



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Aprofundamento

Regressão Linear Simples

Modelos com desfecho contínuo

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

Sumário

- 1 **Modelagem**
 - Modelos em geral
 - Trailer
- 2 **Regressão Linear Simples**
 - Introdução
 - A regressão
 - Coeficiente de Determinação r^2
 - Exercício
 - Bônus: preditor categórico
 - Resumo
- 3 **Aprofundamento**
 - Aprofundamento



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Aprofundamento

Discussão da aula passada



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Aprofundamento

Discussão da leitura obrigatória da aula passada

Modelos



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Modelos em geral
Trailer

Regressão

Aprofundamento

Definição

Versão simplificada da realidade, adequada ao fim pretendido.

Modelos servem para:

- representar fenômenos, experimentos, dados, etc. de **forma tratável**;
- avaliar cenários controlados, menos complexos que a realidade;
- extrapolar resultados e conclusões.



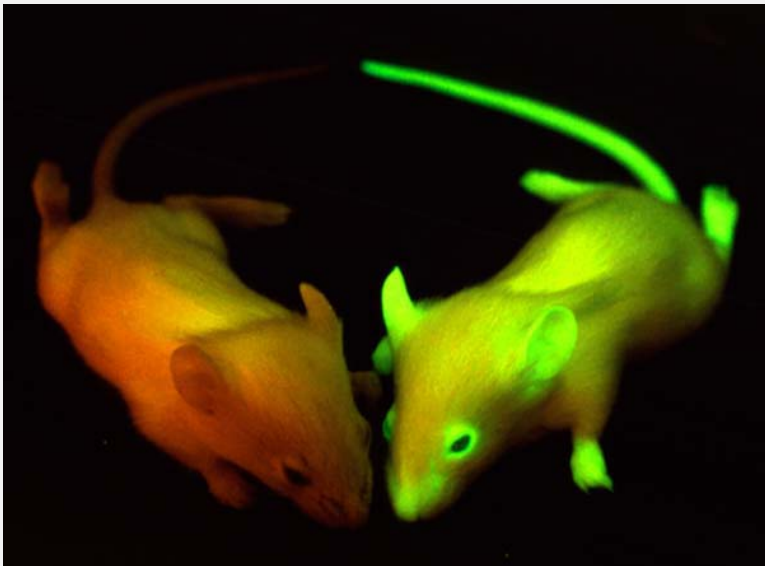
Regressão
Linear
Simples
Felipe
Figueiredo

Modelagem
Modelos em geral
Trailer
Regressão
Aprofundamento



Regressão
Linear
Simples
Felipe
Figueiredo

Modelagem
Modelos em geral
Trailer
Regressão
Aprofundamento



Regressão
Linear
Simples
Felipe
Figueiredo

Modelagem
Modelos em geral
Trailer
Regressão
Aprofundamento

- Distribuições de probabilidade servem como modelo para a distribuição dos dados (teórico x empírico)
- Modelos de regressão servem como um *framework* para testar hipóteses **específicas** sobre a relação presumida entre variáveis

Modelo de regressão

Formulação explícita de uma hipótese sobre a associação entre o desfecho (contínuo, neste contexto) e o preditor

Regressão
Linear
Simples
Felipe
Figueiredo

Modelagem
Modelos em geral
Trailer
Regressão
Aprofundamento

Modelo explicativo/explanatório

Verificação ou teste de hipóteses sobre a relação entre as variáveis avaliadas.

Modelo preditivo

Estimativa do resultado esperado, mesmo para dados que não foram testados...

... restrito ao intervalo testado.

TABLE 1

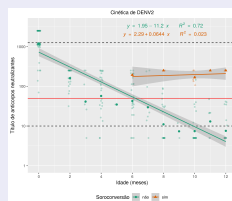
Regression models

| | 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) | |

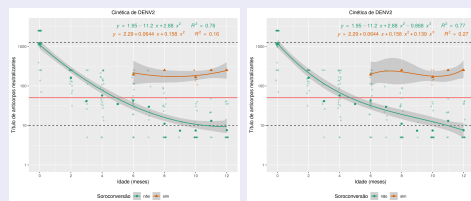
Deutsches Ärzteblatt International | Dtsch Arztebl Int 2010; 107(44): 776–82

Decaimento de anticorpos de neonatos recebidos da mãe

Modelo linear – polinômio de grau 1



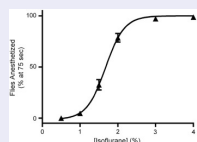
Modelos polinomiais de grau 2 e 3



E você pensando...



Modelo de regressão logística 4 parâmetros



$$\hat{Y} = a + \frac{b - a}{1 + \left(\frac{c}{x}\right)^d}$$

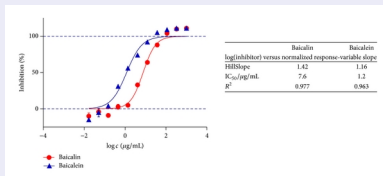
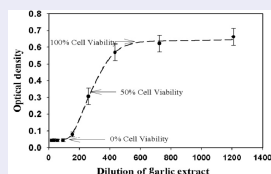
Gadagkar, Call, 2015; J. Pharmacol. Toxicol. Methods

Regressão
Linear
SimplesFelipe
FigueiredoModelagem
Modelos em geral
Trailer

Regressão

Aprofundamento

Aplicações (EC50, IC50, ED50, TD50, LD50, ...)



[1] Gupta, Lee, 2013; [2] Jelic, et al., 2016

Vamos começar pelo modelo mais simples

(Hoje, apenas desfecho contínuo!)

Regressão
Linear
SimplesFelipe
FigueiredoModelagem
Modelos em geral
Trailer

Regressão

Aprofundamento

Modelo de regressão linear simples

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

1 Tendência

"Níveis de insulina em jejum tendem a aumentar com a idade?"

2 Ajuste de curva

"Qual é o EC₅₀ de uma nova droga?"

3 Predição

"Como prever o risco de infarto do miocárdio, sabendo-se a idade, pressão e nível de colesterol?"

Regressão
Linear
SimplesFelipe
Figueiredo

Modelagem

Regressão

Introdução
A regressão
r²Exercício
Bônus: preditor
categórico
Resumo

Aprofundamento

Depois dos comerciais...

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.,¹ ELIZABETH R. BERTONE-JOHNSON, Sc.D.,²
PHILIP C. NASCA, Ph.D.,² FREDERICK HOOVEN, Ph.D.,³
and PENELOPE S. PEKOW, Ph.D.²

Regressão
Linear
SimplesFelipe
Figueiredo

Modelagem

Regressão

Introdução
A regressão
r²Exercício
Bônus: preditor
categórico
Resumo

Aprofundamento

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.

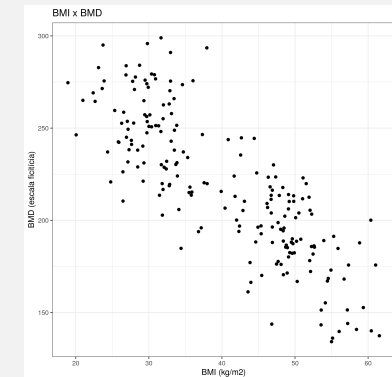
Methods: We conducted a cross-sectional study among women aged 50–84 years referred by their physicians for a bone mineral density (BMD) examination at Baystate Medical Center between October 1998 and September 2000. BMI was determined prior to the BMD examination in the clinic. Information on other 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.

ento

- Dados simulados, inspirados no paper.
- Existe uma tendência? Ela é linear?
- Podemos estimar BMD sabendo o IMC?



- Dependente: BMD (contínua)
 - Sinônimos: desfecho, resposta
- Independente: BMI (contínua)
 - Sinônimos¹: preditor, fator

Esta relação pode ser expressa como

$$\text{BMD} \sim \text{BMI}$$

¹ Em alguns contextos também covariável/cofator (quando há mais de uma V.I.)

- A equação de uma reta é definida pela fórmula

$$\text{BMD} = a \times \text{BMI} + b$$

- Duas “variáveis” e dois parâmetros
 - BMD é a variável dependente (dados)
 - BMI é a variável independente (dados)
 - b é o intercepto (*intercept*)
 - a é a inclinação (*slope*)

Inversão, em relação à matemática básica

- Note que aqui os “dados” já foram coletados (fixos)
- Nosso objetivo é estimar os parâmetros da reta b e a

- O intercepto é o valor (hipotético) de BMD quando BMI = 0
- A inclinação é quanto BMD altera² quando aumentamos o BMI em 1 unidade

Atenção

Para estas interpretações serem válidas, a relação deve ser linear (proporcional).

²na média!

Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

Definição

Uma **reta de regressão** é a reta para a qual a soma dos erros quadráticos dos resíduos (ε) é o mínimo.

- Também chamada de reta de melhor ajuste
- Minimiza os resíduos (erros aleatórios ε)³
- Erros aleatórios ε em torno de zero
- Dados observados: Y (desfecho contínuo) e X (preditor)
- Parâmetros estimados (β_0 e β_1)

Formulação

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

³ Método dos mínimos quadrados

Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

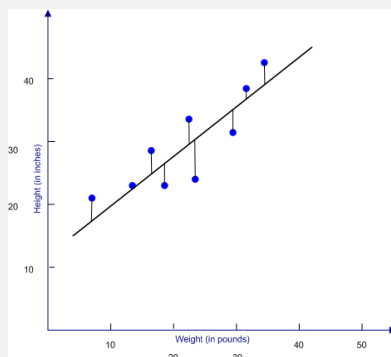
Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento



Definição

Resíduos são a distância entre o dado observado e a reta.

Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

- 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**

Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

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

Exemplo 17.1

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

Pergunta

Podemos explicar o “comportamento” e a variabilidade da insulina sabendo a composição lipídica?

Quais são as variáveis?

- Dependente: insulina (contínua)
- Independente: conteúdo lipídico (contínua)

Esta relação pode ser expressa como

$\text{insulina} \sim \text{conteúdo lipídico}$

Componentes da regressão linear simples

Versão simplificada (apenas variáveis)

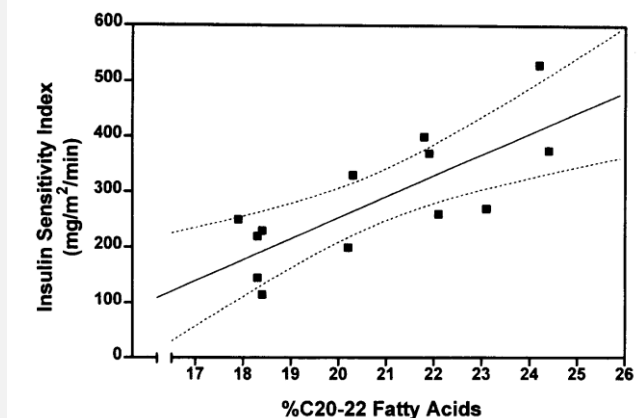
$\text{insulina} \sim \text{conteúdo lipídico}$

Modelo completo

$\text{insulina} = \beta_0 + \beta_1(\text{conteúdo lipídico}) + \varepsilon$

Hipótese: ε é um erro aleatório ⁴ normalmente distribuído e centrado em zero – a incerteza que não pode ser controlada.

⁴ residual – não é explicado pela relação entre as variáveis do modelo



Fonte: Motulsky, 1995

Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

\hat{r}^2

Exercício

Bônus: preditor

categórico

Resumo

Aprofundamento

Regressão
Linear
Simples

Felipe
Figueiredo

| 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 ≈ 37.2
- Isto significa que:

Interpretação da inclinação

para cada unidade aumentada no %C20-22...

... teremos um aumento proporcional de aproximadamente 37.2 $\text{mg/m}^2/\text{min}$ na sensibilidade à insulina

Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

\hat{r}^2

Exercício

Bônus: preditor

categórico

Resumo

Aprofundamento

Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

\hat{r}^2

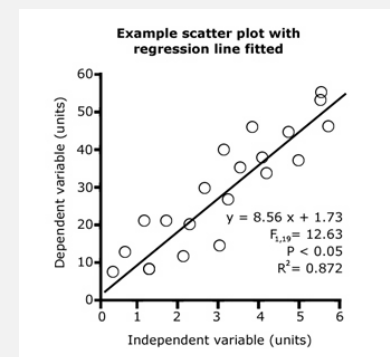
Exercício

Bônus: preditor

categórico

Resumo

Aprofundamento



Uma forma **simplicista** de aferir a qualidade do ajuste do modelo⁵ é o Coeficiente de Determinação r^2 .

(r^2 corresponde ao quadrado de r !)

⁵ Também chamada de Goodness of Fit (GoF)

Definição

O **coeficiente de determinação** r^2 é a razão entre variância explicada e a variância total observada.

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

- Lembrando: r^2 é o quadrado de r !

- Qual é a porcentagem da variância dos dados pode ser explicada pela reta regressora?
- O coeficiente r^2 é a fração da variância que é compartilhada entre X e Y .

- Obs: Como r está sempre entre -1 e 1
 - $|r|$ está sempre entre 0 e 1
 - r^2 está sempre entre 0 e 1

Além disso, $r^2 \leq |r|$

Por que?

Compare os seguintes números entre 0 e 1:

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

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

Exemplo 17.1

Na aula de correlação linear produto-momento de Pearson, vimos que para o exemplo 17.1, $r = 0.77$.

$$r^2 = 0.77^2 = 0.59$$

Interpretação do Coeficiente de Determinação r^2

Podemos explicar 59% da variância da insulina considerando apenas o conteúdo lipídico.

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.,¹ ELIZABETH R. BERTONE-JOHNSON, Sc.D.,²
PHILIP C. NASCA, Ph.D.,² FREDERICK HOOVEN, Ph.D.,³
and PENELOPE S. PEKOW, Ph.D.²

Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

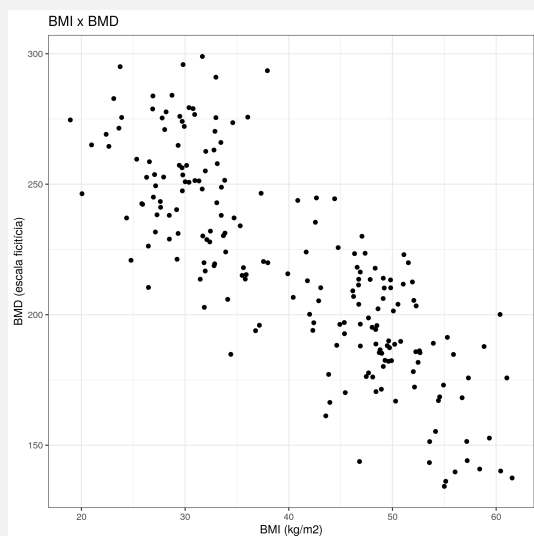
ABSTRACT

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.



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

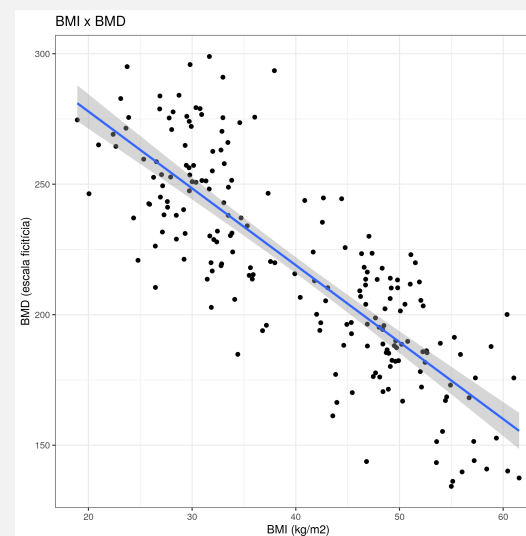
Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

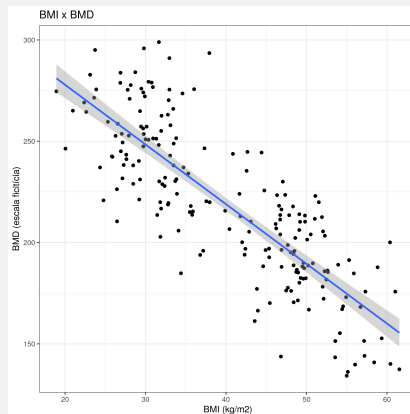
Bônus: preditor

categorico

Resumo

Aprofundamento

- Se o modelo é adequado, podemos substituir isto...
- ... por isto
- Como saber se o modelo representa bem os dados?



Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

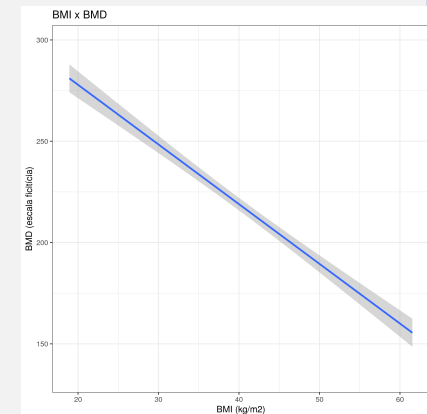
Bônus: preditor

categorico

Resumo

Aprofundamento

- Se o modelo é adequado, podemos substituir isto...
- ... por isto
- Como saber se o modelo representa bem os dados?



Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

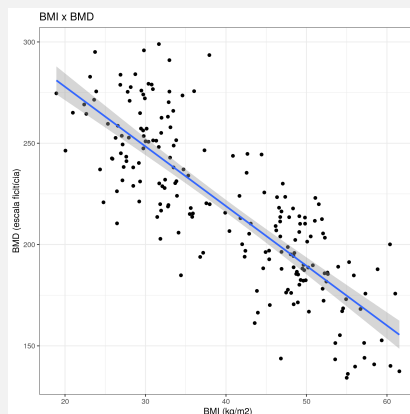
Bônus: preditor

categorico

Resumo

Aprofundamento

- A dispersão em torno da reta é aprox. aleatória?
- Observe o formato faixa de confiança em torno da reta
- A dispersão do desfecho pode ser explicada pela variável independente?
- Toda? Parte? Quanta?



Modelagem

Regressão

Introdução

A regressão

r^2

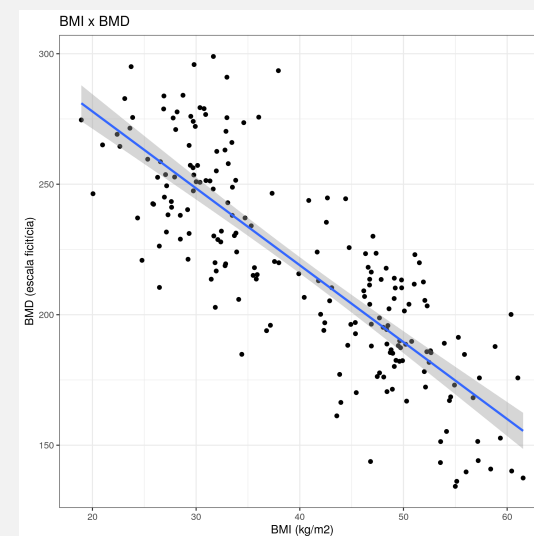
Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento



Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

Perguntas

- Os resíduos são aprox. normais?
- Quantos % de variância podem ser explicados pelo modelo?
- Qual é o BMD predito para um hipotético BMI = 0?
- Quanto o BMD muda, para cada unidade de BMI?

Saída típica de um programa de análise

```
Residuals:
    Min       1Q   Median       3Q      Max
-55.11 -14.86  -1.63   17.49   68.47

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  336.7947     6.0993   55.22  <2e-16 ***
BMI          -2.9461     0.1467  -20.08  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.17 on 198 degrees of freedom
Multiple R-squared:  0.6707, Adjusted R-squared:  0.669
F-statistic: 403.2 on 1 and 198 DF,  p-value: < 2.2e-16
```

Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

- Como vimos, os resíduos são erros aleatórios
(em torno da reta)
- Erros que não podem ser explicados pelo modelo
- Devem ser normalmente distribuídos em torno de zero
(reta como referência)

Saída típica de um programa de análise

```
Residuals:
    Min       1Q   Median       3Q      Max
-55.11 -14.86  -1.63   17.49   68.47
```

Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

Podemos também verificar esta premissa
visualmente

Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

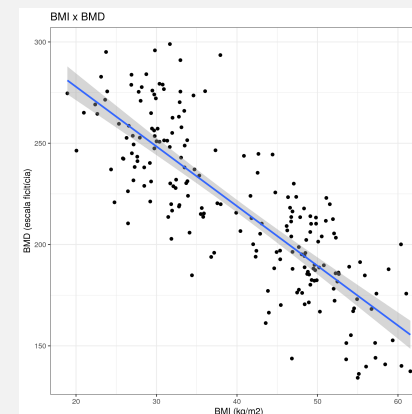
categorico

Resumo

Aprofundamento

A distribuição dos resíduos é
aprox. Normal?

- A dispersão em torno da reta
é aprox. aleatória?
- A dispersão dos resíduos
aumenta ou diminui ao longo
da faixa considerada?



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

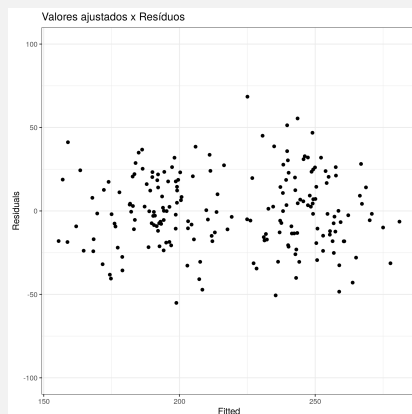
categorico

Resumo

Aprofundamento

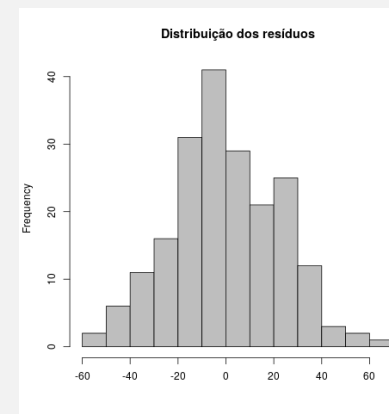
A distribuição dos resíduos é aprox. Normal?

- A dispersão em torno de 0 é aprox. aleatória?
- A dispersão dos resíduos aumenta ou diminui ao longo da faixa considerada?



A distribuição dos resíduos é aprox. Normal?

- A dispersão em torno de 0 é aprox. aleatória?
- A dispersão dos resíduos aumenta ou diminui ao longo da faixa considerada?



Perguntas

- Os resíduos são aprox. normais?

Resposta

- Sim (probably...)

Saída típica de um programa de análise

Residuals:

| Min | 1Q | Median | 3Q | Max |
|--------|--------|--------|-------|-------|
| -55.11 | -14.86 | -1.63 | 17.49 | 68.47 |

Perguntas

- Quantos % de variância podem ser explicados pelo modelo?

Resposta

- Podemos explicar $r^2 = 67\%$ da variância observada no BMD (considerando **apenas** o BMI)
- 33% são devidos a outros fatores

Saída típica de um programa de análise

Multiple R-squared: 0.6707, Adjusted R-squared: 0.669

E os parâmetros da reta estimados a partir dos dados?

Perguntas

- Qual é o BMD predito para um hipotético BMI = 0?

Resposta

- BMD = 337 unidades (IC = [324.8, 348.8])

Saída típica de um programa de análise

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 336.7947      6.0993   55.22  <2e-16 ***
```

⁶ H_0 : Intercepto não é igual a 0 (modelo nulo)

Perguntas

- Quanto o BMD muda, para cada unidade de BMI?

Resposta

- Decréscimo de 3 unidades de BMD (IC = [-3.2, -2.7])

(para cada incremento unitário de BMI)

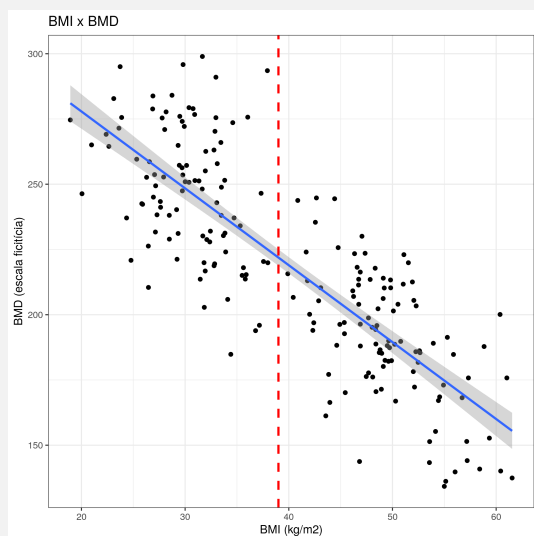
Saída típica de um programa de análise

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
BMI          -2.9461      0.1467  -20.08  <2e-16 ***
```

⁷ H_0 : Inclinação não é igual a 0 (modelo nulo)

Vamos agora fazer predições sobre valores
não observados

E o BMI = 39?



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

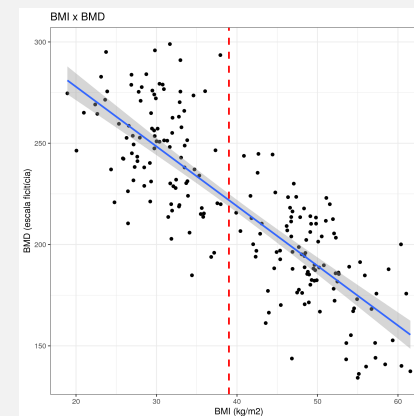
Bônus: preditor

categorico

Resumo

Aprofundamento

E o BMI = 39?



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

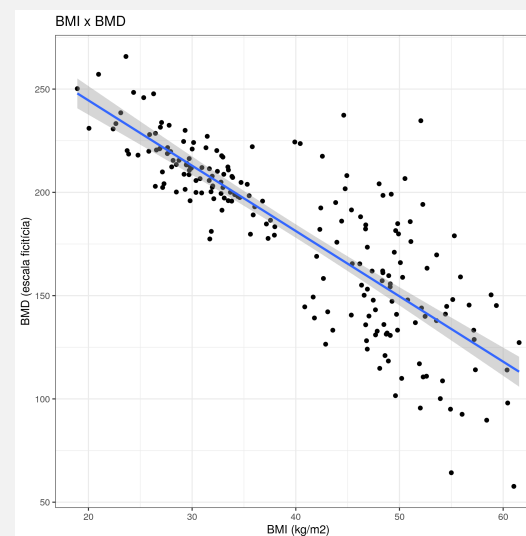
categorico

Resumo

Aprofundamento

- o valor predito pelo modelo é 221.9 (IC: [218.8, 225.0])
- P: O que isto significa?

Dispersão em torno da reta **crece** ao longo da faixa ($r^2 = 68\%$)



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

E quando os resíduos não são aleatórios em
torno da reta?

(casos extremos)



Regressão
Linear
Simples

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

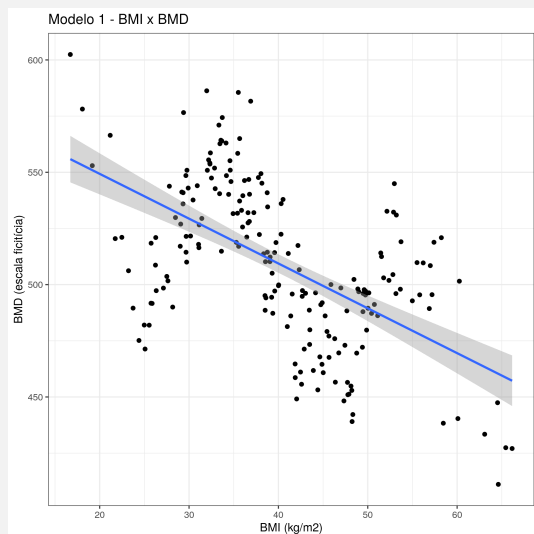
Bônus: preditor

categorico

Resumo

Aprofundamento

Dispersão em torno da reta **varia** ao longo da faixa ($r^2 = 32\%$)



Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

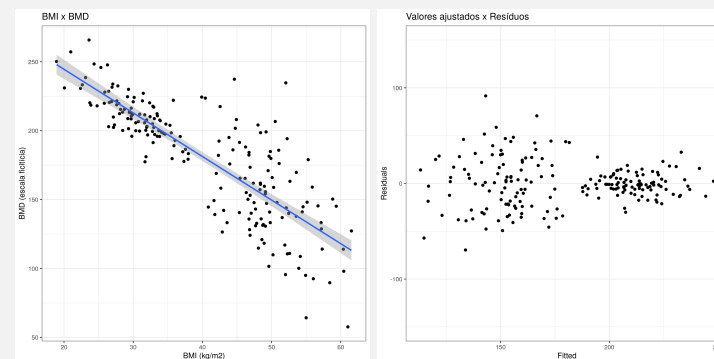
Bônus: preditor

categorico

Resumo

Aprofundamento

Análise de resíduos dos 2 últimos exemplos



Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

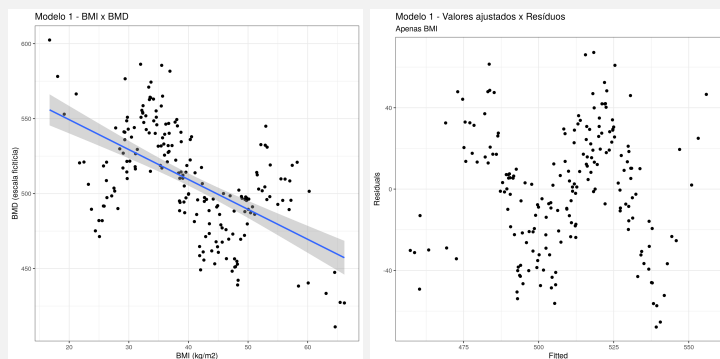
categorico

Resumo

Aprofundamento

Lembre-se: Ao ler um artigo, você não terá acesso a estas visualizações!

Análise de resíduos dos 2 últimos exemplos



Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

Lembre-se: Ao ler um artigo, você não terá acesso a estas visualizações!

Regressão linear simples x múltipla⁸



Outros fatores

Nesses casos, não podemos explicar a variância do BMD apenas com o BMI.

É evidente que algum outro fator *deveria* ter sido considerado no modelo

Isto permite **ajustar** a heterogeneidade da variância observada com outros cofatores (além do BMI).

Regressão
Linear
Simple

Felipe
Figueiredo

Modelagem

Regressão

Introdução

A regressão

r^2

Exercício

Bônus: preditor

categorico

Resumo

Aprofundamento

(sai a regressão linear **simples** e entra a **múltipla** – Cap 31)

⁸ Ajustar para outros fatores: como visto no abstract do exercício!

Queremos avaliar a eficiência de uma nova dieta reduzida em gordura no tratamento de obesidade.

Selecionamos aleatoriamente 100 pessoas obesas para o grupo 1, que receberão a dieta com pouca gordura. Selecionamos outras 100 pessoas obesas para o grupo 2 que receberão a mesma quantidade de comida, com proporção normal de gordura. O estudo durou 4 meses.

A perda de peso média no grupo 1 foi de 9.33 lbs ($s=4.72$) e no grupo 2 foi de 7.58 lbs ($s=3.90$).

Essa nova dieta é eficaz na perda de peso?

Fonte: Khan Academy

Saída típica de um programa

```
Residuals:
    Min       1Q   Median       3Q      Max
-13.7754  -3.1275  -0.2171   3.0112  11.9957

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.3340     0.4332   21.548  < 2e-16 ***
GrupoGrupo2  -1.7587     0.6126   -2.871   0.00454 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.332 on 198 degrees of freedom
Multiple R-squared:  0.03996, Adjusted R-squared:  0.03512
F-statistic: 8.242 on 1 and 198 DF,  p-value: 0.004537
```

Interpretação (assumindo pareamento)

- Perda média do grupo 1 (referência): 9.33 lbs (IC=[8.48, 10.19]).
- Perda média do grupo 2 em relação à referência: -1.76 lbs (IC=[-2.97, -0.55]).

- Quão bem a reta regressora se ajusta aos dados?
- O que pode explicar a relação observada?
- Qual proporção (porcentagem) da variância pode ser explicada pelas variáveis analisadas?
- É necessário investigar a relação entre as variáveis!
- O modelo de RLS permite preditor categórico (com qualquer número de níveis!)

Aprofundamento

Leitura obrigatória

- Capítulo 18
- Capítulo 19, pular as seções:
 - regressão linear como método de mínimos quadrados
 - calculando a regressão linear

Leitura recomendada

- **Capítulo 31** — fortemente recomendado para a aula que vem!
- Schneider A, Hommel G, Blettner M, 2010.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2992018/>