MEDIDAS DE DISPERSÃO:

A função dessas medidas é avaliar o quanto estão dispersos os valores observados numa distribuição de freqüência ou de probabilidades, ou seja, o grau de afastamento ou de concentração entre os valores.

Como principais medidas de dispersão temos:

- Amplitude Total ou Range de uma distribuição
- Desvio Médio Absoluto
- Variância
- Desvio Padrão
- Variância Relativa
- Coeficiente de Variação

A primeira medida citada, **Amplitude Total ou Range**, já vimos anteriormente e podemos dizer que não é uma boa medida, pois sendo apenas a diferença entre o maior e o menor valor observado, não dá a noção de quanto os valores intermediários estão afastados ou concentrados.

Podemos dar um exemplo elementar disto. Suponhamos duas distribuições A e B:

A → 2; 3; 4; 13; 18 => $R_A = 18 - 2 = 16$

B → 2; 4; 7; 9; 18 => $R_B = 18 - 2 = 16$

Ambas as distribuições tem a mesma Amplitude Total, mas em qual delas a dispersão é maior?

Primeiramente vamos encontrar uma medida de posição que já conhecemos: a Média.

$$\overline{X}_A = \frac{\sum X_i}{n} \Rightarrow \overline{X}_A = \frac{2+3+4+13+18}{5} \Rightarrow \overline{X}_A = \frac{40}{5} \Rightarrow \overline{X}_A = 8$$

$$\overline{X}_B = \frac{\sum X_i}{n} \Rightarrow \overline{X}_B = \frac{2+4+7+9+18}{5} \Rightarrow \overline{X}_B = \frac{40}{5} \Rightarrow \overline{X}_B = 8$$

Construindo uma tabela para cada uma das distribuições, vamos criar uma coluna para o desvio (ou distância) dos valores observados em relação à Média $\left(X_i - \overline{X}\right)$.

Distribuição A

i	X_i	$\left(X_i - \overline{X}\right)$
1	2	2 – 8 = – 6
2	3	3 – 8 = – 5
3	4	4 – 8 = – 4
4	13	13 – 8 = 5
5	18	18 – 8 = 10
Σ	40	0

Distribuição B

I	X_i	$\left(X_i - \overline{X}\right)$
1	2	2 – 8 = – 6
2	4	4 – 8 = – 4
3	7	7 – 8 = – 1
4	9	9 – 8 = 1
5	18	18 – 8 = 10
Σ	40	0

Vemos então, que em ambas as distribuições, o somatório dos desvios em relação à média aritmética é igual a zero. Isto acontecerá em toda e qualquer série de observações, pois “a soma dos desvios em relação à média aritmética sempre será nula”. Como avaliar então a dispersão entre os valores observados?

Uma das formas de resolvermos este problema é somando em módulo todos os desvios. Este somatório, dividido pelo número de observações, nos dará a segunda medida citada no início, o **Desvio Médio Absoluto (DMA)**, que nada mais é do que a média aritmética dos desvios considerados em módulos (valor absoluto).

Teremos então para as duas distribuições:

Distribuição A

i	X_i	$\left\| X_i - \overline{X} \right\|$
1	2	|2 – 8| = 6
2	3	|3 – 8| = 5
3	4	|4 – 8| = 4
4	13	|13 – 8| = 5
5	18	|18 – 8| = 10
Σ	40	30

$$DMA_A = \frac{\sum\left|X_i - \overline{X}\right|}{n} = \frac{30}{5} = 6$$

Distribuição B

i	X_i	$\left\| X_i - \overline{X} \right\|$
1	2	|2 – 8| = 6
2	4	|4 – 8| = 4
3	7	|7 – 8| = 1
4	9	|9 – 8| = 1
5	18	|18 – 8| = 10
Σ	40	22

$$DMA_B = \frac{\sum\left|X_i - \overline{X}\right|}{n} = \frac{22}{5} = 4{,}4$$

Como o Desvio Médio Absoluto da Distribuição B é menor do que o da Distribuição A, podemos afirmar que a dispersão entre os valores de B é menor.

Observação: Se tivermos uma distribuição para dados agrupados teremos de multiplicar cada módulo pela respectiva freqüência e dividir pela soma das freqüências. Assim, a fórmula para cálculo do DMA será:

$$\boxed{DMA = \frac{\sum\left|X_i - \overline{X}\right| \cdot F_i}{\sum F_i}}$$

Lembrando ainda que, se os dados se apresentarem em intervalos de classe, consideraremos X_i como o ponto médio do intervalo de classe.

A outra maneira de eliminarmos o problema do somatório nulo dos desvios em relação à Média, é elevando-os ao quadrado. O somatório destes quadrados dividido pelo número de observações, nos dará a terceira medida citada no início, a **Variância** (absoluta).

Teremos então para as duas distribuições:

Distribuição A

i	X_i	$\left(X_i - \overline{X}\right)^2$
1	2	$(-6)^2 = 36$
2	3	$(-5)^2 = 25$
3	4	$(-4)^2 = 16$
4	13	$(5)^2 = 25$
5	18	$(10)^2 = 100$
Σ	40	202

Distribuição B

i	X_i	$\left(X_i - \overline{X}\right)^2$
1	2	$(-6)^2 = 36$
2	4	$(-4)^2 = 16$
3	7	$(-1)^2 = 1$
4	9	$(1)^2 = 1$
5	18	$(10)^2 = 100$
Σ	40	154

Valores obtidos para as Variâncias (σ^2):

$$\sigma_A^2 = \frac{\sum\left(X_i - \overline{X}\right)^2}{n} = \frac{202}{5} = 40{,}4 \qquad \text{e} \qquad \sigma_B^2 = \frac{\sum\left(X_i - \overline{X}\right)^2}{n} = \frac{154}{5} = 30{,}8$$

Novamente lembramos que, se tivermos uma distribuição para dados agrupados, teremos de multiplicar o quadrado de cada desvio pela respectiva freqüência, dividindo a seguir pela soma das freqüências e se os dados se apresentarem em intervalos de classe, X_i será o ponto médio do intervalo de classe. Assim, a fórmula para cálculo da Variância será:

$$\boxed{\sigma^2 = \frac{\sum (X_i - \overline{X})^2 \cdot F_i}{\sum F_i}}$$

Para o exemplo dado, pudemos constatar também pela Variância, que a dispersão é menor entre os valores da distribuição B.

No exemplo aqui citado, a Média é um valor exato e temos apenas 5 valores observados. Em outros casos, será mais trabalhoso calcularmos a Variância através dessa forma, principalmente se a Média for uma dízima periódica. Por exemplo: se $\overline{X} = 40/9 = 4{,}444...$

Neste caso, além de trabalhoso, teremos imprecisão devido ao necessário arredondamento dos valores.

A outra fórmula para o cálculo da Variância e bem mais utilizada do que a primeira é:

$$\boxed{\sigma^2 = \frac{1}{n}\left\{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right\}}$$ para o cálculo da Variância Populacional ou:

$$\boxed{S^2 = \frac{1}{n-1}\left\{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right\}}$$ se quisermos calcular a Variância Amostral

Observe que, se estivermos trabalhando com uma Amostra, basta subtrairmos 1 (grau de liberdade) do denominador da fração 1/*n*. Esta regra vale também para os cálculos pelo processo anterior, ou seja, se usássemos o processo anterior (somatório dos quadrados dos desvios) para uma amostra, ao invés de dividirmos por *n*, dividiremos por *n –1*, então:

$$\boxed{S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}}$$ ou $$\boxed{S^2 = \frac{\sum (X_i - \overline{X})^2 \cdot F_i}{\sum F_i - 1}}$$ para dados agrupados

No exemplo dado, obteremos para as 2 distribuições os mesmos valores de Variância, tanto pela primeira como pela segunda fórmula (mais utilizada).

Vamos novamente construir as tabelas, criando além da coluna dos valores observados (X_i) uma coluna para os quadrados desses valores a fim de obtermos o somatório para usar na fórmula:

Distribuição A

i	X_i	X_i^2
1	2	$2^2 = 4$
2	3	$3^2 = 9$
3	4	$4^2 = 16$
4	13	$13^2 = 169$
5	18	$18^2 = 324$
Σ	40	522

Distribuição B

i	X_i	X_i^2
1	2	$2^2 = 4$
2	4	$4^2 = 16$
3	7	$7^2 = 49$
4	9	$9^2 = 81$
5	18	$18^2 = 324$
Σ	40	474

Valores obtidos para as Variâncias:

$$\sigma_A^2 = \frac{1}{n}\left\{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right\} = \frac{1}{5}\left(522 - \frac{40^2}{5}\right) = \frac{1}{5}\left(522 - \frac{1600}{5}\right) = \frac{1}{5}(522-320) = \frac{202}{5} = 40{,}4$$

$$\sigma_B^2 = \frac{1}{n}\left\{\sum X_i^2 - \frac{\left(\sum X_i\right)^2}{n}\right\} = \frac{1}{5}\left(474 - \frac{40^2}{5}\right) = \frac{1}{5}\left(474 - \frac{1600}{5}\right) = \frac{1}{5}(474-320) = \frac{154}{5} = 30{,}8$$

Ou seja, os mesmos valores obtidos pela fórmula anterior.

Para dados agrupados consideraremos a freqüência no cálculo da Variância e quando os dados se apresentarem em intervalos de classe, X_i será o ponto médio do intervalo de classe. Logo, a fórmula será:

$$\boxed{\sigma^2 = \frac{1}{\sum F_i}\left\{\sum X_i^2 \cdot F_i - \frac{\left(\sum X_i \cdot F_i\right)^2}{\sum F_i}\right\}}$$ para a Variância Populacional e

$$\boxed{S^2 = \frac{1}{\sum F_i - 1}\left\{\sum X_i^2 \cdot F_i - \frac{\left(\sum X_i \cdot F_i\right)^2}{\sum F_i}\right\}}$$ para a Variância Amostral

Agora, raciocinando logicamente e observando as fórmulas para cálculo da Variância, notamos tratar-se de uma soma de quadrados. Portanto, se a unidade de medida da variável for, por exemplo, *metro*, teremos como resultado para a Variância *metros quadrados*. Para voltarmos à unidade original, precisamos definir outra medida de dispersão, a mais utilizada das medidas citadas: o **Desvio Padrão**, que nada mais é do que a raiz quadrada positiva da Variância.

Assim, temos:

$\boxed{\sigma = \sqrt{\sigma^2}}$ que é o Desvio Padrão para a população e;

$\boxed{S = \sqrt{S^2}}$ que é o Desvio Padrão para uma amostra

Logo, para encontrarmos o Desvio Padrão, deveremos primeiro calcular o valor da Variância e depois extrair a raiz quadrada positiva deste resultado.

No exemplo dado, teremos:

$\sigma_A = \sqrt{\sigma_A^2} = \sqrt{40{,}4} \cong 6{,}36$ e $\sigma_B = \sqrt{\sigma_B^2} = \sqrt{30{,}8} \cong 5{,}55$

A quinta medida de dispersão que veremos é a **Variância Relativa (VR)**, que é relativa porque leva em consideração no seu cálculo o quadrado da Média. Portanto, essa variância será mais diretamente influenciada pelo valor da Média e nem sempre a distribuição que apresenta a maior (ou menor) Variância Absoluta, necessariamente apresentará a maior (ou menor) Variância Relativa, pois dependerá do valor da Média.

Entenda por quê:

$\boxed{VR = \frac{\sigma^2}{\overline{X}^2}}$ ou seja, é o resultado da divisão da Variância pelo quadrado da Média.

No exemplo que demos aqui, as Médias das distribuições A e B são iguais (8), então se a Variância de A é maior do que a de B, a VR_A também será maior.

No caso $VR_A = 40{,}4/64 = 0{,}63125$ e $VR_B = 30{,}8/64 = 0{,}48125$

Mas vamos imaginar outro exemplo:

Suponhamos agora uma distribuição Y com Média 10 e Variância igual a 16. A Variância Relativa será igual a: 16/100 = 0,16. Suponhamos agora uma distribuição Z com Média 3 e Variância igual a 4. A Variância Relativa será igual a: 4/9 = 0,444...

A segunda distribuição tem uma Variância bem menor (25% da variância de Y), mas no entanto a sua Média não é menor na mesma proporção (30% da Média de Y), estando bem mais próxima da variância do que a primeira. Logo, embora a segunda distribuição tenha uma Variância (absoluta) menor, terá uma Variância Relativa maior.

A última medida de dispersão que veremos é o **<u>Coeficiente de Variação (CV)</u>**, que nada mais é do que simplesmente o valor positivo da raiz quadrada da Variância Relativa. Logo:

$$CV = \sqrt{VR} = \sqrt{\frac{\sigma^2}{\overline{X}^2}} = \frac{\sqrt{\sigma^2}}{\sqrt{\overline{X}^2}} = \frac{\sigma}{\overline{X}}$$

Então, o CV também pode ser expresso como sendo o resultado da divisão do Desvio Padrão pela Média.

Vemos que se trata igualmente de uma medida relativa de dispersão, útil para a comparação em termos relativos do grau de concentração em torno da média de séries distintas. Portanto, é uma medida que serve para avaliar a homogeneidade de séries estatísticas e pode ser expresso na forma unitária ou percentual.

<u>Exemplo</u>: Numa empresa, o Salário Médio dos homens é de R$ 400,00 com Desvio Padrão de R$ 150,00, enquanto a Média do salário das mulheres é R$ 300,00 com Desvio Padrão de R$ 120,00. Qual dos grupos apresenta a maior dispersão relativa?

Podemos notar que, embora haja maior dispersão absoluta para o salário dos homens, a dispersão relativa será maior para o salário das mulheres, pois:

$CV_H = 150/400 = 0{,}375$ ou 37,5%

$CV_M = 120/300 = 0{,}40$ ou 40%

Considera-se que um CV superior a 50% indica alto grau de dispersão e conseqüentemente pequena representatividade da média, enquanto para um CV inferior a 50% a média será tanto mais representativa quanto menor for o valor do CV, ou seja, quanto menor for o CV mais homogênea será considerada a série e quanto maior for o CV, mais heterogênea.