

VARIÂNCIAS E COVARIÂNCIAS ASSINTÓTICAS DOS SEGUNDOS MOMENTOS AMOSTRAIS NO MODELO DE CALIBRAÇÃO ESTRUTURAL ABSOLUTA SEM SUPOSIÇÃO DE NORMALIDADE

Manoel DOMINGOS FILHO¹
Claudia Regina Oliveira de Paiva LIMA²
Mônica Carneiro SANDOVAL³
Altemir da Silva BRAGA¹

- RESUMO: Um fato que se constata com frequência quando se trabalha com os modelos de calibração, é a necessidade de calcular a esperança, a variância e a covariância dos momentos amostrais. Não muito raro, a obtenção dessas medidas de posição e de dispersão, envolve expressões que dependem de cálculos laboriosos e cujos resultados finais são na verdade expressões assintóticas. Neste artigo, temos como objetivo principal, a obtenção de expressões assintóticas, para as medidas de dispersão (variância e covariância), no modelo de calibração estrutural sem a suposição de normalidade nas variáveis e nos erros de mensuração. Os resultados são apresentados em forma de teoremas, cujas demonstrações levam em consideração o cálculo de momentos centrados. A obtenção dessas estatísticas é necessária, pois elas são incorporadas nas expressões dos previsores, clássico e inverso no modelo de calibração.
- PALAVRAS-CHAVE: Calibração; expressões assintóticas; previsores.

1 Introdução

O problema da calibração é usualmente caracterizado por duas variáveis, digamos X e Y , supostamente relacionadas por meio de uma função f conhecida,

¹Universidade Federal do Acre – UFAC, Departamento de Matemática e Estatística, CEP: 69915-900, Rio Branco, AC, Brasil. E-mail: *manoelufsc@yahoo.com.br* / *altemirbraga@ufac.br*

²Universidade Federal de Pernambuco – UFPE, Departamento de Estatística, CEP: 50740-540, Recife, PE, Brasil. E-mail: *claudia@ufpe.br*

³Universidade de São Paulo – USP, Instituto de Matemática e Estatística – IME, Departamento de Estatística, CEP: 05315-970, São Paulo, SP, Brasil. E-mail: *monica@ime.usp.br*

de tal forma que Y seja um valor obtido com um erro ε de observação, e X é uma quantidade fixa e controlada sem erro de medida, definida em um intervalo I denominado por intervalo de calibração. Na primeira etapa do experimento de calibração, amostram-se n observações da variável aleatória Y a partir de valores prefixados de X , de modo a se estimar a função f que relaciona as duas variáveis. Portanto, X e Y podem-se relacionar através do modelo $Y = f(X) + \varepsilon$. Na segunda etapa, “a calibração propriamente dita”, seleciona-se uma amostra aleatória de tamanho k ($k \geq 1$) da variável Y digamos Y_0 , correspondente a um único valor de X desconhecido, digamos X_0 . Deseja-se, então, estimar este valor desconhecido X_0 , baseado nas informações do ponto de interesse e da função estimada no experimento de calibração.

Osborne (1991) descreveu que procedimentos desta natureza são utilizados quando existem duas maneiras distintas para se obter uma medida de um determinado fenômeno. Uma usualmente denominada por X , a qual se constitui numa maneira exata, mas demorada e de alto custo, e outra denominada Y , que pode ser obtida de forma rápida e de baixo custo, porém não muito precisa. Sob a suposição de que as duas medidas se relacionam de alguma forma, o experimentador deseja, então, determinar o valor exato da medida, usando, para isto, um método impreciso e rápido. O problema de calibração é, na verdade, uma forma de predição inversa, e muitos autores o denominam de regressão inversa, predição inversa ou problema de discriminação, segundo se constata em Hoadley (1970). Na literatura, o termo calibração foi usado sem distinção para dois tipos de problemas que possuem conceitos diferentes, conduzindo, assim, a modelos estatísticos diferentes. Williams (1969) enfatizou esse problema, os classificando em “Calibração Absoluta” e “Calibração Comparativa”.

Na calibração absoluta, um instrumento ou uma técnica de medida sujeita a erro, é calibrado versus outra medida padrão, que é conhecida, ou então, é feita com erro experimental desprezível. Na calibração comparativa, vários instrumentos ou técnicas de medida são calibrados, e nenhum deles é considerado como medida padrão. Nesse tipo de calibração, o experimentador pretende determinar a relação entre diversos testes ou instrumentos, que fornecem medidas indiretas similares, não existindo, porém, medidas padrões como referência. A teoria desenvolvida nesse artigo, leva em consideração apenas a calibração linear absoluta estrutural, que será chamada simplesmente de calibração.

Formalmente, o modelo de calibração absoluta com erros nas variáveis pode ser definido da seguinte forma:

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

$$Y_{0i} = \alpha + \beta X_0 + \varepsilon_{0i}, \quad i = n + 1, \dots, n + k, \quad (2)$$

em que os erros $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ e $\varepsilon_{1n+1}, \varepsilon_{1n+2}, \dots, \varepsilon_{1n+k}$, são mutuamente independentes e identicamente distribuídos com média zero e variância σ^2 . X_1, X_2, \dots, X_n , são consideradas constantes conhecidas enquanto que α, β, X_0 e σ^2 são parâmetros desconhecidos. Neste caso, os valores de X são obtidos supostamente sem erros de mensuração. Entretanto, se considerarmos que os valores de X

contenham erros de medidas, então o modelo de calibração linear pode ser definido pelas expressões:

$$\begin{aligned} y_i &= \alpha + \beta U_i + \varepsilon_i, \quad i = 1, 2, \dots, n \\ X_i &= U_i + \delta_i, \quad i = 1, 2, \dots, n \\ Y_{0i} &= \alpha + \beta U_0 + \varepsilon_{0i}, \quad i = n + 1, \dots, n + k, \end{aligned} \quad (3)$$

em que os erros $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ e $\varepsilon_{1n+1}, \varepsilon_{1n+2}, \dots, \varepsilon_{1n+k}$ são considerados mutuamente independentes e identicamente distribuídos com média zero e variância σ_ε^2 . Os erros $\delta_1, \delta_2, \dots, \delta_n$ também são mutuamente independentes e identicamente distribuídos com média zero e variância σ_δ^2 .

A equação definida no modelo (1), assim como as duas primeiras equações definidas no modelo (3) dizem respeito apenas à primeira etapa de calibração, enquanto que a equação definida no modelo da equação (2) juntamente com a terceira equação do modelo (3), só diz respeito à segunda etapa de calibração. O modelo (3) é conhecido como modelo de calibração estrutural, quando a variável independente for considerada uma variável aleatória, definida por alguma distribuição de probabilidade. Ou seja, o modelo de calibração estrutural se caracteriza essencialmente, pelo fato das quantidades desconhecidas serem consideradas variáveis aleatórias, e será este o modelo considerado nesse artigo.

Além das condições impostas sobre os erros de mensuração nas variáveis X e Y , estabelecendo que eles sejam mutuamente independentes, faz-se ainda as seguintes suposições adicionais ao modelo (3): (a) os primeiros momentos dos erros de mensuração são todos nulos, isto é, $E(\varepsilon_i) = E(\varepsilon_{0n+1}) = E(\delta_i) = 0$ e, (b) os momentos, $E(\varepsilon_i^j) = m_j^\varepsilon$, $E(\delta_i^j) = m_j^\delta$ para $j = 2, 3, 4$ e $E(U_i^j) = E(U_{0i}^j) = m_j^U$, $j = 1, 2, 3$ e 4 existem e são finitos.

Como o modelo de calibração estrutural é caracterizado pelo fato da variável independente ser considerada uma variável aleatória, é usual referir-se aos estimadores da quantidade desconhecida U_0 como sendo previsores e não estimadores. Desta forma, os previsores, clássico e inverso da quantidade desconhecida U_0 são dados por:

$$\begin{aligned} \hat{U}_{0CQ} &= \frac{Y_0 - \hat{\alpha}_Q}{\hat{\beta}_Q} \\ \hat{U}_{0IQ} &= \hat{\gamma}_Q + \hat{\phi}_Q \bar{Y}_0, \end{aligned} \quad (4)$$

em que $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$ e $\hat{\phi}$ são os estimadores de mínimos quadrados ordinários dos parâmetros α , β , γ e ϕ , respectivamente.

Um critério encontrado na literatura que tem sido utilizado para comparar qual dos previsores acima é o melhor para estimar a quantidade desconhecida U_0 , é o erro quadrático médio, como pode ser visto em Shukla (1972), que é função das variâncias e covariâncias dos segundos momentos amostrais.

2 Metodologia

Para obter os resultados apresentados neste artigo, foi considerado o modelo de calibração linear absoluta estrutural e seus previsores, clássico e inverso, como aqueles definidos em Shukla (1972). Como esses previsores são funções dos segundos momentos amostrais, como pode ser visto em Shukla (1972), Lima (1996) e Domingos Filho (2000), e que o erro quadrático médio dos previsores depende das variâncias e covariâncias amostrais desses segundos momentos, foi então necessário obter tais variâncias e covariâncias. Isto se fez, abrindo os somatórios que envolvem as variáveis, dependente e independente, além dos erros de mensuração. Uma vez resolvido a questão dos somatórios, aplicou-se a definição de esperança matemática para obter os resultados desejados. Pelo fato de se trabalhar com momentos amostrais centrados, foi possível utilizar a função geradora de momentos da distribuição normal, facilitando assim, as manipulações algébricas com as esperanças matemática dos somatórios. Os resultados obtidos são todos assintóticos, sendo que as expressões foram truncadas nos termos de primeira ordem, isto é, os termos com ordem de magnitude n^{-2} foram desprezados.

3 Resultados e discussão

Neste item, serão apresentadas em forma de teoremas, as expressões assintóticas da esperança, variância e covariância dos segundos momentos amostrais no modelo de calibração absoluta estrutural sem normalidade.

Teorema 1: Sob o modelo da equação (3), tem-se que:

$$i) E(S_{xx}) = (\sigma_u^2 + m_2^\delta) + O(n^{-1}),$$

$$\text{Var}(S_{xx}) = \frac{\Gamma_1}{n} + O(n^{-2}) \text{ em que } \Gamma_1 = m_4^u + 4\sigma_u^2 m_2^\delta + m_4^\delta - 4m_1^u m_3^u + 8m_2^u (m_1^u)^2 - 4(m_1^u)^4 - (m_2^u)^2 - (m_2^\delta)^2;$$

$$ii) E(S_{yy}) = (\sigma_u^2 + m_2^\delta) + O(n^{-1}),$$

$$\text{Var}(S_{yy}) = \frac{4\beta^2 \sigma_u^2 m_2^\delta + \beta^4 (m_4^u - (m_2^u)^2) - (m_2^\delta)^2 - 4\beta^4 m_1^u (m_3^u - 2m_1^u m_2^u + (m_1^u)^3) + m_4^\delta}{n} + O(n^{-2});$$

$$iii) E(S_{xy}) = \beta \sigma_u^2 + O(n^{-1}),$$

$$\text{Var}(S_{xy}) = \frac{\beta^2 [m_4^u - 4m_1^u m_3^u + 8m_2^u (m_1^u)^2 - 4m_1^u - (m_2^u)^2] + \beta m_2^u m_2^\delta - \beta m_2^\delta (m_1^u)^2}{n} + \frac{m_2^u m_2^\delta - m_2^\delta (m_1^u)^2 + m_2^\delta m_2^\delta}{n} + O(n^{-2}).$$

Demonstração do item *i*): Será mostrada inicialmente, a expressão da esperança do segundo momento amostral. Note que o segundo momento amostral da variável independente X pode ser escrito como:

$$S_{xx} = \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right).$$

Portanto, segue-se que a esperança matemática da soma acima pode então ser escrita da seguinte forma:

$$E(S_{xx}) = \frac{1}{n} E \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right).$$

Lembrando que a variável independente X é obtida com erros de medidas sendo definida no modelo (3), então com a substituição desses valores na expressão acima, obtém-se a seguinte expressão para a esperança:

$$E(S_{xx}) = n^{-1} E \left(\sum_{i=1}^n (U_i + \delta_i)^2 - n \left\{ n^{-1} \sum_{i=1}^n U_i + n^{-1} \sum_{i=1}^n \delta_i \right\}^2 \right).$$

Para demonstrar a forma funcional da esperança, utilizamos as suposições adicionais do modelo (3), as propriedades da esperança e após algumas manipulações algébricas, chega-se ao resultado desejado.

Note que o segundo momento amostral de X pode também ser escrito da seguinte forma:

$$S_{xx} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = M'_1 - \bar{X}^2,$$

em que, $M'_1 = \frac{1}{n} \sum_{i=1}^n X_i^2$ e $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, fazendo a origem coincidir com a média da população, isto é, tomando $\mu_1 = 0$ obtém-se a seguinte relação:

$$S_{xx} = (M'_1)^2 - 2M'_1\bar{X} + \bar{X}^4. \quad (5)$$

Note também que a variância assintótica do segundo amostral pode ser escrita como:

$$Var(S_{xx}) = E((S_{xx}^2)) - (E(S_{xx}))^2. \quad (6)$$

Além disso, a esperança do segundo momento amostral elevado ao quadrado pode ser obtida, aplicando-se as propriedades da esperança na expressão (5), isto é:

$$E(S_{xx}^2) = E((M'_1)^2) - E(2M'_1\bar{X}) + E(\bar{X}^4).$$

Após algumas álgebras, obtém-se os seguintes resultados da esperança:

$$E((M'_1)^2) = \frac{\mu_4 + (n-1)\mu_2^2}{n},$$

$$E(2M'_1\bar{X}) = \frac{2(\mu_4 + (n-1)\mu_2^2)}{n^2},$$

$$E(\bar{X}^4) = \frac{\mu_4 + 3(n-1)\mu_2^2}{n^3}.$$

Portanto, com a substituição desses resultados na expressão (6) acima, obtém-se a seguinte expressão para a variância:

$$\begin{aligned} \text{Var}(S_{xx}) &= \frac{\mu_4 + (n-1)\mu_2^2}{n} - \frac{2(\mu_4 + (n-1)\mu_2^2)}{n^2} + \\ &\frac{\mu_4 + 3(n-1)\mu_2^2}{n^3} + \frac{\mu_4 + 3(n-1)\mu_2^2}{n^3} - \\ &\left(\sigma^2 - \frac{\sigma^2}{n}\right)^2, \end{aligned}$$

que após algumas simplificações algébricas, se reduz a:

$$\text{Var}(S_{xx}) = \frac{\mu_4 - \mu_2^2}{n} + O(n^{-2}). \quad (7)$$

Assim, precisa-se determinar quem são os valores de μ_4 e μ_2^2 para obter-se uma expressão explícita da variância. Como está se trabalhando com os momentos amostrais centrados, obtêm-se as seguintes relações:

$$\begin{aligned} \mu_1 &= E(X - \mu) = E(X) - \mu = 0 \Rightarrow E(X) = \mu_1 \\ \mu_2 &= E(X - \mu)^2 = E(X^2) - \mu_1^2 \Rightarrow E(X^2) = \mu_2 + \mu_1^2 \\ \mu_3 &= E(X - \mu)^3 = E(X^3) - 3\mu_1\mu_2 - \mu_1^3 \Rightarrow E(X^3) = \mu_3 + 3\mu_1\mu_2 + \mu_1^3 \\ \mu_4 &= E(X - \mu)^4 = E(X^4) - 4\mu_1\mu_3 - 6\mu_1^2\mu_2 - \mu_1^4 \Rightarrow \\ E(X^4) &= \mu_4 + 4\mu_1\mu_3 + 6\mu_1^2\mu_2 + \mu_1^4. \end{aligned}$$

Portanto, o numerador da expressão da variância definida em (3), resulta na seguinte expressão:

$$\mu_4 - \mu_2^2 = E(X^4) - 4\mu_1E(X^3) + 8\mu_1^2E(X^2) - 4\mu_1^4 - [E(X^2)]^2. \quad (8)$$

Observando que os valores da variável independente são obtidos com erros de medidas, isto é, $X_i = U_i + \delta_i$, que o primeiro momento centrado da variável U_i é definido como $E(U) = \mu$ e que os erros de mensuração da variável independente são variáveis aleatórias com primeiro momento nulo, ou seja, $E(\delta) = 0$ e lembrando-se das suposições adicionais apresentadas no modelo (3) tem-se:

$$\begin{aligned} \mu_4 - \mu_2^2 &= E(U_i^4) + 4E(U_i^2)E(\delta_i^2) + E(\delta_i^4) - 4\mu E(U_i^3) - 4\mu^2 E(\delta_i^2) + \\ &+ 8\mu^2 E(U_i^2) - 4\mu^4 - (E(U_i^2))^2 - (E(\delta_i^2))^2. \end{aligned}$$

Assim, calculando cada um dos momentos da expressão acima e substituindo-os na expressão (8) obtém-se o resultado desejado, o qual é:

$$\begin{aligned} \text{Var}(S_{xx}) &= \frac{m_4^u + 4m_2^\delta \sigma_u^2 + m_4^\delta - 4m_1^u m_3^u + 8m_2^u (m_1^u)^2 - 4(m_1^u)^4 - (m_2^u)^2 - (m_2^\delta)^2}{n} \\ &+ O(n^{-2}). \end{aligned}$$

Demonstração de (ii). Assim como fizemos no item anterior, demonstraremos primeiramente, a expressão da esperança, observando-se que o segundo amostral da variável dependente Y poder ser escrito como:

$$S_{yy} = \frac{1}{n} \left(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \right),$$

e que seu valor esperado é dado pela expressão:

$$E(S_{yy}) = \frac{1}{n} E \left(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \right).$$

Lembrando que a variável dependente apresentada no modelo (3) dada por: $Y_i = \alpha + \beta U_i + \varepsilon_i$, tem seu valor esperado escrito da seguinte forma:

$$E(S_{yy}) = \frac{1}{n} E \left(\sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \right),$$

e que seu valor esperado é dado pela seguinte expressão:

$$E(S_{yy}) = n^{-1} E \left(\sum_{i=1}^n (\alpha + \beta U_i + \varepsilon_i)^2 - n \{ \alpha + n^{-1} \beta \sum_{i=1}^n U_i + n^{-1} \sum_{i=1}^n \varepsilon_i \}^2 \right).$$

Utilizando as suposições adicionais do modelo (3), as propriedades da esperança e após algumas manipulações algébricas, chega-se ao resultado desejado.

Para demonstrar a expressão da variância, observa-se que a forma geral para se obter sua expressão pode ser escrita como:

$$Var(S^2) = \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3},$$

ou, simplesmente, despreza-se os termos de ordens maiores ou iguais a 2 para obter:

$$Var(S^2) = \frac{\mu_4 - \mu_2^2}{n} + O(n^{-2}).$$

Se fizer $S^2 = S_{yy}$, obte-se a variância do segundo momento amostral de Y , em função do segundo e do quarto momento centrados, a qual pode ser escrita como:

$$Var(S_{yy}) = \frac{\mu_4 - \mu_2^2}{n} + O(n^{-2}). \quad (9)$$

Como $\mu_4 - \mu_2^2 = E(Y^4) - 4\mu E(Y^3) + 8\mu^2 E(Y^2) - 4\mu^4 - (E(Y^2))^2$, então, seguindo o mesmo raciocínio utilizado para encontrar a variância do segundo amostral da variável independente X , S_{xx} tem-se:

$$\begin{aligned}\mu_1 &= E(Y), \\ \mu_2 &= E(Y^2) - \mu_1^2 \Rightarrow E(Y^2) = \mu_2 + \mu_1^2, \\ \mu_3 &= E(Y^3) - 3\mu_1\mu_2 - \mu_1^3, \Rightarrow E(Y^3) = \mu_3 + 3\mu_1\mu_2 + \mu_1^3 \\ \mu_4 &= E(Y^4) - 4\mu_1\mu_3 - 6\mu_1^2\mu_2 - \mu_1^4 \Rightarrow E(Y^4) = \mu_4 + 4\mu_1\mu_3 + 6\mu_1^2\mu_2 + \mu_1^4.\end{aligned}$$

Substituindo os momentos de Y por $\alpha + \beta U_i + \varepsilon_i$ em $\mu_4 - \mu_2^2$ e aplicando as propriedades da esperança obtém-se após alguma álgebra, a seguinte expressão:

$$\mu_4 - \mu_2^2 = 4\beta^2 E(U^2)E(\varepsilon^2) + \beta^4 E(U^4) - \beta^4 [E(U^2)]^2 - [E(\varepsilon^2)]^2 - 4\beta^4 \mu_u^4 - 4\beta^4 \mu_u E(U^3) - 4\beta^2 \mu_u^2 E(\varepsilon^2) + 8\beta^4 \mu_u^2 E(U^2) + E(\varepsilon^4).$$

Utilizando-se as suposições adicionais já anunciadas anteriormente, e após algumas manipulações algébricas conclui-se a demonstração.

Demonstração de (iii). Idêntica às demonstrações dos itens (i) e (ii).

Teorema 2: Sob o modelo (3) tem-se que as covariâncias entre os segundos momentos amostrais são dadas pelas seguintes expressões:

$$i) Cov(S_{xx}, S_{xy}) = \frac{\beta(m_4^u - 4m_1^u m_3^u + 8m_2^u (m_1^u)^2 - 4(m_1^u)^4 - (m_2^u)^2) + 2m_2^u \delta \sigma_u^2}{n} + O(n^{-2})$$

$$ii) Cov(S_{yy}, S_{xy}) = \frac{\beta^3(m_4^u - 4m_1^u (m_3^u - 2m_1^u m_2^u + (m_1^u)^3) - (m_2^u)^2) + 2\beta m_2^u \varepsilon \sigma_u^2}{n} + O(n^{-2})$$

$$iii) Cov(S_{xx}, S_{yy}) = \frac{\beta^2(m_4^u - 4m_1^u m_3^u + 8(m_1^u)^2 m_2^u - 4(m_1^u)^4 - (m_2^u)^2)}{n} + O(n^{-2}).$$

Demonstração de (i). Escrevendo a expressão da covariância em função dos valores esperados dos segundos momentos amostrais tem-se:

$$Cov(S_{xx}, S_{xy}) = E(S_{xx} S_{xy}) - E(S_{xx}) E(S_{xy}).$$

Desenvolvendo o produto envolvido no cálculo da primeira esperança na igualdade acima, obte-se a seguinte expressão:

$$\begin{aligned}S_{xx} S_{xy} &= n^{-2} \sum X_i^2 \sum X_i Y_i - n^{-3} \sum X_i^2 \sum Y_i \sum X_i - n^3 (\sum X_i)^2 \sum X_i Y_i \\ &+ n^{-4} (\sum X_i)^2 \sum Y_i \sum X_i.\end{aligned}$$

Colocando mais uma vez a variável independente em função de seus erros de mensuração, $X_i = U_i + \delta_i$, escrevendo a variável dependente da forma como definida no modelo (3), $Y_i = \alpha + \beta U_i + \varepsilon_i$ e substituindo esses valores na expressão acima, obtemos, após passar a esperança e utilizando a função geradora de momentos, a seguinte expressão:

$$E(S_{xx} S_{xy}) = \beta \mu_{2u}^2 + \beta \mu_{2u} \mu_{2\delta} + n^{-1} (\beta \mu_{4u} - 3\beta \mu_{2u}^2) + O(n^{-2}).$$

Note que as esperanças matemática dos momentos amostrais S_{xx} e S_{xy} podem ser escritas como:

$$E(S_{xx}) = \sigma_u^2 + \sigma_\delta^2 - \frac{\sigma_u^2 + \sigma_\delta^2}{n} = \mu_{2u} + \mu_{2\delta} - \frac{\mu_{2u} + \mu_{2\delta}}{n},$$

$$E(S_{xy}) = \beta\sigma_u^2 - \frac{\beta\sigma_u^2}{n} = \beta\mu_{2u} - \frac{\beta\mu_{2u}}{n}.$$

Com substituição desses valores na expressão (6), obtém-se o seguinte resultado para a expressão assintótica da covariância amostral.

$$Cov(S_{xx}, S_{xy}) = \frac{\beta(\mu_{4u} - \mu_{2u}^2) + 2\beta\sigma_u^2\sigma_\delta^2}{n} + O(n^{-2}),$$

pois $\mu_2 = \sigma^2$.

Escrevendo o segundo e o quarto momento de U da forma como foi definido no modelo (3), conclui-se a prova. Demonstração de (ii). Procedimento análogo ao item (i), simplesmente substituindo S_{xx} por S_{xy} . Demonstração de (iii). Procedimento análogo ao item (i), simplesmente substituindo S_{xy} por S_{xy} .

3.1 Simulação Monte Carlo

O estudo via simulação de dados tem por objetivo a verificação da robustez dos resultados obtidos analiticamente, uma vez que retiramos a suposição de normalidade tanto nas variáveis como nos erros de mensuração. Utilizando as expressões das variâncias e covariâncias dos segundos momentos amostrais demonstradas nos teoremas (1) e (2) e considerando o modelo de calibração estrutural sem suposição de normalidade nas variáveis e nos erros de mensuração, simulamos os pares (X_i, Y_i) através do modelo $Y_i = \alpha + \beta U_i + \varepsilon_i$. Para tanto, os valores da variável independente X_i foram considerados constantes fixas e definidos por: $X_i = (1, 2, 3, \dots, n)$, onde n é o tamanho da amostra. Os vetores $\varepsilon_i = (\varepsilon_1, \dots, \varepsilon_n, \varepsilon_{0n+1}, \dots, \varepsilon_{0n+k})$ e $\delta_i = (\delta_1, \dots, \delta_n)$ foram gerados a partir de uma distribuição (simétrica) t-Student, ambos com 5 graus de liberdade e com tamanho de amostra $n = 20, 30$ e 40 . O vetor $U_i = (U_1, \dots, U_n)$ foi gerado a partir da diferença entre os valores de X_i e os valores de δ_i , fazendo com que a variável U_i seja também uma variável aleatória com distribuição t-Student. Os valores da variável dependente Y_i foram obtidos a partir da equação de Berkson (1969) a qual é dada pela expressão $Y_i = \alpha + \beta U_i + \varepsilon_i$, sendo que em cada par ordenado os valores de Y_i , correspondam à média aritmética de 5000 replicações, isto é, $Y_i = \frac{\sum_{j=1}^{5000} Y_j}{5000}$. Além disso, os valores das variâncias e covariâncias simuladas são as médias de 5000 replicações. Os resultados da simulação desse processo estão mostrados nas tabelas e gráficos a seguir.

Na Figura 1 são mostrados os formatos das relações entre as variáveis X_i e Y_i , onde se percebe que em todos os casos, que as variáveis realmente se relacionam linearmente como era desejada que acontecesse.

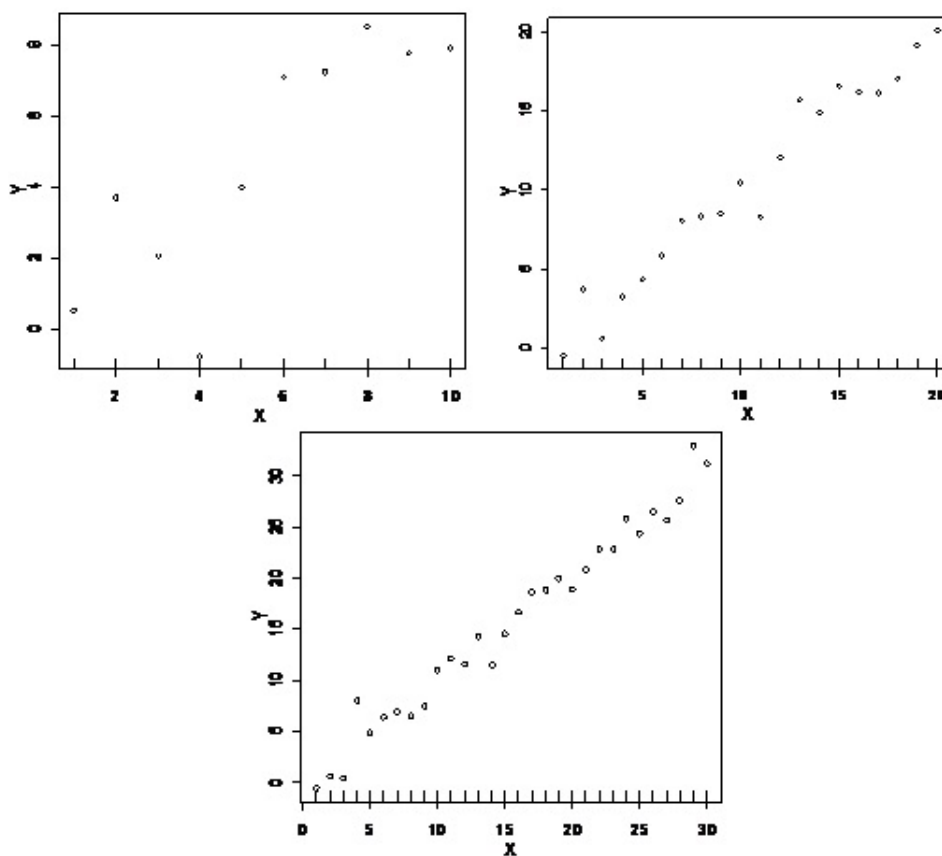


Figura 1 - O gráfico superior à esquerda representa o diagrama de dispersão com tamanho da amostra $n = 10$, o gráfico superior à direita representa o diagrama de dispersão com tamanho da amostra $n = 20$ e o gráfico central representa o diagrama de dispersão com tamanho da amostra $n = 30$.

Observando as Tabelas 1 e 2 se percebe que os valores aproximados (simulados) das variâncias e covariâncias convergem rapidamente para os valores empíricos, mesmo quando o tamanho da amostra é pequeno. Observa-se também que de modo geral a aproximação é pela esquerda, isto é, os valores simulados são menores que os valores empíricos. Pode ser observado também que a maior diferença nas convergências acontece para a covariância entre S_{xy} e S_{yy} , e mesmo neste caso, a aproximação acontece rapidamente com o aumento no tamanho da amostra, o que é desejável. Esses resultados da simulação mostram que as expressões obtidas analiticamente estão realmente corretas, além de revelar a robustez das expressões obtidas no modelo de calibração sem a suposição de normalidade nos dados.

Tabela 1 - Variâncias empíricas e simuladas com $\alpha = 0$, $\beta = 1$ e $\sigma_u^2 = 2$

n	$Var(S_{xx})$		$Var(S_{xy})$		$Var(S_{yy})$	
	Empírico	Aproximado	Empírico	Aproximado	Empírico	Aproximado
10	0,8880	0,8480	1,1038	1,0931	2,0720	2,0269
20	0,4440	0,4340	0,5526	0,5422	1,0378	1,0428
30	0,2960	0,2927	0,3655	0,3631	0,6846	0,6725

Tabela 2 - Covariâncias empíricas e simuladas com $\alpha = 0$, $\beta = 1$ e $\sigma_u^2 = 2$

n	$Cov(S_{xx}, S_{xy})$		$Cov(S_{xx}, S_{yy})$		$Cov(S_{xy}, S_{yy})$	
	Empírico	Aproximado	Empírico	Aproximado	Empírico	Aproximado
10	0,9207	0,9013	0,9696	0,9709	1,5078	1,2900
20	0,4607	0,4487	0,4855	0,4812	0,7552	0,6374
30	0,3058	0,2999	0,3208	0,3224	0,4980	0,4277

Conclusões

A obtenção desses resultados é importante, por que além de propiciar encontrar resultados mais fortes como, por exemplo, o erro quadrático médio dos previsores clássico e inverso de U_0 , eles cobrem uma lacuna existente na literatura, uma vez que esses resultados ainda não haviam sido demonstrados em nenhuma literatura. Na verdade, a literatura disponha apenas de expressões como as aqui demonstradas, mas com a suposição de normalidade, como pode ser observado em Fuller (1987). Consequentemente, obtivemos neste trabalho, resultados mais gerais uma vez que é suficiente ter modelos onde os quatro primeiros momentos existem e sejam finitos, acompanhados das suposições adicionais já mencionadas no corpo do artigo

DOMINGOS FILHO, M.; LIMA, C. R. P.; SANDOVAL, M. C.; BRAGA, A. S. Variances and covarinces asymptotics of the second moments samples in the model of absolute structural calibration without normality assumption. *Rev. Bras. Biom.*, São Paulo, v.29, n.4, p.571-582, 2011.

- **ABSTRACT:** A fact that evidences frequently when works with the calibration models, is the necessity to calculate the hope, the variance and the covariance of the moments samples. Not very rare, the attainment of these measures of position and dispersion, involves expressions that depend on laborious calculations and whose final results they are in the truth asymptotic expressions. In this article, we have as objective main, the attainment of asymptotics expressions, for the measures of dispersion (variance and covariance), in the model of structural calibration without the assumption of normality in the variable and the errors of measures. The results are presented in form of theorems, whose demonstrations take in consideration the calculation of centered moments. The attainment of these statistics is necessary therefore they are incorporated in the expressions of the predictors, classic and inverse in the calibration model.

▪ **KEYWORDS:** *Calibration; asymptotics expressions; predictors.*

Referências

BERKSON, J. Estimation of a linear function for a calibration line: consideration of a recent proposal. *Technometrics*, Washington, v.11, p.649-660, 1969.

DOMINGOS FILHO, M. *Calibração absoluta estrutural sem a suposição de normalidade*. 2000. Dissertação (Mestrado) – Centro de Ciências Exatas e da Natureza, Universidade Federal de Pernambuco, Recife, 2000.

FULLER, W. A. *Measurement error models*. New York: Wiley, 1987. 435p.

HOADLEY, B. A Bayesian look at inverse linear regression. *J. Am. Stat. Assoc.*, Baltimore, v.65, p.356-369, 1970.

LIMA, C. R. O. P. *Calibração absoluta com erros nas variáveis*. 1996. 223f. Tese (Doutorado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1996.

OSBORNE, C. Statistical calibration: a review. *Int. Stat. Rev.*, Edinburgh, v.59, n.3, p.309-336, 1991.

SHUKLA, G. K. On the problem of calibration. *Technometrics*, Washington, v.14, n.4, p.537-553, 1972.

WILLIAMS, E. J. Regression methods in calibration problems. *Bull. Int. Stat. Inst.*, Rome, v.43, p.17-28, 1969.

Recebido em 22.08.2011.

Aprovado após revisão em 23.02.2012.