Métodos Estatísticos Básicos

Aula 4 - Medidas de dispersão

Prof. Regis Augusto Ely

Departamento de Economia
Universidade Federal de Pelotas (UFPel)

Abril de 2014

Amplitude total

- **Amplitude total:** $AT = X_{max} - X_{min}$. É a única medida de dispersão que não tem na média o ponto de referência.
- Para dados agrupados sem intervalos de classe, a fórmula é a mesma acima.
- Para dados com intervalos de classe, $AT = L_{max} - l_{min}$, onde l_{min} é o menor limite inferior das classes e L_{max} o maior limite superior.
- Obs: a amplitude total desconsidera valores intermediários.

- **Desvio quartil:** $D_q = \frac{(Q3-Q1)}{2}$. É também chamado de amplitude semi-interquartílica.
- Usamos o desvio quartil preferencialmente quando a medida de tendência central utilizada é a mediana.
- O desvio quartil não é tao afetado por valores extremos.
  Ex: $\{40, 45, 48, 62, 70\}$
  $Q1 = \frac{40+45}{2} = 42,5$ e $Q3 = \frac{62+70}{2} = 66$
  $D_q = \frac{66-42,5}{2} = 11,75$

Desvio médio absoluto

- **Desvio médio absoluto:** $D_m = \frac{\sum_{i=1}^{n} |Xi - \bar{X}|}{n}$; $D_{me} = \frac{\sum_{i=1}^{n} |Xi - Me|}{n}$. É a média aritmética dos valores absolutos dos desvios tomados em relação à média ou à mediana.
  Ex: {-4, -3, -2, 3, 5}
  $\bar{X} = -0,2$ e $M_e = 2$.
  $D_m = \frac{|-4+0,2|+|-3+0,2|+|-2+0,2|+|3+0,2|+|5+0,2|}{5} = 3,36$
  $D_{me} = \frac{|-4+2|+|-3+2|+|-2+2|+|3+2|+|5+2|}{5} = 3$
- Para dados agrupados devemos utilizar as frequências,
  $D_m = \frac{\sum_{i=1}^{n} fi.(Xi - \bar{X})}{\sum_{i=1}^{n} fi}$, e se tivermos intervalos de classe, então Xi será o ponto médio de cada classe.

Diferença média

- **Diferença média:** $\triangle = \frac{1}{n^2}.\sum_{i=1}^{n}\sum_{j=1}^{n}|Xi - Xj|$. É o desvio absoluto em relação a todos os dados entre si.
- Essa expressão pode ser simplificada para $\triangle = \frac{4}{n^2}.\sum_{i=1}^{n} i.Xi - 2\bar{X}(1 + \frac{1}{n})$ (ver pag. 51 de *Hoffman, R. Estatística para economistas*).

Desvio-padrão

- **Desvio padrão:** $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \overline{X})^2}{n}}$. É a raíz da média aritmética dos quadrados dos desvios.
  Ex: {-4, -3, -2, 3, 5}
  $\bar{X} = -0,2$
  $\sigma = \sqrt{\frac{(-4+0,2)^2+(-3+0,2)^2+(-2+0,2)^2+(3+0,2)^2+(5+0,2)^2}{5}} = \sqrt{12,56} = 3,54$.

- **Desvio padrão amostral:** $S = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \bar{X})^2}{n-1}}$. Utilizamos essa pequena correção no caso de termos apenas uma amostra da população completa.

- Quando temos dados agrupados, devemos ponderar o desvio padrão pelas frequências:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}[(Xi - \overline{X})^2 . fi]}{\sum_{i=1}^{n} fi}}$$ quando se trata da população inteira.

$$S = \sqrt{\frac{\sum_{i=1}^{n}[(Xi - \overline{X})^2 . fi]}{\sum_{i=1}^{n} fi - 1}}$$ quando se trata de uma amostra.

- Com intervalos de classe, Xi será o ponto médio da classe.

- As principais propriedades do desvio padrão são:

1. Somando (ou subtraindo) uma constante a todos os valores de uma variável, o desvio-padrão não se altera.
2. Multiplicando (ou dividindo) todos os valores de uma variável por uma constante (diferente de zero), o desvio-padrão será multiplicado (ou dividido) por essa constante.

- **Variância** (σ^2ou S^2) **:** é o desvio-padrão elevado ao quadrado. A propriedade 1 continua válida para a variância, mas a propriedade 2 se altera, pois se multiplicarmos todos os valores por uma constante (diferente de zero), a variância será multiplicada por essa mesma constante elevada ao quadrado.

- **Coeficiente de variação de Pearson (CVP):** $CVP = \frac{\sigma}{\bar{X}} \times 100$. Caracteriza a dispersão dos dados em relação ao seu valor médio.
- Um desvio padrão de 2 pode ser grande para dados cuja média é 20, mas pequeno se a média é 200. O CVP padroniza as variações, possibilitando a comparação entre dados distintos.
- Ex:

variável	média	desvio
altura	175cm	5,0cm
peso	68kg	2,0kg

Qual série é mais homogênea?

$$=> CVP_{altura} = \frac{5{,}0}{175} x100 = 2,85\%. \; <=$$

$$CVP_{peso} = \frac{2{,}0}{68} x100 = 2,94\%.$$

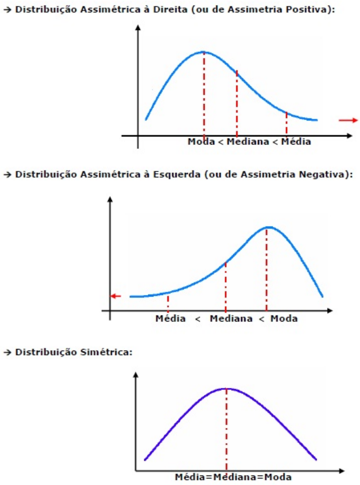
- **Coeficiente de variação de Thorndike (CVT):** $CVT = \frac{S}{Me} x100$. Utilizamos a mediana para o cálculo.

Medidas de assimetria

- **Distribuição simétrica:** dizemos que os dados tem uma distribuição simétrica quando Média = Mediana = Moda.
- **Distribuição assimétrica à esquerda:** é a assimétrica negativa, que ocorre quando Média < Mediana < Moda.
- **Distribuição assimétrica à direita:** é a assimétrica positiva, que ocorre quando Média > Mediana > Moda.
- **Coeficiente de assimetria de Pearson:** $CAP_{Me} = \frac{3.(\bar{X} - Me)}{\sigma}$ e $CAP_{Mo} = \frac{\bar{X} - Mo}{\sigma}$. Compara graus de assimetria entre distribuições diferentes.
- **Classificação:**
  1. $|CAP| < 0,15 \Rightarrow$ Assimetria pequena.
  2. $0,15 < |CAP| < 1 \Rightarrow$ Assimetria moderada.
  3. $|CAP| > 1 \Rightarrow$ Assimetria elevada.

Medidas de assimetria

- Se $CAP = 0$, os dados tem distribuição simétrica. Se $CAP < 0$ a assimetria é negativa. Se $CAP > 0$ a assimetria é positiva.



- **Curtose:** é o grau de achatamento de uma distribuição em relação à distribuição normal (em forma de sino).
- **Distribuição leptocúrtica:** apresenta uma distribuição mais alongada do que a normal.
- **Distribuição platicúrtica:** apresenta uma distribuição mais achatada do que a normal.
- **Distribuição mesocúrtica:** distribuição não é nem achatada nem alongada (igual a da normal).

Medidas de curtose

- **Percentílico de curtose:** $C1 = \frac{(Q3-Q1)}{2(P90-P10)}$.
  $C1 = 0,263 \Rightarrow$ curva mesocúrtica.
  $C1 < 0,263 \Rightarrow$ curva leptocúrtica.
  $C1 > 0,263 \Rightarrow$ curva platicúrtica.
- **Momento de curtose:** $K = \frac{\sum_{i=1}(Xi-\overline{X})^4.fi}{\sum_{i=1}^{n} fi}\frac{1}{S^4}$.
- $K = 3 \Rightarrow$ curva mesocúrtica.
- $K > 3 \Rightarrow$ curva leptocúrtica.
- $K < 3 \Rightarrow$ curva platicúrtica.

- Os valores dos coeficientes de curtose determinam o grau de achatamento da distribuição. Graficamente,