

Modelos de Regressão:

aplicações em dados de contagem e proporções

08 de agosto de 2017

André F. B. Menezes

Departamento de Estatística
Universidade Estadual de Maringá – UEM





Organização

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Resultados e
Discussões

Referências

Introdução

Modelos de Regressão

Modelos para dados de contagem

Modelos para dados no intervalo unitário

Cr terios para compara  o de modelos

Materiais

Recursos computacionais

Conjuntos de dados

N mero de publica  es produzidas por Ph.D. em Bioqu mica

Vota  o presidencial de 2010

Resultados e Discuss es

An lise do n mero de publica  es produzidas por Ph.D.
em Bioqu mica

An lise da vota  o presidencial de 2010

Refer ncias



Introdução

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Resultados e
Discussões

Referências

2

Motivação e Objetivos

- ▶ Trabalho final para a disciplina Modelos Lineares Generalizados;
- ▶ Realizar a modelagem de dois conjuntos de dados:
 1. **Discreto:** número de publicações dos Ph.D. em Bioquímica (LONG, 1990);
 2. **Contínuo:** proporção de votos da ex-presidenta Dilma na eleição de 2010 no Paraná.



Modelos de Regressão

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

3

Modelos para dados de
contagem

Modelos para dados no
intervalo unitário

Crítérios para comparação
de modelos

Materiais

Resultados e
Discussões

Referências

Conceito

Permitem a inclusão de variáveis independentes (covariáveis) para:

- ▶ Descrever a relação entre a variável resposta e as variáveis preditoras;
- ▶ Realizar predições por meio do modelo estabelecido.

Um modelo de regressão pode ser expresso por:

$$\begin{aligned} Y | \mathbf{X} &\sim f(\theta) \\ Q(Y | \mathbf{X}) &= g(\mathbf{X} | \beta) \end{aligned} \tag{1}$$



Modelos Lineares Generalizados

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Modelos para dados de contagem

Modelos para dados no intervalo unitário

Crítérios para comparação de modelos

Materiais

Resultados e Discussões

Referências

4

Definição

Dada uma amostra aleatória de n observações (y_i, \mathbf{x}_i) , em que $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$, os MLG's são definidos por:

(i) **Componente aleatório:**

$$f(y_i | \theta_i, \phi) = \exp \left\{ \frac{1}{a(\phi)} [y_i \phi + b(\theta_i)] + c(y_i, \phi) \right\} \quad (2)$$

(ii) **Componente sistemático:**

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X} \boldsymbol{\beta} \quad (3)$$

(iii) **Função de ligação:**

$$\eta_i = g(\mu_i) \quad (4)$$



Modelos para dados de contagem

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Modelos para dados de
contagem

Modelos para dados no
intervalo unitário

Crítérios para comparação
de modelos

Materiais

Resultados e
Discussões

Referências

5

Modelos considerados

- ▶ Modelo Poisson;
- ▶ Modelo Quase-Poisson;
- ▶ Modelo Binomial Negativo.

Modelos alternativos

- ▶ Poisson Generalizada;
- ▶ COM-Poisson;
- ▶ Gamma-Count;
- ▶ Weibull discreta;
- ▶ Lindley discreta.



Modelos para dados de contagem

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Modelos para dados de
contagem

Modelos para dados no
intervalo unitário

Crêterios para comparaçã
de modelos

Materiais

Resultados e
Discussões

Referências

6

Modelo Poisson

- Função massa de probabilidade:

$$P(Y = y \mid \mu) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \dots \quad (5)$$

- Especificação:

$$Y_i \mid \mathbf{x}_i \sim \text{Poisson}(\mu_i) \quad \text{com} \quad \log(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (6)$$

- Função log-verossimilhança:

$$\ell(\boldsymbol{\beta} \mid y_i, \mathbf{x}_i) = \sum_{i=1}^n y_i \mathbf{x}_i^\top \boldsymbol{\beta} - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) - \log(y_i!) \quad (7)$$



Modelos para dados de contagem

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Modelos para dados de
contagem

Modelos para dados no
intervalo unitário

Crerios para comparaço
de modelos

Materiais

Resultados e
Discussões

Referências

7

Modelo Quase-Poisson

- Função quase-verossimilhança:

$$Q(\mu_i | y_i) = \int_{y_i}^{\mu_i} \frac{y_i - t}{\sigma^2 V(t)} dt. \quad (8)$$

em que $E(Y_i) = \mu_i$ e $\text{Var}(Y_i) = \sigma^2 V(\mu_i)$.

- Estimao do parâmetro extra:

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \quad (9)$$

Se $\sigma^2 < 1$ têm-se subdispersão e $\sigma^2 > 1$ têm-se superdispersão.



Modelos para dados de contagem

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Modelos para dados de
contagem

Modelos para dados no
intervalo unitário

Crítérios para comparação
de modelos

Materiais

Resultados e
Discussões

Referências

8

Modelo Binomial Negativo

- ▶ Função massa de probabilidade:

$$P(Y = y \mid \mu, \theta) = \frac{\Gamma(\theta + y)}{\Gamma(y + 1) \Gamma(\theta)} \left(\frac{\mu}{\mu + \theta} \right)^y \left(\frac{\theta}{\mu + \theta} \right)^\theta, \quad (10)$$

- ▶ Foi utilizado a parametrização $\theta = \phi^{-1}$ sugerida por Lawless (1987).
- ▶ Acomoda casos de superdispersão, sendo caracterizado quando $\phi > 0$.
- ▶ Especificação:

$$Y_i \mid \mathbf{x}_i \sim \text{BN}(\mu_i, \phi) \quad \text{com} \quad g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (11)$$



Modelos para dados de contagem

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Modelos para dados de contagem

Modelos para dados no intervalo unitário

Crítérios para comparação de modelos

Materiais

Resultados e Discussões

Referências

9

Observações

- ▶ A função log foi escolhida como função de ligação, isto é,

$$g(\mu_i) = \eta_i = \log(\mu_i) \rightarrow \mu_i = \exp[\mathbf{x}_i^\top \boldsymbol{\beta}]$$

- ▶ No modelo Quase-Poisson temos que $V(\mu_i) = \mu_i$, logo $\text{Var}(Y_i) = \sigma^2 \mu_i$, portanto as estimativas serão idênticas.
- ▶ Propriedade dos modelos de quase-verossimilhança:

$$-E\left(\frac{\partial^2}{\partial \mu_i^2} Q(\mu_i | y_i)\right) \leq -E\left(\frac{\partial^2}{\partial \mu_i^2} \ell(\mu_i | y_i)\right) \quad (12)$$



Modelos para dados no intervalo unitário

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Modelos para dados de
contagem

Modelos para dados no
intervalo unitário

Critérios para comparação
de modelos

Materiais

Resultados e
Discussões

Referências

10

Distribuições consideradas

- ▶ Beta;
- ▶ Simplex.

Distribuições alternativas

- ▶ Kumarasawamy;
- ▶ Johnson S_B ;
- ▶ unit-Logistic;
- ▶ unit-Gamma.



Modelos para dados no intervalo unitário

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Modelos para dados de contagem

Modelos para dados no intervalo unitário

Crítérios para comparação de modelos

Materiais

Resultados e Discussões

Referências

11

Modelo de Regressão Beta

- ▶ Formalmente introduzido por Ferrari e Cribari-Neto (2004).
- ▶ Função densidade de probabilidade:

$$f(y \mid \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1} \quad (13)$$

- ▶ Especificação:

$$Y_i \mid \mathbf{x}_i \sim \text{Beta}(\mu_i, \phi) \quad \text{com} \quad g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (14)$$



Modelos para dados no intervalo unitário

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Modelos para dados de contagem

Modelos para dados no intervalo unitário

Critérios para comparação de modelos

Materiais

Resultados e Discussões

Referências

12

Modelo de Regressão Simplex

- ▶ Proposto por Barndorf-Nielsen e Jørgensen (1991).
- ▶ Função densidade de probabilidade:

$$f(y \mid \mu, \phi) = [2 \pi \phi^2 \{y(1-y)\}^3]^{-1/2} \exp \left\{ -\frac{1}{2 \phi^2} d(y; \mu) \right\}$$

em que

$$d(y; \mu) = \frac{(y - \mu)^2}{y(1-y) \mu^2 (1 - \mu)^2}$$

- ▶ Especificação:

$$Y_i \mid \mathbf{x}_i \sim \text{Simplex}(\mu_i, \phi) \quad \text{com} \quad g(\mu_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta} \quad (15)$$

- ▶ Vantagem: os parâmetros μ e ϕ são ortogonais.



Modelos para dados no intervalo unitário

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Modelos para dados de contagem

Modelos para dados no intervalo unitário

Critérios para comparação de modelos

Materiais

Resultados e Discussões

Referências

13

Observações

- ▶ A função *logit* foi escolhida como função de ligação, isto é,

$$g(\mu_i) = \eta_i = \log \left(\frac{\mu_i}{1 - \mu_i} \right) \rightarrow \mu_i = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}$$

- ▶ Os parâmetros de precisão (Beta) e dispersão (Simplex) foram considerados constantes.



Cr terios para compara  o de modelos

Modelos de Regress o

Andr  F. B. Menezes

Introdu  o

Modelos de
Regress o

Modelos para dados de
contagem

Modelos para dados no
intervalo unit rio

Cr terios para compara  o
de modelos

14

Materiais

Resultados e
Discuss es

Refer ncias

- Teste da raz o de verossimilhan as (deviance):

$$S_{LR} = 2 \left[\ell(\hat{\theta}_q \mid y_i) - \ell(\hat{\theta}_p \mid y_i) \right]$$

$$S_{LR} \sim \chi^2_{p-q}$$

- Cr terio de Informa  o de Akaike:

$$AIC = 2k - 2\ell(\hat{\theta} \mid y_i)$$

- Cr terio de Informa  o de Bayesiano:

$$BIC = \log(n)k - 2\ell(\hat{\theta} \mid y_i)$$

- An lise dos res duos via gr fico normal de probabilidades com envelope simulado.



Cr terios para compara  o de modelos

Modelos de Regress  o

Andr  F. B. Menezes

Introdu  o

Modelos de
Regress  o

Modelos para dados de
contagem

Modelos para dados no
intervalo unit rio

Cr terios para compara  o
de modelos

15

Materiais

Resultados e
Discuss  es

Refer ncias

- Teste para modelos n o aninhados (VUONG, 1989).
- \mathcal{H}_0 : modelos s o equivalentes.

$$T_{LR,NN} = \frac{1}{\hat{\omega}^2 \sqrt{n}} \sum_{i=1}^n \log \frac{f(y_i | \mathbf{x}_i, \hat{\theta})}{g(y_i | \mathbf{x}_i, \hat{\gamma})}$$

em que

$$\hat{\omega}^2 = \frac{1}{n} \sum_{i=1}^n \left(\log \frac{f(y_i | \mathbf{x}_i, \hat{\theta})}{g(y_i | \mathbf{x}_i, \hat{\gamma})} \right)^2 - \left(\frac{1}{n} \sum_{i=1}^n \left(\log \frac{f(y_i | \mathbf{x}_i, \hat{\theta})}{g(y_i | \mathbf{x}_i, \hat{\gamma})} \right) \right)^2$$

$$T_{LR,NN} \xrightarrow{D} N(0, 1)$$

- Rejeita-se \mathcal{H}_0 se $|T_{LR,NN}| < z_{\alpha/2}$.
- $f(y_i | \mathbf{x}_i, \hat{\theta})$ sendo “melhor”(“pior”) do que $g(y_i | \mathbf{x}_i, \hat{\gamma})$ se $T_{LR,NN} > z_{\alpha}$ (ou $T_{LR,NN} < -z_{\alpha}$).



Recursos Computacionais

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Recursos computacionais

16

Conjuntos de dados

Número de publicações
produzidas por Ph.D. em
Bioquímica

Votação presidencial de
2010

Resultados e
Discussões

Referências

Software SAS® 9.4

Procedimentos utilizados:

- ▶ PROC GENMOD: ajuste dos modelos Poisson e Binomial Negativo;
- ▶ PROC COUNTREG: ajuste dos modelos Poisson, Binomial Negativo, ZIP e ZINB;
- ▶ PROC GLIMMIX: ajuste dos modelos Quase-Poisson e Beta;
- ▶ PROC NLMIXED: ajuste dos modelos Beta e Simplex;



Recursos Computacionais

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações
produzidas por Ph.D. em
Bioquímica

Votação presidencial de
2010

Resultados e
Discussões

Referências

17

Software R, versão 3.3.2

Bibliotecas auxiliares:

- ▶ `betareg`: ajuste do modelo Beta;
- ▶ `simplexreg`: ajuste do modelo Simplex;
- ▶ `hnp`: gráficos (meio) normais com envelope simulado;
- ▶ `ggplot2`: recursos gráficos;



Conjuntos de dados

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações
produzidas por Ph.D. em
Bioquímica

Votação presidencial de
2010

Resultados e
Discussões

Referências

18

Número de publicações produzidas por Ph.D. em Bioquímica

- ▶ Inicialmente analisado por Long (1990) consiste de um estudo observacional;
- ▶ A população foi definida como todos os bioquímicos que receberam o título de Ph.D. nos EUA durante 1956–1958 e 1961–1963;
- ▶ Principal objetivo foi identificar os efeitos de determinadas covariáveis sobre a produtividade dos bioquímicos durante seu Ph.D.



Conjuntos de dados

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações
produzidas por Ph.D. em
Bioquímica

Votação presidencial de
2010

Resultados e
Discussões

Referências

19

Número de publicações produzidas por Ph.D. em Bioquímica

Levantamento contou com 915 observações contendo informações das seguintes variáveis:

Variável	Definição
art	Número de publicações em revistas científicas nos últimos três anos do Ph.D.
fem	Sexo do individuo
mar	Estado civil (casado ou solteiro)
kid5	Número de filhos menores que 6 anos
phd	Prestigio do programa de Ph.D.
ment	número de artigos publicados pelo orientador nos últimos três anos



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Recursos computacionais

Conjuntos de dados

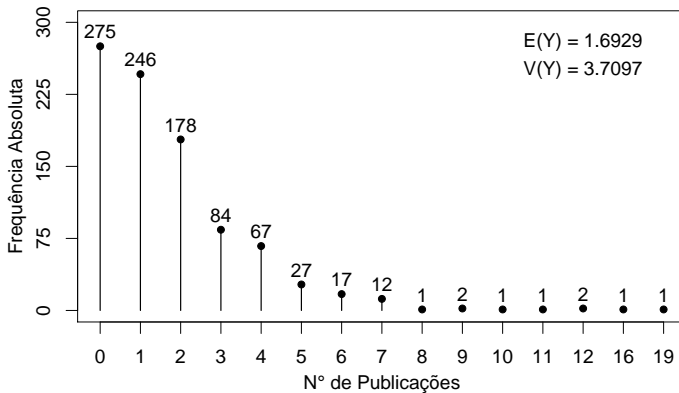
Número de publicações produzidas por Ph.D. em Bioquímica

Votação presidencial de 2010

Resultados e Discussões

Referências

20





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações produzidas por Ph.D. em Bioquímica

Votação presidencial de 2010

Resultados e Discussões

Referências

21

Medidas amostrais do número de publicações dos(as) bioquímicos(as) Ph.D. conforme gênero e estado civil.

Gênero	Estado Civil	N	Mínimo	Máximo	Média	Variância	CV (%)
Homem	Solteiro	113	0	7	1.9469	4.0507	103.3766
	Casado	381	0	19	1.8635	4.9655	119.5775
Mulher	Solteira	196	0	7	1.3878	2.2796	108.7979
	Casada	225	0	10	1.5422	2.5172	102.8751



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações produzidas por Ph.D. em Bioquímica

Votação presidencial de 2010

Resultados e Discussões

Referências

22

Medidas amostrais do número de publicações dos(as) bioquímicos(as) Ph.D. conforme número de filhos.

Nº de Filhos	N	Mínimo	Máximo	Média	Variância	CV (%)
0	599	0	19	1.7212	3.7365	112.3057
1	195	0	12	1.7590	4.1942	116.4298
2	105	0	11	1.5429	3.0198	112.6320
3	16	0	3	0.8125	0.8292	112.0721



Conjuntos de dados

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações
produzidas por Ph.D. em
Bioquímica

Votação presidencial de
2010

Resultados e
Discussões

Referências

23

Votação presidencial de 2010

- ▶ Variáveis relacionadas a eleição de 2010 → Furriel (2017).
- ▶ Variáveis demográficas do censo de 2010 → <http://www.atlasbrasil.org.br>.
- ▶ Objetivo: verificar quais foram os fatores que influenciaram o percentual de votos recebidos por Dilma no estado do Paraná no segundo turno da eleição de 2010.



Conjuntos de dados

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações
produzidas por Ph.D. em
Bioquímica

Votação presidencial de
2010

Resultados e
Discussões

Referências

24

Votação presidencial de 2010

Os dados cobrem 395 municípios do Paraná.

Variável	Definição
pt2010	Proporção de votos válidos do PT no segundo turno da eleição de 2010
pt2006	Proporção de votos válidos do PT no segundo turno da eleição de 2006
gini	Índice de Gini
idhm_e	Índice de Desenvolvimento Humano Municipal – Dimensão Educação
idhm_l	Índice de Desenvolvimento Humano Municipal – Dimensão Longevidade
idhm_r	Índice de Desenvolvimento Humano Municipal – Dimensão Renda
urb	Proporção da população vivendo em áreas urbanas
des	Proporção da PEA com 18 ou mais de idade em 2010 que estava desocupada
pib	PIB per capita do município



Votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Recursos computacionais

Conjuntos de dados

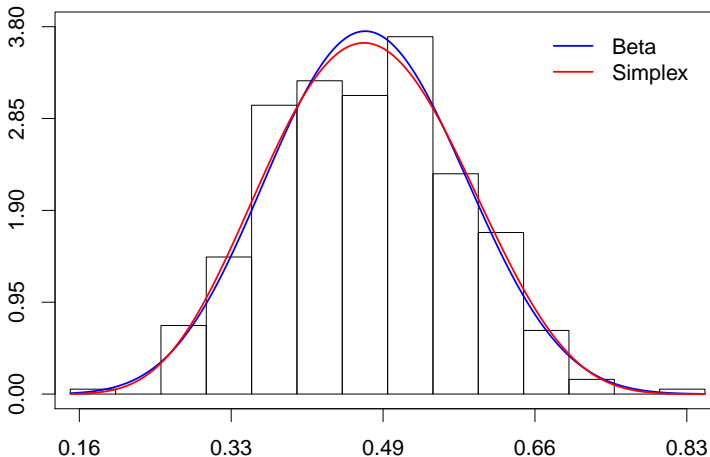
Número de publicações
produzidas por Ph.D. em
Bioquímica

Votação presidencial de
2010

25

Resultados e
Discussões

Referências





Votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Recursos computacionais

Conjuntos de dados

Número de publicações produzidas por Ph.D. em Bioquímica

Votação presidencial de 2010

26

Resultados e Discussões

Referências

Estatísticas descritivas das variáveis consideradas.

Variável	Mínimo	$Q_{0.25}$	Mediana	Média	$Q_{0.75}$	Máximo	CV%
pt2010	0.1877	0.3995	0.4784	0.4782	0.5492	0.8098	21.5757
pt2006	0.1426	0.3611	0.4271	0.4291	0.5063	0.7539	22.8908
gini	0.3300	0.4300	0.4700	0.4657	0.5000	0.6600	12.2630
idhm_e	0.3620	0.5760	0.6210	0.6110	0.6550	0.7680	10.2870
idhm_l	0.7650	0.8050	0.8210	0.8205	0.8360	0.8700	2.5558
idhm_r	0.5700	0.6690	0.6920	0.6919	0.7150	0.8500	5.5289
urb	0.0935	0.5527	0.7190	0.6839	0.8423	1.0000	29.6121
des	0.0036	0.0289	0.0392	0.0410	0.0501	0.1013	40.6904
pib	5.8737	10.4045	12.9056	14.6271	16.2779	103.8509	57.5005



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

27

Modelagem

- Preditor considerado:

$$\log(\mu_i) = \beta_0 + \beta_1 \text{ fem} + \beta_2 \text{ mar} + \beta_3 \text{ kid5} + \beta_4 \text{ phd} + \beta_5 \text{ ment.}$$

- Modelos considerados:
 - Poisson(μ_i);
 - Quase-Poisson(μ_i, σ^2);
 - Binomial Negativo(μ_i, ϕ).



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

28

Estimativas e Erro Padrão

Parâmetro	Poisson		Quase-Poisson		Binomial Negativo	
	Estimativa	EP	Estimativa	EP	Estimativa	EP
β_0	0.0800	0.0986	0.0800	0.1334	0.0397	0.1328
β_1	0.2246	0.0546	0.2246	0.0739	0.2164	0.0727
β_2	0.1552	0.0614	0.1552	0.0830	0.1505	0.0821
β_3	-0.1849	0.0401	-0.1849	0.0543	-0.1764	0.0531
β_4	0.0128	0.0264	0.0128	0.0357	0.0153	0.0360
β_5	0.0255	0.0020	0.0255	0.0027	0.0291	0.0035
σ^2, ϕ	1.0000	—	1.8290	—	0.4416	0.0530



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

29

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

Avaliação e Comparação entre Modelos

Modelo	G.I. Deviance	Pearson χ^2	ℓ	AIC	BIC	
Poisson	909	1634.3710	1662.5466	-642.0261	3314.1126	3343.1
Binomial Negativo	909	1004.2815	944.5494	-551.9281	3135.9167	3169.1
Quase-Poisson	909	1284.0522	1296.0522	—	—	—



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

30

Modelo considerado

$$Y_i \mid \mathbf{x}_i \sim \text{BN}(\mu_i, \phi)$$

$$\log(\mu_i) = \beta_0 + \beta_1 \text{ fem} + \beta_2 \text{ mar} + \beta_3 \text{ kid5} + \beta_4 \text{ ment.}$$



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

31

Ajuste modelo Binomial Negativo

Parâmetro	Estimativa	EP	LI-Wald	LS-Wald	S_W	$P(S_W > \chi^2)$
β_0	0.0867	0.0731	-0.0566	0.2299	1.41	0.2359
β_1	0.2167	0.0727	0.0742	0.3591	8.89	0.0029
β_2	0.1469	0.0817	-0.0131	0.3070	3.24	0.0720
β_3	-0.1768	0.0531	-0.2808	-0.0728	11.10	0.0009
β_4	0.0294	0.0034	0.0228	0.0360	75.98	<.0001
ϕ	0.4417	0.0530	0.3491	0.5587	—	—



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

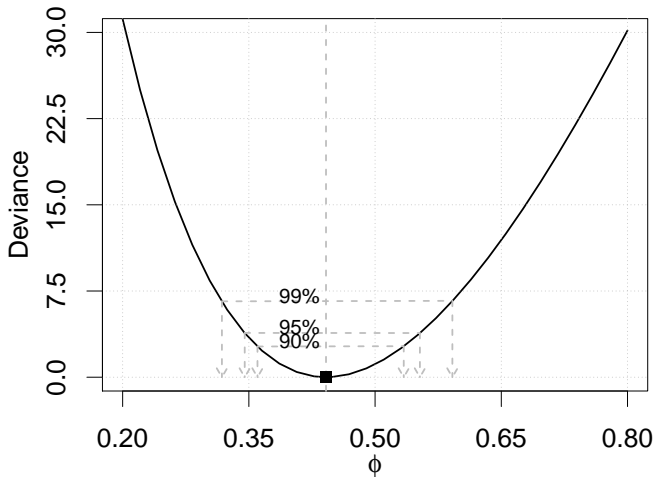
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

32





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

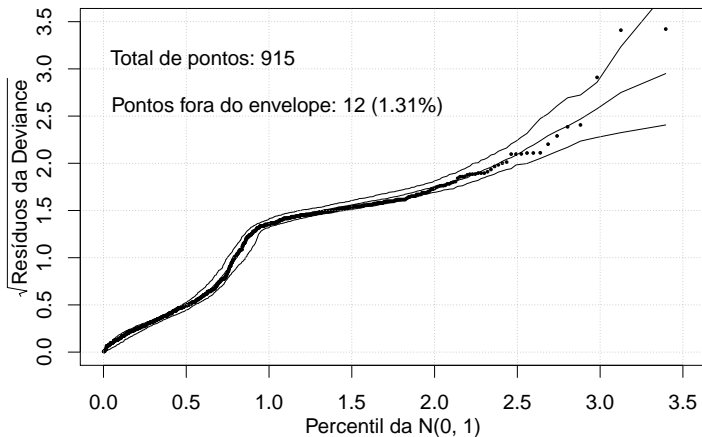
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

33





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

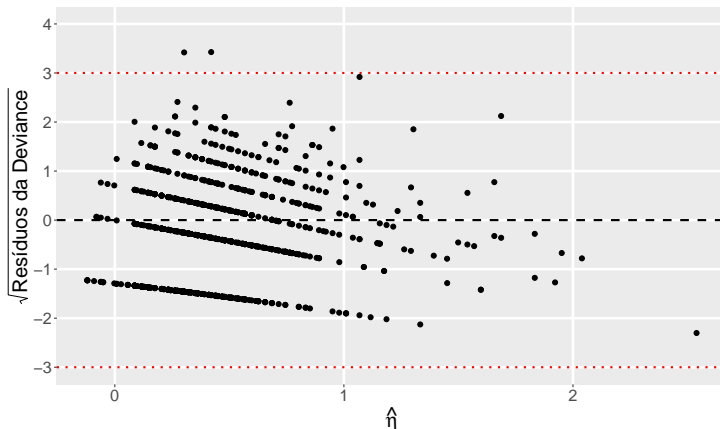
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

34





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

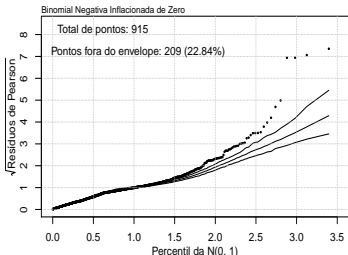
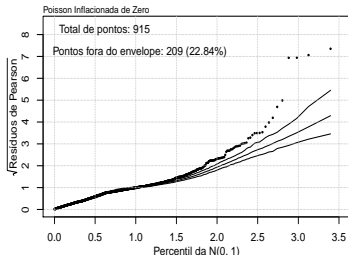
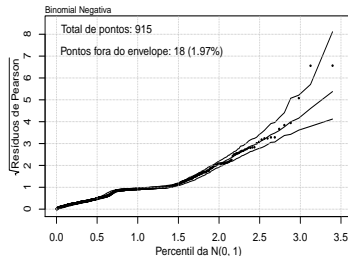
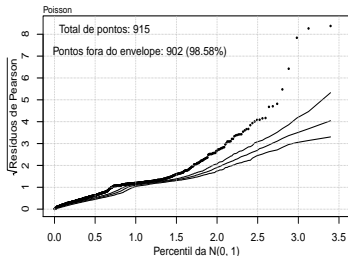
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

35





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

36

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

Discriminação entre modelos

Modelo	np	AIC	BIC	Vuong (valor-p)
BN	6	3134	3163	—
Poisson	5	3312	3336	4.3423 (< 0.001)
ZIP	10	3230	3278	2.4882 (0.0128)
ZINB	12	3122	3175	-6.5000 (0.6916)



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

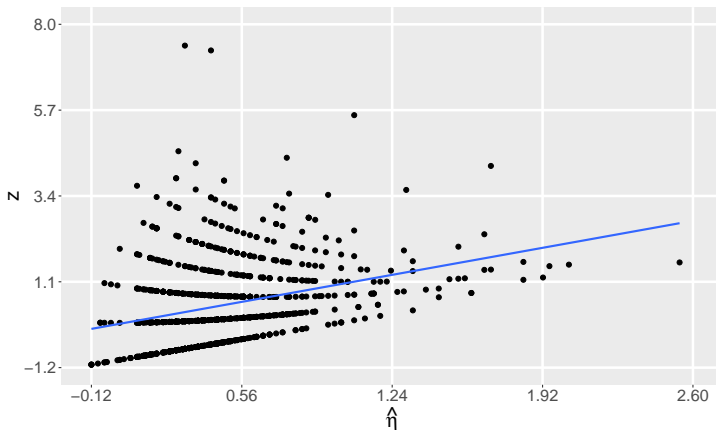
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

37



► Rejeita-se a hipótese nula do teste RESET.



Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

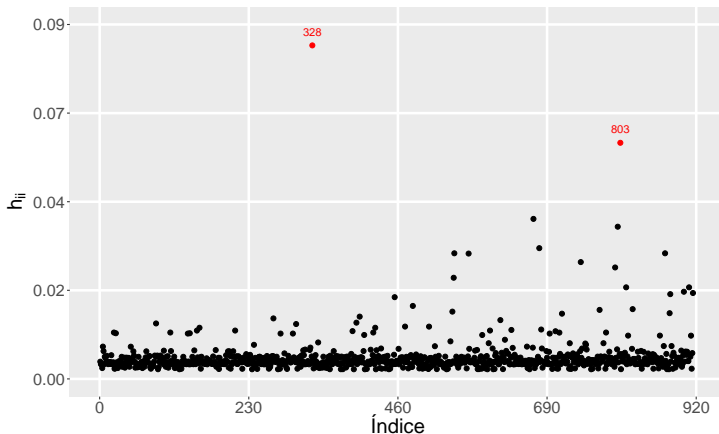
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

38





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

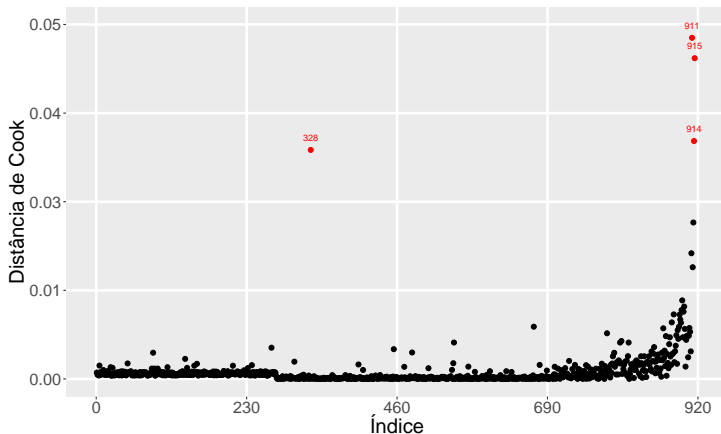
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

39





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

40

Considerações Finais

- ▶ É esperado que o número de publicações dos homens seja $\exp(0.2167) \simeq 1.2394$ maior do que das mulheres;
- ▶ Relação negativa entre o número médio de publicações e o número de filhos;
- ▶ Número de artigos publicados pelo orientador nos últimos três anos tem efeito positivo na produtividade do pesquisador;
- ▶ Modelos alternativos devem ser avaliados, uma vez que o modelo Binomial Negativo não apresentou um ajuste adequado.



Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Resultados e
Discussões

Análise do número de
publicações produzidas por
Ph.D. em Bioquímica

Análise da votação
presidencial de 2010

Referências

41

Modelagem

- Preditor considerado:

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \text{pt2006} + \beta_2 \text{gini} + \beta_3 \text{idhm_e} + \beta_4 \text{idhm_l} \\ + \beta_5 \text{idhm_r} + \beta_6 \text{urb} + \beta_7 \text{des} + \beta_8 \text{pib}$$

- Modelos considerados:

- $\text{Beta}(\mu_i, \phi)$;
- $\text{Simplex}(\mu_i, \phi^2)$



Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Resultados e
Discussões

Análise do número de
publicações produzidas por
Ph.D. em Bioquímica

Análise da votação
presidencial de 2010

Referências

42

Estimativas e Erro Padrão (EP)

Parâmetro	Beta		Simplex	
	Estimativa	EP	Estimativa	EP
β_0	-1.1352	0.5211	-1.1305	0.5161
β_1 (pt2006)	3.0309	0.1255	3.0343	0.1205
β_2 (gini)	-0.7275	0.2634	-0.7397	0.2632
β_3 (idhm_e)	-0.9978	0.3041	-0.9828	0.2998
β_4 (idhm_l)	0.0316	0.6582	0.0597	0.6556
β_5 (idhm_r)	1.8597	0.6001	1.8277	0.5966
β_6 (urb)	-0.7461	0.0933	-0.7621	0.0922
β_7 (des)	-2.4220	0.8331	-2.3907	0.8244
β_8 (pib)	-0.0008	0.0015	-0.0009	0.0015
ϕ	78.5775	5.5570	0.4719	0.0168



Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

43

Discriminação ente Modelos

Modelo	$-2 \hat{\ell}$	AIC	BIC	Vuong
Beta	-1169.4060	-1149.4060	-1109.6172	—
Simplex	-1171.3461	-1151.3461	-1111.5573	-0.8206 (0.7940)



Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

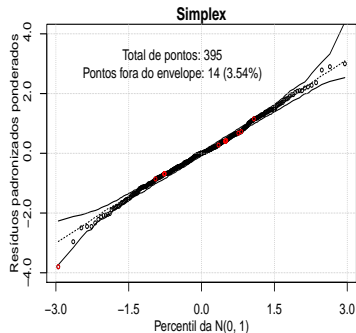
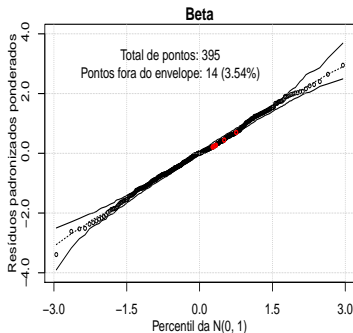
Resultados e
Discussões

Análise do número de
publicações produzidas por
Ph.D. em Bioquímica

Análise da votação
presidencial de 2010

Referências

44





Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Resultados e
Discussões

Análise do número de
publicações produzidas por
Ph.D. em Bioquímica

Análise da votação
presidencial de 2010

Referências

45

Modelo considerado

$$Y_i | \mathbf{x}_i \sim \text{Simplex}(\mu_i, \phi)$$

$$\begin{aligned} \text{logit}(\mu_i) = & \beta_0 + \beta_1 \text{pt2006} + \beta_2 \text{gini} + \beta_3 \text{idhm_e} + \beta_4 \text{idhm_r} \\ & + \beta_5 \text{urb} + \beta_6 \text{des} \end{aligned}$$



Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

46

Ajuste Modelo Simplex

Parâmetro	Estimativa	EP	LI	LS	t	$P(T > t)$
β_0	-1.0463	0.2685	-1.5725	-0.5201	-3.8974	0.0001
β_1 (pt2006)	3.0302	0.1202	2.7947	3.2658	25.2130	<0.000
β_2 (gini)	-0.7283	0.2628	-1.2434	-0.2131	-2.7708	0.0056
β_3 (idhm_e)	-0.9832	0.2984	-1.5681	-0.3984	-3.2951	0.0010
β_4 (idhm_r)	1.7495	0.5684	0.6355	2.8635	3.0780	0.0021
β_5 (urb)	-0.7558	0.0915	-0.9352	-0.5765	-8.2599	<0.000
β_6 (des)	-2.4523	0.8190	-4.0575	-0.8471	-2.9943	0.0028
ϕ	0.4722	0.0168	0.4392	0.5051	—	—

► Não rejeita-se a hipótese nula do teste RESET.



Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

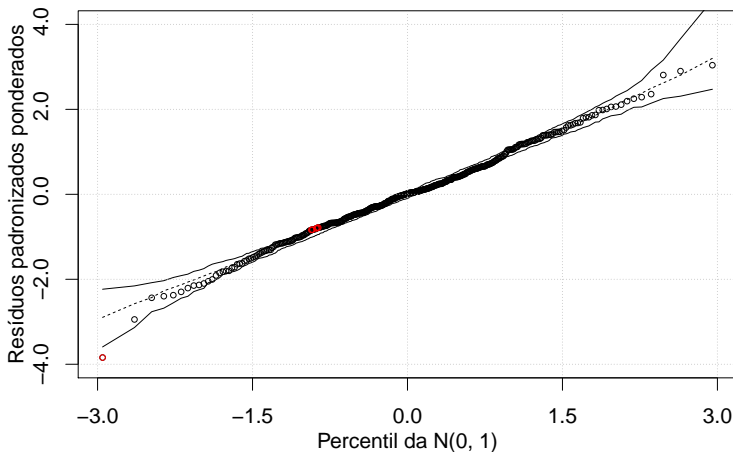
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

47





Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

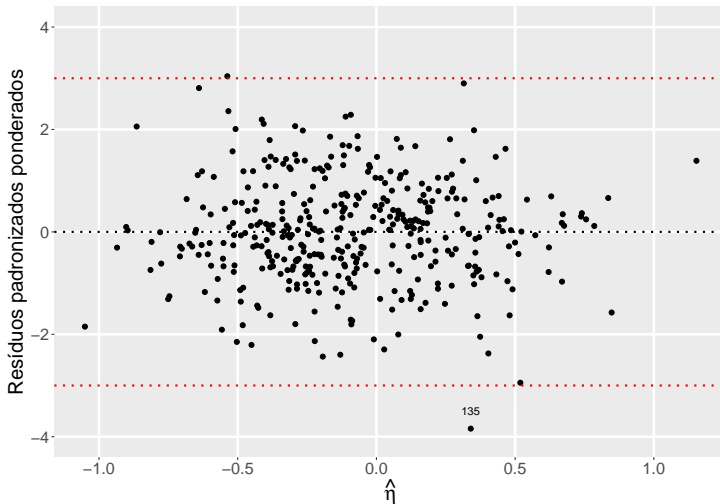
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

48





Análise da votação presidencial de 2010

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

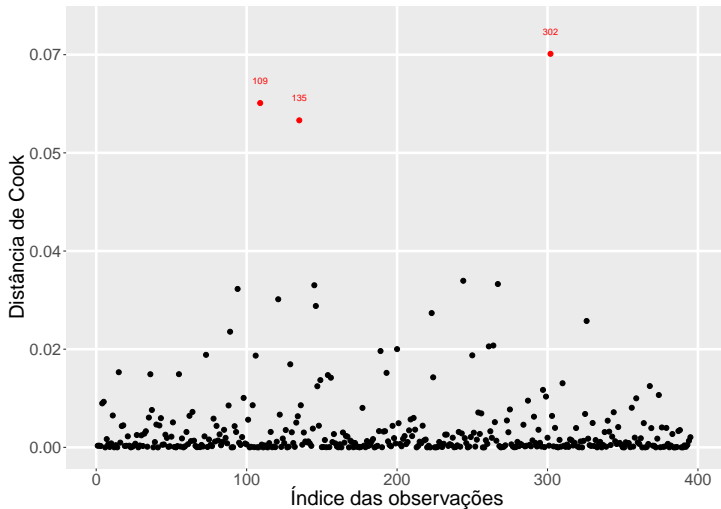
Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

49





Número de publicações produzidas por Ph.D. em Bioquímica

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de Regressão

Materiais

Resultados e Discussões

Análise do número de publicações produzidas por Ph.D. em Bioquímica

Análise da votação presidencial de 2010

Referências

50

Considerações Finais

- ▶ Ambos os modelos Beta e Simplex apresentaram ajuste satisfatório;
- ▶ Conforme critérios discutidos escolheu-se pelo modelo Simplex;
- ▶ As covariáveis `gini`, `idhm_e`, `urb` e `des` exercem efeito negativo sobre a proporção de votos;
- ▶ Tiveram uma influencia positiva na proporção de votos da Dilma as covariáveis `pt2006` e `idhm_r`.



Referências

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Resultados e
Discussões

Referências

51

[1] BARNDORFF-NIELSEN, O.; JØRGENSEN, B. Some parametric models on the Simplex. *Journal of Multivariate Analysis*, v. 39, n. 1, p. 106–116, 1991.

[2] FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, v. 31, n. 7, p. 799–815, 2004.

[3] FURRIEL, O. W. *Determinantes do voto à presidência: análise espacial das eleições gerais no Brasil no período de 1994 a 2014*. 2017. PIC – Universidade Estadual de Maringá.

[4] LONG, J. S. The origins of sex differences in science. *Social Forces*, Oxford University Press, v. 68, n. 4, p. 1297–1316, 1990.



Referências

Modelos de Regressão

André F. B. Menezes

Introdução

Modelos de
Regressão

Materiais

Resultados e
Discussões

Referências

52

[6] MIYASHIRO, E. S. *Modelos de regressão Beta e Simplex para a análise de proporções*. Dissertação (Mestrado) — Universidade de São Paulo - USP, 2008.

[7] NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, [Royal Statistical Society, Wiley], v. 135, n. 3, p. 370–384, 1972.

[8] RAMSEY, J. B. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 31, n. 2, p. 350–371, 1969.

[9] VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, [Wiley, Econometric Society], v. 57, n. 2, p. 307–333, 1989.