Anexo I

Estatística Básica Aplicada a um Conjunto de Números

Podemos extrair uma quantidade maior de informações de nossas medições efetuando alguns cálculos estatísticos básicos com elas, como veremos a seguir.

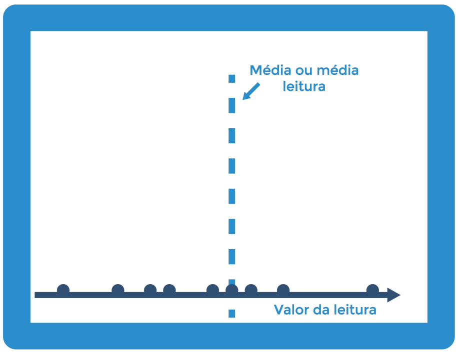
1. Média Aritmética

Se repetidas medições dão origem a resultados um pouco diferentes entre si, não significa que estejamos medindo errado. Isso pode ser devido a variações naturais no que está acontecendo. Por exemplo, o valor da velocidade do vento na rua normalmente não será constante. Além disso, os instrumentos de medição muitas vezes não apresentam um comportamento completamente estável (e uma trena pode deformar-se e fornecer resultados diferentes). Se houver variação entre as leituras quando elas são repetidas, é melhor calcular a **média aritmética**.

A média aritmética nos fornece uma estimativa do valor "verdadeiro" do mensurando. Ela é normalmente representada por um símbolo com uma barra em cima.

A figura 1 ilustra um conjunto de valores e sua média aritmética. O exemplo 1 mostra como calcular uma média aritmética.

Figura 1- Ilustração de um conjunto de valores e da sua média aritmética



Exemplo 1 – Cálculo da média de um conjunto de dados

Consideremos um conjunto de 10 medições. Para calcular a média aritmética, devemos somar todos os dados e dividir o total pelo número de medições (neste caso, 10). Medições: 16, 19, 18, 16, 17, 19, 20, 15, 17, 13. O somatório dos dez valores resulta em 170. Então a média aritmética do conjunto de dados em questão é:

$$\frac{170}{10} = 17$$

De maneira geral, **quanto mais resultados tivermos, melhor será a estimativa que faremos do valor "verdadeiro"**. O ideal seria calcular a média aritmética de um conjunto infinito de valores, pois, quanto mais resultados utilizarmos, mais perto da estimativa ideal estaremos. Contudo, a realização de muitas medições nem sempre compensa o esforço despendido. Em geral, **quatro a dez medições são suficientes**.

2. Desvio-padrão

Quando repetidas medições nos fornecem diferentes resultados, é importante termos conhecimento da dispersão desses dados. **A dispersão dos valores nos dá informação sobre a incerteza da medição**. Sabendo a magnitude da dispersão, podemos começar a julgar a qualidade de nossas medições.

A forma usual de quantificarmos a dispersão é através do cálculo do desvio-padrão. O desvio-padrão de um conjunto de dados nos informa sobre a distância entre cada dado e a média aritmética do conjunto.

De forma genérica, aproximadamente dois terços dos dados de um determinado conjunto estarão compreendidos entre mais ou menos um desvio-padrão da média aritmética; aproximadamente 95% dos dados estarão entre mais ou menos dois desvios-padrão. Essas constatações são de grande aplicabilidade, mas não são universais.

O valor “verdadeiro” para o desvio-padrão apenas pode ser calculado a partir de um conjunto muito grande (infinito) de dados. Na prática, apenas uma estimativa do desvio-padrão pode ser encontrada. A letra “s” normalmente é utilizada para representar o desvio-padrão estimado. O exemplo 2 mostra como calcular o desvio-padrão estimado.

Exemplo 2 – Cálculo do desvio-padrão estimado de um conjunto de dados

Vamos utilizar o conjunto de dados do exemplo 1: 16, 19, 18, 16, 17, 19, 20, 15, 17, 13.

A primeira coisa que temos a fazer é calcular a média aritmética, que, de acordo com o exemplo 1, é 17. Depois temos que calcular a diferença entre cada valor e a média aritmética:

-1, +2, +1, -1, 0, +2, +3, -2, 0 e -4.

E elevar cada diferença ao quadrado:

1, 4, 1, 1, 0, 4, 9, 4, 0 e 16

Após, somamos os quadrados obtidos e dividimos por n-1 (neste caso n é 10, então n - 1 = 9):

$$\frac{1+4+1+1+0+4+9+4+0+16}{10-1}=\frac{40}{9}=4,44$$

Por fim, o desvio-padrão estimado “s” é encontrado extraindo a raiz quadrada do resultado obtido na operação aqui realizada:

$$s = \sqrt{4,44} = 2,1$$

Todas as operações que acabamos de realizar podem ser resumidas, para um conjunto de n dados, na equação 1:

$$s = \frac{\sqrt{\Sigma_i^n = 1\left(x_i - \overline{x}\right)^2}}{\left(n-1\right)} \qquad (1)$$

Onde xi é o iésimo resultado de um conjunto de medições e é a média aritmética dos n resultados em questão.

3. Distribuições de Probabilidade

A dispersão de um conjunto de valores pode assumir diferentes formas ou distribuições de probabilidade.

Uma distribuição de probabilidade é uma relação entre o resultado de um experimento estatístico e a probabilidade de sua ocorrência.

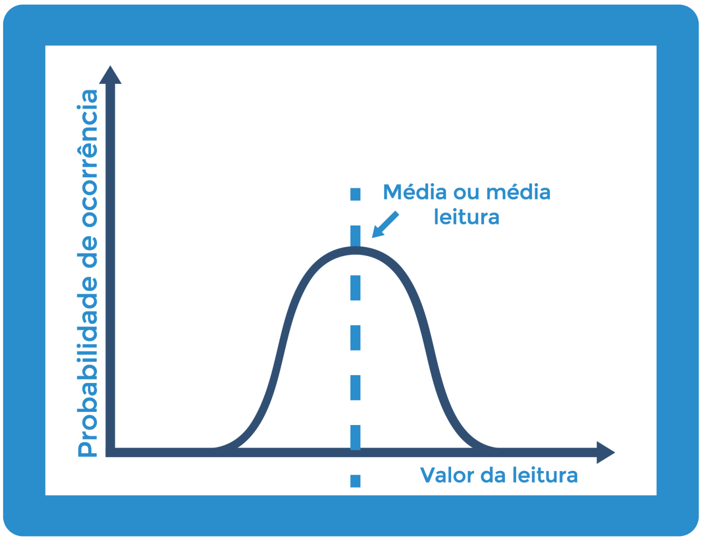
As distribuições de probabilidade podem assumir diferentes formas quando representadas graficamente, conforme veremos a seguir.

a) Distribuição Normal

Em um conjunto de leituras, algumas vezes os valores apresentam maior probabilidade de estarem próximos à média aritmética do que distantes dela. Essa característica é típica de uma **distribuição normal** ou **gaussiana**. Podemos observar esse tipo de distribuição se examinarmos as alturas de indivíduos de uma amostra representativa de homens.

A maioria dos homens tem a altura próxima à média aritmética das alturas. Poucos são muito altos ou muito baixos. Um esboço de uma distribuição normal é mostrado na figura 2.

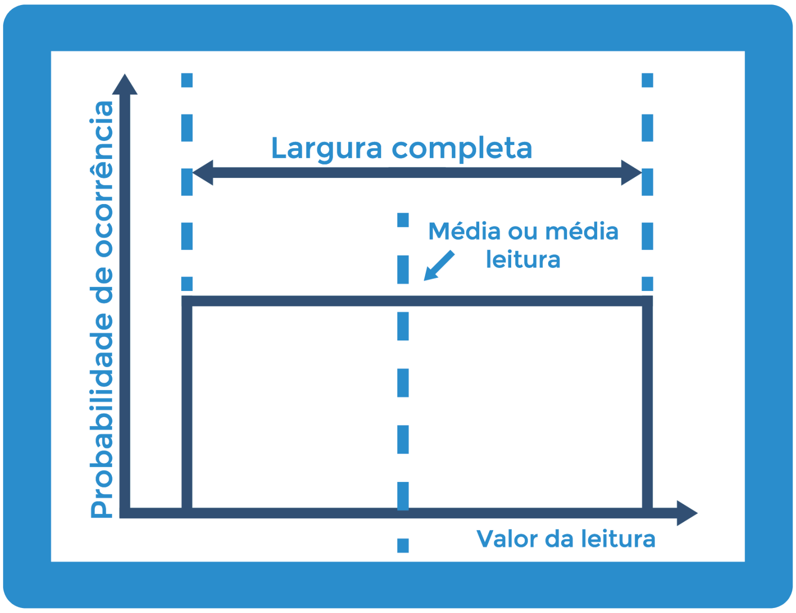
Figura 2 - Exemplo de uma distribuição normal para a medida de altura de uma quantidade de pessoas (a linha tracejada corresponde à média aritmética das alturas medidas dos indivíduos)



b) Distribuição Retangular ou Uniforme

Quando os valores medidos estão espalhados de forma "parelha", entre o maior e o menor valor medido, temos uma distribuição retangular ou uniforme. Isso pode ser observado ao examinarmos a forma como gotas de chuva caem em um fio de luz, fino e longo, por exemplo. A probabilidade de elas caírem em uma parte ou outra do fio é a mesma. Um esboço de uma distribuição uniforme é apresentado na figura 3.

Figura 3 - Exemplo de distribuição retangular ou uniforme



c) Outras Distribuições

Mais raramente, outras distribuições de probabilidade podem ocorrer, como a triangular e a bimodal.

Glossário

Distribuição normal ou gaussiana: distribuição estatística das probabilidades, contínua, parametrizada por um número real e o desvio padrão. Este tipo de distribuição se apresenta em forma de sino, unimodal, simétrica em relação a sua média.