

Regressão Logística

Felipe Figueiredo

Tópicos em

Regressão

Tópicos em Regressão Logística Modelos com desfecho categórico binário

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

Discussão da aula passada

Discussão da leitura obrigatória da aula passada



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Múltipla

Sumário



Regressão Linear Múltipla

2 Regressão Logística

Regressão Logística

Aprofundamento

JOURNAL OF WOMEN'S HEALTH © Mary Ann Liebert, Inc.

Aprofundamento



Tópicos em

Regressão

Logística

Felipe

Figueiredo

Regressão Logística

Felipe Figueiredo

Regressão Múltipla

Regressão

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

The Association between Body Mass Index and

Osteoporosis in Patients Řeferred for a Bone Mineral Density Examination

Tópicos em

Regressão Linear Múltipla



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear

Regressão

Aprofundamento

Hoje vamos interpretar os resultados do abstract



Tópicos em Regressão Logística

Felipe Figueiredo

Linear
Múltipla
Regressão Linear
Múltipla

Regressão

Anrofundamento

Para isto selecionaram 100 mulheres brancas e 100 mulheres pardas, de meia idade. Mensuraram a BMD e calcularam o BMI delas.

Os pesquisadores querem investigar se a etnicidade das participantes

tem algum efeito detectável na associação entre a densidade mineral

óssea (BMD) e o índice de massa corpórea (BMI).

Enunciado 1



Tópicos em Regressão

ABSTRACT

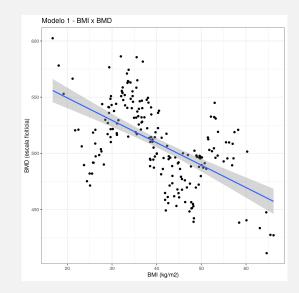
Purpose: Osteoporosis affects 4–6 million (13%–18%) postmenopausal white women in the United States. Most studies to date on risk factors for osteoporosis have considered body mass index (BMI) only as a possible confounder. In this study, we assess the direct relationship between BMI and osteoporosis.

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 (\sim 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.

Modelo 1





Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundamento

ento

Quais são as variáveis?

Tópicos em Regressão Logística

Felipe Figueiredo

Múltipla

Regressão Linear

Logística

Esta relação pode ser expressa como

Dependente: BMD (contínua)

Independente: BMI (contínua)

 $BMD \sim BMI$

Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Múltipla Regressão Linear

Modelo 1

Residuals: Min 1Q Median 3Q Max -67.833 -21.767 2.178 20.743 67.185 Coefficients: Estimate Std. Error t value Pr(>|t|) (Intercept) 589.199 8.499 69.322 <2e-16 ***
BMI -1.995 0.205 -9.732 <2e-16 ***

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

Residual standard error: 29.41 on 198 degrees of freedom Multiple R-squared: 0.3236, Adjusted R-squared: 0.3201 F-statistic: 94.71 on 1 and 198 DF, p-value: < 2.2e-16

Modelo 1 completo

 $BMD = 589.20 - 1.99 \times BMI$

Componentes do modelo 1

Versão simplificada (apenas variáveis)

 $BMD \sim BMI$

Modelo completo

 $\mathsf{BMD} = \beta_0 + \beta_1(\mathsf{BMI}) + \varepsilon$

Hipótese: ε é um erro aleatório 1 normalmente distribuído e centrado em zero – a incerteza que não pode ser controlada.

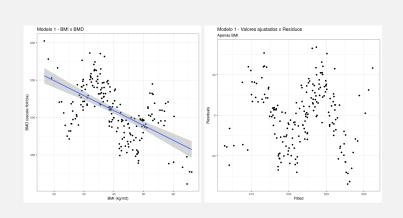
Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Múltipla

Regressão Linear Múltipla







Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Múltipla Regressão Linear Múltipla

¹residual – não é explicado pela relação entre as variáveis do modelo



Modelo 1 completo

 $BMD = 589.20 - 1.99 \times BMI$

Interpretação

As participantes perdem, na média, 1.99 unidades de BMD para cada incremento unitário do BMI.

Este é o chamado resultado bruto. Agora vamos ajustá-lo com outros preditores.

Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla

Regressão Linear Múltipla Regressão Logística

Aprofundamento

Agora vamos ver se a etnia tem algum efeito



Tópicos em Regressão Logística

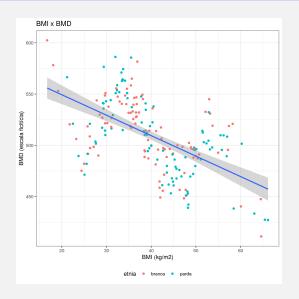
> Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundamento

Modelo 2





Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundamento

Quais são as variáveis?

- Dependente: BMD (contínua)
- Independente: BMI (contínua)
- Independente: etnia (categórica binária)

Esta relação pode ser expressa como

 $\mathsf{BMD} \sim \mathsf{BMI} + \mathsf{etnia}$



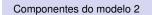
Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundament



Versão simplificada (apenas variáveis)



Tópicos em Regressão Logística

> Felipe Figueiredo

Regressão Linear Múltipla

Regressão Linear Múltipla

Logística

Aprofundamento

Modelo completo

$$\mathsf{BMD} = \beta_0 + \beta_1(\mathsf{BMI}) + \beta_2(\mathsf{etnia}) + \varepsilon$$

 $\mathsf{BMD} \sim \mathsf{BMI} + \mathsf{etnia}$

Modelo 2

```
Residuals:
   Min
            10 Median
                           3Q
-67.357 -22.005 1.801 20.785 67.616
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 589.4360 8.6055 68.495 <2e-16 ***
BMI
            -1.9905
                       0.2067 -9.632 <2e-16 ***
etniaparda -0.8206
                     4.1938 -0.196
                                       0.845
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 29.49 on 197 degrees of freedom
Multiple R-squared: 0.3237, Adjusted R-squared: 0.3168
F-statistic: 47.14 on 2 and 197 DF, p-value: < 2.2e-16
```

Que outra variável os pesquisadores deveriam ter investigado?



Tópicos em Regressão Logística

Felipe Figueiredo

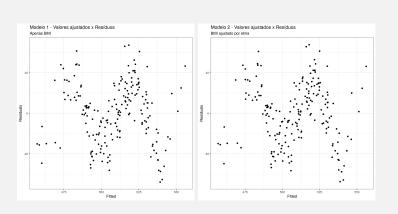
Regressão Linear Múltipla

Regressão Linear Múltipla

Logística

Aprofundamento

Modelo 2





Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundamento



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundament

Modelo 2.1 - idade

BMI x BMD



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Múltipla Regressão Linear Múltipla

Logística

Aprofundamento



Tópicos em Regressão

Felipe Figueiredo

Regressão Múltipla

Componentes do modelo 2.1

Versão simplificada (apenas variáveis)



Logística

Regressão Linear Múltipla

Modelo completo

 $BMD = \beta_0 + \beta_1(BMI) + \beta_2(idade) + \varepsilon$

 $BMD \sim BMI + idade$

BMI (kg/m2)

Quais são as variáveis?

Dependente: BMD (contínua) Independente: BMI (contínua)

Independente: idade (contínua)

Esta relação pode ser expressa como

 $BMD \sim BMI + idade$



Regressão Logística

Felipe Figueiredo

Regressão Múltipla Regressão Linear Múltipla

Logística

Modelo 2.1

Residuals:

1Q Median 3Q -52.039 -24.688 -0.058 23.599 42.146

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 780.0347 22.2560 35.048 <2e-16 *** -2.0350 0.1727 -11.781 <2e-16 *** BMI idade -3.1429 0.3471 -9.056 <2e-16 ***

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

Residual standard error: 24.78 on 197 degrees of freedom Multiple R-squared: 0.5224, Adjusted R-squared: 0.5175 F-statistic: 107.7 on 2 and 197 DF, p-value: < 2.2e-16

Modelo 2.1 completo

 $BMD = 780.03 - 2.04 \times BMI - 3.02 \times idade$

Tópicos em Regressão Logística

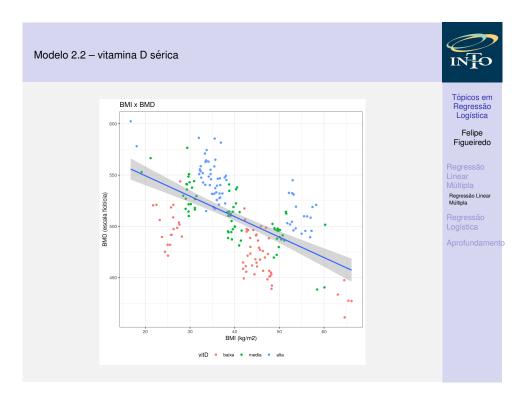
Felipe Figueiredo

Múltipla Regressão Linear Múltipla

Modelo 2.1 Tópicos em Regressão Logistica Felipe Figueiredo Regressão Linear Múltipla Regressão Logistica Aprofundamento







Quais são as variáveis?

Dependente: BMD (contínua)Independente: BMI (contínua)

● Independente: vitamina D sérica (categórica – 3 níveis)

Esta relação pode ser expressa como

 $BMD \sim BMI + vitD$

INTO

Tópicos em Regressão Logística

Felipe Figueiredo

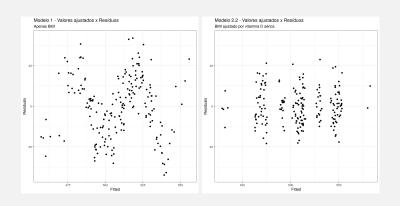
Regressão Linear Múltipla

Regressão Linear Múltipla

Logística

Aprofundamento

Modelo 2.2





Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundamento

Componentes do modelo 2.2

Modelo completo

Versão simplificada (apenas variáveis)



Tópicos em Regressão Logística

> Felipe Figueiredo

Regressão Linear Múltipla

Regressão Linear Múltipla

Regressão Logística

Aprofundamento

Modelo 2.2

 $BMD \sim BMI + vitD$

 $\mathsf{BMD} = \beta_0 + \beta_1(\mathsf{BMI}) + \beta_2(\mathsf{vitD}) + \varepsilon$

Modelo 2.2 completo

 $\mathsf{BMD} = 555.45 - 1.91 \times \mathsf{BMI} + \beta_2 \times \mathsf{vitD}$



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundament



Modelo 1 completo

 $\mathsf{BMD} = \mathsf{589.20} - \mathsf{1.99} \times \mathsf{BMI}$

Modelo 2.2 completo

 $\mathsf{BMD} = 555.45 - 1.91 \times \mathsf{BMI} + \beta_2 \times \mathsf{vitD}$

Interpretação

As participantes perdem, na média, 1.99 unidades de BMD para cada incremento unitário do BMI (resultado bruto).

Após ajustar pelo nível sérico de vitamina D, o resultado é 1.91.

Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla

Regressão Linear Múltipla

Logística

Aprofundamento

Componentes do modelo 3

Versão simplificada (apenas variáveis)

 $BMD \sim BMI + idade + vitD$

Modelo completo

 $BMD = \beta_0 + \beta_1(BMI) + \beta_2(idade) + \beta_3(vitD) + \varepsilon$



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear

Múltipla
Regressão

Anrofundamento



Tópicos em Regressão Logística

> Felipe Figueiredo

Regressão Linear Múltipla

Regressão Linear Múltipla

Regressão Logística

Aprofundamento

Modelo 3

```
Residuals:
             1Q Median
                             3Q
   Min
-19.7284 -4.4059 -0.3755 4.7776 24.0977
           Estimate Std. Error t value Pr(>|t|)
(Intercept) 740.47234 6.41200 115.48 <2e-16 ***
           -1.94958 0.04933 -39.52 <2e-16 ***
BMI
           -3.04192 0.09904 -30.72 <2e-16 ***
vitDmedia 31.14768 1.22729 25.38 <2e-16 ***
vitDalta
          58.07642 1.23165 47.15 <2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 7.068 on 195 degrees of freedom
Multiple R-squared: 0.9615, Adjusted R-squared: 0.9607
F-statistic: 1218 on 4 and 195 DF, p-value: < 2.2e-16
```

Agora um modelo maior (ajustando para todas as variáveis

relevantes)

Modelo 3 completo

 $\mathsf{BMD} = 740.47 - 1.95 \times \mathsf{BMI} - 3.04 \times \mathsf{idade} + \beta_3 \times \mathsf{vitD}$



Tópicos em Regressão Logística

> Felipe Figueiredo

Regressão Linear Múltipla Regressão Linear Múltipla

Regressão Logística

Aprofundamen

Modelo 3



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Múltipla

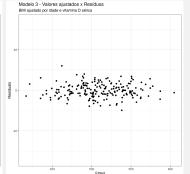
Regressão Linear Múltipla

Aprofundamento

Modelo 3 - Valores ajustados x Resíduos BMI ajustado por idade e vitamina D sérica



Modelo 1 - Valores ajustados x Resíduos Apenas BMI



Desfecho binário

Vamos discretizar os dados em duas categorias

Osteoporose

• Sadio: BMD >= 500 Osteoporose: BMD < 500

Idoso

Não idoso: idade < 60 Idoso: idade >= 60

Obeso

Não obeso: BMI < 30 Obeso: BMI >= 30



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão

Regressão Logística



Modelo 1 completo

 $\mathsf{BMD} = \mathsf{589.20} - \mathsf{1.99} \times \mathsf{BMI}$

Modelo 3 completo

 $\mathsf{BMD} = 740.47 - 1.95 \times \mathsf{BMI} - 3.04 \times \mathsf{idade} + \beta_3 \times \mathsf{vitD}$

Interpretação

As participantes perdem, na média, 1.99 unidades de BMD para cada incremento unitário do BMI (resultado bruto).

Após ajustar pela idade e pelo nível sérico de vitamina D, o resultado é 1.95.

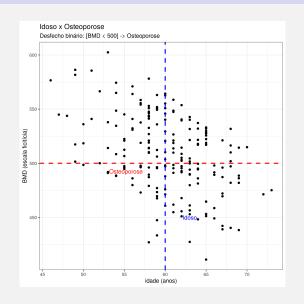
Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Regressão Linear Múltipla

Logística

Modelo 4





Tópicos em Regressão Logística

Felipe Figueiredo

Regressão

Regressão Logística

Quais são as variáveis?

- INTO
- Tópicos em Regressão Logística
- Felipe Figueiredo
- Regressão Linear Múltipla
- Regressão
- Logistica Regressão Logistica

Aprofundamento

Esta relação pode ser expressa como

Dependente: Osteoporose (categórica – binária)

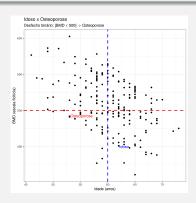
Independente: Idoso (categórica – binária)

Osteoporose \sim Idoso

INTO

Tabela de contingência Idoso x Osteoporose

osteo										
idoso		Sadio	Osteoporose							
Nao	Idoso	98	6							
Idos	30	68	28							



Tópicos em Regressão Logística

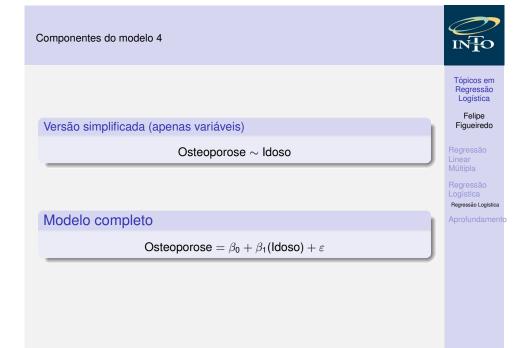
Felipe Figueiredo

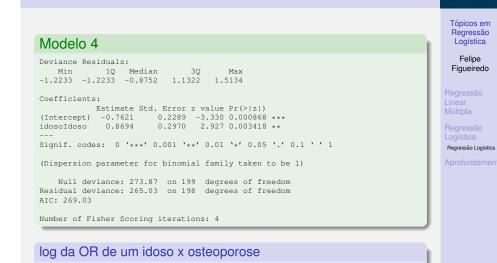
Regressão Linear Múltipla

ogística

Regressão Logística

Aprofundament





log(OR) = 0.8694



Tópicos em Regressão Logística

Felipe

Figueiredo

Regressão Logística

Regressão

Transformando o log da OR na OR

 $log(OR) \approx 0.87...$

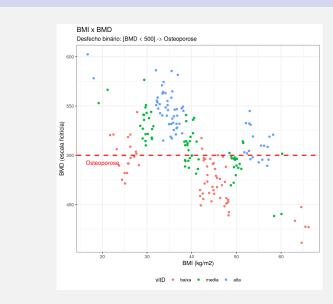
... portanto...

 $OR = e^{0.87} \approx 2.4$

Resultado

• (Idoso) OR: 2.39, IC: [1.33, 4.27]

Modelo 5



INTO

Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla

Regressão

Regressão Logística

Aprofundamento

Quais são as variáveis?

- Dependente: Osteoporose (categórica binária)
- Independente: BMI (contínua)
- Independente: idade (contínua)
- Independente: vitamina D sérica (categórica 3 níveis)

Esta relação pode ser expressa como

Osteoporose \sim BMI + idade + vitD



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla

Regressão

Regressão Logística

Aprofundamento

Componentes do modelo 5

Versão simplificada (apenas variáveis)

INTO

Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla

Regressão

Regressão Logística

Modelo completo

Osteoporose = $\beta_0 + \beta_1(BMI) + \beta_2(idade) + \beta_3(vitD) + \varepsilon$

Osteoporose \sim BMI + idade + vitD



Modelo 5

```
Deviance Residuals:
   Min
             10
                   Median
                                 3Q
                                         Max
-2.61379 -0.07713 -0.00244 0.08301
Coefficients:
           Estimate Std. Error z value Pr(>|z|)
(Intercept) -53.07482 10.80588 -4.912 9.03e-07 ***
            0.43097 0.08301 5.191 2.09e-07 ***
idade
            0.67277 0.14175 4.746 2.07e-06 ***
           -6.13676
                      1.28900 -4.761 1.93e-06 ***
vitDmedia
vitDalta -12.59502 2.38905 -5.272 1.35e-07 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
(Dispersion parameter for binomial family taken to be 1)
   Null deviance: 273.869 on 199 degrees of freedom
Residual deviance: 69.365 on 195 degrees of freedom
AIC: 79.365
```

Modelo 5 completo

 $log(Osteoporose) = -77.43 + 0.46 \times BMI + 0.83 \times idade + \beta_3 \times vitD$



Tópicos em

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.



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Logística



Modelo 5 completo

 $log(Osteoporose) = -77.43 + 0.46 \times BMI + 0.83 \times idade + \beta_3 \times vitD$

Resultado

• (BMI) OR: 1.54, IC: [1.31, 1.81]

(idade) OR: 1.96, IC: [1.48, 2.59]

• (vitD média x baixa) **OR**: **0.002161**, **IC**: [**0.0001728**, **0.0270419**]

(vitD alta x baixa) OR: 3.388847e-06, IC: [3.13678e-08, 3.66116e-04]

Interpretação

Após ajustar pela idade e pelo nível sérico de vitamina D, as participantes tem chance aumentada de desenvolver osteoporose para cada incremento unitário do BMI.

Como tabular os dados necessários para esta análise?

Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Logística

Bônus



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão

Regressão Logística

Bônus



Como tabular os dados necessários para esta análise?

	Α	В	С	D	E	F	G	Н
1	BMI 🔻	BMD ▼	etnia 🔻	idade 🔻	vitD ▼	osteo 🔻	obeso 🔻	idoso ▼
2	26,24		branca	62	baixa			ldoso
3	31,10	516,43	parda	68	media	Sadio	Obeso	Idoso
4	24,99		branca		baixa			ldoso
5	39,57	514,25	branca	58	media	Sadio	Obeso	Nao Idoso
6	31,98	586,27			alta			Nao Idoso
7	25,08	471,34	parda	72	baixa			Idoso
8	32,92		branca		alta			ldoso
9	34,43	540,17	branca	63	alta	Sadio	Obeso	Idoso
10	33,45	562,70	parda	60	alta			ldoso
11	28,17	490,01	branca	63	baixa			Idoso
12	39,07	512,27	parda	59	media	Sadio	Obeso	Nao Idoso
13	32,34	553,67		62	alta	Sadio	Obeso	Idoso
14	26,27	520,95	parda	58	baixa	Sadio	Nao Obeso	Nao Idoso
15	16,71	602,43	branca	53	alta	Sadio	Nao Obeso	Nao Idoso
16	36,75	528,11	branca	65	alta	Sadio		Idoso
17	29.73	510.03	parda	68	media	Sadio	Nao Obeso	Idoso

Dados coletados

BMI, BMD, etnia, idade, vitD

Dados calculados

osteo = BMD < 500; obeso = BMI > 30; idoso = idade > 60

Aprofundamento



Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla

Regressão Logística

Regressão Logística

Aprofundamento

Tópicos em Regressão Logística

Felipe Figueiredo

Regressão Linear Múltipla

Regressão Logística

Aprofundamento

Aprofundamento

Leitura recomendada

Leitura obrigatória

Capítulo 31

Capítulo 32

Capítulo 25: seção teste t de uma razão (sobre o uso do logaritmo)