Revista Eletrônica Paulista de Matemática

ISSN 2316-9664
v. 23, n. 2, dez. 2023
Artigo de Pesquisa

Cássio Pinho dos Reis
Universidade Federal de Mato Grosso do Sul
cassio.reis@ufms.br

Silvia dos Santos de Almeida
Universidade Federal do Pará
salmeida@ufpa.br

Uma aplicação do problema da estimação em modelos de regressão com erros nas variáveis

An application of the estimation problem in regression models with variable errors

Resumo

A exigência de um melhor controle dos procedimentos de mensuração tem se intensificado, levando pesquisadores a buscar novas técnicas para obter melhores resultados, com o auxílio de algum instrumento de medição, embora possa haver erros de mensuração. Dessa forma, o objetivo deste trabalho é mostrar o problema que se tem, quando a variável preditora do modelo de regressão apresenta algum tipo de erro na obtenção dos dados, fazendo com que o modelo de regressão com erro nas variáveis o mais apropriado para este caso. Utiliza-se como aplicação, os dados referentes ao índice de massa corporal e circunferência da cintura de 80 pessoas. Como as amostras são sujeitas a erros, o mais indicado é a utilização do modelo funcional, pois na estimação dos seus parâmetros, os erros são levados em consideração. A presença de erros de medida, influenciam diretamente na precisão dos estimadores dos modelos, fazendo com que à medida em que o erro ligado a variável independente aumenta, o erro padrão também aumenta.

Palavras-chave: análise de regressão; modelo com erros nas variáveis; estimadores.

Abstract

The demand for better control of measurement procedures has intensified, leading researchers to seek new techniques to obtain better results, with the aid of some measuring instrument, although there may be measurement errors. Thus, the objective of this work is to show the problem that exists when the predictor variable of the regression model presents some type of error in obtaining the data, making the regression model with error in the variables as appropriate for this case.. As an application, data referring to the body mass index and waist circumference of 80 people are used. As the samples are subject to errors, the most indicated is the use of the model functional, because in the estimation of its parameters, the errors are taken into account. The presence of measurement erros, influence the accuracy of the models' estimators, meaning that as the error linked to the independent variable increases, the standard error also increases.

Keywords: regression analysis, models with erros in variables, estimators.

Artigo recebido em abr. 2023 e aceito em jul. 2023

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. De acordo com Lima (1996), 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 (Reis; Torres; Almeida, 2009). O uso de boas práticas, tais como a calibração do instrumento, o cálculo cuidadoso, boa manutenção de registros, a verificação constante do instrumento, pode reduzir as incertezas de medição, mas nunca a eliminar por completo.

Carrasco (2012) cita que há diversas classes de modelos de regressão, como por exemplo, heterocedásticos ou homocedásticos (de acordo com a variância do modelo), lineares ou não lineares (como por exemplo, quadrática, exponencial, logarítmica, etc), e com ou sem erros nas variáveis. Em cada uma dessas classes, existe suposições que obrigatoriamente devem ser consideradas, para que a inferência estatística seja coerente (Rosner; Willett; Spiegelman, 1989; Stefanski, 1985; Stefanski; Carroll, 1985).

Bell (1999) cita que nenhuma medida pode ser considerada exata. Quando a quantidade é medida, o resultado depende do sistema de medição, o procedimento de medição, a habilidade do operador, do ambiente e outros efeitos. Mesmo que a quantidade fosse medida por diversas vezes, da mesma maneira e nas mesmas circunstâncias, um valor diferente seria obtido a cada vez.

Podem existir dois tipos de erros de medição, como pode ver em Reis, Torres e Almeida (2009), o sistemático e o aleatório. Um erro sistemático está associado com o fato de que um valor medido contém um deslocamento. Em geral, um erro sistemático, considerado como uma quantidade, é um componente do erro que se mantém constante ou depende de uma forma específica de outra quantidade, ou seja, sofre influência do mundo externo. Um erro aleatório está associado ao fato de que quando uma medição é repetida, geralmente vai fornecer um valor de medição que é diferente do valor anterior, ou seja, é aleatório pois o próximo valor medido não pode ser previsto exatamente anteriormente a tais valores.

Os modelos com erros nas variáveis são utilizados quando as variáveis preditoras não são observadas diretamente ou estão sujeitas a erros de medição. Neste sentido, este trabalho tem como objetivo mostrar o problema que se tem, quando a variável preditora do modelo de regressão apresenta algum tipo de erro de mensuração na hora da obtenção dos dados, fazendo com que o modelo de regressão clássico se torne impreciso. Nesses casos, o modelo de regressão com erro nas variáveis se torna o apropriado para este caso.

Para obtenção dos resultados, utilizou-se informações referentes ao índice de massa corporal (IMC) e a circunferência da cintura (CC) dos moradores do município de Barcarena – Pará, presentes no setor de avaliação nutricional durante o evento Ação Global que ocorreu neste mesmo município, em maio de 2010. Este trabalho se dispõe a contribuir com o assunto teórico e com exemplos práticos, para a divulgação da metodologia que se apresenta, sendo, portanto, justificado certamente pela relevância para a sociedade como um todo.

Fuller (1987), fez um experimento para ilustrar como os erros nas variáveis podem ser empregados. O objeto do experimento foi avaliar o rendimento da produção de determinado cereal (Y), em função da percentagem de nitrogênio no solo (X). Apesar das amostras terem sido levadas para laboratório, não é realista supor que a análise química do solo não esteja associada a erros de medição.

Aoki, Bolfarine e Singer (2001) realizaram um estudo em que o interesse é comparar a eficácia de dois tipos de escovas de dentes na remoção de placa bacteriana; a covariável é o índice de placa antes da escovação e a variável resposta é o índice de placa após a escovação. Neste caso, é razoável supor que a covariável está sujeita a erros de medição, pois a quantidade de placa bacteriana é avaliada imprecisamente e é determinada de forma semelhante antes e após a escovação.

Na prática, é comum encontrar situações em que as covariáveis não são observadas diretamente, e consequentemente, os erros de medição podem ocorrer. Exemplos clássicos podem ser vistos em Buonaccorsi (2010), Carroll *et al* (2006) e Cheng e Van Ness (1994).

2 Metodologia

Para a obtenção dos resultados e aplicação de toda metodologia feita, utilizou-se informações referentes ao índice de massa corporal (IMC) e a circunferência da cintura de 80 moradores do município de Barcarena presentes no setor de avaliação nutricional durante o evento Ação Global que ocorreu em maio de 2010, no Município de Barcarena. Assim, construiu-se dois modelos de regressão (um modelo clássico e outro com erro nas variáveis), sendo que em ambos, a variável preditora é a circunferência da cintura (CC), enquanto que a variável de interesse (resposta) é o IMC. Todas as variáveis em estudo foram obtidas por meio de um instrumento de medição. O peso foi mensurado a partir de uma balança eletrônica com capacidade máxima de 150 kg, com divisões de 100g e precisão de 0,1kg. A balança foi calibrada previamente onde o indivíduo posicionou-se em pé, no centro da base da balança, descalço e com roupas leves. A estatura foi mensurada com o auxílio de um estadiômetro compacto com campo de medição de 0 a 200 cm e dispositivo para medição de altura fixo a parede. Onde o indivíduo ficou em pé e descalço em posição ortostática, com o corpo erguido em extensão máxima e a cabeça ereta, olhando para frente, em posição de Frankfut, com as costas e a parte posterior dos joelhos encostado ao antropômetro e com os calcanhares juntos. Por fim, o índice de massa corporal foi obtido a partir do peso e da altura. (Cuppari, 2018; Duarte, 2007).

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, Martins e Toledo (2010), 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.

Segundo Freire *et al* (2008), 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 a partir de uma amostra de valores de X e Y, onde o modelo linear pode ser descrito formalmente como

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

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, Martins e Toledo (2010), eles podem ser estimados utilizando o método dos mínimos quadrados, sendo que

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

e

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

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

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

Dessa forma, o modelo estimado é dado por:

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

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

$$EP(\hat{Y}) = \sqrt{\sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n-2}}, \quad (9) \qquad EP(\hat{\alpha}) = \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{S_{XX}}}, \quad (10)$$

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 ideia de que os *Xi's* são medidas sem erro, e, portanto, não pode ser aplicada a modelos que contenham este tipo de variável.

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 U_i e y_i são duas quantidades não observáveis e estão relacionadas por meio de uma equação linear como apresentada em Moran (1971),

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

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

$$Y_i = y_i + \varepsilon_i, \quad i = 1,\dots,n \quad (13) \quad \text{e} \quad X_i = U_i + \delta_i, \quad i = 1,\dots,n \tag{14}$$

sendo os $\varepsilon_i's$ e $\delta_i's$, $i = 1,\dots,n$ 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 $\nu_i = \varepsilon_i - \beta\delta_i$, $\quad i = 1,\dots,n$.

Observa-se que este não é um modelo de regressão clássico, pois X_i é aleatório, , 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, dependendo do seu comportamento. Ou os valores desta variável são fixados ou supõe-se que sejam variáveis aleatórias independentes com alguma distribuição de probabilidade. 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 $i = 1,\dots,n$. 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. 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 (1999) 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

$$Y_i = \alpha + \beta U_i + \varepsilon_i \qquad (16) \qquad \text{e} \qquad X_i = U_i + \delta_i, \qquad i = 1,\ldots,n, \qquad (17)$$

sendo que U_i, $i = 1,\ldots,n$ são as constantes fixadas (parâmetros incidentais), os erros de medição (ε_i, δ_i), normais, independentes e identicamente distribuídos, ambos 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

$$\hat{Y} = \hat{\alpha} + \hat{\beta}\hat{U}_i, \qquad i = 1,\ldots,n \qquad (18)$$

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

$$\hat{\beta} = \frac{S_{XY}}{S_{XX} - \sigma_\delta^2} \qquad (19) \qquad \text{e} \qquad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}. \qquad (20)$$

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

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

ou ainda

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

(22)

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

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

onde $\hat{\sigma}_\varepsilon^2 = S_{YY} - \hat{\beta}_{XY}$, os valores de $\bar{X}$, $\bar{Y}$, 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, pois já se tem conhecimento deste valor proveniente 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

$$EP(\hat{\alpha}) = \sqrt{\frac{n\bar{X}^2 \hat{V}ar(\hat{\beta}) + S_v^2}{n}} \qquad (24) \qquad \text{e} \qquad EP(\hat{\beta}) = \sqrt{\frac{(S_{XX} S_v^2 + \hat{\beta}^2 \sigma_\delta^4)}{l^2 (n-1)}}, \qquad (25)$$

onde o valor de $\bar{X}$, S_{XX}, $\hat{\beta}$ e S_v^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, $EP\left(\hat{Y}_i\right)$, e é dada por

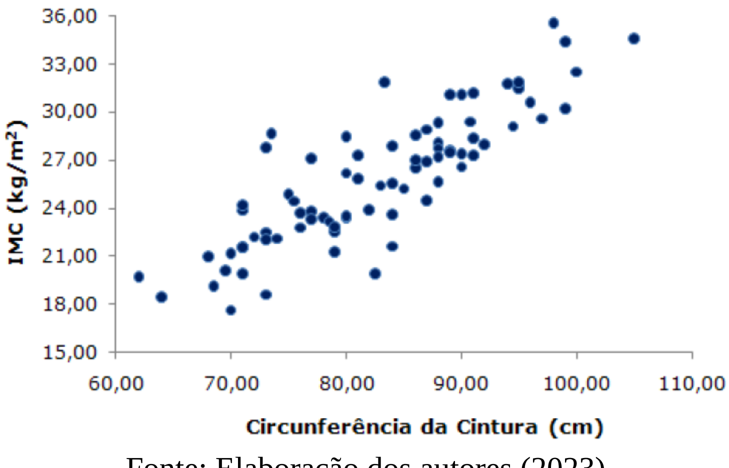
$$\hat{E}P\left(\hat{Y}_i\right) = \sqrt{S_v^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y} - \left(X_i - \overline{X}\right)\hat{\beta}\right)^2}{(n-2)}}. \qquad (27)$$

3 Resultados e discussões

Para se construir o modelo de regressão linear, é necessário verificar se as variáveis estudadas possuem uma relação linear entre elas. A Figura 1 apresenta o diagrama de dispersão das variáveis em estudo. Nela, pode-se verificar uma tendência de relação linear positiva entre o índice de massa corporal (IMC) e a circunferência da cintura, porém não há a garantia de existir uma causalidade.

Para confirmar a suposição de que as variáveis apresentam uma correlação linear positiva, foi calculado o coeficiente de correlação linear de Pearson, cujo valor foi de 0,85, confirmando assim a existência de uma forte correlação linear positiva, além disso, o nível descritivo (valor-p) é <0,0001, confirmando a significância do valor 0,85, uma vez que o valor <0,0001 está dentro da região crítica, rejeitando assim a hipótese nula de que o valor da correlação linear de Pearson não é significativo. Também é necessário verificar se a variável resposta (índice de massa corporal), segue uma distribuição normal. Para tanto, foi feito o teste de aderência de Kolmogorov-Smirnov para o IMC.

Figura 1 - Diagrama de Dispersão para as Variáveis índice de Massa Corporal e Circunferência da Cintura, dos moradores do município de Barcarena presentes no setor de avaliação nutricional do Ação Global em Maio de 2010.



Fonte: Elaboração dos autores (2023)

Satisfeitas as suposições de normalidade da variável resposta Yi e de correlação entre Xi e Yi é possível calcular as estimativas dos parâmetros desconhecidos α, β e seus respectivos erros padrão para o modelo clássico. A Tabela 1 apresenta as estimativas do modelo de regressão linear clássico e seus respectivos erros padrão.

Tabela 1 - Estimativas Obtidas no Modelo de Regressão Clássico.

$\hat{\alpha}$	$\hat{\beta}$	EP($\hat{\alpha}$)	EP($\hat{\beta}$)
-4,8516	0,3721	8,80	0,23

Fonte: Elaboração dos autores (2023)

Portanto, o modelo estimado é dado por

$$\widehat{Y}_i = -4,8516 + 0,3721X_i$$
.

Como foi evidenciado que as duas variáveis podem conter erros de medição, é preferível utilizar o modelo de regressão com erros nas variáveis, porque além de levar em consideração algum tipo de erro associado à variável resposta, que neste caso é o índice de massa corporal, ele leva em consideração também os erros associados a variável preditora (circunferência da cintura). Então, será construído o modelo de regressão linear com erro nas variáveis.

De acordo com os resultados das estimativas dos parâmetros, juntamente com seus erros padrão, para o modelo funcional, apresentada na Tabela 2, pode-se destacar que a presença de erros de mensuração na variável preditora afeta a precisão dos estimadores. Pode-se observar que, à medida em que a variância aumenta, ou seja, o erro associado a variável X cresce, a estimativa dos erros padrão tanto de α, quanto de β também crescem.

Tabela 2 - Estimativas Obtidas por Meio do Modelo de Regressão Funcional com Diferentes Valores para σ_δ^2.

σ_δ^2	$\hat{\alpha}$	$\hat{\beta}$	EP($\hat{\alpha}$)	EP($\hat{\beta}$)
1	-5,204	0,376	2,179	0,026
5	-6,697	0,394	2,218	0,027
10	-8,777	0,420	2,361	0,028
20	-13,850	0,481	3,184	0,038
30	-20,658	0,563	5,607	0,068
50	-44,907	0,856	36,409	0,440

Fonte: Elaboração dos autores (2023)

Pode-se observar que à medida em que a variabilidade da variável preditora, que neste caso é a circunferência da cintura aumenta, ou seja, à medida em que o erro na hora da mensuração da circunferência da cintura aumenta, não só os erros padrão das estimativas aumentam, mas também, os pressupostos necessários para a validação do modelo funcional passam a não ser validados, ou seja, quanto maior o valor desta variância, os resíduos padronizados passam a tornar dependentes um com os outros e passam a ter variâncias diferentes, se tornam heterocedásticos.

4 Conclusão

Este trabalho teve como objetivo mostrar o problema que se tem, quando a variável preditora do modelo de regressão apresenta algum tipo de erro de mensuração na hora da obtenção dos dados, fazendo com que o modelo de regressão clássico se torne impreciso, e consequentemente, o modelo de regressão com erro nas variáveis o apropriado para este caso.

Construído os dois modelos de regressão linear (um modelo clássico e um modelo com erro nas variáveis), verificou-se que todos os pressupostos necessários para a validação dos modelos foram atendidos, como a normalidade dos resíduos e da variável resposta, além da independência e da variância constante. Observou-se também com a obtenção dos erros padrão das estimativas, que a questão de que o modelo de regressão com erros nas variáveis é o melhor modelo para se usar neste caso em que as variáveis são obtidas por meio de um instrumento de medição, é comprovada nos modelos em estudo, uma vez que o erro padrão dos estimadores alfa e beta são inferiores em relação ao erro padrão dos estimadores do modelo clássico.

Pode-se concluir também que a presença de erros de mensuração na variável preditora afeta a precisão dos estimadores, uma vez que a medida em que a variância cresce, ou seja, o erro associado a variável X cresce, a estimativa dos erros padrão tanto de α, quanto de β também crescem. Além de que também, os pressupostos deixam de ser válidos a partir do momento em que este erro cresce.

5 Referências

ALMEIDA, Silvia dos Santos de. **Desenvolvimento de gráficos de controle aplicados ao modelo funcional de regressão**. Orientador: Robert Wayne Samohyl. 2003. 116 f. Tese (Doutorado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2003. Disponível em: https://repositorio.ufsc.br/xmlui/handle/123456789/85317. Acesso em: 13 mar. 2023.

ALMEIDA, Silvia dos Santos. **Calibração absoluta funcional sem a suposição de normalidade**. Orientadora: Claudia Regina Oliveira de Paiva Lima. 1999. 77 f. Dissertação (Mestrado em Estatística) - Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco, Recife, 1999.

AOKI, Reiko; BOLFARINE, Heleno; SINGER, Julio M. Null intercept measurement error regression models. **Test**, v. 10, n. 2, p. 441-457, 2001. DOI 10.1007/BF02595707. Disponível em: https://ideas.repec.org/a/spr/testjl/v10y2001i2p441-457.html. Acesso em: 13 mar. 2023.

BARTLETT, Maurice Stevenson. Fitting a straight line when both variables are subject to error. **Biometrics**, v. 5, n. 3, p. 207-212, 1949. DOI 10.2307/3001936. Disponível em: https://www.jstor.org/stable/3001936. Acesso em: 13 mar. 2023.

BELL, Stephanie. **Measurement good practice guide No. 11 (issue 2)**: a beginner's guide to uncertainty of measurement. Teddington: Crown, 1999. Disponível em: http://www.demarcheiso17025.com/private/A%20Beginner%92s%20Guide%20to%20Uncertainty%20of%20Measurement%20.pdf. Acesso em: 13 dez. 2021.

BUONACCORSI, Jonh P. **Measurement error:** models, methods and applications. Boca Raton: CRC Press, 2010.

CARRASCO, Jalmar Manuel Farfán. **Modelos de regressão beta com erro nas variáveis**. Orientadora: Silvia Lopes de Paula Ferrari. 2012. 117 f. Tese (Doutorado em Ciências) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2012. DOI

10.11606/T.45.2012.tde-15082012-093632. Disponível em: https://teses.usp.br/teses/disponiveis/45/45133/tde-15082012-093632/pt-br.php. Acesso em: 13 mar. 2023.

CARROLL, Raymond J. *et al*. **Measurement error in nonlinear models:** a modern perspective. 2nd ed. Boca Raton: Chapman & Hall/CRC, 2006.

CHENG, Chi-Lun; VAN NESS, John W. On estimating linear relationships when both variables are subject to errors. **Journal of the Royal Statistical Society**. Series B (Methodological), v. 56, n.1, p. 167-183. Disponível em: https://www.jstor.org/stable/2346037. Acesso em: 13 mar. 2023.

CHENG, Chi-Lun; VAN NESS, John W. **Statistical regression with measurement error**. London: Oxford University Press. 1999.

CUPPARI, Lílian. **Nutrição clínica no adulto**. 3. ed. São Paulo: Manole, 2018.

DUARTE, Antônio Cláudio Goulart. **Avaliação nutricional**: aspectos clínicos e laboratoriais. São Paulo: Atheneu, 2007.

FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade; TOLEDO, Geraldo Luciano. **Estatística aplicada**. 3. ed. São Paulo: Atlas, 2010.

FREIRE, Clarice Azevedo de Luna *et al*. **Análise de modelos de regressão linear**: com aplicações. 2. ed. Campinas: Editora Unicamp, 2008.

FULLER, Wayne. **Measurement error models**. New York: John Wiley, 1987.

LIMA, Claudia Regina Oliveira de Paiva. **Calibração absoluta com erros nas variáveis**. Orientador: Heleno Bolfarine. 1996. 223 f. Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1996. DOI 10.11606/T.45.1996.tde-20210729-010547. Disponível em: https://www.teses.usp.br/teses/disponiveis/45/45133/tde-20210729-010547/en.php. Acesso em: 13 mar. 2023.

LINNA, Kenneth W.; WOODALL, William H. Effect of measurement error on shewhart control charts. **Journal of Quality Technology**, v. 33, n. 2, p. 213-222, 2001. DOI 10.1080/00224065.2001.11980068. Disponível em: https://www.tandfonline.com/doi/abs/10.1080/00224065.2001.11980068. Acesso em: 13 mar. 2023.

MORAN, Patrick A. Estimating structural and functional relationships, **Journal of Multivariate Analysis**, v. 1, n. 2 p. 232-255, 1971. Disponível em: https://www.sciencedirect.com/science/article/pii/0047259X71900133. Acesso em: 13 mar. 2023.

REIS, Cássio Pinho dos; TORRES, Rodrigo Valente; ALMEIDA, Silvia dos Santos de. O problema da estimação em modelos de calibração linear com erro nas variáveis. *In*: SIMPÓSIO DE PESQUISA OPERACIONAL E LOGÍSTICA DA MARINHA, 12., 2009, Rio de Janeiro. Anais **[...]**. Rio de Janeiro: CASNAV, 2009. Disponível em: https://www.marinha.mil.br/spolm/sites/www.marinha.mil.br.spolm/files/080_0.pdf.. Acesso em: 13 mar. 2023.

ROSNER, Bernad; WILLETT, Walter C.; SPIEGELMAN, Donna. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. **Statistics in Medicine**, v. 8, n. 9, p. 1051-1073, 1989. DOI 10.1002/sim.4780080905. Disponível em: https://pubmed.ncbi.nlm.nih.gov/2799131 /. Acesso em: 13 mar. 2023.

STEFANSKI, Leonard A. The effects of measurement error on parameter estimation. **Biometrika**, v. 72, n. 3 p. 583–592, 1985.

STEFANSKI, Leonard A.; CARROLL, Raymond J. Covariate measurement error in logistic regression. **The Annals of Statistics**, v. 13, n. 4, p. 1335-1351, 1985. DOI 10.1214/aos/1176349741. Disponível em: https://projecteuclid.org/journals/annals-of-statistics/volume-13/issue-4/Covariate-Measurement-Error-in-Logistic-Regression/10.1214/aos/1176349741.full. Acesso em: 13 mar. 2023.

WALD, Abraham. The fitting of straight lines if both variables are subject to error. **The Annals of Mathematical Statistics**, v.11, n. 3, p. 284-300, 1940. DOI 10.1214/aoms/1177731868. Disponível em: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-11/issue-3/The-Fitting-of-Straight-Lines-if-Both-Variables-are-Subject/10.1214/aoms/1177731868.full. Acesso em: 13 mar. 2023.