Revista da Estatística UFOP, Vol III(3), 2014, ISSN 2237-8111
Edição especial: 59ª Reunião Anual da Regional Brasileira da Sociedade Internacional de Biometria - RBRAS
Trabalhos apresentados na versão poster.

Estudo da temperatura do ar a partir do modelo funcional de regressão

Eudmar Paiva de Almeida [1]
Danielle da Silva Pompeu [1]
Silvia dos Santos de Almeida [2]
Vanderly Janeiro [3]

1 Introdução

A qualidade principal de um instrumento de medição é a de medir com o mínimo erro, isto é, um instrumento de medição de boa qualidade deve ser capaz de apresentar resultados com pequenos erros de medição. Entretanto, por melhores que sejam as características de um instrumento, este sempre poderá apresentar erros. Muitos são os setores da ciência onde é necessária uma boa precisão, como é o caso da Meteorologia (que tem como foco o estudo dos processos atmosféricos e a previsão do tempo). Neste sentido, este trabalho apresenta o método de estimação no Modelo Funcional para a Temperatura do Ar (Y) em função da Radiação Global (X), onde as duas variáveis envolvidas possuem erros de medição.

2 Material e métodos

2.1 Dados referentes ao trabalho

Foram obtidos junto a Faculdade de Meteorologia da Universidade Federal do Pará (UFPA), e se referem a 80 medidas de Temperatura do ar e Radiação Global que foram coletadas sobre uma cultura de soja na região de Paragominas, região nordeste do Estado do Pará.

2.2 Modelo de regressão com erros nas variáveis (MEV)

Para Linna e Woodall (2001), erros de medição significantes frequentemente existem em diversas aplicações. Portanto, sabe-se que fisicamente, toda medição é passível de erros, por isso, este é o argumento mais forte para acreditar que estão presentes, quer sejam provocados pelo instrumento de medição, pelo operador ou por fatores externos (sol, chuva, vento, entre outros). Os modelos com erros nas variáveis que na verdade é uma generalização dos modelos de regressão padrão (clássico), são adequados para estes casos.

[1]PBE - UEM. e-mail: *eudmar@gmail.com, danielle11silva@gmail.com*
[2]FAEST - UFPA. e-mail: *salmeida@ufpa.br*
[3]DES - UEM. e-mail: *vjaneiro@uem.br*

Suponha um modelo de regressão, onde as variáveis U_i e Y_i são duas quantidades não observáveis e estão relacionadas a partir de uma equação linear como apresentada por Moran (1971) que pode ser visto como

$$y_i = \alpha + \beta\, U_i,\ i = 1, ..., n \tag{1}$$

em que α e β são parâmetros desconhecidos. Porém, nem y e nem U são observados diretamente, de modo que os valores observados são X_i e Y_i, em que $Y_i = y_i + \varepsilon_i$ e

$$X_i = U_i + \delta_i \tag{2}$$

onde $i = 1, ..., n$ sendo os ε_i e σ_i erros independentes e identicamente distribuídos, com média zero e variâncias finitas dada por σ^2_ε e σ^2_δ , respectivamente. Assim, o modelo dado em (1), pode ser reescritos como $y_i = \alpha + \beta\, U_i + \varepsilon_i$ para $i = 1, ..., n$.

É importante ressaltar que no modelo com erro nas variáveis, as variáveis $U_i\ 's$ podem se apresentar de maneiras diferentes, Fuller (1987), afirma que quando $U_i\ 's$ são considerados constantes tem-se o modelo funcional, e com isto, o número de parâmetros cresce de acordo com o tamanho da amostra, fazendo com que no modelo existam $n+4$ parâmetros ao todo, sendo eles: α , β , σ^2_ε , σ^2_δ e U_i , com $i = 1, ..., n$. Já quando $U_i\ 's$ são considerados variáveis aleatórias, tem-se o modelo estrutural μ_μ e variância σ^2_μ, neste caso, têm-se seis parâmetros no modelo, sendo eles α , β , μ_μ , σ^2_ε , σ^2_δ e σ^2_μ.

Neste trabalho, o modelo com erros nas variáveis utilizado é o modelo funcional. Almeida (2003) destaca que a estimação dos parâmetros da regressão no modelo funcional só é possível se forem feitas suposições adicionais aos parâmetros.

Seja o modelo funcional formalmente definido por Almeida (1999 e 2003) dado pelas equações (1) e (2) sendo que U_i são as constantes fixadas (parâmetros incidentais), e ε_i , δ_i os erros de medição considerados normais, independentes e identicamente distribuídos, ambos com média zero e variâncias constantes dados por σ^2_ε e σ^2_δ , respectivamente, sendo a equação de regressão estimada no modelo com erro nas variáveis dada pela esperança da equação de regressão populacional de Y, isto é, por

$$\widehat{Y} = \widehat{\alpha} + \widehat{\beta}\, \widehat{U}_i;\ i = 1, ..., n \tag{3}$$

sendo os estimadores da regressão, $\widehat{\alpha}$ e $\widehat{\beta}$ dados por

$$\widehat{\beta} = \frac{S_{XY}}{S_{XX} - \sigma^2_\delta} \tag{4}$$

e $\widehat{\alpha} = \overline{Y} - \widehat{\beta}\, \overline{X}$ em que, $\overline{X}$ e $\overline{Y}$ são suas respectivas médias, S_{XX} e S_{XY} são

$$S_{XX} = \frac{1}{n-1} \sum (X_i - \overline{X})^2 \tag{5}$$

$$S_{XY} = \frac{1}{n-1}\sum(X_i - \overline{X})(Y_i - \overline{Y}). \quad (6)$$

O estimador não viesado da variância do erro do modelo funcional é igual a variância do modelo, que é dado por

$$\widehat{\sigma_Y}^2 = \frac{1}{n-2}\sum(Y_i - \widehat{Y_i}). \quad (7)$$

Fuller (1987), Almeida (2003) e Carvalho Jr. (2007) mostram que o melhor estimador linear não viesado de U_i é dado por $\widehat{U_i} = \frac{\sigma_\delta^2\widehat{\beta}(Y_i - \widehat{\alpha}) + \sigma_\varepsilon^2}{\widehat{\beta}^2\widehat{\sigma}_\delta^2 + \widehat{\sigma}_\varepsilon^2}$ no qual $\widehat{\sigma}_\varepsilon^2 = S_{YY} - \widehat{\beta}_{XY}$. O valor de σ_δ^2 é determinado pelo conhecimento da variabilidade de processos anteriores. Geralmente, sua fixação não é considerado um problema, pois é comum ter conhecimento deste valor, a partir da observação de processos passados, de especificações de equipamento de medição ou ainda da própria experiência. A partir desses resultados, é possível obter para o modelo funcional, as variâncias e seus erros padrão da distribuição aproximada dos estimadores de α e β dado por $EP(\widehat{\alpha}) = \sqrt{\frac{n\overline{X}^2\widehat{Var}(\widehat{\beta}) + \widehat{\sigma}_Y^2}{n}}$ e $EP(\widehat{\beta}) = \sqrt{\frac{S_{XX}\widehat{\sigma}_Y^2 + \widehat{\beta}^2\sigma_\delta^2}{l^2(n-1)}}$, no qual o valor de $\overline{X}$,$\widehat{\beta}$, S_{XX}, e $\widehat{\sigma}_Y^2$ são dados pela média de X e pelas equações (4), (5) e (7), respectivamente. E o valor de l pode ser obtido por $l = \frac{1}{n-1}\sum(U_i - \overline{U})^2$.

Quando se estuda duas ou mais variáveis, cuja relação é expressa pela Equação (1), o desvio padrão de Y_i (reta de regressão) passa a ser o afastamento médio e mínimo existente entre cada ponto observado e a reta estimada ($\widehat{Y_i}$), e é chamado de Erro Padrão da reta de Regressão, que no caso do modelo funcional, um estimador pode ser dado pela raiz quadrada de (7), resultando em $\widehat{\sigma}_Y^2 = EP(Y_i) = \sqrt{\frac{1}{n-2}\sum(Y_i - \widehat{Y_i})}$, onde $\widehat{Y_i}$ é dada pela equação (3).

3 Resultados e discussão

Inicialmente buscam-se confirmar a suposição de que as variáveis apresentam uma boa correlação linear, para tal, foi calculado o coeficiente de correlação de Pearson (r = 0,606) e o valor do nível descritivo (p-valor = 0,000), confirmando assim a existência de uma boa correlação linear. O modelo de regressão, é necessário verificar se a variável resposta, que neste caso é a Temperatura do Ar, segue uma distribuição normal. Para tanto foi feito o teste de aderência de Kolmogorov-Smirnov para a Temperatura do Ar, onde pode-se verificar pelo (p-valor ¿0,150), que a variável resposta segue uma distribuição normal. Como mencionado anteriormente, o valor de σ_δ^2 pode ser determinado pelo conhecimento da variabilidade de processos anteriores, que está neste trabalho é de $\sigma_\delta^2 = 1$. A partir desses resultados, é possível obter para o modelo funcional, as variâncias e seus erros padrão da distribuição aproximada dos estimadores de α e β . Na Tabela 1 é apresentado os valores dos estimadores obtidos, seus respectivos erros-padrão e o modelo estimado.

A presença de erros de mensuração afeta a precisão dos estimadores, de forma à medida

Tabela 1: Estimativa Obtida por Meio do Modelo de Regressão Funcional para $\sigma^2_\delta = 1$.

σ^2_δ	$\widehat{\alpha}$	$\widehat{\beta}$	$EP(\widehat{\alpha})$	$EP(\widehat{\beta})$	$EP(Y)$	Modelo Estimado($\widehat{Y_i}$)
1	23,493678	0,008453	0,113459	0,000505	0,402309	23,493678+0,008453U_i

que cresce, ou seja, o erro associado a variável X cresce, a estimativa do erro padrão do modelo (Y) também cresce. Também destaca que em amostras sujeitas a erros a utilização do modelo clássico de regressão pode levar a problemas como: estimadores distintos da realidade (inferências erradas), pois quanto maior o erro associado a variável , refletido em , maior será o erro padrão dos seus estimadores. Aplicando o modelo encontrado para exemplificar uma estimação da Temperatura do Ar (em °C) em função da Radiação Global (para um valor de 191,1737223 Joules/m2), tem-se: $Temperatura\ do\ Ar = 23,493678 + 0,0084553 * 191,1737223 = 25,1096695°C$.

4 Conclusão

Este trabalho teve como objetivo apresentar um estudo prático da estimação dos parâmetros do Modelo (capaz de estimar a Temperatura do Ar a partir da Radiação Global) com Erros nas Variáveis (Funcional), tendo em vista a pouca divulgação deste tipo de modelagem, contribuindo desta forma para um melhor planejamento dos processos nas diversas áreas do conhecimento, que necessitem de modelagem estatística.

Referências

[1] ALMEIDA, S. S. **Calibração Absoluta Funcional Sem a Suposição de Normalidade**. Recife, 1999. 77p. Dissertação (Mestrado em Estatística). Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco.

[2] ALMEIDA, S. S. **Desenvolvimento de Gráficos de Controle Aplicados ao Modelo Funcional de Regressão**. Florianópolis, 2003. 116p. Tese (Doutorado em Engenharia de Produção). PPGEP, Universidade Federal de Santa Catarina.

[3] CARVALHO JR., J. G.; ALMEIDA, S. S.; RAMOS, E. M. L. S. **. Gráfico de Controle de Regressão Estrutural**. Tendência em Matemática Aplicada e Computacional, São Carlos, v.8, p. 361-367, 2007.

[4] CHENG, C. L. e VAN NESS, J.W. **On estimating linear relationships when both variables are subjects to erros**. Royal Statistical Society, v. 56, n°1, p. 167- 183, 1994.

[5] FONSECA, J. S. ; MARTINS, G. A. ; TOLEDO, G. L. **Estatística Aplicada**. 2. ed., São Paulo: Atlas, 1988.

[6] FULLER, W. A. **Measurement Error Models**. JOHN WILEY, New York, 1987.

[7] LINNA, K. W. e WOODALL, W. H. **Effect of Measurement Error on Shewhart Control Charts, Journal of Quality Technology**. v. 33, n° 2, p. 213-222, 2001.

[8] MORAN, P. A. P. **Estimating structural and functional relationships**. Journal of Multivariate Analysis, v. 1, p. 232-255, 1971.

[9] NETER, J. ; NACHTSHEIM, C. J. ; KUTNER, M. H. **Applied Linear Statistical Models**. 5.ed. Columbus: Editora McGraw-Hill/Irwin, 2005.