

Estimação do coeficiente de correlação intraclasse via amostragem por conjuntos ordenados - um estudo por simulação

Cesar Augusto Taconeli ¹, José Luiz Padilha da Silva ², Idemauro Antonio Rodrigues de Lara ³

Abstract: *Ranked set sampling is a cost-effective alternative to simple random sampling. In this study we evaluate the performance of ranked set sampling for estimating the intraclass correlation coefficient. Simulated results showed that ranked set sampling produces less biased, more efficient and robust estimates than its simple random sampling counterpart.*

Keywords: Mean squared error; Monte Carlo simulation; Non-normality; Repeatability; Variance components.

1 Introdução

É comum estudos nas áreas de Ciências da Saúde e Ciências Agrárias envolverem medidas em diferentes escalas. A acurácia dessas medidas, e das consequentes conclusões, depende diretamente da consistência dos medidores (avaliadores) e dos instrumentos de medida. Nesse sentido, a análise da consistência dos medidores é fundamental para a obtenção de medidas confiáveis, sujeitas à menor variação possível (MONTGOMERY, 2009). Em particular, é importante analisar a variação das medidas produzidas por diferentes avaliadores (repetibilidade), e a variação das medidas realizadas por um mesmo avaliador em diferentes momentos (reprodutibilidade). Em ambos os casos, menor variação indica maior consistência.

A análise da consistência de medidores depende da escala dos dados, de tal forma que há métodos específicos nos casos de escalas dicotômicas, ordinais, discretas e contínuas. Há uma considerável variedade de métodos paramétricos e não paramétricos aplicados com tal finalidade, produzindo coeficientes com interpretações e aplicações variadas (BARNHART; HABER; LIN, 2007). Nesse contexto, um importante índice é o coeficiente de correlação intraclasse (*intraclass correlation*-ICC, (BARTKO, 1966)), usualmente considerado na avaliação de medidores submetidos a uma escala contínua de mensuração. Em geral, assume-se distribuição Normal para os efeitos de cada fonte de variação atuante sobre o processo de medida, de tal forma que o coeficiente de correlação intraclasse é calculado como uma função dos respectivos componentes de variância, e as inferências baseadas na suposição de normalidade.

Em alguns estudos, a suposição de normalidade para o efeito de indivíduo pode não se aplicar. Isso pode ocorrer devido à existência de outliers, assimetria, excesso de curtose ou mesmo pela existência de subpopulações de indivíduos. Nessas situações, a suposição de normalidade implica na especificação incorreta do modelo, introduzindo incerteza adicional ao processo de estimação, podendo resultar em maior viés e menor precisão das estimativas produzidas. Como alternativa, poderia-se recorrer a algum modelo probabilístico alternativo, ou recorrer a procedimentos não-paramétricos. De qualquer forma, a identificação do não atendimento da suposição de normalidade nem sempre é algo imediato, particularmente quando o número de indivíduos na amostra é reduzido.

O objetivo deste trabalho é analisar o desempenho da amostragem por conjuntos ordenados (ACO), um delineamento amostral alternativo à amostragem aleatória simples (AAS), na análise da consistência entre avaliadores usando o coeficiente de correlação intraclasse. A amostragem por conjuntos ordenados foi introduzida em McIntyre (1952), aplicada à estimação da produção média de pastagens. É um delineamento amostral recomendável quando a avaliação da variável de interesse é cara, ou de difícil obtenção, mas a ordenação das unidades em pequenas amostras (conjuntos), segundo os possíveis valores da variável de interesse, pode ser feita de maneira simples e econômica. O processo de ordenação pode ser baseado, por exemplo, nos valores de alguma variável concomitante, fortemente correlacionada à variável de interesse, ou por meio do julgamento de um especialista. Nosso foco é comparar as eficiências dos dois delineamentos (i) sob a suposição de normalidade para os efeitos de indivíduos, e (ii) considerando diferentes tipos de desvios dessa suposição.

¹Departamento de Estatística - Universidade Federal do Paraná. e-mail: taconeli@ufpr.br

²Departamento de Estatística - Universidade Federal do Paraná. e-mail: jlpadilha@ufpr.br

³Departamento de Ciências Exatas - ESALQ-USP. e-mail: idemauro@usp.br

2 Amostragem por conjuntos ordenados

A seleção de uma amostra por conjuntos ordenados pode ser descrita conforme os seguintes passos:

1. Selecione n^2 elementos da população de interesse e divida-os aleatoriamente em n conjuntos de tamanho n ;
2. Ranqueie as unidades amostrais dentro de cada conjunto de acordo com o critério de ordenação (barato e acessível) estabelecido;
3. Identifique a unidade ranqueada na posição i do i -ésimo conjunto, for $i = 1, 2, \dots, n$, para compor a amostra final. Somente essas observações devem ser efetivamente medidas em relação à variável de interesse;
4. Os passos 1 a 3 podem ser repetidos m vezes para a obtenção de uma amostra final de tamanho $N = mn$.

Diversos estudos comprovam a maior eficiência da ACO em relação à AAS na estimação de parâmetros como a média, a variância, a proporção e quantis populacionais, dentre vários outros. Alguns desenvolvimentos recentes da ACO e delineamentos amostrais relacionados podem ser vistos em Bouza-Herrera e Al-Omari (2018).

3 Coeficiente de correlação intraclass

Considere um experimento em que J avaliadores (juizes) são requisitados a mensurar K indivíduos. Vamos admitir que cada avaliador produza uma única aferição para cada indivíduo, e que o interesse seja investigar a consistência inter-avaliadores, isto é, a repetibilidade. Adicionalmente, assumimos que os efeitos de avaliadores e indivíduos são aleatórios, resultando no seguinte modelo:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad (1)$$

em que $\alpha_i \sim N(0, \sigma_{ind}^2)$, $\beta_j \sim N(0, \sigma_{juiz}^2)$ e $\epsilon_{ij} \sim N(0, \sigma_{erro}^2)$ denotam, respectivamente, os efeitos aleatórios de indivíduo, juiz, e o erro. Nesta especificação, estamos assumindo, adicionalmente, a não existência de interação indivíduo-juiz, e a independência mútua entre os efeitos aleatórios.

O coeficiente de correlação intraclass, baseado no modelo de efeitos aleatórios apresentado fica definido por:

$$ICC = \frac{\sigma_{ind}^2}{\sigma_{ind}^2 + \sigma_{juiz}^2 + \sigma_{erro}^2},$$

sendo usualmente estimado por:

$$\widehat{ICC} = \frac{QM_{ind} - QM_{erro}}{QM_{ind} + (J - 1)QM_{erro} + (QM_{juiz} - QM_{erro})/K}, \quad (2)$$

em que:

$$QM_{ind} = \frac{SQ_{ind}}{K - 1}; \quad QM_{juiz} = \frac{SQ_{juiz}}{J - 1}; \quad QM_{erro} = \frac{SQ_{erro}}{(K - 1)(J - 1)}, \quad (3)$$

sendo SQ_{ind} , SQ_{juiz} e SQ_{erro} as somas de quadrados associadas a cada fonte de variação.

4 Simulação Monte Carlo

Para investigar a performance da amostragem por conjuntos ordenados na estimação do coeficiente de correlação intraclass, um estudo de simulação Monte Carlo foi conduzido, em que foram considerados diferentes:

- Tamanhos de conjuntos: $n = 2, 3, 4$ e 5 ;
- Número de ciclos: $m = 3, 5$ e 10 ;
- Distribuições para o efeito de indivíduo: normal, t -Student, gamma e normal contaminada;

- Níveis de repetibilidade: $ICC = 0.50, 0.70$ e 0.90 .

Os níveis de repetibilidade especificados foram atingidos variando, simultaneamente, σ_{ind}^2 , σ_{juiz}^2 e σ_{erro}^2 , de maneira que menor correlação intraclasse foi obtida diminuindo σ_{juiz}^2 e σ_{erro}^2 , e incrementando σ_{ind}^2 . Para os casos em que a distribuição para o efeito dos indivíduos não é normal, os parâmetros da verdadeira distribuição foram modificados de maneira a se estabelecer os valores especificados de ICC.

Para cada combinação dos parâmetros de simulação apresentados, nós simulamos 10.000 amostras via ACO e via AAS (fixando o mesmo tamanho final de amostra $N = mn$, para fins de comparação), e a cada simulação as estimativas de ICC produzidas por ambos os delineamentos foram calculadas. Para isso, o modelo apresentado em (1) foi ajustado aos dados simulados a cada rodada. Ajustado o modelo, obtivemos ICC usando o estimador apresentado em (2). Os resultados foram sumarizados por meio do erro quadrático médio e viés, calculados com base nos resultados simulados. Todas as simulações foram realizadas no software estatístico R (R Core Team, 2017), e o pacote lme4 (BATES et al., 2015) foi usado para ajuste dos modelos e estimação da correlação intraclasse.

5 Resultados e discussão

As Tabelas 1 a 4 apresentam os resultados da simulação para efeitos aleatórios de indivíduos seguindo distribuição normal, t-Student, gamma e normal contaminada, respectivamente.

Tabela 1: Erros quadráticos médios (e vícios) simulados para o caso de efeitos de indivíduos com distribuição normal

m	n	ICC=0.50		ICC=0.75		ICC=0.90	
		SRS	RSS	SRS	RSS	SRS	RSS
3	2	0.0317 (-0.0276)	0.0312 (-0.0140)	0.0276 (-0.0483)	0.0248 (-0.0357)	0.0117 (-0.0365)	0.0106 (-0.0294)
	3	0.0235 (-0.0118)	0.0219 (0.0038)	0.0174 (-0.0288)	0.0145 (-0.0158)	0.0059 (-0.0211)	0.0047 (-0.0121)
	4	0.0191 (-0.0041)	0.0172 (0.0115)	0.0131 (-0.0179)	0.0106 (-0.0051)	0.0040 (-0.0138)	0.0031 (-0.0068)
	5	0.0169 (-0.0006)	0.0151 (0.0150)	0.0109 (-0.0145)	0.0086 (-0.0012)	0.0033 (-0.0116)	0.0024 (-0.0043)
5	2	0.0214 (-0.0096)	0.0211 (-0.0013)	0.0153 (-0.0237)	0.0143 (-0.0158)	0.0053 (-0.0193)	0.0045 (-0.0144)
	3	0.0164 (-0.0006)	0.0159 (0.0090)	0.0107 (-0.0127)	0.0098 (-0.0045)	0.0033 (-0.0117)	0.0028 (-0.0063)
	4	0.0143 (0.0055)	0.0137 (0.0140)	0.0090 (-0.0083)	0.0078 (-0.0008)	0.0025 (-0.0079)	0.0021 (-0.0037)
	5	0.0134 (0.0042)	0.0124 (0.0144)	0.0078 (-0.0051)	0.0068 (0.0027)	0.0021 (-0.0056)	0.0018 (-0.0018)
10	2	0.0145 (0.0039)	0.0144 (0.0096)	0.0091 (-0.0078)	0.0087 (-0.0044)	0.0025 (-0.0081)	0.0025 (-0.0066)
	3	0.0125 (0.0082)	0.0122 (0.0123)	0.0076 (-0.0034)	0.0069 (0.0004)	0.0019 (-0.0043)	0.0017 (-0.0020)
	4	0.0115 (0.0097)	0.0109 (0.0151)	0.0064 (-0.0004)	0.0061 (0.0029)	0.0017 (-0.0028)	0.0016 (-0.0010)
	5	0.0109 (0.0111)	0.0105 (0.0170)	0.0059 (0.0020)	0.0057 (0.0042)	0.0016 (-0.0021)	0.0014 (-0.0002)

Tabela 2: Erros quadráticos médios (e vícios) simulados para o caso de efeitos de indivíduos com distribuição t-Student

m	n	ICC=0.50		ICC=0.75		ICC=0.90	
		SRS	RSS	SRS	RSS	SRS	RSS
3	2	0.0364 (-0.0418)	0.0350 (-0.0251)	0.0380 (-0.0765)	0.0335 (-0.0620)	0.0267 (-0.0835)	0.0230 (-0.0721)
	3	0.0261 (-0.0233)	0.0244 (-0.0029)	0.0236 (-0.0484)	0.0202 (-0.0344)	0.0151 (-0.0577)	0.0124 (-0.0483)
	4	0.0219 (-0.0116)	0.0201 (0.0029)	0.0182 (-0.0343)	0.0151 (-0.0210)	0.0102 (-0.0439)	0.0079 (-0.0348)
	5	0.0192 (-0.0076)	0.0168 (0.0079)	0.0150 (-0.0286)	0.0122 (-0.0149)	0.0082 (-0.0370)	0.0063 (-0.0289)
5	2	0.0247 (-0.0191)	0.0242 (-0.0102)	0.0216 (-0.0436)	0.0205 (-0.0374)	0.0129 (-0.0510)	0.0118 (-0.0469)
	3	0.0190 (-0.0065)	0.0179 (0.0050)	0.0150 (-0.0276)	0.0136 (-0.0179)	0.0083 (-0.0376)	0.0072 (-0.0323)
	4	0.0165 (-0.0002)	0.0151 (0.0076)	0.0124 (-0.0193)	0.0107 (-0.0125)	0.0065 (-0.0300)	0.0052 (-0.0245)
	5	0.0146 (0.0021)	0.0137 (0.0094)	0.0105 (-0.0137)	0.0090 (-0.0074)	0.0051 (-0.0261)	0.0040 (-0.0195)
10	2	0.0166 (-0.0017)	0.0164 (0.0030)	0.0122 (-0.0192)	0.0121 (-0.0174)	0.0060 (-0.0290)	0.0059 (-0.0272)
	3	0.0140 (0.0043)	0.0135 (0.0083)	0.0095 (-0.0098)	0.0089 (-0.0076)	0.0044 (-0.0213)	0.0040 (-0.0193)
	4	0.0127 (0.0077)	0.0122 (0.0131)	0.0083 (-0.0079)	0.0077 (-0.0024)	0.0035 (-0.0171)	0.0031 (-0.0146)
	5	0.0116 (0.0074)	0.0111 (0.0124)	0.0076 (-0.0036)	0.0069 (0.0004)	0.0031 (-0.0142)	0.0027 (-0.0105)

Tabela 3: Erros quadráticos médios (e vícios) simulados para o caso de efeitos de indivíduos com distribuição gamma

m	n	ICC=0.50		ICC=0.75		ICC=0.90	
		SRS	RSS	SRS	RSS	SRS	RSS
3	2	0.0344 (-0.0336)	0.0327 (-0.0191)	0.0308 (-0.0542)	0.0279 (-0.0430)	0.0150 (-0.0465)	0.0119 (-0.0351)
	3	0.0243 (-0.0160)	0.0225 (0.0009)	0.0202 (-0.0364)	0.0162 (-0.0200)	0.0076 (-0.0280)	0.0057 (-0.0178)
	4	0.0195 (-0.0067)	0.0180 (0.0094)	0.0149 (-0.0247)	0.0118 (-0.0083)	0.0052 (-0.0189)	0.0034 (-0.0094)
	5	0.0177 (-0.0021)	0.0155 (0.0106)	0.0126 (-0.0195)	0.0098 (-0.0048)	0.0040 (-0.0148)	0.0028 (-0.0063)
5	2	0.0226 (-0.0134)	0.0221 (-0.0047)	0.0171 (-0.0288)	0.0167 (-0.0222)	0.0066 (-0.0243)	0.0059 (-0.0199)
	3	0.0176 (-0.0031)	0.0165 (0.0064)	0.0124 (-0.0175)	0.0110 (-0.0106)	0.0040 (-0.0148)	0.0033 (-0.0099)
	4	0.0151 (0.0025)	0.0137 (0.0101)	0.0101 (-0.0107)	0.0087 (-0.0024)	0.0030 (-0.0107)	0.0025 (-0.0057)
	5	0.0138 (0.0062)	0.0128 (0.0148)	0.0090 (-0.0098)	0.0073 (-0.0001)	0.0025 (-0.0072)	0.0021 (-0.0038)
10	2	0.0155 (0.0022)	0.0149 (0.0061)	0.0100 (-0.0103)	0.0097 (-0.0057)	0.0030 (-0.0100)	0.0027 (-0.0074)
	3	0.0126 (0.0053)	0.0126 (0.0098)	0.0079 (-0.0052)	0.0075 (-0.0013)	0.0021 (-0.0058)	0.0020 (-0.0040)
	4	0.0116 (0.0102)	0.0112 (0.0155)	0.0069 (-0.0019)	0.0064 (0.0008)	0.0019 (-0.0040)	0.0017 (-0.0017)
	5	0.0113 (0.0111)	0.0107 (0.0150)	0.0066 (-0.0007)	0.0061 (0.0027)	0.0017 (-0.0032)	0.0015 (-0.0009)

Tabela 4: Erros quadráticos médios (e vícios) simulados para o caso de efeitos de indivíduos com distribuição normal contaminada

m	n	ICC=0.50		ICC=0.75		ICC=0.90	
		SRS	RSS	SRS	RSS	SRS	RSS
3	2	0.0357 (-0.0676)	0.0332 (-0.0518)	0.0388 (-0.0773)	0.0362 (-0.0668)	0.0230 (-0.0725)	0.0203 (-0.0632)
	3	0.0277 (-0.0117)	0.0261 (0.0034)	0.0268 (-0.0573)	0.0219 (-0.0421)	0.0133 (-0.0502)	0.0110 (-0.0408)
	4	0.0237 (-0.0070)	0.0211 (0.0049)	0.0197 (-0.0513)	0.0159 (-0.0341)	0.0072 (-0.0248)	0.0054 (-0.0162)
	5	0.0186 (-0.0211)	0.0164 (-0.0070)	0.0129 (-0.0262)	0.0106 (-0.0132)	0.0049 (-0.0171)	0.0035 (-0.0091)
5	2	0.0263 (-0.0117)	0.0253 (-0.0039)	0.0215 (-0.0332)	0.0204 (-0.0253)	0.0085 (-0.0303)	0.0082 (-0.0271)
	3	0.0206 (-0.0034)	0.0196 (0.0052)	0.0168 (-0.0298)	0.0151 (-0.0211)	0.0050 (-0.0189)	0.0043 (-0.0140)
	4	0.0174 (-0.0109)	0.0158 (-0.0025)	0.0111 (-0.0010)	0.0102 (0.0057)	0.0044 (-0.0190)	0.0036 (-0.0140)
	5	0.0166 (0.0290)	0.0156 (0.0373)	0.0094 (-0.0143)	0.0080 (-0.0068)	0.0033 (-0.0105)	0.0025 (-0.0056)
10	2	0.0183 (-0.0033)	0.0188 (0.0005)	0.0130 (-0.0232)	0.0125 (-0.0193)	0.0038 (-0.0161)	0.0038 (-0.0150)
	3	0.0153 (0.0008)	0.0148 (0.0057)	0.0097 (-0.0195)	0.0092 (-0.0169)	0.0028 (-0.0049)	0.0026 (-0.0033)
	4	0.0130 (0.0076)	0.0129 (0.0132)	0.0084 (-0.0066)	0.0078 (-0.0024)	0.0020 (-0.0012)	0.0018 (0.0005)
	5	0.0133 (0.0171)	0.0126 (0.0229)	0.0073 (0.0017)	0.0068 (0.0044)	0.0019 (-0.0035)	0.0017 (-0.0012)

Alguns resultados merecem destaque:

- O estimador do ICC via ACO consistentemente apresentou menor erro quadrático médio que o correspondente estimador obtido via AAS, tanto para o caso em que os indivíduos têm distribuição normal quanto para os outros cenários, baseados nas demais distribuições;
- Na maior parte dos casos o ICC é subestimado, independente da distribuição dos indivíduos. No entanto, pode-se verificar o menor viés produzido via ACO em relação à estimação via AAS;
- O erro quadrático médio, sob ambos os delineamentos, diminui à medida que o valor de ICC se afasta de 0.5. O mesmo não é verificado em relação ao viés;
- Sob ACO, erro quadrático médio e viés diminuem à medida que tanto n quanto m aumentam.

6 Considerações finais

Neste trabalho tratamos da estimação do coeficiente de correlação intraclasse quando os indivíduos são selecionados para compor a amostra via amostragem por conjuntos ordenados. Diferentes distribuições de probabilidades foram consideradas para o efeito de indivíduo. Através de um extenso estudo de simulação, pudemos verificar que a ACO supera a AAS na estimação de ICC, produzindo estimativas com menor erro quadrático médio e menos viesadas, para todas as distribuições consideradas. Os resultados indicam o potencial da ACO em estudos envolvendo a análise da acurácia de medidores e instrumentos de medida. Como temas para estudos futuros, sugere-se a avaliação de intervalos de confiança para ICC usando ACO, a avaliação do impacto de erros no processo de ordenação dos conjuntos, e a investigação da performance da ACO na estimação de outras medidas de consistência, como a reprodutibilidade (consistência intra-avaliador). Adicionalmente, recomenda-se a aplicação da metodologia em bases de dados reais, para fins ilustrativos.

Referências

- BARNHART, H. X.; HABER, M. J.; LIN, L. I. An overview on assessing agreement with continuous measurements. *Journal of biopharmaceutical statistics*, Taylor & Francis, v. 17, n. 4, p. 529–569, 2007.
- BARTKO, J. J. The intraclass correlation coefficient as a measure of reliability. *Psychological reports*, SAGE Publications Sage CA: Los Angeles, CA, v. 19, n. 1, p. 3–11, 1966.
- BATES, D. et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, v. 67, n. 1, p. 1–48, 2015.
- BOUZA-HERRERA, C. N.; AL-OMARI, A. I. F. *Ranked Set Sampling: 65 Years Improving the Accuracy in Data Gathering*. [S.l.]: Academic Press, 2018.
- MCINTYRE, G. A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, CSIRO, v. 3, n. 4, p. 385–390, 1952.
- MONTGOMERY, D. C. *Statistical quality control*. [S.l.]: Wiley New York, 2009. v. 7.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>.