Computação das Medidas de Tendência Central e Dispersão Intervalares em Java

Laís M. Neves, Eliaquim L. S. Neto, Eduardo W. M. Almeida, Daniele P. Santos
Thiago A. S. Nascimento, João G. Q. L. Martins, Marcília A. Campos

Centro de Informática, CIn, UFPE,
50732-970, Recife, PE
(lmm3, elsn, ewma, dps, tasn, jgqlm, mac)@fêb.unesp.br

Aline B. Loreto

Depto de Matemática, UNISC,
96815-900, Santa Cruz do Sul, RS
abl@unisc.br

Resumo: *O presente trabalho tem como objetivo computar as Medidas de Tendência Central média, mediana e moda e as Medidas de Dispersão amplitude, variância, desvio padrão, coeficiente de variação, covariância e coeficiente de correlação a nível intervalar. Para isso utiliza-se a biblioteca de operações aritméticas intervalares (+, -, *, /) desenvolvida no ambiente de programação Java [2].*

Palavras-chave: indicadores estatísticos, matemática intervalar, Java, aritmética de exatidão máxima.

Introdução

Uma forma de se considerar não somente os erros numéricos, mas também os erros aleatórios inerentes aos valores $x_1, \dots, x_n$ e $y_1, \dots, y_n$ é considerar que, para $i = 1, \dots, n$ e $j = 1, \dots, n$, $x_i \in [\underline{x}_i, \bar{x}_i]$ e $y_j \in [\underline{y}_j, \bar{y}_j]$. Neste caso, consideram-se os dados de entrada (as amostras) $\{x_1, \dots, x_n\}$, $x_1 \in X_1, \dots, x_n \in X_n$, ou seja, domínios intervalares onde $X_1 = [\underline{x}_1, \bar{x}_1], \dots, X_n = [\underline{x}_n, \bar{x}_n]$.

Nestas situações, para diferentes valores possíveis $x_i \in X_i$, obtém-se diferentes valores da média me , moda mo , mediana md , amplitude am , variância va , desvio padrão dp , coeficiente de variação cv , covariância co e coeficiente de correlação cc .

Considerando que os dados de entrada são valores intervalares, deseja-se conhecer quais são os intervalos média intervalar **MEv**, mediana intervalar **MDv**, moda intervalar **MOv**, variância intervalar **VAv**, desvio padrão intervalar **DPv**, coeficiente de variação intervalar **CVv**, covariância intervalar **Cov**, coeficiente de correlação intervalar **CCv** e amplitude

intervalar **AMv** dos possíveis valores da média me , moda mo , mediana md , variância va , desvio padrão dp , coeficiente de variação cv , covariância co , coeficiente de correlação cc e amplitude am .

As definições dos indicadores média intervalar, moda intervalar, variância intervalar e desvio padrão intervalar foram retiradas de Campos e Faria [1], os indicadores covariância intervalar e coeficiente de correlação intervalar foram pesquisadas em [7] e [5]. A definição do indicador coeficiente de variação intervalar foi formulada por [4].

Exemplo 1. Sejam os valores reais $\{1; 2; 2; 4; 5\}$ e $\{2; 3; 1; 4; 5\}$ amostras aleatórias de uma população. Representando estes valores de forma intervalar, com precisão $\epsilon = 10^{-2}$ temos como dados de entrada os valores intervalares $\{[0;99; 1;01]; [1;99; 2;01]; [1;99; 2;01]; [3;99; 4;01], [4;99; 5;01]\}$ e $\{[1;99; 2;01]; [2;99; 3;01]; [0;99; 1;01]; [3;99; 4;01]; [4;99; 5;01]\}$. Vamos considerar estes dados intervalares para exemplificar os cálculos de cada indicador intervalar definido a seguir.

As operações intervalares envolvidas nos conceitos seguintes podem ser encontradas em Moore [5].

Os cálculos dos exemplos, de cada indicador intervalar, foram realizados utilizando sistema de ponto flutuante F(10, 3, -10,10) com arredondamento direcionado [3].

1. Média aritmética Intervalar

Considere

$$\sum_{i=1}^n X_i = \sum_{i=1}^n [\underline{x}_i, \bar{x}_i] = [\sum_{i=1}^n \underline{x}_i, \sum_{i=1}^n \bar{x}_i]$$

Logo

$$MEv = \frac{1}{n-1} \sum_{i=1}^n X_i = \frac{1}{n} [\sum_{i=1}^n \underline{x}_i, \sum_{i=1}^n \bar{x}_i]$$

Esta última expressão pode ser resolvida de duas formas distintas:

- i) usando a multiplicação de um real por um intervalo, como $1/n > 0$ não haverá alteração nos limites do intervalo resultante e
- ii) definindo o intervalo degenerado $[1/n; 1/n]$ e usando a multiplicação de intervalos. Em ambos os casos tem-se a

Definição 1.1. MÉDIA ARITMÉTICA INTERVALAR

$$ME_v = [me(\underline{x}), me(\overline{x})] = [\underline{me}, \overline{me}].$$

Se for necessário um único valor para a média, calcular o ponto médio do intervalo, usando

$$m(x) = (\underline{x} + \overline{x})/2$$

Exemplo 2. Considerando os valores intervalares $X_1 = [0,99; 1,01]$, $X_2 = [1,99; 2,01]$, $X_3 = [1,99; 2,01]$, $X_4 = [3,99; 4,01]$ e $X_5 = [4,99; 5,01]$, obtemos $ME = [2,78; 2,82]$.

2. Mediana Intervalar

Primeiramente deve-se ordenar os dados intervalares de acordo com os seguintes casos:

- **TODOS OS INTERVALOS DISJUNTOS:** considera-se a ordem intervalar de Kulisch [3]:

$$X_1 \leq X_2 \leq \dots \leq X_n, \text{ onde}$$

$$X_1 \leq X_2 \Leftrightarrow \underline{x}_1 \leq \underline{x}_2 \wedge \overline{x}_1 \leq \overline{x}_2$$

- **TODOS OS INTERVALOS ENCAIXADOS:** neste caso, para a ordenação dos elementos, considera-se a ordem de inclusão [3]:

$$X_{(1)} \subseteq X_{(2)} \subseteq \dots \subseteq X_{(n)}, \text{ onde}$$

$$X_{(1)} \subseteq X_{(2)} \Leftrightarrow \underline{x}_1 \geq \underline{x}_2 \wedge \overline{x}_1 \geq \overline{x}_2$$

A mediana intervalar MD corresponderá ao intervalo mediano. A posição deste intervalo é calculada como em valores reais.

Definição 1.2. MEDIANA INTERVALAR

$$ME_v = [me(\underline{x}), me(\overline{x})] = [\underline{me}, \overline{me}]$$

3. Moda Intervalar

Define-se o intervalo modal por

Definição 1.3. MODA INTERVALAR

Se existe um valor modal para os dados reais, então

$$MO_v = [mo_{0 \leq i \leq n}\{\underline{x}_i\}, mo_{0 \leq i \leq n}\{\overline{x}_i\}] = [\underline{mo}, \overline{mo}]$$

Exemplo 3. Com os valores $X_1 = [0,99; 1,01]$, $X_2 = [1,99; 2,01]$, $X_3 = [1,99; 2,01]$, $X_4 = [3,99; 4,01]$ e $X_5 = [4,99; 5,01]$, obtemos $MO = [1,99; 2,01]$.

Exemplo 4. Com os valores intervalares $X_1 = [1:99; 2:01]$, $X_2 = [2:99; 3:01]$, $X_3 = [0:99; 1:01]$, $X_4 = [3:99; 4:01]$ e $X_5 = [4:99; 5:01]$, não obtemos valor modal intervalar MO pois não existe, para estes dados intervalares, intervalos repetidos.

4. Variância Intervalar

Temos que

$$x_i - ME = [\underline{x}_i, \overline{x}_i] - [\underline{me}, \overline{me}] = [\underline{x}_i - \overline{me}, \overline{x}_i - \underline{me}]$$

$$(x_i - ME)^2 = \begin{cases} [(\underline{x}_i - \overline{me})^2, (\overline{x}_i - \underline{me})^2]; \\ \text{se } \underline{x}_i - \overline{me} \geq 0 \\ [(\overline{x}_i - \underline{me})^2, (\underline{x}_i - \overline{me})^2]; \\ \text{se } \underline{x}_i - \overline{me} < 0 \\ [0, \max((\underline{x}_i - \overline{me})^2, (\overline{x}_i - \underline{me})^2)]; \\ \text{se } \underline{x}_i - \overline{me} < 0 \leq \overline{x}_i - \underline{me} \end{cases}$$

Então definimos a variância intervalar como

Definição 1.4. VARIÂNCIA INTERVALAR

$$VA_v = [\underline{va}, \overline{va}] = \frac{1}{n-1} \sum_{i=1}^n (x_i - ME)^2$$

Exemplo 5. Considerando os valores intervalares do Exemplo 1 $X_1 = [0,99; 1,01]$, $X_2 = [1,99; 2,01]$, $X_3 = [1,99; 2,01]$, $X_4 = [3,99; 4,01]$ e $X_5 = [4,99; 5,01]$, obtemos $ME = [2,78; 2,82]$ e $VA = [2,08; 2,24]$.

5. Desvio Padrão Intervalar

Como o desvio padrão é a raiz quadrada positiva da variância, a obtenção da raiz quadrada de cada limite do intervalo VA, fornecerá o intervalo desvio padrão DP:

Definição 1.5. DESVIO PADRÃO INTERVALAR

$$DP_v = \sqrt{VA} = +\sqrt{[\underline{va}, \overline{va}]} = [+ \sqrt{\underline{va}}, + \sqrt{\overline{va}}]$$

Exemplo 6. Considerando o valor da variância intervalar, calculado no exemplo 4, $VA = [2,08; 2,24]$ obtemos $DP = [1,44; 1,50]$.

6. Coeficiente de Variação Intervalar

Definição 1.6. COEFICIENTE DE VARIAÇÃO INTERVALAR

Se $0 \notin [\underline{me}, \overline{me}]$, então

$$CV_v = [\underline{cv}, \overline{cv}] = \frac{DP_v}{ME_v} = \frac{[dp, \overline{dp}]}{[\underline{me}, \overline{me}]} = [dp, \overline{dp}] * \frac{1}{[\underline{me}, \overline{me}]}$$

Exemplo 7. Considerando os mesmos valores intervalares do exemplo da variância intervalar, Exemplo 4, e

o intervalo do desvio padrão intervalar $DP = [1,44; 1,50]$, calculado no Exemplo 5, obtemos $ME = [2,78; 2,82]$ e $CV = [0,51; 0,54]$.

7. Covariância Intervalar

Temos que:

$$x_i - ME_x = [\underline{x}_i, \bar{x}_i] - [\underline{me}_x, \bar{me}_x] = [\underline{x}_i - \bar{me}_x, \bar{x}_i - \underline{me}_x]$$

e

$$y_i - ME_y = [\underline{y}_i, \bar{y}_i] - [\underline{me}_y, \bar{me}_y] = [\underline{y}_i - \bar{me}_y, \bar{y}_i - \underline{me}_y]$$

onde ME_x é a média intervalar dos valores de X e

ME_y a média intervalar dos valores de Y. Então, definimos a covariância intervalar como:

Definição 1.7. COVARIÂNCIA INTERVALAR

$$CO_v = [\underline{co}, \bar{co}] = \frac{1}{n-1} \sum_{i=1}^n (x_i - ME_x)(y_i - ME_y)$$

Exemplo 8. Neste caso precisamos de dois conjuntos de dados intervalares. Consideramos os mesmos valores intervalares do Exemplo 4 da variância intervalar $X_1 = [0,99; 1,01]$, $X_2 = [1,99; 2,01]$, $X_3 = [1,99; 2,01]$, $X_4 = [3,99; 4,01]$ e $X_5 = [4,99; 5,01]$, 1.4.4, e os seguintes dados intervalares $Y_1 = [1,99; 2,01]$, $Y_2 = [2,99; 3,01]$, $Y_3 = [0,99; 1,01]$, $Y_4 = [3,99; 4,01]$ e $Y_5 = [4,99; 5,01]$. Obtemos $ME_x = [2,78; 2,82]$, $ME_y = [2,98; 3,00]$ e $CO = [1,74; 1,87]$.

8. Coeficiente de Correlação Intervalar

O coeficiente de correlação intervalar é definido da seguinte maneira:

Definição 1.8. COEFICIENTE DE CORRELAÇÃO INTERVALAR

Se $0 \notin DP_x DP_y$ então

$$CC_v = [\underline{cc}, \bar{cc}] = \frac{CO_v}{DP_x DP_y} = [\underline{co}, \bar{co}] * \left(\frac{1}{[\underline{dp}_x, \bar{dp}_x][\underline{dp}_y, \bar{dp}_y]} \right)$$

(onde CO é covariância intervalar, DP_x o desvio padrão intervalar dos valores de X e DP_y o desvio padrão intervalar dos valores de Y).

Exemplo 9. Considere os mesmos conjuntos de valores intervalares do Exemplo 7 da covariância intervalar. Obtemos os valores intervalares de $CO = [1,74; 1,87]$, $DP_x DP_y = [2,02; 2,14]$ e $CC = [0,81; 0,93]$.

9. Amplitude

Com os dados intervalares ordenados segundo a ordem para intervalos disjuntos " $\leq$ " ou intervalos encaixados " $\subseteq$ " (Item 2.), considera-se :

- Mínimo Intervalar:

$$MI = [\underline{mi}_{1 \leq i \leq n}\{\underline{x}_i\}, \underline{mi}_{1 \leq i \leq n}\{\bar{x}_i\}] = [\underline{mi}, \bar{mi}]$$

- Máximo Intervalar

$$MO = [mo_{1 \leq i \leq n}\{\underline{x}_i\}, mo_{1 \leq i \leq n}\{\bar{x}_i\}] = [mo, \bar{mo}]$$

Define-se a amplitude total intervalar como sendo:

Definição 3.4. AMPLITUDE TOTAL INTERVALAR

$$AT_v = [\underline{at}, \bar{at}] = \begin{cases} [\underline{ma} - \underline{mi}, \bar{ma} - \bar{mi}], & \text{se } \underline{ma} \geq \bar{mi} \\ [0, \bar{ma} - \bar{mi}], & \text{se } \underline{ma} < \bar{mi} \end{cases}$$

10. Medidas de Erro

Para verificação de resultados e certificação da não ocorrência de intervalos superestimados (intervalos com amplitude grande), consideram-se as seguintes medidas de erros [7]:

$$i) \quad |x_k - m(X_k)| < \varepsilon, \quad m(X_k) = \left(\frac{x_k + \bar{x}_k}{2} \right), \text{ onde } m(X_k) = \bar{x}_k + \underline{x}_k / 2;$$

$$ii) \quad |x_k - m(X_k)| < \frac{w(X_k)}{2}, \text{ onde } w(X_k) = \bar{x}_k - \underline{x}_k;$$

$$iii) \quad \left(\frac{x_k - m(x_k)}{x_k} \right) \leq \frac{w(X_k)}{2 \min |X_k|}, \text{ se}$$

$0 \notin X_k$, e realiza-se uma comparação com os valores reais dos indicadores estatísticos descritos acima.

Estudo de Caso

Como estudo de caso, vamos considerar o cálculo do IMC (Índice de Massa Corporal) intervalar dos alunos do ensino fundamental de 5a. a 8a. série da escola Nossa Senhora Imaculada Conceição, localizada na cidade de Cachoeira do Sul, RS.

O índice de Massa Corporal (IMC) é uma fórmula que indica se uma pessoa está acima do peso, se está obeso ou com peso ideal considerado saudável. A fórmula para calcular o Índice de Massa Corporal é: $IMC = peso/(altura)^2$.

Tabela 1.1: Idade, IMC Meninas e IMC Meninos.

Idade	IMC Meninas			IMC Meninos		
	Saudável	Sobrepeso	Obeso	Saudável	Sobrepeso	Obeso
9	< 19.1	19.1-22.7	≥ 22.8	< 19.1	19.1-22.8	≥ 22.8
10	< 19.9	19.9-24.0	≥ 24.1	< 19.8	19.8-23.9	≥ 24.0
11	< 20.7	20.7-25.3	≥ 25.4	< 20.6	20.6-25.0	≥ 25.1
12	< 21.7	21.7-26.6	≥ 26.7	< 21.2	21.2-25.9	≥ 26.0
13	< 22.6	22.6-27.7	≥ 27.8	< 21.9	21.9-26.7	≥ 26.8
14	< 23.3	23.3-28.5	≥ 28.6	< 22.6	22.6-27.5	≥ 27.6
15	< 23.9	23.9-29.0	≥ 29.1	< 23.3	23.3-28.2	≥ 28.3
16	< 24.4	24.4-29.3	≥ 29.4	< 23.9	23.9-28.8	≥ 28.9

Considera-se o IMC da turma de 21 alunos da 5a. série da escola Imaculada Conceição.

IMC das meninas: {16.8786, 14.2687, 20.7834, 18.2961, 19.8179, 27.0258, 22.6562, 22.0711, 26.9127, 21.0517, 24.6094, 25.1095, 16.0697, 21.8146}.

Representação intervalar: {[16.874, 16.884], [14.264, 14.274], [20.778, 20.788], [18.291, 18.301], [19.813, 19.823], [27.021, 27.031], [22.651, 22.661], [22.066, 22.076], [26.908, 26.918], [21.047, 21.057], [24.604, 24.614], [25.104, 25.114], [16.065, 16.075], [21.810, 21.820]}.

Para calcular os indicadores estatísticos covariância e coeficiente de correlação precisa-se de dois conjuntos de dados. Dessa forma, considera-se o IMC dos meninos da 5a. série da mesma escola (e alguns valores do IMC das meninas para completar o total de valores necessários): {16.8786, 15.1506, 20.7834, 19.0342, 19.8179, 17.3469, 22.6562, 15.7342, 26.9127, 20.8889, 21.0517, 25.1095, 16.0697, 21.8146}.

Representa-se estes valores em intervalos com a mesma precisão $\pm = 0:005$: {[16.874, 16.884], [15.146, 15.156], [20.778, 20.788], [19.029, 19.039], [19.813, 19.823], [17.342, 17.352], [22.651, 22.661], [15.729, 15.739], [26.908, 26.918], [20.884, 20.894], [21.047, 21.057], [25.104, 25.114], [16.065, 16.075], [21.810, 21.820]}.

Na Tabela 1.2 apresentam-se os resultados dos indicadores estatísticos, reais e intervalares, com IMC das meninas da 5a. série.

Na Tabela 1.3 apresenta-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 1.2: Indicadores Estatísticos, dados com IMC reais (meninas da 5a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninas da 5a. série)	Dados com IMC Intervalares
Média	22.874	[22.868, 22.880]
Mediana		
Moda	-	-
Amplitude total		
Variância	18.282	[18.206, 18.357]
Desvio Padrão	4.2756	[4.2668, 4.2845]
Coef. de Variação	0.18692	[0.18649, 0.18735]
Covariância	11.835	[11.762, 11.906]
Coef. de Correlação	0.71544	[0.70802, 0.72290]

Tabela 1.3: Indicadores Estatísticos e erro relativo (meninas 5a. série).

Indicadores Estatísticos	Erro Relativo $ \frac{x_k - m(x_k)}{x_k} \leq \frac{w(x_k)}{2min x_k }$
Média	0.0000033662 $\leq$ 0.00023545
Mediana	
Moda	-
Amplitude total	
Variância	0.000012618 $\leq$ 0.0041414
Desvio Padrão	0.000013243 $\leq$ 0.0020664
Coef. de Variação	0.000015568 $\leq$ 0.0023029
Covariância	0.000052657 $\leq$ 0.0060996
Coef. de Correlação	0.000028769 $\leq$ 0.010508

Tabela 1.5: Indicadores Estatísticos e erro relativo (meninos 5a. série).

Indicadores Estatísticos	Erro Relativo $ \frac{x_k - m(x_k)}{x_k} \leq \frac{w(x_k)}{2min x_k }$
Média	0.0000016884 $\leq$ 0.00029549
Mediana	
Moda	-
Amplitude total	
Variância	0.000012878 $\leq$ 0.0055226
Desvio Padrão	0.000013893 $\leq$ 0.0027537
Coef. de Variação	0.0017629 $\leq$ 0.0030508
Covariância	0.000027560 $\leq$ 0.0064737
Coef. de Correlação	0.000024846 $\leq$ 0.011152

IMC dos meninos: {15.1506, 14.9515, 19.0342, 15.3739, 17.3469, 15.7342, 20.8889}. Representação intervalar: {[15.146, 15.156], [14.946, 14.956], [19.029, 19.039], [15.369, 15.379], [17.342, 17.352], [15.729, 15.739], [20.884, 20.894]}.

Para os indicadores estatísticos covariância e coeficiente de correlação considera-se o IMC das meninas da 5a. série da mesma escola: {16.8786, 4.2687, 20.7834, 18.2961, 19.8179, 27.0258, 22.6562}. Representa-se estes valores em intervalos com a mesma precisão $\pm = 0:005$: {[16.874, 16.884], [14.264, 14.274], [20.778, 20.788], [18.291, 18.301], [19.813, 19.823], [27.021, 27.031], [22.651, 22.661]}.

Na Tabela 1.4 apresentam-se os resultados dos indicadores estatísticos, reais e intervalares, com IMC dos meninos da 5a. série. Na Tabela 1.5 apresentam-se os erros obtidos ao calcular os indicadores estatísticos com valores intervalares.

Tabela 1.4: Indicadores Estatísticos, dados com IMC reais (meninos da 5a. série) e dados com IMC intervalares.

Indicadores Estatísticos	Dados com IMC Reais (meninos da 5a. série)	Dados com IMC Intervalares
Média	19.747	[19.740, 19.753]
Mediana		
Moda	-	-
Amplitude total		
Variância	14.485	[14.405, 14.565]
Desvio Padrão	3.8059	[3.7955, 3.8164]
Coef. de Variação	0.1924	[0.19215, 0.19333]
Covariância	14.82073	[14.725, 14.916]
Coef. de Correlação	0.71092	[0.70309, 0.71878]

Implementação

As operações intervalares aqui apresentadas, foram implementadas em linguagem de programação Java. Para realizar tais operações foi desenvolvida uma classe onde seus métodos recebem como entrada um conjunto de intervalos e retornam como resultado um intervalo que representa a resposta da respectiva operação realizada.

É importante ressaltar algumas peculiaridades da implementação, tais como o cálculo da Covariância e do Coeficiente de Correlação Intervalar, onde a entrada de dados para essas operações consiste em dois conjuntos de intervalo. No cálculo da Mediana e da Amplitude Intervalar, devemos assumir que os intervalos pertencentes ao conjunto de entrada enquadram-se em um dos casos de ordenação vistos no Item 2. No cálculo da Moda Intervalar, quando a mesma não existir, será retornada uma saída nula.

Agradecimentos

Às professoras Marcília A. Campos e Aline B. Loreto por todo o apoio dado ao nosso trabalho. Às nossas famílias.

Referências

- [1] CAMPOS, M. A., FARIA, R. A. Definição de alguns indicadores estatísticos usando intervalos. In XI Congresso Nacional de Matemática Aplicada e Computacional, Ouro Preto, MG, 1988.
- [2] FERREIRA, R. V., FERNANDES, B. T., CAMPOS, M. A., Evaluating the floating-point in Java Virtual Machine, In Proceedings ENNEMAC 2004.
- [3] KULISCH, U. W. , MIRANKER, W. L. , Computer Arithmetic in Theory and Practice. Academic Press, New York, New York, 1981.
- [4] LORETO, A. B., CAMPOS, M. A., CLAUDIO, D. M., TOSCANI, L. V. Analisando a Complexidade Computacional de Problemas de Medidas de Tendência Central e Dispersão, CNMAC 2005.
- [5] MOORE, R. E. , Interval Analysis. Prentice-Hall, Englewood Cliffs, New Jersey, 1966.

[6] MOORE, R. E. Methods and Applications of Interval Analysis. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1979.

[7] RATSCHKE, H., ROKNE, J., New Computer Methods for Global Optimization, Collis Horwood Limited, Great Britain, 1988.