

Conceitos e Aplicações Básicas de Estatística

Descrição

A estatística é a ciência que se dedica à coleta, organização, análise e interpretação de dados. Para o candidato a concursos públicos, dominar seus conceitos é imprescindível, pois permeia desde análises de políticas públicas até gestão administrativa e elaboração de diagnósticos institucionais.

Conforme afirma o reconhecido estatístico Mário Triola, em sua obra "Introdução à Estatística": "A estatística é a arte e a ciência de coletar, analisar, apresentar e interpretar dados. Ela transforma dados brutos em informação útil para tomada de decisão."

A estatística divide-se em duas vertentes principais: **estatística descritiva** (que organiza e resume dados) e **estatística inferencial** (que extrai conclusões sobre populações a partir de amostras). Este material concentra-se na estatística descritiva, fundamental para qualquer candidato.

População, Universo, Amostra: Os Três Pilares Iniciais

População e Universo: Definições e Distinções

Na linguagem estatística, **população** e **universo** são frequentemente tratados como sinônimos, embora alguns autores façam sutis diferenciações técnicas. Para fins práticos e de concursos públicos, ambos referem-se ao **conjunto completo de elementos que compartilham características comuns e são objeto de investigação estatística**.

A população é definida pela característica que se deseja estudar, não pelo número de pessoas ou objetos envolvidos. Uma população é sempre finita ou infinita em relação ao tamanho potencial de elementos.

População Finita: é aquela cujos elementos podem ser enumerados, ainda que teoricamente. Exemplos concretos incluem: todos os servidores públicos do Tribunal de Justiça do Estado de São Paulo em determinada data; todas as notas fiscais emitidas por uma empresa em um ano fiscal; todos os processos administrativos de um órgão

p blico em um per odo espec fico.

Popula  o Infinita:   aquela cuja enumera  o de todos os elementos   imposs vel. Exemplos incluem: todas as possibilidades de resultados no lan amento infinito de um dado; todos os comprimentos de pe as que uma m quina poderia potencialmente produzir indefinidamente; a popula  o de mol culas de ar em uma sala.

Na pr tica administrativa, frequentemente lidam-se com popula  es teoricamente infinitas que s o tratadas estatisticamente como finitas para fins de amostragem. Esta distin o aparece recorrentemente em quest es de metodologia.

Amostra: Representatividade e Confiabilidade

A amostra   um subconjunto selecionado da popula  o, mediante procedimentos espec ficos, que permite realizar infer ncias sobre a popula  o inteira com margem de erro calcul vel e probabilidade conhecida.

A utiliza  o de amostras justifica-se por raz es multifacetadas:

- **Raz es econ micas:** Redu o significativa de custos operacionais, particularmente em institui es p blicas com or amentos limitados
- **Raz es temporais:** Economia consider vel de tempo em coleta e an lise de dados
- **Raz es t cnicas:** Menor probabilidade de erros quando volume de dados   reduzido
- **Raz es de viabilidade:** Possibilita estudos em popula  es infinitas ou muito grandes
- **Raz es metodol gicas:** Permite estudos destrutivos onde a an lise destr  o elemento (teste de resist ncia de materiais, por exemplo)

A validade de qualquer an lise estat stica repousa fundamentalmente na **representatividade da amostra**. Uma amostra   representativa quando reproduz, em propor es semelhantes, as caracter sticas da popula  o da qual foi extra da. Amostras enviesadas ou n o representativas produzem resultados n o confi veis e podem levar a conclus es completamente equivocadas sobre pol ticas p blicas e decis es administrativas.

A rela  o entre tamanho da amostra e confiabilidade dos resultados   inversamente proporcional ao erro amostral (margem de erro). Quanto maior a amostra, menor o erro amostral, maior a precis o das estimativas sobre a popula  o.

Amostragem: M todos e Procedimentos Estat sticos

A amostragem é o processo sistemático de seleção dos elementos que constituem a amostra. A escolha do método de amostragem é decisiva metodológica crítica que determina diretamente a confiabilidade dos resultados e a validade das inferências estatísticas.

Métodos Probabilísticos de Amostragem

Os métodos probabilísticos (ou aleatórios) oferecem confiabilidade estatística superior porque garantem que cada elemento da população possui probabilidade conhecida, diferente de zero, de ser selecionado. Esta característica permite o cálculo preciso do erro amostral.

Amostragem Aleatória Simples: Este é o método mais fundamental e puro. Todo elemento da população possui a mesma probabilidade de ser selecionado. É análogo ao sorteio de loteria onde cada número tem chance idêntica de sair. Utiliza-se tabelas de números aleatórios geradas por computador, ou métodos de sorteio controlado.

Procedimento: Numera-se todos os elementos da população de 1 a N. Utiliza-se gerador de números aleatórios para selecionar n números. Os elementos correspondentes aos números selecionados compõem a amostra.

Aplicação prática em contexto público: Selecionar 50 processos dentre 2.000 processos administrativos para auditoria, numerando-os de 1 a 2.000 e selecionando 50 números aleatoriamente.

Amostragem Sistemática: Método que combina aleatoriedade com sistematização. Após ordenar os elementos da população, seleciona-se o primeiro elemento aleatoriamente dentro de um intervalo, e em seguida seleciona-se cada k-ésimo elemento subsequente, onde k é calculado como N/n .

Fórmula do intervalo:

$$k = N/n$$

, onde N é o tamanho da população e n é o tamanho desejado da amostra.

Aplicação: Para amostra de 100 elementos de população de 5.000, $k = 50$. Seleciona-se aleatoriamente um número entre 1 e 50 (digamos 27), depois seleciona-se os elementos: 27, 77, 127, 177, 227, etc.

Vantagem: Praticidade em processos administrativos e quando a população está organizada em listas.

Amostragem Estratificada: A população é dividida em estratos (subgrupos internamente homogêneos), e amostras são extraídas proporcionalmente de cada estrato. Esta metodologia garante que grupos importantes não sejam negligenciados.

Exemplo institucional: Uma auditoria em instituição pública com 3.000 servidores (1.000 nível superior, 1.200 nível médio, 800 operacional). Para amostra de 300 servidores, extrai-se: 100 de nível superior, 120 de nível médio, 80 operacional (proporções 1:1,2:0,8).

Vantagem: Aumenta precisão das estimativas quando há heterogeneidade entre estratos.

Amostragem por Conglomerados: A população é dividida em grupos (conglomerados) que são internamente heterogêneos mas similares entre si. Seleciona-se aleatoriamente alguns conglomerados completos para análise.

Aplicação: Para pesquisa sobre satisfação de servidores em instituição federal com 50 unidades regionais, seleciona-se aleatoriamente 5 unidades e entrevista-se todos os servidores dessas 5 unidades.

Vantagem: Extremamente útil quando população está geograficamente dispersa, reduzindo custos de deslocamento.

Observação importante: Amostragem estratificada e por conglomerados parecem similares mas diferem fundamentalmente: na estratificada, extrai-se amostra de cada estrato; na conglomerados, seleciona-se conglomerados inteiros.

Métodos Não-Probabilísticos de Amostragem

Estes métodos não garantem representatividade estatística rigorosa nem permitem cálculo de erro amostral. Utilizados quando limitações práticas impedem amostragem probabilística ou em pesquisas exploratórias.

Amostragem por Conveniência:

Seleciona-se elementos de fácil acesso ao pesquisador. Exemplo: entrevistar apenas servidores que voluntariamente comparecem.

Advertência crítica:

Este método introduz vieses significativos e deve ser absolutamente evitado em estudos que buscam precisão estatística e confiabilidade de resultados. Frequentemente aparece em questões de concursos como exemplo de metodologia inadequada.

Amostragem Intencional ou por Amostragem por Quotas:

Julgamento: O pesquisador seleciona elementos que julga serem representativos baseado em conhecimento prévio ou expertise.

Similar à estratificada em aparência, mas sem seleção aleatória dentro dos estratos. O pesquisador preenche quotas de cada grupo utilizando critério de conveniência.

Questões frequentemente testam a capacidade de distinguir entre amostragem probabilística (que permite inferência estatística válida) e não-probabilística (que não permite). Esta diferença é fundamental para avaliar validade de pesquisas.

Variáveis: Tipologia e Caracterização Estatística

Uma **variável** é qualquer característica, atributo, propriedade ou fenômeno mensurável do elemento pesquisado que pode assumir diferentes valores ou categorias. É aquilo que se mede, observa ou registra em uma investigação estatística.

A compreensão da natureza das variáveis é fundamental porque determina quais técnicas estatísticas são apropriadas para sua análise.

Variáveis Qualitativas ou Categóricas

Expressam atributos, qualidades ou categorias sem relação intrínseca de ordem ou magnitude numérica, embora possam ser representadas numericamente para fins de processamento.

Variáveis Qualitativas Nominais: Não existe ordenação natural ou lógica entre as categorias. Cada categoria é simplesmente diferente das demais, não sendo melhor, pior, maior ou menor.

Exemplos: cor dos olhos (azul, castanho, verde), estado civil (solteiro, casado, divorciado, viúvo), região geográfica (Norte, Nordeste, Sudeste, Sul, Centro-Oeste), partido político de filiação, departamento de lotação de servidor público, tipo de processo administrativo.

Características essenciais: Operações matemáticas (soma, subtração, multiplicação) não possuem sentido ou validade. Não se pode dizer que "viúvo é maior que divorciado" ou que "Sudeste menos Sul é igual a Nordeste".

Variáveis Qualitativas Ordinais: Existe uma ordenação natural ou hierárquica entre as categorias, embora não se quantifique a distância entre elas. As categorias possuem relação de precedência.

Exemplos: grau de escolaridade (fundamental incompleto, fundamental completo, médio incompleto, médio completo, superior incompleto, superior completo), classe social (baixa, média, alta), nível de satisfação em escala Likert (totalmente insatisfeito, insatisfeito, neutro, satisfeito, totalmente satisfeito), classificação de desempenho funcional (insuficiente, regular, bom, excelente), prioridade de processo (baixa, média, alta).

Característica crucial: Embora haja ordem, não se pode afirmar que a distância (intervalo) entre "insatisfeito" e "neutro" é idêntica à distância entre "neutro" e "satisfeito". As categorias não são equidistantes.

A ordenação permite aplicar técnicas estatísticas um pouco mais sofisticadas (como mediana e percentis) que em nominais, mas ainda com restrições comparado a variáveis quantitativas.

Variáveis Quantitativas ou Numéricas

Expressam valores numéricos reais sobre os quais operações matemáticas possuem significado prático e teórico. Permitem aplicação de toda gama de técnicas estatísticas.

Variáveis Quantitativas Discretas: Assumem apenas valores inteiros e específicos, resultando de processos de contagem. Entre dois valores discretos consecutivos não existem valores intermediários possíveis.

Exemplos: número de filhos de um servidor, número de processos tramitados em um mês, número de servidores em um departamento, número de ausências injustificadas, número de cursos realizados, quantidade de denúncias recebidas.

Características: Cada valor é isolado; não se pode ter 2,5 filhos ou 3,7 processos.

Variáveis Quantitativas Contínuas: Podem teoricamente assumir qualquer valor dentro de um intervalo contínuo, resultando de processos de medição. Entre dois valores contínuos sempre existem infinitos valores intermediários possíveis.

Exemplos: peso de um indivíduo, altura, temperatura, tempo de duração de um atendimento ao público, renda mensal, velocidade de processamento de dados, consumo de energia elétrica.

Características: Os valores são teoricamente infinitos em um intervalo (peso pode ser 70kg, 70,5kg, 70,51kg, 70,512kg, etc.).

Ponto de atenção prático: Em pesquisas práticas e relatos administrativos, variáveis contínuas frequentemente aparecem discretizadas ou arredondadas (peso registrado em quilograma inteiro, tempo em minutos completos), mas mantêm sua natureza contínua conceitual. Isso é importante para escolha de técnicas analíticas apropriadas.

Observação para concursos: Questões frequentemente apresentam situações práticas e exigem classificação correta da variável. Exemplo: número de horas trabalhadas? É discreto (contagem de horas inteiras) ou contínuo (pode ter 8,5 horas)? A resposta depende de como é medido na instituição.

Medidas de Tendência Central: Sintetizando o Centro dos Dados

As medidas de tendência central descrevem o **valor central, típico ou representativo** de um conjunto de dados. São indicadores sintetizadores que buscam representar, em um único número, o comportamento geral ou típico de todos os dados.

Estas medidas são fundamentais porque reduzem grande volume de informações a um valor interpretável, facilitando comunicação e comparação entre grupos diferentes.

Média Aritmética: A Medida Clássica

A **média aritmética** (ou simplesmente **média**) é a soma de todos os valores dividida pelo número total de observações:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Onde:

$$\bar{x}$$

representa a média

$$\sum_{i=1}^n x_i$$

denota a soma de todos os valores de x

n

o número total de observações

Características fundamentais:

- É a medida mais intuitiva, amplamente utilizada e compreendida
- Leva em consideração absolutamente todos os valores do conjunto
- É sensível a valores extremos ou atípicos (outliers); um único valor muito alto ou muito baixo pode distorcer significativamente a média
- Existe uma única média para cada conjunto de dados
- É apropriada para dados quantitativos (discretos ou contínuos), nunca para dados qualitativos

Exemplo ilustrativo prático: Considere analista de recursos humanos analisando salários de cinco servidores:

- Servidor A: R\$ 3.000
- Servidor B: R\$ 3.200
- Servidor C: R\$ 3.500
- Servidor D: R\$ 3.800
- Servidor E: R\$ 15.000 (diretor, valor atípico)

$$\bar{x} = (3.000 + 3.200 + 3.500 + 3.800 + 15.000) / 5 = \text{R\$ } 5.700$$

Note como o salário atípico de R\$ 15.000 distorce dramaticamente a média para cima. R\$ 5.700 não representa adequadamente o "salário típico" que está na verdade entre R\$ 3.000 e R\$ 3.800.

Média Ponderada: Quando valores possuem pesos diferentes (importâncias diferentes), utiliza-se média ponderada:

$$\bar{x}_p = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i}$$

Onde w_i representa o peso atribuído a cada valor x_i .

Aplicação prática: Cálculo de desempenho quando diferentes avaliações possuem pesos distintos. Se um servidor possui: Avaliação 1 (nota 8, peso 2), Avaliação 2 (nota 9, peso 3),

Avaliação 3 (nota 7, peso 1):

$$\bar{x}_p = \frac{8(2) + 9(3) + 7(1)}{2 + 3 + 1} = \frac{16 + 27 + 7}{6} = \frac{50}{6} = 8,33$$

Mediana: A Medida de Simetria

A **mediana** é o valor que **divide exatamente ao meio o conjunto ordenado de dados**, deixando 50% dos valores abaixo e 50% dos valores acima. É o valor central quando os dados estão organizados em ordem.

Procedimento de cálculo:

Ordene todos os valores em ordem crescente (ou decrescente, o resultado é o mesmo)

Se o número de observações (n) é **ímpar**: A mediana é o valor que ocupa a posição central, especificamente a posição:

$$n + 1 \frac{n + 1}{2}$$

Se o número de observações (n) é **par**: A mediana é a média aritmética dos dois valores centrais, nas posições:

$$n \frac{n}{2}$$

e

$$n + 1 \frac{n}{2} + 1$$

Exemplo com número ímpar de observações: Conjunto: 7, 12, 15, 18, 22 (n=5) Mediana está na posição $(5+1)/2 = 3$, sendo o valor 15

Exemplo com número par de observações: Conjunto: 7, 12, 15, 18, 22, 25 (n=6) Mediana = (valor na posição 3 + valor na posição 4) / 2 = (15 + 18) / 2 = 16,5

Características essenciais:

- Não é afetada por valores extremos; é uma medida robusta
- Ideal especialmente quando existem outliers ou valores atípicos
- Divide a distribuição em duas partes equiprováveis
- Menos sensível a alterações nos extremos que a média
- Aplicável a dados quantitativos e também a dados qualitativos ordinais

Comparação ilustrativa com a média: Retornando ao exemplo de salários (R\$ 3.000, R\$ 3.200, R\$ 3.500, R\$ 3.800, R\$ 15.000):

- Média = R\$ 5.700 (distorcida para cima pelo salário atípico)
- Mediana = R\$ 3.500 (valor do meio, muito mais representativo da realidade salarial)

A mediana de R\$ 3.500 reflete muito mais adequadamente o "salário típico" deste grupo que a média de R\$ 5.700.

Ponto de atenção para concursos: Questões frequentemente comparam moda e mediana para testar compreensão sobre influência de outliers em distribuições. A mediana é preferida quando dados contêm valores extremos.

Moda: A Medida de Frequência

A **moda** é o valor (ou categoria) que **ocorre com maior frequência** no conjunto de dados. É simplesmente o valor mais frequente.

Identificação da moda: Conta-se a frequência com que cada valor aparece. O(s) valor(es) com maior frequência é (são) a(s) moda(s).

Exemplo prático: Em pesquisa sobre nível de satisfação (escala 1-5) de 20 servidores:

- Nota 1: 2 servidores
- Nota 2: 3 servidores
- Nota 3: 5 servidores
- Nota 4: 7 servidores
- Nota 5: 3 servidores

A moda é 4, pois é a nota com maior frequência (7 servidores).

Características importantes:

- Pode não existir (quando todos os valores têm a mesma frequência, distribuição amodal)
- Pode haver mais de uma moda (distribuição bimodal quando dois valores têm máxima frequência igual, multimodal quando mais de dois)
- É aplicável a dados qualitativos nominais, ordinais e quantitativos
- Não é afetada por valores extremos
- Em distribuições simétricas e unimodais, moda, mediana e moda coincidem aproximadamente

Observação importante: Uma distribuição amodal (sem moda definida) ou multimodal (com várias modas) frequentemente indica heterogeneidade nos dados, sugerindo possível divisão em subgrupos distintos.

Relação entre Moda, Mediana e Moda em Diferentes Distribuições

Em distribuição simétrica e unimodal: Os três valores coincidem aproximadamente

$\text{Moda} = \text{Mediana} = \text{Moda}$

Em distribuição assimétrica à direita (cauda positiva, valores extremos altos):

$\text{Moda} > \text{Mediana} > \text{Moda}$

A média é puxada para cima pelos valores altos. Exemplo: salários em setor onde alguns ganham muito mais elevam a média acima da mediana.

Em distribuição assimétrica esquerda (cauda negativa, valores extremos baixos):

$$\text{Média} < \text{Mediana} < \text{Moda}$$

A média é puxada para baixo pelos valores baixos.

Implicação crítica para análise: Esta relação permite avaliar o tipo de distribuição apenas comparando as três medidas. Isto é especialmente útil quando a distribuição gráfica não está disponível.

Escolha da Medida de Tendência Central Apropriada

Utilize a média aritmética quando:

- Os dados são quantitativos e aproximadamente simétricos
- Não existem outliers significativos
- Deseja-se uma medida que considere todos os valores
- Planeja realizar análises inferenciais posteriores

Utilize a mediana quando:

- Dados quantitativos contêm outliers ou valores atípicos
- A distribuição é assimétrica
- Dados são ordinais
- Deseja representação mais robusta não influenciada por extremos

Utilize a moda quando:

- Dados são qualitativos nominais
- Deseja identificar categoria mais frequente ou típica
- Há interesse em valor mais comum, não valor central
- Distribuição é claramente multimodal e isso é relevante para análise

Medidas de Dispersão: Quantificando a Variabilidade

Enquanto as medidas de tendência central descrevem onde está o centro dos dados, as medidas de dispersão descrevem o **grau de variação, espalhamento ou heterogeneidade dos dados em relação ao valor central**. São essenciais para compreender se os dados estão concentrados (homogêneos) ou dispersos (heterogêneos).

Dois conjuntos podem ter a mesma média mas dispersões completamente diferentes, levando a conclusões analíticas radicalmente distintas.

Amplitude: A Medida Mais Elementar

A **amplitude** (ou amplitude total) é a **diferença entre o maior valor (máximo) e o menor valor (mínimo)** do conjunto de dados:

$$A = x_{\max} - x_{\min} = x_{\{máx\}} - x_{\{mín\}}$$

Características:

- Medida mais simples de calcular e mais rápida
- Utiliza apenas dois valores: o máximo e o mínimo
- Ignora completamente todos os valores intermediários
- Sensível a valores extremos isolados
- Pouco precisa para representar a dispersão real dos dados

Exemplo prático: Em conjunto de tempos de atendimento (em minutos): 5, 12, 15, 18, 22

$$\text{Amplitude} = 22 - 5 = 17 \text{ minutos}$$

Porém, esta informação não nos diz se a maioria dos atendimentos está agrupada próximo a 5 minutos, próximo a 22 minutos, ou espalhada por todo o intervalo. Daí a necessidade de medidas mais sofisticadas.

Conclusão: A amplitude é informativa apenas como visão inicial; raramente é suficiente como medida única de dispersão.

Variância: A Medida Fundamental de Dispersão

A **variância** mede o **desvio quadrático médio de todos os valores em relação à média aritmética**. Quantifica o quanto, em média, cada observação está afastada da média.

Para uma população completa:

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

Para uma amostra:

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

Onde:

x_i = cada observação individual
 μ = a média populacional ou \bar{x} = a média amostral
 \bar{x} = a média amostral

N = o tamanho total da população
 n = o tamanho da amostra

N = o tamanho total da população
 n = o tamanho da amostra

Observação absolutamente crítica: Note a diferença fundamental no divisor:

Para população: divide-se por N
 Para amostra: divide-se por n

Para amostra: divide-se por $n-1$, não por n

Esta diferença ($n-1$) é chamada de **correção de Bessel** e existe porque, ao usar uma amostra para estimar a variância da população, o uso de n subestimaria a variância populacional. O divisor $n-1$ (chamado de graus de liberdade) garante que a variância amostral seja um **estimador não enviesado** da variância populacional. **Este detalhe aparece frequentemente em questões de concursos públicos.**

Processo de cálculo:

Calcula-se a diferença de cada valor em relação à média: $(x_i - \bar{x})$

Eleva-se ao quadrado cada diferença: $(x_i - \bar{x})^2$

Soma-se todos os quadrados: $\sum (x_i - \bar{x})^2$

Divide-se por $n-1$ (amostra) ou N (população) ou N

Características essenciais:

- Quanto maior a variância, maior a dispersão dos dados
- Variância zero indica que todos os valores são idênticos (sem variação)
- Expressa-se na **unidade original ao quadrado**, o que torna interpretação prática difícil

Desvantagem crítica: A elevação ao quadrado na fórmula torna a unidade final incompreensível. Se medimos renda em reais, a variância fica em reais ao quadrado, que não tem significado prático direto.

Desvio Padrão: A Medida Interpretável

O **desvio padrão** é a **raiz quadrada positiva da variância**, trazendo a medida de dispersão de volta à unidade original, tornando-a interpretável.

Para uma população:

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

Para uma amostra:

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

Relação fundamental:

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

O desvio padrão é a raiz da variância.

Características essenciais:

- Medida mais intuitiva e interpretável que a variância
- Está na mesma unidade que os dados originais
- Permite interpretar prática: quanto maior o desvio padrão, maior a variabilidade
- Fundamental para construir intervalos de confiança
- Essencial para aplicações envolvendo Distribuição Normal
- Aproximadamente 68% dos dados de uma distribuição normal caem dentro de ± 1 desvio padrão da média; 95% dentro de ± 2 desvios padrão

Exemplo prático interpretável: Se o tempo de atendimento tem média 15 minutos e desvio padrão 3 minutos, podemos afirmar que os tempos variam tipicamente 3 minutos acima ou abaixo da média. A maioria dos atendimentos (cerca de 68%) dura entre 12 e 18 minutos.

Comparação de dispersão:

Conjunto A: Média 50, desvio padrão 2 Conjunto B: Média 50, desvio padrão 10

Ambos têm a mesma média, mas Conjunto B é muito mais disperso. O desvio padrão imediatamente revela isto.

Ponto de atenção crítico para concursos: Dominar completamente a diferença entre:

σ (desvio padrão populacional) versus s (desvio padrão amostral)
 Divisor N versus $n-1$

Frequentemente questões cobram exatamente estes detalhes.

Coeficiente de Variação: Dispersão Relativa

O **coeficiente de variação (CV)** é uma medida **relativa** de dispersão que expressa o desvio padrão como percentagem da média:

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

$$\text{Ou equivalentemente: } CV = \frac{\sigma}{\mu} \times 100\%$$

Utilidade fundamental:

- Permite comparar dispersão entre grupos que têm **médias diferentes**
- Permite comparar dispersão entre **variáveis com unidades diferentes**
- Quanto menor o CV, menor a variabilidade relativa; quanto maior, maior a variabilidade relativa
- É adimensional (sem unidade), permitindo comparações universais

Exemplo comparativo prático:

Máquina A produz parafusos com peso médio 100g e desvio padrão 2g: $CV_A = \frac{2}{100} \times 100\% = 2\%$

Máquina B produz correntes com peso médio 1.000g e desvio padrão 50g: $CV_B = \frac{50}{1.000} \times 100\% = 5\%$

Embora o desvio absoluto de B (50g) seja muito maior que o de A (2g), a máquina A tem melhor consistência relativa. A máquina A produz com 2% de variação, enquanto B tem 5% de variação.

Interpretação prática do CV:

- CV < 15%: baixa dispersão relativa, dados homogêneos
- CV entre 15% e 30%: dispersão moderada
- CV > 30%: alta dispersão relativa, dados heterogêneos

Aplicação em administração pública: Comparar variabilidade de custos entre departamentos com orçamentos diferentes, ou variabilidade de tempos de processamento entre tipos diferentes de processos.

Intervalo Interquartil: Medida Robusta

O **intervalo interquartil (IIQ)**, também chamado amplitude interquartil, é a diferença entre o terceiro quartil e o primeiro quartil:

$$IIQ = Q_3 - Q_1$$

Conceitos básicos:

- Quartis dividem os dados em quatro partes iguais
- Primeiro quartil (Q_1): valor abaixo do qual caem 25% dos dados
- Segundo quartil (Q_2): valor abaixo do qual caem 50% dos dados (é a mediana)
- Terceiro quartil (Q_3): valor abaixo do qual caem 75% dos dados

Características fundamentais:

- Contém exatamente os 50% dos dados centrais
- Não é afetado por valores extremos; é altamente robusta
- Particularmente útil para identificação de outliers
- Menos precisa que desvio padrão em distribuições normais, mas mais robusta em distribuições anormais

Método de Tukey para Identificação de Outliers:

Um valor é considerado outlier (valor atípico ou extremo) se:

- For inferior a $Q_1 - 1,5 \times IIQ$, ou
- For superior a $Q_3 + 1,5 \times IIQ$

Exemplo prático: Se $Q_1 = 30$, $Q_3 = 50$, então: $IIQ = 50 - 30 = 20$
 $\text{Limite inferior} = 30 - 1,5(20) = 30 - 30 = 0$
 $\text{Limite superior} = 50 + 1,5(20) = 50 + 30 = 80$

Qualquer valor menor que 0 ou maior que 80 seria considerado outlier.

Observação importante: O intervalo interquartil é especialmente útil em análises exploratórias de dados para detectar observações anômalas antes de realizar análises mais sofisticadas.

Porcentagem: Fundamento Essencial em Estatística Aplicada

A **porcentagem** é a representação proporcional de uma quantidade expressa como fração de 100. Embora conceitualmente simples, é absolutamente fundamental em estatística aplicada e frequentemente em questões de concursos públicos.

Conceito e Fórmula Fundamental

Uma porcentagem representa quantas partes de cada 100 uma quantidade corresponde:

$$P\% = \frac{\text{parte}}{\text{total}} \times 100$$

$$\text{Equivalentemente: } \text{parte} = \text{total} \times \frac{P}{100}$$

A porcentagem é uma razão onde o denominador sempre é 100.

Cálculos Básicos de Porcentagem

Encontrar a porcentagem: Se 35 de 140 pessoas responderam a uma pesquisa, qual a porcentagem?

$$P = \frac{35}{140} \times 100 = 0,25 \times 100 = 25\%$$

Encontrar o valor correspondente a uma porcentagem: Se 25% de 140 é quanto?

$$\text{Valor} = 140 \times \frac{25}{100} = 140 \times 0,25 = 35$$

Encontrar o total quando conhece-se a porcentagem: Se 35 pessoas representam 25% do total, qual é o total?

$$\text{Total} = \frac{35}{0,25} = \frac{35}{25/100} = \frac{35 \times 100}{25} = 140$$

Aumentos e Diminuições Percentuais

Aumento percentual: Quando um valor x aumenta de p por cento, o novo valor é:

$$\text{Novo valor} = x \times \left(1 + \frac{p}{100}\right) = x \times (1 + 0,0p)$$

O multiplicador $(1 + p/100)$ é chamado **fator de aumento**.

Exemplo prático: Um servidor público com salário de R\$ 3.000 recebe aumento de 10%:

$$\text{\$}\text{\text{Novo salário}} = 3.000 \times (1 + 0,10) = 3.000 \times 1,10 = \text{\$} 3.300\text{\$}$$

Aumento em reais: R\$ 300

Diminuição percentual: Quando um valor x diminui de p por cento, o novo valor \hat{x} :

$$\text{Novo valor} = x \cdot (1 - p/100) = x \cdot (1 - 0,0p) \quad \text{Novo valor} = x \cdot (1 - p/100) = x \cdot (1 - 0,0p)$$

O multiplicador $(1 - p/100)$ $\hat{\text{c}}$ chamado **fator de diminuição**.

Exemplo prático: Um orçamento de R\$ 100.000 sofre redução de 15%:

$$\text{\$}\text{\text{Novo orçamento}} = 100.000 \times (1 - 0,15) = 100.000 \times 0,85 = \text{\$} 85.000\text{\$}$$

Redução em reais: R\$ 15.000

Aumentos ou diminuições sucessivos: Se ocorrem múltiplos aumentos ou diminuições, multiplicam-se os fatores:

Aumento de 10%, depois aumento de 5%: $\text{\$}\text{\text{Valor final}} = x \times 1,10 \times 1,05 = x \times 1,155\text{\$}$ Equivalente a aumento único de 15,5% (não 15%)

Ponto de atenção crítico: Aumentos e diminuições percentuais sucessivos não se somam. A multiplicação dos fatores revela o efeito real.

Porcentagem de Porcentagem

Frequentemente em questões aparecem situações onde se calcula porcentagem de uma porcentagem. O modo correto $\hat{\text{c}}$ multiplicar as proporções (em forma decimal):

Exemplo: Se 60% de um grupo total são mulheres, e 40% das mulheres trabalham em área administrativa, qual porcentagem do grupo total são mulheres administrativas?

$$\text{\$}\text{\text{Proporção}} = 0,60 \times 0,40 = 0,24 = 24\%\text{\$}$$

Ponto crítico: Não se somam porcentagens (60% + 40% $\hat{=}$ 100% neste contexto). Multiplicam-se as proporções decimais.

Distribuição de Frequências em Porcentagem

Em uma distribuição de frequências, expressa-se as frequências absolutas (contagens) como porcentagens (frequências relativas) para permitir melhor comparação e interpretação:

Exemplo prático em contexto de concurso público:

Uma instituição avaliou desempenho de 80 servidores:

Nível de Desempenho	Frequência Absoluta	Frequência Relativa (%)
Insatisfatório	12	$12/80 \cdot 100 = 15\%$

Nível de Desempenho	Frequência Absoluta	Frequência Relativa (%)
Satisfatório	48	$48/80 \cdot 100 = 60\%$
Excelente	20	$20/80 \cdot 100 = 25\%$
Total	80	100%

Observação crítica: As porcentagens devem **sempre totalizar 100%** (ou muito próximo, considerando arredondamentos). Se não totalizarem 100%, há erro nos cálculos.

Variação Percentual: Medindo Mudanças

Quando se compara um valor em dois momentos diferentes, calcula-se a **variação percentual** (ou taxa de variação):

$$\text{Variação \%} = \frac{\text{Valor Final} - \text{Valor Inicial}}{\text{Valor Inicial}} \times 100$$

Exemplo prático: Um indicador de satisfação de servidores passou de 70% em 2022 para 82% em 2023:

$$\text{Variação} = \frac{82 - 70}{70} \cdot 100 = 17,14\%$$

Este indicador **creceu 17,14%** em relação ao valor inicial.

Ponto de atenção absolutamente crítico: Não confundir **variação percentual** (calculada acima) com **diferença em pontos percentuais**:

- Diferença em pontos percentuais: $82\% - 70\% = 12$ pontos percentuais (diferença simples)
- Variação percentual: 17,14% (crescimento relativo ao valor inicial)

Estes são conceitos diferentes frequentemente cobrados em concursos para testar compreensão profunda.

Quando um indicador passa de 50% para 60%:

- Diferença em p.p.: 10 pontos percentuais
- Variação relativa: $(60-50)/50 \cdot 100 = 20\%$

Conversão entre Frações, Decimais e Porcentagens

Frequentemente necessário converter entre estas representações:

De fração para porcentagem:

$$3/4 = 0,75 = 75\%$$

De decimal para porcentagem:

$$0,25 = 25\% \text{ (multiplica-se por 100)}$$

De porcentagem para decimal:

$75\% = 0,75$ (divide-se por 100) $75\% = 0,75$ \text{(divide-se por 100)}

De porcentagem para fração:

$25\% = \frac{25}{100} = \frac{1}{4}$

(simplifica-se quando possível)

Integração Prática dos Conceitos: Um Caso Real

Para demonstrar como todos estes conceitos integram-se em análise estatística prática realista, consideremos um cenário típico de concurso público:

Uma instituição federal de grande porte deseja avaliar a satisfação de seus 2.000 servidores ativos. Como questionar todos é inviável, extrai-se uma **amostra** de 200 servidores mediante **amostragem estratificada** por nível hierárquico (garantindo representação proporcional de cada nível).

Para cada servidor entrevistado, coleta-se:

Variáveis qualitativas nominais:

- Departamento de lotação (Administração, Técnico, Operacional)
- Formação profissional (Ensino Médio, Ensino Superior, Especialização)

Variáveis qualitativas ordinais:

- Nível de satisfação com ambiente (1 = Muito Insatisfeito até 5 = Muito Satisfeito)
- Avaliação do clima organizacional (Ruim, Regular, Bom, Excelente)

Variáveis quantitativas discretas:

- Tempo de serviço em anos (números inteiros)
- Número de cursos de capacitação realizados

Variáveis quantitativas contínuas:

- Jornada semanal de trabalho em horas
- Renda mensal bruta

Após compilação dos dados sobre **nível de satisfação** (escala 1-5) de todos os 200 respondentes:

Distribuição: 15 pessoas com nota 1; 25 pessoas com nota 2; 40 pessoas com nota 3; 85 pessoas com nota 4; 35 pessoas com nota 5

Medidas de tendência central:

- **Média:** $(15 \cdot 1 + 25 \cdot 2 + 40 \cdot 3 + 85 \cdot 4 + 35 \cdot 5) / 200 = (15 + 50 + 120 + 340 + 175) / 200 = 700 / 200 = 3,5$ Interpretação: Satisfação média é 3,5, tendendo ao neutro/satisfação moderada.
- **Mediana:** Valor na posição 100,5 (média das posições 100 e 101). Ordenando: 100º e 101º valores correspondem a nota 4 Interpretação: Metade dos servidores tem satisfação ≥ 4 , metade tem satisfação < 4 .
- **Moda:** Nota 4 (frequência máxima de 85 ocorrências) Interpretação: Satisfação mais frequente é 4.

Observação analítica: Média (3,5) < Mediana (4) < Moda (4) não ocorre, mas temos Média < mediana/moda, sugerindo leve assimetria à esquerda (algumas insatisfações puxam a média para baixo).

Medidas de dispersão:

Cálculos mostram:

- Variância: $s^2 \approx 1,28$
- Desvio padrão: $s \approx 1,13$

Interpretação: Respostas variam tipicamente 1,13 pontos em relação à média 3,5. A maioria das respostas está entre 2,37 e 4,63 (média ± 1 desvio padrão).

Análise percentual:

Nível	Frequência Absoluta	Frequência Relativa
Muito Insatisfeito (1-2)	40	20%
Neutro (3)	40	20%
Satisfeito (4-5)	120	60%
Total	200	100%

60% dos servidores manifestam satisfação (notas 4-5), enquanto 20% manifestam insatisfação (notas 1-2) e 20% são neutros.

Esta análise integrada permite conclusões robustas para subsidiar decisões administrativas sobre clima organizacional, políticas de recursos humanos, e intervenções para melhorar satisfação dos insatisfeitos.

Pontos Críticos para Sucesso em Concursos Públicos

Ao estudar estatística para concursos, mantenha atenção especial em:

Definições Precisas: Cada termo possui definição exata no contexto estatístico. Confundir população com amostra, média com mediana, população com universo, ou variância com

desvio padrão compromete respostas inteiras. Concursos testam precisão conceitual.

Mãotodos Apropriados: Nem toda medida é apropriada para todo tipo de dado. Não se calcula média de cores, não se calcula moda de alturas (em regra), não se aplica amostragem não-probabilística quando é exigida confiabilidade estatística. A escolha correta do método é frequentemente cobrada.

Interpretação Crítica: Estatísticas podem ser apresentadas enganosamente ou interpretadas incorretamente. Compreensão profunda permite identificar análises enviesadas, conclusões indevidas, ou manipulações. Questões exigem discernimento crítico, não apenas cálculo.

Precisão Aritmética: Erros aritméticos simples (confundir divisão por n com divisão por $n-1$, esquecer de elevar ao quadrado na variância) comprometem respostas. Pratique cálculos.

Contexto Institucional: Muitas questões apresentam cenários de administração pública, políticas governamentais, ou avaliação de programas. Compreender como estatística fundamenta tomada de decisão pública enriquece sua capacidade analítica e de interpretação.

Distinção Conceitual Crítica: Compreenda profundamente:

- Diferença entre amostragem probabilística e não-probabilística
- Por que se usa $n-1$ em amostra vs N em população
- Como outliers afetam diferentemente média vs mediana
- Por que não se somam porcentagens simples (multiplicam-se fatores)
- Relação entre variância e desvio padrão

Nota sobre Fontes e Fundamentação:

Esta exposição baseia-se em princípios consolidados da estatística descritiva conforme ensinados em cursos de metodologia científica, estatística e análise de dados em instituições de ensino superior brasileiras. Os conceitos apresentados refletem padronização internacional de terminologia estatística, conforme estabelecido em obras de referência como "Introduction to Statistical Quality Control" (Montgomery), "Statistics for Business and Economics" (Anderson, Sweeney, Williams) e normativas técnicas brasileiras de estatística.

Observa-se que não foram incluídas Súmulas do STF ou STJ pois não existem súmulas destes tribunais diretamente relacionadas a conceitos fundamentais de estatística descritiva. Súmulas referem-se tipicamente a interpretações jurídicas, não a conteúdo técnico de metodologia estatística.

Data de criação

03/20/2025

Autor

admin