1 Medidas de dispersão

As medidas de posição são importantes para caracterizar um conjunto de dados, mas não são suficientes para caracterizar completamente a distribuição dos dados. Para isso é necessário obter as medidas de dispersão, que medem a variabilidade dos dados.

Por exemplo: Considere as amostras referentes a altura, em cm, de dois grupos de pessoas.

Grupo A: 185 185 185

Grupo B: 187 183 185

A média para os dois grupos é a mesma $\overline{X}_A = 185$ e $\overline{X}_B = 185$.

Os 2 conjuntos não diferem entre si e consideramos somente a média, pois se basearmos somente por essa medida os dois grupos são considerados como de mesma altura. Entretanto o grupo A tem todas as observações iguais a média. Já no grupo B ocorre uma certa dispersão nos dados.

As medidas de variabilidade ou dispersão possibilitam que façamos distinção entre os conjuntos quanto à sua homogeneidade, isto é, o grau de concentração em torno de uma medida de tendência central.

1.1 Amplitude Total

Amplitude Total (A) é a diferença entre o maior e o menor valor da amostra. Essa medida é bastante simples, e obtida pela expressão:

$$A = Max - Min$$

Para dados agrupados a amplitude total é a diferença entre o ponto médio da última e da primeira classe.

Para expressar variabilidade a amplitude total não é muito usada, pois baseia-se em apenas dois dados.

1.2 Variância e Desvio Padrão

A variância é baseada pela quadrado dos desvios dos dados em relação à média. Esta medida é expressa na unidade dos dados ao quadrado.

- Para a população a variância é representada por

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

em que N é o tamanho da população

- Para a amostra a variância é representada por

$$S^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}{n-1}$$

em que n é o tamanho da população

Para dados agrupados, a variância é obtida por meio da expressão:

- Para a população a variância é representada por

$$\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2 fa_i}{\sum_{i=1}^{k} fa_i}$$

- Para a amostra a variância é representada por

$$S^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2 fa_i}{\sum_{i=1}^{k} fa_i - 1}$$

O desvio padrão é a raíz quadrada positiva da variância. Esta medida é expressa na mesma unidade dos dados.

- Para a população o desvio padrão é representada por

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

- Para a amostra o desvio padrão é representada por

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

em que n é o tamanho da população

Nota:

- O desvio padrão e a variância são medidas de dispersão ou variabilidade, a opção do uso de um ou outro, depende da finalidade da informação.
- A variância tem pouca utilidade na estatística descritiva, porém é muito importante na inferência estatística e em combinações de amostras.
- O desvio padrão é muito usado na estatística descritiva.
- É importante notar que, se os dados representarem uma amostra e não toda a população, a expressão matemática da variância deve ter $(n - 1)$ no denominador em substituição ao fator n, esta mudança é chamada de fator de correção de Bessel ou conforme os estatísticos, número de graus de liberdade. Dessa forma temos a variância da amostra.

1.2.1 Propriedades da Variância

A variância apresenta um conjunto vasto de propriedades, todas elas, sem dúvida, de grande utilidade no cálculo do seu valor.

1. A variância de uma constante k é nula;

$$\begin{aligned}
S^{2^*} &= \frac{\sum_{i=1}^{n}\left(k-\overline{X}\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}(k-k)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}(0)^2}{n-1} \\
&= 0
\end{aligned}$$

2. Somando-se ou subtraindo-se uma constante k a todos os dados a variância não se altera.

$$\begin{aligned}
X_i^* &= X_i + k \\
\overline{X}^* &= \overline{X} + k \\
S^{2^*} &= \frac{\sum_{i=1}^{n}\left(X_i^*-\overline{X}^*\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}\left(X_i+k-(\overline{X}+k)\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}\left(X_i+k-\overline{X}-k\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1} \\
&= S^2
\end{aligned}$$

3. Multiplicando-se todos os dados por uma constante k, a variância fica multiplicada por k^2.

$$\begin{aligned}
X_i^* &= kX_i \\
\overline{X}^* &= k\overline{X} \\
S^{2^*} &= \frac{\sum_{i=1}^{n}\left(X_i^* - \overline{X}^*\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}\left(kX_i - k\overline{X}\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}\left(k\left(X_i - \overline{X}\right)\right)^2}{n-1} \\
&= \frac{\sum_{i=1}^{n}k^2\left(X_i - \overline{X}\right)^2}{n-1} \\
&= k^2S^2
\end{aligned}$$

1.2.2 Propriedades do Desvio Padrão

1. Somando-se ou subtraindo-se uma constante k a todos os dados o desvio padrão não se altera.

$$\begin{aligned}
X_i^* &= X_i + k \\
S^{2^*} &= S^2 \\
S &= \sqrt{S^2}
\end{aligned}$$

2. Multiplicando-se todos os dados por uma constante k, a variância fica multiplicada por k^2.

$$\begin{aligned}
X_i^* &= kX_i \\
S^{2^*} &= k^2S^2 \\
S &= \sqrt{k^2S^2} = kS
\end{aligned}$$

1.3 Coeficiente de Variação

O coeficiente de variação (CV) é uma medida de dispersão que expressa o desvio padrão em termos da média de forma percentual

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

Se as amostras tiverem unidade diferentes ou médias diferentes o CV pode ser utilizado para comparar a variabilidade entre duas amostras.

1.4 Erro Padrão da Média

O erro padrão da média é uma medida de dispersão que dá a precisão com que a média populacional está sendo estimada. É obtido pela fórmula

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

em que:

- S é o desvio padrão da amostra;
- n é o tamanho da amostra.

2 Exemplos

Sejam dados referentes a um levantamento onde observou-se o numero de peças defeituosas em 25 maquinas de uma empresas.

Tabela 1: Número de peças defeituosas em 25 maquinas de uma empresa

1	3	4	5	6
2	3	4	5	6
2	3	4	5	6
2	3	5	5	7
2	4	5	5	8

A amplitude total

$$A = Max - Min = 8 - 1 = 7$$

Temos que a média é $\overline{X} = 4$ e como se trata de uma amostra temos:

$$S^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2}{n-1} = \frac{\left((1-4)^2 + (2-4)^2 + ... + (8-4)^2\right)}{25-1} = 3,041666667 \cong 3,04$$

O desvio padrão

$$S = \sqrt{S^2} = \sqrt{3,04} = 1,7435595 \cong 2$$

O coeficiente de variação

$$CV = 100\frac{S}{\overline{\overline{X}}} = 100\frac{2}{4} = 50\%$$

O erro padrão da médio

$$S(\overline{X}) = \frac{S}{\sqrt{n}} = \frac{2}{\sqrt{25}} = 0,4$$

2.1 Dados Agrupados

Assim, Amplitude total

$$A = Max - Min = 12,34 - 6,94 = 5,40$$

Tabela 2: Resumo da distribuição de freqüências, relativa ao tempo em segundos para carga de um aplicativo num sistema compartilhado (30 observações)

Classes			x	Frequencia Absoluta (fa)	$x - \overline{X}$	$(x - \overline{X})fa$
6,27	$\vdash$	7,62	6,94	3	7,5076	22,5228
7,62	$\vdash$	8,97	8,29	7	1,9321	13,5247
8,97	$\vdash$	10,32	9,64	10	0,0016	0,016
10,32	$\vdash$	11,67	10,99	6	1,7161	10,2966
11,67	$\vdash$	13,02	12,34	4	7,0756	28,3024
Total				30		74,6625

Temos que a média é $\overline{X} = 9,68$ e como se trata de uma amostra temos:

$$
\begin{aligned}
S^2 &= \frac{\sum_{i=1}^{n} \left(x_i - \overline{X}\right)^2 fa_i}{\sum_{i=1}^{k} fa_i - 1} \\
&= \frac{74,6625}{29} \\
&= 2,5745689 \cong 2,5746
\end{aligned}
$$

O desvio padrão

$$S = \sqrt{S^2} = \sqrt{2,5746} = 1,604556 \cong 1,60$$

O coeficiente de variação

$$CV = 100\frac{S}{\overline{X}} = 100\frac{1,60}{9,68} = 16,53\%$$

O erro padrão da média

$$S(\overline{X}) = \frac{S}{\sqrt{n}} = \frac{1,60}{\sqrt{30}} = 0,29$$