SPOLM 2009

ISSN 2175-6295 Rio de Janeiro- Brasil, 05 e 06 de agosto de 2009.

080/2009 - O PROBLEMA DA ESTIMAÇÃO EM MODELOS COM ERROS NAS VARIÁVEIS: ESTUDO DO VOLUME DE ÁRVORES QUARUBA

Cássio Pinho dos Reis
Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
cassioreisufpa@gmail.com

Rodrigo Valente Torres
Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
rodrigo174@gmail.com

Silvia dos Santos de Almeida
Universidade Federal do Pará – UFPA
Rua Augusto Correa, 01, Guamá – 66075-110, Belém-PA
salmeida@ufpa.br

Resumo

O objetivo deste trabalho é mostrar os problemas dos estimadores nos modelos de regressão quando as variáveis, independente (*X*) e dependente (*Y*), possuem erros, indicando alternativas para este tipo de modelagem. Para tanto, utiliza-se 32 medidas do diâmetro (cm) da altura do peito (*X*) e do volume (m^3) de árvores Quaruba (*Y*). Os resultados mostram que na presença de erros na variável independente (*X*) é mais coerente a utilização do Modelo com Erros nas Variáveis (MEV), pois na estimação de seus parâmetros, os erros associados as variáveis *X*, também são levados em consideração, fato que não ocorre na estimação dos parâmetros no Modelo de Regressão Clássico, onde este erro é considerado nulo, gerando neste caso estimadores não precisos, podendo acarretar com isso inferências erradas.

Palavras-Chaves: Estimadores, Regressão Clássica, Regressão Funcional.

Abstract

The aim of this work is to show the problems of the estimators in regression models when the independent (X) and dependent (Y) variables have errors, suggesting alternatives to this type of modeling. For this, we use 32 measures of the diameter (cm), of the height of the chest (X) and of the volume (m3) of Quaruba trees (Y). The results show that in the presence of errors in the independent variable (X) is more consistent the use of the Model with Errors in Variables (MEV), because in the estimation of its parameters, the errors associated with the variables X are also taken into account, which does not occur in the estimation of parameters in the Classic Regression Model, where this error is considered null, generating in this case not accurate estimators and which can lead to erroneous inferences.

Keywords: Estimators, Classic Regression, Functional Regression.

1. INTRODUÇÃO

Em geral, os instrumentos de medição são feitos para dar a máxima precisão necessária para a execução de inúmeras atividades tanto profissionais quanto cotidianas que requerem atenção. 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 que necessitam dessa minimização dos erros de medida dos instrumentos de medição, como é o caso da Dandometria (ciência que estuda a altura e o diâmetro das árvores). Vochysia Maxima Oucke, ou Quaruba, como é o nome usual, é uma árvore típica da região amazônica e das Guianas, porém pode-se encontrá-la desde o sul do México até o norte do Brasil. Sua variabilidade é enorme, mas é comum encontrar árvores de 30 metros de altura e 60 centímetros de diâmetro, entretanto já foi encontrada árvores de 57 metros de altura e 180 centímetros de diâmetro.

A madeira da Quaruba é classificada como de características médias quanto ao peso específico, resistência mecânica e retratibilidade, e possui baixa resistência natural ao apodrecimento. Sua secagem ao ar é reportada como moderada, com tendência a empenamentos e rachaduras. Apesar de ser classificada como mediana, a madeira é fácil de trabalhar, sendo bastante utilizada em acabamentos internos na construção civil, tais como, forros, guarnições, molduras e similares. Utilizada também na marcenaria em geral como armação de móveis, compensados, embalagens leves, tábuas em geral, formas para concreto, brinquedos, etc.

Atualmente, nos sistemas de inventário florestal, é imprescindível o estudo detalhado de equações volumétricas visando dar representatividade às informações sobre o volume de madeira que será obtido para as mais diversas atividades. Estas pesquisas necessitam de base cadastral de informações e recursos computacionais compatíveis com os níveis de precisão almejados.

Neste sentido, este trabalho apresenta dois métodos de estimação do volume (Y) da Quaruba em relação ao seu diâmetro (X), onde no primeiro método, se supõem que a variável diâmetro (X) é medida sem erro (modelo clássico de regressão), o que na prática é uma suposição fraca, devido a mesma ser resultante de processo de medida, logo passível de erro. No segundo método, esse erro é considerado (modelo com erro na variável), e é observado o seu efeito nas estimativas dos parâmetros.

2. METODOLOGIA

2.1. DADOS

Os dados utilizados neste trabalho foram obtidos junto a Universidade Federal Rural da Amazônia (UFRA), e se referem a 32 amostras de árvores da espécie Vochysia Maxima Oucke (Quaruba), onde observa-se as variáveis volume (m^3) e o diâmetro da árvore na altura do peito (cm).

2.2. ANÁLISE DE REGRESSÃO

A análise de regressão é feita a partir de um conjunto de métodos para o estabelecimento de fórmulas que interpretam a relação funcional entre variáveis com boa aproximação. Para Fonseca et al (1988), ela é feita para que se possa encontrar que: (a) se há alguma relação entre as variáveis e, caso a resposta seja positiva, dizer se ela é fraca ou forte; (b) caso essa relação exista, se há como estabelecer um modelo matemático que interprete a relação entre as variáveis e (c) construído o modelo, pode ser utilizado para fins de predição.

2.2.1. Modelos de Regressão Linear Clássico

Segundo Charnet et al (1999), o termo regressão foi empregado pela primeira vez por Francis Galton (1822 - 1911) num estudo da relação entre as alturas dos pais e filhos. O modelo de regressão é um dos métodos estatísticos mais usados para investigar a relação entre variáveis. Seu principal objetivo é modelar o relacionamento entre diversas variáveis preditoras e uma variável resposta. Este relacionamento pode ser por uma equação linear ou uma função não linear.

Sendo X uma variável independente com valores fixados, e Y a variável dependente, pode-se determinar uma relação funcional entre as mesmas como $Y = f(x)$ a partir de uma amostra de valores de X e Y, onde o modelo pode ser descrito formalmente como

$$Y_i = \alpha + \beta X_i + \varepsilon_i; \qquad i = 1,...,n, \tag{1}$$

sendo que $\varepsilon_i's$ são os erros aleatórios independentes e identicamente distribuídos com média zero e variância constante. Em geral, os valores de α e β são desconhecidos. Segundo Fonseca *et al* (1988), eles podem ser estimados utilizando o método dos mínimos quadrados, sendo que

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}} \tag{2}$$

e

$$\hat{\alpha} = \overline{Y} - \beta\overline{X}\ . \tag{3}$$

Os valores de $\overline{X}, \overline{Y}, S_{XX}$ e S_{XY} são dados pelas equações a seguir

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \tag{4}$$

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}, \tag{5}$$

$$S_{XX} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} \tag{6}$$

e

$$S_{XY} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{n-1}. \tag{7}$$

O modelo estimado é dado por:

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i, \qquad i = 1,...,n. \tag{8}$$

Sendo $\hat{\beta}$ e $\hat{\alpha}$ dados pelas equações (2) e (3), respectivamente.

Deve-se lembrar que esses estimadores são função dos dados amostrais e variam, portanto, de amostra para amostra. Assim, genericamente, o erro padrão é o desvio padrão da distribuição dos estimadores em diversas amostragens. A determinante principal da precisão é a quantidade de dispersão na população: quanto maior a dispersão (erro padrão), menor a precisão das estimativas obtidas. Portanto os estimadores do erro padrão de $\hat{Y}$, $\hat{\alpha}$ e $\hat{\beta}$ podem ser calculados por

$$EP(\hat{Y}) = \sqrt{\sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)}{n-2}} = \hat{\sigma}_Y \,, \tag{9}$$

$$EP(\hat{\alpha}) = \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{S_{XX}}} \,, \tag{10}$$

e

$$EP(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}_Y{}^2}{S_{XX}}} \tag{11}$$

onde $\hat{Y}$, $\overline{X}$, S_{XX} e $\hat{\sigma}_Y$ são obtidos pelas equações (8), (4), (6) e (9) respectivamente. Entretanto, sabe-se da teoria clássica de regressão que toda essa formulação está baseada na idéia de que os $xi's$ são medidas sem erro, e, portanto não pode ser aplicada a modelos que contenham este tipo de variável.

2.2.2. Modelos de Regressão com Erros nas Variáveis (MEV)

Na prática, apesar dos constantes avanços tecnológicos estarem tornando cada vez mais precisos os procedimentos de mensuração, não é realista supor que a variável X seja medida sem nenhum erro, O mais comum é não se ter acesso aos seus verdadeiros valores que foram mensurados durante o processo de amostragem, Neste caso, têm-se os chamados Modelos 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 eles estão presentes, quer sejam eles provocados pelo instrumento de medição, pelo operador ou por fatores externos (sol, chuva, vento). Este tipo de problema é bastante antigo na análise de regressão linear quando Wald (1940) e Bartlett (1949) se interessaram pelo assunto. Na regressão clássica, sabe-se que o modelo considera variáveis independentes $X_i's$, são medidas sem erros, com isso as formulações não podem ser aplicadas a modelos que contenham esses tipos de erros. Os modelos com erros nas variáveis (MEV) são uma generalização dos modelos de regressão padrão (clássico). Suponha um modelo de regressão, onde as variáveis *Ui* e y_i são duas quantidades não observáveis e estão relacionadas por meio de uma equação linear como apresentada em Moran (1971),

$$y_i = \alpha + \beta U_i \qquad ; i = 1,\ldots,n \,, \tag{12}$$

onde 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, onde

$$Y_i = y_i + \varepsilon_i, \qquad i = 1,...,n \tag{13}$$

e

$$X_i = U_i + \delta_i, \qquad i = 1,...,n \tag{14}$$

sendo os $\varepsilon_i's$ e $\delta_i's$, $i = 1,...,n$ erros, e considerados independentes e identicamente distribuídos com média zero e variâncias finitas σ_ε^2 e σ_δ^2, respectivamente. Assim, podem-se reescrever os modelos dados anteriormente (12) – (14), como

$$Y_i = \alpha + \beta X_i + \nu_i, \tag{15}$$

onde $\nu_i = \varepsilon_i - \beta\delta_i$, $i = 1,...,n$.

Observa-se que este não é um modleo de regressão clássico, pois X_i é aleatório, para qualquer modelo com erros, e também X está correlacionado com o erro ν_i, onde $\text{cov}(X,\nu) = -\beta\sigma_\delta^2$. Neste caso, Almeida (2003) mostra que se forem utilizados os estimadores de regressão clássica (mínimos quadrados) em dados com erro em X, serão obtidas estimativas inconsistentes.

É importante ressaltar que no modelo com erros nas variáveis, as variáveis $U_i's$ podem se apresentar de duas maneiras. Para Fuller (1987), o modelo funcional é aquele que considera os $U_i's$ como valores constantes, e com isto, o número de parâmetros cresce de acordo com o tamanho da amostra, fazendo com que no modelo, existem n + 4 parâmetros ao todo, sendo eles α, β, σ_ε^2, σ_δ^2 e U_i com $i = 1,...,n$. Já no modelo estrutural, a principal característica que o modelo apresenta é o fato de que os $U_i's$ são agora, em vez de valores constantes, variáveis aleatórias independentes entre si e também de para todo, Possuí média μ_μ e variância σ_μ^2 . Portanto, têm-se seis parâmetros no modelo, sendo eles α, β, μ_μ, σ_ε^2, σ_δ^2 e σ_μ^2, Neste estudo específico, 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 será possível se forem feitas suposições adicionais aos parâmetros. Como por exemplo, o conhecimento de uma das variâncias do erro, digamos, σ_δ^2 (variância do erro da variável X). Cheng e Van Ness (1994) afirmam que o procedimento comumente utilizado no modelo de regressão funcional para estimação de α e β, quando a variância dos erros de medição σ_δ^2 é conhecida, é adotar os estimadores obtidos pelo método dos momentos no modelo estrutural, como os correspondentes no modelo funcional, e que nesse caso também são estimadores consistentes.

Seja o modelo funcional formalmente definido em Almeida (2003) como

$$Y_i = \alpha + \beta U_i + \varepsilon_i \tag{16}$$

$$X_i = U_i + \delta_i, \qquad i = 1,...,n, \tag{17}$$

sendo que U_i, $i = 1,...,n$ são as constantes fixadas (parâmetros incidentais), os erros de medição (ε_i, δ_i), normais, independentes e identicamente distribuídos, ambas com média zero e variâncias constantes σ_ε^2 e σ_δ^2, respectivamente.

Almeida (2003) mostra que a equação de regressão estimada no modelo com erros nas variáveis é dada pela esperança da equação de regressão populacional de *Y*, da seguinte maneira

$$\hat{Y} = \hat{\alpha} + \hat{\beta}\hat{U}_i, \qquad i = 1,...,n \tag{18}$$

onde no caso do modelo funcional com uma das variâncias conhecidas, σ_δ^2, os estimadores da regressão, $\hat{\alpha}$ e $\hat{\beta}$ são obtidos

$$\hat{\beta} = \frac{S_{XY}}{S_{XX} - \sigma_\delta^2} \tag{19}$$

e

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}\,. \tag{20}$$

O estimador não viesado da variância do erro do modelo funcional é dado por:

$$\hat{\sigma}_{v_i}^2 = \frac{\sum (Y_i - \overline{Y} - (X_i - \overline{X})\hat{\beta})^2}{(n-2)} = \hat{\sigma}_Y^2 \tag{21}$$

ou ainda

$$\hat{\sigma}_Y^2 = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}\,. \tag{22}$$

Fuller (1987) mostra que o mehor estimador linear não viesado de U_i é dado por

$$\hat{U}_i = \frac{\hat{\sigma}_\delta^2\hat{\beta}(Y_i - \hat{\alpha}) + \hat{\sigma}_\varepsilon^2 X_i}{\hat{\beta}^2\hat{\sigma}_\delta^2 + \hat{\sigma}_\varepsilon^2}, \tag{23}$$

onde $\hat{\sigma}_\varepsilon^2 = S_{YY} - \hat{\beta}_{XY}$

onde os valores de $\overline{X}$, $\overline{Y}$, S_{XX} e S_{XY}, são dados pelas equações (4), (5), (6) e (7), respectivamente. O valor de σ_δ^2 é determinado pelo conhecimento da variabilidade de processos anteriores. Geralmente sua fixação não é considerado um problema, pois geralmente já se tem conhecimento deste valor provenientes de dados 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, os seus erros padrão da distribuição aproximada dos estimadores de α e β. Eles podem ser obtidos a partir de

$$EP(\hat{\alpha}) = \sqrt{\frac{n\overline{X}^2 \hat{V}ar(\hat{\beta}) + S_\nu^2}{n}} \qquad (24)$$

e

$$EP(\hat{\beta}) = \sqrt{\frac{\left(S_{XX} S_\nu^2 + \hat{\beta}^2 \sigma_\delta^4\right)}{l^2 (n-1)}} \ , \qquad (25)$$

onde o valor de $\overline{X}$, S_{XX}, $\hat{\beta}$ e S_ν^2 são dados pelas equações (4), (6), (19) e (21) respectivamente. E o valor de l pode ser obtido por

$$l = \frac{\sum_{i=1}^{n} \left(U_i - \overline{U}\right)^2}{n-1} . \qquad (26)$$

Quando se estuda duas ou mais variáveis, cuja relação é expressa por uma equação do tipo (8), o desvio padrão de Y (reta de regressão) passa a ser o afastamento médio e mínimo existente entre cada ponto observado e a reta estimada, e é chamado de Erro Padrão da Linha de Regressão, que no caso do modelo funcional, um estimador pode ser dado pela raiz quadrada de (10). Porém, quando se deseja estimar o desvio padrão ou erro padrão da estimativa da reta de regressão, $\hat{Y}_i$, ela é chamada de Erro Padrão da Linha de Regressão Estimada, $EP(\hat{Y}_i)$, e é dada por

$$\hat{E}P(\hat{Y}_i) = \sqrt{S_\nu^2} = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_i - \overline{Y} - \left(X_i - \overline{X}\right)\hat{\beta}\right)^2}{(n-2)}} . \qquad (27)$$

2.2.3. Desenvolvimento de uma Ferramenta Computacional do MEV

Como não existe um programa ou pacote computacional que realize automaticamente os cálculos do modelo com erros nas variáveis, se fez necessário o desenvolvimento desta ferramenta computacional, para facilitar o processo, e para se ter uma melhor precisão nos resultados, uma vez que assim, usa-se todas as casas decimais possíveis. A Figura 1 apresenta esta ferramenta computacional.

	A	B	C	D	E	F	G	H	I	J
1	Diâmetro(X)	Volume(Y)	Sxx		Syy		Sxy	Ui1	S^2_v1	I1
2			Xi-Xbarra	(Xi-Xbarra)^2	Yi-Ybarra	(Yi-Ybarra)^2	(Xi-Xbarra)(Yi-Ybarra)	Valor total do 1° erro	pela 2° formula	
3	20,626500	0,210540	-9,554819	91,294561	-0,295058	0,087059	2,819226905	20,46559601	0,001869538	94,39526841
4	36,382800	0,647080	6,201481	38,458370	0,141482	0,020017	0,877397195	36,30107957	0,000482239	37,45147256
5	24,414400	0,392550	-5,766919	33,257352	-0,113048	0,012780	0,651939352	24,55931155	0,001516375	31,606965
6	33,900000	0,440950	3,718681	13,828590	-0,064648	0,004179	-0,24040577	33,2947049	0,026456658	9,693173293
7	31,216200	0,517600	1,034881	1,070979	0,012002	0,000144	0,012420515	31,15936499	0,000233256	0,956574453
8	40,425400	0,651820	10,244081	104,941201	0,146222	0,021381	1,497908768	39,96483247	0,015317479	95,71714069
9	37,815200	0,667950	7,633881	58,276143	0,162352	0,026358	1,239374934	37,67065844	0,001508642	56,09020892
10	24,509900	0,319010	-5,671419	32,164991	-0,186588	0,034815	1,058219391	24,37177765	0,001377617	33,75076784
11	36,669300	0,709180	6,487981	42,093901	0,203582	0,041446	1,320835388	36,79057588	0,001062061	43,68227985
12	46,982500	1,074500	16,801181	282,279691	0,568902	0,323649	9,558223515	47,4517686	0,015901704	298,268438
13	49,497200	0,784230	19,315881	373,103268	0,278632	0,077636	5,38202021	48,6396401	0,053104289	340,7096271
14	22,027000	0,308730	-8,154319	66,492914	-0,196868	0,038757	1,605325443	22,09413774	0,000325487	65,40249664
15	21,008500	0,277000	-9,172819	84,140604	-0,228598	0,052257	2,096889167	21,05745101	0,000173031	83,24496251
16	39,629600	0,801490	9,448281	89,270019	0,295892	0,087552	2,795669655	39,80405555	0,002197709	92,59706344
17	31,608200	0,562920	1,426881	2,035990	0,057322	0,003286	0,081791509	31,68157004	0,000388721	2,250753939
18	33,008700	0,771360	2,827381	7,994085	0,265762	0,070629	0,751410142	33,72038981	0,036574817	12,52502397
19	37,242300	0,658750	7,060981	49,857456	0,153152	0,023455	1,081402518	37,11971042	0,001085195	48,141279
20	36,637500	0,752290	6,456181	41,682276	0,246692	0,060857	1,592687458	36,92232182	0,00585797	45,44112235
21	12,792100	0,081980	-17,389219	302,384929	-0,423618	0,179452	7,366388242	12,92115501	0,001202681	297,9132524
22	29,857500	0,581350	-0,323819	0,104859	0,075752	0,005738	-0,024529877	30,17115787	0,007104165	0,000103243
23	24,732700	0,314260	-5,448619	29,687446	-0,191338	0,036610	1,042528495	24,5550497	0,002278938	31,6549034
24	13,082500	0,093880	-17,098819	292,369603	-0,411718	0,169512	7,039893595	13,22735741	0,001515242	287,4368051
25	19,671600	0,184800	-10,509719	110,454188	-0,320798	0,102911	3,371498069	19,50856242	0,001919447	113,9077277
26	20,053500	0,224540	-10,127819	102,572713	-0,281058	0,078994	2,846505748	20,00089295	0,000199843	103,6410694
27	24,446200	0,272150	-5,735119	32,891587	-0,233448	0,054498	1,338852719	24,13994303	0,006772868	36,49822053
28	21,931600	0,231800	-8,249719	68,057859	-0,273798	0,074965	2,258757526	21,7218114	0,003178077	71,56326462
29	14,928700	0,124460	-15,252619	232,642379	-0,381138	0,145266	5,813354512	15,00628623	0,00043468	230,2816121
30	16,297500	0,183250	-13,883819	192,760423	-0,322348	0,103908	4,475422942	16,4596161	0,001897811	188,2851236
31	32,531300	0,687570	2,349981	5,522412	0,181972	0,033114	0,427630494	32,97800049	0,014408989	7,821428742
32	50,738600	1,129270	20,557281	422,601812	0,623672	0,388967	12,82099814	51,04329934	0,00670415	435,2222342
33	50,356600	1,072850	20,175281	407,041974	0,567252	0,321775	11,44446612	50,48880662	0,001262139	412,3940635
34	30,780600	0,449030	0,599281	0,359138	-0,056568	0,003200	-0,033900217	30,51131489	0,00523631	0,108897449
35	**Média de X**	**Média de Y**		Somatório		Somatório	Somatório	**Média de U**	Somatório	Somatório
36	**30,181318750**	**0,5056**		3611,693712		2,68516908	94,37020280	**30,181318750**	0,21954813	3608,65332397
37			**Resultado Final**	Dividindo por 31	**Resultado Final**	Dividindo por 31	**Resultado Final**		**Resultado Final**	**Resultado Final**
38				**116,5062488**		**0,086618357**				
39							Dividindo por 31		Dividindo por 30	Dividindo por 31
40							**3,04420009**		**0,007318271**	**116,4081717**
42		sigma"	alfa	beta	erro padrao alfa	erro padrao beta	erro padrao Y	sigma1^2_epsolon		
43		1	-0,2898391	0,026355285	0,045596803	0,00142525	0,085546894	0,006387597		
45			sigma	Var(Alfachapeu)	Var(Betachapeu)	Var(Ychapeu)				
46			1	0,0020790684	0,0000020313					

Figura 1: Ferramenta Computacional Usada para o Desenvolvimento e Geração dos Estimadores do Modelo com Erros nas Variáveis.

3. RESULTADOS E DISCUSSÕES

Em Dandometria (Setor da ciência que estuda a altura e o diâmetro das árvores), a medição de altura de árvores pode ser necessária para determinar o volume da madeira de um povoamento florestal ou para acompanhar o desenvolvimento florestal de uma região (REIS *et al*, 2009). A modificação dos modelos, assim como o surgimento de novas marcas, justifica a contínua avaliação da precisão de instrumentos para medição da altura de árvores. Os erros de medição de altura influem diretamente na precisão da estimativa do volume de árvores individuais e consequentemente do povoamento florestal.

Com base na importância dos estudos da dandometria, e nos possíveis erros de medição que poderá ter ao medir o diâmetro na altura do peito das árvores, pretende-se estimar o volume das árvores Quaruba, e verificar seu comportamento à medida que se aumenta o erro na variável independente (X). Para isso será feita a verificação das estimativas obtidas para o modelo funcional para diferentes valores de erros. Para tanto, utiliza-se como aplicação os dados de diâmetro na altura do peito e volume de uma amostra de árvores Quaruba.

Analisando a Figura 2, percebe-se visualmente a existência de uma forte correlação positiva entre as variáveis diâmetro (X) e volume (Y), das árvores Quaruba, indicando que um modelo linear pode ser utilizado para estimar o volume das árvores. É importante ressaltar que os valores são resultantes de processos de medição, ou seja, ambas podem ser sujeita a erros.

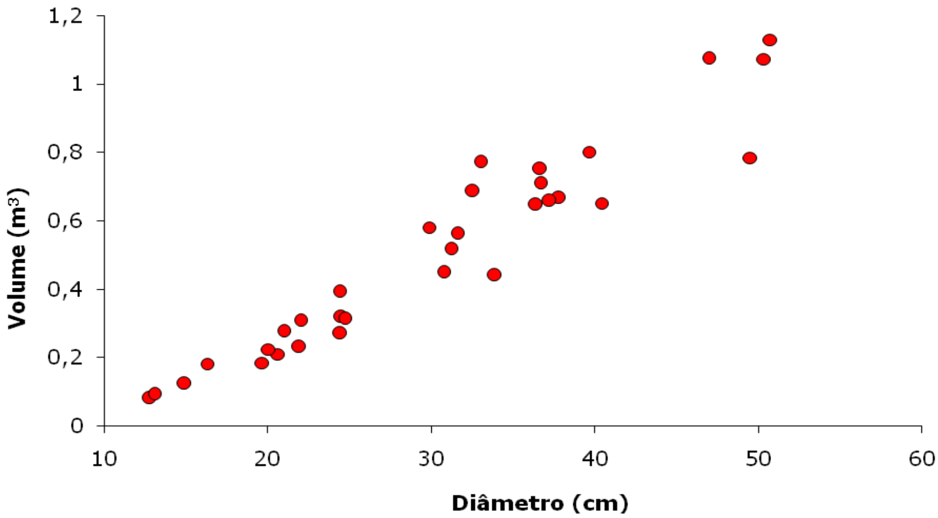

Figura 2: Gráfico de Dispersão para as Variáveis Diâmetro na Altura do Peito (DAP) e Volume das Árvores Quaruba.

Os resultados das estimativas dos parâmetros, juntamente com seus erros padrão, tanto para o modelo clássico quanto para o modelo funcional, apresentada nas Tabelas 1 e 2, respectivamente. De onde, pode-se destacar que a presença de erros de mensuração afeta a precisão dos estimadores, observe que a medida em que σ_{δ}^{2} cresce, ou seja, o erro associado a variável X cresce, a estimativa do erro padrão do modelo (Y) também crescem.

Tabela 1: Estimativas Obtidas no Modelo de Regressão Clássico.

α	β	EP(α)	EP(β)	EP(Y)
-0,283	0,026	0,046	0,001	0,086

Portanto, o modelo estimado é dado por

$$\hat{Y} = -0{,}283 + 0{,}026 X_i$$

Tabela 2: Estimativas Obtidas por Meio do Modelo de Regressão Funcional com Diferentes Valores para σ_{δ}^{2}.

σ_{δ}^{2}	α	β	EP(α)	EP(β)	EP(Y)	Modelo Estimado ($\hat{Y}$)
1	-0,2898	0,0263	0,0456	0,0014	0,0855	-0,2898+0,0263U_i
5	-0,3183	0,0273	0,0474	0,0015	0,0865	-0,3183+0,0273U_i
10	-0,3571	0,0286	0,0540	0,0017	0,0896	-0,3571+0,0286U_i
20	-0,4464	0,0315	0,0908	0,0029	0,1041	-0,4464+0,0315U_i
30	-0,5565	0,0352	0,1946	0,0064	0,1311	-0,5565+0,0352U_i
40	-0,6953	0,0398	0,4653	0,0154	0,1726	-0,6953+0,0398U_i
50	-0,8759	0,0458	1,1611	0,0384	0,2319	-0,8759+0,0458U_i

75	-1,7080	0,0733	13,8546	0,4590	0,5251	-1,7080+0,0733U_i
100	-5,0606	0,1844	506,3538	16,7771	1,7390	-5,0606+0,1844U_i
110	-13,6159	0,4679	10975,7968	363,6620	4,8479	-13,6159+0,4679U_i

Analisando-se ambos os métodos (clássico e funcional), pode-se destacar que em amostras sujeitas a erros, como no caso das medições do diâmetro na altura do peito de árvore Quaruba (X), 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 X, refletido em σ_δ^2 , maior será o erro padrão dos seus estimadores, conforme mostra a Tabela 2.

4. CONCLUSÃO

Este trabalho teve como objetivo comparar as estimativas geradas nos modelos de regressão clássico e funcional, verificando-se seus comportamentos a medida em que se aumenta o erro na variável X. Inicialmente observou-se o diagrama de dispersão, de onde conclui-se que o diâmetro e o volume são altamente correlacionados comprovando a existência de uma relação linear, o que proporciona a modelagem por meio de um modelo de regressão, modelo este capaz de estimar o volume da arvore Quaruba a partir de seu diâmetro.

Foi necessário o desenvolvimento de uma ferramenta computacional, pois não se tem ainda um pacote computacional (programa) que realize automaticamente este tipo de modelagem, foi feita a comparação dos estimadores dos modelos de regressão clássico e funcional. Observa-se que, como as variáveis das árvores de Quaruba são resultantes de processos de medição, ou seja, as amostras estão sujeitas a erros, a utilização do modelo clássico pode trazer problemas, pois neste tipo de modelagem, o erro existente na variável *X* é considerado nulo o que faz com que, quanto maior o erro associado a variável *X*, o erro padrão de seus estimadores também será maior e o modelo irá gerar estimativas distantes da realidade. Então, neste caso, o mais viável é a utilização do modelo funcional, pois ao estimar os seus parâmetros, os erros na variável *X* são levados em consideração, logo esses estimadores serão consistentes e o modelo mais preciso.

5. REFERÊNCIAS BIBLIOGRÁFICAS

[1] ALMEIDA, Silvia dos Santos, 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.

[2] ALMEIDA, Silvia dos Santos, 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.

[3] BARTLETT, V. D. Fitting a Straight Line When Both Variables are Subject to Error, *Biometrics*, v. 5, *p.* 207-212, 1949.

[4] CARVALHO JÚNIOR, J. G., ALMEIDA, S. S. e RAMOS, E. M. L. S. Gráfico de Controle de Regressão Estrutural, TEMA Tend. Mat. Apl. Comput., 8, No. 3, 361-370, 2007.

[5] CHARNET, Reinaldo *et al*. *Análise de Modelos de Regressão Linear*. 2.ed., Campinas: Editora UNICAMP, 1999.

[6] 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.

[7] FONSECA, J. S. *et al*. *Estatística Aplicada*. 2. ed., São Paulo: Atlas, 1988.

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

[9] HOADLEY, B. A Bayesian and Application of Linear Model, *Jornal of the American Statistical Association*, *p*, 356-269, 1970.

[10] LIMA, Cláudia Regina Oliveira de Paiva, Calibração Absoluta com Erros nas Variáveis. São Paulo, 1996. 223p. Tese (Doutorado em Estatística). IME, Universidade de São Paulo.

[11] 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.

[12] REIS, C. P; TORRES, R.V; ALMEIDA, S.S. *Comparação dos Estimadores de Regressão Clássica e Funcional Aplicado ao Estudo das Árvores Quaruba*. XI Escola de Modelos de Regressão, Recife, pg. 131, mar.2009.

[13] MORAN, P. A. P. *Estimating Structural and Functional Relationships,* Journal of Multivariate Analysis, v. 1, p. 232-255, 1971.

[14] WALD, A. *The Fitting of Straight Lines if Both Variables are Subject to Error*, Annals of Mathematical Statistics, v.11, *p.* 284-300, 1940.

[15] WILLIAMS, E.J. A *Note on Regression Methods in Calibration*, Technometrics, p, 189-192, 1969.