Aula 04: Medidas de Dispersão

Prof. Daniel Caetano

Objetivo: Apresentar os conceitos de medidas de dispersão de um conjunto de dados.

Bibliografia:

- FONSECA, J. S; MARTINS, G. A. Curso de Estatística. São Paulo: Atlas, 1996.
- NETO, P. L. O. C. Estatística. São Paulo: Edgard Blücher, 2002.
- NETO, P. L. O. C; CYMBALISTA, M. Probabilidades. São Paulo: Edgard Blücher, 1974.

1. MEDIDAS DE DISPERSÃO

Medidas de Dispersão são valores estatísticos que ofereçam algumas informações sobre a variação do valor de uma variável aleatória (ou sobre a amostra à qual ela corresponde).

1.1. Variância

A variância é uma medida do quanto o valor da variável varia em torno da média. É definida pela expressão, para dados populacionais:

$$Var[X] = \sigma_x^2 = \frac{\sum_{i=0}^{n}(X_i - \bar{X})^2}{n}$$

ou:

$$Var[X] = \sigma_x^2 = \sum_{i=0}^{n}(X_i - \bar{X})^2 . FR_i$$

Para dados de amostra, não se usa como divisor o valor "n", e sim "n-1" e não se calcula variância para uma amostra de um único elemento:

$$Var[X] = S_x^2 = \frac{\sum_{i=0}^{n}(X_i - \bar{X})^2}{n-1}$$

Em todos os casos, note que as diferenças entre os valores de X e a média de X é elevada ao quadrado, para evitar que erros positivos cancelem erros negativos. O lado negativo desta estratégia é que os erros ficam elevados ao quadrado.

Quando os dados estão agrupados em classes, usa-se como valor Xi o valor central de cada classe.

Propriedades:

1. A variância de uma constante é zero.
2. Multiplicando uma variável aleatória por ***c***, sua variância fica multiplicada por c^2.
3. Somando-se ou subtraindo-se um valor de uma variável aleatória, sua variância não se altera.
4. A variância da soma ou da subtração de duas variáveis aleatórias independentes é a soma ou subtração das suas respectivas variâncias.

1.2. Desvio Padrão

O desvio padrão também mede a variação de uma variável em torno de sua média, mas sem o problema de estar elevado ao quadrado. O desvio padrão é definido como:

$$\sigma_x = \sqrt{\sigma_x^2}$$

Para o desvio padrão populacional, se usa, como indicado nas fórmulas, a letra sigma. Para o desvio padrão da amostra, usa-se a letra **S**.

Características:

1. Quanto menor o desvio padrão, mais os valores estão próximos à média.
2. Em oposição, quanto mais os valores se afastarem da média, maior o desvio padrão.
3. Duas amostras com a mesma média mas com desvios-padrão diferentes terão distribuições de valores diferentes, isto é, uma delas tendo valores mais próximos da média e outra com valores mais distantes da média.

Propriedades:

1. Somando ou subtraindo uma mesma constante ***c*** de cada elemento da amostra não irá alterar o desvio padrão.
2. Multiplicando ou dividindo os valores da amostra por uma constante ***c*** fará com que o desvio padrão seja multiplicado ou dividido por essa mesma constante.

1.3. Coeficiente de Variação de Pearson

Indica uma relação porcentual entre o desvio padrão e a média, podendo ser calculado, a depender de serem dados populacionais ou amostrais:

$$CV = \frac{\sigma}{\overline{X}}.100\%$$

ou

$$CV = \frac{S}{\overline{X}}.100\%$$

A dispersão é:

a) Baixa para 0 ≤ CV ≤ 15%

b) Média para 15 < CV ≤ 30%

c) Alta para CV > 30%

2. INTERPRETAÇÃO DO DESVIO PADRÃO

Em amostras/populações que obedecem uma distribuição normal, isto é, possuem o formato de um sino, possuindo um valor "verdadeiro" e outros se distribuindo ao redor desse (em geral várias medidas de uma mesma coisa, por exemplo, várias medidas de massa de um corpo, ou várias medidas de distância entre duas cidades) é possível dar uma interpretação ao desvio padrão.

Nesses casos, tanto para amostras quanto para populações, pode-se dizer que aproximadamente 68% dos valores/medidas se situam no intervalo entre a média e um desvio padrão, para cima ou para baixo.

Da mesma maneira, pode-se dizer que aproximadamente 95% dos valores/medidas se situam no intervalo entre a média e dois desvios padrão, para cima ou para baixo. Finalmente, pode-se dizer que 99,7% dos valores/medidas, se situam no intervalo entre a média e três desvios padrão, para cima ou para baixo.

Essas porcentagens são obtidas pela área sob a curva do histograma das frequências de cada valor/categoria considerada. Essas porcentagens significam a "probabilidade" de ocorrência de um determinado valor.

3. BIBLIOGRAFIA

MARINHO, P. Probabilidade e Estatística Aplicada à Engenharia. Rio de Janeiro, SESES, 2016.
FONSECA, J. S; MARTINS, G. A. Curso de Estatística. São Paulo: Atlas, 1996.
NETO, P. L. O. C. Estatística. São Paulo: Edgard Blücher, 2002.
NETO, P. L. O. C; CYMBALISTA, M. Probabilidades. São Paulo: Edgard Blücher, 1974.