Desvio Padrão

From Wikipedia, the free encyclopedia

probabilidade e estatística, o **desvio padrão** de uma distribuição de probabilidade, de uma variável aleatória, ou população é uma medida do espalhamento dos seus valores. Ele é geralmente denotado pela letra σ (sigma minúsculo). Ele é definido como a raiz quadrada da variância.

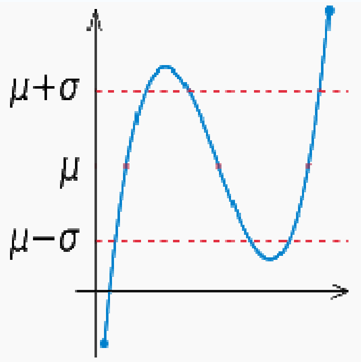
Para entender o desvio padrão, lembre-se que a variância é a *média* das diferenças ao quadrado entre os pontos dados e a média. A variância está tabulada em unidades quadradas. O desvio padrão, sendo a raiz quadrada daquela quantidade, portanto mede o espalhamento dos dados ao redor da média, medida na mesma unidade que os dados.

Dito mais formalmente, o desvio padrão é a raiz quadrática média (RMS) dos desvios dos valores de sua média aritmética.

Por exemplo, na população {4, 8}, a média é 6 e os desvios da média são {−2, 2}. Estes desvios quadráticos são {4, 4} a média dos quais (a variância) é 4. Portanto, o desvio padrão é 2. Neste caso 100% dos valores da população estão a um desvio padrão da média.

Oe desvio padrão é a medida mais comum da dispersão estatística, medindo quão largamente estão espalhados os valores num conjunto de dados. Se muitos pontos dados estão próximos à média, então o desvio padrão é pequeno; se muitos pontos dados estão longe da média, então o desvio padrão é grande. Se todos os valores dados forem iguais, então o desvio padrão é zero.

Para uma população, o desvio padrão pode ser estimado por um desvio padrão modificado (*s*) de uma amostra. As fórmulas são dadas abaixo.



Dada uma variável aleatória (em azul), o desvio padrão σ é uma medida do espalhamento dos valores da variável aleatória para longe de sua média μ.

Contents

[hide]

1 Definition and calculation

1.1 A simple example

1.2 Desvio padrão of a variável aleatória

1.3 Estimating população desvio padrão from amostra desvio padrão

1.4 Desvio padrão of a continuous variável aleatória

2 Example

3 Interpretation and application

3.1 Real-life examples

- 3.1.1 Weather
- 3.1.2 Sports
- 3.1.3 Finance

3.2 Geometric interpretation

3.3 Rules for normally distributed data

3.4 Chebyshev's inequality

4 Relationship between desvio padrão and mean

5 Rapid calculation methods

6 See also

7 External links

Definição e cálculo

Um exemplo simples

Suponha que desejamos encontrar o desvio padrão do conjunto dos números 4 e 8.

passo1: encontre média aritmética (ou média) de 4 e 8,

$$(4 + 8) / 2 = 6.$$

Passo 2: Encontre a diferença entre cada número e a média,

$$4 - 6 = - 2$$
$$8 - 6 = 2.$$

Passo 3: Eleve ao quadrado cada uma das diferenças

$$(- 2)^2 = 4$$
$$2^2 = 4.$$

Passo 4: some as diferenças obtidas,

$4 + 4 = 8.$

Passo 5: divide a soma pela quantia de pontos (aqui temos dois números),

$8 / 2 = 4.$

Passo 6: tome a raiz quadrada não negativa do quociente,

$$\sqrt{4} = 2.$$

Assim, o desvio padrão é 2.

Desvio padrão de uma variável aleatória

O desvio padrão de uma variável aleatória *X* é definido como:

$$\sigma = \sqrt{\operatorname{E}((X - \operatorname{E}(X))^2)} = \sqrt{\operatorname{E}(X^2) - (\operatorname{E}(X))^2}$$

onde E(*X*) é o valor esperado de *X*.

Nem *todas* variáveis aleatórias tem um desvio padrão, pois estes valores esperados não precisam existir. Por exemplo, o desvio padrão de uma variável aleatória que segue uma distribuição de Cauchy é indefinida porque seu E(*X*) é indefinido.

Se a variável aleatória *X* assume os valores $x_1, \dots, x_N$(os quais são números reais) com igual probabilidade, então seu desvio padrão pode ser calculado como segue. Primeiro, a média de *X*, $\overline{x}$, é definida como um somatório:

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

onde *N* é o número de amostras tomadas. Depois, o desvio padrão simplifica para

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2}.$$

Em outras palavras, o desvio padrão de uma variável aleatória discreta uniforme *X* pode ser calculado como segue:

1. Para cada valor x_i calcule a diferença $x_i - \overline{x}$ entre x_i e o valor médio $\overline{x}$.
2. Calcule os quadrados destas diferenças.
3. Encontre a média das diferenças quadradas. Esta quantidade é a variância σ^2.
4. Tome a raiz quadrada da variância.

A expressão acima pode também ser trocada com

$$\sigma = \sqrt{\frac{1}{N}\left(\sum_{i=1}^{N} x_i^2 - N\overline{x}^2\right)}.$$

A igualdade destas duas expressões pode ser mostrada com um pouco de álgebra:

$$\begin{aligned}\sum_{i=1}^{N}(x_i - \overline{x})^2 &= \sum_{i=1}^{N}(x_i^2 - 2x_i\overline{x} + \overline{x}^2)\\ &= \left(\sum_{i=1}^{N} x_i^2\right) - \left(2\overline{x}\sum_{i=1}^{N} x_i\right) + N\overline{x}^2\\ &= \left(\sum_{i=1}^{N} x_i^2\right) - 2\overline{x}(N\overline{x}) + N\overline{x}^2\\ &= \left(\sum_{i=1}^{N} x_i^2\right) - N\overline{x}^2.\end{aligned}$$

Estimando o desvio padrão da população a partir do desvio padrão amostra

No mundo real, encontrar o desvio padrão de uma população toda não é realística, exceto em certos casos, tais como standardized testing, onde cada membro de uma população é tirado da amostra. Na maioria dos casos, o desvio padrão é estimado examinando uma amostra aleatória tirada da população. A medida mais comum usada é o *desvio padrão da amostra*, que é definido por

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

onde $\{x_1, x_2, ..., x_N\}$é a amostra e $\overline{x}$ é a média da amostra. O denominador *N* − 1 é o número de graus de liberdade no vetor $(x_1 - \overline{x}, ..., x_N - \overline{x})$.

A razão para esta definição é que s^2 é um estimador imparcial para a variância σ^2 da população subjacente, se esta variância existe e os valores da amostra são extraídos independentemente com reposição. Entretanto, *s não* é um estimado imparcial para o desvio padrão σ; ele tende a subestimar o desvio padrão da população. Embora um estimador imparcial para σ é conhecido quando a variável aleatória está distribuída normalmente, a fórmula é complicada e acrescenta uma correção menor.

Além disso , a imparcialidade, neste sentido da palavra, não é sempre desejável; ver influências de um estimador.

Outro estimador algumas vezes usado é a expressão semelhante

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}\ .$$

Esta forma tem um mean squared error uniformemente menor do que o estimador imparcial, e é o maximum-likelihood estimate quando a população for normalmente distribuída.

Desvio padrão de uma variável aleatória contínua

Distribuições contínuas usualmente dão uma fórmula para se calcular o desvio padrão como uma função dos parâmetros da distribuição. Em geral, o desvio padrão de uma variável aleatória contínua *X* com função de densidade de probabilidade *p*(*x*) é

$$\sigma = \sqrt{\int (x-\mu)^2\, p(x)\, dx}$$

Onde

$$\mu = \int x\, p(x)\, dx$$

Exemplo

Mostraremos como calcular o desvio padrão de uma população. Nosso exemplo usará as idades de quatro crianças: { 5, 6, 8, 9 }.

passo1. Calcule a mean average, $\overline{x}$:

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Nós temos *N* = 4 porque existem quatro pontos de dados:

$$x_1 = 5$$
$$x_2 = 6$$
$$x_3 = 8$$
$$x_4 = 9$$

$$\overline{x} = \frac{1}{4}\sum_{i=1}^{4} x_i$$ Trocando *N* com 4

$$\overline{x} = \frac{1}{4}(x_1 + x_2 + x_3 + x_4)$$

$$\overline{x} = \frac{1}{4}\left(5+6+8+9\right)$$

$\overline{x} = 7$ Isto é a média.

Passo 2. Calcule o desvio padrão, σ. (Desde que os quatro valores representam a população toda, nós não usamos a fórmula para desvio padrão estimados neste caso):

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$\sigma = \sqrt{\frac{1}{4}\sum_{i=1}^{4}(x_i - \overline{x})^2}$$

Trocando *N* com 4

$$\sigma = \sqrt{\frac{1}{4}\sum_{i=1}^{4}(x_i - 7)^2}$$

Trocando $\overline{x}$ com 7

$$\sigma = \sqrt{\frac{1}{4}\left[(x_1-7)^2 + (x_2-7)^2 + (x_3-7)^2 + (x_4-7)^2\right]}$$

$$\sigma = \sqrt{\frac{1}{4}\left[(5-7)^2 + (6-7)^2 + (8-7)^2 + (9-7)^2\right]}$$

$$\sigma = \sqrt{\frac{1}{4}\left((-2)^2 + (-1)^2 + 1^2 + 2^2\right)}$$

$$\sigma = \sqrt{\frac{1}{4}\left(4+1+1+4\right)}$$

$$\sigma = \sqrt{\frac{10}{4}}$$

$$\sigma = \sqrt{2.5} \approx 1.58$$

$$R = (\overline{x}, \overline{x}, \overline{x})$$

cujas coordenadas são a média dos valores que iniciamos com eles. Um pouco de álgebra mostra que a distância entre *P* e *R* (que é o mesmo que a distância entre *P* e a linha *L*) é dado por σ√*3*. Uma fórmula análoga (com 3 trocado por *N*) é também válida para uma população de *N* valores; então temos de trabalhar no $\mathbf{R}^N$.

No mínimo 50% dos valores estão dentro de 1.41 desvio padrões da média.

No mínimo 75% dos valores estão dentro de 2 desvio padrões da média.

No mínimo 89% dos valores estão dentro de 3 desvio padrões da média.

No mínimo 94% dos valores estão dentro de 4 desvio padrões da média.

No mínimo 96% dos valores estão dentro de 5 desvio padrões da média.

No mínimo 97% dos valores estão dentro de 6 desvio padrões da média.

No mínimo 98% dos valores estão dentro de 7 desvio padrões da média.

No mínimo $(1 - 1/k^2) \times 100\%$ dos valores estão dentro de *k* desvio padrões da média.

$$\sigma(r) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - r)^2}$$

$$r = \overline{x}.$$

$$\sigma = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N} x_i{}^2\right) - \overline{x}^2} = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N} x_i{}^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2} = \frac{1}{N}\sqrt{N\left(\sum_{i=1}^{N} x_i{}^2\right.}$$

$$\sigma = \frac{1}{s_0}\sqrt{s_0 s_2 - s_1^2}$$

$$s_j = \sum_{k=1}^{N} x_k^j.$$

$$s = \sqrt{\frac{\sum_{i=1}^{N} {x_i}^2 - N\overline{x}^2}{N-1}}\ .$$

$$s = \sqrt{\frac{N\sum_{i=1}^{N} {x_i}^2 - \left(\sum_{i=1}^{N} x_i\right)^2}{N(N-1)}}.$$