

## PRECISÃO RELATIVA PARA ESTIMADORES DA VARIÂNCIA DE UMA DISTRIBUIÇÃO NORMAL SOB O DELINEAMENTO POR CONJUNTOS ORDENADOS

Clécia Aparecida GARCIA<sup>1</sup>  
Maria Cecília Mendes BARRETO<sup>2</sup>

- **RESUMO:** Em muitos levantamentos, os indivíduos selecionados para compor a amostra são difíceis de serem mensurados ou o processo de mensuração apresenta alto custo. Quando antes da mensuração é possível se fazer uma ordenação de custo reduzido nos indivíduos selecionados, a amostragem por conjuntos ordenados (McIntyre, 1952) apresenta uma eficiência maior quando comparada a amostragem aleatória simples na estimação de diversos parâmetros, entre eles a média populacional, os quantis e a função densidade. Neste trabalho, é apresentado um estudo da precisão relativa simulada de três estimadores da variância populacional de uma distribuição normal, obtidos a partir da amostragem por conjuntos ordenados: o estimador não-paramétrico (Stokes, 1980), o estimador natural (Barnett & Moore, 1997) e o estimador da Stokes modificado (Yu et al., 1999). É apresentada também uma aplicação a dados reais referente a produção diária de leite de vacas de mesma raça, mesma parição e mesma nutrição. A precisão relativa também foi calculada para este conjunto de dados a partir de 14 amostras por conjuntos ordenados utilizando os três estimadores. Para a precisão relativa simulada, os resultados mostraram que o estimador modificado da Stokes, por ser não viciado e apresentar a menor variância, tem o melhor desempenho para amostras pequenas. Entretanto, conforme aumenta o tamanho amostral, os estimadores natural e não paramétrico apresentam um vício simulado praticamente igual a zero e erro médio quadrático bem próximo do estimador modificado da Stokes. Pode-se dizer que o estimador da Stokes modificado apresentou a maior eficiência simulada para os diversos tamanhos de amostras e repetições estudadas. Os resultados obtidos no conjunto de dados sobre a produção de leite, mesmo com ordenação imperfeita, foram similares ao estudo por simulação.
- **PALAVRAS-CHAVE:** Amostragem por conjuntos ordenados; estimadores da variância; distribuição normal; precisão relativa.

---

<sup>1</sup>Centro de Métodos Quantitativos, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo - USP, CEP 14049-900, Ribeirão Preto, SP, Brasil. E-mail: *cleciagarcia@yahoo.com.br*

<sup>2</sup>Departamento de Estatística, Universidade Federal de São Carlos - UFSCar, Caixa Postal 676, CEP 13565-905, São Carlos, SP, Brasil. E-mail: *cbarreto@power.ufscar.br*

## 1 Introdução

O delineamento por conjuntos ordenados ou amostragem por conjuntos ordenados (*ranked set sampling* - RSS) foi elaborado por McIntyre em 1952. Por meio desse delineamento obteve-se um estimador não viciado da média populacional,  $\mu_x$ , sendo este mais preciso do que o obtido a partir de uma amostra aleatória simples.

Posteriormente, Stokes & Sager (1988) apresentaram um estimador da densidade populacional que é não viciado e com menor variância que o utilizado em amostragem aleatória simples. No caso da estimação de quantis, Chen (2000) apresentou um estimador que é assintoticamente não viciado e com erro médio quadrático menor que o utilizado em amostragem aleatória simples.

A amostragem por conjuntos ordenados pode ser descrita em três etapas:

Primeiro, selecionam-se  $n$  amostras aleatórias simples, cada uma contendo  $n$  elementos; a seguir considera-se em cada amostra sua ordenação crescente por meio de inspeção visual ou por alguma variável concomitante; após a ordenação, mensura-se um elemento de cada amostra: na primeira amostra, o menor elemento, na segunda, o segundo menor e assim por diante, até que o maior seja mensurado na  $n$ -ésima amostra.

Este procedimento pode ser repetido  $m$  vezes, até que um total de  $mn^2$  elementos tenham sido retirados da população, mas somente  $N = mn$  tenham sido mensurados. Uma amostra obtida a partir da amostragem por conjuntos ordenados pode ser indicada por  $X_{(11)1}, X_{(22)1}, \dots, X_{(nn)1}, \dots, X_{(11)m}, X_{(22)m}, \dots, X_{(nm)m}$ , onde  $X_{(ii)j}$  corresponde a  $j$ -ésima réplica da  $i$ -ésima estatística de ordem na  $n$ -ésima amostra,  $i = 1, \dots, n$  e  $j = 1, \dots, m$ . A amostra por conjuntos ordenados (ACO) corresponde a um conjunto de estatísticas de ordem, cujos elementos são independentes, já que provêm de amostras escolhidas independentemente, porém não são identicamente distribuídos.

Na maioria das vezes, supõe-se que o processo de ordenação é acurado, ou seja, a ordenação é perfeita. Isso equivale a dizer que, se todos os  $n$  elementos da  $i$ -ésima amostra fossem de fato mensurados, o julgamento do  $i$ -ésimo menor elemento estaria correto (Barreto & Barnett, 1999). Se isto não ocorre tem-se a ordenação imperfeita.

Quando o objetivo é estimar a variância de uma população normal, pode-se utilizar o estimador de Stokes (1980), o estimador natural de Barnett & Moore (1997) e o estimador modificado de Yu et al. (1999).

## 2 Desenvolvimento

Sob o delineamento por conjuntos ordenados em uma população normal, a estimação da variância pode ser feita por três estimadores:

### 1. Estimador não paramétrico da Stokes (1980)

$$\hat{\sigma}^2 = \sum_{j=1}^m \sum_{i=1}^n \frac{(X_{(ii)j} - \mu'_{rep})^2}{mn - 1},$$

onde

$$\mu'_{rep} = \sum_{j=1}^m \sum_{i=1}^n X_{(ii)j} / (mn).$$

Stokes (1980) mostrou que este estimador é assintoticamente não viciado para a variância populacional  $\sigma_x^2$ ;

### 2. Estimador natural baseado em Barnett & Moore (1997)

Em busca de maior eficiência, vários autores obtiveram estimadores lineares não viciados ótimos (BLUES - *Best linear unbiased estimators*), que na classe dos estimadores que são combinações lineares dos termos  $X_{(ii)}$  possuem menor variância.

Barnett & Moore (1997) encontraram os BLUES dos parâmetros da família de distribuição locação e escala. Seguindo o mesmo procedimento de Barnett & Moore (1997) para a obtenção dos BLUES de  $\mu$  e  $\sigma$ , Raminelli (2003) estendeu os resultados considerando réplicas. Para o caso específico do parâmetro de escala,  $\sigma$ , o estimador é dado por

$$\sigma_{rep}^* = \sum_{i=1}^n \eta_i \bar{X}_{i(\cdot)},$$

onde

$$\bar{X}_{i(\cdot)} = \sum_{j=1}^m X_{(ii)j} / m \text{ e } \eta_i = \frac{\frac{1}{\nu_i} \left[ \alpha_i \left( \sum_{j=1}^n \frac{1}{\nu_j} \right) - \left( \sum_{j=1}^n \frac{\alpha_j}{\nu_j} \right) \right]}{\Delta},$$

sendo

$$\Delta = \left( \sum_{j=1}^n \frac{1}{\nu_j} \right) \left( \sum_{j=1}^n \frac{\alpha_j^2}{\nu_j} \right) - \left( \sum_{j=1}^n \frac{\alpha_j}{\nu_j} \right)^2$$

e,  $\alpha_i$  e  $\nu_i$  são, respectivamente, a esperança e a variância da  $i$ -ésima estatística de ordem padronizada em uma população com distribuição normal padronizada.

Como o interesse deste trabalho está na estimação da variância populacional de uma distribuição normal, então um estimador natural pode ser dado por  $\sigma_{rep}^{*2}$ . Garcia (2004) mostrou numericamente que este estimador é assintoticamente não viciado para variância populacional;

### 3. Estimador da variância modificado

Considerando que o estimador  $\hat{\sigma}^2$ , introduzido por Stokes (1980), é assintoticamente não viciado, Yu et al. (1999) propuseram uma versão não viciada que é dada por

$$\hat{\sigma}_{MS}^2 = \frac{1}{mn - 1 + \sum_{i=1}^n \frac{\alpha_i^2}{n}} \sum_{j=1}^m \sum_{i=1}^n (X_{(ii)j} - \mu'_{rep})^2,$$

onde

$$\mu'_{rep} = \sum_{j=1}^m \sum_{i=1}^n X_{(ii)j} / (mn)$$

e  $\alpha_{(i)}$  é a esperança da  $i$ -ésima estatística de ordem padronizada em uma população com distribuição normal padronizada. Pode-se notar que a diferença entre  $\hat{\sigma}^2$  e  $\hat{\sigma}_{MS}^2$  está no denominador, com o acréscimo de  $\sum_{i=1}^n \alpha_i^2/n$  ao denominador de  $\hat{\sigma}^2$ .

O desempenho dos três estimadores da variância pode ser baseado na precisão relativa simulada. Dell & Clutter (1972), definiram a precisão relativa do estimador  $\hat{\theta}_1$  em relação a  $\hat{\theta}_2$ , como a razão entre a variância de  $\hat{\theta}_2$  em relação a variância de  $\hat{\theta}_1$ . Stokes (1980) definiu a precisão relativa em termos da razão entre o erro médio quadrático no caso em que um dos estimadores ou ambos,  $\hat{\theta}_1$  e  $\hat{\theta}_2$ , forem viciados. Assim, adotou-se aqui, como precisão relativa simulada do estimador  $\hat{\theta}_1$  em relação a  $\hat{\theta}_2$ , tanto a razão entre as variâncias simuladas como a razão entre os erros médios quadráticos simulados.

### 3 Resultados da precisão relativa simulada

Antes de apresentar o estudo por simulação da precisão relativa dos estimadores da variância populacional, é apresentado um estudo para verificar quando o tamanho da simulação estabiliza o valor médio simulado dos três estimadores  $\hat{\sigma}^2$ ,  $\sigma_{rep}^{*2}$  e  $\hat{\sigma}_{MS}^2$ . Na Figura 1 (ver Apêndice) é observado o valor médio simulado para os estimadores,  $\hat{\sigma}^2$ ,  $\sigma_{rep}^{*2}$  e  $\hat{\sigma}_{MS}^2$ , *versus* tamanho das simulações, considerando uma ACO de tamanho  $n = 2$  e  $m = 3$  para uma  $N(0, 1)$ .

De acordo com a Figura 1 concluí-se que a partir de 400 simulações, o valor médio simulado para os estimadores  $\hat{\sigma}^2$ ,  $\sigma_{rep}^{*2}$  e  $\hat{\sigma}_{MS}^2$ , se estabiliza em torno do seu verdadeiro valor. Geralmente, nos estudos por simulação costuma-se gerar 1.000 valores. Logo, para nosso trabalho, considerando a distribuição normal com parâmetros  $\mu = 0$  e  $\sigma^2 = 1$ , o estudo por simulação da precisão relativa consistiu na geração de 1.000 amostras por conjuntos ordenados, para diferentes valores de  $n$  e  $m$ . Os tamanhos de amostras e réplicas estudados foram, respectivamente,  $n = 3, 4, 5, 6$  e  $10$  e  $m = 2, 3, 4, 5$  e  $6$ . A partir das amostras simuladas foram calculados o valor médio, o vício, a variância e o erro médio quadrático de cada um dos estimadores. Em seguida, foram obtidas as precisões relativas:  $PR(\hat{\sigma}^2, \sigma_{rep}^{*2}) = PR1$ ,  $PR(\hat{\sigma}^2, \hat{\sigma}_{MS}^2) = PR2$ ,  $PR(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2) = PR3$ ,  $PREMQ(\hat{\sigma}^2, \sigma_{rep}^{*2}) = PREMQ1$ ,  $PREMQ(\hat{\sigma}^2, \hat{\sigma}_{MS}^2) = PREMQ2$  e  $PREMQ(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2) = PREMQ3$ .

Observa-se na Tabela 1 as precisões relativas simuladas considerando a razão entre as variâncias simuladas e os erros médios quadráticos simulados para diversos tamanhos de amostras e réplicas.

De acordo com a Tabela 1 tem-se que o estimador  $\hat{\sigma}_{MS}^2$  é sempre melhor que  $\hat{\sigma}^2$  e  $\sigma_{rep}^{*2}$  para diferentes tamanhos de  $n$  e  $m$ , porém quanto maior  $n$  e  $m$  mais próximo  $\hat{\sigma}^2$  e  $\sigma_{rep}^{*2}$  estão de  $\hat{\sigma}_{MS}^2$ , ou seja,  $\hat{\sigma}^2$  e  $\sigma_{rep}^{*2}$  são assintoticamente não viciados. Pode-se dizer que o estimador da Stokes modificado,  $\hat{\sigma}_{MS}^2$ , apresentou a maior eficiência simulada para os diversos tamanhos de amostras e repetições estudados.

Tabela 1 - Precisões relativas em função da razão de variâncias  $PR1$ ,  $PR2$  e  $PR3$  e em função da razão dos erros médios quadráticos  $PREMQ1$ ,  $PREMQ2$  e  $PREMQ3$

$n$	$m$	$PR1$	$PR2$	$PR3$	$PREMQ1$	$PREMQ2$	$PREMQ3$
3	2	2,2356	0,8333	0,3727	2,2862	0,8107	0,3546
	3	2,4414	0,8906	0,3648	2,4963	0,8691	0,3482
	4	2,3381	0,9185	0,3929	2,3680	0,9046	0,3820
	5	2,4137	0,9352	0,3874	2,4444	0,9218	0,3772
	6	2,4594	0,9461	0,3847	2,4844	0,9313	0,3748
4	2	1,7590	0,8542	0,4856	1,7565	0,8261	0,4703
	3	1,8793	0,9033	0,4808	1,8755	0,8846	0,4717
	4	1,8054	0,9276	0,5138	1,8005	0,9124	0,5068
	5	1,8090	0,9422	0,5209	1,7966	0,9301	0,5168
	6	1,9249	0,9519	0,4946	1,9197	0,9416	0,4905
5	2	1,5256	0,8718	0,5715	1,5117	0,8660	0,5722
	3	1,6018	0,9147	0,5710	1,5954	0,9152	0,5737
	4	1,6110	0,9359	0,5718	1,6351	0,9427	0,5765
	5	1,5513	0,9489	0,6149	1,5489	0,9437	0,6143
	6	1,6311	0,9574	0,5870	1,6308	0,9630	0,5905
6	2	1,3877	0,8861	0,6384	1,3717	0,8660	0,6313
	3	1,4386	0,9239	0,6421	1,4290	0,9110	0,6375
	4	1,4397	0,9430	0,6550	1,4312	0,9301	0,6499
	5	1,4407	0,9545	0,6625	1,4355	0,9437	0,6575
	6	1,4364	0,9619	0,6697	1,4320	0,9534	0,6660
10	2	1,1287	0,9217	0,8166	1,0977	0,8935	0,8139
	3	1,1335	0,9475	0,8358	1,1193	0,9359	0,8361
	4	1,1468	0,9605	0,8376	1,1329	0,9483	0,8365
	5	1,1088	0,9684	0,8734	1,0979	0,9583	0,8728
	6	1,1393	0,9738	0,8547	1,1306	0,9651	0,8537

Nas Figuras 2 e 3 (ver Apêndice) são observadas, respectivamente, as precisões relativas de  $\hat{\sigma}^2$  em relação a  $\sigma_{rep}^{*2}$  ( $PREMQ1$ ),  $\hat{\sigma}^2$  em relação a  $\hat{\sigma}_{MS}^2$  ( $PREMQ2$ ) e  $\sigma_{rep}^{*2}$  em relação a  $\hat{\sigma}_{MS}^2$  ( $PREMQ3$ ) versus tamanho amostral  $n$  considerando  $m = 2$  e  $m = 6$ . A partir destas, nota-se que conforme aumenta o tamanho amostral as precisões relativas se aproximam de um, ou seja, conforme o tamanho amostral

$n$  ou tamanho  $m$  de réplicas aumenta, os estimadores praticamente igualam-se. Verifica-se que o estimador  $\hat{\sigma}_{MS}^2$  é sempre melhor para qualquer  $n$  e para qualquer  $m$ . As figuras, considerando a precisão relativa com relação a variância, foram omitidas, pois têm forma semelhante a das Figuras 2 e 3.

No estudo da precisão relativa simulada foi constatada a presença de estimativas negativas de  $\sigma_{rep}^*$ . Porém, à medida que o tamanho amostral aumenta foi observado que todos os valores de  $\sigma_{rep}^*$  foram admissíveis. No caso em que o tamanho amostral foi pequeno, os valores de  $\sigma_{rep}^*$  negativos foram desconsiderados para o cálculo do valor médio e demais quantidades apresentadas.

De certa forma esse tipo de situação era esperado uma vez que Cesário (2001) relatou a presença de estimativas negativas de  $\sigma_{rep}^*$  para a construção dos intervalos de confiança *bootstrap* para a média de uma população normal sob RSS, considerando 1.000 simulações.

## 4 Aplicação a dados reais

Com o objetivo de investigar a eficácia do delineamento por conjuntos ordenados, Al-Saleh & Al-Shrafat (2001) estimaram a produção média de leite a partir de um estudo de 402 ovelhas. Os animais foram divididos em 134 conjuntos de tamanho 3 e depois ordenados visualmente (pelo tamanho do úbere do animal) por dois observadores, um experiente e outro inexperiente. A precisão relativa do estimador da média populacional,  $\mu_x$ , obtido a partir da RSS com relação ao obtido a partir da amostragem aleatória simples, foi calculada considerando a ordenação perfeita.

Com base no trabalho de Al-Saleh & Al-Shrafat (2001) e em parceria com o Setor de Bovinocultura de Leite da Fazenda Experimental “Marcelo Mesquita Serva”, pertencente à Universidade de Marília (Unimar), Marília, SP, para o presente trabalho foi elaborado um delineamento por conjuntos ordenados com o objetivo de estimar a variância da produção diária de leite, com base em 22 vacas Puras de Origem (PO) e primíparas (primeira cria).

Os animais de mesma raça, mesma parição e mesma nutrição são alimentados com 12 kg/dia de ração concentrada, composta por casca de soja, farelo de algodão, resíduo de panificação de biscoito, milho moído, uréia, núcleo mineral e vitamínico, com a mistura contendo teores médios de 22% de proteína bruta (PB) e 73% de nitrogênio digestíveis totais (NDT).

As vacas são ordenhadas três vezes ao dia, sendo às 4, 12 e 20 horas, por um sistema de ordenha tipo Tandem - duplo 4 em fosso, linha baixa e canalizada. Os animais possuem identificação individual eletrônica por transponder, com pesagem eletrônica do leite, pelo sistema Alpro-DeLaval.

A partir dessas informações e da disponibilidade dos pesquisadores da referida instituição decidiu-se selecionar amostras de tamanho  $N = 6$  ( $n = 3$  e  $m = 2$ ), uma para cada um dos 29 dias possíveis para a coleta de dados.

Para fazer a ordenação, os animais foram numerados de 1 a 22 e alojados em sistema de *free stall*. Em seguida, para obter uma amostra por conjuntos de

tamanho  $n = 3$  com  $m = 2$  réplicas, foram selecionados  $3 \times 2 = 6$  conjuntos cada um de tamanho 3, totalizando 18 animais. Em cada um dos 6 conjuntos, os 3 animais foram ordenados a partir do tamanho do úbere (inspeção visual) por um especialista. Para compor a amostra por conjuntos ordenados, escolheu-se: o animal que ficou na posição 1 no conjunto 1, o animal que ficou na posição 2 no conjunto 2, o animal que ficou na posição 3 no conjunto 3, o animal que ficou na posição 1 no conjunto 4, o animal que ficou na posição 2 no conjunto 5 e o animal que ficou na posição 3 no conjunto 6.

Esse procedimento foi repetido mais 28 vezes, um para cada dia de coleta de dados.

Uma consideração importante se refere ao fato de que as informações de 29 dias distintos de 22 animais estão sendo considerados como uma população normal.

Para este estudo estão sendo desprezadas as correlações entre as produções do mesmo animal que eventualmente foi selecionado em dias distintos. Um outro aspecto desprezado se refere ao fato de que a produção de leite de um animal ao longo do tempo obedece uma curva de lactação. Assim, para os 29 dias estão sendo consideradas produções homogêneas.

Observa-se na Tabela 2 as 29 amostras por conjuntos ordenados da produção diária de leite, ou seja, a soma da produção de leite nas três ordenhas diárias, considerando o tamanho amostral  $N = 6$  ( $n = 3$  e  $m = 2$ ).

Como toda a produção de leite é registrada, foi possível verificar se a ordenação dos animais foi correta ou não. Os valores da Tabela 2 assinalados com asterisco indicam as observações em que a ordenação foi imperfeita. Uma explicação para esse fato, é que a ordenação feita visualmente foi realizada alguns dias antes da efetiva mensuração, pois o pesquisador não tinha disponibilidade para fazer a ordenação todos os dias. Portanto, como a ordenação não foi realizada no dia da coleta da amostra, o especialista não pôde levar em consideração a variabilidade diária do tamanho do úbere em cada animal e dentro de cada dia, a variação em cada ordenha.

Com base nesse conjunto de dados, estudou-se a precisão relativa para estimadores da variância em uma população normal, considerando amostras por conjuntos ordenados e ordenação perfeita. Apesar desta última suposição não ter sido verificada, mesmo assim realizou-se o estudo numérico que será descrito a seguir.

Com relação a suposição de distribuição normal, foi realizada uma análise descritiva dos dados.

O conjunto de dados analisado corresponde às 638 produções diárias de leite pesadas em 29 dias para as 22 vacas. Na Figura 4 (ver Apêndice) são observados o histograma e o gráfico quantil-quantil normal das 638 observações.

De acordo com o histograma tem-se que os dados parecem ter distribuição normal, entretanto o gráfico quantil-quantil mostra alguns pontos fora da reta, o que implica numa possível violação na suposição de normalidade dos dados.

Levantada a suspeita de não-normalidade, foi aplicado aos dados dois testes: Kolmogorov-Smirnov e Shapiro-Wilk. Os testes resultaram, respectivamente, em  $p$ -valores de 0,0015 e 0,0002. Fixando um nível de significância de 0,05, rejeita-se

a hipótese de que os dados têm distribuição normal com média 29,2198 e variância 17,0547.

Tabela 2 - Amostras por conjuntos ordenados (ACO) da produção diária de leite, considerando o tamanho amostral  $N = 6$

dias	ACO					
1	28,4	30,2	24,9*	28,1	29,3*	30,9
2	22,4	28,9*	27,5*	29,7*	28,9	26,7*
<b>3</b>	31,2	31,4*	28,0*	26,8	28,9	26,6*
4	26,0	29,9*	26,8*	28,8*	28,0	33,2
5	25,6	31,0	34,9*	25,9	29,4*	34,2
6	28,2	32,0*	35,5	30,5*	27,3	25,3*
<b>7</b>	26,8*	30,4	28,6*	29,8*	32,8*	31,2
<b>8</b>	28,0*	35,3*	27,7*	32,0*	33,1*	30,1
<b>9</b>	27,1*	31,3	35,9	36,0*	33,0*	29,9*
10	25,1*	25,7*	27,7	27,9	28,6*	33,8
11	25,7	28,2*	30,4*	26,4	32,3	24,5*
<b>12</b>	29,9*	32,6*	31,8*	31,3	27,6	32,7
<b>13</b>	27,1	29,6	34,0	27,6	27,0*	32,0
<b>14</b>	26,9	32,6*	27,8	30,3	30,1	28,6*
15	29,3*	30,7*	25,0	24,7	29,1*	29,8*
16	33,4	29,6*	29,8	27,1	29,6	24,9*
<b>17</b>	30,9*	29,9*	31,6	31,8	30,7*	31,9
18	26,3*	26,5*	34,4	24,3	34,9*	41,4
19	24,6	33,0*	42,8	29,2*	26,2*	35,0
20	30,8*	29,6*	31,2	27,4	20,0*	28,0*
21	34,2*	23,9*	29,2*	21,3	24,5*	32,6
<b>22</b>	29,9*	28,7	29,1*	31,2*	20,4*	32,6
<b>23</b>	26,6*	31,1	26,5*	25,9*	29,1	35,4
24	25,9	29,4*	20,8*	27,4	30,0	35,5
<b>25</b>	25,1	33,7*	29,1	31,2*	30,4*	27,6*
26	18,4	25,6	34,6*	29,5*	26,9	31,3
<b>27</b>	25,0*	30,9	25,9*	30,3*	27,8	29,2*
<b>28</b>	25,6	31,2*	32,2	27,5*	28,4*	36,1
<b>29</b>	28,4	31,2	32,4	30,0*	28,0	32,6

\* valores em que a ordenação foi imperfeita.

Este resultado fez que se levantassem suspeitas de que a população não é homogênea, mesmo as vacas sendo da mesma raça, mesma parição e mesma nutrição. Além disso, em geral, a produção de leite tem distribuição normal. Sendo assim, foi realizada uma análise descritiva para os 22 animais separadamente,



calculando sua média e variância. A partir dessa análise, foi verificado que três vacas possuíam desvio padrão maior que as demais e valores médios distantes da média geral. Com isso, chegou-se a conclusão de que essas vacas faziam que o conjunto de 638 observações não fosse homogêneo. Portanto, optou-se em retirar esses três animais, produzindo um conjunto de dados de 551 produções diárias de leite pesadas em 29 dias para as 19 vacas.

Observa-se na Figura 5 o histograma e o gráfico quantil-quantil da produção diária de leite de 551 observações.

Os testes de Kolmogorov e Shapiro foram novamente aplicados às 551 observações e seus  $p$ -valores são 0,2340 e 0,2597, respectivamente. Fixando um nível de significância de 0,05, tem-se que os dados tem distribuição normal com média 29,4933 e variância 9,8069. Note que, tirando os três animais, a média praticamente se manteve, mas a variabilidade foi reduzida.

Tendo retirado três animais, conseqüentemente a amostra por conjuntos ordenados é reduzida e a Tabela 2 fica com apenas 14 amostras por conjuntos ordenados (dias em negrito).

## 5 Precisão relativa utilizando os dados reais

Considerando que os dados da produção diária de leite das 551 observações tem distribuição normal (com média 29,4933 e variância 9,8069), o estudo da precisão relativa consistiu de 14 amostras por conjuntos ordenados, cada uma de tamanho  $N = 6$  ( $n = 3$  e  $m = 2$ ). A partir dessas 14 ACO foram calculados o valor médio, o vício, a variância e o erro médio quadrático de cada um dos estimadores,  $\hat{\sigma}^2$ ,  $\sigma_{rep}^{*2}$  e  $\hat{\sigma}_{MS}^2$ . Em seguida, foram obtidas as precisões relativas:  $PR(\hat{\sigma}^2, \sigma_{rep}^{*2})$ ,  $PR(\hat{\sigma}^2, \hat{\sigma}_{MS}^2)$ ,  $PR(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2)$ ,  $PREMQ(\hat{\sigma}^2, \sigma_{rep}^{*2})$ ,  $PREMQ(\hat{\sigma}^2, \hat{\sigma}_{MS}^2)$  e  $PREMQ(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2)$ .

Nas Tabelas 3 e 4 são observados, respectivamente, os resultados para o valor médio, vício, vício relativo, variância e erro médio quadrático e as precisões relativas.

Tabela 3 - Valor médio (VM), vício (V), vício relativo (VR), variância (Va) e erro médio quadrático (EMQ) para os estimadores (E)  $\hat{\sigma}^2$ ,  $\sigma_{rep}^{*2}$  e  $\hat{\sigma}_{MS}^2$

E	VM	V	VR	Va	EMQ
$\hat{\sigma}^2$	7,9545	-1,8527	-0,1889	25,5016	28,9340
$\sigma_{rep}^{*2}$	4,5399	-5,2673	-0,5371	44,5100	72,2550
$\hat{\sigma}_{MS}^2$	7,2611	-2,5460	-0,2596	21,2495	27,7319

Tabela 4 - Precisões relativas:  $PR(\hat{\sigma}^2, \sigma_{rep}^{*2})$ ,  $PR(\hat{\sigma}^2, \hat{\sigma}_{MS}^2)$ ,  $PR(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2)$ ,  $PREMQ(\hat{\sigma}^2, \sigma_{rep}^{*2})$ ,  $PREMQ(\hat{\sigma}^2, \hat{\sigma}_{MS}^2)$  e  $PREMQ(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2)$

PR	$(\hat{\sigma}^2, \sigma_{rep}^{*2})$	$(\hat{\sigma}^2, \hat{\sigma}_{MS}^2)$	$(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2)$
	1,7454	0,8333	0,4774
PREMQ	$(\hat{\sigma}^2, \sigma_{rep}^{*2})$	$(\hat{\sigma}^2, \hat{\sigma}_{MS}^2)$	$(\sigma_{rep}^{*2}, \hat{\sigma}_{MS}^2)$
	2,4972	0,9584	0,3838

De acordo com a Tabela 3 tem-se que o valor médio do estimador  $\hat{\sigma}^2$  está mais próximo do verdadeiro (9,8069) além de apresentar o menor vício absoluto, seguido do valor médio do estimador  $\hat{\sigma}_{MS}^2$ , com segundo menor vício absoluto. Entretanto, a variância do estimador  $\hat{\sigma}_{MS}^2$  é menor e mesmo levando em conta o seu vício, seu erro médio quadrático é inferior aos demais. O estimador  $\sigma_{rep}^{*2}$  apresenta maior vício absoluto, maior variabilidade e o pior desempenho comparado a  $\hat{\sigma}^2$  e  $\hat{\sigma}_{MS}^2$ .

Observa-se na Tabela 4 que para os dados da produção diária de leite, considerando tamanho amostral  $N = 6$ , o estimador  $\hat{\sigma}_{MS}^2$  é melhor que  $\hat{\sigma}^2$  e  $\sigma_{rep}^{*2}$ , sendo  $\sigma_{rep}^{*2}$  o pior estimador. Esses resultados são compatíveis aos apresentados na Seção 3.

## Conclusões

Na estimação da variância de uma população normal sob o delineamento de amostras por conjuntos ordenados, o estimador  $\hat{\sigma}_{MS}^2$  apresentou o melhor desempenho em termos da precisão relativa, quando comparado com os estimadores  $\hat{\sigma}^2$  e  $\sigma_{rep}^{*2}$ .

Utilizando os dados da produção diária de leite, o estudo da precisão relativa com 14 amostras por conjuntos ordenados mesmo com ordenação imperfeita foram similares ao estudo por simulação.

## Agradecimentos

Os autores agradecem o pesquisador Prof. Dr. Cledson Augusto Garcia, do Departamento de Zootecnia, Faculdade de Ciências Agrárias (FCA)-Unimar e os funcionários João Carlos e José Roberto, do Setor de Bovinocultura da Universidade de Marília-SP, por ter realizado o experimento usando amostragem por conjuntos ordenados, no período de julho a setembro de 2003.

GARCIA, C. A.; BARRETO, M. C. M. Relative precision for estimators of the variance of normal distribution under ranked set sampling *Rev. Mat. Estat.*, São Paulo, v.23, n.1, p.67-80, 2005.

■ **ABSTRACT:** Ranked set sampling (McIntyre, 1952) is indicated when the cost of measuring sample units is high, but it is possible to rank them before their effective measurement. We present in this paper a simulation study of the relative precision of estimators of variance normal population under the design of ranked set sampling. Three estimators are investigated: the nonparametric estimator developed by Stokes (1980), the natural estimator by Barnett & Moore (1997) and the Stokes' modified estimator (Yu et al., 1999). The simulation results have shown that the unbiased Stokes' modified estimator has the shortest variance and the best relative precision performance for small sample size. The natural estimator and the nonparametric estimator have small simulated bias and mean square error similar to those of the modified Stokes' estimators. Using a ranked set sampling data for average milk yield, the simulation results are illustrated.

■ **KEYWORDS:** Ranked set sampling; estimators of the variance; normal distribution; relative precision.

## Referências

AL-SALEH, M. F.; AL-SHRAFAT, K. Estimation of average milk yield using ranked set sampling. *Environmetrics*, Chichester, v.12, p.395-399, 2001.

BARNETT, V.; MOORE, K. Best linear unbiased estimates in ranked set sampling with particular reference to imperfect ordering. *J. Appl. Stat.*, Abington, v.24, n.6, p.697-710, 1997.

BARRETO, M. C. M.; BARNETT, V. Best linear unbiased estimators for the simple linear regression model using ranked set sampling. *Environ. Ecol. Stat.*, Norwell, v.6, p.119-133, 1999.

CESÁRIO, L. C. *Intervalos de confiança bootstrap em amostragem por conjuntos ordenados*. 2001. 103f. Dissertação (Mestrado em Estatística) - Centro de Ciências Exatas e Tecnologia, Universidade Federal de São Carlos, São Carlos, 2001.

CHEN, Z. On ranked set sample quantiles and their applications. *J. Stat. Plann. Infer.*, Amsterdam, v.83, p.125-135, 2000.

GARCIA, C. A. *Precisão relativa e intervalos de confiança bootstrap para a variância populacional de uma distribuição normal baseados na amostragem por conjuntos ordenados*. 2004. 127f. Dissertação (Mestrado em Estatística) - Centro de Ciências Exatas e Tecnologia, Universidade Federal de São Carlos, São Carlos, 2004.

DELL, T. R.; CLUTTER, J. L. Ranked set sampling theory with order statistics background. *Biometrics*, Washington, v.28, p.545-555, 1972.

MCINTYRE, G. A. A method for unbiased selective sampling using ranked sets. *Austr. J. Agric. Res.*, Collingwood, v.3, p.385-390, 1952.

RAMINELLI, J. A. *Procedimentos de estimação intervalar de um quantil populacional em amostragem por conjuntos ordenados*. 2003. 81f. Dissertação (Mestrado em Estatística) - Centro de Ciências Exatas e Tecnologia, Universidade Federal de São Carlos, São Carlos, 2003.

STOKES, S. L. Estimation of variance using judgement ordered ranked set samples. *Biometrics*, Washington, v.36, p.35-42, 1980.

STOKES, L. S.; SAGER, T. W. Characterization of a ranked set sample with application to estimating distributions functions. *J. Am. Stat. Assoc.*, New York, v.83, p.374-381, 1988.

YU, P. L. H.; LAM, K.; SINHA, B. K. Estimation of normal variance basead on balanced and unbalanced ranked set samples. *Environ. Ecol. Stat.*, Norwell, v.6, p.23-46, 1999.

## Apêndice

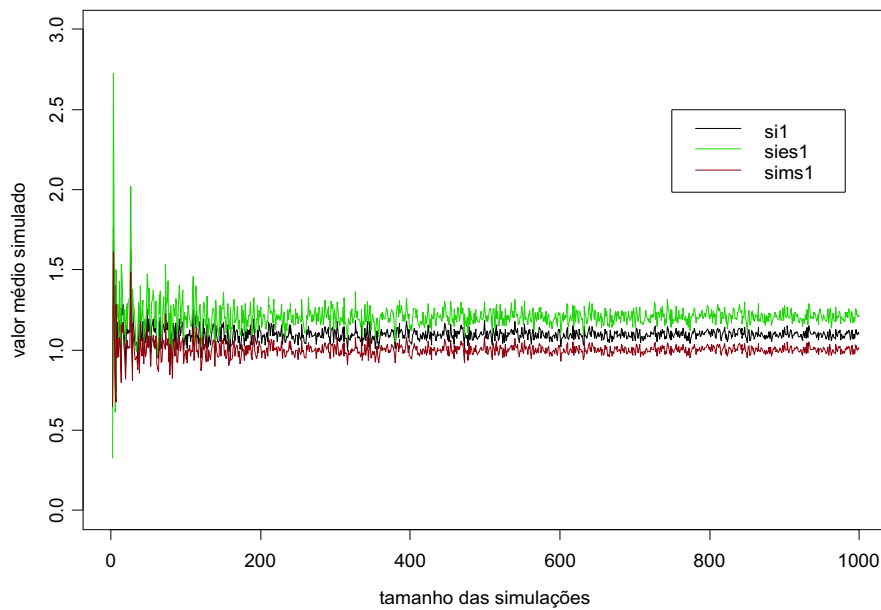


FIGURA 1 - Valor médio simulado para os estimadores  $\hat{\sigma}^2$  (si1),  $\sigma_{rep}^{*2}$  (sies1) e  $\hat{\sigma}_{MS}^2$  (sims1) versus tamanho das simulações.

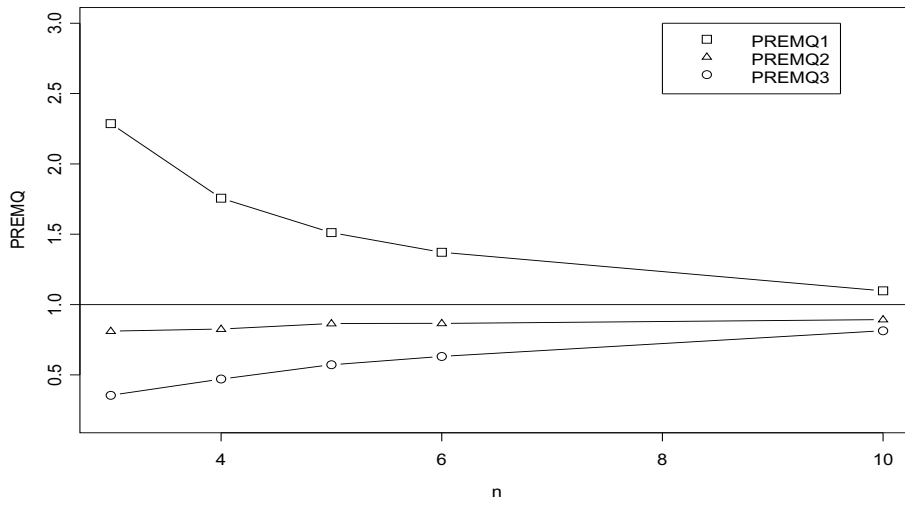


FIGURA 2 - Precisões relativas de  $\hat{\sigma}^2$  em relação a  $\sigma_{rep}^{*2}$  (PREMQ1),  $\hat{\sigma}^2$  em relação a  $\hat{\sigma}_{MS}^2$  (PREMQ2) e  $\sigma_{rep}^{*2}$  em relação a  $\hat{\sigma}_{MS}^2$  (PREMQ3) versus tamanho amostral  $n$  considerando  $m = 2$ .

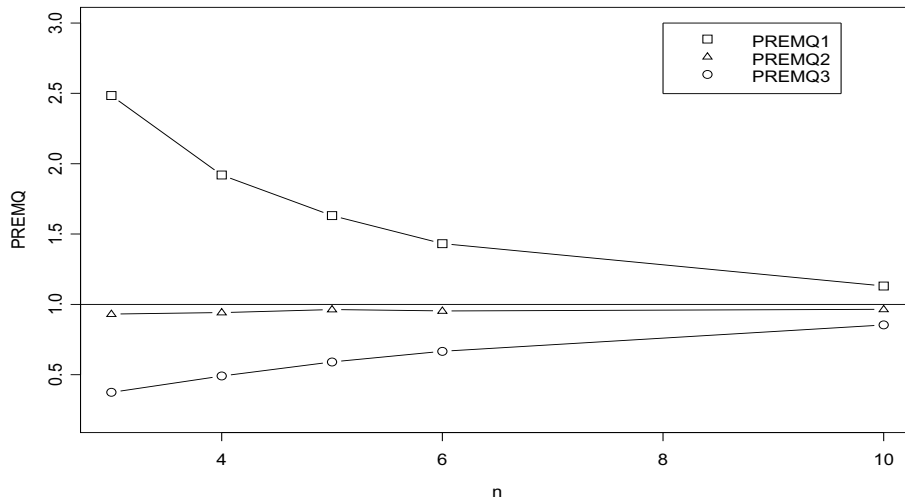


FIGURA 3 - Precisões relativas de  $\hat{\sigma}^2$  em relação a  $\sigma_{rep}^{*2}$  (PREMQ1),  $\hat{\sigma}^2$  em relação a  $\hat{\sigma}_{MS}^2$  (PREMQ2) e  $\sigma_{rep}^{*2}$  em relação a  $\hat{\sigma}_{MS}^2$  (PREMQ3) versus tamanho amostral  $n$  considerando  $m = 6$ .

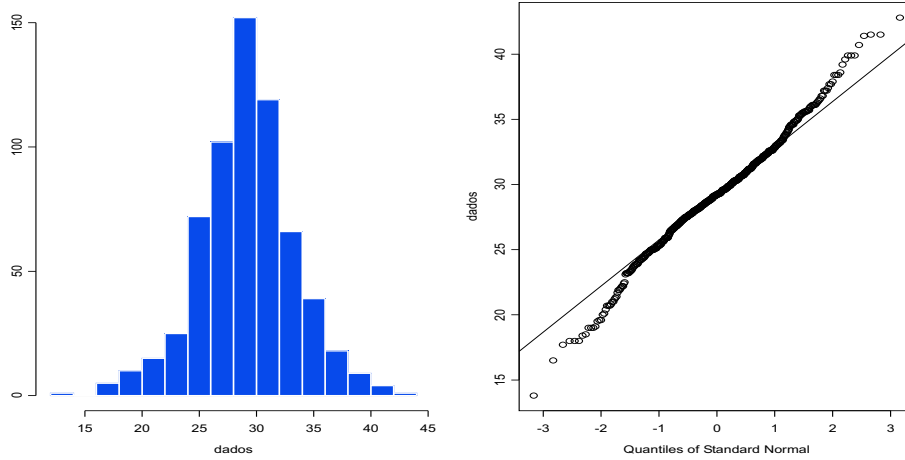


FIGURA 4 - Histograma e gráfico quantil-quantil da produção diária de leite de 638 observações.

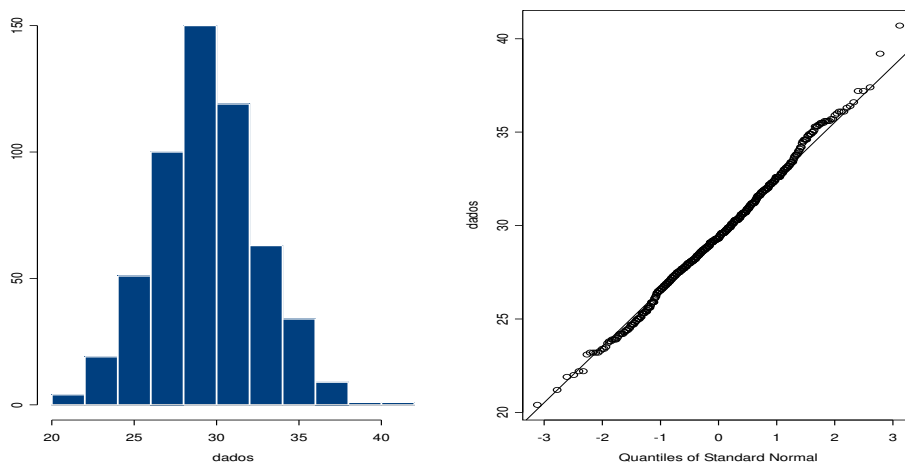


FIGURA 5 - Histograma e o gráfico quantil-quantil da produção diária de leite de 551 observações.

Recebido em 04.08.2004.

Aprovado após revisão em 01.04.2005.