Cruzeiro do Sul
Educacional

Medidas de Dispersão ou Variação

Conteudista
Prof.ª Dra. Rosangela Maura C. Bonici

Revisão Textual
Juliane Khun

Sumário

Objetivos da Unidade

- Trabalhar com as medidas de dispersão;
- Aprender como calcular a variância e o desvio padrão de dados brutos, de dados agrupados em distribuições de frequência variável discreta e de dados agrupados em distribuições de frequência variável contínua;
- Compreender também qual o significado do desvio padrão e em quais situações práticas ele poderá ser empregado.

Atenção, estudante! Aqui, reforçamos o acesso ao conteúdo *on-line* para que você assista à videoaula. Será muito importante para o entendimento do conteúdo.

Este arquivo PDF contém o mesmo conteúdo visto *on-line*. Sua disponibilização é para consulta *off-line* e possibilidade de impressão. No entanto, recomendamos que acesse o conteúdo *on-line* para melhor aproveitamento.

—

VOCÊ SABE RESPONDER?

Para darmos início à unidade, contudo, convidamos você a refletir acerca da questão: Você sabe o que é desvio padrão?

Contextualização

Entenda o que é Média e Desvio Padrão de uma Prova

O entendimento natural que grande parte dos candidatos utiliza para avaliar seus desempenhos nas provas é: "acertei mais ou menos questões do que a média?". Claro que a premissa vigente é a de que os candidatos mais bem preparados superam, em número de acertos, a média das provas.

O resultado de uma prova, normalmente, é conhecido por meio de informações como média e desvio padrão, bem como pela distribuição de frequências do número de acertos dos candidatos, expressos em forma gráfica.

> Esse gráfico, denominado histograma, mostra, no eixo dos X (abscissas), o número de questões e, no eixo dos Y, o número de candidatos que acertaram o referido número de questões.

O que significam essas informações?

Numa distribuição de frequências, há três medidas importantes: a moda, a mediana e a média. A primeira é o "pico", isto é, o ponto no eixo das abscissas de maior frequência.

A mediana é o ponto, no eixo das abscissas, que divide as ocorrências em duas frações iguais, cada uma com 50% das frequências.

A média é o ponto, no eixo das abscissas, que faria com que o gráfico ficasse equilibrado, não inclinando nem para a esquerda nem para a direita; em suma, a média é o ponto, no eixo das abscissas, situado na vertical que passa pelo centro de gravidade da figura.

O que se deseja em uma prova do Concurso Vestibular é um histograma formando uma curva simétrica, distribuído entre 0 e 25 acertos, concentrando moda, mediana e média próximas a 15 acertos, exibindo distribuição balanceada de acertos, tanto à esquerda quanto à direita do centro de distribuição, de acordo com o gráfico apresentado na Figura da página a seguir.

O Gráfico mostra uma distribuição normal rigorosamente simétrica. No centro da distribuição, coincidem média, mediana e moda. Uma curva de distribuição normal (ou Curva de Gauss) tem como característica englobar 99,73% das ocorrências no intervalo compreendido entre a média e ± 3 desvios padrão, conforme detalhado no mesmo Gráfico.

O desvio padrão de uma prova mede o grau de dispersão dos candidatos em relação à média, isto é, o quanto o conjunto de candidatos se distanciou da média, tanto além como aquém do centro de distribuição.

Isso significa que os escores obtidos por 99,73% dos candidatos estarão compreendidos entre a média e ± 3 desvios padrão, ou seja, salvo raras exceções, todos os candidatos estarão neste intervalo.

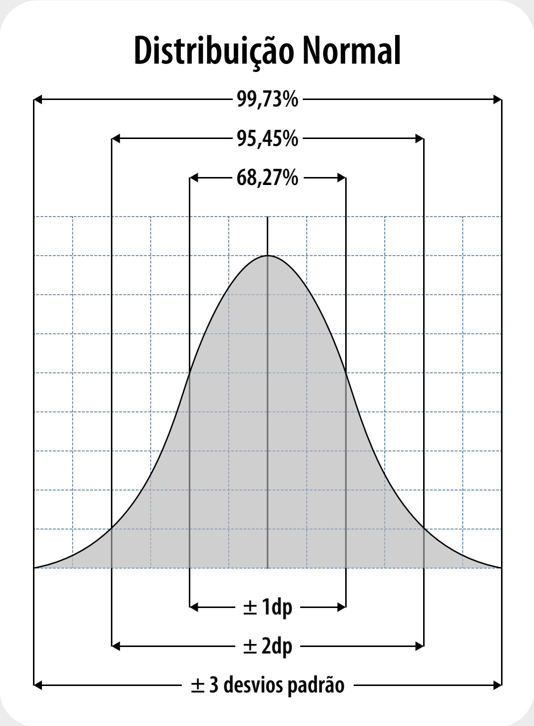

Figura 1 – Distribuição Normal
Fonte: Acervo do conteudista

#ParaTodosVerem: gráfico em que se demonstram as seguintes informações: Distribuição normal; 99,73% que compreende +/- 3 desvios padrão; 95,45%, compreendendo +/- 2 desvios padrão; 68,27%, com +/- 1 desvio padrão. As informações estão indicadas por setas pretas, que se sobrepõem a uma linha que tem movimento ascendente e depois decrescente. Fim da descrição.

Ao se elaborar uma prova, espera-se que o resultado da aplicação dessa prova gere uma "curva de distribuição normal", isto é, essa prova deve gerar uma média de 15 acertos e os candidatos devem estar distribuídos simetricamente entre zero e 30 acertos (ou entre dois limites internos desse intervalo, equidistantes de 15).

A obtenção de uma distribuição simétrica com 100% das ocorrências entre 0 e 30 acertos, em uma prova de 30 questões e média de 15 acertos, pode ser possível quando se obtém um desvio padrão de 5 acertos. Nesse caso, se o histograma assumir formato semelhante ao da curva normal, todos os escores possíveis de serem obtidos pelos candidatos ficariam simetricamente contidos no intervalo entre a média e ± 3 desvios padrão, ou seja, entre 0 acertos (15 - (3 x 5)) e 30 acertos (15 + (3 x 5)).

Infelizmente, não é fácil obter uma curva de distribuição normal. O resultado obtido pelos candidatos em uma prova depende de muitos fatores, entre os quais pode ser destacada a preparação deles e o grau de dificuldade da prova. Por isso mesmo, é importante poder avaliar uma prova pelo resultado obtido na sua aplicação.

A informação da média permite verificar o grau de facilidade da prova para a população que a realizou. Quanto menor a média (abaixo de 15 acertos), menor a facilidade dos candidatos com as questões. Quanto maior a média (acima dos 15 acertos), maior a facilidade dos candidatos com as questões propostas.

Uma prova, com histograma normal, com média de 15 acertos e desvio padrão de 4 acertos significa que 99,73% dos candidatos serão encontrados entre os escores 3 e 27 acertos (15 - (3 x 4)) e (15 + (3 x 4)). Isso significa que os candidatos com escores 0, 1, 2, 28, 29 e 30 acertos serão 0,23% da população, portanto, pouquíssimos, conforme mostrado na figura a seguir, correspondente à distribuição com desvio 4.

Enquanto que numa prova, também com histograma normal, com média de 15 acertos, mas desvio padrão de 2 acertos, significa que 99,73% dos candidatos serão encontrados entre os escores 9 e 21 acertos (15 - (3 x 2)) e (15 + (3 x 2)). Isto é, também haverá poucos candidatos (0,23%) com escores de 0 até 8 e de 22 até 30 acertos, conforme mostrado na figura correspondente à distribuição com desvio 2.

Analisando as duas Figuras que seguem, é possível concluir que quanto maior for o desvio padrão, mais aberta é a curva (maior dispersão), ou seja, maior variedade de escores obtidos pelos candidatos e melhores condições de discriminar a qualificação dos candidatos.

Curvas muito fechadas (pequeno desvio padrão) significam menor dispersão, ou seja, grande concentração de escores e menor variedade deles (muitos empates). Em outras palavras, se houver muitos empates, a prova poderá não avaliar devidamente a preparação dos candidatos. Ao mesmo tempo, provas muito difíceis não diferenciam escores obtidos unicamente por meio de acerto casual.

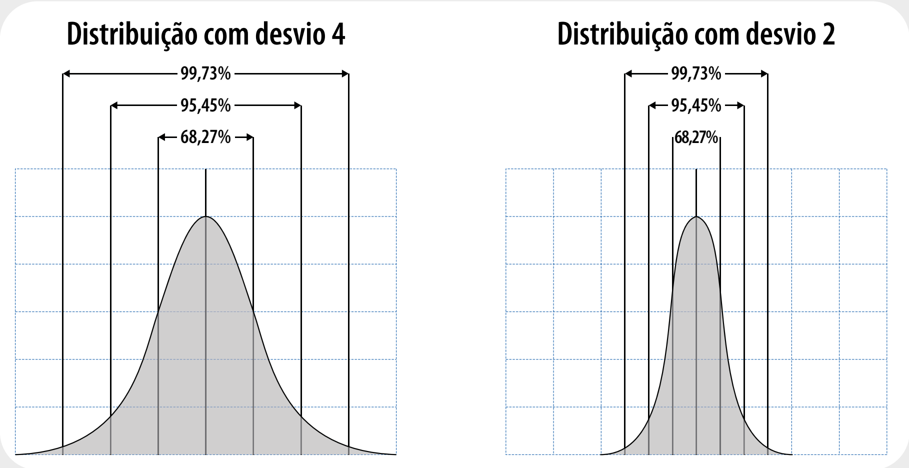

Figura 2 – Distribuição com desvio 4 e 2

Fonte: Vestibular da UFRGS 2005

#ParaTodosVerem: dois gráficos em que se demonstram as seguintes informações: à esquerda: Distribuição com desvio 4; 99,73%; 95,45%; 68,27%. À direita: distribuição com desvio 2; 99,73%; 95,45%; 68,27%. As informações estão indicadas por setas pretas, que se sobrepõem a uma linha que tem movimento ascendente e depois decrescente. Fim da descrição.

Introdução

As medidas de variação ou dispersão avaliam a dispersão ou a variabilidade da sequência numérica em análise. São medidas que fornecem informações complementares à informação da média aritmética. As principais medidas de dispersão são a variância e o desvio padrão.

Usaremos as letras s^2 para denotar a variância de uma amostra e s para denotar o seu desvio padrão.

Cálculo da Variância e Desvio Padrão

Para calcular a variância e o desvio padrão, vamos analisar três casos:

I. Quando os dados ainda não foram agrupados em tabelas de frequência, ou seja, estão na forma de dados brutos ou rol;

II. Quando os dados estão agrupados em distribuições de frequência variável discreta; e

III. Quando os dados estão agrupados em distribuições de frequência variável contínua.

Dados Brutos ou Rol

Para podermos calcular a variância e o desvio padrão de dados brutos, vamos usar as fórmulas a seguir.

Fórmula para o Cálculo da Variância de Dados Brutos

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

- s^2 = Variância;
- x_i = Cada um dos valores assumidos pela variável;
- x = Média aritmética dos dados brutos;
- n = Total de elementos observados.

Fórmula para o Cálculo do Desvio padrão de Dados Brutos

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

- s = Desvio-padrão;
- s^2 = Variância.

Vejamos um exemplo de utilização da variância e desvio padrão. Calcule a variância e o desvio padrão das notas de três turmas de estudantes.

Quadro 1 – Notas de estudantes das Turmas A, B e C

Turma	Nota dos alunos								Média	Desvio padrão
A	4	5	5	6	6	7	7	8	6	1,31
B	1	2	4	6	6	9	10	10	6	3,51
C	0	6	7	7	7	7,5	7,5		6	2,69

Observe no Quadro que a média e o desvio padrão das notas já estão calculados.

Vamos ver como isso foi feito.

O desvio padrão da turma A foi calculado da seguinte forma:

1º) Determinar a média aritmética das notas, pois a variância depende dela. Como são dados brutos, vamos relembrar a fórmula para o cálculo da média.

$$\bar{x} = \frac{\sum x_i}{n}$$

Usando as notas da turma A para fazer os cálculos, temos:

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

$$\bar{x} = \frac{48}{8} = 6$$

Concluímos que a média aritmética das notas vale 6.

2º) Vamos calcular a variância das notas da turma A. Para isso, vamos usar a fórmula para o cálculo da variância de dados brutos, que é:

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

Vamos entender o que a fórmula está dizendo...

$\sum (x_i - \bar{x})^2$ (faça a diferença entre cada nota e a média aritmética e eleve ao quadrado; depois, some cada uma dessas diferenças).

Depois, divida o valor que encontrou pelo total de notas menos 1.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n \quad 1} = s^2 = \frac{(4-6)^2 + (5-6)^2 + (5-6)^2 + (6-6)^2 + (6-6)^2 + (7-6)^2 + (7-6)^2 + (8-6)^2}{8-1}$$

$$= \frac{4+1+1+0+0+1+1+4}{7} = \frac{12}{7} \cong 1,71$$

Temos que a variância das notas vale 1,71.

3º) Vamos calcular o desvio padrão das notas usando a fórmula:

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

Substituindo a variância na fórmula e fazendo o cálculo, temos:

$$s \quad \sqrt{1,71} = 1,31$$

Temos que o desvio padrão vale 1,31.

Para calcular o desvio padrão das turmas B e C, procedeu-se da mesma forma.

Considerações

Tabela 1 – Notas de estudantes das Turmas A, B e C

Turma	Nota dos alunos								Média	Desvio padrão
A	4	5	5	6	6	7	7	8	6	1,31
B	1	2	4	6	6	9	10	10	6	3,51
C	0	6	7	7	7	7,5	7,5		6	2,69

Observando a Tabela 1, podemos fazer as seguintes considerações:

- As notas que geraram média 6 nas três turmas são bastante diferentes;
- Os desvios padrão são bem diferentes. O menor está na turma A, o intermediário na turma C e o maior na turma B;
- O desvio padrão nos mostra a variabilidade dos dados em relação à média. Grosso modo, dizemos que o desvio padrão nos mostra se a média aritmética sofreu pouca ou muita influência dos valores extremos (muito grandes ou muito pequenos). Nesse caso, podemos afirmar que:
    - A turma A foi a menos influenciada por valores extremos;
    - A turma C foi medianamente influenciada por valores extremos;
    - A turma B foi a mais influenciada por valores extremos.

Distribuição de Frequência Variável Discreta

Para calcular a variância e o desvio padrão de uma distribuição de frequência variável discreta, vamos usar as fórmulas a seguir:

Fórmula para o cálculo da variância da distribuição de frequência variável discreta

$$s^2 = \frac{\sum (x_i - \underline{x})^2 . f_i}{\sum f_i - 1}$$

- s^2 = Variância;
- x_i = Cada um dos valores assumidos pela variável;
- f_i = frequência absoluta;
- $\underline{x}$ = Média aritmética da variável discreta;
- $\sum f_i - 1$ = Soma do total de elementos observados menos 1.

Fórmula para o cálculo do desvio padrão da distribuição de frequência variável discreta

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

- s = Desvio-padrão;
- s^2 = Variância.

Vejamos um exemplo.

A Tabela 2 representa as notas de Matemática. Calcule a variância e o desvio padrão.

Tabela 2 – Notas de Matemática

Notas de Matemática (xi)	fi
2	3
3	5

Notas de Matemática (xi)	fi
4	8
5	4
Totais	**20**

As notas de Matemática estão agrupadas em uma distribuição de frequência variável discreta. Para calcular a variância e o desvio padrão, temos de usar as fórmulas correspondentes.

1° Vamos calcular a variância usando a fórmula.

$$s^2 = \frac{\sum (x_i - \bar{x})^2 . f_i}{\sum f_i - 1}$$

Vamos entender o que ela significa

$\sum (x_i - \bar{x})^2 . f_i$ = devemos subtrair cada nota da média aritmética. Esse resultado deve ser elevado ao quadrado. Depois, deve ser multiplicado pela respectiva frequência.

Ao final, fazer o somatório desses valores

$\sum f_i - 1$ = somar o total de notas e subtrair 1.

Primeiro, devemos calcular a média aritmética, para podermos depois usar a fórmula da variância.

Lembra-se da fórmula da média aritmética ponderada? É ela que iremos usar!

$$\bar{X} = \frac{\sum x_i . f_i}{\sum f_i}$$

Vamos usar a distribuição das notas de Matemática e abrir uma coluna para podermos multiplicar x_i por f_i e calcular a média.

Tabela 3 – Notas de Matemática – média aritmética ponderada

Notas de Matemática (x_i)	f_i	$x_i.f_i$
2	3	2.3 = 6
3	5	3.5 = 15
4	8	4.8 = 32
5	4	5.4 = 20
Totais	**20**	**73**

Calculando a média, temos:

$$\overline{X} = \frac{73}{20} = 3,65$$

A média aritmética das notas de Matemática é **3,65**.

Vamos calcular agora a Variância, usando a fórmula.

Para podermos fazer $\sum(x_i - \overline{x})^2.f_i$, vamos abrir uma nova coluna na distribuição de frequência das notas de Matemática, para poder facilitar nossos cálculos.

Tabela 4 – Notas de Matemática – Variância

Notas de Matemática (x_i)	f_i	$(x_i - \overline{x})^2 \cdot f$
2	3	$(2-3,65)^2 \cdot 3 = 8,17$
3	5	$(3-3,65)^2 \cdot 5 = 2,11$
4	8	$(4-3,65)^2 \cdot 8 = 0,987$
5	4	$(5-3,65)^2 \cdot 4 = 7,29$
Totais	**20**	**18,55**

Concluímos daí que $\sum(x_i - \bar{x})^2.f_i$ vale 18,55, completando a resolução.

$$\sum f_i - 1 = 20 - 1 = 19$$

Calculando, temos $\frac{18,55}{19} = 0,98$

A variância das notas de Matemática vale 0,98.

2° Vamos calcular o desvio padrão usando a fórmula.

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

- s = Desvio-padrão;
- s^2 = Variância.

$$s = \sqrt{0,98} = 0,99 \quad \text{(desvio padrão)}$$

Considerações

Podemos concluir pelos cálculos que o desvio padrão vale 0,99, o que nos demonstra uma variabilidade pequena nas notas de Matemática.

Cálculo da Variância e do Desvio Padrão da Distribuição de Frequência Variável Contínua

Para calcular a variância e o desvio padrão de variáveis contínuas, devemos proceder como para as variáveis discretas, tomando somente o cuidado de substituir o **xi pelos pontos médios de cada classe**, uma vez que a variável está agrupada com intervalos de classe.

Fórmula para o cálculo da variância da distribuição de frequência variável contínua

$$= \frac{\sum (x_i - \bar{x})^2 . f_i}{\sum f_i - 1}$$

- s^2 = Variância;
- x_i = Cada um dos valores assumidos pela variável;
- f_i = frequência absoluta;
- x = Média aritmética da variável discreta;
- $\sum f_i - 1$ = Soma do total de elementos observados menos 1.

Fórmula para o cálculo do desvio padrão da distribuição de frequência variável discreta

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

- s = Desvio-padrão;
- s^2 = Variância.

Vamos ver um exemplo.

A Tabela 5 representa um banco de horas de uma pequena empresa. Calcule a variância e o desvio padrão.

Tabela 5 – Banco de horas dos empregados de uma empresa

Banco de horas (h)	f_i
0 |- 4	1
4 |- 8	3
8 |- 12	5
12 |- 16	1
Total	**10**

1º) Para calcular a variância, a primeira coisa que temos de conhecer é a média aritmética desse banco de horas, caso contrário, não há como usar a fórmula da variância.

Importante

Lembra-se da fórmula da média aritmética ponderada? É ela que iremos usar!

$$\overline{X} = \frac{\sum x_i . f_i}{\sum f_i}$$

Na variável contínua, para podermos calcular a média, temos de fazer aparecer os x_i, calculando o ponto médio entre cada uma das horas. Para isso, vamos abrir uma coluna para distribuição, para colocar o ponto médio, e outra, para podermos multiplicar x_i por f_i.

Tabela 6 – Banco de horas dos empregados de uma empresa – Distribuição

Banco de horas (h)	f_i	x_i (ponto médio)	$x_i \cdot f_i$
0 |- 4	1	$\frac{0+4}{2}=2$	2 . 1 = 2
4 |- 8	3	$\frac{4+8}{2}=6$	3 . 6 = 8
8 |- 12	5	$\frac{8+12}{2}=10$	5 . 10 = 50
12 |- 16	1	$\frac{12+16}{2}=14$	1 . 14 = 14
Total	**10**	**-**	**84**

$$\overline{X} = \frac{\sum x_i . f_i}{\sum f_i} \Rightarrow \overline{X} = \frac{84}{10} = 8,4$$

Temos que a média do banco de horas é 8,4h.

Agora sim, estamos em condições de calcular a variância.

$$s^2 = \frac{\sum (x_i - \bar{x})^2 . f_i}{\sum f_i - 1}$$

Vamos usar a distribuição e abrir uma coluna para podermos calcular.

$$\sum (x_i - \bar{x})_2 . f_i$$

Tabela 7 – Banco de horas dos empregados de uma empresa – Variância

Banco de horas (h)	f_i	x_i (ponto médio)	$(x_i - \underline{x})^2 \cdot f_i$
0 |- 4	1	2	$(2 - 8{,}4)^2 \cdot 1 = 40{,}96$
4 |- 8	3	6	$(6 - 8{,}4)^2 \cdot 3 = 17{,}28$
8 |- 12	5	10	$(10 - 8{,}4)^2 \cdot 5 = 12{,}80$
12 |- 16	1	14	$(14 - 8{,}4)^2 \cdot 1 = 31{,}36$
Total	**10**	**-**	**102,4**

Temos que $\sum (x_i - \bar{x})^2 . f = 102{,}4_i$ e $\sum f_i - 1 = 10 - 1 = 9$

Aplicando os valores na fórmula, vem:

$$s^2 = \frac{\sum (x_i - \bar{x})^2 . f_i}{\sum f_i - 1} \Rightarrow s^2 = \frac{102{,}4}{9} = 11{,}38$$

Chegamos à conclusão de que a variância vale 11,38.

2° Agora vamos calcular o desvio padrão usando $s = \sqrt{s^2}$

Substituindo os valores, temos: $s = \sqrt{11{,}38} = 3{,}37$

Considerações

Feitos os cálculos, verificamos que a variância do banco de horas é 3,37, o que demonstra uma variabilidade média nas horas.

Importante

Quanto maior o desvio padrão, maior a variação ou dispersão dos dados.

Quanto menor o desvio padrão, menor a variação ou dispersão dos dados.

Em Síntese

Finalizamos mais uma Unidade na qual aprendemos a calcular a variância, calcular e interpretar o desvio padrão.

Como vimos, o desvio padrão fornece informações que complementam a informação da média aritmética, mostrando se a variação dos dados que geraram a média aritmética é pequena, média ou grande.

Só conseguimos identificar se um desvio padrão é pequeno ou grande se tivermos dois conjuntos que tenham médias iguais para podermos comparar seus desvios padrão.

Estou confiante e tenho certeza de que vocês conseguiram acompanhar e que estão satisfeitos por terem conseguido vencer mais esta etapa.

Agradeço a todos, continuem se esforçando sempre e até a próxima!

Atividades de Fixação

1. O que a variância representa em estatística?

a) A média dos valores em um conjunto de dados.
b) A medida da dispersão ou espalhamento dos dados.
c) A soma de todos os valores em um conjunto de dados.
d) O valor mais frequente em um conjunto de dados.
e) A mediana dos valores em um conjunto de dados.

2. Qual das seguintes afirmações descreve melhor o desvio padrão em estatística?

a) O valor mais frequente em um conjunto de dados.
b) A medida da tendência central em um conjunto de dados.
c) Uma medida de dispersão que indica o grau de variabilidade em um conjunto de dados.
d) A soma de todos os valores em um conjunto de dados.
e) A mediana dos valores em um conjunto de dados.

Atenção, estudante! Veja o gabarito desta atividade de fixação no fim deste conteúdo.

Material Complementar

Sites

UOL Educação
Este *site* oferece uma gama diversificada de conteúdos educativos, desde notícias sobre educação até artigos especializados em diferentes disciplinas. Se você está em busca de informações confiáveis, recursos didáticos, dicas de estudo ou novidades no campo da educação, o UOL Educação é o seu destino ideal.
http://bit.ly/49rZphM

ADVFN
Com uma vasta gama de informações, análises, notícias atualizadas e ferramentas para análise de mercado, é o local ideal para investidores e interessados em compreender melhor as nuances do mercado financeiro.
https://bit.ly/49ggyv8

Leitura

Variância e Desvio Padrão
O *site* oferece uma ampla variedade de conteúdos educacionais, cobrindo diversas áreas do saber, desde ciências naturais até humanas, passando por disciplinas acadêmicas e curiosidades do mundo.
http://bit.ly/49gvIjY

Vídeo

Média, Mediana e Desvio Padrão – Professora Angela – Matemática
Este vídeo oferece uma explicação acessível, passo a passo, para ajudar na compreensão desses conceitos fundamentais.
https://youtu.be/5-TERv1ky_c

Referências

CRESPO A. A. **Estatística Fácil**. 11. ed. São Paulo: Saraiva, 1994.

DOWNING, D. **Estatística Aplicada**. 2. ed. São Paulo: Saraiva, 2002.

MORETTIN, L. G. **Estatística Básica**. 7. ed. São Paulo: Pearson, 2000.

NEUFELD, J. L. **Estatística Aplicada à Administração Usando o Excel**. São Paulo: Pearson, 2003.

SPIEGEL, M. R. **Estatística**. 3. ed. São Paulo: Pearson, 1994. (Coleção Schaum).

SPIEGEL, M. R. **Probabilidade e Estatística**. São Paulo: Pearson, 1977. (Coleção Schaum).

SILVA, E. M. **Estatística para os Cursos de Economia, Administração e Ciências Contábeis**. 3. ed. São Paulo: Atlas, 1999.

Gabarito

Questão 1

b) A medida da dispersão ou espalhamento dos dados.

Justificativa: A variância é uma medida estatística que representa o grau de dispersão ou espalhamento dos dados em um conjunto. Ela indica o quão distantes os pontos de dados estão da média (ou valor esperado) desse conjunto. Quanto maior a variância, maior a dispersão dos dados, e quanto menor a variância, mais próximos os dados estão da média. Portanto, a variância é uma medida fundamental para compreender a variabilidade em um conjunto de dados, o que a torna a resposta correta para a questão.

Questão 2

c) Uma medida de dispersão que indica o grau de variabilidade em um conjunto de dados.

Justificativa: O desvio padrão é uma medida estatística que quantifica o grau de dispersão ou variabilidade dos dados em um conjunto. Ele indica o quanto os valores se afastam da média (tendência central) de um conjunto de dados. Quanto maior o desvio padrão, maior a dispersão dos dados; quanto menor o desvio padrão, menos dispersão os dados apresentam.