Estatística aplicada: Medidas de Dispersão

Apresentação

A estatística está relacionada aos métodos científicos para coleta, organização, resumo, apresentação e análise de dados. A função principal da estatística é ajudar na tomada de decisões, dando subsídios a interpretação de dados fruto de pesquisas, experimentos, a própria realidade de uma amostra, um contexto, uma população. Estas informações produzidas na estatística, ajuda-nos a conhecer o objeto de estudo, não sendo um fim, más meio.

Nesta Unidade de Aprendizagem você conhecerá as medidas de dispersão que constituem um ponto importante da análise estatística de dados obtidos por medição.

Bons estudos.

Ao final desta Unidade de Aprendizagem, você deve apresentar os seguintes aprendizados:

- Reconhecer medidas de dispersão.
- Diferenciar medidas de dispersão e medidas de tendência central.
- Calcular a amplitude, desvio padrão, variância e coeficiente de variação de um conjunto de dados.

Desafio

Joaquim é o encarregado da produção e responsável pela qualidade dos produtos. Na figura vemos ele tirando as medidas de peça metálica. A peça foi dobrada e deve atender medidas previamente definidas para evitar problemas de montagem que pode ser montada encaixada a juntas em curvas, ponteiras, sistemas de fixação, etc. Desta maneira ter um processo confiável é necessário para se ter produtos conformes.

Agora você foi solicitado(a) para certificar através das medições feitas na largura - medida básica de verificação se está conforme ou não conforme - utilizando um paquímetro, obtendo 5 leituras das quais foram:

- M1: 53,32 mm
- M2: 53,33 mm
- M3: 53,33 mm
- M4: 53,31 mm
- M5: 53,34 mm

Sabendo que a dispersão neste caso é um dado importante, calcule a amplitude e o desvio padrão e explique qual dos dois resultados você usaria e justifique a utilização deste resultado.

Infográfico

Tratamos sobre as medidas de dispersão, no entanto, um fator muito importante relacionado com as medidas de dispersão é o ato de realizar uma medida, seja ela uma medida de temperatura, pressão, força, massa, tensão, corrente, etc. Todas estas medidas possuem incertezas, algumas inerentes ao processo de medida e outras associadas aos instrumentos de medição. No Brasil é o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) quem garante que estas medidas ou essas incertezas estão corretas.

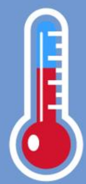
A seguir você vai conhecer um pouco mais a fundo a respeito deste laboratório, como ele opera em âmbito nacional e como os laboratórios associados ao INMETRO prestam serviços a comunidade.

INMETRO

O INMETRO é uma autarquia (parte da administração pública) que tem por objetivo fortalecer as empresas nacionais, aumentando a sua produtividade por meio da adoção de mecanismos destinados à melhoria da qualidade e da segurança de produtos e serviços.

É o órgão do governo responsável pela qualidade dos produtos produzidos no país e pelos importados.

A qualidade aqui referida está na medição das variáveis que compõem o processo de produção, como, por exemplo, a calibração de instrumentos de medição de temperatura (termômetros, pirômetros, termopares, etc.).



O INMETRO é o laboratório de referência do Brasil. No entanto, apenas o INMETRO não é capaz de atender toda a demanda de calibração do país.

Para isso, existe a Rede Brasileira de Calibração (RBC), que congrega competências técnicas e capacitações vinculadas às indústrias, universidades e institutos tecnológicos, habilitados à realização de serviços de calibração.

A forma com a qual cada laboratório de calibração expressa a sua qualidade está nas incertezas associadas aos seus padrões. Tais valores podem ser pesquisados no site do INMETRO.

Normalmente, os laboratórios apresentam as faixas de calibração e as suas correspondentes incertezas de medição.

Por exemplo, um determinado laboratório RBC apresenta em seu escopo (conjunto de atividades) o serviço de Termorresistência, e as calibrações podem ser feitas na faixa de -30 a 200 °C com incertezas de 0,20 a 0,38 °C.

Assim, essas incertezas carregam informações calculadas a partir das medidas de dispersão estudadas nesta unidade.

Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Conteúdo do livro

Inserir introduçãotilizam-se as medidas de dispersão para indicar o grau de variação dos elementos de um conjunto numérico em relação a sua média. Conhecer o grau de variabilidade de um conjunto de dados torna a análise mais confiável já que muitas vezes as medidas de tendência central, como a média, a moda e a mediana não deixam evidente a homogeneidade ou não dos dados. Constantemente nos deparamos com situações em que dados nos são apresentados, sejam em pesquisas eleitorais, na formulação e teste de um novo medicamento, na disseminação e casos de morte por uma doença, no crescimento populacional ou situação da economia em determinado setor, entre milhares de outras. Para todas estas e diversas outras situações problema, interessa conhecer o comportamento do conjunto de dados, e este é o papel da estatística.

No capítulo Estatística aplicada: Medidas de Dispersão, você conhecerá as quatro principais medidas de dispersão, aprofundando cada uma delas e diferenciando-as das medidas de tendência central, além de estabelecer relações destas com problemas aplicados. Serão fornecidos subsídios para que você amplie o conhecimento e perceba a relação entre as definições e os conceitos teóricos das medidas estudadas com a sua aplicabilidade.

Boa leitura.

METROLOGIA

sagah
SOLUÇÕES
EDUCACIONAIS
INTEGRADAS

Estatística aplicada: medidas de dispersão

Cristiane da Silva

OBJETIVOS DE APRENDIZAGEM

> Reconhecer medidas de dispersão.
> Diferenciar medidas de dispersão e medidas de tendência central.
> Calcular amplitude, desvio-padrão, variância e coeficiente de variação de um conjunto de dados.

Introdução

Cada vez mais recebemos uma gama de informações advindas de meios de comunicação, mídias sociais, leituras, publicidade, etc. E assim como essa informação é importante para nos inserirmos no meio em que vivemos, também se torna fundamental desenvolver a habilidade de lê-la. Ou seja, o domínio de métodos quantitativos como as medidas de dispersão e sua interpretação é essencial para que tenhamos uma visão mais clara dos fatos e dados observados.

Nesse contexto, espera-se, por meio deste capítulo, apresentar elementos que fundamentem as interpretações de dados que forem analisados nas mais diversas áreas do conhecimento, permitindo que o leitor conheça formas de cálculo simples, práticas e relevantes para o seu cotidiano.

Neste capítulo, você vai conhecer as medidas de dispersão. Inicia-se apresentando os conceitos de dispersão, variação e amplitude de forma mais abrangente; na segunda seção, aprofundam-se as definições, diferenciando, ainda, as medidas de dispersão das medidas de tendência central, e destacam-se as características de cada uma delas. Por fim, evidencia-se como calcular algumas das principais medidas de dispersão por meio de um *software* livre.

A variabilidade de um conjunto de dados

Sempre que iniciamos uma pesquisa, nos interessa conhecer a variabilidade do conjunto de dados em análise. Isso porque detectar variabilidades nos leva a buscar explicações para elas — fazer conexões, tentar compreender o que está causando esse efeito de variabilidade. Para ficar mais claro, pense em uma pandemia, como a da Covid-19. O que causou a variabilidade no número de mortes pela doença? Será que foi o comportamento das pessoas? A elaboração de vacinas e o avanço do processo de vacinação? Foram as mutações da doença? As políticas públicas adotadas ou não? Houve alguma relação com o sexo, a idade, a escolaridade, a renda ou a região em que reside a população? Bom, aqui parece haver alguns elementos relevantes para tentar compreender a variabilidade observada, bem como o comportamento da doença ao longo do tempo. Becker (2015) destaca que se não houver variabilidade em um sentido absoluto, não é possível detectar determinado fenômeno.

Algumas das medidas importantes de variação são a amplitude, o desvio-padrão, a variância e o coeficiente de variação. Para além de encontrar seus valores numéricos, é fundamental desenvolver a habilidade de interpretar e compreender esses valores (TRIOLA, 2017). Uma ilustração visual da variação pode ser observada na Figura 1.

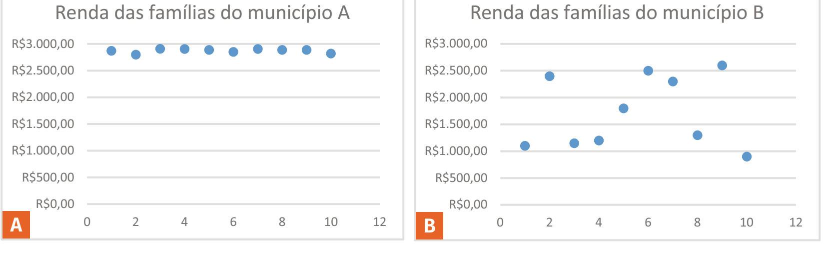


Figura 1. Gráficos de dispersão da renda das famílias.

Na Figura 1a são apresentadas as rendas de dez famílias do município A e, na Figura 1b, as rendas de dez famílias do município B. É possível observar que a renda das famílias do município B tem mais variação do que a renda das famílias do município A. O gráfico à direita mostra mais espalhamento do que o gráfico à esquerda. Essa característica de espalhamento — ou variação, ou dispersão — é tão importante que a medimos com números.

As medidas de dispersão indicam o grau de variabilidade dos dados observados. Em outras palavras, o quanto os valores observados são próximos ou distantes entre si. A amplitude total é a mais simples dessas medidas, pois ela é a diferença entre o maior e o menor valor observado (KUYVEN, 2010). Um exemplo dessa medida pode ser a amplitude térmica, a variação que costuma ocorrer ao longo de um dia na temperatura.

O desvio-padrão é a medida de dispersão mais utilizada, porque leva em consideração a totalidade dos valores da variável em estudo. O desvio-padrão considera os desvios em torno da média aritmética, e a sua fórmula básica pode ser expressa como a raiz quadrada da média aritmética dos quadrados dos desvios (KUYVEN, 2010). É importante mencionar que as fórmulas de cálculo são diferentes (em seu denominador) quando se trata de desvio-padrão amostral ou populacional. Triola (2017) explica que o desvio-padrão é uma medida de quanto os valores de dados se afastam da média.

A variância, por sua vez, é o desvio-padrão elevado ao quadrado. É uma medida de pouca utilidade para descrever um conjunto de dados, mas é muito importante na inferência estatística e em combinações de amostras (KUYVEN, 2010; TRIOLA, 2017). Conforme Triola (2017), a variância tem a desvantagem de usar unidades diferentes das unidades dos dados originais, o que torna difícil entender como ela se relaciona com o conjunto de dados original.

Para compreender melhor o desvio-padrão, é importante destacar que, ao se medir a variação em um conjunto de dados amostrais, faz sentido começar com as quantidades individuais pelas quais cada valor se afasta da média. Para um valor particular x, a quantidade de desvio é $x - \bar{x}$, que é a diferença entre o valor individual x e a média. Para obter uma estatística que meça a variação, é preciso evitar o cancelamento de números positivos e negativos. Sendo assim, deve-se somar os valores absolutos como em $\Sigma|x - \bar{x}|$. Se for encontrada a média correspondente a essa soma, obtêm-se o desvio médio absoluto, que é a distância média dos dados até a média (TRIOLA, 2017).

No entanto, como o desvio médio absoluto requer o uso de valores absolutos, ele usa uma operação que não é algébrica, o que cria dificuldades nos métodos de inferência estatística. Outro fator relevante é que ele é um estimador viesado, o que significa dizer que, quando encontrados os desvios médios absolutos para amostras não se tende a atingir o desvio médio absoluto da população. Além disso, como o desvio-padrão baseia-se em uma raiz quadrada de uma soma de quadrados, se assemelha às fórmulas de distância encontradas na álgebra, o que é mais vantajoso (TRIOLA, 2017).

Outra medida que merece destaque é a amplitude semi-interquartílica, também conhecida como desvio quartílico, que é indicada por Q e é definida por:

$$Q = \frac{Q_3 - Q_1}{2}$$

em que Q_1 e Q_3 são o primeiro e o terceiro quartis. A amplitude interquartílica $Q_3 - Q_1$ é usada algumas vezes, mas a amplitude semi-interquartílica é mais comum como medida de dispersão (SPIEGEL; STEPHENS, 2009).

Já a amplitude entre os percentis 10 e 90 é definida por $P_{90} - P_{10}$, em que P_{10} e P_{90} são o 10° e 90° percentis referentes aos dados. A semiamplitude percentílica entre 10 e 90, $\frac{1}{2}(P_{90} - P_{10})$, também pode ser usada, mas normalmente isso não ocorre (SPIEGEL; STEPHENS, 2009).

Nesse contexto, a regra empírica ajuda a interpretar o valor de um desvio-padrão. Conforme Triola (2017), essa regra estabelece que, para conjuntos de dados que tenham uma distribuição aproximadamente normal, aplicam-se as seguintes propriedades: i) cerca de 68% de todos os valores ficam a até um desvio-padrão da média; ii) cerca de 95% de todos os valores ficam a até dois desvios-padrão da média; e iii) cerca de 99,7% de todos os valores ficam a até três desvios-padrão da média, como mostra a Figura 2.

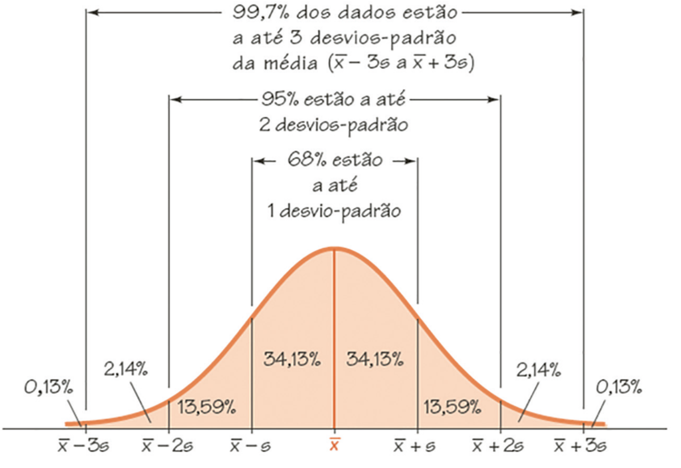


Figura 2. Regra empírica.
Fonte: Adaptada de Triola (2017).

Cabe ressaltar que quando comparada a variação em dois conjuntos de dados diferentes, os desvios-padrão só devem ser comparados se os dois conjuntos de dados usarem as mesmas escala e unidades de medida e tiverem, aproximadamente, a mesma média. Quando as médias forem diferentes ou as amostras usarem escalas ou unidades de medida diferentes, utiliza-se o coeficiente de variação (TRIOLA, 2017). Além de contornar esses problemas, para caracterizar a dispersão ou a variabilidade dos dados em termos relativos ao seu valor médio, utiliza-se o coeficiente de variação.

Saiba mais

Você pode saber mais sobre as medidas de dispersão consultando o Capítulo 4 da obra *Estatística*, de Spiegel e Stephens (2009).

Nesta seção, você conheceu os conceitos de variação, dispersão, amplitude entre percentis 10 e 90 e a lógica do cálculo do desvio-padrão. As ideias apresentadas buscaram estabelecer conexão com o cotidiano por meio de situações com as quais nos deparamos. A seguir, você aprofundará o estudo ao diferenciar as medidas de tendência central das medidas de dispersão.

Medidas de tendência central e de dispersão

Para conhecer as fórmulas matemáticas das medidas de tendência central e de dispersão, é importante primeiramente diferenciá-las. Conforme Kuyven (2010), as medidas de tendência central indicam o padrão de respostas de uma variável e mostram o valor que se localiza no meio de um conjunto de dados ordenados. Já as medidas de dispersão informam o quão distantes estão os dados uns dos outros.

Dentre as medidas de tendência central que apresentaremos estão a média, a mediana e a moda. Dentre as medidas de dispersão estão a amplitude total, o desvio-padrão, a variância e o coeficiente de variação.

Conforme Larson e Farber (2015), a média de um conjunto de dados é dada pela soma dos valores dos dados dividida pelo número de observações. Downing e Clark (2010) apresentam a fórmula para o cálculo da média, considerando n números x_1, x_2, x_3, ..., x_n. A média amostral é expressa pelo símbolo $\bar{X}$ (enquanto a média populacional é expressa por μ) e tem a seguinte representação matemática:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n} \qquad (1)$$

Já a mediana é o ponto, ou elemento, a meio caminho dos dados, ou seja, metade dos números está acima dela e metade abaixo. Um procedimento importante para o cálculo da mediana é ordenar a lista de valores (DOWNING; CLARK, 2010). Nesse contexto, Kuyven (2010) define a mediana como um valor que está no meio dos dados quando o conjunto de dados está ordenado. A mediana indica o centro de um conjunto de dados ordenado, de modo que, dividido em duas partes com quantidades iguais de valores, tem-se o resultado da mediana. Havendo no conjunto de dados um número ímpar de observações, a mediana é o elemento do meio; havendo um número par de observações, a mediana é a média dos dois elementos que ocupam as posições centrais.

Existem duas fórmulas para encontrar a posição da mediana, que são muito úteis quando o conjunto de dados é extenso. Observe:

Se *n* for ímpar:

$$Me = X_{\left(\frac{n+1}{2}\right)} \tag{2}$$

Se *n* for par:

$$Me = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} \tag{3}$$

onde *n* é o tamanho da amostra e *X* é a variável de interesse no estudo.

A moda é caracterizada pelo valor que ocorre com maior frequência. Havendo mais de um valor nessas condições, todos eles serão denominados modas. Em outras palavras, uma distribuição pode ter mais do que uma moda; no caso em que houver duas modas, esta será chamada de distribuição bimodal (DOWNING; CLARK, 2010; LARSON; FARBER, 2015).

A amplitude total é uma medida de dispersão. O termo "dispersão" indica o grau de afastamento de um conjunto de números em relação à sua média. Uma das maneiras de medir a dispersão consiste simplesmente em fazer a diferença entre o maior e o menor valor. Essa grandeza é denominada amplitude (DOWNING; CLARK, 2010).

O desvio-padrão mede a variação dos dados com relação à média, tendo a mesma unidade de medida que o conjunto de dados. Ele é sempre maior ou igual a zero; nesse caso, igual a zero significa que o conjunto de dados não apresenta variação, ou seja, todos os elementos têm o mesmo valor. À medida que os valores se afastam da média, mais dispersos são os dados, portanto, maior será o desvio-padrão (LARSON; FARBER, 2015). A fórmula do desvio-padrão amostral é dada por:

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} \tag{4}$$

Dito de outra maneira, essa fórmula indica a realização dos seguintes passos.

1. Calcular a média do conjunto de dados amostrais.
2. Subtrair a média de cada valor de *x*.
3. Elevar ao quadrado cada resultado do passo 2.
4. Somar todos os resultados do passo 3.

5. Dividir a soma do passo 4 por (n – 1), quando os dados forem amostrais. Observação: para dados populacionais divide-se a soma por N.
6. Extrair a raiz quadrada do resultado da soma.

Fique atento

O cálculo do desvio-padrão amostral difere do desvio-padrão populacional. No caso do desvio-padrão amostral, depois de encontrar todos os valores individuais $(X - \bar{X})^2$, eles são combinados de acordo com sua soma e, a seguir, divididos por n – 1, porque há apenas n – 1 valores independentes. Dada uma média de n elementos, apenas n – 1 deles podem ser associados a qualquer número antes que o último valor seja determinado. A divisão por n – 1 faz com que a variância amostral s^2 tenda para o valor da variância populacional σ^2, ao passo que a divisão apenas por N (como é o caso do desvio-padrão populacional) faz com que a variância amostral s^2 subestime a variância populacional σ^2 (TRIOLA, 2017).

A variância é o desvio-padrão elevado ao quadrado, sendo simbolizada por σ^2 quando se refere à população e por s^2 quando se refere à amostra. A variância tem pouca utilidade como estatística descritiva, porém é muito importante na inferência estatística e em combinações de amostras. A fórmula da variância amostral é dada por:

$$S^2 = \frac{\sum(X - \bar{X})^2}{n - 1} \tag{5}$$

O coeficiente de variação é a razão entre o desvio-padrão e a média referentes a dados de uma mesma série. Sua fórmula, em percentual, é dada por:

$$CV = \frac{\sigma}{\mu} \cdot 100 \text{ para populações} \tag{6}$$

$$CV = \frac{S}{\bar{X}} \cdot 100 \text{ para amostras} \tag{7}$$

Nesta seção, vimos a diferença entre medida de tendência central e medida de dispersão. Conhecemos as principais delas em seus conceitos, fórmulas e pontos de atenção. A seguir, aprofundaremos o estudo das medidas de dispersão especialmente no que se refere ao cálculo da amplitude, do desvio-padrão, da variância e do coeficiente de variação de um conjunto de dados.

Conhecendo o comportamento dos dados

Como vimos, para conhecer o comportamento de um conjunto de dados fazemos uso das medidas estatísticas. Nosso enfoque aqui será para as medidas de dispersão: amplitude, desvio-padrão, variância e coeficiente de variação. Os dados utilizados são fictícios e referem-se a uma amostra da renda de dez famílias dos municípios A e B. O Quadro 1 apresenta o conjunto de dados.

Quadro 1. Renda das famílias

Renda das famílias do município A	Renda das famílias do município B
R$2.870,00	R$1.100,00
R$2.800,00	R$2.400,00
R$2.910,00	R$1.150,00
R$2.905,00	R$1.200,00
R$2.890,00	R$1.800,00
R$2.850,00	R$2.500,00
R$2.905,00	R$2.300,00
R$2.890,00	R$1.300,00
R$2.890,00	R$2.600,00
R$2.820,00	R$900,00

Para calcular as medidas de dispersão, será utilizado o Excel. Como a amplitude total não possui uma fórmula específica para o seu cálculo no Excel, buscamos, por meio das funções do Excel, quais são os valores máximo e mínimo do conjunto de dados. Para encontrar o valor máximo utiliza-se a função = MÁXIMO para cada conjunto de dados separadamente, ou seja, selecionam-se os dados da renda das famílias do município A e depois os dados da renda das famílias do município B, como mostra a Figura 3.

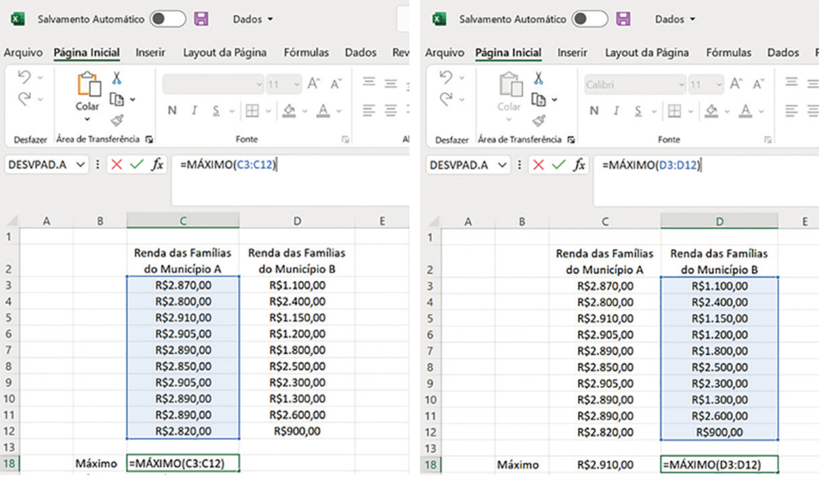


Figura 3. Encontrando o valor máximo dos conjuntos de dados.

O cálculo do valor mínimo é análogo, ou seja, utiliza-se a função = MÍNIMO para cada conjunto de dados separadamente, como mostra a Figura 4.

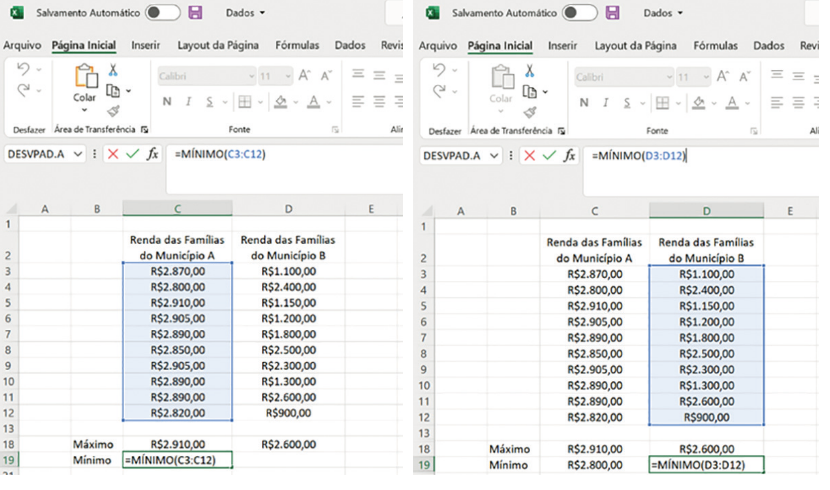


Figura 4. Encontrando o valor mínimo dos conjuntos de dados.

Agora é possível calcular a amplitude dos dados, que será a diferença entre o maior valor do conjunto de dados e o menor valor do mesmo conjunto, como mostra a Figura 5.

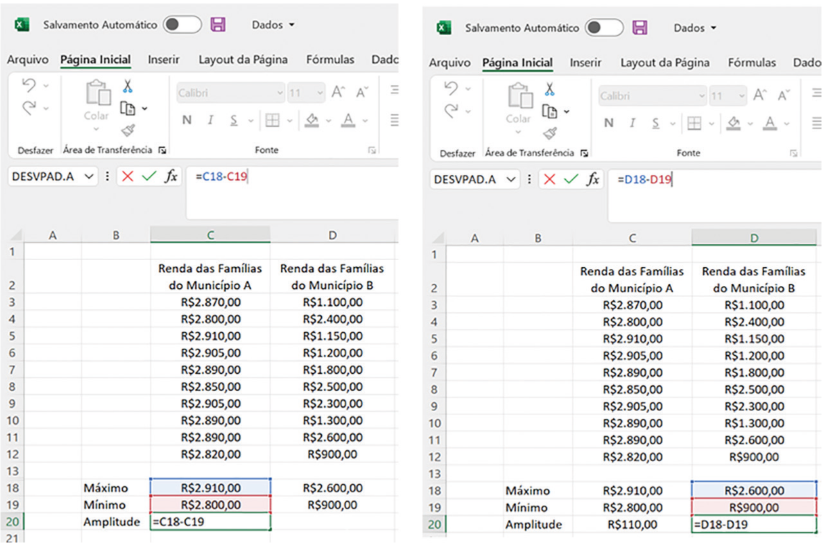


Figura 5. Calculando a amplitude total dos conjuntos de dados.

Com isso, constata-se que a amplitude da renda das famílias do município A — que é igual a R$ 110,00 — é menor do que a observada no município B — que é de R$ 1.700,00. Isso significa dizer que existe uma grande variação no município B comparativamente ao A; em outras palavras, os valores das rendas das famílias do município B são mais distantes umas das outras.

Seguimos com o cálculo do desvio-padrão para cada conjunto de dados (A e B). O cálculo do desvio-padrão é realizado utilizando a função = DESVPAD.A, o que indica que será calculado o desvio-padrão amostral; isso está indicado por “(.A)” na fórmula. Quando o estudo envolve dados de toda a população, deve-se escolher a opção DESVPAD.P. Observe a Figura 6.

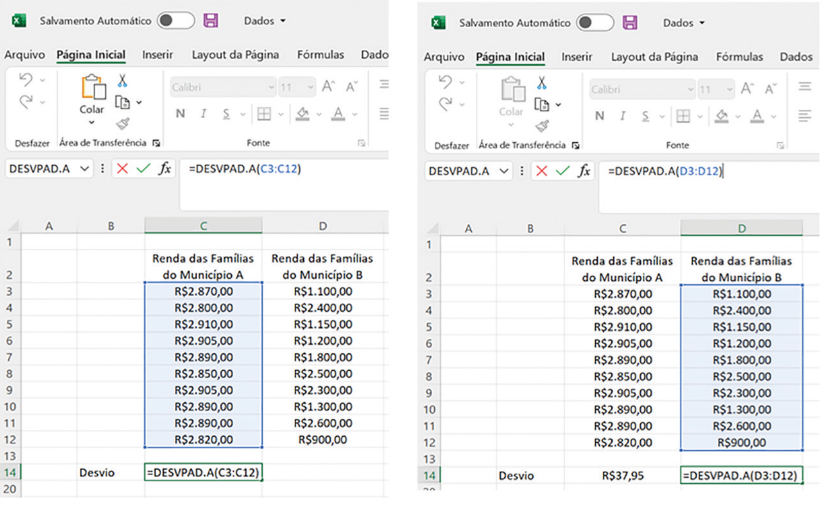


Figura 6. Calculando o desvio-padrão dos conjuntos de dados.

O desvio-padrão nos permite observar que há um afastamento de R$ 37,95 na renda das famílias do município A em relação à média e um afastamento de R$ 668,02 na renda das famílias do município B em relação à média.

A variância é encontrada utilizando-se a função = VAR.A, o que indica que será calculada a variância amostral, indicado por "(.A)" na fórmula. Quando o estudo envolve dados de toda a população, deve-se escolher a opção VAR.P. Observe a Figura 7.

A interpretação da variância absoluta não tem muito sentido prático, por se tratar de uma medida quadrática. Isso pode ser observado pelos resultados retornados pelo Excel, sendo R$ 1.440,00 a variância da renda das famílias do município A e R$ 446.250,00 a variância da renda das famílias do município B.

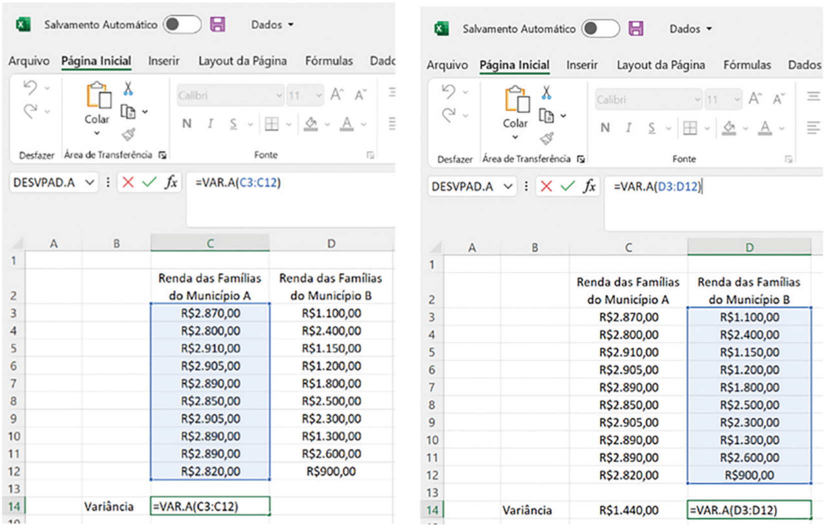


Figura 7. Calculando a variância dos conjuntos de dados.

Assim como a amplitude total, o coeficiente de variação também não possui uma fórmula específica para o seu cálculo no Excel, mas é sabido que ele é calculado pela divisão do desvio-padrão pela média do conjunto de dados e, por fim, multiplicado por 100 para ter a resposta em percentual, como mostra a Figura 8.

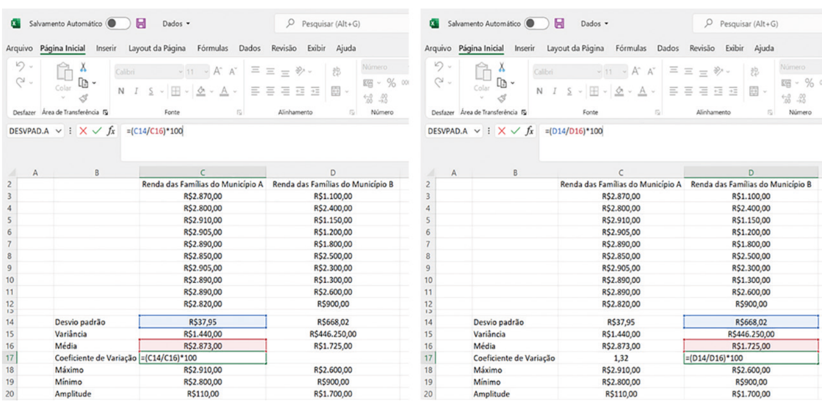


Figura 8. Calculando o coeficiente de variação dos conjuntos de dados.

O coeficiente de variação analisa a dispersão em termos relativos, por isso é dado em percentual. Quanto menor for o valor do coeficiente de variação, mais homogêneos serão os dados, ou seja, menor será a dispersão em torno da média. Analisando a renda das famílias do município A, verifica-se que há baixa dispersão nas rendas; os dados são homogêneos, pois o valor retornado foi de 1,32% para o CV. Já no caso da renda das famílias do município B, observa-se alta dispersão, e os dados são heterogêneos, pois o valor retornado foi de 38,73% para o CV.

Saiba mais

Você pode saber mais sobre outro *software* utilizado para calcular as medidas de dispersão consultando o Capítulo 3 da obra *Método quantitativo com o uso de software*, de Raupp (2012).

Nesta seção, você percebeu a aplicabilidade dos conceitos teóricos estudados nas duas primeiras seções, especialmente no que se refere às medidas de dispersão. Pode, assim, constatar que é possível realizar os cálculos de forma mais eficiente e rápida por meio do Excel, bem como identificar a interpretação prática das medidas estatísticas.

Esta obra apresentou técnicas importantes para a atividade profissional nas mais diversas áreas do conhecimento, pois a estatística está presente em tudo o que nos acompanha diariamente. Para compreender o mundo que nos cerca precisamos entender o que as informações que nos chegam estão efetivamente expressando. É claro que o assunto não se esgota aqui, por isso sugere-se a continuidade dos estudos.

Referências

BECKER, J. L. *Estatística básica*: transformando dados em informação. Porto Alegre: Bookman, 2015. (Série Métodos de Pesquisa). *E-book*.

DOWNING, D.; CLARK, J. *Estatística aplicada*. 2. ed. São Paulo: Saraiva, 2010.

KUYVEN, P. S. *Raciocínio lógico e métodos quantitativos*. São Leopoldo: Unisinos, 2010.

LARSON, R.; FARBER, B. *Estatística aplicada*. São Paulo: Pearson, 2015.

SPIEGEL, M. R.; STEPHENS, L. J. *Estatística*. 4. ed. Porto Alegre: Bookman, 2009. (Coleção Schaum). *E-book*.

TRIOLA, M. F. *Introdução à estatística*. 12. ed. Rio de Janeiro: LTC, 2017.

Leitura recomendada

RAUPP, C. A. F. *Método quantitativo com o uso de software*. São Leopoldo: Unisinos, 2012.

Conteúdo:

SOLUÇÕES EDUCACIONAIS INTEGRADAS

Dica do professor

A Estatística no sentido mais restrito, é destinada a obtenção de dados, derivados de uma amostra, a exemplo de médias, medianas, amplitudes, desvio padrão da população toda, desvio padrão da amostra da população, dentre outros. Assim a estatística é aplicada desde medidas de controle em peças, doenças em populações, dados financeiros, tendo infinitas aplicações.

No vídeo da Dica do Professor, você vai ter a introdução, passando pelas medidas de centro e de dispersão, aprenderá a diferenciá-las e também verá como são calculadas.

Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Exercícios

1) **O grau para o qual os dados numéricos tendem a dispersar-se em torno de um valor médio é denominado de variação ou dispersão dos dados. Existem várias medidas de dispersão ou variação, sendo as mais comuns:**

A) Média, moda e mediana.

B) Amplitude, média e desvio padrão.

C) Amplitude, variância e desvio padrão.

D) Variância, coeficiente de variação e mediana.

E) Moda, desvio médio e desvio padrão.

2) **Quanto as medidas de dispersão, pode-se dizer que as afirmações a seguir:**
I) A variável que mede o desvio em relação à média, em unidades de desvio padrão, é denominada variável padronizada e é uma quantidade adimensional (ou seja, independe das unidades usadas).
II) Às vezes, o desvio padrão correspondente aos dados de uma amostra é definido com (N – 3), em lugar de N nos denominadores das Equações, porque o valor resultante representa uma estimativa melhor do desvio padrão da população da qual a amostra foi extraída.
III) Dentre outras medidas de dispersão ou variação, podemos encontrar: o desvio médio, a semi-amplitude interquartílica e a amplitude entre os percentis 10 e 90.

A) Todas estão corretas.

B) Somente I está correta.

C) I e III estão corretas.

D) I e II estão corretas.

E) Somente II está correta.

3) **Um pesquisador fez uma pesquisa de altura em uma turma do ensino primário, composta por 10 alunos, e obteve os seguintes resultados em centímetros: 99, 99, 103, 107, 106, 107, 99, 102, 100, 107. Qual é a amplitude das alturas dos alunos dessa sala?**

A) 6.

B) 107.

C) 8.

D) 102,5.

E) 7.

4) **Qual o desvio padrão amostral do conjunto B: { 1,1,3,4,5,1,3 } .**

A) 2,57.

B) 4.

C) 1,49.

D) 5.

E) 1,61.

5) **A metrologia é uma ciência que trabalha aspectos tanto teóricos quanto a prática da medição e também os princípios de incertezas envolvidos. Em que atividades metrológicas são aplicadas as medidas de dispersão?**

A) Estimativa de incerteza de medição somente.

B) Estimativa da incerteza de medição, leitura de instrumentos, estudos de confiabilidade metrológica.

C) Estimativa de incerteza de medição, estudos de confiabilidade metrológica e no controle estatístico de processo.

D) Calibração, rastreabilidade e leitura de instrumentos.

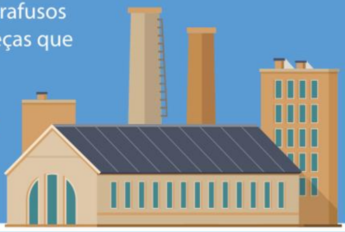
E) Leitura de paquímetro, leitura de micrômetro e leitura de relógio comparador.

Na prática

A seguir, você irá ver de forma prática como as medidas de dispersão podem auxiliar no controle de qualidade de uma empresa. Será apresentada de forma prática a implementação das medidas de dispersão mais utilizadas; a amplitude, a variância e o desvio padrão além da média. Suponha então uma empresa que fabrique parafusos. Uma medida importante para garantir a qualidade do produto fabricado é a medida de comprimento. Para esta empresa é importante garantir que o tamanho dos seus parafusos não varie fora de seus valores aceitáveis.

Vamos supor que uma empresa que produz parafusos quer fazer uma análise inicial em um lote de peças que acabaram de ser produzidas.

Neste caso, o lote foi de um parafuso de 1,0cm de comprimento. Foram tomadas 20 medidas e anotadas na tabela a seguir:



Comprimento do parafuso em cm	
0,9912	1,0100
0,9874	0,9822
0,9757	0,9932
1,0149	1,0073
1,0090	1,0092
1,0220	1,0028
0,9974	1,0000
0,9891	0,9906
1,0126	0,9955
0,9942	0,9985

Foram feitas três análises de medidas diferentes. Veja a seguir as suas características e exemplos:

A amplitude nos dados é dada pela diferença entre o maior valor observado (1,0220) e o menor (0,9757):

$$A = X_{max} - X_{min} = 1,0220 - 0,9757 = 0,0463cm$$

Segundo o manual da qualidade da empresa, se a amplitude de um lote de parafusos de 1,0cm for maior do que 0,0500cm, o lote é descartado.

Neste exemplo, a amplitude foi de 0,0463cm. Logo, neste primeiro teste, o lote foi aprovado.

Agora, vamos observar a variância dos dados, que é a média dos quadrados das diferenças entre cada uma das medidas e a média aritmética do conjunto de amostras.

$$VAR = \frac{\sum_{i=1}^{20} x_i - \bar{x}}{20}$$

VAR = 0,000139cm²
Uma baixa variância indica que os dados são homogêneos e têm pouca variação entre as medidas.

Outra medida que pode ser feita é com relação ao desvio padrão dos comprimentos. O desvio padrão é dado pela raiz quadrada da variância:

$$Desvio\ padrão = \sqrt{VAR}$$

Dp = 0,01178cm

O valor médio dos comprimentos dos parafusos será: $\bar{x}$=0,9991±0,0117cm

Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Saiba mais

Para ampliar o seu conhecimento a respeito desse assunto, veja abaixo as sugestões do professor:

Aplicação da Análise de Variância na Implantação do CEP

Este artigo apresenta uma aplicação da Análise de Variância (ANOVA) na identificação das fontes de variabilidade em um processo de laminação de barras de aço, utilizado em uma empresa metalmecânica sediada no Rio Grande do Sul. O estudo das possíveis fontes de variabilidade serviu de base para o estabelecimento do controle estatístico do processo (CEP), o qual é descrito neste trabalho juntamente com os respectivos estudos de estabilidade e capacidade.

Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Medidas de dispersão: os valores estão próximos entre si ou variam muito?

Neste artigo, os autores buscam responder à questão: Para além de expressar através de um único valor em torno do qual tende a se concentrar um conjunto de dados numéricos, importa saber como estas observações estão distribuídas em nossa população de estudo – são elas bastante próximas entre si ou variam muito? Sendo excelente fonte para aplicação de estudos de dados.

Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Medições na vida cotidiana – Inmetro

Neste vídeo comemorativo ao Dia Mundial da Metrologia (20 de maio). Destaca a importância das medições tanto para as tarefas cotidianas, como as compras do mês e o abastecimento dos carros, quanto para a fabricação de bens na indústria.

Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.

Estatística (Média, Mediana, Moda, Variância e Desvio Padrão)

Veja em resolução de exercícios com a aplicação de estatística, trabalhando a média, a mediana, a moda, a variância e desvio padrão.

Aponte a câmera para o código e acesse o link do conteúdo ou clique no código para acessar.