Estatística Descritiva

Amostras univariadas

Estatísticas de localização

	Dados não agrupados	Dados discretos agrupados	Dados contínuos agrupados
Média amostral ($\bar{x}$)	$\bar{x} = \frac{1}{N} \cdot \sum_{n=1}^{N} x_n$ N: número de dados da amostra; x_n: valor de cada um dos dados amostrais, $n = 1, \ldots, N$.	$\bar{x} = \sum_{k=1}^{K} f_k \cdot x_k$ K: número de células; f_k: frequência relativa (N_k/N) de cada valor observado.	$\bar{x} = \sum_{k=1}^{K} f_k \cdot M_k$ M_k: ponto central da célula k.
Mediana amostral (*Med*) (ou percentil 50 ou segundo quartil)	Considere-se que os dados da amostra são colocados por ordem crescente dos seus valores, formando um vector $(x_1^*, x_2^*, x_3^*, \ldots, x_N^*)$. - Se o número de dados que constituem a amostra (*N*) for ímpar, a mediana toma o valor do dado que naquele vector ocupa a posição central: $(Med = x^*_{(N+1)/2})$. - Se o número de dados for par, a mediana toma o valor médio dos dois termos cujas localizações no vector mais se aproximam da posição central: $\left(\mathrm{Med} = \frac{x^*_{N/2} + x^*_{(N/2)+1}}{2}\right)$.		$Med \cong LI_{med} + \frac{0.5 - fa^-_{med}}{f_{med}} \cdot \Delta_{med}$ LI_{med}: limite inferior da célula mediana (i.e., da célula para a qual a frequência acumulada passam de um valor inferior a 0.5 para um valor superior a 0.5); fa^-_{med}: frequência relativa acumulada correspondente à célula que precede a célula mediana; f_{med}: freq. relativa (não acumulada) correspondente à célula mediana; Δ_{med}: amplitude da célula mediana
Moda (*Mod*)	A moda é o valor dos dados que ocorre com mais frequência (se existirem 2 ou mais valores adjacentes para os quais a frequência seja máxima, a moda será a média desses valores).		$Mod = LI_{med} + \frac{d_1}{d_1 + d_2} \cdot \Delta$ LI: limite inferior da célula modal (i.e., da célula com maior freq. relativa); d_1: diferença entre a frequência relativa da célula modal e a frequência relativa da célula precedente; d_2: diferença entre a frequência relativa da célula modal e a frequência relativa da célula seguinte; Δ: amplitude da célula modal.

Estatísticas de dispersão

	Dados não agrupados	Dados discretos agrupados	Dados contínuos agrupados
Amplitude da amostra (*A*)	Diferença entre o valor máximo e o valor mínimo dos dados.		Diferença entre o ponto central da célula com o valor máximo e o ponto central da célula com o valor mínimo.
Intervalo Interquartis (*IIQ*)	Intervalo cujos extremos são o primeiro quartil (ou percentil 25) e o terceiro quartil (ou percentil 75).		Intervalo cujos extremos são o primeiro quartil (ou percentil 25) e o terceiro quartil (ou percentil 75). $Primeiro\ quartil \cong LI_{1ºQ} + \frac{0.25 - fa^-_{1ªQ}}{f_{1ªQ}} \cdot \Delta_{1ªQ}$ $Terceiro\ quartil \cong LI_{3ºQ} + \frac{0.75 - fa^-_{3ªQ}}{f_{3ªQ}} \cdot \Delta_{3ªQ}$ $LI_{1ªQ}$ ($LI_{3ªQ}$): limite inferior da célula correspondente ao 1ª quartil (3º quartil), i.e., da célula onde as freq. acumuladas passam de um valor inferior a 0.25 (0.75) para um valor superior a 0.25 (0.75); $fa^-_{1ºQ}$ ($fa^-_{3ºQ}$): frequência relativa acumulada correspondente à célula que precede a célula correspondente ao 1º quartil (3º quartil); $f_{1ºQ}$ ($f_{3ºQ}$): frequência relativa (não acumulada) correspondente à célula do 1º quartil (3º quartil); $\Delta_{1ºQ}$ ($\Delta_{3ºQ}$): amplitude da célula do 1º quartil (3º quartil).
Desvio absoluto médio amostral (*DAM*)	$DAM = \frac{1}{N} \cdot \sum_{n=1}^{N} \lvert x_n - \bar{x} \rvert$	$DAM = \sum_{k=1}^{K} f_k \cdot \lvert x_k - \bar{x} \rvert$	$DAM = \sum_{k=1}^{K} f_k \cdot \lvert M_k - \bar{x} \rvert$
Desvio quadrático médio amostral (*DQM*)	$DQM = \frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \bar{x})^2$	$DQM = \sum_{k=1}^{K} f_k \cdot (x_k - \bar{x})^2$	$DQM = \sum_{k=1}^{K} f_k \cdot (M_k - \bar{x})^2$
Variância amostral (s^2)	$s^2 = \frac{1}{N-1} \cdot \sum_{n=1}^{N} (x_n - \bar{x})^2$	$s^2 = \frac{N}{N-1} \sum_{k=1}^{K} f_k \cdot (x_k - \bar{x})^2$	$s^2 = \frac{N}{N-1} \sum_{k=1}^{K} f_k \cdot (M_k - \bar{x})^2$
Desvio padrão amostral (*s*)	$s = \sqrt{\frac{1}{N-1} \cdot \sum_{n=1}^{N} (x_n - \bar{x})^2}$	$s = \sqrt{\frac{N}{N-1} \sum_{k=1}^{K} f_k \cdot (x_k - \bar{x})^2}$	$s = \sqrt{\frac{N}{N-1} \sum_{k=1}^{K} f_k \cdot (M_k - \bar{x})^2}$

Outras estatísticas

	Dados não agrupados	**Dados discretos agrupados**	**Dados contínuos agrupados**
Momento ordinário de ordem i (m'_i)	$m'_i = \frac{1}{N} \cdot \sum_{n=1}^{N} (x_n)^i$ $(i = 1,2, \ldots)$	$m'_i = \sum_{k=1}^{K} f_k \cdot (x_k)^i$ $(i = 1,2, \ldots)$	$m'_i = \sum_{k=1}^{K} f_k \cdot (M_k)^i$ $(i = 1,2, \ldots)$
Momento centrado de ordem i (m_i)	$m_i = \frac{1}{N} \cdot \sum_{n=1}^{N} (x_n - \bar{x})^i$ $(i = 1,2, \ldots)$	$m_i = \sum_{k=1}^{K} f_k \cdot (x_k - \bar{x})^i$ $(i = 1,2, \ldots)$	$m_i = \sum_{k=1}^{K} f_k \cdot (M_k - \bar{x})^i$ $(i = 1,2, \ldots)$
Medida da assimetria (k_3) (grande população; amostra limitada)	$k_3 = \frac{N^2}{(N-1) \cdot (N-2)} \cdot m_3$		
Coeficiente de assimetria amostral (g_1)	$g_1 = \frac{k_3}{s^3}$		
Medida de *kurtose* (k_4) (grande população; amostra limitada)	$k_4 = \frac{N^2}{(N-1) \cdot (N-2) \cdot (N-3)} \cdot [(N+1) \cdot m_4 - 3 \cdot (N-1) \cdot m_2^2]$		
Coeficiente de *kurtose* amostral (g_2)	$g_2 = \frac{k_4}{s^4}$		

Fórmulas de cálculo alternativas:

$$\sum_{n=1}^{N} (x_n - \bar{x})^2 = \left(\sum_{n=1}^{N} x_n^2 \right) - N\bar{x}^2$$

Adição de uma constante c a todos os valores da amostra ($y_n = x_n + c$):

$$\bar{y} = \bar{x} + c$$

$$y_n - \bar{y} = x_n + c - (\bar{x} + c) = x_n - \bar{x}$$

Amostras bivariadas

Ajuste de uma relação linear ($y = a' + b \cdot x$) (Método dos mínimos quadrados)	$a' = \bar{y} - b \cdot \bar{x}$ $b = \frac{s_{XY}}{s_{XX}}$	onde: $s_{XY} = \sum_{n=1}^{N} (x_n - \bar{x}) \cdot (y_n - \bar{y})$ $s_{XX} = \sum_{n=1}^{N} (x_n - \bar{x})^2$
Covariância amostral (c_{XY})	$c_{xy} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x}) \cdot (y_n - \bar{y})$	
Coeficiente de correlação amostral (r_{XY})	$r_{XY} = \frac{c_{XY}}{s_X \cdot s_Y}$ fórmula de cálculo alternativa: $r_{xy} = \frac{s_{xy}}{\sqrt{s_{XX} \cdot s_{YY}}}$	onde: s_X é o desvio padrão amostral da variável *X* s_Y é o desvio padrão amostral da variável *Y* $s_{YY} = \sum_{n=1}^{N} (y_n - \bar{y})^2$
Coeficiente de determinação amostral (r^2_{XY})	$r^2_{XY} = \frac{b^2 \cdot s_{XX}}{s_{YY}}$	
Coeficiente de determinação amostral corrigido (r^2_{XY} (corrigido))	$r^2_{XY}(\text{corrigido}) = 1 - \frac{\sum_{n=1}^{N} e_n^2 \,/\, (N-2)}{s_{YY}/(N-1)}$	onde: $\sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = \sum_{n=1}^{N} (y_n - a' - b \cdot x_n)^2 = \sum_{n=1}^{N} e_n^2$

Fórmulas de cálculo alternativas:

$$s_{XX} = \sum_{n=1}^{N} (x_n - \bar{x})^2 = \left(\sum_{n=1}^{N} x_n^2 \right) - N\bar{x}^2$$

$$s_{XY} = \sum_{n=1}^{N} (x_n - \bar{x})(y_n - \bar{y}) = \left(\sum_{n=1}^{N} x_n y_n \right) - N\bar{x}\bar{y}$$

Formulário adaptado de:
Estatística
Rui Campos Guimarães, José A. Sarsfield Cabral
Verlag Dashöfer