OBSCOOP
Observatório do Cooperativismo

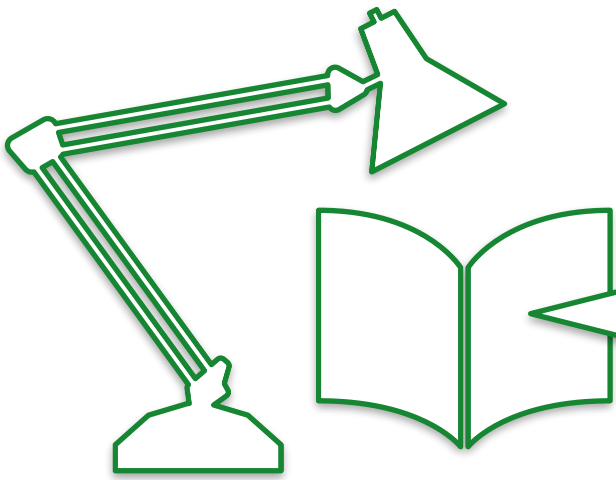
Aula 03 –(Aula Síncrona)

(30/08)

Davi R. de Moura Costa

USP
fea-RP

Tópicos



<u>Medidas-Resumo</u>

- Medidas de posição
- Medidas de dispersão

Aula 30 e 31/08

- Quantis
- Box Plots
- Transformações

Aula 13 e 14/09

Referências

Leitura obrigatória

- Bussab & Morettin - Cap. 3

Estatística descritiva

Medidas de Posição

Medidas de Dispersão

Quantis e Box Plots

Transformações

Aula 30 e 31/08

Aula 13 e 14/09

Média
Moda
Mediana

Variância e Desvios

Outliers, efeitos e Tratamentos

Medidas de Posição

Foi verificado no RH da Empresa o número de filhos de cada funcionário casado . Os dados são os que seguem:

Variável Quantitativa - Discreta

Nº de filhos z_i	Freqüência n_i	Porcentagem $100\,f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Medidas de Posição

Foi verificado no RH da Empresa o número de filhos de cada funcionário casado . Os dados são os que seguem:

Nº de filhos z_i	Freqüência n_i	Porcentagem $100\ f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Moda

$$(0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 3,\ 3,\ 3,\ 5)$$

Moda

$$Mo = 2$$

Medidas de Posição

Nº de filhos z_i	Freqüência n_i	Porcentagem $100\,f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Mediana

$$Md(X) = \begin{cases} X_{\frac{n+1}{2}}, & \text{se } n \text{ é impar;} \\ X_{\frac{(\frac{n}{2})+X(\frac{n}{2}+1)}{2}}, & \text{se } n \text{ é par.} \end{cases}$$

$$x_{(\frac{n}{2})} = \frac{20}{2} = x_{10} \Rightarrow x_{10} = 2$$

$$x_{(\frac{n}{2}+1)} = x_{(\frac{20}{2}+1)} = x_{11} \Rightarrow x_{11} = 2$$

$$\begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} & x_{17} & x_{18} & x_{19} & x_{20} \\ (0, & 0, & 0, & 0, & 1, & 1, & 1, & 1, & 1, & 2, & 2, & 2, & 2, & 2, & 2, & 2, & 3, & 3, & 3, & 5) \end{matrix}$$

Mediana

$$Md = \frac{2+2}{2} = 2$$

Medidas de Posição

Foi verificado no RH da Empresa o número de filhos de cada funcionário casado . Os dados são os que seguem:

Nº de filhos z_i	Freqüência n_i	Porcentagem $100\ f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Média

$$(0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 1,\ 1,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 3,\ 3,\ 3,\ 5)$$

Média simples

$$Me = \frac{\sum_{i=1}^{n} x_i}{n} \qquad Me = \frac{(0*4)+(1*5)+(2*7)+(3*3)+(5*1)}{20} = \frac{33}{20} = 1{,}65$$

Medidas de Posição

Média simples (só é válida para variáveis quantitativas)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad Me = \frac{\sum_{i=1}^{n} x_i}{n}$$

Se temos $n = n_1 + n_2 + \ldots + n_j \quad \Rightarrow \quad \bar{X} = \frac{(n_1 * x_1) + (n_2 * x_2) + \ldots + (n_k * x_k)}{n} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i$

Fazendo $f_i = \frac{n_1}{n}$ a frequência relativa da observação x_i teremos:

$$\bar{X} = \sum_{i=1}^{k} f_i x_i$$

$$Me = \frac{\sum_{i=1}^{n} x_i}{n} \qquad Me = \frac{(0 * 4) + (1 * 5) + (2 * 7) + (3 * 3) + (5 * 1)}{20} = \frac{33}{20} = 1{,}65$$

Medidas de Posição

Foi verificado no RH da Empresa o número de filhos de cada funcionário casado . Os dados são os que seguem:

Variável Quantitativa - Contínua

Classes de salários	Ponto médio s_i	Freqüência n_i	Porcentagem $100\ f_i$
4,00 ⊢ 8,00	6,00	10	27,78
8,00 ⊢ 12,00	10,00	12	33,33
12,00 ⊢ 16,00	14,00	8	22,22
16,00 ⊢ 20,00	18,00	5	13,89
20,00 ⊢ 24,00	22,00	1	2,78
Total	—	36	100,00

Medidas de Posição

Foi verificado no RH da Empresa o a faixa de salário de cada funcionário casado . Os dados são os que seguem:

Classes de salários	Ponto médio s_i	Freqüência n_i	Porcentagem $100\ f_i$
4,00 ⊢ 8,00	6,00	10	27,78
8,00 ⊢ 12,00	10,00	12	33,33
12,00 ⊢ 16,00	14,00	8	22,22
16,00 ⊢ 20,00	18,00	5	13,89
20,00 ⊢ 24,00	22,00	1	2,78
Total	—	36	100,00

Moda

$$s = \begin{bmatrix} 10(6) & 12(10) & 8(14) \\ 5(18) & 1(22) & . \end{bmatrix}$$

Considerei - ponto médio como representativo

Moda

$$Mo \simeq 10$$

Mediana

Foi verificado no RH da Empresa o a faixa de salário de cada funcionário casado . Os dados são os que seguem:

Classes de salários	Ponto médio s_i	Freqüência n_i	Porcentagem $100\ f_i$
4,00 ⊢ 8,00	6,00	10	27,78
8,00 ⊢ 12,00	10,00	12	33,33
12,00 ⊢ 16,00	14,00	8	22,22
16,00 ⊢ 20,00	18,00	5	13,89
20,00 ⊢ 24,00	22,00	1	2,78
Total	—	36	100,00

$$s = \begin{bmatrix} 10(6) & 12(10) & 8(14) \\ 5(18) & 1(22) & . \end{bmatrix}$$

Considerei - ponto médio como representativo

$$Md(S) \simeq \begin{cases} s_{\frac{n+1}{2}}, & \text{se } n \text{ é impar;} \\ \dfrac{s_{(\frac{n}{2})} + s_{(\frac{n}{2}+1)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

$$s_{(\frac{n}{2})} = \frac{36}{2} = s_{18} \Rightarrow s_{18} = 10$$

$$s_{(\frac{n}{2}+1)} = s_{(\frac{36}{2}+1)} = s_{19} \Rightarrow s_{19} = 10$$

$$Md(S) = \frac{10+10}{2} \simeq 10$$

Média Aproximada

Foi verificado no RH da Empresa o a faixa de salário de cada funcionário casado . Os dados são os que seguem:

Classes de salários	Ponto médio s_i	Freqüência n_i	Porcentagem $100\ f_i$
4,00 ⊢ 8,00	6,00	10	27,78
8,00 ⊢ 12,00	10,00	12	33,33
12,00 ⊢ 16,00	14,00	8	22,22
16,00 ⊢ 20,00	18,00	5	13,89
20,00 ⊢ 24,00	22,00	1	2,78
Total	—	36	100,00

$$s = \begin{bmatrix} 10(6) & 12(10) & 8(14) \\ 5(18) & 1(22) & . \end{bmatrix}$$

$$\bar{s} = \frac{(n_1 * s_1) + (n_2 * s_2) + \ldots + (n_k * s_k)}{n} = \frac{1}{n}\sum_{i=1}^{n} n_i s_i$$

$$\bar{s} \simeq \frac{(10 * 6) + (12 * 10) + (8 * 14) + (5 * 18) + (1 * 22)}{36} = 11{,}22$$

Mediana Verdadeira? - **Tabela 2.1**

$$s_{\left(\frac{n}{2}\right)} = \frac{36}{2} = s_{18} \Rightarrow s_{18} = 9{,}8$$

$$s_{\left(\frac{n}{2}+1\right)} = s_{\left(\frac{36}{2}+1\right)} = s_{19} \Rightarrow s_{19} = 10{,}53$$

$$Md(S) = \frac{9{,}8 + 10{,}53}{2} = 10{,}16$$

Qual a Média e a Moda verdadeira ?

Medidas de Dispersão

Apontar a heterogeneidade dos dados (indivíduos)
(Relação com o Risco)

Prêmio das ações das empresas em diferentes indústrias

grupo A (variável X): 3, 4, 5, 6, 7

grupo B (variável Y): 1, 3, 5, 7, 9

grupo C (variável Z): 5, 5, 5, 5, 5

grupo D (variável W): 3, 5, 5, 7

grupo E (variável V): 3, 5, 5, 6, 6

Olhando o prêmio médio

$$\bar{x} = \bar{y} = \bar{z} = \bar{w} = \bar{v} = 5{,}0$$

Não informa nada sobre o risco

Algumas medidas de interesse:

Desvio médio $$dm(x) = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Variância $$Var(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

Desvio padrão $$dp(x) = \sqrt{var(x)}$$

Grupo A

$$dm(x) = \frac{|(3-5)+(4-1)+(5-5)+(6-5)+(7-5)|}{5}$$

$$dm(x) = \frac{2+1+1+2}{5} = \frac{6}{5} = 1{,}2$$

$$Var(x) = \frac{(3-5)^2+(4-1)^2+(5-5)^2+(6-5)^2+(7-5)^2}{5}$$

$$Var(x) = \frac{(2)^2+(1)^2+(0)^2+(1)^2+(2)^2}{5} = \frac{4+1+1+4}{5} = \frac{10}{5} = 2$$

$$dp(x) = \sqrt{var(x)} = 1{,}41$$