Estatística I

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Universidade Federal de Viçosa
Campus UFV - Viçosa

Sumário

Medidas de Dispersão

O resumo de um conjunto de dados por uma única medida representativa de posição central esconde toda a informação sobre a variabilidade do conjunto de observações. Por exemplo, suponhamos que cinco grupos de alunos submeteram-se a um teste, obtendo-se as seguintes notas:

- 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

Vemos que $\bar{x} = \bar{y} = \bar{z} = \bar{w} = \bar{v} = 5$. A identificação de cada uma destas séries por sua média (5, em todos os casos) nada informa sobre suas diferentes variabilidades. Notamos, então, a conveniência de serem criadas medidas que sumarizem a variabilidade de um conjunto de observações e que nos permita, por exemplo, comparar conjuntos diferentes de valores, como os dados acima, segundo algum critério estabelecido.

Um critério frequentemente usado para tal fim é aquele que mede a dispersão dos dados em torno de sua média. A principal e mais conhecida é a variância amostral.

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

O princípio básico é analisar os desvios das observações em relação à média dessas observações. Desvio é interpretado como o afastamento de uma observação em relação a uma determinada medida de posição.

Variância Amostral

$$S^2(X) = \frac{SQD_X}{n-1} = \frac{\sum\limits_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1} = \frac{\sum\limits_{i=1}^{n} X_i^2 - \frac{\left(\sum\limits_{i=1}^{n} X_i\right)^2}{n}}{n-1} = \frac{\sum\limits_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

Variância Amostral

$$S^2(X) = \frac{SQD_X}{n-1} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

Medidas foram tomadas de 3 grupos, calcule a variância:

- Grupo 1 (em Kg): 1,3,5,7,9
- Grupo 2 (em metros): 1.8,3.8,5.8,7.8,9.8
- Grupo 3 (em R$): 1002,1004,1006,1008,1010

Sendo a variância uma medida de dimensão igual ao quadrado da dimensão dos dados (por exemplo, se os dados são expressos em *cm*, a variância será expressa em cm^2), ela pode causar problemas de interpretação. Costuma-se usar, então, o desvio padrão, que é definido como a raiz quadrada positiva da variância.

$$dp(X) = S(X) = \sqrt{Var(X)} \tag{1}$$

Ambas as medidas de dispersão (*Dm* e *dp*) indicam em média qual será o “erro” (desvio) cometido ao tentar substituir cada observação pela medida resumo do conjunto de dados (no caso, a média). Resolvido, portanto, o problema da unidade de medida.

Coeficiente de Variação

$$CV(X) = \frac{dp(X)}{\bar{X}} \cdot 100 = \frac{S(X)}{\bar{X}} \cdot 100 \qquad (2)$$

Mesmo o DP pode induzir à conclusões errôneas com relação à variabilidade. Suponha dois conjuntos de dados $D_1 = \{10, 20, 30\}$ e $D_2 = \{10000, 10010, 10020\}$. Note que nestes casos $\bar{x}_1 = 20, dp(x) = 10, \bar{x}_2 = 10010$ e $dp(x_2) = 10$. Porém, em termos percentuais, o primeiro conjunto de dados é mais heterogênio.

Coeficiente de Variação

Obs.: O C.V. é utilizado para avaliar qual o percentual da média que o desvio-padrão representa. Isso é chamado de homogeneidade. Na situação em que as amostras possuem a mesma média, a conclusão pode ser feita a partir da comparação de suas variâncias. Para amostras com médias diferentes, aquela que apresentar menor CV, é a mais homogênea.

Erro Padrão da Média

$$S(\overline{X}) = \sqrt{\frac{S^2(X)}{n}}$$

Obs.: É uma medida utilizada para avaliar a precisão da média.

Erro Padrão da Média

$$S(\overline{X}) = \sqrt{\frac{S^2(X)}{n}}$$

Obs.: É uma medida utilizada para avaliar a precisão da média.

Exemplo

Considere duas amostras de tamanhos $n = 6$, em que $S_A^2 = 5,6$ e $S_B^2 = 28$. Temos que:

$$S(\bar{X}_A) = \sqrt{\frac{5,6}{6}} = 0,966; \quad S(\bar{X}_B) = \sqrt{\frac{28}{6}} = 2,1602$$

e, portanto, a amostra A forneceu uma estimativa de média associada à uma maior precisão.

Erro Padrão da Média

Notem que o erro padrão da média é:

- Inversamente proporcional ao tamanho da amostra;
- Diretamente proporcional à variância da amostra.

Amplitude Total

A amplitude total (AT) é dada pela diferença entre o maior e o menor valor de uma amostra ou de um conjunto de dados. Se $X_1, X_2, X_3, \cdots, X_n$ é uma amostra de valores da variável X, então:

$$AT_X = X_{(n)} - X_{(1)}$$

Recorde que a notação $X_{(i)}$ indica estatísticas de ordem da amostra, isto é: $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq \cdots \leq X_{(n)}$. Portanto, a amplitude total indica que o desvio entre duas observações quaisquer é no máximo igual a AT.

Referências

- Bastos, Fernando de Souza (2025). ***Apostila Interativa.*** Disponível online: https://ufvest.shinyapps.io/ApostilaInterativa/.
- Ferreira, Eric Batista e Marcelo Silva de Oliveira (2020). ***Introdução à Estatística com R.*** Editora Universidade Federal de Alfenas. URL: https://www.unifal-mg.edu.br/bibliotecas/wp-content/uploads/sites/125/2021/12/32-EBR_Unifal.pdf.
- Montgomery, D. C. e G. C Runger (2016). ***Estatística Aplicada E Probabilidade Para Engenheiros.*** 6ª ed. São Paulo: Grupo Gen-LTC.

- Morettin, P.A. e W.O Bussab (2023). ***Estatística básica***. 10ª ed. São Paulo: Editora Saraiva.
- Peternelli, Luiz Alexandre (s.d.). ***Apostila (EST 106)***. Formato slide Disponível no PVANet - Moodle.