

Monitoramento da *Vochysia inundata Ducke* por meio de Gráfico de Controle de Regressão Funcional

Marcello Neiva de Mello, neivamarcello@gmail.com
Edson Marcos Leal Soares Ramos, edson@ufpa.br
Adrilayne dos Reis Araújo, adrilayne@ufpa.br
Sílvia dos Santos de Almeida, salmeida@ufpa.br
Wagner Rogério Ferreira Pinheiro, wagner2235@gmail.com

Universidade Federal do Pará

Resumo. O gráfico de controle de regressão funcional é uma ferramenta do controle estatístico da qualidade, bastante útil, porém, pouco difundida para monitorar processos onde estes admitem erro na variável independente (X), geralmente, oriundos da não calibração de aparelhos, ou até mesmo de erro humano. Desta forma, o objetivo da utilização desta ferramenta estatística é monitorar o comportamento das características da qualidade que apresentem relação estatística, para este caso, diâmetro e volume da árvore *Vochysia inundata Ducke*. Para modelagem a partir da análise de regressão funcional, fez-se necessário a aplicação da transformação de Box-Cox, pois o modelo inicial viola o pressuposto de homocedasticidade. A partir do monitoramento das características da qualidade da árvore *Vochysia inundata Ducke*, observou-se por meio do gráfico de controle de regressão funcional que o processo deve ser investigado, pois nota-se a existência de pontos amostrais acima do limite superior de controle (LSC), ou seja, mostrando árvores com valores acima do esperado, sendo possivelmente altas e de diâmetro um pouco menor, o que torna inviável o corte comercial da árvore.

Palavras-chave: Gráfico de Controle de Regressão Funcional, Árvore *Vochysia inundata Ducke*, Transformação de Box-Cox.

1. Introdução

A utilização de técnicas estatísticas no controle da qualidade de processos tornou-se indispensável para identificação e análise dos problemas. Dentre as principais ferramentas para monitoramento de processos, pode-se destacar o Gráfico de Controle de Regressão, apresentado por DiPaola (1945), para descrever o controle simultâneo de duas variáveis. Para Almeida (2003) o objetivo do gráfico de controle de regressão é controlar uma variação média em lugar de uma constante, onde a linha de regressão obtida a partir da equação de regressão clássica resume o relacionamento linear entre as duas variáveis de interesse e ainda cita que em torno desta linha de regressão é estabelecido um intervalo de confiança, dentro do qual se espera que o valor real da variável dependente esteja contido. Além dos modelos clássicos, existem modelos de regressão que admitem erros na variável independente, geralmente oriundos da não calibração de aparelhos, ou até mesmo de erro humano, estes modelos são conhecidos como Modelo de Regressão Funcional, Estrutural e Ultra-Estrutural.

Almeida (2003), a partir da combinação das técnicas estatísticas de análise de regressão com erro nas variáveis, mais especificamente análise de regressão funcional, combinada com a metodologia de gráficos de controle, desenvolveu uma ferramenta denominada gráfico de controle de regressão funcional. Porém, devido a sua recente descoberta o gráfico de controle de regressão funcional é pouco conhecido e, conseqüentemente, existem raras aplicações deste na literatura especializada.

Neste sentido, este trabalho visa construir e analisar o desempenho de gráficos de controle de regressão funcional em alguns processos oriundos de medições, mostrando os benefícios da utilização desta poderosa ferramenta estatística no controle e avaliação destes processos. Para tanto

utilizou-se os dados referentes ao Volume e Diâmetro de Árvores Quaruba, do Município de Bragança-PA, em Janeiro de 2009.

2. Metodologia

2.1 Modelo de Regressão Funcional

Fuller (1987) mostra que o modelo funcional admite os U_i 's, parâmetros incidentais como constantes, ou seja, valores que crescem junto com o tamanho da amostra. Os demais parâmetros da amostra são denominados de estruturais. Admiti-se $n + 4$ parâmetros ao todo neste modelo, sendo eles $\alpha, \beta, \sigma_\varepsilon^2, \sigma_\delta^2$ e U_i , com $i = 1, \dots, n$.

Considerando a suposição adicional de normalidade, o logaritmo (l) da correspondente função de verossimilhança e utilizando a notação de Kendall e Stuart (1979) tem-se,

$$-\frac{2}{n}l = \log \sigma_\delta^2 + \log \sigma_\varepsilon^2 + \frac{S_1}{\sigma_\delta^2} + \frac{S_2}{\sigma_\varepsilon^2}, \text{ onde } S_1 = \frac{1}{n} \sum_{i=1}^n (X_i - U_i)^2 \text{ e } S_2 = \frac{1}{n} \sum_{i=1}^n (Y_i - (\alpha + \beta U_i))^2.$$

Solari (1969) mostra que a estimação por verossimilhança só será possível se forem feitas suposições adicionais aos parâmetros. As suposições mais comuns adotadas para contornar o problema de estimação no modelo funcional são: (i) σ_ε^2 ou σ_δ^2 conhecido; (ii) σ_ε^2 e σ_δ^2 conhecidos; (iii) a razão $\sigma_\varepsilon^2 / \sigma_\delta^2$ conhecida ou (iv) o α conhecido.

2.2 Gráfico de Controle

Desde sua introdução por Shewhart, os gráficos de controle têm sido a ferramenta do CEQ mais utilizada para monitorar e manter o controle estatístico dos processos. Os gráficos de controle são aplicados para distinguir num processo as variações que não podem ser economicamente identificadas e corrigidas (causas aleatórias), daquelas que podem ser causas assinaláveis. Existem diversos tipos de gráficos de controle, dentre os muitos tipos pode-se destacar: Gráficos de Controle para Variáveis e Gráficos de Controle para Atributos.

Os gráficos de controle para variáveis são, certamente, os mais empregados no controle estatístico da qualidade. Especificamente, têm-se os gráficos desenvolvidos para monitorar o nível do processo, como o gráfico da média e os gráficos desenvolvidos para monitorar a dispersão do processo, por exemplo, o gráfico do desvio padrão. De uma maneira geral, o controle estatístico da qualidade não é simplesmente uma coleção de técnicas estatísticas, mas uma forma de pensar que é essencial para a melhoria contínua da qualidade.

2.3 Gráfico de Controle de Regressão Funcional

O gráfico de controle tem como objetivo controlar uma variação média em lugar de uma constante, em que a linha de regressão, obtida a partir da equação de regressão clássica, resume o relacionamento linear entre as duas variáveis de interesse. E, em torno desta linha de regressão, é estabelecido um intervalo de confiança, dentro do qual se espera que o valor real da variável dependente esteja contido com certa probabilidade.

Mitra (1988) destaca que o valor central do gráfico de controle de regressão é uma reta inclinada e as linhas dos limites de controle são paralelas à linha de regressão, além disso, os cálculos dos seus limites são um pouco mais complexos e mais demorados que os dos gráficos de controle convencionais.

Sejam X e Y variáveis sujeitas a erro de medição, então, o modelo de regressão funcional é expresso por,

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}U_i + \varepsilon_i,$$

onde, $U_i = X_i - \delta_i$ e $i = 1, \dots, n$. Os U_i são desconhecidos, então, são estimados a partir do método dos mínimos quadrados, por $\hat{U}_i = \frac{\sigma_\delta^2 \beta (Y_i - \hat{\alpha}) + \hat{\sigma}_\varepsilon^2}{\hat{\beta} \sigma_\delta^2 + \hat{\sigma}_\varepsilon^2}$, em que, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$; $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$;

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}); \quad S_{xx} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2; \quad S_{yy} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2; \quad \hat{\sigma}_\varepsilon^2 = S_{yy} - \hat{\beta} S_{xy};$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx} - \sigma_\delta^2} \text{ e } \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}.$$

Os erros de medição (ε_i , δ_i) seguem distribuição normal com médias zero e variâncias constantes σ_ε^2 e σ_δ^2 , respectivamente. Assim, *LSC*, *LC* e *LIC* são obtidos, respectivamente, por

$$\text{Limite Superior de Controle: } LSC = \hat{Y}_i + k\hat{\sigma},$$

$$\text{Linha Central da Regressão Estrutural: } LC = \hat{Y}_i = \hat{\alpha} + \hat{\beta}\hat{U}_i$$

e

$$\text{Limite Inferior de Controle: } LIC = \hat{Y}_i - k\hat{\sigma},$$

$$\text{em que a estimativa de } \sigma \text{ é obtida por } \hat{\sigma} = \hat{EP}(Y) = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y} - (X_i - \bar{X})\hat{\beta})^2}{n-2}}.$$

2.4 Transformação Box-Cox

Quando um dos pressupostos necessários a aplicação da técnica Análise de Regressão é violado, torna-se necessário a utilização de algum procedimento visando à correção desta falha. Por exemplo, quando a suposição de normalidade da variável explicativa (Y) não é satisfeita, pode-se utilizar algum tipo de transformação para tentar normalizá-la. De fato, transformações lineares são artifícios comumente utilizados para melhorar a adequação do modelo de regressão aos dados. A violação de pressupostos, como homocedasticidade, independência e normalidade dos resíduos comprometem sobremaneira a validação do modelo de regressão obtido. Neste sentido, a transformação de Box-Cox surge como uma das mais eficientes transformações lineares para se atingir as condições necessárias a validação do modelo estatístico.

De acordo com Box e Cox (1964), a transformação de Box-Cox faz parte da família de transformações paramétricas, na qual a variável dependente (Y) passa a ser $Y_i(\lambda) = \alpha + \beta X_i + \varepsilon_i$, sendo $Y_i(\lambda) = (Y^\lambda - 1)/\lambda$, em que Y é a variável observada, $Y(\lambda)$ é a variável transformada e λ é o parâmetro de transformação que maximiza a função.

3. Resultados

3.1 Árvores *Vochysia inundata* Ducke

Dentre as muitas espécies da floresta amazônica encontra-se a Quaruba, nome vulgar da *Vochysia inundata* Ducke, uma árvore existente nas matas de terra firme e várzea alta, não inundada, principalmente nos estados do Pará e Amapá e ainda, nas Guianas. Além disso, é utilizada principalmente em construção civil e naval, marcenaria, caixotaria, tábuas, ripas, forros, parte interior de móveis, etc.

Devido à importância da Quaruba, neste trabalho utilizam-se técnicas estatísticas para o monitoramento das características (variáveis), para analisar os padrões exigidos pelas empresas consumidoras desta variedade de madeira. A área de estudo localiza-se na comunidade Benjamin Constant, no município de Bragança, nordeste paraense em uma propriedade de um pequeno agricultor, onde foram mensuradas 31 indivíduos de *Vochysia inundata* Ducke em 1 hectare de floresta, de aproximadamente 40 anos.

3.2 Modelagem das Características Associadas às Árvores Quaruba

Para modelagem a partir da análise de regressão são necessários alguns pressupostos, como correlação e normalidade da variável Y . Foram obtidos 31 pares de dados das variáveis: diâmetro (X) e volume (Y) de árvores Quaruba, no município de Bragança, em Janeiro de 2009. Ambas, resultantes de processos de medição, sendo, portanto, sujeitas a erros. As variáveis diâmetro e volume apresentam forte correlação positiva, com o coeficiente de correlação de Pearson, $r = 0,969$, o que demonstra que um modelo linear pode ser adotado. Devido o modelo inicial ter apresentado a violação do pressuposto de homocedasticidade, fez-se necessário aplicar a transformação de Box-Cox à variável Y , possibilitando a modelagem a partir da análise de regressão. Para tanto, se faz necessário a estimação do parâmetro desconhecido λ . A partir do pacote estatístico MINITAB 14, pôde-se estimar o valor do parâmetro que maximiza a função, sendo neste caso $\lambda = 0,5$. O modelo obtido a partir dos novos valores calculados de Y , na transformação, é dado por

$$\hat{Y}_i(\lambda) = -0,0632 + 0,0249U_i, \quad (1)$$

3.4 Análise de Resíduos Após a Transformação Box-Cox Aplicada à Variável Y

i) Normalidade dos Resíduos Após a Transformação Box-Cox

Para verificação da normalidade dos resíduos realizou-se o teste de *Kolmogorov-Smirnov*, as hipóteses em teste são: H_0 : Os resíduos apresentam distribuição Normal *versus* H_1 : Os resíduos não apresentam distribuição Normal. Onde se obteve um nível descritivo $p = 0,150$ para a estatística $KS = 0,128$. Logo, não há evidências para rejeitar a hipótese nula, ou seja, os resíduos apresentam distribuição Normal.

ii) Independência e Homocedasticidade dos Resíduos

A Figura 2 apresenta o gráfico dos resíduos padronizados *versus* as observações ordenadas, obtidos a partir do Modelo (1) para as variáveis volume e diâmetro. Nela, pode-se observar que os resíduos apresentam aleatoriedade em torno do zero, admitindo-se independência. A Figura 2 apresenta o gráfico de resíduos padronizados *versus* os valores ajustados, obtidos a partir do Modelo (1) para a variável volume. Nela, pode-se observar a validação do pressuposto de homocedasticidade, assumindo assim que os resíduos mostram-se aleatórios e apresentam variâncias constantes.

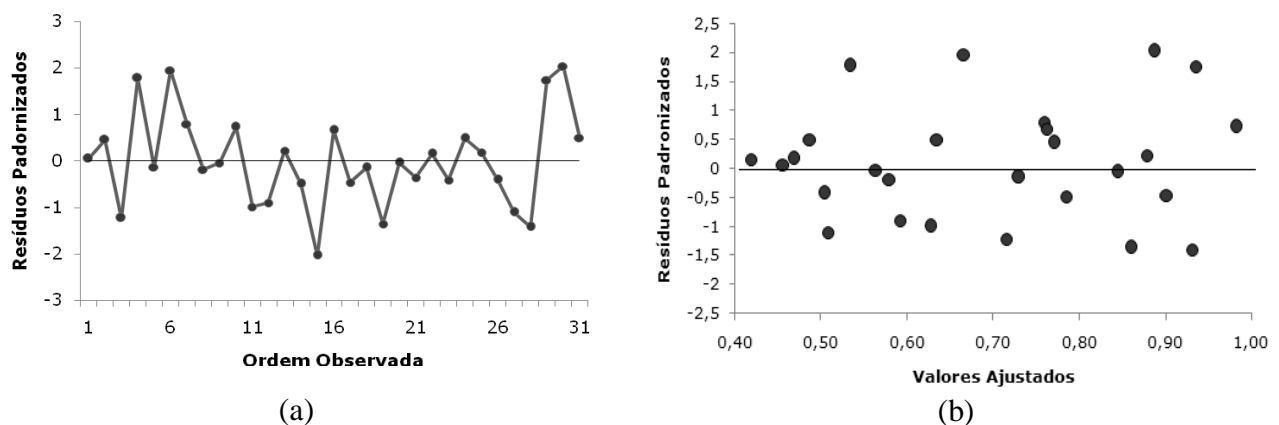


Figura 1. Resíduos Padronizados *versus* as Observações Ordenadas (a) e Resíduos Padronizados *versus* os Valores Ajustados (b), Obtidos a partir do Modelo (1) para Volume e Diâmetro de Árvores Quaruba, do Município de Bragança-PA, em Janeiro de 2009.

Com a transformação da variável Y , esta fica sem unidade de medida, para obter o verdadeiro valor da variável, com sua respectiva unidade de medida, é necessário utilizar o modelo dado por

$$\hat{Y}_i(\lambda) = (-0,0632 + 0,0249U_i)^2.$$

3.5 Gráfico de Controle de Regressão Funcional para Árvores Quaruba

Atendido aos pressupostos necessários a modelagem via Análise de Regressão, o Modelo (1) é considerado adequado para prever novos valores da variável diâmetro em função da variável volume. Fixa-se inicialmente a variância relacionada ao erro, contido na variável independente, $\sigma^2_\delta = 20$, pois este valor é proveniente do conhecimento de análises realizadas em amostras anteriores e $k = 3$. Portanto, o limite superior de controle, a linha central e o limite inferior de controle são obtidos, respectivamente, por

$$LSC = \hat{Y}_i(\lambda) + 3 \times 0,0716,$$
$$LC = \hat{Y}_i(\lambda) = -0,0632 + 0,0249U_i$$

e

$$LIC = \hat{Y}_i(\lambda) - 3 \times 0,0716.$$

A Figura 3 apresenta o Gráfico de Controle de Regressão Funcional para as variáveis volume e diâmetro de árvores Quaruba, do município de Bragança-PA, em janeiro de 2009, com $k = 3$. Nela, observa-se a existência de pontos fora dos limites de controle, dessa maneira pode-se afirmar que o processo não está sob controle estatístico.

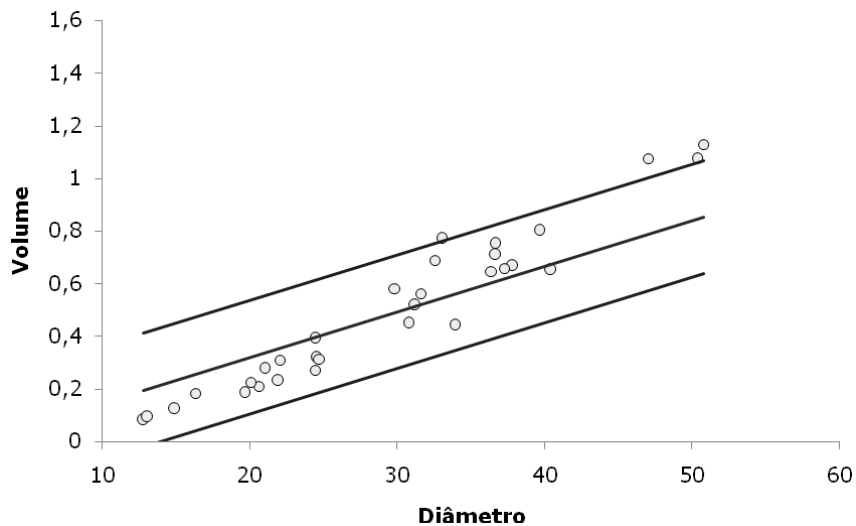


Figura 3: Gráfico de Controle de Regressão Funcional das Variáveis Volume e Diâmetro, com $\sigma^2_\delta = 20$, de Árvores Quaruba, do município de Bragança-PA, em Janeiro de 2009, com $k = 3$.

4. Conclusão

Este trabalho teve como objetivo desenvolver gráficos de controle de regressão funcional, para tanto, utilizou-se os dados referentes ao Volume e Diâmetro de Árvores Quaruba, do Município de Bragança-PA, em Janeiro de 2009. A partir do monitoramento das árvores constatou-se que o processo deve ser investigado, pois nota-se a existência de pontos amostrais acima do limite superior de controle (LSC), ou seja, mostrando árvores com valores acima do esperado, sendo possivelmente altas e de diâmetro um pouco menor. Isto ocorre provavelmente pelo sombreamento da área em que se encontram. Sendo assim, faz-se necessária a investigação do processo. Caso um ponto amostral apareça abaixo dos limites de controle, a árvore apresentará um valor considerado alto de diâmetro e um valor baixo referente ao volume, sendo uma árvore com pequena estatura e porte mais grosso. Isto pode ocorrer devido a ataques de pragas no período vegetativo da planta (primeiros 5 anos), deficiência hídrica ou até mesmo alta exposição ao sol, comprometendo assim o desenvolvimento da mesma. Caso o ponto amostral esteja acima do limite superior de controle, a árvore apresentará um volume maior, sendo alta e de diâmetro um pouco menor, este fenômeno pode ocorrer devido à

disposição das árvores no terreno, pois aquelas que apresentam maior sombreamento tendem a crescer mais verticalmente para buscar a luz solar.

5. Referências

- [1] DIPAOLA, P. P. Use of Correlation in Quality Control. *Industrial Quality Control*, v. 2, n. 1, July, p. 10-14, 1945.
- [2] ALMEIDA, S.S. Desenvolvimento de Gráficos de Controle Aplicados ao Modelo Funcional de Regressão. 2003. 166 f. Tese (Doutorado em Engenharia de Produção), Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis - SC.
- [3] FULLER, W. A. *Measurement Error Models*, JOHN WILEY, New York, 1987.
- [4] MITRA, A. *Quality Control and Improvement*, 2.ed. Upper Saddle River, New Jersey: Prentice Hall, 1988.
- [5] BOX, G. E. P and COX, D. R. *An Analyses of Transformation*. *Journal of Royal Statistical Society. B*, 39, 211-252, 1964.
- [6] LILLIEFORE, H. W. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, *Journal of the American Statistical Association*, n. 62, p. 399-402, 1967.