Aula 04

Banco do Brasil (Escriturário - Agente de Tecnologia) Probabilidade e Estatística - 2023 (Pós-Edital)

Autor:

Equipe Exatas Estratégia Concursos

28 de Dezembro de 2022

Índice

1) Medidas de Dispersão 4
2) Amplitude Total 7
3) Amplitude Interquartílica 14
4) Desvios em Relação à Média Aritmética e Mediana 18
5) Desvio Absoluto Médio 28
6) Variância 40
7) Desvio-Padrão 51
8) Coeficiente de Variação (ou Dispersão Relativa) 63
9) Variância Relativa 70
10) Questões Comentadas - Amplitude Total - Cesgranrio 74
11) Questões Comentadas - Amplitude Interquartílica - Cesgranrio 77
12) Questões Comentadas - Variância - Cesgranrio 83
13) Questões Comentadas - Desvio-Padrão - Cesgranrio 89
14) Questões Comentadas - Coeficiente de Variação (ou Dispersão Relativa) - Cesgranrio 92
15) Aviso importante - Orientação de estudo 101
16) Questões Comentadas - Medidas de Dispersão - Inéditas 102
17) Questões Comentadas - Amplitude Total - Inéditas 104
18) Questões Comentadas - Amplitude Interquartílica - Inéditas 106
19) Questões Comentadas - Desvio Absoluto Médio - Inéditas 110
20) Questões Comentadas - Desvio-Padrão - Inéditas 114
21) Questões Comentadas - Variância Relativa - Inéditas 116
22) Lista de Questões - Amplitude Total - Cesgranrio 119
23) Lista de Questões - Amplitude Interquartílica - Cesgranrio 122
24) Lista de Questões - Variância - Cesgranrio 126
25) Lista de Questões - Desvio-Padrão - Cesgranrio 130
26) Lista de Questões - Coeficiente de Variação (ou Dispersão Relativa) - Cesgranrio 133
27) Lista de Questões - Medidas de Dispersão - Inéditas 138
28) Lista de Questões - Amplitude Total - Inéditas 140

Índice

29) Lista de Questões - Amplitude Interquartílica - Inéditas 142

30) Lista de Questões - Desvio Absoluto Médio - Inéditas 145

31) Lista de Questões - Desvio-Padrão - Inéditas 148

32) Lista de Questões - Variância Relativa - Inéditas 150

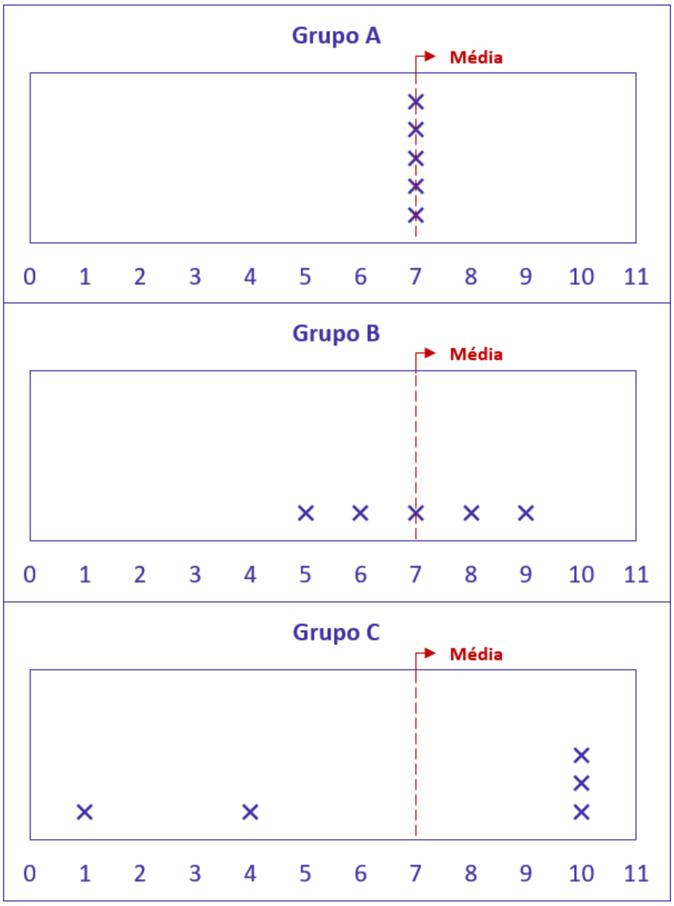
MEDIDAS DE VARIABILIDADE

Nas aulas anteriores, estudamos mecanismos para encontrar valores (média, mediana e moda) que sintetizam o comportamento dos elementos de um conjunto de dados. Esses valores fornecem parâmetros significativos para uma análise dos dados, porém, também é importante identificarmos como variam ou como se diferenciam as características dos elementos de um conjunto.

Imagine, por exemplo, que você precise avaliar três grupos de alunos, cada um com cinco elementos, no que diz respeito ao domínio de uma determinada matéria. Os testes mostraram os seguintes resultados:

Grupos
$\boldsymbol{A = 7, 7, 7, 7, 7}$
$\boldsymbol{B = 5, 6, 7, 8, 9}$
$\boldsymbol{C = 1, 4, 10, 10, 10}$

Para analisar esses dados, podemos, inicialmente, calcular a média aritmética dos três grupos. Concluímos, então, que todos possuem a mesma média aritmética ($\bar{x} = 7$). Contudo, ao observarmos a variação dos dados, percebemos que os grupos se comportam de maneira diferente, apesar de todos possuírem a mesma média.



Nesse caso, **a média**, embora seja uma medida representativa do conjunto, **não indica o grau de homogeneidade ou heterogeneidade existente entre os valores que compõem o conjunto**. Desse modo, precisamos recorrer a procedimentos matemáticos que possibilitem a compreensão da discrepância existente entre os valores do conjunto.

As medidas de dispersão (ou variabilidade) são justamente métricas que mostram a variação dos dados de um conjunto. Elas podem ser divididas em **dois grupos**:

a) medidas de dispersão absoluta:
- amplitude total;
- amplitude interquartílica;
- desvio médio;
- variância; e
- desvio-padrão.

b) medidas de dispersão relativa:
- coeficiente de variação (de Pearson); e
- variância relativa.

Nessa aula, aprenderemos a medir o grau de concentração ou dispersão dos dados em torno da média. Para isso, estudaremos as principais medidas de dispersão, que são: amplitude total, amplitude interquartílica, desvio médio, variância, desvio padrão, coeficiente de variação e variância relativa.

(COC-UFAC/UFAC/2019) Analise as seguintes assertivas:

I. A moda e o desvio padrão são medidas de dispersão,

II. O desvio médio e a média são medidas de dispersão,

III. O coeficiente de variação e a variância são medidas de dispersão,

IV. A moda, a média e o desvio padrão são medidas de posição.

Pode-se afirmar que estão corretas:

a) Apenas I. e II.

b) Apenas II. e III.

c) Apenas III.

d) Apenas IV.

e) Apenas I. e IV.

Comentários:

As medidas de posição consistem em valores que representam a tendência de concentração dos dados observados. As medidas de posição mais importantes são as medidas de tendência central. Nesse grupo, encontram-se as medidas mais utilizadas: média aritmética, moda e mediana.

Já as medidas de dispersão medem o grau de variabilidade dos elementos de uma distribuição. A dispersão aumenta à proporção que o valor da medida de dispersão também aumenta. As principais medidas de dispersão são amplitude, desvio médio, variância, desvio padrão e coeficiente de variação.

Gabarito: C.

(VUNESP/MPE-SP/2016) Na estatística, são considerados medidas de dispersão:

a) média e moda.

b) percentil e coeficiente de variação.

c) amplitude total e percentil.

d) amplitude total e desvio padrão.

e) variância e média.

Comentários:

As medidas de tendência central estudam o centro da amostra. As medidas de tendência central mais utilizadas são a média aritmética, a mediana e a moda.

Por sua vez, as medidas de separatrizes dividem os dados em grupos com a mesma quantidade de elementos, sendo representadas pelos quartis, decis e percentis.

Por fim, as medidas de dispersão têm a finalidade de identificar o quanto os dados estão dispersos em torno da média de uma amostra. São dadas pelos coeficientes de variação, desvio padrão, amplitude e variância.

Gabarito: D.

AMPLITUDE TOTAL

A **amplitude total** (ou simplesmente amplitude) é a **diferença entre os valores extremos de um conjunto de observações**, ou seja**, a diferença entre o maior e o menor elemento desse conjunto**:

$$\boxed{\boldsymbol{A_T = x_{máx} - x_{mín}}}$$

Essa medida de dispersão chama atenção por ser extremamente simples e muito fácil de se calcular. Contudo, há uma certa restrição quanto ao seu uso por conta de sua grande instabilidade, vez que leva em consideração apenas os valores extremos da série.

Por exemplo, vamos comparar os conjuntos A e B da tabela a seguir:

Conjunto	Média	Amplitude total
$\boldsymbol{A = 5, 7, 8, 9, 10, 11, 55}$	$\bar{x} = 15$	$A_T = 55 - 5 = 50$
$\boldsymbol{B = 12, 13, 14, 15, 16, 17, 18}$	$\bar{x} = 15$	$A_T = 18 - 12 = 6$

Reparem que as médias aritméticas dos dois conjuntos são iguais a 15. Portanto, no que diz respeito a essa medida de posição, podemos considerá-los idênticos. Porém, ao calcularmos a amplitude total, verificamos que os valores do conjunto A apresentam um grau de dispersão bem maior que os do conjunto B.

Isso acontece porque, **no cálculo da amplitude total, desconsideramos os valores da série que se encontram entre os extremos, o que pode conduzir a interpretações equivocadas**. Com frequência, um valor discrepante pode afetar a medida de maneira acentuada. É o caso, por exemplo, do último valor (55) do conjunto A, sensivelmente maior que seu antecessor (11), que elevou a magnitude da amplitude total para 50.

Além disso, **a amplitude total também é sensível ao tamanho de amostra**. Normalmente, a amplitude total tende a aumentar com o incremento do tamanho da amostra, ainda que não proporcionalmente. Ainda, **a amplitude total pode apresentar muita variação de uma amostra para outra, ainda que extraídas de uma mesma população**.

Apesar das limitações dessa medida, há situações em que ela pode ser aplicada de forma satisfatória. É o caso, por exemplo, da variação da temperatura em um dia. Também é o caso de quando uma compreensão rápida dos dados é mais relevante que a exatidão de um procedimento complexo.

Amplitude Total para dados não-agrupados

Para dados não agrupados, o cálculo da amplitude total pode ser expresso pela seguinte fórmula:

$$\boxed{\boldsymbol{A_T = x_{máx} - x_{mín}}}$$

em que $x_{máx}$ é o maior elemento; e $x_{mín}$ é o menor elemento do conjunto.

EXEMPLIFICANDO

Calcular a amplitude total dos conjuntos apresentados a seguir:

$$A = 50, 50, 50, 50, 50, 50, 50$$

$$B = 47, 48, 49, 50, 51, 52, 53$$

$$C = 20, 30, 40, 50, 60, 70, 80$$

Aplicando a fórmula anterior para esses dados, obtemos os seguintes resultados:

$$A_{T_A} = x_{máx} - x_{mín} = 50 - 50 = 0$$

$$A_{T_B} = x_{máx} - x_{mín} = 53 - 47 = 6$$

$$A_{T_C} = x_{máx} - x_{mín} = 80 - 20 = 60$$

Nesse caso, podemos observar que o conjunto A obteve uma amplitude total igual a 0, ou seja, uma dispersão nula. Então, significa que os valores não variam entre si. O conjunto B, por sua vez, obteve uma amplitude igual a 6. Já a variável C teve uma amplitude total igual a 60.

Embora o valor da amplitude total seja diferente para os conjuntos A, B e C, todos possuem a mesma média aritmética (50). Independentemente da média, verificamos que o conjunto A possui elementos mais homogêneos do que os conjuntos B e C. E, também, que os elementos do conjunto B são mais homogêneos do que os do conjunto C.

Amplitude Total para dados agrupados sem intervalos de classes

Para dados **agrupados SEM intervalos de classe**, a fórmula usada para a identificação da **amplitude total** é similar à adotada para dados não-agrupados. **A única diferença consiste na identificação dos valores mínimo e máximo, que agora ocorre por meio de uma tabela de frequências**.

EXEMPLIFICANDO

Calcular a amplitude total da tabela de frequências apresentada a seguir.

x_i	f_i
1	10
3	15
5	10
7	8
9	7

Nesse caso, como 1 e 9 são os valores mínimo e máximo da variável x_i, temos o seguinte resultado:

$$A_T = x_{máx} - x_{mín}$$

$$A_T = 9 - 1 = 8$$

É importante ressaltar que esses valores foram selecionados independentemente da frequência associada a eles.

Amplitude Total para dados agrupados em classes

Para dados **agrupados em intervalos de classe**, podemos definir a **amplitude total** de duas formas:

1) pela diferença entre o limite superior da última classe (L_{sup}) e o limite inferior da primeira classe (l_{inf}), conforme expresso na fórmula a seguir:

$$\boldsymbol{A = L_{sup} - l_{inf}}$$

2) pela diferença entre o ponto médio da última classe ($PM_{\text{últ}}$) e o ponto médio da primeira classe (PM_{pri}), conforme expresso na fórmula a seguir:

$$\boldsymbol{A = PM_{\text{últ}} - PM_{pri}}$$

EXEMPLIFICANDO

Calcular a amplitude total da distribuição de frequências apresentada a seguir:

Classes	PM_i	f_i
1 ⊢ 5	3	5
5 ⊢ 9	7	10
9 ⊢ 13	11	15
13 ⊢ 17	15	10
17 ⊢ 21	19	5
Total		**45**

Pelo primeiro método, temos que o limite superior da última classe é 21, enquanto o limite inferior da primeira classe é 1. Portanto, temos a seguinte amplitude:

$$A = L_{sup} - l_{inf}$$

$$A = 21 - 1 = 20$$

Pelo segundo método, temos que o ponto médio da última classe é 19, enquanto o ponto médio da primeira classe é 3. Portanto, temos a seguinte amplitude:

$$A = PM_{últ} - PM_{pri}$$

$$A = 19 - 3 = 16$$

Observe que a amplitude é menor pelo segundo método, porque os extremos da distribuição são desconsiderados.

HORA DE PRATICAR!

(COPEVE (UFAL)/Pref. Maceió/2012) Um registro em saúde epidemiológica apresenta os dados: 3, 4, 7, 8 e 8. Se calcularmos 8 – 3 = 5, estaremos determinando:

a) a amplitude total.

b) o primeiro quartil.

c) o desvio médio.

d) a distância interquartílica.

e) o terceiro quartil.

Comentários:

A amplitude total (ou simplesmente amplitude) é a diferença entre os valores extremos de um conjunto de observações, ou seja, a diferença entre o maior e o menor elemento desse conjunto.

Gabarito: A.

(VUNESP/Pref. de São José dos Campos/2012) A diferença entre o maior e o menor valor em um conjunto de dados é denominado (a)

a) curva normal.

b) amplitude total.

c) média.

d) média ponderada.

e) moda.

Comentários:

A diferença entre o maior e o menor valor em um conjunto de dados é denominada de amplitude (ou amplitude total).

Gabarito: B.

Propriedades da Amplitude Total

Nesse tópico, estudaremos as principais propriedades da amplitude total:

1ª Propriedade

- **Somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, a amplitude do conjunto não é alterada.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{3, 6, 8, 9, 10\}$, cuja amplitude total é:

$$A = x_{máx} - x_{mín}$$
$$A = 10 - 3 = 7$$

Se adicionarmos o número 5 a cada um dos termos da sequência, obteremos uma nova lista $\{y_n\} = \{x_n + 5\} = \{8, 11, 13, 14, 15\}$, cuja amplitude total é:

$$A = y_{máx} - y_{mín}$$
$$A = 15 - 8 = 7$$

Logo, a adição do número 5 a cada um dos termos da sequência fez com que a amplitude permanecesse inalterada.

2ª Propriedade

- **Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, a amplitude do conjunto fica multiplicada (ou dividida) por essa constante.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{8, 11, 13, 14, 15\}$, cuja amplitude total é:

$$A = x_{máx} - x_{mín}$$

$$A = 15 - 8 = 7$$

Se multiplicarmos cada um dos termos da sequência por 5, obteremos uma nova lista $\{y_n\} = \{x_n \times 5\} = \{40, 55, 65, 70, 75\}$, cuja amplitude total é:

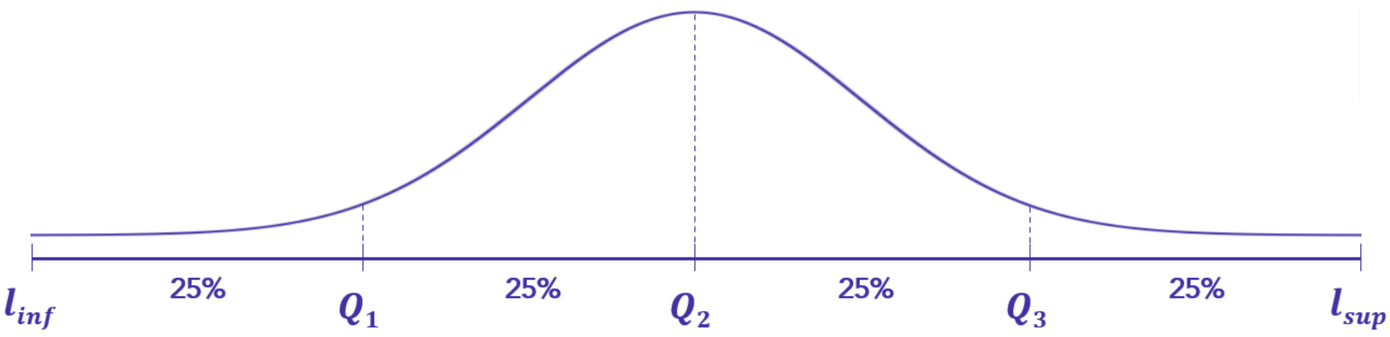
$$A = y_{máx} - y_{mín}$$

$$A = 75 - 40 = 35$$

Logo, a multiplicação de cada um dos termos da sequência por 5 fez com que a amplitude total do conjunto também fosse multiplicada por 5.

AMPLITUDE INTERQUARTÍLICA

Como já sabemos, denominamos de **quartis** os valores de uma série que a dividem em **quatro partes iguais**, isto é, **quatro partes contendo o mesmo número de elementos (25%).** A imagem a seguir mostra os quartis de uma distribuição hipotética:



Temos, então, **3 quartis** (Q_1 , Q_2 e Q_3) para dividir uma série em **quatro partes** iguais:

- Q_1: o **primeiro quartil** corresponde à separação dos primeiros 25% de elementos da série;
- Q_2: o **segundo quartil** corresponde à separação de metade dos elementos da série, **coincidindo com a mediana ($Q_2 = M_d$)**;
- Q_3: o **terceiro quartil** corresponde à separação dos primeiros 75% de elementos da série, ou dos últimos 25% de elementos da série.

A **amplitude interquartílica** (ou distância interquartílica, ou intervalo interquartílico) é o resultado da subtração entre o terceiro quartil e o primeiro quartil:

$$\boxed{A_{IQ} = Q_3 - Q_1}$$

A **amplitude semi-interquartílica** (ou desvio quartílico) é definida como a metade desse valor, sendo calculada pela expressão apresentada a seguir:

$$\boxed{D_Q = \frac{Q_3 - Q_1}{2}}$$

FIQUE ATENTO!

Reparem que a fórmula da **amplitude interquartílica** (ou distância interquartílica) é muito parecida com a fórmula da **amplitude semi-interquartílico (ou desvio quartílico)**, podendo ser facilmente confundida.

HORA DE PRATICAR!

(AOCP/SUSIPE-PA/2018) Quartis são valores que dividem os dados de uma amostra em quatro grupos, cada um deles contendo 1/4 do tamanho total da amostra. Em relação ao assunto, informe se é verdadeiro (V) ou falso (F) o que se afirma a seguir e assinale a alternativa com a sequência correta.

( ) O primeiro quartil Q1 tem 1/4 dos dados acima dele e 3/4 dos dados abaixo dele.

( ) O terceiro quartil Q3 tem 3/4 dos dados abaixo dele e 1/4 dos dados acima dele.

( ) O quartil Q3 é a própria mediana.

( ) A distância interquartílica é dada por DIQ = Q3 – Q1.

a) V – F – F – V.

b) F – V – F – V.

c) F – V – V – V.

d) V – V – F – V.

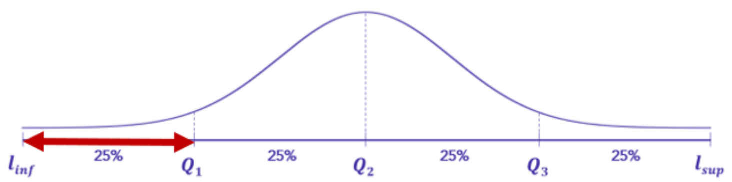
e) F – V – F – F.

Comentários:

Vamos analisar cada assertiva:

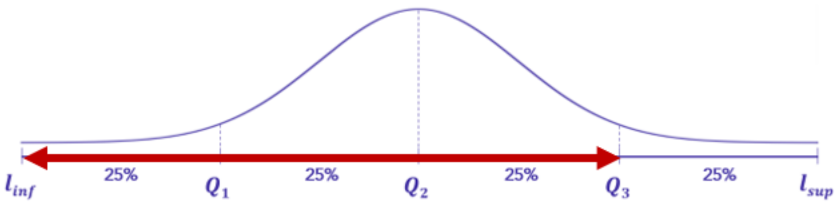
Item 1 - O primeiro quartil Q1 tem 1/4 dos dados acima dele e 3/4 dos dados abaixo dele.

Falso. O primeiro quartil (Q1) tem 1/4 (25%) dos dados **abaixo** dele e 3/4 (75%) **acima** dele.



Item 2 - O terceiro quartil Q3 tem 3/4 dos dados abaixo dele e 1/4 dos dados acima dele.

Verdadeiro. De fato, o terceiro quartil (Q3) tem 3/4 (75%) dos dados abaixo dele e 1/4 (25%) acima dele



Item 3 - O quartil Q3 é a própria mediana.

Falso. A terceira assertiva é falsa, pois a mediana é equivalente ao segundo quartil (Q2).

4 - A distância interquartílica é dada por $DIQ = Q_3 - Q_1$.

Verdadeiro. Essa é a exata definição de distância interquartílica.

Gabarito: B.

Propriedades da Amplitude Interquartílica

A seguir, veremos que a amplitude interquartílica e o desvio quartílico possuem as mesmas propriedades da amplitude total.

1ª Propriedade

- **Somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, a amplitude interquartílica (e o desvio quartílico) do conjunto não é alterada.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9, 11, 13\}$, cuja amplitude interquartílica é:

$$A_{IQ} = Q_3 - Q_1$$

$$A_{IQ} = 11 - 3 = 8$$

Se adicionarmos o número 5 a cada um dos termos da sequência, iremos obter uma nova lista $\{y_n\} = \{x_n + 5\} = \{6, 8, 10, 12, 14, 16, 18\}$, cuja amplitude interquartílica é:

$$A_{IQ} = Q_3 - Q_1$$

$$A_{IQ} = 16 - 8 = 8$$

Logo, a adição do número 5 a cada um dos termos da sequência fez com que a amplitude interquartílica permanecesse inalterada.

2ª Propriedade

- **Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, a amplitude interquartílica (e o desvio quartílico) do conjunto fica multiplicada (ou dividida) por essa constante.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9, 11, 13\}$, cuja amplitude interquartílica é:

$$A_{IQ} = Q_3 - Q_1$$

$$A_{IQ} = 11 - 3 = 8$$

Se multiplicarmos cada um dos termos da sequência por 5, iremos obter uma nova lista $\{y_n\} = \{x_n \times 5\} = \{5, 15, 25, 35, 45, 55, 65\}$, cuja amplitude interquartílica é:

$$A_{IQ} = Q_3 - Q_1$$

$$A_{IQ} = 55 - 15 = 40$$

Logo, a multiplicação de cada um dos termos da sequência por 5 fez com que a amplitude interquartílica do conjunto também fosse multiplicada por 5.

Desvios em Relação à Média Aritmética e Mediana

Antes de apresentarmos as fórmulas para o cálculo do desvio médio e da variância, precisamos compreender qual o conceito de desvio em estatística. **Um desvio é a distância entre qualquer observação do conjunto de dados e uma medida descritiva desse conjunto**:

$$desvio = observação - medida$$

Em especial, destacamos os desvios em relação à média aritmética e em relação à mediana:

$$\boxed{\boldsymbol{d_i = x - \overline{x}} \quad (média)}$$

ou

$$\boxed{\boldsymbol{d_i = x - M_d} \quad (mediana)}$$

É natural pensarmos que, **quando os desvios em relação a uma medida descritiva são pequenos, as observações estão concentradas em torno dessa medida e, portanto, a variabilidade dos dados é pequena**. Agora, **quando os desvios são maiores, significa que as observações estão dispersas e, portanto, a variabilidade dos dados é grande**.

(VUNESP/TJ-SP/2015) Leia o texto a seguir para responder à questão. Uma pequena empresa que emprega apenas cinco funcionários paga os seguintes salários mensais (em mil reais):

0,9	1,2	1,4	1,5	2,0

Considerando-se a média dos salários, o valor do desvio do salário de quem ganha R$ 1.400,00 mensais é

a) –1.000.

b) –400.

c) 0.

d) 200.

e) 400.

Comentários:

Para responder a questão, primeiro teremos que calcular a média:

$$\bar{x} = \frac{0,9 + 1,2 + 1,4 + 1,5 + 2}{5}$$

$$\bar{x} = \frac{7}{5} = 1,4\,mil$$

Então, o desvio em relação ao salário de R$ 1.400 é:

$$d_i = x_i - \bar{x}$$

$$1400 - 1400 = 0$$

Gabarito: C.

(VUNESP/TJ-SP/2015) Leia o texto a seguir para responder à questão. Uma pequena empresa que emprega apenas cinco funcionários paga os seguintes salários mensais (em mil reais):

0,9	1,2	1,4	1,5	2,0

Somando-se os valores absolutos dos desvios individuais dos salários tomados em relação à média, encontra-se o valor de

a) 1.400,00.

b) 1.200,00.

c) 1.000,00.

d) 800,00.

e) 0.

Comentários:

Como vimos na questão anterior, a média dos salários é:

$$\bar{x} = \frac{0,9 + 1,2 + 1,4 + 1,5 + 2}{5}$$

$$\bar{x} = \frac{7}{5} \Longrightarrow \bar{x} = 1,4\,mil$$

Agora, calcularemos os desvios para cada valor apresentado:

Valor	Desvio	Desvio absoluto
0,9	$0{,}9 - 1{,}4 = -0{,}50$	0,5
1,2	$1{,}2 - 1{,}4 = -0{,}2$	0,2
1,4	$1{,}4 - 1{,}4 = 0$	0
1,5	$1{,}5 - 1{,}4 = 0{,}1$	0,1
2	$2 - 1{,}4 = 0{,}6$	0,6
Total		**1,4**

Portanto, a soma dos desvios absolutos é 1,4 mil.

Gabarito: A.

Propriedades dos Desvios em Relação à Média Aritmética e Mediana

Nesse tópico, revisaremos algumas propriedades importantes dos desvios sobre as quais discutimos quando estudamos sobre a média e a mediana.

1ª Propriedade

- **A soma algébrica dos desvios em relação à média é nula.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 2, 3, 4, 5, 6, 7\}$, com média $\bar{x} = 4$. O desvio em relação à média é a diferença entre cada elemento da sequência e a média aritmética. Como a sequência possui 7 elementos, teremos o mesmo número de desvios para calcular. Logo, basta encontrarmos a diferença entre cada elemento e a média:

$$d_1 = x_1 - \bar{x} = 1 - 4 = -3$$

$$d_2 = x_2 - \bar{x} = 2 - 4 = -2$$

$$d_3 = x_3 - \bar{x} = 3 - 4 = -1$$

$$d_4 = x_4 - \bar{x} = 4 - 4 = 0$$

$$d_5 = x_5 - \bar{x} = 5 - 4 = 1$$

$$d_6 = x_6 - \bar{x} = 6 - 4 = 2$$

$$d_7 = x_7 - \bar{x} = 7 - 4 = 3$$

Agora, somaremos todos esses desvios:

$$\sum_{i=1}^{7} d_i = d_1 + d_2 + d_3 + d_4 + d_5 + d_6 + d_7$$

$$\sum_{i=1}^{7} d_i = (-3) + (-2) + (-1) + 0 + 1 + 2 + 3$$

$$\sum_{i=1}^{7} d_i = 0$$

Portanto, não importa qual a sequência de números, a soma dos desvios em relação à média é sempre igual a zero.

2ª Propriedade

- **A soma dos quadrados dos desvios da sequência de números $\{x_i\}$, em relação a um número a, é mínima se a for a média aritmética dos números.**

EXEMPLIFICANDO

Novamente, vamos tomar como exemplo a sequência $\{x_n\} = \{1, 2, 3, 4, 5, 6, 7\}$, com média $\bar{x} = 4$. Já calculamos os desvios desses números em relação à média:

$$d_1 = x_1 - \bar{x} = 1 - 4 = -3$$

$$d_2 = x_2 - \bar{x} = 2 - 4 = -2$$

$$d_3 = x_3 - \bar{x} = 3 - 4 = -1$$

$$d_4 = x_4 - \bar{x} = 4 - 4 = 0$$

$$d_5 = x_5 - \bar{x} = 5 - 4 = 1$$

$$d_6 = x_6 - \bar{x} = 6 - 4 = 2$$

$$d_7 = x_7 - \bar{x} = 7 - 4 = 3$$

Na propriedade anterior, vimos que a soma dos desvios é sempre igual a zero. Agora, calcularemos a soma dos quadrados desses desvios. Em outras palavras, vamos elevar cada um deles ao quadrado e somar todos os resultados:

$$\sum_{i=1}^{7} d_i = d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 + d_7^2$$

$$\sum_{i=1}^{7} d_i = (-3)^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2 + 3^2$$

$$\sum_{i=1}^{7} d_i = 9 + 4 + 1 + 0 + 1 + 4 + 9$$

$$\sum_{i=1}^{7} d_i = 28$$

A propriedade nos garante que, para essa sequência numérica, o valor 28 é o menor valor possível. Isto é, se encontrarmos os desvios em relação a outro número (diferente da média) e, em seguida, calcularmos a soma dos quadrados dos desvios, o valor obtido será maior que 28. Vamos ver o que acontece ao calcularmos o desvio em relação ao número 6:

$$d_1 = x_1 - \bar{x} = 1 - 6 = -5$$

$$d_2 = x_2 - \bar{x} = 2 - 6 = -4$$

$$d_3 = x_3 - \bar{x} = 3 - 6 = -3$$

$$d_4 = x_4 - \bar{x} = 4 - 6 = -2$$

$$d_5 = x_5 - \bar{x} = 5 - 6 = -1$$

$$d_6 = x_6 - \bar{x} = 6 - 6 = 0$$

$$d_7 = x_7 - \bar{x} = 7 - 6 = 1$$

Agora, calcularemos a soma dos quadrados desses números:

$$\sum_{i=1}^{7} d_i = (-5)^2 + (-4)^2 + (-3)^2 + (-2)^2 + (-1)^2 + 0^2 + 1^2$$

$$\sum_{i=1}^{7} d_i = 25 + 16 + 9 + 4 + 1 + 0 + 1 = 56$$

Como esperávamos, o resultado foi maior do que 28.

3ª Propriedade

- **A soma dos desvios absolutos de uma sequência de números, em relação a um número a, é mínima quando a é a mediana dos números.**

EXEMPLIFICANDO

Vamos tomar como exemplo a série $\{x_n\} = \{1, 2, 3, 4, 6, 7, 8, 9\}$. Como o número de termos é par, a mediana será, por convenção, a média aritmética dos dois termos centrais:

$$M_d = \frac{4+6}{2} = 5.$$

O desvio em relação à mediana corresponde à diferença entre cada elemento da sequência e a mediana. Como são 8 números, temos a mesma quantidade de desvios para calcular. Logo, basta encontrarmos a diferença entre cada número e a mediana:

$$d_1 = x_1 - M_d = 1 - 5 = -4$$

$$d_2 = x_2 - M_d = 2 - 5 = -3$$

$$d_3 = x_3 - M_d = 3 - 5 = -2$$

$$d_4 = x_4 - M_d = 4 - 5 = -1$$

$$d_5 = x_5 - M_d = 6 - 5 = 1$$

$$d_6 = x_6 - M_d = 7 - 5 = 2$$

$$d_7 = x_7 - M_d = 8 - 5 = 3$$

$$d_8 = x_8 - M_d = 9 - 5 = 4$$

Agora, precisamos somar os valores absolutos desses desvios:

$$\sum_{i=1}^{8} |d_i| = |d_1| + |d_2| + |d_3| + |d_4| + |d_5| + |d_6| + |d_7| + |d_8|$$

$$\sum_{i=1}^{8} |d_i| = |-4| + |-3| + |-2| + |-1| + |1| + |2| + |3| + |4|$$

$$\sum_{i=1}^{8} |d_i| = 4 + 3 + 2 + 1 + 1 + 2 + 3 + 4 = 20$$

A propriedade garante que, ao calcularmos a soma dos desvios absolutos em relação à mediana, o menor valor que encontraremos para essa sequência será 20.

Há um detalhe importante que precisamos esclarecer. Como vimos anteriormente, **quando o número de elementos do conjunto é ímpar, o valor da mediana é único e igual ao termo central**. Porém, **quando o número de elementos é par, a mediana pode ser qualquer valor entre os termos centrais, havendo infinitos valores possíveis para a mediana**. **Por convenção, contudo, adotamos a média aritmética dos valores centrais**.

Certo, o que isso tem a ver com a propriedade que estamos estudando? Significa dizer que, se calcularmos a soma dos desvios absolutos para qualquer valor entre 4 e 6, que são os termos centrais, o valor dos desvios absolutos em relação a mediana também será mínimo. A título exemplificativo, vamos calcular os desvios em relação ao valor 4,5:

$$d_1 = x_1 - 4{,}5 = 1 - 4{,}5 = -3{,}5$$
$$d_2 = x_2 - 4{,}5 = 2 - 4{,}5 = -2{,}5$$
$$d_3 = x_3 - 4{,}5 = 3 - 4{,}5 = -1{,}5$$
$$d_4 = x_4 - 4{,}5 = 4 - 4{,}5 = -0{,}5$$
$$d_5 = x_5 - 4{,}5 = 6 - 4{,}5 = 1{,}5$$
$$d_6 = x_6 - 4{,}5 = 7 - 4{,}5 = 2{,}5$$
$$d_7 = x_7 - 4{,}5 = 8 - 4{,}5 = 3{,}5$$
$$d_8 = x_8 - 4{,}5 = 9 - 4{,}5 = 4{,}5$$

Somando os valores absolutos desses desvios:

$$\sum_{i=1}^{8} |d_i| = |d_1| + |d_2| + |d_3| + |d_4| + |d_5| + |d_6| + |d_7| + |d_8|$$

$$\sum_{i=1}^{8}|d_i| = |-3,5| + |-2,5| + |-1,5| + |-0,5| + |1,5| + |2,5| + |3,5| + |4,5|$$

$$\sum_{i=1}^{8}|d_i| = 3,5 + 2,5 + 1,5 + 0,5 + 1,5 + 2,5 + 3,5 + 4,5 = 20$$

Como havíamos previsto, o valor também foi igual ao valor mínimo, 20.

Por último, a propriedade também garante que, para qualquer valor fora do intervalo entre 4 e 6, encontraremos um valor maior que o mínimo. Por exemplo, vamos calcular os desvios em relação ao valor 7:

$$d_1 = x_1 - 7 = 1 - 7 = -6$$
$$d_2 = x_2 - 7 = 2 - 7 = -5$$
$$d_3 = x_3 - 7 = 3 - 7 = -4$$
$$d_4 = x_4 - 7 = 4 - 7 = -3$$
$$d_5 = x_5 - 7 = 6 - 7 = -1$$
$$d_6 = x_6 - 7 = 7 - 7 = 0$$
$$d_7 = x_7 - 7 = 8 - 7 = 1$$
$$d_8 = x_8 - 7 = 9 - 7 = 2$$

Somando os valores absolutos desses desvios:

$$\sum_{i=1}^{8}|d_i| = |d_1| + |d_2| + |d_3| + |d_4| + |d_5| + |d_6| + |d_7| + |d_8|$$

$$\sum_{i=1}^{8}|d_i| = |-6| + |-5| + |-4| + |-3| + |-1| + |0| + |1| + |2|$$

$$\sum_{i=1}^{8}|d_i| = 6 + 5 + 4 + 3 + 1 + 0 + 1 + 2 = 22$$

Portanto, como havíamos previsto anteriormente, o valor foi maior que o mínimo.

RESUMINDO

Podemos resumir as propriedades dos desvios da seguinte forma:

1ª) a soma dos desvios em relação à média aritmética é sempre nula;

2ª) a soma dos quadrados dos desvios em relação à média aritmética é mínima; e

3ª) a soma dos módulos dos desvios em relação à mediana é mínima.

INDO MAIS FUNDO!

Caso o número de elemento seja par, a soma dos módulos também será mínima se os desvios forem calculados em relação a um dos valores centrais. Isto é, também será mínima a soma dos módulos dos desvios calculados em relação a qualquer termo no intervalo $\left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$, em que $x_{\frac{n}{2}}$ e $x_{\frac{n}{2}+1}$ são os termos centrais.

HORA DE PRATICAR!

(FCC/TRE-SP/2012) Dado um conjunto de observações, indicadas por X_i $(i = 1, 2, 3, \dots, n)$, o desvio $e1$ da i-ésima observação em relação a um valor α é $e_i = X_i - \alpha$ e $|e_i|$ é o valor absoluto de ei. Considere as seguintes afirmações para qualquer conjunto de observações:

I. O valor de $\sum e_i^{\ 2}$ é mínimo se α for igual à média aritmética das observações.

II. O valor de $\sum |e_i|$ é mínimo se α for igual à mediana das observações.

III. O valor de $\sum e_i$ é nulo se α for igual à moda das observações.

IV. O valor de $\sum |e_i|$ é nulo se α for igual à média aritmética das observações.

Então, são corretas APENAS

a) I e II.

b) I e III.

c) II e III.

d) II e IV.

e) II, III e IV.

▪

Comentários:

Vamos analisar cada assertiva:

A sentença I é verdadeira, pois a soma dos quadrados dos desvios é mínima quando os desvios são calculados em relação à média aritmética.

A sentença II também é verdadeira, pois a soma dos módulos dos desvios é mínima quando os desvios são calculados em relação à mediana. Em qualquer situação, quando o desvio é calculado em relação à mediana, a soma dos desvios absolutas é mínima.

A sentença III é falsa, vez que a soma dos módulos dos desvios é nula se os desvios são calculados em relação à média. Somente seria verdadeira caso a moda fosse igual à média.

A sentença IV é falsa, pois a soma dos desvios absolutos em relação à média somente é nula quando todos os desvios também são nulos, ou seja, se todos os números fossem iguais e não houvesse dispersão dos dados.

Gabarito: A.

Desvio Absoluto Médio

O **desvio absoluto médio**, ou simplesmente desvio médio, **mede a dispersão entre os valores da distribuição e a média dos dados coletados**. Para compreender essa medida, vamos supor que o Estratégia Concursos tenha realizado uma semana de revisão para estudantes da área fiscal, obtendo os seguintes números de visualizações:

Dia da semana	Número de visualizações
Domingo	2.000
Segunda	4.000
Terça	5.200
Quarta	6.300
Quinta	5.400
Sexta	4.100
Sábado	2.400
Total	$\sum f_i = 29.400$

Isso significa que a semana de revisão teve uma média diária de 4.200 visualizações. Esse resultado, porém, não retrata a realidade com fidedignidade, pois alguns dias tiveram mais visualizações do que a média; enquanto outros não. Por isso, é importante sabermos o quão distante a média está em relação aos valores reais por ela representados.

Para calculá-los, basta subtrairmos o valor da média de cada observação, conforme mostrado a seguir:

Dia da semana	Número de visualizações	$x_i - \bar{x}$
Domingo	2.000	$2.000 - 4.200 = -2.200$
Segunda	4.000	$4.000 - 4.200 = -200$
Terça	5.200	$5.200 - 4.200 = 1000$
Quarta	6.300	$6.300 - 4.200 = 2.100$
Quinta	5.400	$5.400 - 4.200 = 1.200$
Sexta	4.100	$4.100 - 4.200 = -100$
Sábado	2.400	$2.400 - 4.200 = -1.800$

Total	$\sum f_i = 29.400$	**0**

Notem que, ao calcularmos o desvio médio, obtemos resultados positivos e negativos, que se anulam ao serem somados. Percebam que existem valores de observações que estão muito próximos da média, enquanto outros estão mais distantes.

Como a soma de todos os desvios médios é sempre igual a zero para qualquer conjunto de dados (1.ª propriedade dos desvios), sabemos que $\sum_{i=1}^{n}(x-\bar{x})$ não nos fornecerá nenhuma informação relevante nem nos ajudará a compreender o que está acontecendo com essa variável.

Para superar essa dificuldade, podemos utilizar apenas os resultados positivos dos desvios calculados. A fórmula do cálculo do desvio médio se apresenta da seguinte maneira:

$$\boldsymbol{D_m = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}}$$

em que D_m representa o desvio médio, $|x_i - \bar{x}|$ representa o módulo da diferença entre uma determinada observação e a média calculada, f_i representa a frequência de um determinado valor para a variável da distribuição, e n representa o total de elementos formados pela distribuição.

O desvio médio é uma medida de dispersão mais robusta do que a amplitude total e a amplitude interquartílica, pois leva em consideração todos os valores do conjunto. O inconveniente dessa medida é a operação de módulo, que, por conta de suas características matemáticas, torna difícil o estudo de suas propriedades.

Desvio Médio para dados não-agrupados

O desvio absoluto médio (D_m), de um conjunto de n observações $x_1, \cdots, x_n$, é a média dos valores absolutos das diferenças entre as observações e a média. Isto é,

$$\boldsymbol{D_m = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}}$$

As barras verticais indicam a operação de módulo, que é responsável por transformar qualquer número negativo em um número positivo, isto é, retornar o valor absoluto.

EXEMPLIFICANDO

Calcular o desvio médio do conjunto mostrado a seguir:

$$\{1, 2, 3, 5, 9\}$$

Iniciaremos pelo cálculo da média aritmética:

$$\bar{x} = \frac{1+2+3+5+9}{5} = \frac{20}{5} = 4$$

Vamos montar uma tabela para facilitar o cálculo do desvio médio:

x_i	$(x_i - \bar{x})$	$\lvert x_i - \bar{x}\rvert$
1	$(1-4) = -3$	3
2	$(2-4) = -2$	2
3	$(3-4) = -1$	1
5	$(5-4) = 1$	1
9	$(9-4) = 5$	5
		$\sum \lvert x_i - \bar{x}\rvert = 12$

Aplicando a fórmula do desvio médio, temos:

$$D_m = \frac{\sum_{i=1}^{K} |x_i - \bar{x}|}{n} = \frac{\sum_{i=1}^{K} |x_i - 4|}{n} = \frac{12}{5} = 2{,}4$$

HORA DE PRATICAR!

(CESPE/ANATEL/2004)

meses	fev	mar	abr	mai	jun	jul	ago	set	out	nov
N	100	70	70	60	50	100	50	50	30	20

A tabela acima mostra os números mensais de reclamações (N) feitas por usuários de telefonia fixa, registradas em uma central de atendimento, entre os meses de fevereiro a novembro de 2003. Considerando esses dados, julgue os itens que se seguem.

O maior desvio absoluto dos números mensais de reclamações registradas é superior a 45.

Comentários:

Iniciaremos pelo cálculo da média:

$$\bar{x} = \frac{100 + 70 + 70 + 60 + 50 + 100 + 50 + 50 + 30 + 20}{10} = \frac{600}{10} = 60$$

Agora, calcularemos o módulo (valor absoluto) de cada um dos desvios.

$$|d_1| = |x_1 - \bar{x}| = |100 - 60| = 40$$
$$|d_2| = |x_2 - \bar{x}| = |70 - 60| = 10$$
$$|d_3| = |x_3 - \bar{x}| = |70 - 60| = 10$$
$$|d_4| = |x_4 - \bar{x}| = |60 - 60| = 0$$
$$|d_5| = |x_5 - \bar{x}| = |50 - 60| = 10$$
$$|d_6| = |x_6 - \bar{x}| = |100 - 60| = 40$$
$$|d_7| = |x_7 - \bar{x}| = |50 - 60| = 10$$
$$|d_8| = |x_8 - \bar{x}| = |50 - 60| = 10$$
$$|d_9| = |x_9 - \bar{x}| = |30 - 60| = 30$$
$$|d_{10}| = |x_{10} - \bar{x}| = |20 - 60| = 40$$

O maior desvio absoluto é 40, portanto, o item está incorreto.

Gabarito: Errado.

(CESPE/ANATEL/2004)

meses	fev	mar	abr	mai	jun	jul	ago	set	out	nov
N	100	70	70	60	50	100	50	50	30	20

A tabela acima mostra os números mensais de reclamações (N) feitas por usuários de telefonia fixa, registradas em uma central de atendimento, entre os meses de fevereiro a novembro de 2003. Considerando esses dados, julgue os itens que se seguem.

O desvio médio absoluto da sequência formada pelos números mensais de reclamações é um valor entre 25 e 35.

Comentários:

Para calcular o desvio absoluto médio, temos que encontrar a média dos valores absolutos (módulos) dos desvios.

$$D_m = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

$$D_m = \frac{40 + 10 + 10 + 0 + 10 + 40 + 10 + 10 + 30 + 40}{10} = \frac{200}{10}$$

$$D_m = 20$$

Gabarito: Errado.

Desvio Médio para dados agrupados sem intervalo de classe

Quando os valores vierem dispostos em uma tabela de frequências, o desvio médio será calculado por meio da seguinte fórmula:

$$\boldsymbol{D_m = \frac{\sum_{i=1}^{m} [\, |x_i - \bar{x}| \times f_i \,]}{\sum f_i}}$$

Em que m indica o número de grupos em que os dados estão organizados; e $|x_i - \bar{x}|$ representa o módulo da diferença entre uma determinada observação e a média calculada.

EXEMPLIFICANDO

Durante uma pesquisa, o Estratégia Concursos registrou a quantidade de filhos de seus professores, obtendo a tabela de frequências apresentada a seguir. Vamos calcular o desvio médio dessa distribuição.

Nº de filhos por professor	f_i	$x_i \times f_i$
0	4	$0 \times 4 = 0$
1	8	$1 \times 8 = 8$
2	4	$2 \times 4 = 8$
3	2	$3 \times 2 = 6$
4	2	$4 \times 2 = 8$
*** Pesquisa populacional**	$\sum f_i = 20$	$\sum x_i \times f_i = 30$

Iniciaremos pelo cálculo da média aritmética:

$$\bar{x} = \frac{\sum x_i \times f_i}{\sum f_i} = \frac{30}{20} = 1{,}50 \ filhos\ /\ professor$$

Em seguida, adicionaremos uma nova coluna à tabela anterior, em que calcularemos os produtos dos desvios absolutos por suas respectivas frequências:

Nº de filhos por professor	f_i	$x_i \times f_i$	$\lvert x_i - \bar{x} \rvert \times f_i$
0	4	0	$\lvert 0 - 1{,}5 \rvert \times 4 = 6$
1	8	8	$\lvert 1 - 1{,}5 \rvert \times 8 = 4$
2	4	8	$\lvert 2 - 1{,}5 \rvert \times 4 = 2$
3	2	6	$\lvert 3 - 1{,}5 \rvert \times 2 = 3$
4	2	8	$\lvert 4 - 1{,}5 \rvert \times 2 = 5$
*** Pesquisa populacional**	$\sum f_i = 20$	$\sum x_i \times f_i = 30$	$\sum \lvert x_i - \bar{x} \rvert \times f_i = 20$

Por fim, aplicando a fórmula do desvio médio, temos:

$$D_m = \frac{\sum |x_i - \overline{x}| \times f_i}{\sum f_i} = \frac{20}{20} = 1$$

Desvio Médio para dados agrupados em classes

Se os dados estiverem agrupados em classe, deveremos adotar a mesma convenção que tomamos para o cálculo da média: vamos assumir que todos os valores coincidem com os pontos médios das suas respectivas classes.

EXEMPLIFICANDO

Durante uma pesquisa, o Estratégia Concursos registrou as estaturas de 40 alunos, obtendo a distribuição de frequências apresentada a seguir. Calcule o desvio médio dessa distribuição.

Estaturas	Frequência (f_i)
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
*** Pesquisa amostral**	$\sum f_i = 40$

Inicialmente, construiremos uma tabela como a mostrada a seguir:

Estaturas	Frequência (f_i)	x_i	$x_i \times f_i$	$(x_i - \bar{x})$	$\|x_i - \bar{x}\|$	$\|x_i - \bar{x}\| \times f_i$
150 ⊢ 154	4	152	608	-9	9	36
154 ⊢ 158	9	156	1.404	-5	5	45
158 ⊢ 162	11	160	1.760	-1	1	11
162 ⊢ 166	8	164	1.312	3	3	24
166 ⊢ 170	5	168	840	7	7	35
170 ⊢ 174	3	172	516	11	11	33
*** Pesquisa amostral**	$\sum f_i = 40$		$\sum x_i \times f_i = 6.440$			$\sum \|x_i - \bar{x}\| \times f_i = 184$

Feito isso, podemos calcular a média da distribuição por meio da seguinte fórmula:

$$\bar{x} = \frac{\sum x_i \times f_i}{\sum f_i} = \frac{6.440}{40} = 161$$

Conhecendo a média, completamos a tabela com as diferenças e os produtos necessários para o cálculo do desvio médio. Assim, aplicando a fórmula do desvio médio, temos:

$$D_m = \frac{\sum_{i=1}^{6} |x_i - \bar{x}| \times f_i}{\sum f_i} = \frac{184}{40} = 4{,}6\ cm$$

Portanto, o desvio médio para essa distribuição de estaturas é $4{,}6\ cm$.

HORA DE PRATICAR!

(UEPA/SEFAZ-PA/2013) A tabela abaixo representa as estaturas dos jogadores de voleibol que disputaram a Liga Mundial de 2012.

ESTATURAS (cm)	NÚMERO DE JOADORES
$180 \vdash 190$	10
$190 \vdash 200$	30
$200 \vdash 210$	10
$\sum$	

O desvio médio da estatura dos jogadores é:

a) 2

b) 4

c) 6

d) 8

e) 10

Comentários:

Vamos iniciar pelo cálculo da média. Para isso, construiremos uma coluna com os pontos médios e multiplicaremos cada um pela sua respectiva frequência. Da seguinte forma:

Estaturas	f_i	x_i	$x_i \times f_i$
$180 \vdash 190$	10	185	$10 \times 185 = 1.850$
$190 \vdash 200$	30	195	$30 \times 195 = 5.850$
$200 \vdash 210$	10	205	$10 \times 205 = 2.050$
Total	$\sum f_i = 50$		$\sum x_i \times f_i = 9.750$

Portanto, a média é:

$$\bar{x} = \frac{9.750}{50} = 195$$

Em seguida, adicionaremos uma coluna para calcularmos os módulos dos desvios:

Estaturas	f_i	x_i	$x_i \times f_i$	$\|x_i - \bar{x}\|$
180 ⊢ 190	10	185	$10 \times 185 = 1.850$	$\|185 - 195\| = 10$
190 ⊢ 200	30	195	$30 \times 195 = 5.850$	$\|195 - 195\| = 0$
200 ⊢ 210	10	205	$10 \times 205 = 2.050$	$\|205 - 195\| = 10$
Total	$\sum f_i = 50$		$\sum x_i \times f_i = 9.750$	

Para calcular o desvio médio, devemos multiplicar cada desvio absoluto pela sua respectiva frequência. Depois, basta somar tudo e dividir por n.

Estaturas	f_i	x_i	$x_i \times f_i$	$\|x_i - \bar{x}\|$	$\|x_i - \bar{x}\| \times f_i$
180 ⊢ 190	10	185	$10 \times 185 = 1.850$	$\|185 - 195\| = 10$	$10 \times 10 = 100$
190 ⊢ 200	30	195	$30 \times 195 = 5.850$	$\|195 - 195\| = 0$	$0 \times 30 = 0$
200 ⊢ 210	10	205	$10 \times 205 = 2.050$	$\|205 - 195\| = 10$	$10 \times 10 = 100$
Total	$\sum f_i = 50$		$\sum x_i \times f_i = 9.750$		$\sum \|x_i - \bar{x}\| \times f_i = 200$

▪

Portanto, o desvio médio é:

$$D_m = \frac{\sum |x_i - \bar{x}| \times f_i}{\sum f_i} = \frac{200}{50} = 4$$

Gabarito: B.

Propriedades do Desvio Médio

Nesse tópico, vamos aprender as principais propriedades do desvio médio.

1ª Propriedade

- **Somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, o desvio médio do conjunto não é alterado.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9\}$, cuja desvio médio é:

$$D_m = \frac{|1-5| + |3-5| + |5-5| + |7-5| + |9-5|}{5}$$

$$D_m = \frac{4+2+0+2+4}{5} = \frac{12}{5}$$

Se adicionarmos o número 5 a cada um dos termos da sequência, iremos obter uma nova lista $\{y_n\} = \{x_n + 5\} = \{6, 8, 10, 12, 14\}$, cujo desvio médio é:

$$D_m = \frac{|6-10| + |8-10| + |10-10| + |12-10| + |14-10|}{5}$$

$$D_m = \frac{4+2+0+2+4}{5} = \frac{12}{5}$$

Logo, a adição do número 5 a cada um dos termos da sequência fez com que o desvio médio permanecesse inalterado.

2ª Propriedade

- **Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, o desvio médio do conjunto fica multiplicado (ou dividido) por essa constante.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9\}$, cujo desvio médio é:

$$D_m = \frac{|1-5| + |3-5| + |5-5| + |7-5| + |9-5|}{5}$$

$$D_m = \frac{4+2+0+2+4}{5} = \frac{12}{5}$$

Se multiplicarmos cada um dos termos da sequência por 5, iremos obter uma nova lista $\{y_n\} = \{x_n \times 5\} = \{5, 15, 25, 35, 45\}$, cujo desvio médio é:

$$D_m = \frac{|5-25| + |15-25| + |25-25| + |35-25| + |45-25|}{5}$$

$$D_m = \frac{20+10+0+10+20}{5} = \frac{60}{5}$$

Logo, a multiplicação de cada um dos termos da sequência por 5 fez com que o desvio médio do conjunto também fosse multiplicado por 5.

VARIÂNCIA (σ^2)

Existem outras formas de se eliminar o problema com os números negativos. Além da operação de módulo, podemos trabalhar com potências pares. A utilização de potências de expoente par, como o número dois, além de transformar números negativos em positivos, simplifica o cálculo.

A variância é determinada pela média dos quadrados dos desvios em relação à média aritmética. **Por meio dessa medida de dispersão ou variabilidade, podemos avaliar o quanto os dados estão dispersos em relação à média aritmética**. Nesse sentido, **quanto maior a variância, maior a dispersão dos dados**.

A **variância** leva em consideração a totalidade dos valores da variável em estudo, e não apenas os valores extremos, como faz a amplitude total. Por isso, essa medida de variabilidade **é considerada muito estável**. Além disso, a variância complementa as informações obtidas pelas medidas de tendência central.

Até o momento, as medidas que estudamos não sofriam nenhuma alteração quando o cálculo era realizado para uma amostra. Contudo, para a variância, devemos levar em consideração essa informação, pois há uma **pequena diferença** entre o cálculo da **variância populacional** e da **variância amostral**.

A **variância populacional** é simbolizada pela letra grega σ (sigma), sendo **calculada usando todos os elementos da população**, pela seguinte fórmula:

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

em que: x_i é o valor de ordem i assumido pela variável; μ é a média populacional de x; σ^2 é a variância populacional; e n é o número de dados da população.

A **variância amostral** é simbolizada pela letra s, sendo **calculada a partir de uma amostra da população**, pela seguinte fórmula:

$$\boldsymbol{s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

em que: x_i é o valor de ordem i assumido pela variável; $\bar{x}$ é a média amostral de x; s^2 é a variância amostral; e n é o número de dados da amostra.

Normalmente, uma população possui uma grande quantidade de elementos, o que inviabiliza a realização de um estudo aprofundado de suas medidas, chamadas de **parâmetros populacionais**. Nesse caso, recorremos ao estudo de amostras representativas dessa população, buscando obter indícios do valor correto do parâmetro populacional desconhecido. Esse valor amostral é denominado de **estimador** do parâmetro populacional.

Em nosso caso, a variância populacional cumpre o papel de **parâmetro populacional**, enquanto a variância amostral atua como um **estimador**. Já vimos a variância populacional e a variância amostral são representadas por símbolos diferentes: σ^2 e s^2. O mesmo acontece com a média populacional e a média amostral, que também possuem símbolos diferentes: μ (parâmetro populacional) e $\bar{x}$ (estimador).

Reparem que, quando a variância representa uma descrição da amostra e não da população, caso mais frequente em estatística, o denominador das expressões deve ser $n - 1$, em vez de n. Isso ocorre porque a utilização do divisor $(n - 1)$ resulta em uma melhor estimativa do parâmetro populacional.

Além disso, como a soma dos desvios em relação à média aritmética é sempre nula, apenas $(n - 1)$ dos desvios $(x_i - \bar{x})$ são independentes, vez que $(n - 1)$ desvios determinam automaticamente o valor desconhecido. Para amostras grandes ($n > 30$), não há diferença significativa entre os resultados proporcionados pela utilização de qualquer dos dois divisores, n ou $(n - 1)$.

Em determinadas situações, a aplicação dessas fórmulas pode requerer um esforço considerável. É o caso do que acontece quando a média não é um número natural, situação em que a obtenção da soma dos quadrados dos desvios se torna muito trabalhosa. Por isso, é importante aprendermos outras fórmulas que podem nos ajudar no cálculo da variância.

Já ouviram dizer que **a variância é igual à média dos quadrados menos o quadrado da média**? Pois bem, essa é a fórmula que expressa a **variância populacional**:

$$\boldsymbol{\sigma^2 = \overline{x^2} - (\bar{x})^2}$$

em que $\overline{x^2}$ é a média dos quadrados; e $(\bar{x})^2$ é o quadrado da média.

Como vimos, para encontrarmos a fórmula da variância amostral, basta substituirmos n por $(n - 1)$. Isso é equivalente a multiplicarmos a variância populacional por $\left(\frac{n}{n-1}\right)$. É exatamente o que faremos agora:

$$\boldsymbol{s^2 = \left[\overline{x^2} - (\bar{x})^2\right] \times \left(\frac{n}{n-1}\right)}$$

em que $\overline{x^2}$ é a média dos quadrados; $(\bar{x})^2$ é o quadrado da média; e n é o tamanho da amostra.

Por fim, é importante ressaltarmos que, por ser calculada a partir dos quadrados dos desvios, a **variância é um número em unidade quadrada em relação à variável em questão**, o que pode ser considerado um inconveniente. Por isso, essa medida tem pouca utilidade na estatística descritiva, mas é extremamente importante na inferência estatística e em combinações de amostras. Por exemplo, se os dados estiverem expressos em quilogramas (Kg), a variância estará expressa em quilogramas ao quadrado (Kg²).

TOME NOTA!

Símbolo da variância populacional:

$$\sigma^2$$

Símbolo da variância amostral:

$$s^2$$

FIQUE ATENTO!

A variância de um conjunto é zero quando todos os elementos são iguais. Se todos os elementos são iguais, a média aritmética do conjunto coincide com o valor dos elementos e todos os desvios também são iguais a zero. Logo, a variância também é zero.

A variância é sempre maior ou igual a zero, isto é, sempre tem valor positivo.

RESUMINDO

Fórmula da variância populacional:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} \quad \text{ou} \quad \sigma^2 = \overline{x^2} - (\bar{x})^2$$

Fórmula da variância amostral:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \quad \text{ou} \quad s^2 = \left[\overline{x^2} - (\bar{x})^2\right] \times \left(\frac{n}{n-1}\right)$$

Variância para dados não-agrupados

Para dados não agrupados, a variância pode ser expressa por meio das seguintes fórmulas:

a) para populações

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n} \quad ou \quad \sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n}}$$

b) para amostras

$$\boldsymbol{s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1} \quad ou \quad s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}{n - 1}}$$

A relação entre a variância amostral (s^2) e a variância populacional (σ^2) é dada por:

$$\boldsymbol{s^2 = \left(\frac{n}{n - 1}\right) \times \sigma^2}$$

EXEMPLIFICANDO

Calcular a **variância amostral** do conjunto de números mostrado a seguir:

$$\{1, 2, 3, 5, 9\}$$

Iniciaremos pelo cálculo da média aritmética:

$$\bar{x} = \frac{1 + 2 + 3 + 5 + 9}{5} = \frac{20}{5} = 4$$

Agora, vamos montar uma tabela para facilitar o cálculo da variância:

x_i	$(x_i - \bar{x})^2$
1	$(1-4)^2 = 9$
2	$(2-4)^2 = 4$
3	$(3-4)^2 = 1$
5	$(5-4)^2 = 1$
9	$(9-4)^2 = 25$
	$\sum(x_i - \bar{x})^2 = 40$

Por fim, aplicando a fórmula da **variância amostral**, temos:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1} = \frac{40}{5-1} = 10$$

(VUNESP/TJ-SP/2015) Dados os valores de uma variável: 5, 10, 15, 20, 25, as variâncias amostral e populacional são, respectivamente,

a) 14,7 e 15.

b) 125 e 250.

c) 62,5 e 50.

d) 29,4 e 30,8.

e) 83,3 e 85.

Comentários:

Vamos começar calculando a média:

$$\frac{5+10+15+20+25}{5} = 15$$

Agora, vamos encontrar os desvios em relação à média:

$$d_1 = 5 - 15 = -10$$

$$d_2 = 10 - 15 = -5$$

$$d_3 = 15 - 15 = 0$$
$$d_4 = 20 - 15 = 5$$
$$d_5 = 25 - 15 = 10$$

Para calcular a variância (populacional ou amostral), precisamos calcular a soma dos quadrados dos desvios, isto é:

$$\sum d_i^2 = (-10)^2 + (-5)^2 + 0^2 + 5^2 + 10^2$$
$$\sum d_i^2 = 250$$

Nesse momento, dividiremos esse valor por n para encontrar a variância populacional e por $n - 1$ para encontrar a variância amostral:

$$s^2 = \frac{\sum d_i^2}{n-1} = \frac{250}{5-1} = \frac{250}{4} = 62{,}5 \text{ (variância amostral)}$$

$$\sigma^2 = \frac{\sum d_i^2}{n} = \frac{250}{5} = 50 \text{ (variância populacional)}$$

Gabarito: C.

Variância para dados agrupados sem intervalos de classes

Quando os valores vierem dispostos em uma tabela de frequências, a variância será calculada por meio de uma das seguintes fórmulas:

a) para populações

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{m}(X_i - \mu)^2 \times f_i}{n} \quad ou \quad \sigma^2 = \frac{\sum_{i=1}^{m}\left(X_i^2 \times f_i\right) - \frac{\left(\sum_{i=1}^{m} X_i \times f_i\right)^2}{n}}{n}}$$

b) para amostras

$$\boldsymbol{s^2 = \frac{\sum_{i=1}^{m}(X_i - \bar{x})^2 \times f_i}{n - 1} \quad ou \quad s^2 = \frac{\sum_{i=1}^{m}\left(X_i^2 \times f_i\right) - \frac{\left(\sum_{i=1}^{m} X_i \times f_i\right)^2}{n}}{n - 1}}$$

Em que $n = \sum_{i=1}^{m} f_i$ e $\bar{x} = \frac{\sum_{i=1}^{m} X_i \times f_i}{n}$.

EXEMPLIFICANDO

Durante uma pesquisa, o Estratégia Concursos registrou a quantidade de filhos por professor, obtendo a tabela de frequências apresentada a seguir. Sendo assim, calcule a variância amostral dessa tabela.

Nº de filhos por professor	f_i	$x_i \times f_i$
0	4	$0 \times 4 = 0$
1	8	$1 \times 8 = 8$
2	4	$2 \times 4 = 8$
3	2	$3 \times 2 = 6$
4	2	$4 \times 2 = 8$
*** Pesquisa populacional**	$\sum f_i = 20$	$\sum x_i \times f_i = 30$

Iniciaremos pelo cálculo da média aritmética:

$$\bar{x} = \frac{\sum x_i \times f_i}{\sum f_i} = \frac{30}{20} = 1{,}50\ filhos\ /\ professor$$

Em seguida, adicionaremos uma nova coluna à tabela anterior, em que calcularemos os produtos dos quadrados dos desvios por suas respectivas frequências:

Nº de filhos por professor	f_i	$x_i \times f_i$	$(x_i - \bar{x})^2 \times f_i$
0	4	0	$(0 - 1{,}5)^2 \times 4 = 9$
1	8	8	$(1 - 1{,}5)^2 \times 8 = 2$
2	4	8	$(2 - 1{,}5)^2 \times 4 = 1$
3	2	6	$(3 - 1{,}5)^2 \times 2 = 4{,}5$
4	2	8	$(4 - 1{,}5)^2 \times 2 = 12{,}5$
*** Pesquisa populacional**	$\sum f_i = 20$	$\sum x_i \times f_i = 30$	$\sum (x_i - x)^2 \times f_i = 29$

Por fim, aplicando a fórmula do desvio padrão amostral, temos:

$$s^2 = \frac{\sum(x_i - \bar{x})^2 \times f_i}{n-1} = \frac{29}{19} = 1{,}52$$

Variância para dados agrupados em classes

Para dados contínuos agrupados em classes, a variância é calculada por meio das seguintes expressões:

a) para populações

$$\boldsymbol{\sigma^2 = \frac{\sum_{i=1}^{k}(PM_i - \mu)^2 \times f_i}{n} \quad ou \quad \sigma^2 = \frac{\sum_{i=1}^{k} PM_i^2 \times f_i - \frac{\left(\sum_{i=1}^{k} PM_i \times f_i\right)^2}{n}}{n}}$$

b) para amostras

$$\boldsymbol{s^2 = \frac{\sum_{i=1}^{k}(PM_i - \bar{x})^2 \times f_i}{n - 1} \quad ou \quad s^2 = \frac{\sum_{i=1}^{k} PM_i^2 \times f_i - \frac{\left(\sum_{i=1}^{k} PM_i \times f_i\right)^2}{n}}{n - 1}}$$

Em que $n = \sum_{i=1}^{k} f_i$ e $\bar{x} = \frac{\sum_{i=1}^{k} PM_i \times f_i}{n}$.

Observem que as fórmulas são praticamente iguais as apresentadas no subtópico anterior. A diferença básica é que agora vamos utilizar o ponto médio das k classes.

EXEMPLIFICANDO

Durante uma pesquisa, o Estratégia Concursos registrou as estaturas de 40 alunos, obtendo a distribuição de frequências apresentada a seguir. Vamos calcular a variância amostral dessa distribuição.

Estaturas	Frequência (f_i)
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
*** Pesquisa amostral**	$\sum f_i = 40$

Inicialmente, construiremos uma tabela como a mostrada a seguir:

Estaturas	Frequência (f_i)	x_i	$x_i \times f_i$	$(x_i - \overline{x})$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})^2 \times f_i$
150 ⊢ 154	4	152	608	-9	81	324
154 ⊢ 158	9	156	1.404	-5	25	225
158 ⊢ 162	11	160	1.760	-1	1	11
162 ⊢ 166	8	164	1.312	3	9	72
166 ⊢ 170	5	168	840	7	49	245
170 ⊢ 174	3	172	516	11	121	363
*** Pesquisa populacional**	$\sum f_i = 40$		$\sum x_i \times f_i = 6.440$			$\sum (x_i - \overline{x})^2 \times f_i = 1.240$

Feito isso, podemos calcular a média da distribuição por meio da seguinte fórmula:

$$\bar{x} = \frac{\sum x_i \times f_i}{\sum f_i} = \frac{6.440}{40} = 161$$

Conhecendo a média, completamos a tabela com as diferenças e os produtos necessários para o cálculo da variância. Agora, aplicando a fórmula da variância amostral, temos:

$$s^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \times f_i}{n - 1} = \frac{\sum_{i=1}^{6} (x_i - 161)^2 \times f_i}{40 - 1} = \frac{1.240}{39} = 31{,}79\ cm^2$$

A variância amostral das estaturas é $31,79\ cm^2$.

Propriedades do Variância

Nesse tópico, vamos aprender as principais propriedades da variância.

1ª Propriedade

- **Somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, a variância do conjunto não é alterada.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9\}$, cuja variância é:

$$\sigma^2 = \frac{(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2}{5}$$

$$\sigma^2 = \frac{16+4+0+4+16}{5} = 8$$

Se adicionarmos o número 5 a cada um dos termos da sequência, iremos obter uma nova lista $\{y_n\} = \{x_n + 5\} = \{6, 8, 10, 12, 14\}$, cuja variância é:

$$\sigma^2 = \frac{(6-10)^2 + (8-10)^2 + (10-10)^2 + (12-10)^2 + (14-10)^2}{5}$$

$$\sigma^2 = \frac{16+4+0+4+16}{5} = 8$$

Logo, a adição do número 5 a cada um dos termos da sequência fez com que a variância permanecesse inalterada.

2ª Propriedade

- **Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, a variância do conjunto fica multiplicada (ou dividida) pelo QUADRADO dessa constante.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9\}$, cuja variância é:

$$\sigma^2 = \frac{(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2}{5}$$

$$\sigma^2 = \frac{16 + 4 + 0 + 4 + 16}{5} = 8$$

Se multiplicarmos cada um dos termos da sequência por 5, iremos obter uma nova lista $\{y_n\} = \{x_n \times 5\} = \{5, 15, 25, 35, 45\}$, cuja variância é:

$$\sigma^2 = \frac{(5-25)^2 + (15-25)^2 + (25-25)^2 + (35-25)^2 + (45-25)^2}{5}$$

$$\sigma^2 = \frac{400 + 100 + 0 + 100 + 400}{5} = 200$$

Logo, a multiplicação de cada um dos termos da sequência por 5 fez com que a variância do conjunto fosse multiplicada por $5^2 = 25$.

DESVIO-PADRÃO (σ)

O desvio padrão (s ou σ) é definido como sendo a raiz quadrada da média aritmética dos quadrados dos desvios e, dessa forma, **é determinado pela raiz quadrada da variância**. **É uma das medidas de variabilidade mais utilizadas porque é capaz de apontar de forma mais precisa a dispersão dos valores em relação à média aritmética.**

Valores muito próximos da média resultarão em um desvio-padrão pequeno, enquanto valores mais espalhados levarão a desvios maiores. Essa medida será sempre maior ou igual a zero. Ela será igual a zero quando todos os elementos do conjunto forem iguais.

O desvio padrão é utilizado para comparar a variabilidade de dois conjuntos de dados diferentes quando as médias forem aproximadamente iguais e quando as unidades de medidas para os dois conjuntos forem idênticas.

A fórmula para o cálculo do desvio padrão populacional é:

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

Para o desvio padrão amostral, a fórmula é a seguinte:

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

Como vimos no tópico anterior, a utilização do divisor $(n-1)$ resulta em uma melhor estimativa do parâmetro populacional. Além disso, como a soma dos desvios em relação à média aritmética é sempre nula, apenas $(n-1)$ dos desvios $(x_i - \bar{x})$ são independentes, uma vez que esses $(n-1)$ desvios determinam automaticamente o valor desconhecido.

Por fim, **o desvio-padrão é expresso nas mesmas unidades dos dados originais**. Tanto o desvio padrão como a variância são usados como medidas de dispersão ou variabilidade. O uso de uma medida ou de outra dependerá da finalidade que se tiver em mente.

TOME NOTA!

Símbolo do desvio-padrão populacional:

$$\sigma$$

Símbolo do desvio-padrão amostral:

$$s$$

FIQUE ATENTO!

O desvio-padrão será igual a zero quando todos os elementos forem iguais. Se todos os elementos forem iguais, a média aritmética do conjunto será igual ao valor dos elementos e todos os desvios também serão iguais a zero. Logo, o desvio-padrão também será zero.

O desvio-padrão é sempre maior ou igual a zero, isto é, sempre tem valor positivo.

RESUMINDO

Fórmula do desvio-padrão populacional:

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

Fórmula do desvio-padrão amostral:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

HORA DE PRATICAR!

(VUNESP/ARTESP/2018) Numa série composta por n dados, todos de mesmo valor x ($x \neq 0$), o valor do desvio padrão s é:

a) $s = \frac{n}{x}$

b) $s = 0$

c) $s = \frac{nx}{2}$

d) $s = x$

e) $s = 1$

Comentários:

Como todos os dados são iguais, todos os desvios são nulos. Consequentemente, os quadrados dos desvios também são nulos. Logo, a variância e o desvio-padrão serão iguais a zero.

Gabarito: B.

(UFMT/Pref. de Cáceres-MT/2017) Um conjunto de dados sobre a plaquetopenia de pacientes com dengue tem variância igual a zero. Pode-se concluir que também vale zero

a) a média.

b) o desvio padrão.

c) a mediana.

d) a moda.

Comentários:

O desvio-padrão é a raiz quadrada da variância. Nesse caso, como a variância é igual a zero, então o desvio-padrão vale:

$$\sigma = \sqrt{0} = 0.$$

Gabarito: B.

Desvio-padrão para dados não-agrupados

Para dados não agrupados, o desvio-padrão pode ser expresso por meio das seguintes fórmulas:

a) para populações

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}}$$

b) para amostras

$$\boldsymbol{s = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}}$$

EXEMPLIFICANDO

Vamos calcular o desvio-padrão amostral do conjunto de números mostrado a seguir:

$$\{1, 2, 3, 5, 9\}$$

Iniciaremos pelo cálculo da média aritmética:

$$\bar{x} = \frac{1 + 2 + 3 + 5 + 9}{5} = \frac{20}{5} = 4$$

Em seguida, montaremos uma tabela para facilitar o cálculo do desvio padrão:

x_i	$(x_i - \overline{x})^2$
1	$(1 - 4)^2 = 9$
2	$(2 - 4)^2 = 4$
3	$(3 - 4)^2 = 1$
5	$(5 - 4)^2 = 1$
9	$(9 - 4)^2 = 25$
	$\sum (x_i - \bar{x})^2 = 40$

Por fim, aplicando a fórmula do desvio padrão temos:

$$s = \sqrt{\frac{40}{5-1}} = \sqrt{10} \cong 3{,}16$$

HORA DE PRATICAR!

(FCC/ARTESP/2017) O departamento de operações de uma autarquia do Estado fez um levantamento do número de acidentes em um determinado trecho de rodovia no ano de 2016, conforme tabela a seguir.

Mês	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Nº de Acidentes	36	28	12	5	3	2	2	4	9	11	22	38

Os números indicam que há uma dispersão significativa, portanto, o desvio padrão para esta amostra é representado por

a) 13,30.

b) 14,33.

c) 12,74.

d) 10,40.

e) 11,50.

Comentários:

Vamos iniciar calculando a média:

$$\bar{x} = \frac{36+28+12+5+3+2+2+4+9+11+22+38}{12}$$

$$\bar{x} = \frac{172}{12}$$

$$\bar{x} = \frac{43}{3} = 14{,}33$$

Agora, vamos montar uma tabela para simplificar o cálculo da média dos quadrados:

Valor (x)	X^2
36	1.296
28	784
12	144
5	25
3	9
2	4
2	4
4	16
9	81
11	121
22	484
38	1.444
Total	**4.412**

Portanto, a média dos quadrados é:

$$\overline{x^2} = \frac{4.412}{12} = 367{,}67$$

A variância populacional é dada pela diferença entre a média dos quadrados e o quadrado da média:

$$\sigma^2 = \overline{x^2} - \bar{x}^2$$

$$\sigma^2 = 367{,}67 - (14{,}33)^2$$

$$\sigma^2 = 367{,}67 - 205{,}35$$

$$\sigma^2 = 162{,}32$$

Se multiplicarmos a variância populacional por $\frac{n}{n-1}$, encontraremos a variância amostral:

$$s^2 = 162{,}32 \times \frac{12}{11}$$

$$s^2 = 162{,}32 \times 1{,}09$$

$$s^2 = 177{,}07$$

$$s = \sqrt{177{,}07}$$

$$s = 13{,}30$$

Gabarito: A.

(CESPE/Polícia Federal/2018)

	dia				
	1	2	3	4	5
X (quantidade diária de drogas apreendidas, em kg)	10	22	18	22	28

Tendo em vista que, diariamente, a Polícia Federal apreende uma quantidade X, em kg, de drogas em determinado aeroporto do Brasil, e considerando os dados hipotéticos da tabela precedente, que apresenta os valores observados da variável X em uma amostra aleatória de 5 dias de apreensões no citado aeroporto, julgue o próximo item.

O desvio padrão amostral da variável *X* foi inferior a 7

Comentários:

Começaremos calculando a média:

$$\frac{10+22+18+22+28}{5}=20$$

Agora, vamos encontrar os desvios:

$$d_1=10-20=-10$$
$$d_2=22-20=2$$
$$d_3=18-20=-2$$
$$d_4=22-20=2$$
$$d_5=28-20=8$$

Para calcular a variância (populacional ou amostral), precisamos calcular a soma dos quadrados dos desvios, isto é:

$$\sum d_i^2=(-10)^2+2^2+(-2)^2+2^2+8^2$$
$$\sum d_i^2=176$$

Nesse momento, dividiremos esse valor por $n-1$ para encontrarmos a variância amostral:

$$s^2=\frac{\sum d_i^2}{n-1}=\frac{176}{5-1}=\frac{176}{4}=44$$

E, por fim, o desvio padrão é a raiz quadrada da variância:

$$s=\sqrt{44}$$

O enunciado diz que esse valor é menor do que 7kg. De fato, sabemos que $7^2=49$, logo $\sqrt{44}<7$.

Gabarito: Certo.

Desvio-padrão para dados agrupados sem intervalo de Classe

Quando os valores vierem dispostos em uma tabela de frequências, o desvio-padrão será calculado por meio de uma das seguintes fórmulas:

a) para populações

$$\boldsymbol{\sigma = \sqrt{\frac{\sum_{i=1}^{m}\left(d_i^2 \times f_i\right)}{n}} = \sqrt{\frac{\sum_{i=1}^{m}\left[(X_i - \mu)^2 \times f_i\right]}{n}}}$$

b) para amostras

$$\boldsymbol{s = \sqrt{\frac{\sum_{i=1}^{m}\left(d_i^2 \times f_i\right)}{n-1}} = \sqrt{\frac{\sum_{i=1}^{m}\left[(X_i - \overline{x})^2 \times f_i\right]}{n-1}}}$$

Em que $n = \sum_{i=1}^{m} f_i$ e $\bar{x} = \frac{\sum_{i=1}^{m} X_i \times f_i}{n}$.

EXEMPLIFICANDO

Durante uma pesquisa, o Estratégia Concursos registrou a quantidade de filhos de seus professores, obtendo a tabela de frequências apresentada a seguir. Vamos calcular o desvio-padrão amostral dessa distribuição.

Nº de filhos por professor	f_i	$x_i \times f_i$
0	4	$0 \times 4 = 0$
1	8	$1 \times 8 = 8$
2	4	$2 \times 4 = 8$
3	2	$3 \times 2 = 6$
4	2	$4 \times 2 = 8$
*** Pesquisa populacional**	$\sum f_i = 20$	$\sum x_i \times f_i = 30$

Iniciaremos pelo cálculo da média aritmética:

$$\bar{x} = \frac{\sum X_i \times f_i}{\sum f_i} = \frac{30}{20} = 1{,}50\ filhos\ /\ professor$$

Em seguida, adicionaremos uma nova coluna à tabela anterior, em que calcularemos os produtos dos quadrados dos desvios por suas respectivas frequências:

Nº de filhos por professor	f_i	$x_i \times f_i$	$(x_i - \overline{x})^2 \times f_i$
0	4	0	$(0 - 1{,}5)^2 \times 4 = 9$
1	8	8	$(1 - 1{,}5)^2 \times 8 = 2$
2	4	8	$(2 - 1{,}5)^2 \times 4 = 1$
3	2	6	$(3 - 1{,}5)^2 \times 2 = 4{,}5$
4	2	8	$(4 - 1{,}5)^2 \times 2 = 12{,}5$
*** Pesquisa populacional**	$\sum f_i = 20$	$\sum x_i \times f_i = 30$	$\sum (x_i - x)^2 \times f_i = 29$

Por fim, aplicando a fórmula do desvio padrão amostral, temos:

$$s = \sqrt{\frac{\sum (x_i - \overline{x})^2 \times f_i}{n - 1}}$$

$$s = \sqrt{\frac{29}{19}} = \sqrt{1{,}52} \cong 1{,}23$$

Desvio-padrão para dados agrupados em classes

Quando tivermos que calcular o desvio-padrão para dados agrupados em classes, usaremos as mesmas fórmulas para dados sem intervalos de classes, utilizando para x_i os pontos médios de cada classe, mas adotando os mesmos procedimentos.

EXEMPLIFICANDO

Durante uma pesquisa, o Estratégia Concursos registrou as estaturas de 40 alunos, obtendo a distribuição de frequências apresentada a seguir. Vamos calcular o desvio-padrão amostral dessa distribuição.

Estaturas	Frequência (f_i)
150 ⊢ 154	4
154 ⊢ 158	9
158 ⊢ 162	11
162 ⊢ 166	8
166 ⊢ 170	5
170 ⊢ 174	3
*** Pesquisa amostral**	$\sum f_i = 40$

Inicialmente, construiremos uma tabela como a mostrada a seguir:

Estaturas	Frequência (f_i)	x_i	$x_i \times f_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \times f_i$
150 ⊢ 154	4	152	608	-9	81	324
154 ⊢ 158	9	156	1.404	-5	25	225
158 ⊢ 162	11	160	1.760	-1	1	11
162 ⊢ 166	8	164	1.312	3	9	72
166 ⊢ 170	5	168	840	7	49	245
170 ⊢ 174	3	172	516	11	121	363
*** Pesquisa populacional**	$\sum f_i = 40$		$\sum x_i \times f_i = 6.440$			$\sum (x_i - \bar{x})^2 \times f_i = 1.240$

Feito isso, podemos calcular a média da distribuição por meio da seguinte fórmula:

$$\bar{x} = \frac{\sum PM_i \times f_i}{\sum f_i} = \frac{6.440}{40} = 161$$

Conhecendo a média, completamos a tabela com as diferenças e os produtos necessários para o cálculo do desvio padrão. Agora, aplicando a fórmula do desvio padrão amostral, temos:

$$s = \sqrt{\frac{\sum_{i=1}^{k} (PM_i - \bar{x})^2 \times f_i}{n-1}} = \sqrt{\frac{\sum_{i=1}^{6} (PM_i - 161)^2 \times f_i}{40-1}} = \sqrt{\frac{1.240}{39}} = \sqrt{31{,}79} \cong 5{,}64\ cm$$

O desvio-padrão das estaturas é $5{,}64\ cm$. Vimos anteriormente que o desvio médio, para essa mesma distribuição, foi de $4{,}63\ cm$.

Propriedades do Desvio-padrão

Nesse tópico, vamos estudar as principais propriedades do desvio-padrão.

1ª Propriedade

- **Somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, o desvio-padrão do conjunto não é alterado.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9\}$, cujo desvio-padrão é:

$$\sigma = \sqrt{\frac{(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2}{5}}$$

$$\sigma = \sqrt{\frac{16 + 4 + 0 + 4 + 16}{5}} = 2\sqrt{2}$$

Se adicionarmos o número 5 a cada um dos termos da sequência, iremos obter uma nova lista $\{y_n\} = \{x_n + 5\} = \{6, 8, 10, 12, 14\}$, cujo desvio-padrão é:

$$\sigma = \sqrt{\frac{(6-10)^2 + (8-10)^2 + (10-10)^2 + (12-10)^2 + (14-10)^2}{5}}$$

$$\sigma = \sqrt{\frac{16 + 4 + 0 + 4 + 16}{5}} = 2\sqrt{2}$$

Logo, a adição do número 5 a cada um dos termos da sequência fez com que o desvio-padrão permanecesse inalterado.

2ª Propriedade

- **Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, o desvio-padrão do conjunto fica multiplicado (ou dividido) por essa constante.**

EXEMPLIFICANDO

Vamos tomar como exemplo a sequência $\{x_n\} = \{1, 3, 5, 7, 9\}$, cujo desvio-padrão é:

$$\sigma = \sqrt{\frac{(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2}{5}}$$

$$\sigma = \sqrt{\frac{16 + 4 + 0 + 4 + 16}{5}} = 2\sqrt{2}$$

Se multiplicarmos cada um dos termos da sequência por 5, iremos obter uma nova lista $\{y_n\} = \{x_n \times 5\} = \{5, 15, 25, 35, 45\}$, cujo desvio-padrão é:

$$\sigma = \sqrt{\frac{(5-25)^2 + (15-25)^2 + (25-25)^2 + (35-25)^2 + (45-25)^2}{5}}$$

$$\sigma = \sqrt{\frac{400 + 100 + 0 + 100 + 400}{5}} = \sqrt{\frac{200}{5}} = 10\sqrt{2}$$

Logo, a multiplicação de cada um dos termos da sequência por 5 fez com que o desvio-padrão do conjunto também fosse multiplicado por 5.

COEFICIENTE DE VARIAÇÃO (OU DISPERSÃO RELATIVA)

O desvio-padrão pode ser utilizado para a comparação de duas ou mais séries de valores, no que diz respeito à variabilidade e dispersão, quando os conjuntos possuem a mesma média e estão expressos na mesma unidade de medida (p.ex., os dois conjuntos em centímetros). Porém, quando os conjuntos de dados estão expressos em unidades diferentes (p.ex., quilogramas e centímetros), precisamos de outra medida.

Para contornar essa limitação do desvio-padrão, podemos caracterizar a dispersão ou variabilidade dos dados de maneira relativa ao seu valor médio. Nesse sentido, **o coeficiente de variação é uma medida de dispersão relativa que fornece a variação dos dados em relação à média**, podendo ser calculado como:

a) para populações

$$\boxed{CV = \frac{\sigma}{\mu} \times 100\ (\%)}$$

b) para amostras

$$\boxed{CV = \frac{s}{\bar{x}} \times 100\ (\%)}$$

em que: σ é o desvio-padrão populacional; μ é a média populacional; s é o desvio-padrão amostral; e $\bar{x}$ é a média amostral.

O **coeficiente de variação** pode ser interpretado por meio de algumas regras empíricas:

a) a distribuição tem **baixa dispersão** se $CV < 15\%$;

b) a distribuição tem **média dispersão** se $15\% < CV < 30\%;\ e$

c) a distribuição tem **elevada dispersão** se $CV > 30\%$.

Além disso, quanto menor for o valor do **coeficiente de variação**, mais homogêneos serão os dados, ou seja, menor será a dispersão em torno da média. Por isso, podemos classificar as distribuições em homogêneas ou heterogêneas, da seguinte forma:

a) a distribuição é **homogênea** quando possui dispersão baixa ou média ($CV < 30\%$);

b) a distribuição é **heterogênea** quando possui dispersão elevada ($CV > 30\%$).

EXEMPLIFICANDO

Em uma empresa de tecnologia, o salário médio dos homens é de R$ 1800,00 com desvio-padrão de R$ 810,00 e o salário médio das mulheres é de R$ 1500,00 com desvio padrão de R$ 705,00. A dispersão relativa dos salários dos homens é maior que a das mulheres?

Vamos identificar os dados do problema:

a) para os homens:

$$\begin{cases} \mu_H = 1800 \\ \sigma_H = 810 \end{cases}$$

b) para as mulheres:

$$\begin{cases} \mu_M = 1500 \\ \sigma_M = 705 \end{cases}$$

Agora, vamos calcular os respectivos coeficientes de variação:

a) para os homens:

$$CV = \frac{\sigma_H}{\mu_H} \times 100 = \frac{810}{1800} = 45{,}0\%$$

b) para as mulheres:

$$CV = \frac{\sigma_M}{\mu_M} \times 100 = \frac{705}{1500} = 47{,}0\%$$

Portanto, os salários das mulheres apresentam uma dispersão relativa maior que os salários dos homens. Além disso, as duas distribuições possuem uma alta dispersão ($CV > 30\%$).

HORA DE PRATICAR!

(FCC/ALAP/2020) O número de empregados de uma empresa é igual a 200, sendo que 60% são homens e o restante mulheres. Nesta empresa, a média aritmética dos salários da população formada pelos salários dos homens é igual a 5 mil reais, com um coeficiente de variação igual a 30%, e a média aritmética dos salários da população formada pelos salários das mulheres também é igual a 5 mil reais, porém com um coeficiente de variação igual a 20%. Considerando a população formada por todos os 200 empregados da empresa, obtém-se que a variância, em mil reais ao quadrado, dos respectivos salários é igual a

a) 1,69

b) 1,75

c) 1,30

d) 2,50

e) 3,25

Comentários:

Para responder essa questão, encontraremos os dados considerando separadamente os homens e depois faremos o mesmo processo para as mulheres. Ao final, acharemos o que foi pedido para a população $N = 200$.

Segundo a questão, a população tem tamanho igual a 200, isto é, $N = 200$. Dessa população de empregados, temos que 60% são homens, ou seja:

$$60\% \times 200 \; = \; 120 \; homens.$$

Consequentemente, o número de mulheres será:

$$200 - 120 \; = \; 80 \; mulheres.$$

De acordo com o enunciado, a média aritmética dos salários da população tem coeficiente de variação igual a 30%, isto é, $CV_{homens} = 30\%$. Esse coeficiente é calculado por meio da seguinte fórmula:

$$CV_{homens} = \frac{\sigma}{\mu}$$

A questão nos informou que a média salarial dos homens é de 5 mil reais, ou seja, $\mu = 5 \; (mil \; reais)$. Logo, usando a fórmula acima, conseguiremos encontrar o desvio padrão "populacional" dos homens:

$$30\% = \frac{\sigma_{homens}}{5}$$

$$\sigma_{homens} = 30\% \times 5$$

$$\sigma_{homens} = 1{,}5 \; (mil \; reais)$$

Sabemos que a variância é o quadrado do desvio padrão, então:

$$\sigma^2_{homens} = (1{,}5)^2 = 2{,}25(mil \; reais)^2$$

Adotaremos o mesmo procedimento para as mulheres. A questão nos informou que a média salarial das mulheres é de 5 mil reais, ou seja, $\mu = 5\ (mil\ reais)$. A única diferença é que o coeficiente de variação das mulheres é igual a 20%, $CV_{mulheres} = 20\%$.

$$CV_{mulheres} = \frac{\sigma_{mulheres}}{\mu}$$

$$20\% = \frac{\sigma_{mulheres}}{5}$$

$$\sigma_{mulheres} = 1{,}00\ (mil\ reais)$$

A variância é o quadrado do desvio padrão, então:

$$\sigma^2_{mulheres} = (1{,}00)^2 = 1{,}00(mil\ reais)^2$$

Agora, consideraremos toda a população $N = 200$. A média populacional dos salários dos 200 empregados será 5 mil, já que tanto a média salarial dos homens quanto a média salarial das mulheres é igual é igual a 5 mil reais. Portanto:

$$\bar{x} = 5\ (mil\ reais)$$

Agora, para encontrar a variância, vamos utilizar a fórmula clássica da variância:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

Buscaremos o termo $\sum(x_i - \mu)^2$ para homens e mulheres, lembrando sempre que a média é igual a $5\ (mil\ reais)$, tanto para homens quanto para mulheres.

Calculando para os homens:

$$\sigma^2_{homens} = \frac{\sum(x_i - \bar{x})^2}{N}$$

$$2{,}25 = \frac{\sum(x_i - \bar{x})^2}{120}$$

$$\sum(x_i - \mu)^2 = 2{,}25 \times 120 = 270$$

Calculando para as mulheres:

$$\sigma^2_{mulheres} = \frac{\sum(x_i - \bar{x})^2}{N}$$

$$1{,}00 = \frac{\sum(x_i - \bar{x})^2}{80}$$

$$\sum(x_i - \mu)^2 = 1{,}00 \times 80 = 80$$

Agora, substituiremos esses valores na variância de toda a população, considerando $N = 200$:

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}$$

$$\sigma^2 = \frac{270 + 80}{200} = \frac{350}{200} = 1{,}75\ (mil\ reais)^2$$

Gabarito: B.

(FCC/Metrô-SP/2019) Uma empresa possui 40 funcionários dos quais F1 são mulheres e F2 são homens. Sabe-se que a média salarial das mulheres é de 8 salários mínimos, que a média salarial dos homens é de 10 salários mínimos e que a média salarial de todos os 40 funcionários é de 8,6 salários mínimos. Se a variância dos salários dos funcionários do sexo masculino é igual a (F2 + 4) (salários mínimos)², o coeficiente de variação desses funcionários do sexo masculino é igual a

a) 32%.

b) 25%.

c) 36%.

d) 40%.

e) 15%

Comentários:

Conforme o enunciado, uma empresa possui um total de 40 funcionários, sendo um subtotal F_1 de mulheres e um subtotal F_2 de homens. Logo,

$$F_1 + F_2 = 40 \; (Equação\ 1)$$

De acordo com a questão, a média salarial das mulheres é 8, enquanto a média salarial dos homens é 10.

$$\bar{x}_{mulheres} = 8$$

$$\bar{x}_{homens} = 10$$

$$\bar{x} = 8{,}6$$

Calculando a média dos salários para homens e mulheres :

$$\frac{\Sigma(salário\ mulheres)}{F_1} = 8$$

$$\Sigma(salário\ mulheres) = 8 \times F_1$$

$$\frac{\Sigma(salário\ homens)}{F_2} = 10$$

$$\Sigma(salário\ homens) = 10 \times F_2$$

A média total pode ser calculada por meio da seguinte expressão:

$$\bar{x} = \frac{\Sigma(salário\ mulheres) + \Sigma(salário\ homens)}{F_1 + F_2}$$

$$8{,}6 = \frac{8 \times F_1 + 10 \times F_2}{F_1 + F_2}$$

$$8{,}6 \times F_1 + 8{,}6 \times F_2 = 8 \times F_1 + 10 \times F_2$$

$$0{,}6 \times F_1 = 1{,}4 \times F_2 \; (Equação\ 2)$$

Chegamos, portanto, a uma situação em que temos duas equações e duas incógnitas (F_1 e F_2). Podemos isolar a variável F_1 na Equação 2 e, em seguida, substituí-la na Equação 1, chegando ao valor de F_2.

$$F_1 = \frac{1{,}4 \times F_2}{0{,}6} = \left(\frac{7}{3}\right) \times F_2$$

Substituindo a variável F_1 na Equação 1, chegamos ao valor de F_2.

$$F_1 + F_2 = 40 \; (Equação\ 1)$$

$$\left(\frac{7}{3}\right) \times F_2 + F_2 = 40$$

Multiplicando todos os termos por 3, temos:

$$7 \times F_2 + 3 \times F_2 = 120$$

$$10 \times F_2 = 120$$

$$F_2 = 12$$

Portanto, o número de homens é 12.

O enunciado também forneceu a variância, que é equivalente à expressão (F2+4). Isto é:

$$\sigma^2 = F_2 + 4 = 12 + 4 = 16.$$

Então, o desvio padrão será a raiz quadrada da variância:

$$\sigma = 4.$$

O coeficiente de variação (CV) para os homens será:

$$CV = \frac{\sigma}{\bar{x}_{homens}} = \frac{4}{10} = 40\%$$

Gabarito: D.

(FCC/TRT 20ª Região/2016) Em uma associação de determinada carreira profissional é realizado um censo em que foram apurados os salários de todos os seus 320 associados em número de salários mínimos (S.M.). O coeficiente de variação correspondente foi de 16% e a soma dos quadrados de todos os salários, em (S.M.)2, foi de 8.204,80. O desvio padrão dos salários destes associados é, em S.M., de

a) 0,80

b) 0,64

c) 0,96

d) 0,40

e) 1,60

Comentários:

O coeficiente de variação foi informado na questão. Sabemos que ele é resultado da divisão entre o desvio padrão e a média, então:

$$\frac{16}{100} = \frac{\sigma}{\bar{x}}$$

$$\bar{x} = \frac{100\sigma}{16}$$

A variância resulta da diferença entre a média dos quadrados e o quadrado da média. Vamos aplicar o valor da média na fórmula da variância:

$$\sigma^2 = \overline{x^2} - \bar{x}^2$$

$$\sigma^2 = \frac{8.204,80}{320} - \left(\frac{100\sigma}{16}\right)^2$$

$$\sigma^2 = \frac{8.204,80}{320} - \frac{10.000\sigma^2}{256}$$

$$\sigma^2 = \frac{32.819,2 - 50.000\sigma^2}{1280}$$

$$1280\sigma^2 = 32.819,2 - 50.000\sigma^2$$

$$51280\sigma^2 = 32.819,2$$

$$\sigma^2 = \frac{32.819,2}{51280}$$

$$\sigma^2 = 0,64$$

$$\sigma = \sqrt{0,64}$$

$$\sigma = 0,8$$

Gabarito: A.

VARIÂNCIA RELATIVA

A variância relativa é uma medida de dispersão relativa que resulta do quociente entre a variância absoluta e o quadrado da média. É basicamente o quadrado do coeficiente de variação. Isto é:

a) para populações

$$VR = \left(\frac{\sigma}{\mu}\right)^2 = \frac{\sigma^2}{\mu^2}$$

b) para amostras

$$VR = \left(\frac{s}{\bar{x}}\right)^2 = \frac{s^2}{\bar{x}^2}$$

A variância relativa, assim como o coeficiente de variação, é uma medida adimensional, ou seja, não tem uma unidade de medida. Repare que tanto o numerador (variância) quanto o denominador (quadrado da média) são expressos na mesma unidade de medida, de modo a se cancelarem no momento da divisão.

EXEMPLIFICANDO

Em uma empresa de tecnologia, o salário médio dos homens é de R$ 1800,00 com desvio-padrão de R$ 810,00 e o salário médio das mulheres é de R$ 1500,00 com desvio padrão de R$ 705,00. A **variância relativa** dos salários dos homens é maior que a das mulheres?

Vamos identificar os dados do problema:

a) para os homens:

$$\begin{cases} \mu_H = 1800 \\ \sigma_H = 810 \end{cases}$$

b) para as mulheres:

$$\begin{cases} \mu_M = 1500 \\ \sigma_M = 705 \end{cases}$$

Agora, vamos calcular as respectivas variâncias relativas:

a) para os homens:

$$VR = \left(\frac{\sigma_H}{\mu_H}\right)^2 = \left(\frac{810}{1800}\right)^2 \cong 0{,}20$$

b) para as mulheres:

$$VR = \left(\frac{\sigma_M}{\mu_M}\right)^2 = \left(\frac{705}{1500}\right)^2 \cong 0{,}22$$

Portanto, os salários das mulheres apresentam uma variância relativa maior que os salários dos homens.

HORA DE PRATICAR!

(FCC/SEFAZ-BA/2019) O coeficiente de variação de Pearson correspondente a uma população P1 com média aritmética igual a 20 e tamanho 20 é igual a 30%. Decide-se excluir de P1, em um determinado momento, dois elementos iguais a 11 cada um, formando uma nova população P2. A variância relativa de P2 é igual a

a) 10/147.

b) 4/49.

c) 16/147.

d) 8/49.

e) 4/441.

Comentários:

O coeficiente de variação de Pearson é a razão entre o desvio padrão e a média.

$$CV_{P_1} = \frac{\sigma_1}{\overline{X}_1}$$

$$0{,}3 = \frac{\sigma_1}{20}$$

$$\sigma_1 = 20 \times 0{,}3 = 6$$

Logo, a variância de P_1 é:

$$\sigma_1^2 = 6^2 = 36$$

Como a variância é a média dos quadrados menos o quadrado das médias, temos:

$$\sigma_1^2 = \overline{X_1^2} - (\overline{X_1})^2$$

$$36 = \overline{X_1^2} - 20^2$$

$$36 = \overline{X_1^2} - 400$$

$$\overline{X_1^2} = 436$$

Com isso, podemos calcular a soma dos termos:

$$\overline{X_1} = \frac{\sum X_i}{20}$$

$$\overline{X_1} = \frac{\sum_{i=1}^{20} X_i}{20}$$

$$\sum_{i=1}^{20} X_i = 20 \times \overline{X_1} = 20 \times 20 = 400$$

De igual forma, temos:

$$\overline{X_1^2} = \frac{\sum_{i=1}^{20} X_i^2}{20}$$

$$\sum_{i=1}^{20} X_i^2 = 20 \times \overline{X_1^2} = 20 \times 436 = 8.720$$

O enunciado afirma que dois elementos iguais a 11 serão retirados, formando uma nova população P_2. Dessa forma, as novas somas serão iguais a:

$$\sum_{i=1}^{18} X_i = 400 - 2 \times 11 = 378$$

$$\sum_{i=1}^{18} X_i^2 = 8720 - 2 \times 11^2 = 8.478$$

Assim, as novas médias são iguais a:

$$\overline{X_2} = \frac{\sum_{i=1}^{18} X_i}{18} = \frac{378}{18} = 21$$

$$\overline{X_2^2} = \frac{\sum_{i=1}^{18} X_i^2}{18} = \frac{8.478}{18} = 471$$

De posse dessas informações, podemos calcular a nova variância absoluta:

$$\sigma_2^2 = \overline{X_2^2} - (\overline{X_2})^2$$

$$\sigma_2^2 = 471 - (21)^2$$

$$\sigma_2^2 = 471 - 441$$

$$\sigma_2^2 = 30$$

Finalmente, temos que a variância relativa é a razão entre a variância e o quadrado da média:

$$VR_{P_2} = \frac{\sigma_2^2}{(\overline{X_2})^2} = \frac{30}{441}$$

Simplificando por 3, temos:

$$VR_{P_2} = \frac{30}{441} = \frac{10}{147}$$

Gabarito: A.

QUESTÕES COMENTADAS – CESGRANRIO

Amplitude Total

1. (CESGRANRIO/IBGE/2016) Suponha que, em uma pesquisa on-line sobre as idades dos habitantes de um condomínio, um respondente de 30 anos digite erroneamente sua idade como sendo 300 anos. Considere que esse erro passe despercebido e que não haja outros erros na base de dados.

Nessas condições, a única conclusão que NÃO pode ser formulada é:

a) A média de idades calculada a partir dos dados da base será maior do que a média de idades reais dos respondentes.

b) A mediana de idades calculada a partir dos dados da base será maior do que a mediana de idades reais dos respondentes.

c) A amplitude de idades calculada a partir dos dados da base será maior do que a amplitude de idades reais dos respondentes.

d) O valor máximo das idades calculado a partir dos dados da base será maior do que a idade real do respondente mais velho.

e) A diferença entre as duas maiores idades dos dados da base será maior do que a diferença das idades reais dos dois respondentes mais velhos.

Comentários:

Analisando as alternativas, temos:

Alternativa A: **Correta.** A média é dada pela soma de todos os termos da amostra dividida pelo número de observações. Sendo assim, se o valor no somatório é aumentado, consequentemente, a média também sofre um aumento.

Alternativa B: **Errada.** A mediana corresponde ao termo central da amostra. A depender de quais os termos centrais (ou o termo central) da amostra, a mediana pode mudar ou permanecer a mesma, caso um termo seja aumentado.

Alternativa C: **Correta.** A amplitude é calculada pela diferença entre o maior e o menor valor da amostra.

Alternativa D: **Correta.** O valor máximo das idades será o maior valor informado pelos respondentes.

Alternativa E: **Correta.** Se a maior idade for aumentada, a diferença entre as duas idades também será aumentada.

Gabarito: B.

2. (CESGRANRIO/IBGE/2016) Uma pesquisa em determinado município coletou, dentre outros dados, o número de filhos em cada família. Algumas estatísticas são apresentadas na Tabela abaixo.

Número de filhos	
Média	2
Mediana	1
Moda	0
Desvio-padrão	3
Amplitude	5

Segundo essas estatísticas,

a) metade das famílias tem mais do que 2 filhos.

b) o mais comum é que famílias tenham 2 filhos.

c) mais da metade das famílias não têm filhos.

d) uma família padrão tem em média 3 filhos.

e) de todas as famílias entrevistadas, nenhuma tem 6 filhos.

Comentários:

Vamos analisar cada uma das alternativas:

Alternativa A: **Errada**. Na tabela, temos que a mediana vale 1. Assim, sabemos que metade das famílias tem mais do que 1 filho.

Alternativa B: **Errada**. Na tabela, a moda ou número mais comum vale 0.

Alternativa C: **Errada**. Não podemos afirmar que mais da metade das famílias não têm filhos, pois a mediana do número de filhos vale 1.

Alternativa D: **Errada**. Na tabela, temos que o desvio padrão vale 3. O desvio padrão não pode ser confundido com o conceito de família padrão, que está mais relacionado à moda.

Alternativa E: **Correta**. Temos que a amplitude vale 5, e que o menor número de filhos é 0. Assim, podemos calcular a amplitude:

$$AT = X_{máximo} - X_{mínimo}$$

$$5 = X_{máximo} - 0$$

$$X_{máximo} = 5$$

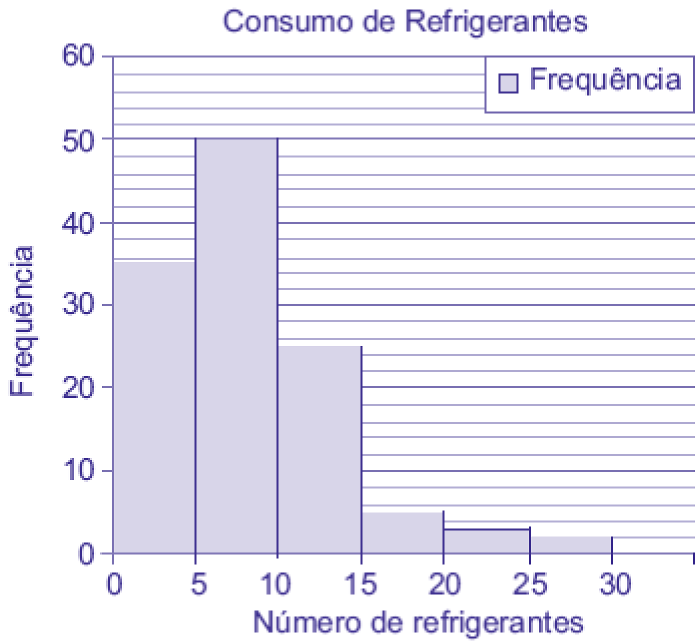
Logo, nenhuma das famílias têm 6 filhos.

Gabarito: E.

QUESTÕES COMENTADAS – CESGRANRIO

Amplitude Interquartílica

1. (CESGRANRIO/BB/2018) Uma escola de Ensino Médio decide pesquisar o comportamento de seus estudantes quanto ao número de refrigerantes consumidos semanalmente por eles. Para isso, uma amostra aleatória de 120 estudantes foi selecionada, e os dados foram sintetizados no histograma abaixo, em classes do tipo [0, 5), [5, 10), [10, 15), [15, 20), [20, 25) e [25, 30].



Qual o valor da amplitude interquartílica, obtido por meio do método de interpolação linear dos dados agrupados em classes?

a) 15

b) $\frac{15}{2}$

c) $\frac{29}{5}$

d) $\frac{47}{7}$

e) 10

Comentários:

Inicialmente, vamos colocar as informações do gráfico em uma tabela para melhor visualizarmos as informações:

Classes	Nº de refrigerantes	Nº acumulado
$0 \vdash 5$	35	35
$5 \vdash 10$	50	85
$10 \vdash 15$	25	110
$15 \vdash 20$	5	115
$20 \vdash 25$	3	118
$25 \vdash 30$	2	120

Sabemos que cada quartil corresponde a 25% da distribuição acumulada, portanto Q_1 tem até 25% das observações e Q_3 tem 75%.

Se temos um total de 120 observações, então temos que Q_1 ocupa o número acumulado de 30 e Q_3 de 90.

$$Q_1 = \frac{120}{4} = 30$$

$$Q_3 = \frac{3 \times 120}{4} = 90$$

Observando a tabela temos que Q_1está na classe [0,5) e Q_3 está na classe [10,15).

Vamos usar o método da interpolação linear para determinar o valor de cada quartil:

$$Q_k = l_{inf_{Q_k}} + \left[\frac{\frac{k \times \sum f_i}{4} - f_{ac_{ant}}}{f_{Q_k}} \right] \times h_{Q_k}$$

em que:

$l_{inf_{Q_k}}$ = limite inferior da classe do quartil considerado;

$f_{ac_{ant}}$ = frequência acumulada da classe anterior à classe do quartil considerado;

h_{Q_k} = amplitude do intervalo de classe do quartil considerado;

f_{Q_k} = frequência simples da classe do quartil considerado.

Aplicando a fórmula acima para o primeiro quartill:

$$Q_1 = l_{inf\,Q_1} + \left[\frac{\frac{1 \times \sum f_i}{4} - fac_{ant}}{f_{Q_1}}\right] \times h_{Q_1}$$

$$Q_1 = 0 + \left[\frac{\left(\frac{1 \times 120}{4}\right) - 0}{35}\right] \times (5 - 0)$$

$$Q_1 = \left[\frac{30 - 0}{35}\right] \times 5$$

$$Q_1 = \frac{150}{35}$$

$$Q_1 = \frac{30}{7}$$

Fazendo o mesmo para o terceiro quartil:

$$Q_3 = l_{inf\,Q_3} + \left[\frac{\frac{3 \times \sum f_i}{4} - fac_{ant}}{f_{Q_1}}\right] \times h_{Q_3}$$

$$Q_3 = 10 + \left[\frac{\left(\frac{3 \times 120}{4}\right) - 85}{25}\right] \times (15 - 10)$$

$$Q_3 = 10 + \left[\frac{90 - 85}{25}\right] \times 5$$

$$Q_3 = 10 + \left(\frac{25}{25}\right)$$

$$Q_3 = 10 + 1$$

$$Q_3 = 11$$

Já temos os valores de Q_3 e Q_1, agora é só fazermos a diferença para sabermos a amplitude interquartílica:

$$Q_3 - Q_1 = 11 - \frac{30}{7} = \frac{77 - 30}{7} = \frac{47}{7}$$

Gabarito: D.

2. (CESGRANRIO/PETROBRAS/2018) Você dispõe de um montante para investir em ações e precisa decidir em que empresa(s) vai alocar esse montante. Três empresas lhe parecem interessantes, e você resolve consultar o desempenho delas nos últimos sessenta meses para minimizar possíveis riscos da sazonalidade no movimento da Bolsa de Valores. Os dados revelaram a seguinte distribuição, em %, das rentabilidades mensais das ações:

Medidas Estatísticas	Empresa A	Empresa B	Empresa C

Rentabilidade média mensal	0,50	0,60	0,40
Desvio padrão	1,00	1,20	0,80
Rentabilidade mínima	-1,80	-2,20	-1,20
Rentabilidade máxima	2,20	2,30	1,80
1º quartil	-0,20	-0,30	-0,10
3º quartil	0,80	0,90	0,70

A alocação dos recursos vai ser feita de acordo com a atitude conservadora de não investir em empresa com rentabilidade considerada outlier, entendendo como tal aquela que apresentar valor além de 1,5 desvio quartílico abaixo ou acima dos quartis 1 e 3.

Com base nesse critério, a escolha do investimento deve recair sobre a(s)

a) empresa A, apenas

b) empresa B, apenas

c) empresa C, apenas

d) empresas A e C, apenas

e) três empresas

Comentários:

Geralmente, os limites para consideração de valores outliers se baseiam na amplitude quartílica, que é somente a diferença entre o terceiro e o primeiro quartil. Assim, temos que a amplitude quartílica é dada por:

$$DIQ = Q_3 - Q_1$$

Vamos calcular a amplitude para cada empresa:

$$DIQ_A = Q_3 - Q_1 = 0{,}8 - (-0{,}2) = 1$$

$$DIQ_B = Q_3 - Q_1 = 0{,}9 - (-0{,}3) = 1{,}2$$

$$DIQ_C = Q_3 - Q_1 = 0{,}7 - (-0{,}1) = 0{,}8$$

Agora, vamos multiplicar a amplitude quartílica de cada empresa por 1,5, conforme enunciado:

$$\Delta_A = 1 \times 1{,}5 = 1{,}5$$

$$\Delta_B = 1{,}2 \times 1{,}5 = 1{,}8$$

$$\Delta_C = 0{,}8 \times 1{,}5 = 1{,}2$$

Pronto, já podemos verificar as rentabilidades mínima e máxima para cada empresa:

Medidas Estatísticas	Empresa A	Empresa B	Empresa C
Rentabilidade mínima	$-0{,}2 - 1{,}5 = -1{,}7$	$-0{,}3 - 1{,}8 = -2{,}1$	$-0{,}1 - 1{,}2 = -1{,}3$
Rentabilidade máxima	$0{,}8 + 1{,}5 = 2{,}3$	$0{,}9 + 1{,}8 = 2{,}7$	$0{,}7 + 1{,}2 = 1{,}9$

Portanto, somente a empresa C estaria apta a receber os investimentos.

Gabarito: C.

3. (CESGRANRIO/EPE/2014) A Tabela a seguir apresenta a vazão média em cada mês para um determinado rio.

Mês	Vazão média mensal (m^3/s)
Janeiro	97
Fevereiro	60
Março	50
Abril	60
Maio	70
Junho	85
Julho	60
Agosto	50
Setembro	68
Outubro	117
Novembro	80
Dezembro	43

De acordo com os dados da Tabela, a mediana e a amplitude interquartílica das vazões valem, respectivamente,

a) 70 e 25

b) 64 e 25

C) 64 e 27,5

d) 68 e 27,5

e) 70 e 27,5

Comentários:

Inicialmente, vamos organizar os dados da tabela em ordem crescente:

$$\underbrace{43 \quad 50 \quad 50} \quad \underbrace{60 \quad 60 \quad 60} \quad \underbrace{68 \quad 70 \quad 80} \quad \underbrace{85 \quad 97 \quad 117}$$

Sabemos que cada quartil corresponde a 25% da distribuição, portanto até Q1 tem 25% das observações e Q3 tem 75%.

Para o primeiro quartil, temos que será a média entre 50 e 60:

$$Q_1 = \frac{50 + 60}{2} = 55$$

Para o segundo quartil, temos que é a mesma mediana da amostra e será a média entre 60 e 68:

$$Q_2 = \frac{60 + 68}{2} = 64$$

Para o terceiro quartil, temos que será a média entre 80 e 85:

$$Q_3 = \frac{80 + 85}{2} = 82{,}5$$

Portanto, já sabemos que a mediana vale 64. Calculando a amplitude:

$$Q_3 - Q_1 = 82{,}5 - 55 = 27{,}5$$

Gabarito: C.

QUESTÕES COMENTADAS – CESGRANRIO

Variância

1. (CESGRANRIO/BB/2018) Uma amostra aleatória de tamanho 5 é retirada de uma população e observa-se que seus valores, quando postos em ordem crescente, obedecem a uma Progressão Aritmética.

Se a variância amostral não viciada vale 40, qual é o valor da razão da Progressão Aritmética?

a) 3

b) $5\sqrt{2}$

c) 4

d) $2\sqrt{5}$

e) 1

Comentários:

O enunciado nos informa que 5 números formam uma P.A. Assim, podemos considerar que a diferença entre eles é r, que é a razão da progressão aritmética:

$$x - 2r; \quad x - r; \quad x; \quad x + r; \quad x + 2r$$

Calculando a média dessa amostra, temos:

$$\bar{x} = \frac{(x - 2r) + (x - r) + x + (x + r) + (x + 2r)}{5}$$

$$\bar{x} = \frac{5x}{5}$$

$$\bar{x} = x$$

Temos que a variância é dada por:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1}$$

$$40 = \frac{\left((x - 2r) - x\right)^2 + \left((x - r) - x\right)^2 + (x - x)^2 + \left((x + r) - x\right)^2 + ((x + 2r) - x)^2}{5 - 1}$$

$$40 = \frac{(-2r)^2 + (-r)^2 + 0^2 + r^2 + (2r)^2}{4}$$

$$4r^2 + r^2 + r^2 + 4r^2 = 40 \times 4$$

$$10r^2 = 160$$

$$r^2 = 16$$

$$r = \sqrt{16}$$

Assim, descobrimos que a razão da progressão aritmética é:

$$r = 4$$

Gabarito: C.

2. (CESGRANRIO/EPE/2014) Uma amostra de tamanho $200, x_1, x_2, \ldots, x_{200}$, foi retirada de uma população, e seus valores foram transformados segundo a função $y_i = 4x_i - 1$ para $i = 1, 2, \ldots, 200$.

Sabendo-se que a média e a variância dos dados transformados $y_1, y_2, \ldots, y_{200}$ são, respectivamente, 3 e 16, os valores da média e da variância dos dados originais são, respectivamente,

a) 1 e 1

b) 1 e 4

c) 3/4 e 63

d) 11 e 64

e) 11 e 256

Comentários:

Ao somarmos ou subtrairmos uma constante de todos os valores de uma variável, a média do conjunto também aumenta ou diminui no valor da constante. A mesma regra se aplica à multiplicação e divisão, a média é multiplicada ou dividida pela constante.

Assim, temos a função dada no enunciado:

$$y_i = 4x_i - 1$$

Calculando a média dos x_i:

$$\bar{x} = \frac{\bar{y} + 1}{4}$$

$$\bar{x} = \frac{3 + 1}{4}$$

$$\bar{x} = 1$$

Para a variância a regra é outra, **somando-se ou subtraindo-se uma constante de todos os valores de uma variável**, **a variância do conjunto não se altera**. Porém, **se multiplicarmos ou dividirmos todos os valores de uma variável por uma constante, a variância fica multiplicada ou dividida pelo quadrado da constante**.

Assim temos:

$$Var(x) = Var\left(\frac{y + 1}{4}\right)$$

$$Var(x) = \frac{1}{4^2} \times Var(y+1)$$

$$Var(x) = \frac{1}{4^2} \times Var(y)$$

$$Var(x) = \frac{16}{4^2}$$

$$Var(x) = \frac{16}{16}$$

$$Var(x) = 1$$

Gabarito: A.

3. (CESGRANRIO/FINEP/2014) O enunciado a seguir deve ser usado para responder à questão.

Abaixo são apresentadas estatísticas das notas brutas obtidas pelos candidatos em um concurso público:

Média aritmética: 78

Variância: 100

A nota de cada candidato foi transformada em nota padronizada, calculada considerando-se a seguinte fórmula:

$$\boldsymbol{Nota\ padronizada = 50 + 5 \times \frac{Nota\ bruta\ do\ candidato\ -\ Media\ aritmetica\ das\ notas\ brutas}{Desvio\ padrao\ das\ notas\ brutas}}$$

A variância das notas padronizadas é

a) 25

b) 50,5

c) 52,5

d) 55

e) 75

Comentários:

Pelas propriedades, sabemos que a variância de uma constante é 0 (zero). Também sabemos que o desvio padrão (σ) não é influenciado pela adição, e que corresponde à raiz da variância. Assim, temos:

$$Var_{N_{padron}} = Var(50) + Var\left(5 \times \left(\frac{N_{bruta} - \bar{x}}{\sigma}\right)\right)$$

$$Var_{N_{padron}} = 0 + Var\left(\frac{5 \times (N_{bruta} - 78)}{10}\right)$$

Agora, **se multiplicarmos ou dividirmos todos os valores de uma variável por uma constante, a variância fica multiplicada ou dividida pelo QUADRADO da constante:**

$$Var(aX + b) = a^2 \times Var(X)$$

Aplicando essa propriedade, temos:

$$Var_{N_{padron}} = Var\left(\left(\frac{5}{10}\right) \times (N_{bruta} - 78)\right)$$

$$Var_{N_{padron}} = \left(\frac{5}{10}\right)^2 \times Var(N_{bruta} - 78)$$

Somando-se ou subtraindo-se uma constante de todos os valores de uma variável, a variância do conjunto não se altera, logo, $Var(N_{bruta} - 78) = Var(N_{bruta})$.

$$Var_{N_{padron}} = \frac{25}{100} \times Var(N_{bruta})$$

$$Var_{N_{padron}} = \frac{25}{100} \times 100$$

$$Var_{N_{padron}} = 25$$

Gabarito: A.

4. (CESGRANRIO/BNDES/2013) Em um departamento de uma empresa, o gerente decide dar um aumento a todos os empregados, dobrando o salário de todos eles.

Em relação às estatísticas dos novos salários, considere as afirmativas abaixo.

I - A média dobra.

II - A variância dobra.

III - A moda dobra.

É correto o que se afirma em

a) I, apenas

b) II, apenas

c) I e III, apenas

d) II e III, apenas

e) I, II e III

Comentários:

Sabemos que, ao somarmos ou subtrairmos uma constante de todos os valores de uma variável, a média do conjunto também **aumenta ou diminui no valor da constante**. A mesma regra se aplica à multiplicação e divisão, a média é **multiplicada ou dividida pela constante**.

Para a variância a regra é outra, somando-se ou subtraindo-se uma constante de todos os valores de uma variável, a variância do conjunto **não se altera**. Porém, se multiplicarmos ou dividirmos todos os valores de uma variável por uma constante, a variância fica **multiplicada ou dividida pelo quadrado da constante**.

Assim, apenas os itens I e III estão corretos.

Gabarito: C.

5. (CESGRANRIO/BNDES/2010) Em uma pesquisa de preços de determinado produto, foram obtidos os valores, em reais, de uma amostra aleatória colhida em 6 estabelecimentos que o comercializam.

Ação da Empresa	Resultado
P	5,00
Q	8,00
R	6,00
S	6,00
T	4,00
U	7,00

A variância dessa amostra é

a) 1,50

b) 1,75

c) 2,00

d) 2,25

e) 2,50

Comentários:

Vamos iniciar calculando a média:

$$\bar{x} = \frac{5+8+6+6+4+7}{6}$$

$$\bar{x} = \frac{36}{6}$$

$$\bar{x} = 6$$

Agora, vamos calcular os desvios de cada nota em relação à média:

(x_i)	5	8	6	6	4	7	**Total**
Desvio em relação à média $(x_i - \bar{x})$	5-6=-1	8-6=2	6-6=0	6-6=0	4-6=-2	7-6=1	
$(x_i - \bar{x})^2$	1	4	0	0	4	1	**10**

Pronto, já podemos calcular a variância. Lembrando que, como se trata de variância amostral, aplicamos n-1.

$$s^2 = \frac{10}{n-1} = \frac{10}{6-1} = \frac{10}{5} = 2$$

Gabarito: C.

QUESTÕES COMENTADAS – CESGRANRIO

Desvio-Padrão

1. (CESGRANRIO/BACEN/2009) Considere as informações a seguir para responder à questão.

A viabilidade financeira do projeto de uma microempresa leva em consideração dados históricos de 100 projetos semelhantes.

A tabela abaixo mostra a distribuição de frequências do VPL - Valor Presente Líquido (valores em milhões de reais) de um conjunto de microempresas similares.

VPL	Frequência relativa
$-10 < x \leq 0$	10%
$0 < x \leq 10$	80%
$10 < x \leq 20$	10%

Segundo os dados históricos, o valor, em milhões de reais, que mais se aproxima do desvio padrão do VPL da microempresa é

a) 1

b) 2

c) 2,5

d) 4

e) 4,5

Comentários:

Queremos calcular o desvio padrão. Vamos reescrever a tabela colocando os pontos médios e multiplicado as frequências:

VPL	Pontos médios (x_i)	x_i^2	Frequência (f)	$x_i^2 \times f$
$-10 < x \leq 0$	-5	25	0,1	2,5
$0 < x \leq 10$	5	25	0,8	20

$\mathbf{10 < x \leq 20}$	15	225	0,1	22,5
	Total		1	45

Vamos iniciar a questão calculando a média. Na tabela percebemos uma distribuição simétrica, portanto a média será o ponto médio da classe central. Assim temos que:

$$\bar{x} = 5$$

$$\overline{x^2} = \frac{45}{1} = 45$$

A variância é dada por:

$$Var = \overline{x^2} - \bar{x}^2$$

$$Var = 45 - 5^2$$

$$Var = 45 - 25$$

$$Var = 20$$

O desvio-padrão será:

$$\sigma = \sqrt{Var}$$

$$\sigma = \sqrt{20}$$

$$\sigma \cong 4{,}47$$

Gabarito: E.

2. (CESGRANRIO/BNDES/2007) O enunciado abaixo refere-se à questão.

Um grupo é formado por 10 pessoas, cujas idades são:

17 19 19 20 20 20 20 21 22 22

Seja μ a média aritmética das idades e σ seu desvio padrão. O número de pessoas desse grupo cujas idades pertencem ao intervalo $[\mu - \sigma, \mu + \sigma]$ é

(Considere $\sqrt{2} = 1,4$)

a) 9

b) 8

c) 7

d) 6

e) 5

Comentários:

Vamos iniciar calculando a média:

$$\bar{x} = \frac{17 + 19 + 19 + 20 + 20 + 20 + 20 + 21 + 22 + 22}{10}$$

$$\bar{x} = \frac{200}{10}$$

$$\bar{x} = 20$$

Agora, vamos calcular os desvios de cada nota em relação à média:

(x_i)	17	19	19	20	20	20	20	21	22	22	**Total**
$(x_i - \bar{x})$	-3	-1	-1	0	0	0	0	1	2	2	
$(x_i - \bar{x})^2$	9	1	1	0	0	0	0	1	4	4	**20**

Pronto, já podemos calcular a variância:

$$\sigma^2 = \frac{20}{10} = 2$$

Calculando o desvio padrão:

$$\sigma = \sqrt{2} = 1{,}4$$

Feito isso, podemos aplicar o intervalo dado no enunciado:

$$\bar{x} - \sigma = 20 - 1{,}4 = 18{,}6$$

$$\bar{x} + \sigma = 20 + 1{,}4 = 21{,}4$$

Assim, temos nesse intervalo:

$$19\ 19\ 20\ 20\ 20\ 20\ 21$$

Gabarito: C.

QUESTÕES COMENTADAS – CESGRANRIO

Coeficiente de Variação (ou Dispersão Relativa)

1. (CESGRANRIO/BB/2021) Um pesquisador recebeu os dados de uma amostra de tamanho 100 de uma população e calculou a média amostral μ, o desvio padrão amostral σ e o coeficiente de variação amostral $CV = \frac{\sigma}{\mu}$. Antes de iniciar a análise, ele foi informado de que os dados dessa amostra estavam todos errados, mas que podiam ser corrigidos somando-se 3 a cada um dos dados que recebeu.

Após fazer tal correção, o valor do coeficiente de variação amostral passou a ser

a) $\frac{3\sigma}{\mu+3}$

b) $\frac{300\sigma}{\mu+300}$

c) $\frac{\sigma}{\mu+3}$

d) $\frac{\sigma}{\mu+300}$

e) $\frac{\sigma}{\mu+0{,}03}$

Comentários:

A média, ao ser somada uma constante k, a nova média também será somada dessa constante, portanto, $\mu + 3$. Já para a variância e desvio padrão, a soma da constante não tem influência. Assim, o valor do coeficiente de variação amostral passará a ser:

$$CV = \frac{\sigma}{\mu + 3}$$

Gabarito: C.

2. (CESGRANRIO/BB/2018) Há dez anos a média das idades, em anos completos, de um grupo de 526 pessoas era de 30 anos, com desvio padrão de 8 anos.

Considerando-se que todas as pessoas desse grupo estão vivas, o quociente entre o desvio padrão e a média das idades, em anos completos, hoje, é

a) 0,45

b) 0,42

c) 0,20

d) 0,27

e) 0,34

Comentários:

Temos no enunciado que a média há 10 anos era 30 anos para um grupo de 526 pessoas. Então, vamos calcular a soma das idades:

$$30 = \frac{x_1 + x_2 + \cdots + x_{526}}{526}$$

$$x_1 + x_2 + \cdots + x_{526} = 526 \times 30$$

$$x_1 + x_2 + \cdots + x_{526} = 15.780$$

Agora, vamos calcular a média atual das idades, considerando que cada pessoa envelheceu 10 anos:

$$\bar{x} = \frac{15.780 + (526 \times 10)}{526}$$

$$\bar{x} = \frac{15.780 + 5.260}{526}$$

$$\bar{x} = \frac{21.040}{526}$$

$$\bar{x} = 40$$

Utilizando as propriedades do desvio padrão, sabemos que essa medida não sofre alteração com o acréscimo de uma constante, como é o caso. Assim, não houve alteração do desvio padrão em relação à media anterior. Desta forma, o quociente entre o desvio padrão e a média das idades atual é:

$$\frac{8}{40} = 0,2$$

Gabarito: C.

3. (CESGRANRIO/LIQUIGÁS/2018) Uma amostra $(x_1, x_2, \ldots, x_{10})$ de tamanho 10 de uma população nos forneceu os seguintes valores

$$\sum_{i=1}^{10} x_i = 20 \ e \ \sum_{i=1}^{10} x_i^2 = 140.$$

Qual o valor do coeficiente de variação amostral?

a) $\frac{7}{20}$

b) $\frac{5}{3}$

c) $\frac{1}{7}$

d) $\frac{1}{6}$

e) 6

Comentários:

O coeficiente de variação é determinado pela divisão entre o desvio padrão e a média aritmética. Por sua vez, a variância populacional (σ^2) é dada pela diferença entre a média dos quadrados e o quadrado da média. Assim, vamos iniciar a resolução calculando o quadrado da média:

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10}$$

$$\bar{x} = \frac{20}{10}$$

$$\bar{x} = 2$$

$$\bar{x}^2 = 2^2 = 4$$

Vamos calcular a média dos quadrados:

$$\overline{x^2} = \frac{\sum_{i=1}^{10} x_i^2}{10}$$

$$\overline{x^2} = \frac{140}{10}$$

$$\overline{x^2} = 14$$

Agora, já podemos calcular a variância populacional:

$$\sigma^2 = \overline{x^2} - \bar{x}^2$$

$$\sigma^2 = 14 - 4$$

$$\sigma^2 = 10$$

Temos a variância populacional, mas precisamos calcular a variância amostral, logo:

$$s^2 = \sigma^2 \times \frac{n}{n-1}$$

$$s^2 = 10 \times \frac{10}{10-1}$$

$$s^2 = 10 \times \frac{10}{9}$$

$$s^2 = \frac{100}{9}$$

$$s = \sqrt{\frac{100}{9}}$$

Assim, já podemos calcular o coeficiente de variação amostral:

$$CV=\frac{s}{\bar{x}}=\frac{\sqrt{\frac{100}{9}}}{2}=\frac{\frac{10}{3}}{2}=\frac{5}{3}$$

Gabarito: B.

4. (CESGRANRIO/EPE/2014) Um pesquisador está interessado em comparar a variabilidade de duas variáveis com médias diferentes e desvios padrões diferentes, presentes num dado estudo estatístico.

A medida estatística adequada a ser usada nesse contexto é a(o)

a) covariância

b) diferença entre a maior e a menor variância

c) razão entre o maior e o menor desvios padrões

d) coeficiente de variação

e) coeficiente de correlação

Comentários:

Das medidas estatísticas, a que mais se adequa aos critérios do enunciado é o **coeficiente de variação**, definido pela razão entre o **desvio padrão** e a **média** da amostra, sendo uma medida adimensional. O coeficiente de variação é usado para expressar o grau de variabilidade dos dados estatísticos, desconsiderando a influência da ordem de grandeza da variável.

Gabarito: D.

5. (CESGRANRIO/IBGE/2014) Os dados a seguir foram obtidos de empregados de uma empresa com três fábricas: I, II e III. A variável de interesse é salário.

Empresa	Média	Desvio padrão
Fábrica I	1185	630,49
Fábrica II	600	355,97
Fábrica III	2150	1106,16

Comparando-se a variabilidade de salários em relação ao salário médio das três fábricas, através de seus coeficientes de variação, conclui-se que a variabilidade da fábrica

a) I é menor apenas do que a da fábrica III.

b) II é menor apenas do que a da fábrica I.

c) II é menor apenas do que a da fábrica III.

d) II é menor do que as das outras duas fábricas.

e) III é menor do que as das outras duas fábricas.

Comentários:

Sabemos que o coeficiente de variação é dado pela divisão entre o desvio padrão e a média. Assim, podemos calcular o coeficiente de variação das três empresas:

Fabrica I: $C_v = \frac{630,49}{1185} = 0,53$

Fabrica II: $C_v = \frac{355,97}{600} = 0,59$

Fabrica III: $C_v = \frac{1106,16}{2150} = 0,51$

Logo, o menor coeficiente de variação é da fábrica III.

Gabarito: E.

6. (CESGRANRIO/IBGE/2013) Seja X uma variável aleatória com distribuição normal cuja média é μ e o desvio padrão é σ.

Se Y = 2X - 1 tem distribuição normal com média 5 e variância 20, o coeficiente de variação populacional $\frac{\sigma}{\mu}$ vale

a) $\frac{\sqrt{42}}{6}$

b) $\frac{\sqrt{21}}{6}$

c) $\frac{\sqrt{5}}{3}$

d) $\frac{\sqrt{39}}{9}$

e) $\frac{4\sqrt{5}}{9}$

Comentários:

Pelas propriedades da variância, sabemos que, se somarmos ou subtrairmos uma constante de todos os valores de uma variável, a variância do conjunto não se altera. Porém, se multiplicarmos ou dividirmos todos os valores de uma variável por uma constante, a variância fica multiplicada ou dividida pelo quadrado da constante. Assim, temos que:

$$Var(Y) = Var(2X - 1)$$

$$Var(Y) = 2^2Var(X) + Var(1)$$

A variância de uma constante é sempre zero, então, $Var(1) = 0$. Dessa forma, temos:

$$20 = 4\sigma^2 + 0$$

$$\sigma^2 = 5$$

$$\sigma = \sqrt{5}$$

Pelas propriedades da média, sabemos que, se somarmos ou subtrairmos uma constante de todos os valores de uma variável, a média do conjunto também **aumenta ou diminui no valor da constante**. A mesma regra se aplica à multiplicação e divisão, a média é **multiplicada ou dividida pela constante**. Assim, temos:

$$\bar{Y} = 2 \times \bar{X} - 1$$

$$5 = 2\mu - 1$$

$$2\mu = 6$$

$$\mu = 3$$

Calculando o coeficiente de variação:

$$C_v = \frac{\sigma}{\mu}$$

$$C_v = \frac{\sqrt{5}}{3}$$

Gabarito: C.

7. (CESGRANRIO/BB/2013) A variância de um conjunto de dados é 4 m^2.

Para o mesmo conjunto de dados foram tomadas mais duas medidas de variabilidade:

a diferença entre o terceiro e o primeiro quartil e o coeficiente de variação.

Esses dois valores caracterizam-se, respectivamente, por

a) possuírem unidades de medida m^2 e m.

b) possuírem unidades de medida m e m^2.

c) ser adimensional e possuir unidade de medida m^2.

d) possuir unidade de medida m e ser adimensional.

e) possuir unidade de medida m^2 e ser adimensional.

Comentários:

Sabemos que a variância é representada no quadrado da unidade de dados original, logo os dados originais da questão estão em metros.

A amplitude interquartílica é dada por $Q_3 - Q_1$. Assim como os dados originais, estão expressos em metros.

O coeficiente de variação é dado por $C_v = \frac{\sigma}{\bar{x}}$, sendo assim, adimensional.

Gabarito: D.

8. (CESGRANRIO/BNDES/2013) Quatro variáveis são utilizadas em um modelo de previsão da quantidade produzida de uma determinada commodity agrícola. São elas:

• temperatura, em graus Celsius

• quantidade de fertilizante, em toneladas

• variação dos preços praticados no mercado internacional, em %

• quantidade produzida de um produto similar, em toneladas

Para determinar qual dessas variáveis apresenta a maior variabilidade, deve-se utilizar

a) apenas a média de cada uma das variáveis

b) apenas a variância de cada uma das variáveis

c) apenas o desvio padrão de cada uma das variáveis

d) a relação $\frac{desvio\ padrão}{média}$ de cada uma das variáveis

e) a relação $\frac{variância}{média}$ de cada uma das variáveis

Comentários:

Vamos analisar as alternativas:

Alternativa A – **Errada.** A média não representa uma medida de variabilidade e sim de tendência central.

Alternativa B – **Errada.** A variância é uma medida de dispersão absoluta, para comparar variáveis diferentes costumamos usar uma medida de dispersão relativa.

Alternativa C - **Errada.** O desvio padrão é também uma medida de dispersão absoluta.

Alternativa D – **Correta.** Conforme já dito, o coeficiente de variação é uma medida de dispersão relativa dada por:

$$Cv = \frac{desvio\ padrão}{média}$$

Alternativa E – **Errada.** O item está errado porque a média no denominador não está elevada ao quadrado.

Gabarito: D.

9. (CESGRANRIO/BACEN/2009) Considere as informações a seguir para responder à questão.

A viabilidade financeira do projeto de uma microempresa leva em consideração dados históricos de 100 projetos semelhantes.

A tabela abaixo mostra a distribuição de frequências do VPL - Valor Presente Líquido (valores em milhões de reais) de um conjunto de microempresas similares.

VPL	Frequência relativa
$-10 < x \leq 0$	10%
$0 < x \leq 10$	80%
$10 < x \leq 20$	10%

Um projeto alternativo para o investidor apresenta um VPL esperado, em reais, de 6 milhões e um risco (desvio padrão) de 2 milhões. Pela ótica do risco relativo, qual o melhor investimento, a microempresa ou o projeto alternativo?

a) A microempresa, pois apresenta um Coeficiente de Variação maior.

b) A microempresa, pois apresenta um Coeficiente de Variação menor.

c) O projeto alternativo, pois apresenta um Coeficiente de Variação maior.

d) O projeto alternativo, pois apresenta um Coeficiente de Variação menor.

e) É indiferente, pois os investimentos apresentam Coeficientes de Variação iguais.

Comentários:

Queremos calcular o coeficiente de variação. Vamos reescrever a tabela colocando os pontos médios e multiplicado as frequências:

VPL	Pontos médios (x_i)	x_i^2	Frequência (f)	$x_i^2 \times f$
$-10 < x \leq 0$	-5	25	0,1	2,5
$0 < x \leq 10$	5	25	0,8	20
$10 < x \leq 20$	15	225	0,1	22,5
	Total		1	45

Vamos iniciar a questão calculando a média. Na tabela, percebemos uma distribuição simétrica. Portanto, a média será o ponto médio da classe central. Assim, temos que:

$$\bar{x} = 5$$

$$\overline{x^2} = \frac{45}{1} = 45$$

A variância é dada por:

$$Var = \overline{x^2} - \bar{x}^2$$

$$Var = 45 - 5^2$$

$$Var = 45 - 25$$

$$Var = 20$$

O desvio-padrão será:

$$\sigma = \sqrt{Var}$$

$$\sigma = \sqrt{20}$$

$$\sigma \cong 4{,}47$$

O coeficiente de variação é dado por:

$$C_v = \frac{\sigma}{\bar{x}}$$

$$C_v = \frac{4{,}47}{5}$$

$$C_v = 0{,}89$$

Considerando o projeto alternativo:

$$C_v = \frac{\sigma}{\bar{x}}$$

$$C_v = \frac{2}{6}$$

$$C_v = 0{,}33$$

Logo, o projeto alternativo tem um coeficiente de variação menor e, consequentemente, menor risco.

Gabarito: D.

Aviso Importante!

Olá, alunos (as)!

Informamos que não temos mais questões da banca, referente ao assunto tratado na aula de hoje, em virtude de baixa cobrança deste tópico ao longo dos anos. No entanto, para complementar o estudo e deixar sua preparação em alto nível, preparamos um caderno de questões inéditas que servirá como treino e aprimoramento do conteúdo.

Em caso de dúvidas, não deixe de nos chamar no Fórum de dúvidas!

Bons estudos!

Estratégia Concursos

QUESTÕES COMENTADAS – INÉDITAS

Medidas de Dispersão

1. (INÉDITA/2022) Sobre medidas de dispersão, julgue os itens:

I – as medidas de dispersão são métricas que mostram a variação dos dados de um conjunto;

II – são medidas de dispersão: amplitude total; amplitude interquartílica; desvio médio; média; variância relativa e mediana.

III – as medidas de dispersão podem ser divididas em dois grupos: medidas de dispersão absoluta e medidas de dispersão relativa.

IV – a média, moda, mediana e o desvio padrão são medidas de posição.

V – a coeficiente de variação e a variância relativa são medidas de dispersão relativa.

Estão corretas as afirmativas:

a) Apenas I. e III.

b) Apenas I. III. e V.

c) Apenas V.

d) Apenas II. e IV.

e) Apenas IV. e V.

Comentários:

As medidas de dispersão medem o grau de variabilidade dos elementos de uma distribuição, ou seja, mostram a variação dos dados de um conjunto. As principais medidas de dispersão podem ser divididas em dois grupos: medidas de dispersão absoluta (amplitude, amplitude interquartílica, desvio médio, variância e desvio padrão) e medidas de dispersão relativa (coeficiente de variação e variância relativa).

Média aritmética, moda e mediana fazem parte do grupo de medidas de posição que sintetizam o comportamento dos elementos de um conjunto de dados.

Gabarito: B.

2. (INÉDITA/2022) Com relação às medidas de tendência central, medidas de posição e medidas de dispersão, assinale a alternativa correta:

a) Dentre as medidas de posição mais utilizadas, temos: a média, mediana, moda e o desvio padrão, sendo a mais importante delas a média aritmética.

b) Sendo a variância uma medida de dispersão, ao dividirmos todos os valores de uma variável por uma constante, a variância do novo conjunto fica dividida pelo quadrado dessa constante.

c) O coeficiente de variação é uma medida de dispersão absoluta que fornece a variação dos dados em relação à média.

d) Sendo a variância uma medida de posição, ao subtrairmos uma constante de todos os valores de uma variável, a variância do conjunto não é alterada.

e) O desvio médio faz parte das medidas de dispersão, ao somarmos uma constante a todos os valores de uma variável, o desvio médio do conjunto também será somado dessa constante.

Comentários:

Vamos analisar cada uma das alternativas:

A - Dentre as medidas de posição mais utilizadas, temos: a média, mediana, moda e o ~~desvio padrão~~, sendo ~~a mais importante delas a média aritmética~~. O desvio padrão faz parte das medidas de dispersão. Além disso, não podemos afirmar que a média aritmética é a mais importante das medidas de posição.

B - Sendo a variância uma medida de dispersão, ao dividirmos todos os valores de uma variável por uma constante, a variância do novo conjunto fica dividida pelo quadrado dessa constante. Perfeito! Uma das propriedades da variância diz que multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, a variância do conjunto fica multiplicada (ou dividida) pelo QUADRADO dessa constante.

C - O coeficiente de variação é uma medida de dispersão ~~absoluta~~ que fornece a variação dos dados em relação à média. Na verdade, o coeficiente de variação é uma medida de dispersão relativa e não absoluta como afirma a questão.

D - Sendo a variância uma ~~medida de posição~~, ao subtrairmos uma constante de todos os valores de uma variável, a variância do conjunto não é alterada. A questão erra ao afirmar que a variância é uma medida de posição, pois trata-se de uma medida de dispersão.

E- O desvio médio faz parte das medidas de dispersão, ao somarmos uma constante a todos os valores de uma variável, o desvio médio do conjunto ~~também será somado dessa constante~~. Na verdade, uma das propriedades do desvio médio diz que somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, o desvio médio do conjunto não é alterado.

Gabarito: B.

Questões Comentadas – INÉDITAS

Amplitude Total

1. (INÉDITA/2022) Uma pesquisa foi realizada em uma escola para avaliar o rendimento dos alunos do terceiro ano. Nessa avaliação foram atribuídos valores de acordo com o rendimento dos alunos nos testes aplicados pelos professores. A tabela de frequências abaixo mostra o resultado da pesquisa:

x_i	nº de alunos
3	4
6	10
8	17
10	13
15	6
Total	**50**

De acordo com as informações da tabela a amplitude total é igual a:

a) 8

b) 10

c) 12

d) 4

e) 9

Comentários:

Para dados agrupados sem intervalos de classe, a fórmula usada para a identificação da amplitude total é expressa pela seguinte fórmula:

$$\boldsymbol{A = x_{máx} - x_{mín}}$$

em que $x_{máx}$ é o maior elemento; e $x_{mín}$ é o menor elemento do conjunto.

De acordo com a tabela apresentada, temos 3 e 15 para os valores mínimo e máximo da variável x_i, assim temos o seguinte resultado:

$$A = x_{máx} - x_{mín}$$

$$A = 15 - 3 = 12$$

Gabarito: C.

QUESTÕES COMENTADAS – INÉDITAS

Amplitude Interquartílica

1. (INÉDITA/2022) Denominamos de quartis os valores de uma série que a dividem em quatro partes iguais, isto é, quatro partes contendo o mesmo número de elementos (25%). A respeito desse assunto, julgue as assertivas e marque a opção correta:

( ) A amplitude interquartílica é o resultado da subtração entre o terceiro quartil e o primeiro quartil, sendo expressa pela seguinte equação: $A_{IQ} = Q_3 - Q_1$

( ) Ao somar ou subtrair uma constante c a/de todos os valores de uma variável, a amplitude interquartílica do conjunto também será somada dessa constante.

( ) O segundo quartil corresponde à separação de metade dos elementos da série, porém, esse valor não coincide com a mediana.

( ) Ao multiplicar ou dividir todos os valores de uma variável por uma constante c, a amplitude interquartílica do conjunto também será multiplicada ou dividida por essa constante.

a) V – V – F – V

b) V – V – F – F

c) V – F – F – V

d) F – F – V – V

e) F – V – V – F

Comentários:

Vamos analisar as assertivas:

Item 1 - A amplitude interquartílica é o resultado da subtração entre o terceiro quartil e o primeiro quartil, sendo expressa pela seguinte equação: $A_{IQ} = Q_3 - Q_1$

De fato, a amplitude interquartílica é dada pela diferença entre o terceiro quartil e o primeiro quartil, assim, para calcularmos essa diferença, usamos a expressão $A_{IQ} = Q_3 - Q_1$. **Item verdadeiro**.

Item 2 - Ao somar ou subtrair uma constante c a/de todos os valores de uma variável, a amplitude interquartílica do conjunto também será somada dessa constante.

Na verdade, uma das propriedades da amplitude interquartílica diz que: Somando-se (ou subtraindo-se) uma constante c a todos os valores de uma variável, a amplitude interquartílica (e o desvio quartílico) do conjunto não é alterada. **Item falso**.

Item 3 - O segundo quartil corresponde à separação de metade dos elementos da série, porém, esse valor não coincide com a mediana.

Os quartis dividem a série em quatro partes iguais, sendo assim, o segundo quartil corresponde a 50% dos elementos da série, portanto, esse valor coincide sim com a mediana, sendo $Q_2 = M_d$. **Item falso**.

Item 4 - Ao multiplicar ou dividir todos os valores de uma variável por uma constante c, a amplitude interquartílica do conjunto também será multiplicada ou dividida por essa constante.

É exatamente isso o que diz uma das propriedades da amplitude interquartílica: multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante c, a amplitude interquartílica (e o desvio quartílico) do conjunto fica multiplicada (ou dividida) por essa constante. **Item verdadeiro**.

Logo, a sequência correta é: V – F – F – V

Gabarito: C.

2. (INÉDITA/2022) Considerando a distribuição de frequências abaixo marque a alternativa correta.

A distribuição dos salários em mil reais dos 100 funcionários de uma empresa está apresentada na tabela abaixo:

Salários (x_i)	Frequência absoluta
$2000 \vdash 3000$	8
$3000 \vdash 4000$	20
$4000 \vdash 5000$	34
$5000 \vdash 6000$	26
$6000 \vdash 7000$	12
Total	**100**

A amplitude interquartílica, dada pela diferença entre o terceiro e o primeiro quartil, é igual a:

a) R$ 2300,00

b) R$ 3500,00

c) R$ 1900,00

d) R$ 1650,00

e) R$ 1800,00

Comentários:

Inicialmente, precisamos calcular as frequências acumuladas de cada classe. Para isso, vamos reescrever a tabela acrescentado essas informações:

Salários (x_i)	Frequência absoluta	Frequência acumulada
$2000 \vdash 3000$	8	8
$3000 \vdash 4000$	20	28
$4000 \vdash 5000$	34	62
$5000 \vdash 6000$	26	88
$6000 \vdash 7000$	12	100
Total	**100**	

Agora, vamos calcular as posições do 1º quartil:

$$P_{Q_k} = \frac{k \times \sum fi}{4}$$

$$P_{Q_1} = \frac{1 \times 100}{4} = 25$$

Logo, o primeiro quartil está na posição 25, o que corresponde à 2ª classe $3000 \vdash 4000$.

Aplicando a fórmula:

$$Q_k = l_{inf\,Q_k} + \left[\frac{\frac{k \times \sum f_i}{4} - fac_{ant}}{f_{Q_k}}\right] \times h_{Q_k}$$

em que:

$l_{inf\,Q_k}$ → Limite inferior à classe quartílica

fac_{ant} → frequência acumulada da classe anterior à classe quartílica

h_{Q_k} → amplitude da classe

f_{Q_k} → frequência da classe

$$Q_1 = 3000 + \left[\frac{25 - 8}{20}\right] \times 1000$$

$$Q_1 = 3000 + 0{,}85 \times 1000$$

$$Q_1 = 3000 + 850$$

$$Q_1 = 3850$$

Calculando a posição do 3º quartil:

$$P_{Q_k} = \frac{k \times \sum fi}{4}$$

$$P_{Q_3} = \frac{3 \times 100}{4} = 75$$

Logo, o terceiro quartil está na posição 75, que corresponde à 4 ª classe $5000 \vdash 6000$.

Calculando o terceiro quartil:

$$Q_k = l_{inf\,Q_k} + \left[\frac{\frac{k \times \sum f_i}{4} - f_{ac_{ant}}}{f_{Q_k}}\right] \times h_{Q_k}$$

$$Q_3 = 5000 + \left[\frac{75 - 62}{26}\right] \times 1000$$

$$Q_3 = 5000 + 0{,}5 \times 1000$$

$$Q_3 = 5000 + 500$$

$$Q_3 = 5500$$

Dessa forma, podemos calcular a amplitude do intervalo que é dada pela diferença entre o maior e o menor:

$$A = 5500 - 3850 = 1650$$

Gabarito: D.

QUESTÕES COMENTADAS – INÉDITAS

Desvio Absoluto Médio

1. (INÉDITA/2022) Em um curso preparatório para concursos, foi realizada uma pesquisa para identificar as idades dos seus alunos. As idades estão distribuídas conforme a tabela de frequências abaixo:

idades (x_i)	Nº de alunos (f_i)
$20 \vdash 25$	150
$25 \vdash 30$	380
$30 \vdash 35$	420
$35 \vdash 40$	150
Total	**1200**

Com base na tabela, o desvio absoluto médio das idades dos alunos é:

a) 5,3

b) 6,2

c) 2,65

d) 4,8

e) 3,62

Comentários:

Vamos iniciar pelo cálculo da média. Para isso, construiremos uma coluna com os pontos médios e multiplicaremos cada um pela sua respectiva frequência. Da seguinte forma:

idades (x_i)	Pontos médios (x_i)	Nº de alunos (f_i)	$x_i \times f_i$
$20 \vdash 25$	22,5	150	3.375
$25 \vdash 30$	27,5	380	10.450
$30 \vdash 35$	32,5	420	13.650

$35 \vdash 40$	37,5	150	5.625
Total		$\sum \boldsymbol{f_i} = \mathbf{1200}$	$\sum \boldsymbol{x_i \times f_i} = \mathbf{33.100}$

Portanto, a média é:

$$\bar{x} = \frac{33100}{1200} = 27{,}58$$

Acrescentando à tabela as informações de módulos dos desvios e da multiplicação dos desvios absolutos por suas respectivas frequências:

(x_i)	(x_i)	(f_i)	$x_i \times f_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert \times f_i$
$20 \vdash 25$	22,5	150	3.375	$\lvert 22{,}5 - 27{,}58 \rvert = 5{,}08$	$5{,}08 \times 150 = 762$
$25 \vdash 30$	27,5	380	10.450	$\lvert 27{,}5 - 27{,}58 \rvert = 0{,}08$	$0{,}08 \times 380 = 30{,}4$
$30 \vdash 35$	32,5	420	13.650	$\lvert 32{,}5 - 27{,}58 \rvert = 4{,}92$	$4{,}92 \times 420 = 2066{,}4$
$35 \vdash 40$	37,5	150	5.625	$\lvert 37{,}5 - 27{,}58 \rvert = 9{,}92$	$9{,}92 \times 150 = 1488$
Total		$\sum f_i = 1200$	$\sum x_i \times f_i = 33.100$		$\sum \lvert x_i - \bar{x} \rvert \times f_i = 4346{,}8$

Assim, o desvio absoluto médio é:

$$D_m = \frac{\sum |x_i - \bar{x}| \times f_i}{n} = \frac{4.346{,}8}{1.200} = 3{,}62$$

Gabarito: E.

2. (INÉDITA/2022) Em um campeonato de basquete, a comissão organizadora fez um levantamento de quantas cestas foram feitas por cada jogador. O resultado é apresentado na tabela de frequências abaixo:

Nº de cestas (x_i)	Nº de jogadores (f_i)
11	5

14	6
23	4
25	3
Total	**18**

O desvio absoluto médio do número de cestas é de:

a) 5,33

b) 6,78

c) 5,2

d) 4,89

e) 3,55

Comentários:

Vamos iniciar a resolução calculando a média de cestas feitas. Para isso, vamos reescrever a tabela multiplicando as quantidades de cestas por suas respectivas frequências:

Nº de cestas (x_i)	Nº de jogadores (f_i)	$x_i \times f_i$
11	5	55
14	6	84
23	4	92
25	3	75
Total	$\sum f_i = 18$	$\sum x_i \times f_i = 306$

A média é dada pela razão entre os dois totais:

$$\bar{x} = \frac{\sum x_i \times f_i}{\sum f_i} = \frac{306}{18} = 17\ cestas/jogador$$

Agora, acrescentamos mais uma coluna para armazenar os módulos dos desvios multiplicados por suas respectivas frequências:

Nº de cestas (x_i)	Nº de jogadores (f_i)	$x_i \times f_i$	$\lvert x_i - \overline{x} \rvert \times f_i$
11	5	55	$\lvert 11 - 17 \rvert \times 5 = 30$
14	6	84	$\lvert 14 - 17 \rvert \times 6 = 18$
23	4	92	$\lvert 23 - 17 \rvert \times 4 = 24$
25	3	75	$\lvert 25 - 17 \rvert \times 3 = 24$
Total	$\sum f_i = 18$	$\sum x_i \times f_i = 306$	$\sum \lvert x_i - \overline{x} \rvert \times f_i = 96$

Agora, basta aplicarmos a fórmula do desvio absoluto médio:

$$D_m = \frac{\sum |x_i - \bar{x}| \times f_i}{n} = \frac{96}{18} = 5{,}33$$

Gabarito: A.

QUESTÕES COMENTADAS – INÉDITAS

Desvio-Padrão

1. (INÉDITA/2022) Em uma empresa de departamento, com 10 funcionários na área de vendas, as quantidades vendidas por cada funcionário no mês foram:

13	14	17	17	17	19	19	21	22	23

O desvio padrão das vendas foi igual a:

a) 4,01

b) 3,09

c) 3,50

d) 2,89

e) 5,2

Comentários:

Para calcularmos o desvio padrão, precisamos inicialmente calcular a média. Assim, temos que:

$$\mu = \frac{13+14+17+17+17+19+19+21+22+23}{10} = \frac{182}{10} = 18{,}2$$

O desvio padrão ($\boldsymbol{s}$ ou $\boldsymbol{\sigma}$) é definido como sendo a raiz quadrada da média aritmética dos quadrados dos desvios, sendo expresso pela fórmula:

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

Calculando o desvio dos dados apresentados, temos:

$$\sigma = \sqrt{\frac{(13-18{,}2)^2 + (14-18{,}2)^2 + (17-18{,}2)^2 + (17-18{,}2)^2 + (17-18{,}2)^2 + (19-18{,}2)^2 + (19-18{,}2)^2 + (21-18{,}2)^2 + (22-18{,}2)^2 + (23-18{,}2)^2}{10}}$$

$$\sigma = \sqrt{\frac{27{,}04 + 17{,}64 + 1{,}44 + 1{,}44 + 1{,}44 + 0{,}64 + 0{,}64 + 7{,}84 + 14{,}44 + 23{,}04}{10}}$$

$$\sigma = \sqrt{\frac{95{,}6}{10}}$$

$$\sigma = \sqrt{9{,}56}$$

$$\sigma = 3{,}09$$

Gabarito: B.

Questões Comentadas – INÉDITAS

Variância Relativa

1. (INÉDITA/2022) Seja uma amostra de candidatos que prestaram concurso para os cargos de assistente e analista. Considere que a média de pontos da prova de analista é de 86 pontos, com desvio padrão de 36; e a média de pontos da prova de assistentes é de 95 pontos, com desvio padrão de 25. A variância relativa de analistas supera a de assistentes em:

a) 0,1

b) 0,5

c) 0,7

d) 0,17

e) 0,3

Comentários:

A variância relativa é dada por:

$$VR = \left(\frac{s}{\bar{x}}\right)^2 = \frac{s^2}{\bar{x}^2}$$

em que s^2 é variância absoluta e $\bar{x}^2$ é o quadrado da média.

A questão nos deu a informação da média para analistas e assistentes, assim como a informação de desvio padrão. Sabemos que o desvio padrão é determinado pela raiz quadrada da variância. Então:

Para analistas:

$$VR = \left(\frac{s_{analistas}}{\bar{x}_{analistas}}\right)^2 = \left(\frac{36}{86}\right)^2 = 0{,}17$$

Para assistente:

$$VR = \left(\frac{s_{assitentes}}{\bar{x}_{assistentes}}\right)^2 = \left(\frac{25}{95}\right)^2 = 0{,}07$$

A questão quer saber o quanto a variação relativa dos analistas é superior à de assistentes, logo temos:

$$VR_{analistas} - VR_{assitentes} = 0{,}17 - 0{,}07 = 0{,}1$$

Gabarito: A.

2. (INÉDITA/2022) Considere uma população X com coeficiente de variação 0,3 e média aritmética igual a 35. E uma população Y com desvio padrão 25 e média aritmética igual a 49. A variância relativa da população X mais a variância relativa da população Y é igual a:

a) 0,25

b) 0,40

c) 0,55

d) 0,28

e) 0,35

Comentários:

Temos que a variância relativa é dada por:

$$VR = \frac{\sigma^2}{(\bar{X})^2}$$

Em que s^2 é variância absoluta e $\bar{x}^2$ é o quadrado da média.

A questão nos forneceu a média e o coeficiente de variação para a população X. O coeficiente de variação é determinado pela razão entre o desvio padrão e a média. Logo:

$$CV_X = \frac{\sigma_X}{\bar{X}_X}$$

$$0{,}3 = \frac{\sigma_X}{35}$$

$$\sigma_X = 35 \times 0{,}3$$

$$\sigma_X = 10{,}5$$

Assim, podemos calcular a variância relativa para a população X:

$$VR_X = \frac{10{,}5^2}{35^2} = \frac{110{,}25}{1225} = 0{,}09$$

Para a população Y, basta aplicarmos:

$$VR_Y = \frac{25^2}{49^2} = \frac{625}{2401} = 0{,}26$$

Agora, basta somarmos as variâncias relativa de X e Y:

$$VR_X + VR_Y = 0{,}09 + 0{,}26 = 0{,}35$$

Gabarito: E.

3. (INÉDITA/2022) Considere uma amostra de tamanho 15, com 270 sendo o somatório dos termos e que tem coeficiente de variação igual a 0,5. A variância relativa dessa amostra é igual a:

a) 0,18

b) 0,50

c) 0,75

d) 0,35

e) 0,25

Comentários:

A variância relativa é dada por:

$$VR = \left(\frac{s}{\bar{x}}\right)^2 = \frac{s^2}{\bar{x}^2}$$

Vamos iniciar a resolução da questão calculando a média da amostra. Ora, sabemos que a média é dada pelo quociente entre a soma de todos os elementos e o número deles. Logo:

$$\bar{x} = \frac{270}{15} = 18$$

A questão nos forneceu o coeficiente de variação, que é calculado pela razão entre o desvio padrão e a média. Logo:

$$CV = \frac{s}{\bar{x}}$$

$$0{,}5 = \frac{s}{18}$$

$$s = 18 \times 0{,}5 = 9$$

Agora, já podemos calcular a variância relativa da amostra:

$$VR = \frac{s^2}{\bar{x}^2} = \frac{9^2}{18^2} = \frac{81}{324} = 0{,}25$$

Gabarito: E.

Lista de Questões – CESGRANRIO

Amplitude Total

1. (CESGRANRIO/IBGE/2016) Suponha que, em uma pesquisa on-line sobre as idades dos habitantes de um condomínio, um respondente de 30 anos digite erroneamente sua idade como sendo 300 anos. Considere que esse erro passe despercebido e que não haja outros erros na base de dados.

Nessas condições, a única conclusão que NÃO pode ser formulada é:

a) A média de idades calculada a partir dos dados da base será maior do que a média de idades reais dos respondentes.

b) A mediana de idades calculada a partir dos dados da base será maior do que a mediana de idades reais dos respondentes.

c) A amplitude de idades calculada a partir dos dados da base será maior do que a amplitude de idades reais dos respondentes.

d) O valor máximo das idades calculado a partir dos dados da base será maior do que a idade real do respondente mais velho.

e) A diferença entre as duas maiores idades dos dados da base será maior do que a diferença das idades reais dos dois respondentes mais velhos.

2. (CESGRANRIO/IBGE/2016) Uma pesquisa em determinado município coletou, dentre outros dados, o número de filhos em cada família. Algumas estatísticas são apresentadas na Tabela abaixo.

Número de filhos	
Média	2
Mediana	1
Moda	0
Desvio-padrão	3
Amplitude	5

Segundo essas estatísticas,

a) metade das famílias tem mais do que 2 filhos.

b) o mais comum é que famílias tenham 2 filhos.

c) mais da metade das famílias não têm filhos.

d) uma família padrão tem em média 3 filhos.

e) de todas as famílias entrevistadas, nenhuma tem 6 filhos.

GABARITO – CESGRANRIO

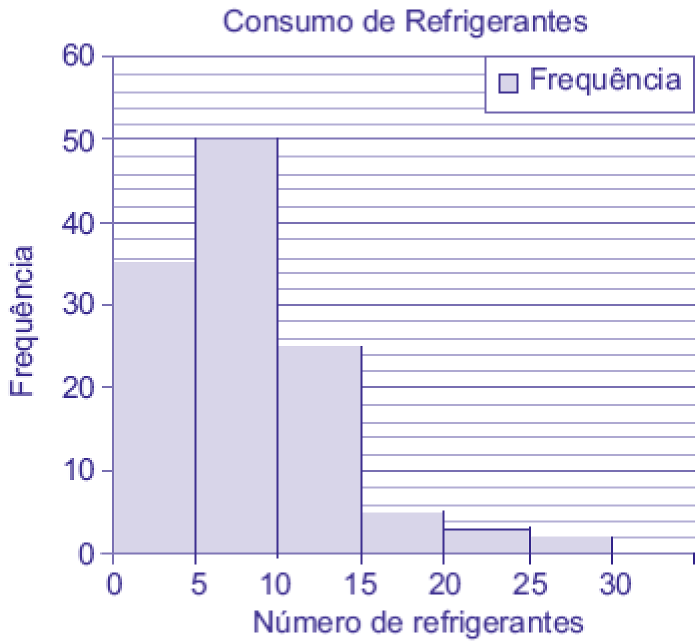
Amplitude Total

1. LETRA B
2. LETRA E

LISTA DE QUESTÕES – CESGRANRIO

Amplitude Interquartílica

1. (CESGRANRIO/BB/2018) Uma escola de Ensino Médio decide pesquisar o comportamento de seus estudantes quanto ao número de refrigerantes consumidos semanalmente por eles. Para isso, uma amostra aleatória de 120 estudantes foi selecionada, e os dados foram sintetizados no histograma abaixo, em classes do tipo [0, 5), [5, 10), [10, 15), [15, 20), [20, 25) e [25, 30].



Qual o valor da amplitude interquartílica, obtido por meio do método de interpolação linear dos dados agrupados em classes?

a) 15

b) $\frac{15}{2}$

c) $\frac{29}{5}$

d) $\frac{47}{7}$

e) 10

2. (CESGRANRIO/PETROBRAS/2018) Você dispõe de um montante para investir em ações e precisa decidir em que empresa(s) vai alocar esse montante. Três empresas lhe parecem interessantes, e você resolve consultar o desempenho delas nos últimos sessenta meses para minimizar possíveis riscos da sazonalidade no movimento da Bolsa de Valores. Os dados revelaram a seguinte distribuição, em %, das rentabilidades mensais das ações:

Medidas Estatísticas	Empresa A	Empresa B	Empresa C
Rentabilidade média mensal	0,50	0,60	0,40
Desvio padrão	1,00	1,20	0,80
Rentabilidade mínima	-1,80	-2,20	-1,20
Rentabilidade máxima	2,20	2,30	1,80
1º quartil	-0,20	-0,30	-0,10
3º quartil	0,80	0,90	0,70

A alocação dos recursos vai ser feita de acordo com a atitude conservadora de não investir em empresa com rentabilidade considerada outlier, entendendo como tal aquela que apresentar valor além de 1,5 desvio quartílico abaixo ou acima dos quartis 1 e 3.

Com base nesse critério, a escolha do investimento deve recair sobre a(s)

a) empresa A, apenas

b) empresa B, apenas

c) empresa C, apenas

d) empresas A e C, apenas

e) três empresas

3. (CESGRANRIO/EPE/2014) A Tabela a seguir apresenta a vazão média em cada mês para um determinado rio.

Mês	Vazão média mensal (m^3/s)
Janeiro	97
Fevereiro	60
Março	50
Abril	60
Maio	70

Junho	85
Julho	60
Agosto	50
Setembro	68
Outubro	117
Novembro	80
Dezembro	43

De acordo com os dados da Tabela, a mediana e a amplitude interquartílica das vazões valem, respectivamente,

a) 70 e 25

b) 64 e 25

C) 64 e 27,5

d) 68 e 27,5

e) 70 e 27,5

GABARITO – CESGRANRIO

Amplitude Interquartílica

1. LETRA D
2. LETRA C
3. LETRA C

LISTA DE QUESTÕES – CESGRANRIO

Variância

1. (CESGRANRIO/BB/2018) Uma amostra aleatória de tamanho 5 é retirada de uma população e observa-se que seus valores, quando postos em ordem crescente, obedecem a uma Progressão Aritmética.

Se a variância amostral não viciada vale 40, qual é o valor da razão da Progressão Aritmética?

a) 3

b) $5\sqrt{2}$

c) 4

d) $2\sqrt{5}$

e) 1

2. (CESGRANRIO/EPE/2014) Uma amostra de tamanho $200, x_1, x_2, \dots, x_{200}$, foi retirada de uma população, e seus valores foram transformados segundo a função $y_i = 4x_i - 1$ para $i = 1, 2, \dots, 200$.

Sabendo-se que a média e a variância dos dados transformados $y_1, y_2, \dots, y_{200}$ são, respectivamente, 3 e 16, os valores da média e da variância dos dados originais são, respectivamente,

a) 1 e 1

b) 1 e 4

c) 3/4 e 63

d) 11 e 64

e) 11 e 256

3. (CESGRANRIO/FINEP/2014) O enunciado a seguir deve ser usado para responder à questão.

Abaixo são apresentadas estatísticas das notas brutas obtidas pelos candidatos em um concurso público:

Média aritmética: 78

Variância: 100

A nota de cada candidato foi transformada em nota padronizada, calculada considerando-se a seguinte fórmula:

$$\textit{Nota padronizada} = 50 + 5 \times \frac{\textit{Nota bruta do candidato} - \textit{Media aritmetica das notas brutas}}{\textit{Desvio padrao das notas brutas}}$$

A variância das notas padronizadas é

a) 25

b) 50,5

c) 52,5

d) 55

e) 75

4. (CESGRANRIO/BNDES/2013) Em um departamento de uma empresa, o gerente decide dar um aumento a todos os empregados, dobrando o salário de todos eles.

Em relação às estatísticas dos novos salários, considere as afirmativas abaixo.

I - A média dobra.

II - A variância dobra.

III - A moda dobra.

É correto o que se afirma em

a) I, apenas

b) II, apenas

c) I e III, apenas

d) II e III, apenas

e) I, II e III

5. (CESGRANRIO/BNDES/2010) Em uma pesquisa de preços de determinado produto, foram obtidos os valores, em reais, de uma amostra aleatória colhida em 6 estabelecimentos que o comercializam.

Ação da Empresa	Resultado
P	5,00
Q	8,00
R	6,00
S	6,00
T	4,00

U	**7,00**

A variância dessa amostra é

a) 1,50

b) 1,75

c) 2,00

d) 2,25

e) 2,50

Gabarito – CESGRANRIO

Variância

1. LETRA C
2. LETRA A
3. LETRA A
4. LETRA C
5. LETRA C

LISTA DE QUESTÕES – CESGRANRIO

Desvio-Padrão

1. (CESGRANRIO/BACEN/2009) Considere as informações a seguir para responder à questão.

A viabilidade financeira do projeto de uma microempresa leva em consideração dados históricos de 100 projetos semelhantes.

A tabela abaixo mostra a distribuição de frequências do VPL - Valor Presente Líquido (valores em milhões de reais) de um conjunto de microempresas similares.

VPL	Frequência relativa
$-10 < x \leq 0$	10%
$0 < x \leq 10$	80%
$10 < x \leq 20$	10%

Segundo os dados históricos, o valor, em milhões de reais, que mais se aproxima do desvio padrão do VPL da microempresa é

a) 1

b) 2

c) 2,5

d) 4

e) 4,5

2. (CESGRANRIO/BNDES/2007) O enunciado abaixo refere-se à questão.

Um grupo é formado por 10 pessoas, cujas idades são:

17 19 19 20 20 20 20 21 22 22

Seja μ a média aritmética das idades e σ seu desvio padrão. O número de pessoas desse grupo cujas idades pertencem ao intervalo $[\mu - \sigma, \mu + \sigma]$ é

(Considere $\sqrt{2} = 1,4$)

a) 9

b) 8

c) 7

d) 6

e) 5

Gabarito – CESGRANRIO

Desvio-Padrão

1. LETRA E
2. LETRA C

LISTA DE QUESTÕES – CESGRANRIO

Coeficiente de Variação (ou Dispersão Relativa)

1. (CESGRANRIO/BB/2021) Um pesquisador recebeu os dados de uma amostra de tamanho 100 de uma população e calculou a média amostral μ, o desvio padrão amostral σ e o coeficiente de variação amostral $CV = \frac{\sigma}{\mu}$. Antes de iniciar a análise, ele foi informado de que os dados dessa amostra estavam todos errados, mas que podiam ser corrigidos somando-se 3 a cada um dos dados que recebeu.

Após fazer tal correção, o valor do coeficiente de variação amostral passou a ser

a) $\frac{3\sigma}{\mu+3}$

b) $\frac{300\sigma}{\mu+300}$

c) $\frac{\sigma}{\mu+3}$

d) $\frac{\sigma}{\mu+300}$

e) $\frac{\sigma}{\mu+0{,}03}$

2. (CESGRANRIO/BB/2018) Há dez anos a média das idades, em anos completos, de um grupo de 526 pessoas era de 30 anos, com desvio padrão de 8 anos.

Considerando-se que todas as pessoas desse grupo estão vivas, o quociente entre o desvio padrão e a média das idades, em anos completos, hoje, é

a) 0,45

b) 0,42

c) 0,20

d) 0,27

e) 0,34

3. (CESGRANRIO/LIQUIGÁS/2018) Uma amostra $(x_1, x_2, ..., x_{10})$ de tamanho 10 de uma população nos forneceu os seguintes valores

$$\sum_{i=1}^{10} x_i = 20 \ e \ \sum_{i=1}^{10} x_i^2 = 140.$$

Qual o valor do coeficiente de variação amostral?

a) $\frac{7}{20}$

b) $\frac{5}{3}$

c) $\frac{1}{7}$

d) $\frac{1}{6}$

e) 6

4. (CESGRANRIO/EPE/2014) Um pesquisador está interessado em comparar a variabilidade de duas variáveis com médias diferentes e desvios padrões diferentes, presentes num dado estudo estatístico.

A medida estatística adequada a ser usada nesse contexto é a(o)

a) covariância

b) diferença entre a maior e a menor variância

c) razão entre o maior e o menor desvios padrões

d) coeficiente de variação

e) coeficiente de correlação

5. (CESGRANRIO/IBGE/2014) Os dados a seguir foram obtidos de empregados de uma empresa com três fábricas: I, II e III. A variável de interesse é salário.

Empresa	Média	Desvio padrão
Fábrica I	1185	630,49
Fábrica II	600	355,97
Fábrica III	2150	1106,16

Comparando-se a variabilidade de salários em relação ao salário médio das três fábricas, através de seus coeficientes de variação, conclui-se que a variabilidade da fábrica

a) I é menor apenas do que a da fábrica III.

b) II é menor apenas do que a da fábrica I.

c) II é menor apenas do que a da fábrica III.

d) II é menor do que as das outras duas fábricas.

e) III é menor do que as das outras duas fábricas.

6. (CESGRANRIO/IBGE/2013) Seja X uma variável aleatória com distribuição normal cuja média é μ e o desvio padrão é σ.

Se Y = 2X - 1 tem distribuição normal com média 5 e variância 20, o coeficiente de variação populacional $\frac{\sigma}{\mu}$ vale

a) $\frac{\sqrt{42}}{6}$

b) $\frac{\sqrt{21}}{6}$

c) $\frac{\sqrt{5}}{3}$

d) $\frac{\sqrt{39}}{9}$

e) $\frac{4\sqrt{5}}{9}$

7. (CESGRANRIO/BB/2013) A variância de um conjunto de dados é 4 m^2.

Para o mesmo conjunto de dados foram tomadas mais duas medidas de variabilidade:

a diferença entre o terceiro e o primeiro quartil e o coeficiente de variação.

Esses dois valores caracterizam-se, respectivamente, por

a) possuírem unidades de medida m^2 e m.

b) possuírem unidades de medida m e m^2.

c) ser adimensional e possuir unidade de medida m^2.

d) possuir unidade de medida m e ser adimensional.

e) possuir unidade de medida m^2 e ser adimensional.

8. (CESGRANRIO/BNDES/2013) Quatro variáveis são utilizadas em um modelo de previsão da quantidade produzida de uma determinada commodity agrícola. São elas:

• temperatura, em graus Celsius

• quantidade de fertilizante, em toneladas

• variação dos preços praticados no mercado internacional, em %

• quantidade produzida de um produto similar, em toneladas

Para determinar qual dessas variáveis apresenta a maior variabilidade, deve-se utilizar

a) apenas a média de cada uma das variáveis

b) apenas a variância de cada uma das variáveis

c) apenas o desvio padrão de cada uma das variáveis

d) a relação $\frac{desvio\ padrão}{média}$ de cada uma das variáveis

e) a relação $\frac{variância}{média}$ de cada uma das variáveis

9. (CESGRANRIO/BACEN/2009) Considere as informações a seguir para responder à questão.

A viabilidade financeira do projeto de uma microempresa leva em consideração dados históricos de 100 projetos semelhantes.

A tabela abaixo mostra a distribuição de frequências do VPL - Valor Presente Líquido (valores em milhões de reais) de um conjunto de microempresas similares.

VPL	Frequência relativa
$-10 < x \leq 0$	10%
$0 < x \leq 10$	80%
$10 < x \leq 20$	10%

Um projeto alternativo para o investidor apresenta um VPL esperado, em reais, de 6 milhões e um risco (desvio padrão) de 2 milhões. Pela ótica do risco relativo, qual o melhor investimento, a microempresa ou o projeto alternativo?

a) A microempresa, pois apresenta um Coeficiente de Variação maior.

b) A microempresa, pois apresenta um Coeficiente de Variação menor.

c) O projeto alternativo, pois apresenta um Coeficiente de Variação maior.

d) O projeto alternativo, pois apresenta um Coeficiente de Variação menor.

e) É indiferente, pois os investimentos apresentam Coeficientes de Variação iguais.

Gabarito – CESGRANRIO

Coeficiente de Variação (ou Dispersão Relativa)

1. LETRA C
2. LETRA C
3. LETRA B
4. LETRA D
5. LETRA E
6. LETRA C
7. LETRA D
8. LETRA D
9. LETRA D

LISTA DE QUESTÕES – INÉDITAS

Medidas de Dispersão

1. (INÉDITA/2022) Sobre medidas de dispersão, julgue os itens:

I – as medidas de dispersão são métricas que mostram a variação dos dados de um conjunto;

II – são medidas de dispersão: amplitude total; amplitude interquartílica; desvio médio; média; variância relativa e mediana.

III – as medidas de dispersão podem ser divididas em dois grupos: medidas de dispersão absoluta e medidas de dispersão relativa.

IV – a média, moda, mediana e o desvio padrão são medidas de posição.

V – a coeficiente de variação e a variância relativa são medidas de dispersão relativa.

Estão corretas as afirmativas:

a) Apenas I. e III.

b) Apenas I. III. e V.

c) Apenas V.

d) Apenas II. e IV.

e) Apenas IV. e V.

2. (INÉDITA/2022) Com relação às medidas de tendência central, medidas de posição e medidas de dispersão, assinale a alternativa correta:

a) Dentre as medidas de posição mais utilizadas, temos: a média, mediana, moda e o desvio padrão, sendo a mais importante delas a média aritmética.

b) Sendo a variância uma medida de dispersão, ao dividirmos todos os valores de uma variável por uma constante, a variância do novo conjunto fica dividida pelo quadrado dessa constante.

c) O coeficiente de variação é uma medida de dispersão absoluta que fornece a variação dos dados em relação à média.

d) Sendo a variância uma medida de posição, ao subtrairmos uma constante de todos os valores de uma variável, a variância do conjunto não é alterada.

e) O desvio médio faz parte das medidas de dispersão, ao somarmos uma constante a todos os valores de uma variável, o desvio médio do conjunto também será somado dessa constante.

GABARITO – INÉDITAS

Medidas de Dispersão

1. LETRA B
2. LETRA B

LISTA DE QUESTÕES – INÉDITAS

Amplitude Total

1. (INÉDITA/2022) Uma pesquisa foi realizada em uma escola para avaliar o rendimento dos alunos do terceiro ano. Nessa avaliação foram atribuídos valores de acordo com o rendimento dos alunos nos testes aplicados pelos professores. A tabela de frequências abaixo mostra o resultado da pesquisa:

x_i	nº de alunos
3	4
6	10
8	17
10	13
15	6
Total	**50**

De acordo com as informações da tabela a amplitude total é igual a:

a) 8

b) 10

c) 12

d) 4

e) 9

GABARITO – INÉDITAS

Amplitude Total

1. LETRA C

LISTA DE QUESTÕES – INÉDITAS

Amplitude Interquartílica

1. (INÉDITA/2022) Denominamos de quartis os valores de uma série que a dividem em quatro partes iguais, isto é, quatro partes contendo o mesmo número de elementos (25%). A respeito desse assunto, julgue as assertivas e marque a opção correta:

( ) A amplitude interquartílica é o resultado da subtração entre o terceiro quartil e o primeiro quartil, sendo expressa pela seguinte equação: $A_{IQ} = Q_3 - Q_1$

( ) Ao somar ou subtrair uma constante c a/de todos os valores de uma variável, a amplitude interquartílica do conjunto também será somada dessa constante.

( ) O segundo quartil corresponde à separação de metade dos elementos da série, porém, esse valor não coincide com a mediana.

( ) Ao multiplicar ou dividir todos os valores de uma variável por uma constante c, a amplitude interquartílica do conjunto também será multiplicada ou dividida por essa constante.

a) V – V – F – V

b) V – V – F – F

c) V – F – F – V

d) F – F – V – V

e) F – V – V – F

2. (INÉDITA/2022) Considerando a distribuição de frequências abaixo marque a alternativa correta.

A distribuição dos salários em mil reais dos 100 funcionários de uma empresa está apresentada na tabela abaixo:

Salários (x_i)	Frequência absoluta
$2000 \vdash 3000$	8
$3000 \vdash 4000$	20
$4000 \vdash 5000$	34
$5000 \vdash 6000$	26

$6000 \vdash 7000$	12
Total	**100**

A amplitude interquartílica, dada pela diferença entre o terceiro e o primeiro quartil, é igual a:

a) R$ 2300,00

b) R$ 3500,00

c) R$ 1900,00

d) R$ 1650,00

e) R$ 1800,00

GABARITO – INÉDITAS

Amplitude Interquartílica

1. LETRA C
2. LETRA D

LISTA DE QUESTÕES – INÉDITAS

Desvio Absoluto Médio

1. (INÉDITA/2022) Em um curso preparatório para concursos, foi realizada uma pesquisa para identificar as idades dos seus alunos. As idades estão distribuídas conforme a tabela de frequências abaixo:

idades (x_i)	Nº de alunos (f_i)
$20 \vdash 25$	150
$25 \vdash 30$	380
$30 \vdash 35$	420
$35 \vdash 40$	150
Total	**1200**

Com base na tabela, o desvio absoluto médio das idades dos alunos é:

a) 5,3

b) 6,2

c) 2,65

d) 4,8

e) 3,62

2. (INÉDITA/2022) Em um campeonato de basquete, a comissão organizadora fez um levantamento de quantas cestas foram feitas por cada jogador. O resultado é apresentado na tabela de frequências abaixo:

Nº de cestas (x_i)	Nº de jogadores (f_i)
11	5
14	6
23	4

25	3
Total	**18**

O desvio absoluto médio do número de cestas é de:

a) 5,33

b) 6,78

c) 5,2

d) 4,89

e) 3,55

GABARITO – INÉDITAS

Desvio Absoluto Médio

1. LETRA E

2. LETRA A

LISTA DE QUESTÕES – INÉDITAS

Desvio-Padrão

1. (INÉDITA/2022) Em uma empresa de departamento, com 10 funcionários na área de vendas, as quantidades vendidas por cada funcionário no mês foram:

13	14	17	17	17	19	19	21	22	23

O desvio padrão das vendas foi igual a:

a) 4,01

b) 3,09

c) 3,50

d) 2,89

e) 5,2

GABARITO – INÉDITAS

Desvio-Padrão

1. LETRA B

LISTA DE QUESTÕES – INÉDITAS

Variância Relativa

1. (INÉDITA/2022) Seja uma amostra de candidatos que prestaram concurso para os cargos de assistente e analista. Considere que a média de pontos da prova de analista é de 86 pontos, com desvio padrão de 36; e a média de pontos da prova de assistentes é de 95 pontos, com desvio padrão de 25. A variância relativa de analistas supera a de assistentes em:

a) 0,1

b) 0,5

c) 0,7

d) 0,17

e) 0,3

2. (INÉDITA/2022) Considere uma população X com coeficiente de variação 0,3 e média aritmética igual a 35. E uma população Y com desvio padrão 25 e média aritmética igual a 49. A variância relativa da população X mais a variância relativa da população Y é igual a:

a) 0,25

b) 0,40

c) 0,55

d) 0,28

e) 0,35

3. (INÉDITA/2022) Considere uma amostra de tamanho 15, com 270 sendo o somatório dos termos e que tem coeficiente de variação igual a 0,5. A variância relativa dessa amostra é igual a:

a) 0,18

b) 0,50

c) 0,75

d) 0,35

e) 0,25

GABARITO – INÉDITAS

Variância Relativa

1. LETRA A **2.** LETRA E **3.** LETRA E

ESSA LEI TODO MUNDO CONHECE:
PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.

1. Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.

2. Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).

3. Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).

4. Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).

5. Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).

6. Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).

7. Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.

8. O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.

Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.