

Abordagem Bayesiana na Proposição de um Modelo de Regressão para a Velocidade do Vento em Fortaleza-CE

Érika F. Morais, Sílvio F. A. Xavier Júnior, Diego V. de S. Ferreira[†]

Programa de Pós-graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco - UFRPE.

Resumo: Neste trabalho utilizamos a Análise de Regressão Linear Múltipla para modelar a velocidade do vento na cidade de Fortaleza-CE, com o objetivo de encontrar um modelo estatístico capaz de tornar possível a realização de previsões para o comportamento dos ventos. Para o estudo foi analisado o banco de dados fornecido pelo INMET. O modelo de regressão linear gerado apresentou um bom ajuste de acordo com o gráfico de resíduos, e as variáveis explicativas deste modelo apresentaram uma correlação linear moderada com a variável de interesse, o que pode ser confirmado com o coeficiente de determinação encontrado que mostra que 53% da variância da velocidade do vento é explicada pela variação da temperatura e da umidade. Além de que o método bayesiano, também indicou que o ajuste em questão é adequado.

Palavras-chave: Análise de Regressão, Velocidade do vento, Distribuição a priori

Abstract: In this paper we use the Multiple Linear Regression Analysis to model the wind speed in the city of Fortaleza, with the goal of finding a statistical model able to make possible the realization of predictions for the behavior of the winds. For the study was analyzed the database provided by INMET. The linear regression model showed a good fit generated according to the residual graph, and the explanatory variables of the model showed a moderate linear correlation with the variable of interest, which can be confirmed with the determination coefficient found showing that 53 % of the variance of the wind speed is explained by variations in temperature and humidity. In addition to the Bayesian method, also indicated that the adjustment in question is appropriate.

Keywords: Regression Analysis, Wind Speed, Prior Distribution

Introdução

Uma melhor compreensão da variabilidade temporal e espacial da velocidade e direção do vento é fundamental para a modelagem de diferentes fenômenos a ele relacionados, incluindo a erosão do solo (JONSSON, 1992), a dispersão do pólen e sementes (FRIEDMAN; BARRET, 2009; JONGEJANS; TELENIUS, 2004), a geração de energia (PIMENTA et al, 2008), etc. Por isso, neste trabalho analisaremos a velocidade dos ventos por meio da Análise de Regressão Linear Múltipla, com o objetivo de encontrar um modelo matemático capaz de descrever de forma satisfatória o real comportamento da velocidade dos ventos em estudo, à fim de que se possam fazer possíveis previsões para tal fenômeno.

Os métodos Bayesianos tem sido aplicados com sucesso nos últimos anos no ajuste de modelos de regressão, como mostram Varona et al. (1997), Blasco et al.(2003), Simões (1981), entre outros. Em uma abordagem Bayesiana, tentaremos encontrar uma distribuição para os parâmetros da regressão que represente probabilisticamente o conhecimento que se tem sobre eles antes da realização do experimento.

A cidade escolhida para este estudo foi Fortaleza-CE, pois os ventos nesta localidade chegam a atingir em média de 12 a 14 Km/h, podendo chegar até 40 Km/h durante os meses mais frios do ano, de acordo com o *site* da Secretaria da Ciência, Tecnologia e Educação Superior do Ceará (SECITECE, 2011). Além de que o estado apresenta um grande potencial em energia eólica.

[†]Autor correspondente: dvsf@ibest.com.br.

Material e métodos

Os dados utilizados neste trabalho são de uma estação meteorológica automática, sobre administração do Instituto Nacional de Meteorologia (INMET, 2012), referentes a cidade de Fortaleza. As 261 observações foram medidas no período de 1 de novembro de 2011 a 26 de janeiro de 2012, colhidas 3 vezes por dia, às 0h, às 12h e às 18h pelo horário oficial de Brasília.

Definição do modelo de regressão

A Análise de Regressão Linear aborda os efeitos lineares produzidos por um conjunto de variáveis. Este estudo é feito através de uma função matemática que pode ser representada por:

$$Y = f(X_1, X_2, \dots, X_k),$$

onde Y é a variável resposta, e os X_j , ($j = 1, \dots, k$) são as variáveis explanatórias. O objetivo da Análise de Regressão é exatamente conhecer esta função matemática para então saber qual é o comportamento da variável resposta “explicado” pela(s) variável(is) explanatória(s), bem como os efeitos que são produzidos entre elas, e assim fazer possíveis previsões para a variável de interesse principal.

As variáveis observadas foram: A velocidade (Y) dos ventos (m/s), a temperatura ($^{\circ}\text{C}$), a umidade (%), a pressão (hPa) e a nebulosidade (Décimos). A Tabela 1 mostra a correlação linear calculada entre as variáveis. E para compor o modelo de regressão, escolhemos apenas as variáveis que apresentam correlação linear, entre as variáveis e a velocidade dos ventos, superior a 30%, neste caso, nosso modelo será composto apenas pelas variáveis Velocidade dos ventos (Y), Temperatura(X_1) e Umidade (X_2).

Tabela 1: Correlação linear entre as variáveis

Correlação linear	
Variáveis	Velocidade dos ventos
Temperatura	0,6164
Umidade	-0,3948
Pressão	0,2315
Nebulosidade	0,1752

Resultados e discussões

Com uso do *Software R* (R CORE TEAM, 2012) encontraremos as estimativas para os parâmetros da regressão, que serão dados pela matriz

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

onde cada elemento corresponderá aos valores dos parâmetros de acordo com as variáveis explanatórias. Dessa forma, calculando a matriz $\hat{\beta}$ teremos:

$$\hat{\beta} = \begin{bmatrix} -0,0799 \\ 0,2607 \\ -0,0548 \end{bmatrix}$$

Em seguida, utilizamos a Análise da Variância para testar se existe ou não regressão linear entre as variáveis, como mostra a Tabela 2.

Tabela 2: Análise de variância da regressão para o modelo estimado – ANOVA.

Varição	Grau de lib.	Soma de quad.	Quadrados médios	F
Regressão	2	208,1109	104,0554	148,6566
Resíduo	259	181,2927	0,6999	
Total	261	389,4035	-	

Desse modo, com um nível de significância de 5% e de acordo com a análise da variância, o teste F mostra que existem evidências a favor da hipótese de que o modelo estimado parece apresentar de fato uma regressão linear com pelo menos um par de parâmetros que diferem estatisticamente de zero e entre si. Essa conclusão pode ser reafirmada com o P-valor do teste, que apresentou valor estatisticamente pequeno ($2,2 \times 10^{-16}$). O modelo de regressão estimado será dado por:

$$\hat{y}_i = -0,0799 + 0,2607x_{1i} - 0,0548x_{2i},$$

com i representando o número de observações das variáveis em estudo.

A Figura 1 mostra o ajuste do modelo por meio da análise dos resíduos e a Figura 2 mostra o gráfico do QQplot. Aparentemente, dada a aleatoriedade dos resíduos no gráfico, o modelo parece estar bem ajustado.

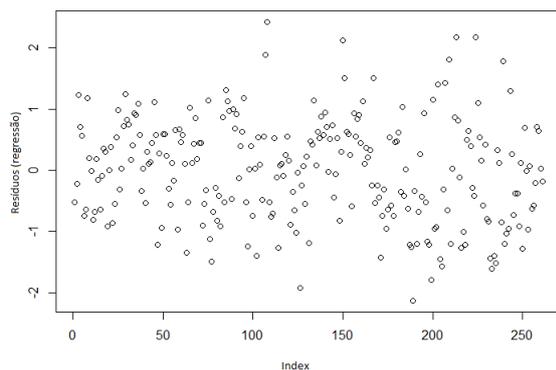


Figura 1: Gráfico de resíduos do modelo.

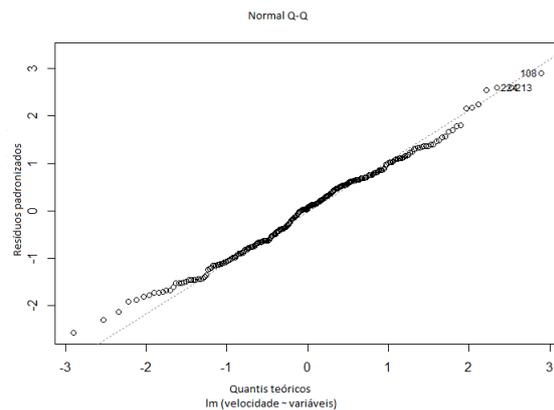


Figura 2: Gráfico QQ plot

O coeficiente de determinação, dado pela estatística $r^2 = \frac{SQ_{reg}}{SQ_{tot}} = 0,5344$, indica a proporção da variância da velocidade dos ventos explicada pela variabilidade das variáveis explicativas do modelo de regressão estimado.

Abordagem Bayesiana

Especificamos a distribuição condicional de y_i dado um vetor de predição \mathbf{X}_i de dimensão $(1 \times k)$ (k é o número de parâmetros da regressão) considerando o modelo de regressão em estudo

$$y_i = \mathbf{X}_i' \boldsymbol{\beta} + \varepsilon_i,$$

em que $\varepsilon_i \sim N(0, \sigma^2)$. Neste caso, teremos a seguinte Função de Verossimilhança para a regressão:

$$\rho(y|x, \boldsymbol{\beta}, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\}.$$

A distribuição a priori hierárquica:

$$\rho(\boldsymbol{\beta}, \sigma^2) = \rho(\sigma^2) \rho(\boldsymbol{\beta} | \sigma^2).$$

Então, finalmente podemos sugerir que $\rho(\sigma^2)$ é uma Distribuição a Priori para σ^2 , onde $\rho(\sigma^2) \sim$ Inversa-Gamma($\sigma^2; a_0, b_0$), com $a_0 = \frac{\nu_0}{2}$ e $b_0 = \frac{\nu_0 s_0^2}{2}$. Assim,

$$\rho(\sigma^2) \propto (\sigma^2)^{-\frac{\nu_0}{2}+1} \exp\left(-\frac{\nu_0 s_0^2}{2\sigma^2}\right).$$

Distribuição da Posteriori

Com a priori agora especificada, a distribuição a posteriori pode ser expressa como

$$\begin{aligned} \rho(\beta, \sigma^2 | Y, X) &\propto \rho(Y|X, \beta, \sigma^2) \rho(\beta | \sigma^2) \rho(\sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} (Y - X\beta)^T (Y - X\beta)\right) \\ &\times (\sigma^2)^{-K/2} \exp\left(-\frac{1}{2\sigma^2} (\beta - \mu_0)^T \Lambda (\beta - \mu_0)\right) \times (\sigma^2)^{(a_0+1)} \exp\left(-\frac{b_0}{2\sigma^2}\right) \end{aligned}$$

A expressão acima pode ser rescrita de maneira que, a média a posteriori μ_n do vetor de parâmetro β pode ser expressa em termos do EMQ de $\hat{\beta}$ e a média da priori μ_0 com a 'força' da priori indicada pela matriz de precisão da priori Λ_0 , onde $\mu_n = (X^T X + \Lambda_0)^{-1} (X^T X \hat{\beta} + \Lambda_0 \mu_0)$

A Figura 3 mostra o comportamento das Posteriors para os parâmetros da regressão, bem como uma série preditiva com 1000 iterações. E a Tabela 3 a seguir mostra os Intervalos Bayesianos com nível de credibilidade de 95% e as médias a posteriori empíricas para os parâmetros da regressão.

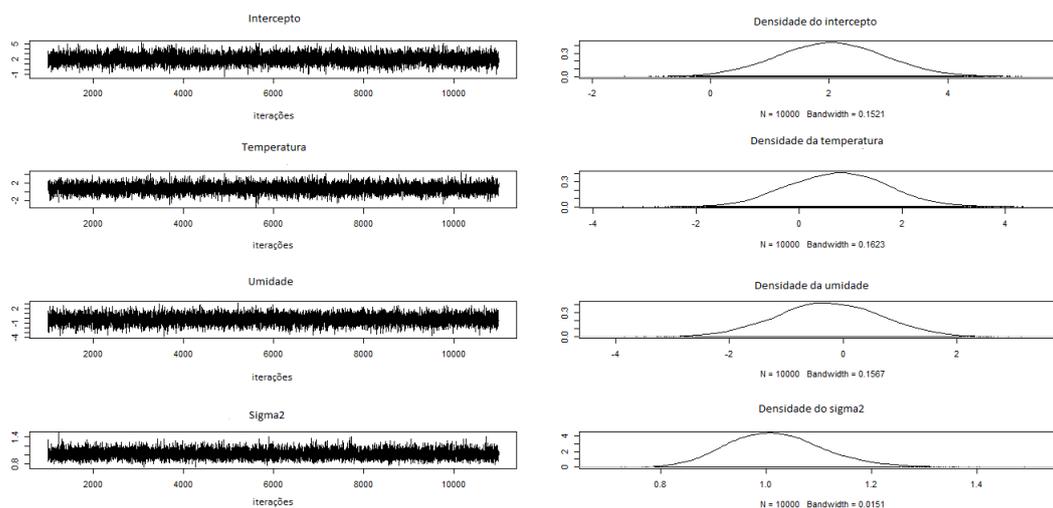


Figura 3: Série de previsão posteriori com 1000 iterações (a esquerda) e Distribuições a Posteriori (a direita)

Tabela 3: Intervalo Bayesiano e média a Posteriori dos parâmetros

Parâmetro da regressão	Intervalo com 95% de credibilidade	Média Posteriori empírica
Intercepto	[0,2835; 3,825]	2,0384
Temperatura	[-1,1272; 2,652]	0,7483
Umidade	[-2,1067; 1,557]	-0,2606
σ^2	[0,8548; 1,216]	1,0186

Conclusões

O modelo de regressão linear foi gerado de forma satisfatória e apresentou um bom ajuste de acordo com o gráfico de resíduos. Este gráfico ilustrou o comportamento dos resíduos e apresentou uma boa

aleatoriedade entre os eles, portanto o ajuste do modelo parece estar adequado. Entretanto, as variáveis que formaram o modelo apresentaram uma correlação linear moderada com a variável de principal interesse, o que pode ser confirmado com o coeficiente de determinação encontrado que mostra que 53% da variância da velocidade dos ventos é explicada pela variação da temperatura e da umidade. Além de que o método bayesiano, também indicou que o ajuste em questão é adequado.

Para estudos futuros, sugerimos a técnica de Anamorfose. Isto, provavelmente, faria com que o modelo fosse melhor ajustado ocasionando previsões mais satisfatórias.

Referências

BLASCO, A.; PILES, M.; VARONA, L. Bayesian Analysis of the effect of selection for growth rate on growth curves in rabbits *Genet. Sel. Evol.*, Valencia, v. 35, p. 21 - 41, 2003.

FRIEDMAN, J.; BARRET, S.C.H. Wind of change: new insights on the ecology and evolution of pollination and mating in wind pollinated plants. *Annals of Botany*, p. 1515-1527, 2009.

INMET - Instituto Nacional de Meteorologia. *Monitoramento das estações convencionais*. Disponível em: <http://www.inmet.gov.br/sim/sonabra/convencionais.php>. Acesso em: 26 de janeiro de 2012.

JONGEJANS, E.; TELENIUS, A. Field experiments on seed dispersal by wind in ten umbelliferous species (Apiaceae). *Plant Ecology*, p. 67-78, 2004.

JONSSON, P. Wind erosion on sugar beet fields in Scania, souther Sweden. *Agricultural and Forest Meteorology*, p. 141-157, 1992.

PIMENTA, F.; KEMPTON, W.; GARVINE, R. Combining meteorological stations and satellite data to evaluate offshore wind power resource of Southeastern Brazil. *Renewable Energy*, p.2375-2387, 2008.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2012. Disponível em: <http://www.R-project.org/>.

SECITECE - Secretaria da Ciência, Tecnologia e Educação Superior do Ceará. Disponível em: <http://www.sct.ce.gov.br/index.php/noticias/57469>. Acesso em: 31 de março de 2011.

SIMÕES, N. R. *Estimadores de Bayes: aplicação ao modelo de regressão linear simples*. Dissertação de Mestrado em Estatística e Experimentação Agronômica - Universidade de São Paulo, Piracicaba, SP., 138 f., 1981.

VARONA, L.; MORENO, C.; GARCIA CORTÉS, L. A.; ALTARRIBA, J. Multiple trait genetic analysis of underlying biological variables of production functions. *Livestock Production Science*, Amsterdam, v. 47, p. 201 - 209, 1997.