UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA
MAT02214 - Estatística Geral 1
ÁREA 1

FORMULÁRIO

Tabela de Frequências

- **Frequências absolutas e relativas:** seja F_j a frequência absoluta do valor/classe j e n o total de elementos do conjunto de dados. A frequência relativa é dada por $f_j = F_j/n$.
- **Número de classes (k):** $k = \sqrt{n}$ (regra empírica); $k = 1 + 3,32\log(n)$, em que k é número de classes e n o número de observações. Ainda, $i = a_t/k$, em que i é a amplitude do intervalo, $a_t = x_{(n)} - x_{(1)}$ é a amplitude total, e $x_{(1)}, x_{(n)}$ são os extremos inferior e superior do conjunto de dados.

Gráficos

- **Histograma:** a densidade de frequência é dada pela razão entre a frequência relativa da classe e a base (comprimento do intervalo) da classe.

Medidas Descritivas

Medidas de localização

- **A média amostral (aritmética simples)** é dada por $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1+x_2+\ldots+x_n}{n}$.
- Suponha que para os dados $x_1, x_2, \ldots, x_n$ possuem os seguintes pesos $p_1, p_2, \ldots, p_n$, então **a média aritmética ponderada** é dada por $\bar{x}_p = \frac{\sum_{i=1}^{n} x_i p_i}{\sum_{i=1}^{n} p_i}$.
  - A média aritmética para dados **agrupados por intervalo de classe** é dada por $\bar{x} = \frac{\sum_j c_j F_j}{n}$ em que, c_j é o ponto médio do intervalo de classe.
- A posição da mediana é dada por $p = \frac{n+1}{2}$. A **mediana** é dada por $Md = x_{(p)}$, quando n é par, e $Md = [x_{(p_1)} + x_{(p_1)}]/2$, quando n é ímpar. As posições dos **quartis** são dadas por $p_1 = \frac{n+1}{4}; p_2 = \frac{2(n+1)}{4}; p_3 = \frac{3(n+1)}{4}$, se n é ímpar, e $p_1 = \frac{n+2}{4}; p_2 = \frac{2n+2}{4}; p_3 = \frac{3n+2}{4}$. Os quartis são dados por $Q_i = x_{(p_i)}$. Se p_i não for um número inteiro, então $Q_i = (x_{(\lceil p_i \rceil)} + x_{(\lfloor p_i \rfloor)})/2$, em que $\lceil p_i \rceil, \lfloor p_i \rfloor$ são o menor e o maior inteiro de p_i.

Medidas de variação ou dispersão

- A **amplitude total** é $a_t = x_{(n)} - x_{(1)}$, em que $x_{(1)}, x_{(n)}$ são os extremos inferior e superior do conjunto de dados.
- A **variância amostral** é dada por $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}$.
  - A variância amostral para dados **agrupados por intervalo de classe** é dada por $s^2 = \frac{\sum_j F_j(c_j - \bar{x})^2}{n-1}$ em que, c_j é o ponto médio do intervalo de classe.
- O **desvio padrão amostral** é igual a raiz quadrada da variância amostral: $s = \sqrt{s^2}$.
- O **coeficiente de varição** é $CV = s/\bar{x}$.

Medidas de formato

- O **coeficiente de assimetria** é $a_3 = m_3/(m_2\sqrt{m_2})$, em que $m_2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$ e $m_3 = \sum_{i=1}^{n}(x_i - \bar{x})^3/n$.
- O **coeficiente de curtose** é $a_4 = m_4/(m_2)^2$, em que $m_2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$ e $m_4 = \sum_{i=1}^{n}(x_i - \bar{x})^4/n$.

Medidas de associação

- A **covariância** amostral entre X e Y é dado por $cov(X,Y) = \frac{s_{xy}}{n}$, em que $s_{xy} = \sum_i (x_i - \overline{x})(y_i - \overline{y}) = \sum_i x_i y_i - [(\sum_i x_i)(\sum_i y_i)]/n$.
- O **coeficiente de correlação** linear amostral entre X e Y é dado por $r = corr(X,Y) = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$, em que
  - $s_{xx} = \sum_i (x_i - \overline{x})^2 = \sum_i x_i^2 - (\sum_i x_i)^2/n$,
  - $s_{yy} = \sum_i (y_i - \overline{y})^2 = \sum_i y_i^2 - (\sum_i y_i)^2/n$ e
  - $s_{xy} = \sum_i (x_i - \overline{x})(y_i - \overline{y}) = \sum_i x_i y_i - [(\sum_i x_i)(\sum_i y_i)]/n$;

  ou

$$r = corr(X,Y) = \frac{cov(X,Y)}{\sqrt{\frac{S_{XX}}{n}\frac{S_{YY}}{n}}}$$