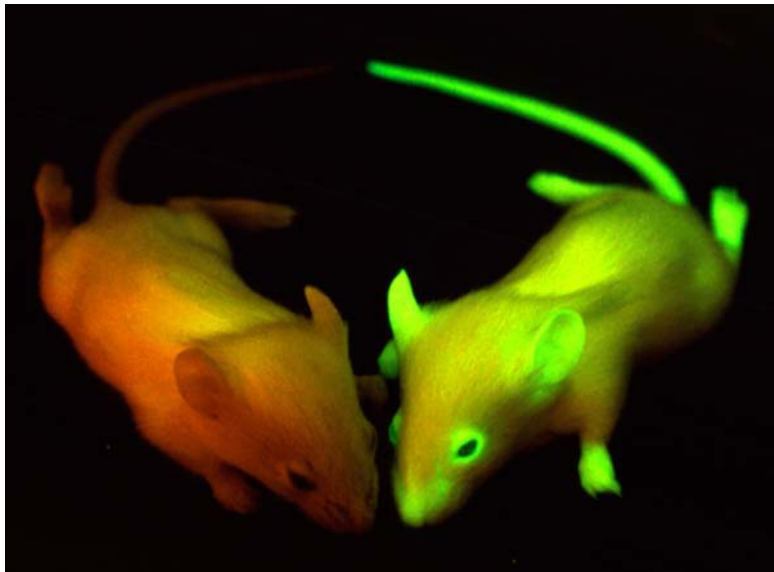
Felipe  
Figueiredo

Modelagem  
Modelos em geral  
Modelos estatísticos

Regressão

Resumo

# Modelos animais



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem  
Modelos em geral  
Modelos estatísticos

Regressão

Resumo

## 1 Modelagem

- Modelos em geral
- Modelos estatísticos

## 2 Regressão Linear Simples

- Introdução
- A regressão
- Coeficiente de Determinação  $r^2$
- Exercício

## 3 Resumo

- Resumo

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Modelos em geral

Modelos estatísticos

Regressão

Resumo

Modelos servem para:

- representar de forma simplificada fenômenos, experimentos, dados, etc;
- possibilitar análise em cenários controlados, menos complexos que a realidade;
- extrapolar resultados e conclusões.



Ao ajustar um modelo aos dados, podemos:

- fazer predições dentro do intervalo observado para dados que não foram obtidos (interpolação)
- fazer predições fora do intervalo observado (extrapolação)

# Para todos os gostos...

**TABLE 1**

**Regression models**

	Application	Dependent variables	Independent variables
Linear regression	Description of a linear relationship	Continuous (weight, blood pressure)	Continuous and/or categorical
Logistic regression	Prediction of the probability of belonging to groups (outcome: yes/no)	Dichotomous (success of treatment: yes/no)	
Proportional hazard regression (Cox regression)	Modeling of survival data	Survival time (time from diagnosis to event)	
Poisson regression	Modeling of counting processes	Counting data: whole numbers representing events in temporal sequence (e.g., the number of times a woman gave birth over a certain period of time)	

1

## Modelagem

- Modelos em geral
- Modelos estatísticos

2

## Regressão Linear Simples

- Introdução
- A regressão
- Coeficiente de Determinação  $r^2$
- Exercício

3

## Resumo

- Resumo

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

# Modelo de regressão linear simples



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

Quando os dados indicam uma relação linear, um modelo de regressão pode ser utilizado para quantificar esta relação com uma **reta de regressão**.

## Exemplo: Algumas aplicações

- Tendência (“Níveis de insulina em jejum tendem a aumentar com a idade?”)
- Ajuste de curva (“Qual é o  $EC_{50}$  de uma nova droga?”)
- Predição (“Como prever o risco de infarto do miocárdio, sabendo-se a idade, pressão e nível de colesterol?”)

JOURNAL OF WOMEN'S HEALTH  
Volume 15, Number 9, 2006  
© Mary Ann Liebert, Inc.

## The Association between Body Mass Index and Osteoporosis in Patients Referred for a Bone Mineral Density Examination

KOFI ASOMANING, M.B.Ch.B., M.S.,<sup>1</sup> ELIZABETH R. BERTONE-JOHNSON, Sc.D.,<sup>2</sup>  
PHILIP C. NASCA, Ph.D.,<sup>2</sup> FREDERICK HOOVEN, Ph.D.,<sup>3</sup>  
and PENELOPE S. PEKOW, Ph.D.<sup>2</sup>

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

## ABSTRACT

*Purpose:* Osteoporosis affects 4–6 million (13%–18%) postmenopausal white women in the United States. Most studies to date on risk factors for osteoporosis have considered body mass index (BMI) only as a possible confounder. In this study, we assess the direct relationship between BMI and osteoporosis.

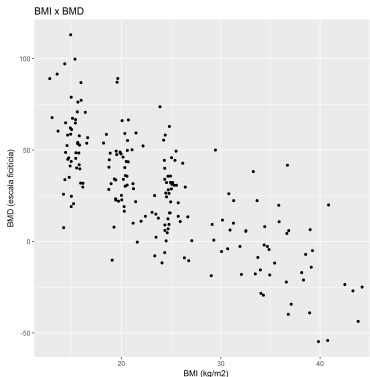
*Methods:* We conducted a cross-sectional study among women aged 50–84 years referred by their physicians for a bone mineral density (BMD) examination at Baystate Medical Center between October 1998 and September 2000. BMI was determined prior to the BMD examination in the clinic. Information on other risk factors was obtained through a mailed questionnaire. Ordinal logistic regression was used to model the association between BMI and osteoporosis, controlling for confounding factors.

*Results:* BMI was inversely associated with BMD status. After adjustment for age, prior hormone replacement therapy (HRT) use, and other factors, odds ratios (OR) for low, high, and obese compared with moderate BMI women were 1.8 (95% CI 1.2–2.7), 0.46 (95% CI 0.29–0.71), and 0.22 (95% CI 0.14–0.36), respectively, with a significant linear trend ( $p < 0.0001$ ) across BMI categories. Evaluating BMI as a continuous variable, the odds of bone loss decreased 12% for each unit increase in BMI (OR = 0.88, 95% CI 0.85–0.91).

*Conclusions:* Women with low BMI are at increased risk of osteoporosis. The change in risk associated with a 1 unit change in BMI (~5–8 lb) is of greater magnitude than most other modifiable risk factors. To help reduce the risk of osteoporosis, patients should be advised to maintain a normal weight.

# Na prática...

- Dados simulados, inspirados no paper.
- Existe uma tendência? Ela é linear?
- Podemos prever a osteoporose a partir do IMC?

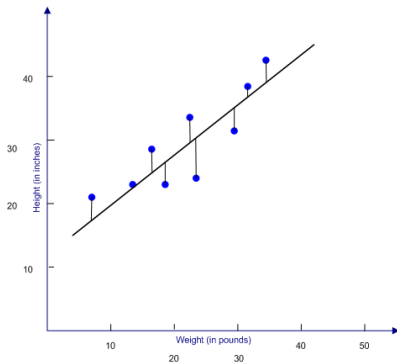


## Definition

Uma **reta de regressão** (também chamada de reta de melhor ajuste) é a reta para a qual a soma dos erros quadráticos dos resíduos é o mínimo.

- É a reta que melhor se ajusta aos dados
- Minimiza os resíduos





## Definition

Resíduos são a distância entre o dado observado e a reta estimada (modelo).

- Para muitos testes presume-se que os dados vem de uma distribuição normal
- Neste caso, não é necessário que os **dados** sejam normais
- **É necessário que os resíduos sejam normais**

- Relembrando: a equação de uma reta é definida pela fórmula

$$\hat{y} = ax + b$$

- No caso da reta regressora:
  - $y$  é a variável dependente
  - $x$  é a variável independente
  - $a$  é a inclinação
  - $b$  é o intercepto
- Assim, o objetivo da análise de regressão é encontrar os valores  $a$  e  $b$

Para determinar a inclinação e o intercepto, usamos:

- as médias de  $X$  e  $Y$
- as variâncias de  $X$  e  $Y$
- o coeficiente de correlação  $r$  entre  $X$  e  $Y$
- o tamanho da amostra  $n$
- ... e algumas operações entre estes termos

- 1 Modelagem
  - Modelos em geral
  - Modelos estatísticos
- 2 Regressão Linear Simples
  - Introdução
  - A regressão
  - Coeficiente de Determinação  $r^2$
  - Exercício
- 3 Resumo
  - Resumo

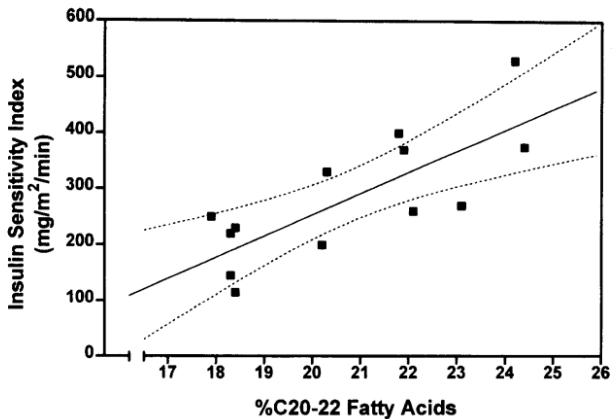
## Example

Voltemos ao exemplo de associar a composição lipídica com a sensibilidade a insulina.

## Pergunta

Qual é o acréscimo na sensibilidade à insulina, para cada unidade aumentada na composição lipídica?

# Exemplo



Fonte: Motulsky, 1995

# Exemplo

## Linear Regression Number of points = 13

Parameter	Expected Value	Standard Error	Lower 95% CI	Upper 95% CI
Slope	37.208	9.296	16.747	57.668
Y intercept	-486.54	193.72	-912.91	-60.173
X intercept	13.076			

$r^2 = 0.5929$

Standard deviation of residuals from line ( $Sy.x$ ) = 75.895

Test: Is the slope significantly different from zero?

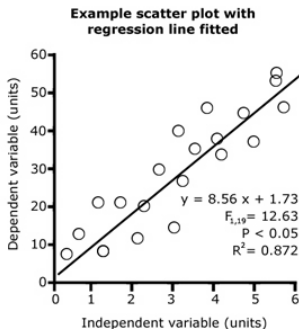
$F = 16.021$

The P value is 0.0021, considered very significant.



- O p-valor é significativo.
- A inclinação é  $\approx 37.2$
- Isto significa que:

para cada unidade aumentada no %C20–22, teremos um aumento proporcional de aproximadamente 37.2 mg/m<sup>2</sup>/min na sensibilidade à insulina



- A qualidade do ajuste do modelo de regressão é determinado pelo **coeficiente de determinação**  $r^2$

- 1 Modelagem
  - Modelos em geral
  - Modelos estatísticos
- 2 **Regressão Linear Simples**
  - Introdução
  - A regressão
  - **Coeficiente de Determinação  $r^2$**
  - Exercício
- 3 Resumo
  - Resumo

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

# Coefficiente de Determinação $r^2$



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

## Definition

O **coeficiente de determinação**  $r^2$  é a relação da variação explicada com a variação total.

$$r^2 = \frac{\text{variação explicada}}{\text{variação total}}$$

- Lembrando:  $r^2$  é o quadrado de  $r$ !

# Coefficiente de Determinação $r^2$



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

- Qual é a porcentagem da variação dos dados pode ser explicada pela reta regressora?
- O coeficiente  $r^2$  é a fração da variância que é compartilhada entre  $X$  e  $Y$ .
- Como  $r$  está sempre entre -1 e 1,  $r^2$  está sempre entre 0 e 1.

# Coefficiente de Determinação $r^2$



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

- Além disso,  $r^2 \leq |r|$
- Por que?

Compare os seguintes números entre 0 e 1:

$$\frac{1}{2} \text{ e } \left(\frac{1}{2}\right)^2 = \frac{1}{4} \Rightarrow \frac{1}{4} \leq \frac{1}{2}$$

$$\frac{1}{3} \text{ e } \left(\frac{1}{3}\right)^2 = \frac{1}{9} \Rightarrow \frac{1}{9} \leq \frac{1}{3}$$

1

## Modelagem

- Modelos em geral
- Modelos estatísticos

2

## Regressão Linear Simples

- Introdução
- A regressão
- Coeficiente de Determinação  $r^2$
- Exercício

3

## Resumo

- Resumo

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

JOURNAL OF WOMEN'S HEALTH  
Volume 15, Number 9, 2006  
© Mary Ann Liebert, Inc.

## The Association between Body Mass Index and Osteoporosis in Patients Referred for a Bone Mineral Density Examination

KOFI ASOMANING, M.B.Ch.B., M.S.,<sup>1</sup> ELIZABETH R. BERTONE-JOHNSON, Sc.D.,<sup>2</sup>  
PHILIP C. NASCA, Ph.D.,<sup>2</sup> FREDERICK HOOVEN, Ph.D.,<sup>3</sup>  
and PENELOPE S. PEKOW, Ph.D.<sup>2</sup>

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo



## ABSTRACT

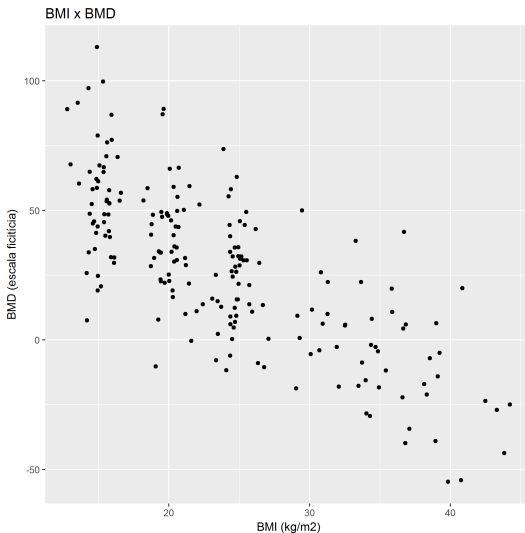
*Purpose:* Osteoporosis affects 4–6 million (13%–18%) postmenopausal white women in the United States. Most studies to date on risk factors for osteoporosis have considered body mass index (BMI) only as a possible confounder. In this study, we assess the direct relationship between BMI and osteoporosis.

*Methods:* We conducted a cross-sectional study among women aged 50–84 years referred by their physicians for a bone mineral density (BMD) examination at Baystate Medical Center between October 1998 and September 2000. BMI was determined prior to the BMD examination in the clinic. Information on other risk factors was obtained through a mailed questionnaire. Ordinal logistic regression was used to model the association between BMI and osteoporosis, controlling for confounding factors.

*Results:* BMI was inversely associated with BMD status. After adjustment for age, prior hormone replacement therapy (HRT) use, and other factors, odds ratios (OR) for low, high, and obese compared with moderate BMI women were 1.8 (95% CI 1.2–2.7), 0.46 (95% CI 0.29–0.71), and 0.22 (95% CI 0.14–0.36), respectively, with a significant linear trend ( $p < 0.0001$ ) across BMI categories. Evaluating BMI as a continuous variable, the odds of bone loss decreased 12% for each unit increase in BMI (OR = 0.88, 95% CI 0.85–0.91).

*Conclusions:* Women with low BMI are at increased risk of osteoporosis. The change in risk associated with a 1 unit change in BMI (~5–8 lb) is of greater magnitude than most other modifiable risk factors. To help reduce the risk of osteoporosis, patients should be advised to maintain a normal weight.

# Na prática...



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

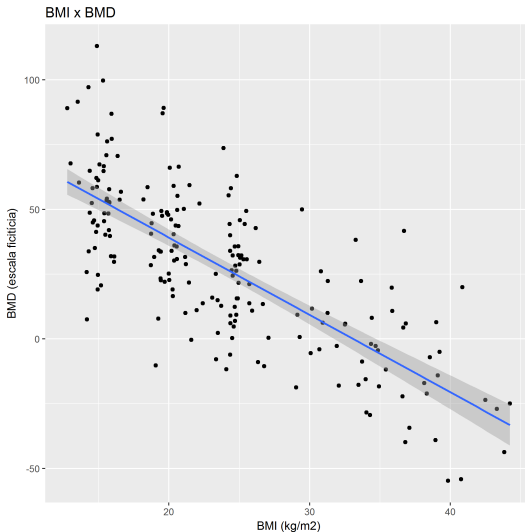
A regressão

$R^2$

Exercício

Resumo

# Na prática...



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

- Os resíduos são aprox. normais?
- Quantos % de variabilidade podem ser explicados pelo modelo?
- Quanto o BMD muda, para cada unidade de BMI?

## Saída típica de um programa de análise

Residuals:

Min	1Q	Median	3Q	Max
-52.097	-13.864	0.762	10.707	58.730

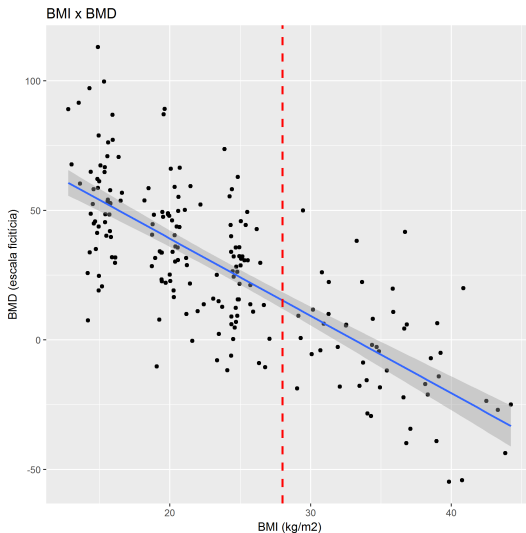
Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	98.8176	4.6281	21.35	<2e-16 ***
BMI	-2.9845	0.1846	-16.17	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.26 on 198 degrees of freedom  
Multiple R-squared: 0.5691, Adjusted R-squared: 0.5669  
F-statistic: 261.5 on 1 and 198 DF, p-value: < 2.2e-16

# E o BMI = 28?



Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

# E o BMI = 28?



Regressão  
Linear  
Simple

Felipe  
Figueiredo

Modelagem

Regressão

Introdução

A regressão

$R^2$

Exercício

Resumo

- o valor predito pelo modelo é 15.25169
- P: O que isto significa?

- 1 Modelagem
  - Modelos em geral
  - Modelos estatísticos
- 2 Regressão Linear Simples
  - Introdução
  - A regressão
  - Coeficiente de Determinação  $r^2$
  - Exercício
- 3 Resumo
  - Resumo

Regressão  
Linear  
Simples

Felipe  
Figueiredo

Modelagem

Regressão

Resumo

Resumo

- É necessário investigar a relação entre as variáveis!
- O que pode explicar a relação observada?
- Qual proporção (porcentagem) da variabilidade pode ser explicada pelas variáveis analisadas?
- Quão bem a reta regressora se ajusta aos dados?



Schneider A, Hommel G, Blettner MDtsch Arztebl Int 2010;  
107(44): 776–82.

[https://www.ncbi.nlm.nih.gov/pmc/articles/  
PMC2992018/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2992018/)