

Aula 00

*Química Analítica p/ Concursos - Curso
Regular 2022*

Autor:
Diego Souza

30 de Janeiro de 2022

Sumário

Estatística aplicada à química (parte 1)	8
1 – Considerações Iniciais	8
2 – Introdução à estatística	8
3 – Estatística descritiva	9
3.1 – População, amostra e amostragem	12
3.2 – Medidas de posição ou de tendência central	17
3.3 – Medidas de dispersão	20
4 – Erros em análises químicas	26
4.1 – Três tipos de erro	26
4.2 – Diferença entre precisão e exatidão	30
4.3 – Propagação do erro aleatório	31
5 – Testes de hipótese	34
5.1 – Histograma de frequência	34
5.2 – Distribuição normal	35
5.3 – Como usar a tabela da área da cauda da distribuição normal padronizada?	37
5.4 – Testes de hipótese: teste Z-score, T-student e teste F	38
5.5 – Aplicando o teste-Z	40
5.6 – Aplicando o teste T-student	42
5.7 – Aplicando o teste F	45
5.8 – Construindo intervalos de confiança para a média populacional	46
Questões Comentadas	49
Lista de Questões da Aula	73
Gabarito	83
Principais Pontos da Aula	84





APRESENTAÇÃO DO CURSO

Olá, pessoal!

Meu nome é **Diego Souza**, Perito Criminal e Professor de Química, e logo mais dedicarei um espaço para uma breve apresentação. É com enorme satisfação que dou início a mais um projeto aqui no Estratégia Concursos, nossos **CURSOS POR ÁREA DA QUÍMICA**, sendo este o curso de **QUÍMICA ANALÍTICA PARA CONCURSOS**. Vamos percorrer os principais tópicos da QUÍMICA ANALÍTICA cobrados em concursos públicos (inclusive em provas práticas, previstas para alguns cargos) na área de química (nível técnico e superior), farmácia e engenharia química, tais como estatística aplicada à química, equilíbrio químico, métodos titulométricos e gravimétricos, marcha analítica para cátions e ânions, métodos eletroanalíticos, espectroscopias no UV-VIS e no infravermelho, noções de física aplicada a laboratórios, análise de água, validação de métodos analíticos, e introdução à quimiometria (este último tópico presente somente em alguns concursos de nível superior).

O curso é uma excelente opção para aqueles que tiveram pouca vivência prática em laboratório, seja porque seu curso possuía poucas aulas experimentais ou porque ainda não teve experiência profissional em laboratório. Trago para os PDFs e videoaulas deste curso muito da minha experiência obtida ao longo dos últimos 13 anos trabalhando em diferentes laboratórios (Universidade, Embrapa, Perícia da PCDF).

Esse novo projeto nasceu a partir de pedidos de muitos alunos e é indicado a **concurseiros, graduandos e alunos de cursos técnicos** nas seguintes situações:

1. Para quem já é nosso aluno de algum curso de química e **deseja complementar o material** de estudo com as aulas previstas para este curso;
2. Para quem é **graduando e deseja estudar os tópicos de química deste curso de maneira simples e didática (praticamente um reforço completo)**, e com foco antecipado na preparação para concursos públicos; ou
3. Para concurseiros que já se sentem totalmente seguros nas demais áreas da química e desejem complementar seus conhecimentos nesta área da química.

Vamos falar rapidamente sobre a sua preparação e de como nosso curso poderá ser determinante nesse processo. **A disponibilidade e a escolha dos materiais são fatores limitantes na preparação para a sua prova**. Se considerarmos a diversidade de conteúdos e a contextualização das provas de concurso, notamos que as informações relevantes para uma boa preparação estão espalhadas em diferentes livros. Nosso curso supera essa dificuldade, pois foi concebido para ser seu **ÚNICO MATERIAL DE ESTUDO**, trazendo exatamente o que precisa [nada a mais e nada a menos]. Tudo isso para **POTENCIALIZAR SEU APRENDIZADO e ABREVIAR SUA JORNADA ATÉ A TÃO SONHADA APROVAÇÃO**.

O nosso curso consistirá de:

- a) **Curso escrito (em PDF) composto por 13 aulas**, além dessa aula demonstrativa, nas quais abordarei a teoria e as aplicações de cada tema previsto no edital, além de cerca de 390 resoluções de questões comentadas;
- b) **Fórum de dúvidas**: um importante canal de comunicação entre os estudantes e os professores. Por ele, passarei explicações detalhadas a respeito das dúvidas que surgirem.



Atenção! Este curso é completo em PDF. Mas algumas videoaulas serão disponibilizadas como complemento no período de vigência do curso, sempre que for possível gravá-las, pois o foco principal será a entrega das aulas em PDF conforme cronograma.

Outro aviso importante: em razão do nosso curso ser mais enxuto em quantidade de aulas quando comparado com os cursos que contemplam as várias áreas da química, nossa aula 00 (aula demonstrativa) consistirá apenas de parte da primeira aula do curso que será integralizada como aula 01. No entanto, lembro que nossas aulas completas apresentam a seguinte estrutura:

- a) Teoria completa intercalada com a resolução comentada de exercícios;
- b) Lista de exercícios comentados ao final da teoria (**entre 30 e 40 questões por aula**);
- c) Sequência de enunciados de todos os exercícios da aula;
- d) Gabarito;
- e) A maioria das aulas contam com recursos extras: tabelas resumo, resumo da aula e/ou mapas mentais.

Em caso de dúvida sobre a estrutura de nossas aulas completas, você pode conferi-la baixando Aulas 00 de outros cursos meus mais extensos, de QUÍMICA, aqui no site do **Estratégia Concursos**.



APRESENTAÇÃO PESSOAL

Por fim, resta uma breve apresentação pessoal. Como já adiantei, meu nome é **Diego Souza**! Sou Doutor em Química pela Universidade Federal de Goiás, atuo como professor de Química no Estratégia Concursos e exerço o cargo de Perito Criminal da Polícia Civil do Distrito Federal (PCDF).

Além da minha esposa Nayara, irmãos, sobrinhos e pais, outra **grande paixão** minha é lecionar aqui no **Estratégia Concursos**, o que perceberão pela dedicação e pela preocupação com seu aprendizado depositadas em cada aula. Por isso, contem comigo sempre que precisarem, seja pelo fórum dos cursos ou pelas redes sociais. Será um prazer respondê-los e ajudá-los.

Estou envolvido com concursos públicos há cerca de 11 anos, conciliei os dois últimos anos de faculdade com meu primeiro cargo público de assistente administrativo. Fui aprovado em concursos na área administrativa e nos seguintes concursos na área de química: Técnico em Química da EMBRAPA (2º colocado), Analista Químico da EMBRAPA (1º), Especialista em Recursos Minerais/Química da ANM (1º) e Perito Criminal/Química da PCDF (2º).

Quanto à atividade de professor, leciono exclusivamente Química para concursos. Você perceberá que trago muito da minha experiência de laboratório para as apostilas, pois muito desse conhecimento prático pode ser cobrado de você em prova. Desde que fui convidado para compor a equipe do Estratégia Concursos, em maio de 2018, já fui responsável por mais de 210 cursos na área de Química.

Deixarei abaixo meus contatos para quaisquer dúvidas ou sugestões. Pelo Instagram e Facebook divulgo dicas, novidades, mapas mentais e dicas sobre química. Por lá e também pelo fórum, terei enorme prazer em orientá-los da melhor forma possível até sua aprovação.

[Instagram](#): Prof.DiegoSouza

[Nosso Telegram](#): t.me/profdiegosouza

[YouTube](#): Prof. Diego Souza



CRONOGRAMA DE AULAS

Vejamos a distribuição das aulas:

Aulas	Conteúdo da aula	Data de entrega
Aula 00	Estatística aplicada à química (parte 1): Introdução à estatística; Estatística Descritiva; Erros em análises químicas; Testes de hipótese.	30/02/2022
Aula 01	Equilíbrio químico (parte 1): Constantes de equilíbrio (K_{ps} , K_w , K_a , K_b); pH e pOH; Solução tampão;	06/02/2022
Aula 02	Equilíbrio químico (parte 2): Equilíbrio químico e termodinâmica; Ácidos e bases dipróticos; Outras constantes; Cálculos importantes envolvendo equilíbrio químico; Atividade e força iônica.	13/02/2022
Aula 03	Gravimetria: Análise gravimétrica; Curva de aquecimento; Pesagem na gravimetria.	20/02/2022
Aula 04	Titulometria: introdução e classificação; Titulação ácido-base; Titulação por precipitação; Titulação de complexação; Titulação por oxirredução.	27/02/2022
Aula 05	Química analítica qualitativa: marcha analítica de cátions e marcha analítica de ânions.	06/03/2022
Aula 06	Métodos eletroanalíticos; Equação de Nernst; Métodos potenciométricos; Eletrogravimetria; Coulometria; Amperometria; Voltametria.	13/03/2022
Aula 07	Técnicas espectroscópicas: fundamentos; Interação luz-matéria na região UV-VIS.	20/03/2022
Aula 08	Instrumentos para análises espectroscópicas no UV-VIS; Aplicações analíticas de espectroscopia UV-VIS;	27/01/2022
Aula 09	Espectroscopia de infravermelho; Espectrometria de massas.	03/04/2022



Aula 10	Noções de física e segurança no laboratório.	10/04/2022
Aula 11	Análise de água: Determinações físico-químicas em amostra de águas; Métodos titulométricos e espectrofotométricos.	17/04/2022
Aula 12	Estatística aplicada à química (parte 2): Método dos mínimos quadrados; Validação analítica: linearidade, sensibilidade, seletividade, robustez, precisão e exatidão.	24/04/2022
Aula 13	Estatística aplicada à química (parte 3): Planejamento fatorial; Análise de componentes principais (PCA).	01/05/2022

Essa é a distribuição dos assuntos ao longo do curso. Eventuais ajustes poderão ocorrer, especialmente por questões didáticas ou para melhor atendê-los em relação a data de prova.



ESTATÍSTICA APLICADA À QUÍMICA (PARTE 1)

1 – Considerações Iniciais

Olá, pessoal, tudo certo?

Nesta aula, começaremos nosso estudo sobre estatística aplicada a química.

Veremos que uma vidraria volumétrica é classificada de acordo com sua precisão e exatidão, que está relacionada ao erro. Desvio padrão, erro e métodos dos mínimos quadrados são assuntos abordados na estatística. Esses são apenas alguns exemplos da utilização de conceitos e ferramentas estatísticas em química, vários outros aparecerão no decorrer do nosso curso. Essa utilização é tão frequente que muitos exercícios abordam, de forma conjunta, conceitos químicos e estatísticos.

Iniciaremos nosso estudo por conceitos mais basilares da estatística, abordando tópicos como estatística descritiva, medidas de posição, medidas de dispersão, e tipos e propagação de erro. Em seguida, ainda nesta aula, estudaremos distribuição normal e testes de hipótese (teste Z, teste t e teste F).

Sem mais, vamos iniciar nossa aula. Desejo-lhe uma boa aula e lembre-se de me procurar caso fique com alguma dúvida. Bons estudos! Forte abraço!

[Instagram](#): Prof.DiegoSouza

[Facebook](#): Prof. Diego Souza

[YouTube](#): Prof. Diego Souza

2 – Introdução à estatística

A estatística está cada vez mais presente no dia-a-dia do químico. Como exemplos de situações em que são aplicados conceitos e ferramentas estatísticas em química, podemos citar:

- Avaliação de erros analíticos;
- Calibração de instrumentos analíticos laboratoriais;
- Comparação de técnicas ou métodos analíticos a nível de precisão, exatidão, recuperação e linearidade;
- Validação de métodos analíticos;
- Otimização de métodos, processos ou reações químicas, definindo as condições experimentais ótimas que podem, de acordo com o desejado, minimizar erros ou maximizar o rendimento;
- Análise exploratória de dados multivariados, na qual é possível avaliar a correlação entre diferentes variáveis experimentais em um espaço multivariado;
- Calibrações multivariadas de instrumentos analíticos, permitindo, por exemplo, a quantificação multielementar a partir de uma varredura espectral;
- Dentre vários outros exemplos.

Não vou delongar na demonstração da importância para química e nem você precisa se preocupar em memorizar essas aplicações.



De maneira mais ampla, *por que estudar estatística?*

Independente da área de atuação, profissionais lidam com fenômenos quantitativos não exatos e precisam, muitas vezes, tomar decisões a respeito desses fenômenos.

Nesse sentido, **medidas experimentais**, a exemplo das determinações analítica em laboratório, sempre estão associadas a alguma variação. Por isso, conclusões não podem ser tiradas com absoluta certeza. Mas não se apavore! Para contornar essa situação, a estatística fornece ferramentas que possibilitam conclusões (decisões) com uma grande probabilidade de estarem corretas e de rejeitar conclusões que sejam improváveis. Veremos isso claramente no estudo dos testes de hipótese, mais adiante.

Exemplos de medidas experimentais: pesquisas eleitorais, pesagens repetitivas de um peso padrão, replicatas de uma amostra cujo analito foi determinado por titulação, replicatas da obtenção da absorbância de uma amostra por espectrofotometria.

De início, trago uma definição esclarecedora de estatística (não se preocupe em decorá-la, apenas a compreenda):

A **estatística** é um conjunto de técnicas que permite, de forma sistemática, não só organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento, como também tomar decisões baseadas nessas análises e interpretação.¹

3 – Estatística descritiva

A **estatística descritiva** consiste na etapa inicial de qualquer análise. Nesta etapa é realizada a sumarização dos dados, deixando-os mais sintéticos e entendíveis de forma mais imediata. Para tanto, se utiliza de gráficos, tabelas e medidas descritivas de posição e dispersão.

Para um melhor entendimento, tomemos como exemplo um estudo, realizado por Rorato e colaboradores², sobre o nível de conhecimento de consumidores de produtos *diet* e *light* na cidade de Curitiba. Durante a pesquisa, foi aplicado o questionário abaixo com 9 questões. Vale destacar que só foram entrevistadas as pessoas que se declararam consumidores de produtos *diet* e *light*.

Considerando que foram entrevistados 80 consumidores, seria uma tarefa bastante difícil tentar extrair informações ou tendências a partir de todos formulários originais, aqueles preenchidos durante coleta de dados. Entretanto, após digitalizarmos esses formulários em uma planilha de Excel, por exemplo, poderíamos gerar tabelas e gráficos que condensem e permitam uma visão mais sintética e geral dos dados.

¹ Adaptado de <http://www.abntouvancouver.com.br/2015/10/a-pesquisa-de-campo-e-analise.html>. Acessado em 20 de jul. 2018.

² RORATO, F.; DEGASPARI, C. H.; MOTTIN, F. Avaliação do nível de conhecimento de consumidores de produtos diet e light que frequentam um supermercado de Curitiba. Visão Acadêmica, v. 7, n. 1, 2006.



E foi exatamente o que os autores do trabalho fizeram por meio de vários gráficos e tabelas distribuídos no artigo.

1. VOCÊ É CONSUMIDOR DE PRODUTOS DIET, LIGHT OU OS DOIS?

☐ DIET ☐ LIGHT ☐ DIET E LIGHT

2. IDADE

☐ 15 A 24 ANOS ☐ 25 A 34 ANOS ☐ 35 A 44 ANOS ☐ 45 A 55 ANOS ☐ >55 ANOS

3. SEXO: ☐ FEMININO ☐ MASCULINO

4. QUAL É A SUA RENDA FAMILIAR?

☐ 1 S.M. ☐ 2 a 3 S.M. ☐ 4 a 5 S.M. ☐ 6 a 7 S.M. ☐ ACIMA DE 8 S.M.

5. QUAL O SEU GRAU DE INSTRUÇÃO?

☐ 1º GRAU INCOMPLETO ☐ 1º GRAU COMPLETO ☐ 2º GRAU INCOMPLETO
☐ 2º GRAU COMPLETO ☐ 3º GRAU INCOMPLETO ☐ 3º GRAU COMPLETO
☐ NENHUMA FORMAÇÃO

6. QUAL É O SEU PESO: PESO _____ ALTURA _____
IMC : _____ kg/m²

7. PORQUE VOCÊ FAZ USO DE PRODUTOS :

DIET	LIGHT
☐ PARA EMAGRECER	☐ PARA EMAGRECER
☐ PARA MANTER A FORMA	☐ PARA MANTER A FORMA
☐ POR ALGUMA DOENÇA	☐ POR ALGUMA DOENÇA
☐ POR HÁBITO	☐ POR HÁBITO
☐ POR SER MAIS SAUDÁVEL	☐ POR SER MAIS SAUDÁVEL
☐ OUTROS _____	☐ OUTROS _____

8. VOCÊ SABE A DIFERENÇA DE UM PRODUTO DIET DE UM PRODUTO LIGHT?

☐ SIM ☐ NÃO

9. ONDE OBTIVE AS INFORMAÇÕES SOBRE ESSES PRODUTOS?

☐ TV ☐ JORNAL ☐ REVISTA ☐ INTERNET ☐ LIVROS
☐ OUTROS _____

Formulário aplicado na pesquisa sobre o nível de conhecimento de consumidores de produtos *diet* e *light* na cidade de Curitiba²

A exemplo dessa sintetização dos dados, têm-se a **tabela** abaixo sobre cruzamento entre faixa etária e consumo, na qual é possível identificar que as faixas etárias compreendidas entre 25 e 44 anos correspondem a maioria (54%) dos consumidores de produtos *diet* e *light*. Além disso, podemos observar no **gráfico de barras**, também apresentado abaixo, que esses produtos são mais consumidos por pessoas com renda familiar mais elevada, 77% dos consumidores apresentam renda familiar superior a 6 salários mínimos.

² RORATO, F.; DEGASPARI, C. H.; MOTTIN, F. Avaliação do nível de conhecimento de consumidores de produtos diet e light que frequentam um supermercado de Curitiba. Visão Acadêmica, v. 7, n. 1, 2006.



Tabela que demonstra o cruzamento entre faixa etária e consumo de produtos *diet* e *light*²

Idade	Light e Diet	Diet	Light	Total
15 a 24 anos	15%	19%	22%	21%
25 a 34 anos	31%	50%	41%	34%
35 a 44 anos	27%	19%	13%	20%
45 a 54 anos	15%	13%	7%	11%
Acima de 55 anos	8%	0%	17%	13%
Sem resposta	4%	0%	0%	1%
Total	100%	100%	100%	100%



Gráfico de barras que demonstra a faixa de renda familiar dos consumidores de produtos diet e light, entrevistados²

Agora que você está convencido que tabelas e gráficos podem sintetizar drasticamente os dados, tornando um emaranhado de números em informações facilmente entendíveis, listo abaixo outros tipos de gráficos que podem aparecer em sua prova. Você não precisa decorar o nome de cada um, apenas saber interpretá-los.



Gráfico de dispersão



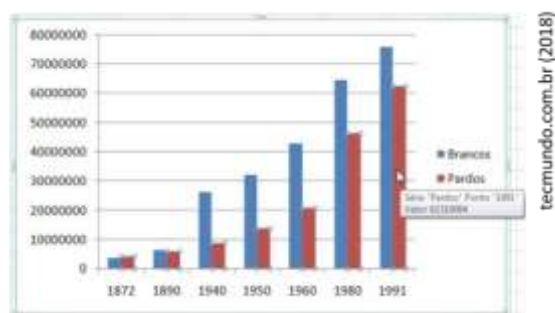
Gráfico de pizza

² RORATO, F.; DEGASPARI, C. H.; MOTTIN, F. Avaliação do nível de conhecimento de consumidores de produtos diet e light que frequentam um supermercado de Curitiba. Visão Acadêmica, v. 7, n. 1, 2006.





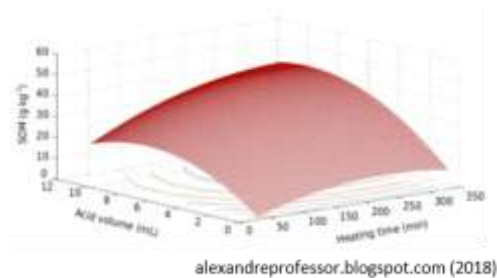
Gráfico em linha para séries temporais



Gráficos de barras para comparação de dados



Histograma de frequência



Superfície de resposta

Falaremos um pouco mais sobre histograma de frequência nessa aula adiante.

3.1 – População, amostra e amostragem

Dois conceitos, segundo Bruns (2010)⁵, importantes da estatística descritiva são:

População: qualquer seleção de indivíduos ou valores, finita ou infinita.

Amostra: uma parte da população, normalmente selecionada com objetivo de se fazer inferências* sobre a população.

**inferência estatística é o ato de fazer afirmações sobre a população ou tomar decisões a partir dos valores obtidos de uma amostra representativa. Para se fazer uma inferência, admite-se que a amostra e a população guardam similaridades fortes entre si quanto às suas características e proporções.*

Se vasculharmos diferentes livros, perceberemos que há variações para esses dois conceitos, mas que na verdade dizem a mesma coisa. É muito comum encontrarmos os termos “universo” e “universo estatístico” como sinônimo de população. Isso porque quando se fala em população precisamos notar que estamos nos referindo a todos os indivíduos (elementos) que guardam uma característica em comum. A delimitação da

⁵ NETO, Benício Barros; SCARMINIO, Ieda Spacino; BRUNS, Roy Edward. Como Fazer Experimentos: Pesquisa e Desenvolvimento na Ciência e na Indústria. Bookman Editora, 2010.



população dependerá do interesse do estudo ou investigação. **Importante notar que a população poder ser finita ou infinita.**

As pessoas que estão realizando esse curso, os moradores de Gramado-RS, e todas as 50.063.860 possibilidades de jogos na MEGA SENA são **exemplos de populações finitas**. Por outro lado, os resultados (cara ou coroa) em sucessivos lançamentos de uma moeda e os números reais são **exemplos de populações infinitas**.

Imaginemos que quiséssemos saber a idade média de uma população: Caso nossa população sejam todas as pessoas que estejam realizando esse curso, então seria fácil coletar os dados e calcular a idade média. Entretanto, imaginemos que nossa população passe a ser todos moradores de Goiânia-GO, isso ficaria muito (muito) mais difícil. Uma alternativa mais viável é selecionarmos uma amostra de moradores de Goiânia e, a partir dessa amostra, estimar a idade média do Goianiense. E é exatamente isso que acontece em pesquisas eleitorais, não é mesmo?

Ora! Se pretendemos fazer estimativas e inferências sobre a população a partir de uma amostra, então o ideal é que tenhamos uma **amostra representativa**, cujo conceito está abaixo. Em laboratórios e em campos experimentais quase sempre é possível obter uma **amostra representativa** a partir de uma **amostragem aleatória**. Portanto o ideal para nós será obter uma **amostra representativa e aleatória**. Veja as definições abaixo, segundo Bruns (2010).

Amostra representativa: apresenta as características relevantes da população na mesma proporção em que elas ocorrem na própria população.

Amostra aleatória: amostra de N valores ou indivíduos obtidos de tal forma que todos os possíveis conjuntos de N valores da população tenham a mesma chance de ser conhecidos.

Para obtenção de uma amostra, se faz necessário realizar uma **amostragem** da população. Vamos desenvolver esse conceito já o relacionando com análises química, pois é nesse contexto que o assunto poderá aparecer em sua prova.

Suponha que haja a suspeita que uma indústria têxtil esteja despejando, em um rio, concentrações de cromo acima do permitido pela legislação. Necessitaríamos, então, realizar uma **amostragem** do líquido efluente (água + contaminantes) da indústria e enviar a amostra para o laboratório. Para tanto, uma estratégia seria coletar pequenos volumes do efluente (amostras) em diferentes horários, **juntar todos os volumes (amostras), formando uma amostra composta** e enviar para o laboratório.

Já no laboratório, o laboratorista não irá utilizar todo líquido para realizar a determinação de cromo. Por isso, ele irá homogeneizar a solução por meio da sua agitação e, em seguida, **retirar uma alíquota (realizar uma amostragem)** para análise. Deste modo, a amostragem está presente desde a coleta da amostra, em campo, até a rotina analítica. Veja na tabela abaixo os diferentes tipos de amostragem.





Tipos de amostragem

Amostragem aleatória ou casual: cada indivíduo da população possui a mesma probabilidade de ser selecionado. Suponhamos que, em uma indústria química, são produzidos diariamente 1000 frascos de hidróxido de sódio P.A. e que se definiu utilizar uma amostragem aleatória de 5 frascos/dia para análise de controle de qualidade. Para tanto, sorteia-se 5 números aleatórios entre 1 e 1000, por ex: 158; 353; 401; 585; e 975, e durante a produção diária, retira-se esses frascos e os envia ao laboratório.

Amostragem sistemática: retira-se indivíduos em intervalos constantes (k), ou seja, a cada k elementos. No mesmo exemplo da produção de hidróxido de sódio, caso fosse definida a utilização de uma amostragem sistemática, com $k = 100$, iniciando a amostragem pelo frasco 1, então seriam coletados os frascos 1, 101, 201, ..., 901. No exemplo do efluente contaminado com cromo da indústria têxtil, estaríamos realizando uma amostragem sistemática caso decidíssemos coletar amostras a cada 1 hora.

Amostragem estratificada: de início, divide-se a população em estratos ou grupos e, na sequência, realiza-se a amostragem em cada grupo separadamente. Por exemplo, um agricultor gostaria de avaliar a fertilidade do solo de sua chácara, antes do plantio, para verificar a necessidade de realizar alguma correção com insumos agrícolas a fim de aumentar sua produtividade. Em sua propriedade rural, é possível identificar três tipos de solos diferentes por meio da coloração: um mais avermelhado (mais argiloso), outro de tonalidade mais escurecida (mais rico em matéria orgânica) e outro mais amarelado (solo mais arenoso). Nesse caso, o ideal é que o agricultor considere cada tipo de solo como um estrato da população e colete amostras separadamente de cada tipo de solo. No campo das ciências sociais, também haverá casos que para se obter uma amostra representativa seja necessária uma amostragem estratificada. Por exemplo, nas pesquisas eleitorais é necessário contemplar proporcionalmente todas as classes sociais de um eleitorado.

Há risco de tendências caso essa regra não seja respeitada.

Amostragem por conglomerados: de início, divide-se a população em seções ou estratos, seleciona-se algumas delas e coleta dados de todos os indivíduos das seções selecionadas. Tem maior aplicabilidade em estudos sociais e quase nenhuma aplicabilidade em química.

Amostragem por conveniência: é definida pela facilidade de coleta de dados. Por exemplo, realizar uma pesquisa entrevistando apenas os seus vizinhos. A amostragem por conveniência também é utilizada em laboratórios. Naquele primeiro exemplo sobre a fabricação de hidróxido de sódio, o laboratorista poderia transferir todo conteúdo do frasco para uma bandeja limpa e retirar pequenas quantidades do reagente em diferentes



pontos, garantindo uma maior representatividade da amostra em relação ao conteúdo total do frasco. No entanto, não raras vezes, o laboratorista uma amostra da parte mais superficial do conteúdo do frasco, pelo bocal, utilizando uma espátula, configurando uma amostragem por conveniência (mais fácil). Nesse caso, não haverá grandes prejuízos da representatividade se considerarmos que o conteúdo do frasco era homogêneo.

Obs: não deixe de estudar a questão abaixo, apresento nela discussões complementares sobre a amostragem de materiais diversos.



(AOCP - Perito Oficial Criminal/Área 05 - PCES - 2019) Os processos de amostragem devem garantir que amostras sólidas, líquidas e gasosas sejam adequadamente tratadas e coletadas a fim de serem representativas. Sobre esse tema, assinale a alternativa correta.

- A) Variações entre a coleta e a análise de líquidos são normais e não têm como ser evitadas devido ao estado líquido da amostra.
- B) Uma das dificuldades de amostragem de sólidos é a não homogeneidade devido a variações de posição das amostras.
- C) O pré-tratamento de amostras biológicas é um processo simples e quase sempre dispensável.
- D) Amostragem composta é quando a amostra contém fases líquida, sólida e gasosa simultaneamente.
- E) Amostras líquidas não devem ser estabilizadas quimicamente para não levar a erros na análise final.

Comentários

Letra A: incorreta. Amostragem de líquido apresentam variações baixas, pois, em geral, os líquidos podem ser homogeneizados antes de sua coleta.

Letra B: correta. A amostragem de sólidos é, em geral, mais crítica devido à heterogeneidade de seus grânulos, o que deve impactar sim em suas posições em um contexto maior. Para ficar mais claro, vamos pensar na amostragem de solo (terra), nele haverá partículas maiores e menores e de diferentes colorações (diferentes composições químicas), o que configura uma amostra heterogênea. Os diferentes tamanhos geram acomodações diferentes, ou seja, partículas menores tendem a decantar e se acomodar mais ao fundo do material, ao passo que as maiores ficam mais superficiais. Vale lembrar ainda que vários tipos de amostras sólidas são heterogêneas, a exemplo de solos e rochas.



Letra C: incorreta. Quase sempre é necessário algum pré-processamento das amostras, sobretudo para sua conservação. Amostras biológicas são, em geral, secas antes de sua armazenagem ou análise, pois, em ambiente úmido com substâncias orgânicas, podem proliferar fungos e bactérias que degradariam o material.

Letra D: incorreta. Amostragem composta é quando se realiza diferentes amostragem de um material ou amostra de grande extensão e reúne essas amostragens para compor uma única amostra denominada amostra composta. Para ilustrar, vamos pensar que chegou à indústria farmacêutica um caminhão de uma determinada matéria prima, podemos realizar amostragens em diferentes pontos dessa carga de matéria prima, reunir esses materiais e uma única amostra composta.

Letra E: incorreta. Em alguns casos é necessário a estabilização para conferir maior homogeneidade à amostra, possibilitando, desta forma, uma amostragem mais representativa.

Resposta: letra B

(IFR-RS - Professor/Química Analítica - IFR-RS - 2015) A amostragem é um processo de extração de uma pequena porção, que seja representativa da composição de todo o material a ser analisado, com vistas a garantir a confiabilidade de um resultado analítico. A amostragem tem relação direta com o estado físico dos materiais, de modo que, para a preparação de uma amostra, diferentes procedimentos são adotados. Na sequência são apresentadas algumas afirmações a respeito desse processo.

I. Amostragem casual é o melhor método de amostragem a ser adotado, pois permite que todo o material seja analisado.

II. Uma amostra representativa permite estimar uma ou mais características para a totalidade de uma população.

III. Em líquidos, a homogeneidade do material a ser amostrado pode ser garantida pelo processo de agitação.

IV. Amostras sólidas sempre são homogêneas e não requerem procedimentos prévios à coleta.

Assinale a alternativa em que todas as afirmativas estão CORRETAS:

- a) Apenas III e IV.
- b) Apenas II e IV.
- c) Apenas I, II e III.
- d) Apenas I, II e IV.
- e) Apenas II e III.

Comentários

Afirmativa I: incorreta. Nem sempre a amostragem aleatória ou casual é o melhor método de amostragem. Em alguns casos, por exemplo, será necessária uma amostragem estratificada



para obter uma maior representatividade da amostra em relação à população. Em suma, a melhor técnica de amostragem dependerá da aplicação.

Afirmativa II: correta. Estudamos que uma amostra representativa apresenta as características relevantes da população na mesma proporção em que elas ocorrem na própria população.

Afirmativa III: correta. Uma das maneiras de se garantir a homogeneidade de uma amostra líquida é por meio da agitação.

Afirmativa IV: incorreta. Suponhamos uma amostra de solo: haverá partículas maiores e menores e de diferentes colorações, o que configura uma amostra heterogênea. Vários tipos de amostras sólidas são heterogêneas, a exemplo de solos e rochas.

Resposta: letra E

3.2 – Medidas de posição ou de tendência central

Conforme já discutido, a estatística descritiva é utilizada para sintetizar e organizar os dados para que fiquem mais facilmente entendíveis. Para tanto, lança mão de variáveis quantitativas (numéricas) denominadas medidas resumo, que se dividem em dois subtipos:

Medidas de posição: são variáveis quantitativas que indicam a posição do conjunto de observações (da amostra).

Medidas de dispersão: são variáveis quantitativas que indicam a dispersão, ou espalhamento, do conjunto de valores da amostra.

Nesta seção vamos nos ater a três das principais **medidas de posição ou de tendência central**: média aritmética, mediana e moda. Bussab e Morettin (2010)⁶ apresentam definições simples e suficientes para essas três medidas de posição. Essas definições estão apresentadas abaixo:



Medidas de posição ou de tendência central

Média aritmética (média, $\bar{x}$): é a soma das observações (x_i), dividida pelo número (n) delas. Usa-se símbolo μ quando a média aritmética for relativa aos dados populacionais

⁶ MORETTIN, P. A.; BUSSAB, W. O. Estatística Básica. Editora Saraiva, 2010.



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

Caso tenhamos repetições dos valores das observações de diferentes indivíduos, do tipo x_1 aparece n_1 , x_2 aparece n_2 , então, a equação acima pode ser reescrita como:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{n} = \frac{1}{n} \sum_{i=1}^n n_i x_i$$

A equação acima também é conhecida como **média aritmética ponderada (média ponderada)**. Em uma prova de concurso, por exemplo, em que questões de conhecimentos específicos apresentam pesos diferentes das questões de conhecimentos gerais, você pode calcular seu desempenho médio substituindo n pelo peso (p) de cada tipo de questão, como segue:

$$\bar{x} = \frac{p_1 x_1 + p_2 x_2 + \cdots + p_k x_k}{p_1 + p_2 + \cdots + p_k} = \frac{1}{p} \sum_{i=1}^n p_i x_i$$

Por fim, se $f_i = n_i/n$ representa a frequência relativa da observação x_i , então podemos obter uma equação para a média que se relacione com a frequência (note que aqui não se divide o somatório por n como nos casos anteriores):

$$\bar{x} = f_1 x_1 + f_2 x_2 + \cdots + f_k x_k = \sum_{i=1}^n f_i x_i$$

Mediana (md): é o valor que ocupa a posição central da série de observações quando estão ordenadas em ordem crescente. Quando o número de observações for par, é necessário calcular a média aritmética das duas observações centrais.

$$\text{Se } n \text{ ímpar, } md = x_{\left(\frac{n+1}{2}\right)}$$

$$\text{Se } n \text{ par, } md = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Moda: é o valor mais frequente do conjunto de valores observados. Uma forma fácil de lembrar do conceito é relacioná-lo a uma roupa que "todo mundo" está usando porque está na moda.



Para melhor entendimento, vamos aplicar esses conceitos em uma situação prática: imaginemos que estamos interessados em analisar a faixa salarial de uma microempresa que têm 10 funcionários cujas remunerações estão dispostas na tabela abaixo.

1000,00	1500,00
1000,00	1500,00
1500,00	2000,00
1500,00	2000,00
1500,00	20.000,00(*)

(*) Salário do gerente.

Utilizando a primeira equação apresentada para média ($\bar{x}$), podemos calcular a média salarial da tabela acima, como segue:

$$\bar{x} = \frac{1000 + 1000 + 1500 + 1500 + 1500 + 1500 + 1500 + 2000 + 2000 + 20000}{10} = 3350 \text{ reais}$$

Também podemos utilizar a segunda equação, contabilizando quantas vezes cada determinado salário aparece, como segue:

$$\bar{x} = \frac{2 \cdot 1000 + 5 \cdot 1500 + 2 \cdot 2000 + 1 \cdot 20000}{2 + 5 + 2 + 1} = 3350 \text{ reais}$$

Por fim, podemos ainda calcular a frequência de cada salário, por exemplo, a frequência do salário R\$1000,00 é 2/10 (2 salários em 10 totais). E, em seguida, calcular a média salarial a partir de sua equação que se relaciona com a frequência, como segue:

$$\bar{x} = \frac{2}{10} \cdot 1000 + \frac{5}{10} \cdot 1500 + \frac{2}{10} \cdot 2000 + \frac{1}{10} \cdot 20000 = 3350 \text{ reais}$$

Note que todas as equações produziram o mesmo resultado. A conveniência definirá a utilização entre uma ou outra fórmula, ou seja, dependerá dos dados fornecidos pelo enunciado.

O proprietário dessa empresa poderia alegar que a média salarial de sua empresa, 3.350,00 reais, está acima da média salarial do brasileiro que foi de 2.112,00 para o ano de 2017, segundo IBGE. Entretanto, avalie com mais cautela a tabela e responda: **você acha que a média obtida representa bem a faixa salarial dessa empresa?**... Se respondeu não, eu concordo com você, pois temos 9 de 10 funcionários recebendo até 2.000,00 reais, abaixo da média salarial do brasileiro. O que está puxando a média salarial para cima é, sem dúvida, o salário do gerente de 20.000,00 reais. Nesse caso, quando há uma amostra muito destoante, podemos estar diante de uma **amostra anômala** ou **outlier**.

Não nos interessam os motivos que levaram o proprietário a remunerar tão bem o gerente da microempresa. Entretanto, se o empresário for justo ao mencionar a faixa salarial de sua empresa, deverá



mencionar a mediana ao invés da média. Então, utilizando a equação da **mediana** e como n é um número par, então obtemos a mediana a partir da média do quinto e sexto salários, em ordem crescente, obtendo 1500,00 reais que é um valor mais condizente com a faixa salarial dos funcionários. Ainda nesse exemplo, a **moda** também é 1500,00 reais por ser o valor que mais se repete.

Em cada situação ou investigação uma dessas 3 medidas de posição pode ser mais adequada.

3.3 – Medidas de dispersão

Para entendermos as **medidas de dispersão**, de início, vamos imaginar que quiséssemos obter medidas que indiquem o grau de flutuação ou dispersão ou de variabilidade dos valores em torno da média. Para tanto, a medida mais imediata é o erro.

Erro (e): é a distância (diferença) entre um dado valor e a média ou entre um valor e o valor real ou verdadeiro (por ex: valor obtido por determinação analítica; valor certificado), conforme demonstrado abaixo.

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

$$e_i = x_i - x_{(\text{real})}$$

Em estatística, o erro também é conhecido como resíduo.

O erro também pode ser expresso em termos percentuais (erro relativo), o qual pode ser calculado pela expressão a seguir:

$$e_i(\%) = \frac{e_i}{x_{\text{real}}} \cdot 100$$

Vamos considerar que se pesou 7 caroços de feijão, retirados de um saco de feijão, cujas massas estão apresentadas na tabela abaixo. Na mesma tabela, adiciono a coluna de erros, também conhecido como resíduos, e outra coluna do resíduo, à qual será importante na discussão que segue.

Pesos de caroços extraídos aleatoriamente de um pacote
de feijão, em gramas.



	Peso (grama)	Erro (grama) $(x_i - \bar{x})$	Erro ² (grama ²) $(x_i - \bar{x})^2$
	0,1188	-0,091	0,008
	0,2673	0,058	0,003
	0,1795	-0,030	0,001
	0,2369	0,027	0,001
	0,1826	-0,027	0,001
	0,1860	-0,023	0,001
	0,2045	-0,005	0,000
Total	1,257	0,000	0,015
Média	0,209	0,000	0,002

Calculando a diferença entre o **valor máximo** (0,2673 g) e **mínimo** (0,1188 g), obtemos a **amplitude** 0,1485 g, que é outra medida de dispersão.

Amplitude (A): é a distância (diferença) entre o valor máximo e o valor mínimo para um conjunto de observações (amostra).

$$A = x_{\text{máx}} - x_{\text{mín}}$$

Retomando a tabela com os pesos de caroços de feijão, note que a coluna **erro** ou **resíduo** nos dá uma noção da variabilidade dos dados em torno da média. Se o erro for positivo ($e > 0$), então o valor observado está acima da média; caso o erro seja negativo ($e < 0$), o valor observado é menor que a média. Entretanto, essa medida de dispersão é útil apenas pontualmente, ou seja, para cada valor (indivíduo). Isso porque sempre obteremos o resultado 0, o caso façamos o somatório dessa coluna. Isto é, isso ocorrerá para qualquer conjunto de valores estudado, pois os resíduos negativos anularão os resíduos positivos, impossibilitando então calcularmos um valor total dos resíduos e, consequentemente, um valor médio.

É muito importante obtermos **medidas de dispersão** média, ou que resumem as séries de dados (valores), para n repetições ou indivíduos ou, no nosso caso, caroços de feijão. Reparemos então na próxima coluna da tabela anterior, **quadrado do resíduo**, $(x_i - \bar{x})^2$. Quando usamos esse artifício de elevar o resíduo ao quadrado, o sinal negativo sumirá dos valores de resíduos negativos e o somatório (total) dessa coluna não será mais zero (0,0). BINGO!!! Agora podemos calcular a média dessa coluna que será então uma **medida de dispersão** que resume uma série de dados, chamada **variância**.

Variância: medida resumo e média da variabilidade dos dados obtida a partir dos quadrados dos resíduos.

Variância amostral (s^2):

$$s^2 = \frac{e_1^2 + e_2^2 + \dots + e_n^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$



Variância populacional (σ^2):

$$\sigma^2 = \frac{e_1^2 + e_2^2 + \dots + e_n^2}{n} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

CUIDADO: Note que o símbolo da variância e o denominador do lado direito são diferentes quando se trata de dados amostrais ou populacionais.

A motivação da diferença no denominador n para $n - 1$ tem a ver com o conceito de **graus de liberdade**, que embora seja um termo muito comum, não é um conceito simples, pois se origina da álgebra linear e está relacionado com independência de vetores. Não se preocupe com esse assunto, apenas lembre-se da diferença sutil entre as duas equações.

Embora a **variância** seja muito apropriada para indicar a dispersão dos dados, ela apresenta inconvenientes de apresentação (*marketing*) e as vezes de seu entendimento numérico. Vamos entender isso na prática: imagine que uma lanchonete venda 4 tipos de bolos (variados por sabor e tipo de cobertura) que custassem 0,80; 0,90; 1,50 e 2,00 reais e um consumidor esteja interessado em saber a faixa que gastaria diariamente comendo um pedaço de bolo por dia. Pois bem, saiba que uma **medida de posição** e uma **medida de dispersão** podem resumir satisfatoriamente a série de dados e responder à pergunta do consumidor. Usemos a **média** seguida da **variância populacional** e obteremos 1,30reais $\pm$ 0,48 reais². Perceba que a dispersão, variância, está em uma dimensão (reais²) diferente da dimensão dos preços (reais). Isso acarreta uma certa confusão de qual faixa de preço se gastaria ao comer um bolo por dia. Isso se torna mais grave para o cliente que não está preocupado com os conceitos estatísticos envolvidos.

Uma alternativa para esse inconveniente é calcular a raiz quadrada da variância e assim a unidade voltará para dimensão original da variável medida, 1,30reais $\pm$ 0,69 reais. A raiz quadrada da variância corresponde a mais uma medida de posição: o **desvio padrão**.

Desvio padrão: raiz quadrada da variância.

Desvio padrão amostral (s):

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

Desvio padrão populacional (σ):

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

Em alguns casos, expressar o desvio padrão em porcentagem pode ser interessante. Para tanto, definiu-se uma nova medida de dispersão chamada de **coeficiente de variação** ou **desvio padrão relativo**.



Coeficiente de variação (CV): é uma medida de dispersão padronizada em relação à média (populacional ou amostral, a depender dos dados fornecidos), expressa em porcentagem. Também conhecido como desvio padrão relativo.

Coeficiente de variação amostral:

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

Coeficiente de variação populacional:

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

NOTA: Considerando que a padronização é realizada em relação à média, então um CV 20% significa que os dados variam, em média, 20% do valor da média em torno dela própria.

Na abordagem das medidas de dispersão, falamos de caroços de feijão e venda de bolos. Agora, vamos ver como as medidas resumo (de posição e de dispersão) se aplicam em química, resolvendo alguns exercícios de concurso.



(NUCEPE - Perito Criminal/Química - PCPI - 2018) A intoxicação por nitrato pode levar a um aumento severo da taxa de oxidação da hemoglobina a metahemoglobina, que é incapaz de ligar-se e transportar oxigênio. Por conta disto, o monitoramento da concentração de nitrato na água destinada ao consumo humano é muito importante. Nesta perspectiva, foram coletadas dez amostras de água de um bebedouro público e procedeu-se à determinação do teor de nitrato, por meio do método simplificado do salicilato, com auxílio de um espectrômetro UV-visível. A partir dos dados da Tabela seguinte, indique a alternativa que apresenta os valores corretos da média da concentração e desvio-padrão, respectivamente.



Amostras	Conc. NO_3^- ($\mu\text{g/mL}$)
1	0,527
2	0,523
3	0,525
4	0,524
5	0,525
6	0,516
7	0,526
8	0,499
9	0,528
10	0,561

- a) 0,525 e 0,015
- b) 0,508 e 0,002
- c) 0,520 e 0,021
- d) 0,512 e 0,009
- e) 0,518 e 0,032

Comentários

De início, é necessário identificar se os dados são amostrais ou populacionais. Neste exercício, o próprio enunciado já afirma que se trata de dados amostrais. Via de regra, valores de ensaios laboratoriais são dados amostrais. Embora o conceito original da estatística diga que uma amostra é um conjunto de indivíduos da população, no âmbito da química, é uma prática consensual utilizar o termo "amostra" sinônimo de "indivíduo". Por isso, não se aflija com isso e não julgue um item incorreto quando o termo amostra aparecer no lugar de indivíduo.

Uma vez identificado que os dados são amostrais, podemos aplicar os valores fornecidos nas equações da média aritmética ($\bar{x}$) e do desvio padrão amostral (s), como segue:

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

$$\bar{x} = \frac{0,527 + 0,523 + 0,525 + 0,524 + 0,525 + 0,516 + 0,526 + 0,499 + 0,528 + 0,561}{10}$$

$$\boxed{\bar{x} = 0,525}$$

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

$$s = \sqrt{\frac{(0,527 - 0,525)^2 + (0,523 - 0,525)^2 + \dots + (0,561 - 0,525)^2}{10 - 1}}$$



$$s = 0,015$$

Resposta: letra A

(Adaptado de UEPI - Perito Criminal/Química - PCPI - 2008) Utilizando a metodologia de volumetria de complexação, foi feita uma análise do leite e determinada a quantidade de cálcio presente no mesmo. Os resultados obtidos, em 100 ml de leite, foram os seguintes:

Quantidade de Cálcio (mg):
128,6
129,4
127,8

O leite trazia na embalagem um valor de 134,0 mg de cálcio em 100 ml de leite. A partir destes dados, calcule o erro absoluto, o erro relativo e o desvio padrão, para a determinação em triplicata, e assinale a alternativa que apresenta, respectivamente, estes valores:

- a) 4,0 %; 5,4 mg; 0,53 mg
- b) 3,2 mg; 5,4 %; 0,8 mg
- c) 5,4 mg; 4,0 %; 0,8 mg
- d) 5,4 mg; 4,0 %; 0,53 mg
- e) 3,2 %; 4,0 mg; 1,6 mg

Comentários

De início, devemos calcular a média ($\bar{x}$) para, em seguida, utilizar seu valor no cálculo do erro absoluto (e), erro relativo ($e(\%)$) e desvio padrão (s), conforme demonstrado abaixo:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\bar{x} = \frac{128,6\text{mg} + 129,4\text{mg} + 127,8\text{mg}}{3}$$

$$\bar{x} = 128,6 \text{ mg}$$

Vale notar que os termos acima e abaixo da média são exatamente simétricos em relação à média, distanciando dela em 0,8 mg. Nesses casos em que os dados são simétricos, a média é igual a mediana. Desta forma, o cálculo pode ser simplificado, mas CUIDADO, para usar o valor da mediana como valor da média você precisará se certificar de que os dados são realmente simétricos. Por exemplo, em uma série com 7 valores ($v_1, v_2, \dots, v_7$), você precisará estar certo que os pares v_1 e v_7 ; v_2 e v_6 ; v_3 e v_5 são simétricos entre si em relação ao termo central v_4 .

Devemos considerar o valor fornecido pela embalagem (134 mg) como sendo o valor esperado ou real, enquanto a média 128,6 mg dos valores determinados como sendo o valor unitário medido. Aplicando à fórmula do erro absoluto e à fórmula do erro relativo, temos:



$$e_i = x_i - x_{(\text{real})}$$

$$e_i = 128,6 \text{ mg} - 134 \text{ mg}$$

$$e_i = -5,4 \text{ mg}$$

$$e_i(\%) = \frac{e_i}{x_{\text{real}}} \cdot 100$$

$$e_i(\%) = \frac{-5,4 \text{ mg}}{134 \text{ mg}} \cdot 100$$

$$e_i(\%) = -4,0\%$$

Por fim, o desvio padrão depende apenas dos valores determinados, pois é uma medida da dispersão dos valores entre repetições ou entre indivíduos. Aplicando a média a média ($\bar{x}$) anteriormente calculada e os valores individuais fornecidos na tabela, temos:

$$s = \sqrt{\frac{(128,6 - 128,6)^2 + (129,4 - 128,6)^2 + (127,8 - 128,6)^2}{3 - 1}}$$

$$s = 0,8 \text{ mg}$$

Resposta: letra C

4 – Erros em análises químicas

Decidi dedicar um capítulo da nossa aula aos erros, pois além de serem muito cobrados, é um tema cujos desdobramentos são relativamente extensos.

4.1 – Três tipos de erro

Algumas literaturas classificam os erros em dois tipos: **sistemático** e **aleatório**. Entendo que incluir um terceiro tipo de erro, o **grosseiro**, seja mais apropriado. Vamos agora entender cada um dos três tipos de erro.

Erro grosseiro: suponhamos que durante a realização de uma titulação ácido-base, o laboratorista se distraia e não adicione o indicador. A mudança de cor da solução (ponto de viragem) nunca ocorrerá. Esse é um baita erro, gentilmente chamado de grosseiro. Podemos citar inúmeros outros exemplos de erros grosseiros que podem acontecer em um laboratório: esquecer de zerar (tarar) a balança antes da pesagem; usar vidraria volumétrica diferente da requerida; usar reagente diferente do requerido, dentre outros. Certamente, todos que trabalham em laboratório poderiam aumentar essa lista de exemplos. Esses erros não são sanáveis (contornáveis), ou seja, os resultados de experimentos realizados em erro grosseiro não



são aproveitáveis. Nesses casos, a saída é repetir o ensaio e ficar mais atento para não incorrer no mesmo tipo de erro novamente.

Erro sistemático: sabemos que uma balança analítica deve estar devidamente nivelada e calibrada para sua utilização rotineira. Imaginemos que uma balança sofra um impacto que a descalibre e, a partir daí, passe a produzir um erro positivo de 0,2 gramas. Se várias pesagens forem realizadas nesse dia, então estará ocorrendo um erro sempre na mesma direção, seja para mais ou para menos, o que denominamos de um **erro sistemático**.

Outros exemplos de erros sistemáticos: reagente com concentração adulterada; utilização do indicador errado; utilização de vidraria volumétrica descalibrada; observação do menisco a partir de um ângulo incorreto que é conhecido como erro de paralaxe. Cada situação dessas irá deslocar os resultados em uma direção.

É claro que o erro sistemático também pode ser evitado, desde que aplicado um pequeno esforço. No primeiro exemplo, uma verificação diária da balança com um conjunto de pesos padrão poderia ter evitado o problema. Para cada situação de erro sistemático, existe uma ação preventiva a ser adotada: utilizar reagentes dentro do prazo de validade, utilizar vidrarias volumétricas calibradas, seguir rigorosamente as orientações de leitura de menisco, dentre outros exemplos. **E aí, seguindo todas as orientações dos manuais e as boas práticas laboratoriais, isentando as análises de erros grosseiros e sistemáticos, temos um final feliz, pois os processos estarão controlados.**

Será que tudo está perdido quando se comete um erro sistemático? Felizmente não! Em algumas situações, é possível recuperar os dados contaminados por erros sistemáticos. Mas isso só é possível quando o experimentador consegue identificar com 100% de certeza quando o erro começou e qual foi o distanciamento do valor experimental. Para o exemplo da balança descalibrada, caso seja identificado que a massa está sendo superestimada em 0,2 gramas, por meio da utilização de pesos padrão, e que esse erro é constante para toda faixa de pesagem, então podem ser realizadas as correções das pesagens daquele dia.

Erro aleatório: também chamado de **erro indeterminado** e de **erro randômico**. Segundo Harris (2013)⁷, esse tipo de erro surge dos efeitos de variáveis que não estão sendo controladas (e que talvez não possam ser controladas) nas medidas. A probabilidade (a chance) de o erro aleatório ser positivo ou negativo é a mesma. Imaginemos que aquela balança foi novamente calibrada, nivelada e tarada ("tudo dentro do conforme"). Cinco pesagens repetitivas do mesmo peso padrão de 1,0 grama geraram os seguintes resultados: 1,0001; 0,9998; 0,9999; 1,0000; 1,0001 gramas. Nesse caso, os valores apresentam uma flutuação mas concentram-se em torno de certo valor intermediário, que nesse caso, deve ser o valor certificado do peso padrão. Essa flutuação natural e intrínseca é, portanto, o **erro aleatório** associado à pesagem dessa balança analítica.

Outros exemplos de erros aleatórios:

- Repetidas aferições de menisco, para um mesmo volume, realizadas pela mesma pessoa, de forma correta. Haverá variações entre as repetições realizada pela mesma pessoa, além do erro associado à vidraria volumétrica;

⁷ HARRIS, DANIEL C. Análise Química Quantitativa, 5ª. Ed., Rio de Janeiro, Editora Livros Técnicos e Científicos, 2001.



- Pequenas variações referentes a repetidas determinações por titulometria para uma amostra líquida homogênea;
- Variações instrumentais decorrente de oscilações da corrente elétrica. Por exemplo, um espectrofotômetro UV-VIS pode apresentar absorbâncias ligeiramente diferentes em leituras repetidas para uma mesma solução.

Cuidado! Caso uma repetição destoe muito das demais, então não estaremos diante de um erro aleatório e sim de um erro sistemático ou grosseiro.



(CESPE - Perito Criminal Federal/Química - PF - 2013) O teor de ferro em dois frascos (A e B) de um medicamento, presente na forma de sulfato ferroso, foi determinado por meio de volumetria de oxidação-redução, utilizando-se KMnO_4 0,01 mol/L como titulante, em meio ácido. O frasco A estava lacrado e o frasco B encontrava-se aberto, com menos da metade de seu volume inicial. Para a análise, 5,00 mL do medicamento de cada frasco foram diluídos em água e o volume, completado para 100,00 mL. Em seguida, alíquotas de 30,00 mL das soluções resultantes foram acidificadas com H_2SO_4 e diretamente tituladas em triplicata com o KMnO_4 . A tabela abaixo mostra os volumes médios de permanganato gastos e seus desvios padrões em mL.

frasco	volume médio de KMnO_4 (mL)	desvio padrão (mL)
A	15,0	$\pm 0,2$
B	12,5	$\pm 0,1$

Com base nessas informações, é correto afirmar que a concentração determinada para o frasco B apresentou um erro sistemático positivo significativo, no nível de 95% de confiança, em relação a obtida para o frasco A.

Comentários

A priori não sabemos qual dos frascos apresenta o resultado correto. Entretanto, o enunciado indica que devemos calcular o erro de B em relação a A, ou seja, devemos considerar o valor de A como verdadeiro. Para simplificar a resolução da questão, vamos calcular o erro em relação ao volume de titulante KMnO_4 , como segue:

$$e_i = x_i - x_{(\text{real})}$$

$$e_i = 12,5 - 15,0 = -2,5 \text{ ml}$$



Para encontrarmos o erro na determinação, deveríamos transformar o erro do volume (-2,5 ml) em concentração por meio de cálculos estequiométricos. Entretanto, nesse exercício, isso não se faz necessário, pois o erro de B em relação A é negativo, o que torna a questão errada.

Resposta: errado

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2011) A realização de dez determinações repetidas da concentração de manganês em um material certificado, cujo teor nominal de Mn é de $1,25 \cdot 10^3 \mu\text{g} \cdot \text{g}^{-1}$, forneceu resultados de $1,20 \cdot 10^3 \mu\text{g} \cdot \text{g}^{-1}$, em média, com desvio padrão de $2,40 \cdot 10 \mu\text{g} \cdot \text{g}^{-1}$. O coeficiente de variação e o erro relativo desse ensaio foram, respectivamente,

- a) 5,0% e 2,0%
- b) 4,0% e 2,0%
- c) 2,0% e 5,0%
- d) 2,0% e 4,0%
- e) 2,0% e 2,0%

Comentários

Aplicando os valores nas equações do erro absoluto e relativo, respectivamente, temos:

$$e_i = X_i - X_{(\text{real})}$$

$$e_i = 1,20 \cdot 10^3 - 1,25 \cdot 10^3$$

$$e_i = 0,05 \cdot 10^3 = 50 \mu\text{g} \cdot \text{g}^{-1}$$

$$e_i(\%) = \frac{e_i}{X_{\text{real}}} \cdot 100$$

$$e_i(\%) = \frac{50}{1,25 \cdot 10^3} \cdot 100$$

$$e_i(\%) = 4,0\%$$

Para o cálculo do **coeficiente de variação amostral**, $CV(\%)$, devemos utilizar o desvio padrão amostral (s) $2,40 \cdot 10 \mu\text{g} \cdot \text{g}^{-1}$ e a média $1,20 \cdot 10^3 \mu\text{g} \cdot \text{g}^{-1}$ dos resultados fornecidos. Aplicando esses valores à equação do CV , temos:

$$CV(\%) = \frac{s}{\bar{X}} \cdot 100$$

$$CV(\%) = \frac{2,40 \cdot 10 \mu\text{g} \cdot \text{g}^{-1}}{1,20 \cdot 10^3 \mu\text{g} \cdot \text{g}^{-1}} \cdot 100$$

$$CV(\%) = 2\%$$



Resposta: letra D

4.2 – Diferença entre precisão e exatidão

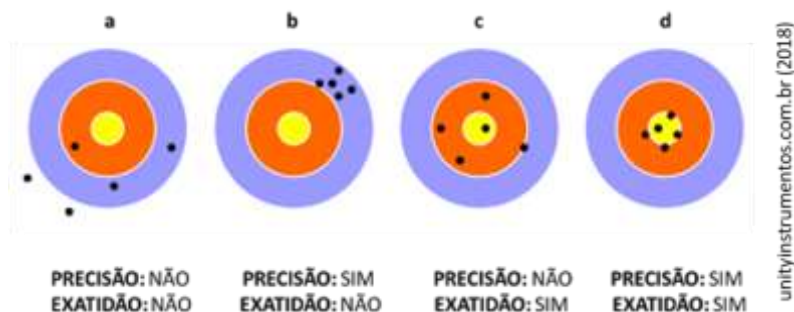
A **precisão** está relacionada à reprodutibilidade de um método analítico ou de medição de um instrumento ou vidraria. Reprodutibilidade, por sua vez, de maneira bem objetiva, corresponde ao quão próximos se encontram as repetições. Por exemplo, estaríamos diante de uma determinação titulométrica precisa (reprodutível) caso fossem gastos 4,92; 4,93 e 4,92 mL de volume do titulante. O **desvio padrão** é um parâmetro que indica a **precisão**, sendo maior a precisão quanto menor for o desvio padrão.

Por outro lado, a **exatidão** está relacionada com a diferença entre o valor medido (experimental) e o valor real ou verdadeiro. Desta forma, a **exatidão** se relaciona com o erro. Maior será a exatidão, quanto menor for o erro. Valores certificados por organizações regulatórias como o NIST (*National Institute of Standards and Technology*) são exemplos de valores considerados verdadeiros. Suponhamos que nosso laboratório desenvolva um novo método de determinação de nitrogênio total em feijão. Podemos, então, comprar do NIST uma amostra certificada de feijão cujo teor de nitrogênio total é conhecido. Analisando esse material certificado em nosso laboratório, podemos calcular o erro (diferença entre os teores determinado e verdadeiro), que indicará a exatidão do novo método. Outro clássico exemplo de valor verdadeiro é a concentração teórica de um padrão primário de alta pureza.

Uma balança analítica precisa e exata poderia apresentar 0,999; 1,000; 0,999; 1,001 g para 4 pesagens consecutivas de um peso de 1,0 g. De outro modo, caso apresentasse 1,251; 1,249; 1,249; 1,251 g, no mesmo exemplo anterior, teríamos uma medição precisa (valores próximos, reprodutíveis), porém, inexata (valor médio 1,250g distante do valor verdadeiro 1,000g). Se os resultados das pesagens fossem 0,948; 1,000; 1,049; 1,000g, então a medição seria imprecisa (desvio padrão mais elevado), porém, exata (resultado médio 0,999g muito próximo do valor verdadeiro). Por fim, a balança pode ser imprecisa e inexata se erro e desvio padrão forem elevados.

Avalie a figura abaixo e relacione os desenhos a, b, c e d com as observações da tabela seguinte. Note que, por meio da análise da precisão e exatidão, é possível identificar a ocorrência de erro sistemático (**Lembre-se que o erro aleatório sempre ocorrerá, mas em situações normais, ele é mantido em um nível baixo ou tolerável**).

Ilustração, em alvos, dos conceitos de precisão e exatidão, seguida de observações pertinentes.



- a) Grande dispersão de resultados. Erros aleatórios elevados. Existência de erros sistemáticos. Resultado não preciso e não exato.
- b) Baixa dispersão de resultados. Erros aleatórios pequenos. Existência de erros sistemáticos. Resultado preciso, mas não exato.



- c) Grande dispersão de resultados. Erros aleatórios elevados. Não existência de erros sistemáticos. Resultado não preciso, mas exato.
- d) Baixa dispersão de resultados. Erros aleatórios pequenos. Não existência de erros sistemáticos. Resultado preciso e exato.

4.3 – Propagação do erro aleatório

Toda medição instrumental possui um **erro aleatório** associado, muitas vezes chamado de **incerteza**. Em geral, são utilizados diferentes instrumentos e vidrarias em um mesmo método laboratorial, produzindo uma incerteza resultante que é uma combinação das certezas individuais. Precisamos compreender como calcular essa incerteza (erro aleatório) resultante.

Há duas maneiras para o cálculo da incerteza resultante: uma para **adição e subtração**; e a outra para **multiplicação e divisão**.

Suponhamos que necessitamos medir 13,0 ml de H_2SO_4 e possuímos três pipetas: uma de 10,0 ml com erro aleatório e_1 ; uma de 2,0 ml com erro e_2 ; e uma de 1,0 ml com e_3 . Caso utilizássemos as 3 pipetas para medição do volume desejado, estaríamos diante de uma **incerteza da adição**, já que os volumes medidos são somados. Nesse caso, podemos calcular a **incerteza resultante** de acordo com a equação abaixo:

$$e = \sqrt{e_1^2 + e_2^2 + e_3^2}$$

O raciocínio acima é válido para adição, conforme apresentado, e também para **subtração**. Por exemplo, o volume do titulante é obtido a partir da subtração de duas leituras de volume (final e inicial) na bureta. Considerando que a bureta apresenta um erro aleatório e_1 , então a **incerteza resultante** associada à medição pode ser calculada como segue:

$$e = \sqrt{e_1^2 + e_1^2}$$

Vamos agora entender quando estaremos diante de uma incerteza da divisão ou da multiplicação. No preparo de uma solução, pesamos uma certa massa, em balança analítica, que apresenta e_1 . Em seguida, solubilizamos a massa pesada em água e aferimos o volume final em um balão volumétrico, que apresenta e_2 . A concentração final é obtida dividindo a massa pelo volume do balão ($C = m/V$). Desta forma, a incerteza resultante está associada a uma divisão de medidas e pode ser calculada como segue:

$$\%e = \sqrt{\%e_1^2 + \%e_2^2}$$

em que $\%e$ corresponde ao erro relativo e, conforme demonstrado anteriormente, pode ser calculado pela equação abaixo:

$$\%e_i = \frac{e_i}{x} \cdot 100$$



O raciocínio é o mesmo para propagar a imprecisão. Portanto, caso quiséssemos propagar a imprecisão (desvio padrão) de três instrumentos ou vidrarias, pelo cálculo da soma ou subtração, teríamos:

$$S = \sqrt{S_1^2 + S_2^2 + S_3^2}$$

Por outro lado, a propagação do desvio padrão (imprecisão) pelo cálculo da propagação do desvio da multiplicação ou divisão, está apresentada abaixo:

$$\left(\frac{S}{X_{\text{resultante}}} \cdot 100 \right)^2 = \left(\frac{S_1}{X_1} \cdot 100 \right)^2 + \left(\frac{S_2}{X_2} \cdot 100 \right)^2 + \left(\frac{S_3}{X_3} \cdot 100 \right)^2$$

$$\left(\frac{S}{X_{\text{resultante}}} \right)^2 = \left(\frac{S_1}{X_1} \right)^2 + \left(\frac{S_2}{X_2} \right)^2 + \left(\frac{S_3}{X_3} \right)^2$$

$$S = X_{\text{resultante}} \sqrt{\left(\frac{S_1}{X_1} \right)^2 + \left(\frac{S_2}{X_2} \right)^2 + \left(\frac{S_3}{X_3} \right)^2}$$

Em uma pesquisa em provas anteriores de diferentes bancas, localizei apenas exercícios sobre propagação da incerteza em provas do CESPE.



Texto referente às questões 06 e 07:

Preparou-se 100 mL de uma solução padrão de biftalato de potássio, de acordo com o seguinte procedimento: (i) pesagem de um béquer vazio em uma balança analítica; (ii) pesagem do béquer contendo 2,0000 g do biftalato na mesma balança; (iii) dissolução do biftalato, por meio da adição de água suficiente ao béquer; (iv) transferência de todo o conteúdo do béquer para um balão volumétrico de 100,00 mL; (v) preenchimento do balão com água até o menisco.

Considerando as informações acima e que os desvios padrão sejam $\pm 0,0001$ g, para uma simples leitura na balança utilizada, e $\pm 0,01$ mL, para uma leitura de menisco no balão, julgue os próximos itens:

(CESPE - Químico - FUB - 2011) Caso a solução de biftalato preparada seja usada, posteriormente, na padronização de uma solução básica, a presença de impurezas no biftalato usado para a preparação de



sua solução gerará um erro aleatório que, conseqüentemente, afetará a precisão dos resultados obtidos durante a padronização.

Comentários

Suponhamos que seja apresentado, no rótulo de um reagente, a pureza de 99,0%, mas que de fato esteja presente 5,0% de impureza, ou seja, a pureza real do reagente é de 95,0%. Caso utilizássemos esse reagente em uma análise, confiando na pureza do rótulo, a solução teria concentração diferente da desejada. Desta forma, ao utilizarmos essa solução, seria produzido um erro sempre na mesma direção, o que corresponde a um erro sistemático. Além disso, o aumento do erro influencia na exatidão e não na precisão do método.

Resposta: errado

(CESPE - Químico - FUB - 2011) O desvio padrão relativo para a concentração da solução de biftalato de potássio preparada é maior que $1,0 \times 10^{-4}$.

Comentários

Este é um exercício sobre propagação de desvio padrão. De início, precisamos identificar se devemos aplicar o cálculo de propagação de desvios da soma e subtração ou da multiplicação e divisão. No caso da concentração, massa, pesada em balança, e volume, medido em balão volumétrico, se relacionam por meio de uma divisão ($C = m/V$), portanto devemos aplicar o cálculo da propagação de desvios da multiplicação e divisão, apresentado na equação abaixo.

$$S = X_{\text{resultante}} \sqrt{\left(\frac{s_1}{x_1}\right)^2 + \left(\frac{s_2}{x_2}\right)^2 + \left(\frac{s_3}{x_3}\right)^2}$$

Nesse caso, $X_{\text{resultante}}$ corresponde à concentração da solução, $C = m/v \rightarrow C = (2g/0,1L) = 20g/L$. Aplicando os valores na fórmula acima, temos:

$$s = 20g \cdot L^{-1} \sqrt{\left(\frac{0,0001g}{2,0g}\right)^2 + \left(\frac{0,01 \cdot 10^{-3}L}{100 \cdot 10^{-3}L}\right)^2}$$
$$s = 2,24 \cdot 10^{-3} g \cdot L^{-1}$$

Por fim, transformando o desvio padrão em desvio padrão relativo (= coeficiente de variação), temos:

$$CV(\%) = \frac{s}{\bar{x}} \cdot 100$$
$$CV(\%) = \frac{2,24 \cdot 10^{-3} g \cdot L^{-1}}{20g \cdot L^{-1}} \cdot 100$$

$$CV(\%) = 1,12 \cdot 10^{-2} \%$$

Resposta: certo



5 – Testes de hipótese

Chegou a hora de aprendermos a tomar decisões apropriadas baseada em probabilidades. Afinal, já alertamos que muitas situações rotineiras lidamos com dados experimentais não exatos, a exemplo disso temos o laboratório. É aí que as **distribuições estatísticas** podem nos auxiliar imensamente. Embora muitas pessoas se queixem da dificuldade de aprender a utilizar as distribuições estatísticas e suas respectivas tabelas, tentaremos simplificar esse aprendizado.

Antes, porém, vamos falar rapidamente de um gráfico (assunto da estatística descritiva) especialmente importante no entendimento das distribuições estatísticas, que é o **histograma de frequência**.

5.1 – Histograma de frequência

Em uma situação hipotética, realizou-se o levantamento das alturas de uma amostra de 90 alunos do Estratégia Concursos. Notou-se que os valores flutuaram em um intervalo restrito. A maior altura foi de 1,98 m, enquanto que a menor, de 1,56 m.

Identificar tendências da distribuição das alturas dos alunos, observando diretamente uma tabela com as 90 alturas, seria uma tarefa árdua. Portanto, podemos dividir todo intervalo de altura (amplitude) em intervalos menores, de 0,05 m (=5 cm), e contabilizar quantos alunos se enquadram em cada faixa de altura, conforme demonstrado na tabela abaixo.

Intervalo (m)	Nº de alunos	Frequência (*)	Frequência acumulada
1,55-- 1,60	3	0,03	0,03
1,60-- 1,65	5	0,06	0,09
1,65-- 1,70	9	0,10	0,19
1,70-- 1,75	18	0,20	0,39
1,75-- 1,80	20	0,22	0,61
1,80-- 1,85	17	0,19	0,80
1,85-- 1,90	10	0,11	0,91
1,90-- 1,95	6	0,07	0,98
1,95-- 2,00	2	0,02	1,00
Total	90	1,00	

(*) Número de alunos no intervalo dividido pelo número total de alunos

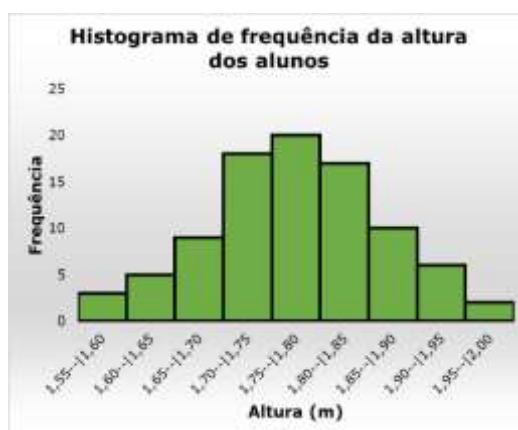
Na tabela acima, o símbolo --| significa intervalo aberto à esquerda e fechado à direita. Por exemplo, 1,55--|1,60 corresponde a altura maior (>) que 1,55 e menor ou igual ($\leq$) a 1,60 m. Em outras palavras, exclui a altura 1,55 m do intervalo e inclui 1,60m. Vale lembrar que os valores, contidos nas colunas de frequência, podem ser transformados em porcentagem caso sejam multiplicados por 100.

As frequências calculadas para os intervalos aparecem nas duas últimas colunas da tabela. É preferível analisarmos a distribuição de valores em termos de frequência, pois assim podemos determinar a probabilidade de determinados valores serem observados. Com essas probabilidades podemos testar hipóteses sobre a população (neste exemplo, a população seria todos os alunos do Estratégia Concursos), como veremos logo mais.



Em uma análise qualitativa da tabela de frequências acima, notamos que o número de alunos em uma determinada faixa aumenta com a proximidade dos valores centrais, ou seja, valores próximos à média, à mediana e à moda. Por outro lado, a frequência decresce à medida que afastamos do centro.

Fica ainda mais ilustrativo e didático analisar os dados graficamente. Portanto, podemos representar a tabela acima por meio de um **histograma de frequência**, representado na figura abaixo. A base de cada retângulo coincide com a largura do próprio intervalo, a altura corresponde ao número de indivíduos em cada intervalo ou classe, e a área é proporcional à sua frequência. No gráfico abaixo, é possível perceber que as alturas se distribuem em torno de 1,77 m e que essa frequência vai diminuindo (mais ou menos) simetricamente de acordo com que se afasta dessa altura central. Essas percepções são vantagens do gráfico em relação aos dados originalmente observados.



5.2 – Distribuição normal

O que seriam as distribuições estatísticas ou distribuições de frequência?

É um modelo (equação matemática) descritivo sobre a chance que uma variável aleatória pode apresentar ao longo de um intervalo de valores, análogo ao comportamento observado em histogramas de frequência. Já sabemos, por exemplo, que a chance de encontrarmos valores próximos à média é maior. A distribuição estatística consegue mensurar o quão maior será essa chance.

A primeira **distribuição estatística** que estudaremos é a **distribuição normal** ou **gaussiana**, justamente por ser a mais importante e a mais utilizada. Ela foi proposta no início do século XIX pelo matemático Karl F. Gauss para estimar a probabilidade de ocorrência de erro em medições. A sua utilização foi tão intensa e difundida que com o tempo foi considerada o comportamento natural (normal) dos erros experimentais de uma forma geral. Obviamente existem outros comportamentos experimentais e, portanto, outras **distribuições estatísticas** que podem ser necessárias ou mais adequadas para explicar o comportamento de uma série de dados. Em laboratórios, quase sempre podemos considerar que os dados se distribuem normalmente, isto é, conforme uma distribuição normal.

A distribuição normal é contínua, ou seja, a variável observada pode apresentar qualquer valor dentro do intervalo de estudo. No caso de uma variável normal, ela poderá assumir qualquer valor no intervalo $(-\infty, +\infty)$ (=“menos infinito a mais infinito”), ou seja, **qualquer** número real. Na prática, sabemos que não é bem assim. Por exemplo, caso estivéssemos estudando o peso das pessoas, mesmo que esses dados se

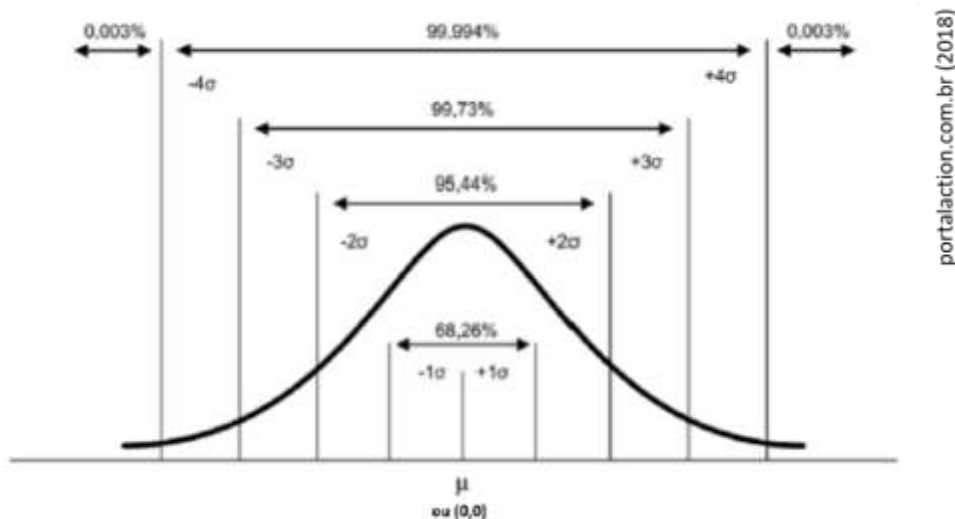


comportem normalmente, sabemos que a massa de um humano não pode crescer indefinidamente, mas isso não diminui a aplicabilidade dessa distribuição.

A equação matemática que descreve uma **distribuição estatística** é entendida como a **densidade de probabilidade**, $f(x)$, e depende sempre de uma quantidade certa de medidas ou parâmetros estatísticos. No caso da distribuição normal depende de apenas duas medidas estatísticas: **média** (μ) e **variância** (σ^2) populacionais. Ou seja, é possível conhecer o comportamento provável ou probabilístico da população em estudo com apenas esses dois parâmetros.

A equação $f(x)$ (não se preocupe em memorizá-la) e a curva obtida por essa equação para a **distribuição normal** estão demonstradas na equação e figura abaixo.

$$f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$



Distribuição de uma variável x em torno da média populacional μ com indicação de afastamento em relação à média e da probabilidade associada a alguns intervalos para x .

Percebemos, na figura acima, que a curva, em formato de sino, é perfeitamente simétrica em torno da média (ponto central). A probabilidade é máxima na média e cai quando nos afastamos dela. Esse comportamento é parecido com o observado para o **histograma de frequência** da altura de 90 alunos do Estratégia Concursos. Por isso, é plausível utilizarmos a **distribuição normal** para prever probabilidades de ocorrências para as alturas para todos os alunos do Estratégia, que é a população naquele caso.

Ainda na figura, perceba que apresentei a possibilidade do valor 0,0 no lugar da média. Isso é possível por meio de um artifício muito utilizado e conhecido em estatística, a **padronização**. A importância da sua aplicação se dá porque a padronização elimina a unidade de medida de uma variável, permitindo assim, uma melhor comparação de variáveis que estavam em escalas totalmente distintas. **Padronizar uma variável aleatória** é subtrair de cada valor a **média populacional** (μ) e dividir pelo **desvio padrão** (σ), obtendo uma nova variável aleatória adimensional chamada **Z** ou **Z_{score}**. O **Z_{score}** por sua vez indica a



distância, em número de desvios padrão, entre um determinado valor e a média, que passa a ser 0,0. Na prática, se o Z_{score} é - 2, então esse valor está abaixo da média numa distância de 2 desvios padrão (2σ). O cálculo do Z_{score} ou da **padronização** de uma variável aleatória (x_i) está representado na equação abaixo:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

Agora que entendemos o que significa a função $f(x)$, **densidade de probabilidade** e como se realiza a **padronização** de uma variável, vamos interpretar as probabilidades percentuais apresentadas na última figura acima. Se deslocarmos apenas 1 desvio padrão (σ) para mais e para menos em torno da média, engloba-se 68,26% da população. Caso façamos o mesmo com 2σ , teremos mais que 95%. Já com 3σ em torno da média, englobaremos quase 100% da população. Por isso, uma regra prática é admitir que qualquer valor que se distancie da média em mais de 3σ é um **outlier** ou **amostra anômala**.

Mas a pergunta que devemos fazer agora é *como calcular essa probabilidade a partir da densidade de probabilidade $f(x)$* ? No histograma de frequência, caso quiséssemos encontrar a probabilidade de ocorrência da altura estar dentro de um intervalo de valores, bastaria somar todas as subáreas dos retângulos contidos dentro desse intervalo. O raciocínio é o mesmo ao utilizar a densidade de probabilidade $f(x)$, só que para encontrar a área sob uma curva, precisamos calcular a integral dentro do intervalo de interesse. Para exemplificar, vamos calcular a **probabilidade (P)** de um dado valor (Z_{score}) estar entre o intervalo $(-3\sigma \text{ e } +3\sigma)$:

$$P(-3\sigma < z < +3\sigma) = P(-3\sigma \leq z \leq +3\sigma) = \int_{-3\sigma}^{+3\sigma} f(x)dx = 0,9973 \text{ (99,73\%)}$$

Isso significa também que a chance é de apenas 0,27% de selecionarmos um valor fora desse intervalo, ou seja, desse valor estar nas caudas da curva normal (distribuição normal). Se dividirmos essa possibilidade por 2, então será 0,135% para cada cauda (uma cauda para valores acima de $+3\sigma$ e outra cauda para valores inferiores a -3σ). Se quiséssemos saber a probabilidade no intervalo $(-\infty, +\infty)$, obviamente o resultado seria 1,00 (100%) que é a probabilidade total ou máxima de um evento acontecer.

Seria algo muito trabalhoso ficar calculando integrais a partir de $f(x)$, equação sigmoideal da distribuição normal, toda hora que precisássemos calcular a probabilidade para um intervalo. Felizmente existe a **tabela da área da cauda da distribuição normal padronizada**, que será fornecida como anexo a esse material e também é fornecida em provas de concurso público, quando seu uso se faz necessário.

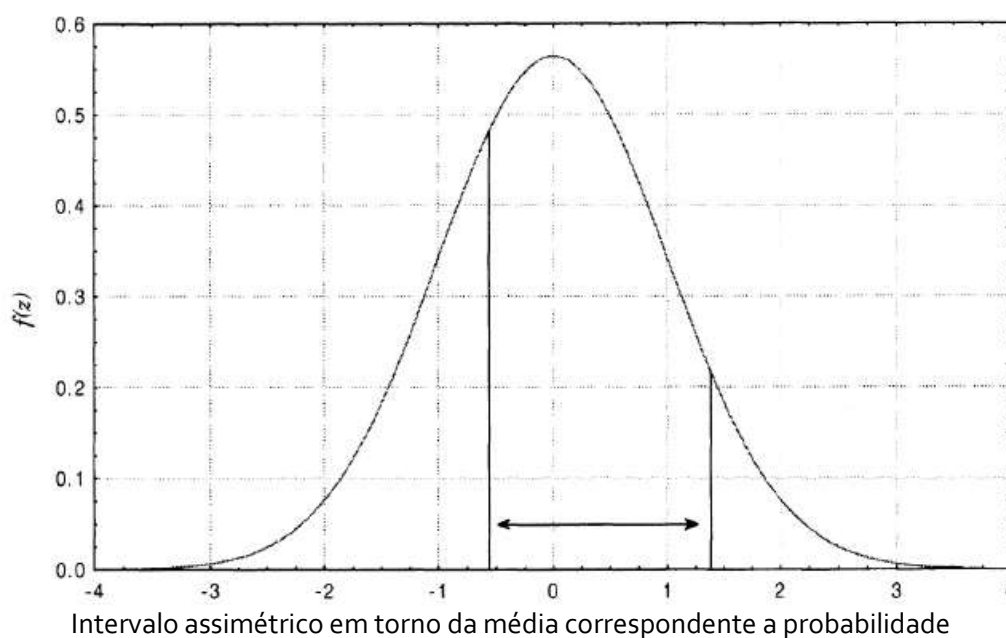
5.3 – Como usar a tabela da área da cauda da distribuição normal padronizada?

A Tabela contém o que se chama de área da cauda (à direita) da distribuição normal padrão para valores de Z que vão de 0,00 a 3,99. Acompanhe a explicação que se segue tendo em mãos a tabela Z (**final da apostila**). A primeira coluna, à esquerda, dá o valor de Z até a primeira casa decimal, enquanto a linha superior da tabela dá a segunda casa. Para saber a área da cauda correspondente a um certo Z , temos que procurar o valor localizado na interseção da linha e da coluna apropriadas.



O valor correspondente a $Z = 1,96$, por exemplo, está na interseção da linha referente a $Z = 1,9$ com a coluna encabeçada por 0,06. Esse valor, **0,0250**, é a fração da área total sob a curva que está localizada à direita de $Z = 1,96$. Como a curva é simétrica em torno da média, uma área idêntica está situada à esquerda de $Z = -1,96$ na outra metade da gaussiana, conforme demonstrado na figura apresentada na seção anterior. A soma dessas duas caudas, a da direita e a da esquerda, resulta em 5% da área total. Daí concluímos que os 95% restantes estão entre $Z = -1,96$ e $Z = 1,96$. Se extrairmos aleatoriamente um valor de Z há uma chance em cada vinte (5%) de que esse valor fique abaixo de $-1,96$, ou acima de $1,96$. Nas outras dezenove vezes a probabilidade é de que ele esteja contido nesse intervalo.

O exemplo acima foi simétrico em relação à curva, mas não necessariamente precisa ser, poderíamos estar interessados na probabilidade $P(-0,62 < Z < 1,31)$ e aí teríamos que subtrair, da área total 1,00 (=100%), as áreas das duas caudas e teríamos: $(1,0 - 0,0951 - 0,2676) = 0,6373$. Lembrando que a área da calda negativa pode ser encontrada na tabela pelo seu valor positivo de Z_{score} , já que a curva normal é simétrica. Portanto, 63,73% dos alunos do Estratégia Concursos devem apresentar alturas padronizadas entre $[-0,62; 1,31]$. Essa área corresponde ao intervalo apontado na figura abaixo.



$$P(-0,62 < Z < 1,31)^8$$

5.4 – Testes de hipótese: teste Z-score, T-student e teste F

Toda discussão introdutória sobre **distribuição normal** foi necessária para entendermos um assunto crucial em estatística, os **testes de hipóteses**. É nessa seção que iremos aprender a tomar decisões e fazer inferências apropriadas com grandes chances de estarem corretas, a partir de dados experimentais.

⁸ NETO, Benício Barros; SCARMINIO, Ieda Spacino; BRUNS, Roy Edward. Como Fazer Experimentos: Pesquisa e Desenvolvimento na Ciência e na Indústria. Bookman Editora, 2010.



Vamos supor 3 situações diferentes, cada uma com 1 pergunta ou hipótese estatística para que você entenda exatamente do que estou falando. Cada um dos três **testes de hipóteses** que discutiremos será apropriado para responder uma das perguntas.

- a) Voltando ao preço do bolo da lanchonete. Imaginemos que historicamente o consumidor saiba que o preço médio μ seja 1,30 reais com desvio padrão (σ) 0,69 reais. Um consumidor atento aos preços praticados, ao pagar 1,90 reais em um bolo, formulou a seguinte questão: **será que houve variação no preço do bolo?**
- b) Um laboratório de análise de grãos está interessado em avaliar a substituição do método colorimétrico (método 1), pelo método por espectroscopia na região do infravermelho próximo (método 2) para determinação de amilose em arroz. Para tanto, analisou-se uma mesma amostra de arroz, 5 vezes, pelos dois métodos, obtendo os seguintes resultados para teor de amilose: $24,5 \pm 1,5 \text{ g.kg}^{-1}$ (método 1); $25,5 \pm 2,0 \text{ g.kg}^{-1}$ (método 2). Com base nesses resultados, o laboratório está interessado em saber: **o resultado médio obtido pelos dois métodos é estatisticamente equivalente ao nível de 95% de confiança ou alpha (α) 0,05?**
- c) Continuando o problema do item "b". O laboratório gostaria de comparar a precisão dos dois métodos, sendo que a precisão está relacionada à variância ou desvio padrão. Portanto, dessa vez, gostaríamos de responder à seguinte questão: **a variância apresentada entre os dois métodos é estatisticamente equivalente ao nível de 95% de confiança ou alpha (α) 0,05?**

Para responder as 3 questões acima, necessitaremos dos seguintes **testes de hipótese: testes Z-score, t-student e F**, respectivamente. Agora, precisamos entender quando aplicar cada um. Isso não é uma tarefa muito complicada.

- i. **Para comparação de valor médio:** utiliza-se **teste Z-score** ou **t-student**. Quando for conhecido o comportamento populacional, ou seja, quando for conhecido μ e σ (média e desvio padrão populacionais), então deve-se utilizar a **distribuição normal** e, portanto, o **teste Z-score**, ou simplesmente **teste Z**. Como nem sempre é possível conhecer o comportamento populacional, então é possível fazer inferências (tomar decisões) a partir de dados amostrais ($\bar{x}$ e s), uma dessas maneiras é utilizar o **teste t**.

Observação: a **distribuição t de Student** tem formato de sino como a **distribuição normal**. A diferença entre elas é que a **função de densidade da distribuição t** é mais alargada que a da **distribuição z** porque é esperado maior variabilidade para dados amostrais, sendo que a largura da função está diretamente relacionada ao desvio padrão.

- ii. **Para comparação de variância ou de precisão de métodos:** utiliza-se o **teste F** que utiliza a **distribuição F de Fisher-Snedecor**. O comportamento da função densidade de probabilidade para essa distribuição é diferente de um sino, veja na figura abaixo.





Curva para função densidade de probabilidade da distribuição F de Fisher-Snedecor.

Antes de aprendermos a usar testes de hipóteses, vamos definir o que é uma **hipótese estatística**:

Hipótese estatística é uma afirmação sobre alguma característica descritiva de uma ou mais populações.

Por exemplo: no segundo problema (b), temos duas possibilidades e, portanto, duas hipóteses. A primeira seria que o teor médio de amilose pelo método 1 seja estatisticamente equivalente ao obtido pelo método 2, ou, seja $\bar{x}_1 = \bar{x}_2$. A segunda hipótese é que essas médias sejam estatisticamente diferentes, ou seja, $\bar{x}_1 \neq \bar{x}_2$.

Desta forma, obtivemos duas **hipóteses** que definiremos como segue:

Hipótese nula (H_0): em que $\bar{x}_1 = \bar{x}_2$. Se passarmos $\bar{x}_2$ para o primeiro lado da igualdade, obteremos $\bar{x}_1 - \bar{x}_2 = 0$ e, por isso, o nome de **hipótese nula**.

Hipótese alternativa (H_1): em que $\bar{x}_1 \neq \bar{x}_2$.

5.5 – Aplicando o teste-Z

Não há melhor maneira de aprender do que colocar a mão na massa, ou nesse caso, no bolo. Vamos aprender a usar os **testes de hipóteses** para responder a dúvida do consumidor: estaria ele pagando um preço diferente do praticado ao pagar 1,90 em um pedaço de bolo? Situação (a). Como são conhecidos o preço médio populacional (μ), 1,30 reais, e o desvio padrão (σ) 0,69 reais, então devemos utilizar o **teste Z**. Nossas hipóteses são:

H_0 : 1,90=1,30, equivalência estatística

H_1 : 1,90 $\neq$ 1,30



Precisamos saber se a diferença entre 1,90 e 1,30 é estatisticamente significativa. Como uma diferença pode assumir tanto valores negativos como positivos, então nosso teste será bilateral. **Se estamos medindo uma diferença entre dois valores, é de se esperar que essa diferença seja significativa caso seu módulo (valor sem sinal) seja alto, ou superior a um valor crítico.**

Como descobrir esse valor crítico? Na tabela! Já discutimos isso na seção anterior. Relembrando: no nosso caso, 95% de confiança ou α 0,05, 0,95 (região de não rejeição de H_0). Nessa região, temos uma equivalência estatística entre os dois valores. Portanto, a região de rejeição, ou de diferença estatística entre os dois valores, é a região complementar. Portanto, $1 - 0,95 = 0,05$ ($=\alpha$). Mas devemos lembrar que nosso teste é bilateral, então cada calda de rejeição terá a metade da área 0,05, ou seja, 0,025. Pronto! Vamos à **Tabela-Z** (final da apostila) localizar o Z_{score} correspondente a essa área de 0,025.... $Z_{score}(\text{crítico}) = 1,96$.

Para decidirmos entre aceitar ou rejeitar H_0 , falta um detalhe: precisamos calcular o valor $Z_{calculado}$ (o preço pago pelo bolo, padronizado), utilizando a equação abaixo:

$$Z_{calculado} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

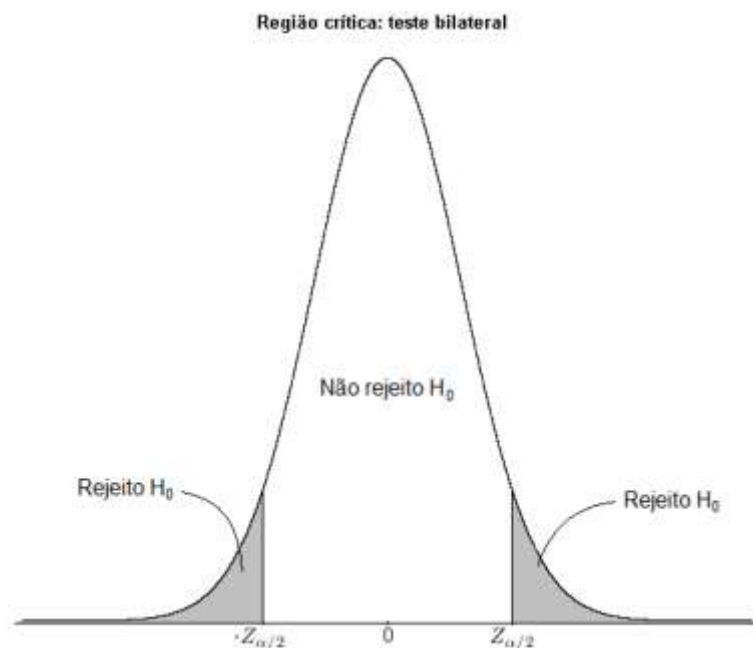
Aplicando os valores do problema, temos:

$$Z_{calculado} = \frac{1,90 - 1,30}{\frac{0,69}{\sqrt{1}}} = 0,87$$

em que n é o número de repetições ou observações. No nosso caso, o consumidor fez a pergunta a partir de uma (1) observação individual: quando pagou 1,90 em um pedaço de bolo. Se tivesse coletado valores de mais de um dia, por exemplo, durante 1 semana (7 dias), então n seria 7.

Nesse caso, o $Z_{calculado}$ (0,87) < $Z_{crítico}$ (1,96), o que significa que nosso valor está na **região de não rejeição de H_0** , que é a região central da curva, ou seja, entre as caldas, representado na figura abaixo. Portanto, a resposta correta ao consumidor seria: **não há evidências estatísticas, ao nível de 95% de confiança, para se acreditar que o preço do bolo mudou**. Ou seja, provavelmente a diferença que o consumidor percebeu deve estar associada à variabilidade normal do preço praticado pela lanchonete.





Indicação das regiões de aceitação e rejeição da **hipótese nula** para o **teste Z**. O mesmo raciocínio de regiões e caudas deve ser utilizado nos outros dois testes de hipóteses que estudaremos.

Se o resultado fosse o contrário, $Z_{\text{calculado}} > Z_{\text{crítico}}$, então rejeitaríamos H_0 e afirmariamos que há evidências estatísticas, **ao nível de 95% de confiança**, para sugerirmos que houve variação no preço do bolo praticado pela lanchonete.

Você precisa entender quando estará diante de um teste de hipótese unilateral ou bilateral, pois isso mudará o valor do $Z_{\text{crítico}}$. Na mesma situação da compra do bolo, imagine que o consumidor estivesse desconfiado que o preço subiu. Nesse caso teríamos um **teste unilateral ou unicaudal**, pois o cliente gostaria de saber se há evidências estatísticas de que o preço atual é maior que o preço populacional. Portanto, as hipóteses seriam:

$$H_0: 1,90 \leq 1,30$$

$$H_1: 1,90 > 1,30$$

A única diferença é que o valor que devemos identificar na **tabela-Z** é o correspondente à área de $1 - 0,95 = 0,05$ e, nesse caso, seria $Z_{\text{crítico}} = 1,645$ (valor interpolado entre as áreas 0,055 e 0,045 da tabela). Lembre-se que estamos diante de um teste unilateral, então não devemos dividir a área por 2. Como nessa situação $Z_{\text{calculado}} (0,87) < Z_{\text{crítico}} (1,645)$, então H_0 é verdadeira (não rejeitada) e não há evidências estatísticas de que o preço do bolo subiu. Espero que, até aqui, você tenha entendido em quais situações se aplica o teste Z e a diferença entre os testes unicaudal e bicaudal.

5.6 – Aplicando o teste T-student

Vamos adiante, lembre-se de que temos mais duas perguntas para responder ao laboratório de análise de grãos que está interessado em comparar dois métodos de determinação de amilose em arroz. **A primeira**



pergunta é se o resultado médio de teor de amilose $2,45 \pm 1,5 \text{ g.kg}^{-1}$ (método 1); $2,55 \pm 2,0 \text{ g.kg}^{-1}$ (método 2) podem ser considerados estatisticamente equivalentes. Nessa situação não sabemos nada sobre dados populacionais, apenas temos médias ($\bar{x}_1$, $\bar{x}_2$) e desvios padrão (s_1 , s_2) amostrais. Portanto, nesse caso, precisamos aplicar o **teste t-student** ou simplesmente **teste t**.

O raciocínio e a sequência na aplicação do **teste t-student** são idênticos aos do **teste Z**. Isto é, calculamos o valor $t_{\text{calculado}}$, e comparamos com o $t_{\text{calculado}}$ ($t_{\text{crítico}}$). Caso $t_{\text{calculado}} < t_{\text{crítico}}$ então nosso valor encontra-se na região de não rejeição de H_0 , o que significaria não haver diferença estatisticamente significativa entre os dois valores comparados, caso o teste fosse bilateral. Se $t_{\text{calculado}} > t_{\text{crítico}}$, rejeita-se H_0 .

Uma única complicação no **teste t** é que há 4 maneiras diferentes de se calcular o $t_{\text{calculado}}$. Vamos entender em quais casos devemos usar cada uma delas:



4 possibilidades para o teste t-student. Quando usar cada uma delas?

Antes de discutirmos as 4 possibilidades, ressalto que é muito improvável que o examinador cobre o conhecimento memorizado das diferentes fórmulas do teste t-student. Entendo que o mais importante é saber identificar cada uma das 4 possibilidades (CUIDADO: como só há uma fórmula mais simples no teste Z e uma no teste F, então essas poderão ser exigidas. Sugiro que as memorize).

(1) Comparação de médias entre uma amostra ($\bar{x}$) e uma população (μ) de variância desconhecida. Nessa situação, o $t_{\text{calculado}}$ e grau de liberdade (gl) são calculados como segue:

$$t_{\text{calculado}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$gl = n - 1$$

Note que não conhecemos a variância populacional (σ^2). Caso fosse conhecido o valor σ^2 , então utilizaríamos o teste-Z.

(2) Comparação de médias de populações com variâncias diferentes e desconhecidas, a partir de médias amostrais $\bar{x}_1$ e $\bar{x}_2$.

$$t_{\text{calculado}} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 \bar{x}_2}}$$



onde $s_{\bar{x}_1\bar{x}_2}$ é o desvio padrão agrupado e é dado por:

$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{\frac{s_1^2}{n_1} + 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{\frac{s_2^2}{n_2} + 1}} - 2$$

(3) Comparação de médias de populações com variâncias desconhecidas e iguais, e tamanhos de amostras iguais ($n_1 = n_2$).

$$t_{calculado} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1\bar{x}_2} \sqrt{\frac{2}{n}}}$$
$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$
$$gl = 2n - 2$$

(4) Comparação de médias de populações com variâncias desconhecidas e iguais, mas tamanhos de amostras diferentes ($n_1 \neq n_2$).

$$t_{calculado} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1\bar{x}_2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$
$$gl = n_1 + n_2 - 2$$

Agora precisamos saber qual das 4 opções se aplica ao nosso exemplo de comparação de métodos. Nesse caso, as variâncias populacionais são certamente desconhecidas porque nossos dados são amostrais. Entretanto, não sabemos se as variâncias associadas a cada método são iguais ou diferentes. Há uma maneira de decidirmos estatisticamente se essas variâncias devem ser consideradas iguais ou diferentes, que será o próximo teste que discutiremos: **teste F**. Por enquanto, vamos considerar que as variâncias são iguais e então nos enquadraremos no **caso (3) do teste-t**: variâncias iguais e tamanhos iguais (experimento em quintuplicata). Utilizando as fórmulas apresentadas no quadro anterior, obteremos os seguintes



resultados (Lembro que os valores que acompanham as médias são os desvios padrão, por exemplo: em $245 \pm 15 \text{ g.kg}^{-1}$, 245 corresponde à média e 15, ao desvio padrão):

$$t_{\text{calculado}} = -0,89$$
$$gl = 8$$

Importante lembrar que o sinal do $t_{\text{calculado}}$ não é relevante, pois ele dependerá de quem o usuário definirá como $\bar{x}_1$ e $\bar{x}_2$. Se $\bar{x}_1$ for o maior valor médio, então o sinal será positivo, caso contrário, negativo. Então vamos considerar apenas o módulo, $|t_{\text{calculado}}| = 0,89$.

Agora, procure na **tabela t** o valor $t_{\text{crítico}}$, utilizando grau de liberdade (gl) 8, que pode ser expresso por v , considerando 95% de confiança que corresponde ao nível de significância (α) 0,05. Como nosso teste é bilateral, devemos dividir α por 2, obtendo a área de probabilidade de 0,025 para cada cauda. Nessas condições, o $t_{\text{crítico}}$ é 2,306.

Como $|t_{\text{calculado}}| < t_{\text{crítico}}$, então a diferença entre as médias está na região de não rejeição de H_0 . Por isso, a **resposta para (b)**: não há evidências estatísticas que as médias, $24,5 \text{ g.kg}^{-1}$ e $25,5 \text{ g.kg}^{-1}$, sejam diferentes, ou seja, os dois métodos quantificam teores de amilose estatisticamente equivalentes. Podemos dizer ainda que a exatidão dos dois métodos é comparável. Nesse caso, a diferença numérica observada deve estar associada ao erro aleatório dos métodos.

5.7 – Aplicando o teste F

Ufa! Só falta uma pergunta a ser respondida, **problema (c)**. Os dois métodos apresentam **precisão comparáveis**? Isso significaria apresentarem variâncias que podem ser consideradas estatisticamente equivalentes (**NOTA**: caso sejam fornecidos os desvios padrão, basta elevar esses valores ao quadrado para obter as respectivas variâncias). Para a comparação de variâncias (duas a duas), o **teste de hipótese** mais utilizado é o **teste F**. Para obter o valor de $F_{\text{calculado}}$, basta dividir a maior variância pela menor variância como descrito na equação abaixo:

$$F_{\text{calculado}} = \frac{s_{\text{maior}}^2}{s_{\text{menor}}^2}$$

Para cada área de probabilidade, no nosso caso 5,0%, existe uma **tabela F**. O **teste F** já considera a condição bicaudal na tabela (**final da apostila**), portanto não é necessário dividir essa área por 2. Para localizarmos o valor de $F_{\text{crítico}}$, na tabela, necessitamos dos seguintes dados: área de probabilidade, graus de liberdade (v_1 e v_2) para as duas amostras. No problema (c), $v_1 = v_2 = 4$. Utilizando esses dados na tabela-F, obtemos $F_{\text{crítico}} = 6,39$. Aplicando as variâncias 2^2 e $1,5^2$ na equação acima, temos $F_{\text{calculado}} = 1,77$. Portanto, $|F_{\text{calculado}}| < F_{\text{crítico}}$, resultado que está na região de não rejeição de H_0 . Por isso, a **resposta para (C)**: não há evidências estatísticas que as variâncias ou a precisão dos dois métodos (1 e 2) sejam diferentes.

Da aplicação dos testes t e F, chegamos à conclusão que os dois métodos investigados apresentam exatidão e precisão comparáveis. Vale lembrar que, em uma situação corriqueira, você deve realizar o teste F, antes do teste t, para descobrir se as variâncias podem ser consideradas iguais ou diferentes e, a partir desse resultado, decidir qual dos 4 **teste t** aplicar.



Uma última informação pertinente: considere média ou desvio padrão históricos (dados coletados no decorrer de um longo período) como sendo dados populacionais.

5.8 – Construindo intervalos de confiança para a média populacional

Em algumas situações e exercícios, é necessário estabelecer os limites de um intervalo de confiança, dentre os quais, existe uma certa probabilidade da média populacional (μ) estar contida. A média populacional é um valor perfeitamente determinável se estivermos dispostos a coletar dados de toda população. Desta forma, não está muito correto associá-la a uma certa probabilidade. A discussão teórica sobre intervalos de confiança para média populacional é um pouco mais extensa, mas não convém tratarmos isso aqui e, por isso, apresentei uma explicação mais sintética no início do parágrafo, mas que será suficiente.

O que você realmente precisa saber é como calcular os intervalos de confiança. Há duas possibilidades: na primeira, o desvio padrão populacional (σ) é conhecido e se utiliza a distribuição normal; na segunda, o desvio padrão populacional não é conhecido e se utiliza a distribuição t-Student.

Intervalo de confiança para μ , a partir de uma observação x_i :

$$x_i - z\sigma < \mu < x_i + z\sigma$$

em que z é o valor da tabela que corresponde ao nível de confiança desejado para o intervalo.

Intervalo de confiança para μ , a partir de uma amostra (conjunto de indivíduos), utilizando a distribuição t-Student:

$$\bar{x} - t_{n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1} \frac{s}{\sqrt{n}}$$

Vamos ver agora como esse assunto já foi cobrado em prova.



(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2018) A concentração de chumbo em um resíduo sólido foi determinada em dois laboratórios, cada um usando um método diferente. As análises foram feitas com quatro réplicas autênticas ($n = 4$), e os resultados (expressos em termos de intervalo de confiança) são indicados abaixo:

- Laboratório A: $160 \pm 52 \mu\text{g kg}^{-1}$
- Laboratório B: $470 \pm 15 \mu\text{g kg}^{-1}$



Analizando os resultados, conclui-se que

- a) eles são estatisticamente iguais.
- b) as precisões são similares.
- c) houve erro grosseiro e/ou interferência que afetou pelo menos um laboratório.
- d) não existiu erro aleatório em nenhum dos resultados.
- e) o coeficiente de variação do laboratório B foi maior.

Comentários

Analizando a situação exposta no enunciado, percebemos que podemos comparar as médias obtidas pelos laboratórios, aplicando o teste t-Student, e comparar os desvios padrão obtidos, aplicando o teste F. Estamos diante daquela típica situação em que devemos realizar o teste F, antes do teste t, para descobrir se as variâncias podem ser consideradas iguais ou diferentes e, a partir desse resultado, decidir qual dos 4 testes t aplicar.

Entretanto, o avaliador apresentou valores para que a utilização de fórmulas não fosse necessária. Isto é, podemos resolver o exercício apenas utilizando nosso conhecimento geral sobre erros e testes de hipótese.

Testes de comparação de média ou de variância são úteis quando os valores a serem comparados são relativamente próximos, por exemplo, $160 \pm 5 \mu\text{g kg}^{-1}$ e $153 \pm 4 \mu\text{g kg}^{-1}$. Nesses casos, o analista de laboratório ficará com dúvida sobre a equivalência estatística das médias obtidas e também do desvio padrão (que elevado ao quadrado corresponde à variância). Diante de resultados muito discrepantes, como é o caso do enunciado, o analista não terá dúvida de que as médias são diferentes e as variâncias ou desvios padrão também. De certo, o resultado dos testes t e F indicarão que os resultados são estatisticamente diferentes.

Após essa breve discussão, vamos à análise das alternativas:

Letras A e B: incorretas. Conforme discussão acima, tanto a média quanto o desvio padrão são estatisticamente diferentes.

Letra C: correta. Não sabemos qual dos laboratórios está determinando chumbo corretamente. Pode até ser que os dois estejam cometendo erros sistemático e/ou aleatório. Considerando a grande discrepância de resultados entre os laboratórios, podemos afirmar que pelo menos um deles está cometendo um erro grosseiro ou que algum interferente esteja presente na análise.

Letra D: incorreta. Erros aleatórios estão sempre presente, podendo apenas ser minimizados por meio da utilização de equipamentos mais precisos.

Letra E: incorreta. Podemos analisar qualitativamente a questão. Observe na equação abaixo do coeficiente de variação (CV), se o laboratório B apresentou maior média (denominador) e menor desvio padrão (numerador), então seu CV será menor que o do laboratório A.

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

Havendo dúvida, podemos aplicar os valores de cada laboratório na equação acima, obtendo CV 32,5% para A e 3,2% para B.

Resposta: letra C



(FDRH - Perito Químico Forense - IGP/RS - 2008) A determinação de Cr em uma amostra foi realizada comparativamente por dois métodos. No método A, após realização de 8 medidas, o valor médio encontrado foi 80 ppm. No método B, o valor médio obtido foi de 78 ppm, resultante de 6 medidas. A variância do método A foi de 8,0 e do método B, de 4,0. O valor de F crítico tabelado para essas condições, em nível de confiança de 95%, é 4,88. Com relação a essa medida de Cr efetuada pelos dois métodos, levando-se em conta a aplicação do teste F, pode-se afirmar que

- a) os desvios padrões devem ter sido devidos a erros aleatórios e independem da amostra.
- b) a aplicação do teste F serviu para selecionar as medidas a serem rejeitadas.
- c) o número de graus de liberdade no método A foi 8.
- d) existe diferença na precisão dos dois métodos.
- e) a aplicação do teste F serviu para avaliar se os dois conjuntos de medidas são estatisticamente diferentes.

Comentários

Conforme estudamos, a comparação da variância ou da precisão de métodos pode ser realizada pelo teste F. O valor de F crítico, 4,88, foi fornecido e o F calculado é obtido pela equação abaixo:

$$F_{\text{calculado}} = \frac{S_{\text{maior}}^2}{S_{\text{menor}}^2}$$

$$F_{\text{calculado}} = \frac{8}{4} = 2$$

Considerando $|F_{\text{calculado}}| < F_{\text{crítico}}$, dizemos que o resultado está na região de não rejeição de H_0 . Portanto, não há evidências estatísticas que as variâncias ou a precisão dos dois métodos sejam diferentes.

Vamos à análise das alternativas:

Letra A: correta. Desvios padrão relativamente baixos estão associados a erros aleatórios, que são inerentes a todas as medidas.

Letra B: incorreta. O teste F foi utilizado para comparação da precisão dos métodos.

Letra C: incorreta. O número de graus de liberdade do método A foi igual a 7 ($gl = 8 - 1$).

Letras D e E: incorreta. De acordo com o teste de hipótese, não há evidências estatísticas que as variâncias ou a precisão dos dois métodos sejam diferentes.

Resposta: letra A

(CESGRANRIO - Téc. Químico de Petróleo - CESGRANRIO - 2018) A determinação de glicose foi feita em quatro réplicas, isto é, $n = 4$, em uma amostra coletada em jejum. Os seguintes resultados foram obtidos: 91,0 mg.dL⁻¹; 95,0 mg.dL⁻¹; 96,0 mg.dL⁻¹ e 98,0 mg.dL⁻¹, com desvio padrão 2,9 mg.dL⁻¹. A distribuição dos dados tende à normalidade (com $t = 3,2$ para 3 graus de liberdade e com 95% de limite de confiança).

Considerando o valor médio e o intervalo de confiança do resultado, o valor mais próximo, em mg.dL⁻¹, da maior concentração admitida para glicose na amostra é

- a) 95,0



- b) 96,5
- c) 98,6
- d) 99,6
- e) 100,2

Comentários

Os dados fornecidos pertinentes para o cálculo do intervalo de confiança são: $\bar{x}=95,0$ mg.dL⁻¹ (Calculado a partir dos valores individuais 91; 95;96 e 98)

$$s=2,9 \text{ mg.dL}^{-1}$$

$$n=4$$

$$t_{0,05,3}=3,20$$

Considerando que os valores conhecidos ($\bar{x}$, s) são amostrais, devemos utilizar a inequação abaixo para calcular o intervalo de confiança a 95%:

$$\bar{x} - t_{n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1} \frac{s}{\sqrt{n}}$$
$$95,0 - 3,2 \frac{2,9}{\sqrt{4}} < \mu < 95,0 + 3,2 \frac{2,9}{\sqrt{4}}$$

$$90,36 < \mu < 99,64$$

Portanto, a maior concentração do intervalo de confiança (maior concentração admitida) para a glicose é 99,64 mg.dL⁻¹.

Resposta: letra D



QUESTÕES COMENTADAS

Texto referente às questões 1 e 2:



elemento	concentração em $\mu\text{g/g}$		desvio-padrão
	casaco	vidraça	s
As	132	122	9,7
Co	0,54	0,61	0,026
La	4,01	3,60	0,20
Sb	2,81	2,77	0,26
Th	0,62	0,75	0,044

Um perito criminal recebeu em seu laboratório, como principal evidência em um caso criminal, pequenos fragmentos de vidro encontrados incrustados no casaco de um suspeito de assassinato. Esses fragmentos são idênticos em composição a uma rara vidraça belga de vidro manchado quebrada durante o crime. O perito decidiu então determinar os elementos As, Co, La, Sb e Th no vidro incrustado no casaco do suspeito para verificar se este era do mesmo material da vidraça belga. A técnica escolhida para essas determinações foi a espectroscopia de absorção atômica. As médias e os desvios-padrão das análises em triplicata desses cinco elementos nas amostras de vidro retiradas do casaco, bem como os valores conhecidos para a vidraça belga são mostrados na tabela acima.

Considerando essa situação hipotética, que $\sqrt{3} = 1,73$ e que o parâmetro t de *Student* para 2 graus de liberdade e 95% de confiança é igual a 4,303, julgue os itens a seguir:

(CESPE - Perito Criminal Federal/Química - PF - 2004) O desvio-padrão s para qualquer um dos elementos analisados pode ser corretamente calculado por meio da equação abaixo, em que x_1 , x_2 e x_3 são os valores individuais de concentração obtidos para cada alíquota analisada.

$$s = \sqrt{\frac{\left(x_1 - \frac{x_1 + x_2 + x_3}{3}\right)^2 + \left(x_2 - \frac{x_1 + x_2 + x_3}{3}\right)^2 + \left(x_3 - \frac{x_1 + x_2 + x_3}{3}\right)^2}{2}}$$

Comentários

Conforme estudamos, o desvio padrão amostral pode ser calculado pela equação abaixo:

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

No exemplo do enunciado, a determinação foi realizada em triplicata ($n=3$). Além disso o termo $(x_1 + x_2 + x_3)/3$, presente na equação do enunciado, corresponde à média ($\bar{x}$). Portanto, o desvio-padrão s para qualquer um dos elementos analisados pode ser corretamente calculado por meio da equação apresentada.

Resposta: certo

(CESPE - Perito Criminal Federal/Química - PF - 2004) Para um nível de confiança de 95% como critério de dúvida, com base nos dados obtidos, o perito tem argumentos suficientes para duvidar da



culpabilidade do acusado, pois pelo menos dois dos elementos analisados possuem concentrações significativamente diferentes nos fragmentos de vidro incrustados no casaco, em comparação com a vidraça de vidro belga.

Comentários

Questão relativamente difícil, pois exige-se do candidato a memorização das equações do teste *t-Student*, além da realização de vários cálculos. De início, temos que identificar qual dos 4 testes *t* devemos aplicar. O número de repetições foi 3 (triplicata) para a análise do vidro encrustado. A variância populacional é desconhecida. As concentrações no vidro Belga são conhecidas e, por isso, devem ser consideradas médias populacionais ou médias de referência. Portanto, estamos diante da situação (1) Comparação de médias entre uma amostra ($\bar{x}$) e uma população (μ) de variância desconhecida. Nessa situação o $t_{calculado}$ e grau de liberdade (gl) são calculados como segue:

$$t_{calculado} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$gl = n - 1$$

Devemos calcular $t_{calculado}$ para cada metal. O $t_{crítico}$ 4,30 foi fornecido pelo enunciado. Caso não tivesse sido fornecido, poderíamos encontrá-lo por meio da tabela: $gl = n - 1 = 2$ e 95% de confiança ($=0,05/2=0,025$, pois o teste bilateral). Para facilitar o entendimento, resumi os resultados na tabela abaixo (deixei em vermelho os valores calculados ou obtidos em tabela).

Elemento	Casaco	Vidraça	s	n	gl	$t_{calculado}$	$t_{crítico}$
As	132	122	9,7	3	2	1,79	4,30
Co	0,54	0,61	0,026	3	2	-4,66*	4,30
La	4,01	3,6	0,2	3	2	3,55	4,30
Sb	2,81	2,77	0,26	3	2	0,27	4,30
Th	0,62	0,75	0,044	3	2	-5,12*	4,30

*Valores que $|t_{calculado}| > t_{crítico}$, região de rejeição de hipótese nula (H_0) e de aceitação da hipótese alternativa (H_1), em que $\bar{x}_1 \neq \bar{x}_2$.

Os resultados apresentados na tabela acima indicam que dois dos elementos analisados possuem concentrações significativamente diferentes nos fragmentos de vidro incrustados no casaco, em comparação com a vidraça de vidro belga ($\bar{x}_1 \neq \bar{x}_2$). Este é um forte indício de que os fragmentos incrustados no casaco do acusado não são oriundos da rara vidraça belga de vidro manchado quebrada durante o crime.

Resposta: certo

(FUNCAB - Perito Criminal/Farmácia - POLITEC-MT - 2013) Uma análise química é o conjunto de técnicas de laboratório utilizado na identificação das espécies químicas e na determinação de



informações químicas ou físicas sobre uma amostra. No que diz respeito à preparação de uma amostra, analise as afirmativas a seguir:

- I. A amostragem é o processo usado para coleta de uma amostra representativa para análise.
- II. Materiais heterogêneos irão sempre produzir amostras heterogêneas para análise também, tendo, obrigatoriamente, a mesma composição do material.
- III. A preparação de uma amostra é o processo pelo qual se converte uma amostra bruta em uma amostra de laboratório homogênea.
- IV. A amostra bruta tem que ser representativa de um lote, e sua escolha deve ser crítica para que a análise seja válida.

Estão corretas apenas as afirmativas:

- a) I, II e III.
- b) I, III e IV.
- c) II, III e IV.
- d) I, II e IV.
- e) I e IV.

Comentários

Afirmativa I: correta. Traz a própria definição de amostragem.

Afirmativa II: incorreta. Materiais heterogêneos são coletados de forma representativa, mas, em seguida, devem passar por um processo chamado de preparação de amostra, tornando-a um material homogêneo. Só após essa etapa, a amostra segue para análise laboratorial. Por exemplo, o solo é amostrado em campo. Em seguida, é submetido aos processos de secagem e moagem, tornando-o um pó seco e homogêneo. Esse preparo da amostra se faz necessário para garantir uma maior representatividade, visto que a maioria dos métodos analíticos utilizam pequenas quantidades do material. Para análise de matéria orgânica no solo, por exemplo, são pesados apenas 0,5g do solo seco e moído.

Afirmativa III: correta. Conforme explicado na afirmativa anterior.

Afirmativa IV: correta. A amostragem é uma etapa crítica para análise laboratorial e deve garantir representatividade da população de interesse.

Resposta: letra B



CESGRANRIO - Téc. de Laboratório/Química - CEFET-RJ - 2014) Os resultados (cinco réplicas) obtidos a partir de uma determinação da concentração de um analito em uma amostra são mostrados na Tabela abaixo.

Réplica	1	2	3	4	5
Resultado (em $\mu\text{g} \cdot \text{g}^{-1}$)	10,5	10,4	10,5	10,5	10,6

Se o resultado verdadeiro para o analito na amostra é $10,0 \mu\text{g} \cdot \text{g}^{-1}$, é possível concluir que o(a)

- a) erro sistemático é $0,5 \mu\text{g} \cdot \text{g}^{-1}$.
- b) desvio padrão é menor que $0,005 \mu\text{g} \cdot \text{g}^{-1}$.
- c) resultado da análise é exato.
- d) valor médio é diferente da mediana.
- e) mediana é $10,4 \mu\text{g} \cdot \text{g}^{-1}$.

Comentários

Letra A: conforme estudamos, o erro sistemático é aquele que desloca o valor observado (experimental) do valor real ou verdadeiro. Devemos então, calcular a média e, em seguida, calcular o erro (=erro sistemático), como segue:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\bar{x} = \frac{10,5 + 10,4 + 10,5 + 10,5 + 10,6}{5}$$

$$\bar{x} = 10,5 \mu\text{g} \cdot \text{g}^{-1}$$

$$e_i = x_i - x_{(\text{real})}$$

$$e_i = 10,5 - 10,0$$

$$e_i = 0,5 \mu\text{g} \cdot \text{g}^{-1}$$

Portanto, a alternativa é verdadeira.

Letra B: incorreta. Aplicando o valor médio e os valores individuais na fórmula do desvio padrão, obtemos $0,071 \mu\text{g} \cdot \text{g}^{-1}$.



Letra C: incorreta. Errado, a exatidão está relacionada com a diferença entre o valor medido (experimental) e o valor real ou verdadeiro, ou seja, com o erro. Já que $10,5 \mu\text{g.g}^{-1}$ é relativamente distante do valor verdadeiro $10 \mu\text{g.g}^{-1}$, podemos dizer que o resultado da análise é inexato.

Letra D: incorreta. O valor médio é $10,5 \mu\text{g.g}^{-1}$. Para obter a mediana, devemos colocar os valores em ordem crescente e identificar o valor central. Quando o número de observações for par, é necessário calcular a média aritmética das duas observações centrais. Seguem abaixo os valores em ordem crescente com destaque do valor central (mediana), que, nesse caso, é igual à média.

10,4 10,5 10,5 10,5 10,6

Letra E: incorreta. Conforme demonstrado na Letra D.

Resposta: letra A

(FUNCAB - Perito Oficial Criminal-Biologia/Química/Farmácia POLITEC-MT - 2013) Calcule o desvio padrão do conjunto de valores (1, 2, 1, 2, 3, 1, 2, 3, 5, 10) e, em seguida, assinale a alternativa correta.

- a) $2,0 < d < 2,1$
- b) $2,1 < d < 2,2$
- c) $2,4 < d < 2,5$
- d) $2,5 < d < 2,6$
- e) $2,6 < d < 2,7$

Comentários

O exercício falha em não dizer e não fornecer indícios sobre os valores serem amostrais ou populacionais. Aplicando a série de valores nas equações do desvio padrão populacional e amostral, obtemos 2,61 e 2,74, respectivamente. Portanto, o único valor que se enquadra em uma das alternativas é 2,61, alternativa E.

Resposta: letra E

(UECE - Perito Criminal/Eng. Química - PCCE - 2003) Na análise volumétrica, um dos problemas é a ocorrência de erro na titulação. Se 30,0 mL de HCl 0,100 M são titulados com NaOH 0,100 M e uma solução de vermelho de metila é usada como indicador, o erro de titulação, admitindo-se $\text{pH} = 4$ no ponto final, é:

Obs.: O ponto final ocorre antes do ponto de equivalência.

- a) +0,02%
- b) -0,10%

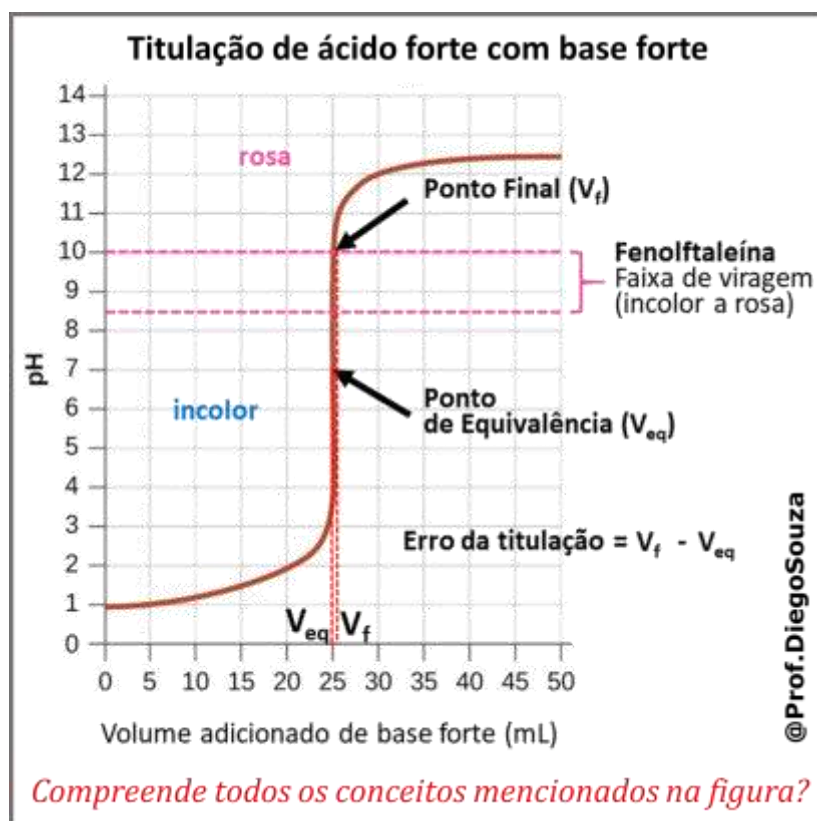


c) $-0,02\%$

d) $-0,20\%$

Comentários: questão um pouco antiga, mas muito interessante, pois exige simultaneamente noções de cálculos estequiométricos e o conceito de erro relativo.

Abro um breve parêntese para revisar rapidamente alguns conceitos importantes sobre titulação. Vamos lá? Observe a figura abaixo que representa o gráfico de uma titulação entre ácido e base fortes, do mesmo modo que acontece no enunciado, e na sequência leia uma breve explicação sobre os conceitos indicados no gráfico.



De início, o pH é ácido e o indicador se apresenta na forma incolor, à medida que a base forte é adicionada, o pH vai aumentando. Note que, na verdade, a fenolftaleína não apresenta um ponto de viragem, mas sim uma faixa de viragem entre pH 8,5 e 10, abaixo de 8,5 ela é incolor e, acima de 10, rosa. Entre esses dois valores, a cor não é muito bem definida. O mesmo acontece para outros indicadores.

Em decorrência dessa faixa de viragem, o ponto final (V_f) só se dá em pH próximo a 10, no exemplo da figura. No caso da nossa questão, a faixa de viragem é em torno de pH 4,0. Já o ponto de equivalência (V_{eq}), indicado no gráfico, é exatamente em pH = 7,00, ponto em que todo ácido foi exatamente consumido pela base adicionada e não há ainda excesso de base. Notamos então que há uma pequena diferença (por meio das linhas verticais pontilhadas em vermelho) entre V_f e V_{eq} . A essa diferença, dá-se o nome de erro da titulação. Vamos agora reforçar por meio de conceitos:

- Ponto final (V_f) ou ponto de viragem:** é o ponto (volume) que se pode determinar experimentalmente. Por exemplo, em uma titulação ácido-base, pode ser utilizado o indicador



fenolftaleína, que se torna rosa quando o pH passa de ácido para básico. Esse ponto de viragem (mudança da cor) é denominado ponto final. Vale reforçar que V_f é diferente de ponto de equivalência V_{eq} .

- ii. **Ponto de equivalência (V_{eq}):** é o volume exato a ser gotejado para que a reação entre titulante e titulado se complete. O ponto de equivalência não pode ser determinado experimentalmente e está relacionado ao número de mols necessários para a reação se completar a partir dos cálculos estequiométricos. Por exemplo, se estivéssemos titulando 1 mol NaOH (presente no erlenmeyer) seria necessário gotejar o volume de um titulante que contivesse 1 mol de um ácido, exemplo HCl.
- iii. **Erro da titulação:** embora V_{eq} e V_f possam apresentar valores próximos, eles são diferentes, o que sempre ocasionará um erro intrínseco à titulação denominado Erro da Titulação (E) $\rightarrow E = V_f - V_{eq}$.

Vamos voltar ao cerne da questão??

Para calcular o erro, precisamos, então, encontrar dois valores: V_{eq} (ponto de equivalência) e V_f (ponto de viragem). O primeiro é relativamente simples, vejamos:

V_{eq}

Para facilitar, note que tanto o ácido (HCl) quanto a base (NaOH) são monoproticos, isto é, liberam, respectivamente apenas 1 H^+ e 1 OH^- . Portanto, 1 mol de NaOH neutraliza 1 mol de HCl, formando 1 mol de água.

Além disso, ambos estão na mesma concentração. Sendo assim, podemos concluir que 1 mL da base neutraliza 1 mL de ácido e, por conseguinte, para neutralizar 30,0 mL de ácido, serão necessários 30,0 mL da base para neutralizá-lo totalmente e atingir o pH = 7,00.

Portanto,

$V_{eq} = 30,0$ mL de base NaOH 0,01 mol/L adicionada.

V_f

"É aqui que a porca torce o rabo". Como determinar o volume de base adicionado até pH = 4,00, ponto final desta titulação específica? A partir do pH, podemos, no máximo, calcular a concentração de H^+ em solução, mas isso vai nos ajudar, como segue:

$$pH = -\log[H^+] \rightarrow 4,0 = -\log[H^+] \rightarrow [H^+] = 10^{-4} \text{ mol} \cdot L^{-1}$$

Sabemos ainda que concentração (C) corresponde à razão entre número de mols (n) e volume da solução (V):

$$C = \frac{n}{V}$$

Ou seja:



$$[H^+] = \frac{n_{H^+}}{V_{\text{total da solução}}}$$

$[H^+]$ já encontramos, mas não sabemos nem número de mols de H^+ e nem o volume total da solução. Vamos, então, escrevê-los em termos de outras variáveis.

Volume total da solução em pH = 4,0:

$$V_{\text{total da solução}} = V_{\text{ácido inicial}} + V_{\text{base titulada}}$$

$$V_{\text{total da solução}} = 30,0 \text{ mL} + V_{\text{base titulada}}$$

Número de mols de H^+ em pH = 4,0:

Se a reação entre o ácido e a base é 1:1, então,

$$\text{Número de mols de } H^+ = \text{número de mols de ácido} - \text{número de mols da base titulada}$$

ou seja,

$$n_{H^+} = n_{\text{ácido}} - n_{\text{base}}$$

Utilizando a fórmula de concentração, podemos reescrever a equação como segue:

$$n_{H^+} = C_{\text{ácido}} \cdot V_{\text{ácido}} - C_{\text{base}} \cdot V_{\text{base}}$$

Conhecemos os valores de concentração de ambos (0,01 mol/L) e o volume de ácido (30,0 mL = $30 \cdot 10^{-3}$ L), então:

$$n_{H^+} = (0,01 \text{ mol} \cdot \text{L}^{-1} \cdot 30 \cdot 10^{-3} \text{ L}) - (0,01 \text{ mol} \cdot \text{L}^{-1} \cdot V_{\text{base}})$$

Podemos reescrever, então, a fórmula da concentração de H^+ como segue:

$$[H^+] = \frac{n_{H^+}}{V_{\text{total da solução}}}$$

$$[H^+] = \frac{(0,01 \text{ mol} \cdot \text{L}^{-1} \cdot 30 \cdot 10^{-3} \text{ L}) - (0,01 \text{ mol} \cdot \text{L}^{-1} \cdot V_{\text{base}})}{30,0 \text{ mL} + V_{\text{base titulada}}}$$



Bizu: utilizamos $30 \cdot 10^{-3} \text{L}$ no lugar de 30 mL para harmonizar com a unidade da concentração molar, no entanto, como há volume em cima e em baixo da razão, se usarmos mL em cima, vamos obter V_{base} já em mL que é mais cômodo.

$$10^{-4} \text{mol.L}^{-1} = \frac{(0,01 \text{mol.L}^{-1} \cdot 30 \text{mL}) - (0,01 \text{mol.L}^{-1} \cdot V_{\text{base}})}{30,0 \text{mL} + V_{\text{base titulada}}}$$

Quem é meu aluno sabe que eu adoro arredondar cálculos para agilizar a resolução de exercícios na hora da prova, mas aqui não dá porque em geral o V_{eq} e V_{f} são próximos. Se arredondássemos, nesse caso, iríamos obter $V_{\text{eq}} = V_{\text{f}}$, o que não é desejável. Portanto, use suas técnicas de cálculo para resolver a equação acima com pelo menos duas casas decimais, beleza?!

Nesse caso, vamos obter:

$$V_{\text{base titulada}} = 29,94 \text{ mL}$$

Finalmente podemos calcular o erro da titulação em termos percentuais.

Erro (%):

$$\text{Erro da titulação (\%)} = \frac{V_{\text{f}} - V_{\text{eq}}}{V_{\text{eq}}} \times 100$$

$$\text{Erro da titulação (\%)} = \frac{29,94 - 30,00}{30,00} \times 100$$

$$\text{Erro da titulação (\%)} = - 0,2\%$$

Resposta: letra D

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2018) Um procedimento gravimétrico foi realizado de acordo com as instruções da literatura. Um precipitado foi obtido após a mistura da solução de amostra com a solução do reagente precipitante (esta última adicionada em quantidade para garantir excesso estequiométrico). Por conta das características do precipitado, uma fração ficou suspensa na solução (suspensão coloidal) e foi perdida na filtragem para recolher o sólido. Após secagem, a massa final do precipitado foi medida em balança calibrada. O procedimento foi repetido três vezes para se obter um valor médio de massa de analito e seu desvio padrão.

Com relação aos erros associados ao resultado, tem-se

a) apenas erro sistemático



- b) apenas erro aleatório
- c) apenas erro instrumental (da balança)
- d) tanto erro aleatório quanto erro sistemático
- e) que não há erro

Comentários

Conforme estudamos, o erro aleatório é intrínseco às medidas, ou seja, ele sempre estará presente, por exemplo, no uso de um balão volumétrico, de um termômetro ou de um instrumento analítico. O que muda é a dimensão do erro aleatório, podemos, por exemplo, diminuí-lo ao trocar a balança utilizada por outra mais sensível e precisa. Além disso, no enunciado é mencionado que “por conta das características do precipitado [...] uma fração [...] foi perdida na filtragem para recolher o sólido”. Portanto, se é uma característica da amostra, haverá perdas semelhantes (erro na mesma direção) nas três repetições, o que configura a ocorrência de erro sistemático. Concluímos, portanto, que estão presentes tanto erro aleatório quanto erro sistemático no experimento.

Resposta: letra D

(CESPE - Perito Criminal/Química - SGA/AC - 2008) Se uma grandeza x qualquer é calculada pela seguinte equação: $x = \frac{a \cdot b}{c}$, em que a , b e c são grandezas medidas com o fim de se obter o valor de x , então o desvio-padrão de x , s_x , pode ser corretamente calculado pela seguinte expressão:

$$s_x = x \sqrt{\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2},$$

em que s_a , s_b e s_c representam, respectivamente, os desvios-padrão das grandezas a , b e c .

Comentários

Item correto. Estudamos que o cálculo da propagação da incerteza é aplicável à propagação do desvio padrão (imprecisão). Segundo o cálculo da multiplicação ou divisão, podemos propagar o desvio padrão por meio da equação abaixo:

$$S = X_{\text{resultante}} \sqrt{\left(\frac{S_1}{X_1}\right)^2 + \left(\frac{S_2}{X_2}\right)^2 + \left(\frac{S_3}{X_3}\right)^2}$$

Resposta: certo



Texto referente aos próximos dois exercícios:

indicador	pH de viragem	número de medidas	t
fenoftaleína	8,2 a 10,0	3	4,30
vermelho de metila	4,4 a 6,2	4	3,18
azul de bromotimol	6,0 a 7,6	5	2,78

Tabela 1

Tabela 2

Considere que uma solução de NaOH foi padronizada por meio da titulação, em quadruplicata, com 20,00 mL de uma solução padrão de biftalato de potássio 0,1000 mol/L e que a concentração média encontrada para a solução de NaOH foi de 0,080 mol/L, com uma estimativa do desvio padrão de $\pm 0,002$ mol/L. Considere, ainda, a tabela 1, em que é apresentado o pH de viragem de alguns indicadores ácido-base, e a tabela 2, em que são apresentados os valores do parâmetro t de *student* para uma probabilidade de 95%.

Com base nessas informações, julgue os itens seguintes:

(CESPE - Químico - FUB - 2011) Considerando-se a inexistência de erros sistemáticos, é correto afirmar que a probabilidade de que o valor real da concentração de NaOH esteja no intervalo compreendido entre 0,078 e 0,082 será maior que 95%.

Comentários

Vale ressaltar que o mencionado “valor real” corresponde à média populacional. Os dados fornecidos pertinentes para o cálculo do intervalo de confiança são:

$$\bar{x} = 0,080 \text{ mol/L}$$

$$s = 0,002 \text{ mol/L}$$

$$n = 4$$

$$t_{0,05,3} = 3,18 \text{ (leia-se: valor } t_{\text{crítico}} \text{ com } \sigma = 0,05 \text{ ou } 95\% \text{ de confiança, e } gl = 3).$$

Considerando que os valores conhecidos ($\bar{x}$, s) são amostrais, devemos utilizar a inequação abaixo para calcular o intervalo de confiança a 95%:

$$\bar{x} - t_{n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1} \frac{s}{\sqrt{n}}$$
$$0,080 - 3,18 \frac{0,002}{\sqrt{4}} < \mu < 0,080 + 3,18 \frac{0,002}{\sqrt{4}}$$



$$0,077 < \mu < 0,083$$

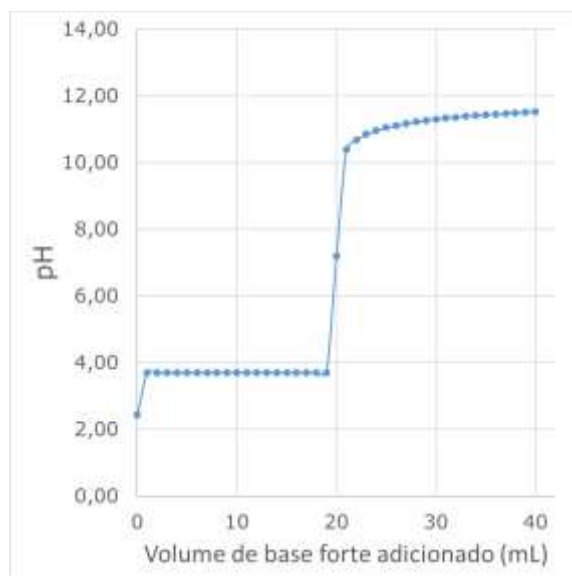
A probabilidade de a média estar contida no intervalo de 0,077 a 0,083 é de 95%. Portanto, podemos concluir que a probabilidade associada ao intervalo apresentado no enunciado (de 0,078 a 0,082) é menor que 95%, já que ele é mais estreito do que o calculado.

Resposta: errado

(CESPE - Químico - FUB - 2011) Entre os indicadores apresentados na tabela 1, o mais adequado à titulação em questão é o vermelho de metila.

Comentários

Note que os conceitos e ferramentas estatísticas são cobrados de forma conjunta. Estudaremos sobre a escolha de indicadores nas aulas sobre equilíbrio químico. Devemos lembrar o formato de uma curva de titulação de ácido fraco com base forte, como segue:



Para essa titulação, o indicador mais adequado é o azul de bromotimol por apresentar a faixa de viragem mais próxima do ponto de equivalência da curva de titulação acima (pH=7).

Resposta: errado

(FUNIVERSA - Papiloscopista - PCGO - 2015) Um estudo realizado com 200 policiais civis, após a realização de testes em que cada profissional deveria tomar decisões importantes referentes a diversos fatores do cotidiano, avaliou características psicológicas atenuantes. Como a variável idade se apresenta como um fator determinante na atitude do profissional, fez-se necessária a elaboração de uma tabela de distribuição de frequência das idades do grupo em questão.



Classe das idades	N.º de policiais
21 - 26	56
26 - 31	48
31 - 36	45
36 - 41	23
41 - 46	15
46 - 51	13
Total	200

Com base nesse caso hipotético, assinale a alternativa que apresenta o ponto médio da classe mediana.

- a) 23,5
- b) 28,5
- c) 33,5
- d) 38,5
- e) 43,5

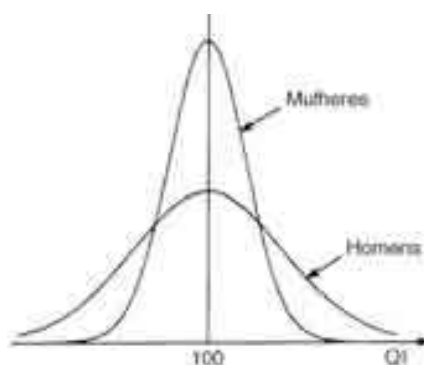
Comentários

A classe mediana é aquela que contém a mediana. Em uma série de 200 valores crescentes, sabemos que o valor mediano será a média entre os valores 100º e 101º. Desta forma, a faixa 26 | - 31 é a classe mediana por conter do 57º ao 104º e, por conseguinte, conter os valores 100º e 101º. Para obtermos o ponto médio de uma classe, devemos calcular a média entre os seus dois limites, como segue:

$$\bar{x} = \frac{26 + 31}{2} = 28,5$$

Resposta: letra B

(CESPE - Perito Criminal/Química - SGA/AC - 2008) As curvas de distribuição normal mostradas abaixo têm aproximadamente o mesmo desvio-padrão e médias significativamente distintas.



Comentários

As curvas acima apresentam a mesma média, 100. A largura de uma curva gaussiana está relacionada ao desvio padrão ou à variância da população. Portanto, podemos afirmar que os desvios padrão são distintos, mas as médias são iguais.

Resposta: errado

(CESPE - Perito Criminal/Química - SGA/AC - 2008) O intervalo de confiança μ de uma medida pode ser corretamente definido pela seguinte expressão:

$$\mu = \bar{x} \pm \frac{ts}{n},$$

em que $\bar{x}$ representa a média, t , o parâmetro de Student, s , o desvio-padrão e n , o número de medidas.

Comentários

Conforme estudamos, o intervalo de confiança, a partir da distribuição *t-Student*, pode ser calculada por:

$$\bar{x} - t_{n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1} \frac{s}{\sqrt{n}}$$

Portanto, o item está errado porque o denominador do último termo seria $\sqrt{n}$ e não n somente.

Resposta: errado

(FUMARC - Analista da Polícia Civil/Estatística - PC-MG - 2013) Sejam $n_1 = 8$ e $n_2 = 13$ os tamanhos de duas amostras coletadas com o objetivo de testar a igualdade entre suas médias. Os desvios-padrão obtidos foram, respectivamente, $s_1 = 10$ e $s_2 = 15$. Para testar a igualdade das variâncias, foi realizado um teste F. Os valores dos graus de liberdade da distribuição F e do valor calculado da estatística de teste F são, respectivamente:

a) 7; 12; 1,50

b) 12; 7; 1,50

c) 12; 7; 2,25

d) 13; 8; 0,44

Comentários

Para obtermos os graus de liberdade, basta subtrairmos 1 do tamanho das amostras ($gl=n-1$), como segue: $gl_1 = 7$ e $gl_2 = 12$. O valor de calculado de F é obtido pela equação abaixo:



$$F_{\text{calculado}} = \frac{s_{\text{maior}}^2}{s_{\text{menor}}^2}$$

$$F_{\text{calculado}} = \frac{15^2}{10^2} = 2,25$$

Resposta: letra C

(FUNCAB - Perito Criminalístico/Eng. Química ou Química ou Química Industrial - POLITEC-SE - 2014)
O teste de *Student* é uma ferramenta estatística. Um conjunto de dados com 9 determinações, cuja média é igual a 5,0, o valor verdadeiro 4,0 e o desvio padrão 2,0 apresenta o valor de igual a:

- a) 5,0
- b) 1,5
- c) 3,0
- d) 0,5
- e) 1,0

Comentários

Se desejamos comparar valores amostrais com um valor verdadeiro (=média populacional), então estamos diante de um teste t-Student da situação (1) Comparação de médias entre uma amostra ($\bar{x}$) e uma população (μ) de variância desconhecida. Nessa situação, o $t_{\text{calculado}}$ é calculado como segue:

$$t_{\text{calculado}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Aplicando os valores fornecidos pelo enunciado, temos:

$$t_{\text{calculado}} = \frac{5,0 - 4,0}{\frac{2}{\sqrt{9}}} = 1,5$$

Resposta: letra B

(CESGRANRIO - Especialista Técnico em Química - ANP - 2016) Considere que o teor de vitamina C foi determinado por dois métodos, sendo a titulação coulométrica usada como referência para testar um método espectrofotométrico.



Usaram-se quatro réplicas autênticas ($n = 4$) para cada método e limite de confiança de 95%, onde $t_{\text{tabelado}} = 3,18$.

Os intervalos de confiança obtidos com os métodos foram:

- Coulometria: $3,4 \pm 0,2 \text{ mg L}^{-1}$
- Espectrofotometria: $5,8 \pm 1,8 \text{ mg L}^{-1}$

Analisando os resultados para o método espectrofotométrico verifica-se que.

- a) houve interferência, caso não tenha ocorrido erro grosseiro do analista.
- b) não existiu erro sistemático, caso não tenha ocorrido erro grosseiro do analista.
- c) não existe possibilidade de ter havido erro grosseiro do analista.
- d) a precisão foi similar ao do método de referência.
- e) o resultado foi estatisticamente igual ao do método de referência.

Comentários

Conforme já discutido em questão anterior, diante de resultados muito discrepantes, como é o caso do enunciado, não haverá dúvidas que as médias são diferentes entre si e as variâncias ou desvios padrão também. De certo, o resultado dos testes t e F indicarão que os resultados são estatisticamente diferentes. Portanto, letras D e E são incorretas.

Como os resultados são distintos, pode ter ocorrido algum erro grosseiro, o que torna a letra C incorreta.

Erro grosseiro e erro sistemático são independentes. Ou seja, um não depende do outro para ocorrer. Portanto, letra B é incorreta.

Por fim, a letra A é correta porque caso não tenha ocorrido erro grosseiro, podemos associar a discrepância de resultado a alguma interferência no método espectrofotométrico, que é o método testado.

Resposta: letra A

(NEC - Químico - UFMA - 2015) Sobre o tratamento estatístico de dados e suas aplicações na análise química, considere as seguintes afirmações:

I - O teste de Fisher (teste F) serve para representar intervalos de confiança e verificar se o desvio padrão de dois resultados são similares.

II - O teste t de Student serve para comparar um grupo de resultados com outro, se são ou não diferentes.



III - O desvio padrão de uma amostra é a raiz quadrada da variância, podendo, então, ser um valor positivo ou negativo.

As afirmações corretas são:

- a) II e III
- b) I e III
- c) somente a III
- d) somente a II
- e) I, II e III

Comentários

Afirmativa I: incorreta. O teste F é utilizado para comparação de variâncias ou desvios padrão.

Afirmativa II: correta. O teste t-Student é utilizado para comparação de grupos de resultados por meio de suas médias.

Afirmativa III: incorreta. O desvio padrão é a raiz quadrada da variância. Entretanto, a questão falha porque o desvio padrão é sempre positivo.

Resposta: letra D

(CESPE - Especialista em Regulação - ANP - 2013) A amostragem é o processo de coletar uma pequena massa de um material cuja composição represente exatamente o todo do material que está sendo amostrado, podendo ser do tipo representativa ou aleatória.

Comentários

A questão peca em utilizar o termo “exatamente”, pois a amostra apresenta características relevantes da população na mesma proporção em que elas ocorrem na própria população. Podemos dizer que amostra representativa possui características semelhantes à população, mas sempre haverá algum pequeno desvio da amostra em relação à população, que é inerente aos processos de amostragem.

Resposta: errado

(FUNIVERSA - Papiloscopista Policial - PCDF - 2015) A distribuição normal é uma das mais importantes distribuições de probabilidade da estatística. Também conhecida como distribuição *gaussiana*, apresenta propriedades importantes para a realização de modelagens e inferências estatísticas sobre diversas variáveis estudadas. Em relação à distribuição normal e a suas propriedades, assinale a alternativa correta.



- a) A distribuição normal é específica para variáveis aleatórias quantitativas discretas.
- b) Para uma variável que possui distribuição normal, a média e a mediana são iguais, mas a moda é diferente.
- c) A média, a mediana e a moda para uma variável que possui esta distribuição são iguais.
- d) A distribuição normal é assimétrica à direita.
- e) A distribuição normal é assimétrica à esquerda.

Comentários

Letra A: incorreta. Estudamos que a distribuição normal é contínua e não discreta.

Letra B: incorreta. Uma das características da distribuição normal é apresentar moda=mediana=média.

Letra C: correta. Conforme discutido na letra B.

Letra D: incorreta. A distribuição normal é simétrica.

Letra E: incorreta. A distribuição normal é simétrica.

Resposta: letra C

(COMPERVE - Téc. de Laboratório/Química - UFRN - 2015) O valor verdadeiro da concentração de uma solução é $0,1085 \text{ mol L}^{-1}$. Mas o valor encontrado por um técnico, na sua determinação, foi de $0,1065 \text{ mol L}^{-1}$. Nesse caso, os valores para o erro absoluto e o erro relativo são, respectivamente:

- a) $+0,0018 \text{ mol L}^{-1}$ e $2,0\%$.
- b) $+0,0020 \text{ mol L}^{-1}$ e $1,84\%$.
- c) $-0,0018 \text{ mol L}^{-1}$ e $2,0\%$.
- d) $-0,0020 \text{ mol L}^{-1}$ e $1,84\%$.

Comentários

Aplicando os valores nas equações do erro absoluto e relativo, respectivamente, temos:

$$e_i = x_i - x_{(\text{real})}$$

$$e_i = 0,1065 - 0,1085$$



$$e_i = -0,002 \text{ mol} \cdot \text{L}^{-1}$$

$$e_i(\%) = \frac{e_i}{x_{\text{real}}} \cdot 100$$

$$e_i(\%) = \frac{-0,002 \text{ mol} \cdot \text{L}^{-1}}{0,1085 \text{ mol} \cdot \text{L}^{-1}} \cdot 100$$

$$e_i(\%) = -1,84\%$$

Resposta: letra D

(PUCPR - Analista em Desenvolvimento Tecnológico - TECPAR - 2014) As ferramentas estatísticas são importantes para a validação de métodos de ensaio. Assinale a alternativa que indica uma ferramenta estatística que possibilita determinar se dois desvios-padrão possuem diferença significativa entre si.

- a) Teste de Stwart.
- b) Teste de Peterson.
- c) Teste N.
- d) Teste F.
- e) Teste de Chapman.

Comentários

Vimos que o teste F é utilizado para comparação de variância. Estudamos ainda que a variância é o quadrado do desvio padrão. Portanto, podemos transformar desvios padrão em variância e compará-los por meio do teste F.

Resposta: letra D

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2012) Sobre os erros associados aos resultados produzidos por um método analítico e as maneiras de avaliá-los, é correto

- a) avaliar o erro sistemático avaliando a variância associada ao resultado.
- b) corrigir um erro sistemático pela calibração do instrumento.
- c) comparar as variâncias de dois resultados usando um teste de rejeição como o teste Q.
- d) eliminar o erro aleatório com a repetição da análise.



e) eliminar o erro sistemático aplicando um teste t de *Student*.

Comentários

Letra A: incorreta. O erro sistemático é obtido comparando um valor com o respectivo valor verdadeiro ou certificado ou teórico. Por outro lado, a variância e o desvio padrão estão relacionados aos resultados de repetições.

Letra B: correta. Conforme exemplificamos, uma balança analítica descalibrada pode produzir sistematicamente o mesmo erro, deslocando o resultado da massa sempre na mesma direção. Nesses casos, a calibração do instrumento é uma maneira de evitar e/ou eliminar o erro sistemático.

Letra C: incorreta. Conforme estudamos, se utiliza o teste F, não o teste Q, para comparar variâncias.

Letra D: incorreta. O erro aleatório é inerente às medidas e, por isso, não pode ser eliminado, apenas, em alguns casos, minimizado. Mesmo os equipamentos mais precisos apresentam um pequeno erro aleatório.

Letra E: incorreta. O teste t de *Student* é utilizado para comparação de médias amostrais ou para comparação de uma média e um valor verdadeiro quando não é conhecida a variância populacional.

Resposta: letra B

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2014) Um método analítico foi desenvolvido e usado por dois técnicos diferentes. Cada um deles realizou três análises independentes para a determinação de um analito (cujo valor certificado é $2,0 \pm 0,1$ mg/g) num material de referência. O primeiro técnico obteve três resultados cujos valores foram 2,5 mg/g, 2,4 mg/g e 2,6 mg/g. O segundo analista obteve 9,9 mg/g, 9,8 mg/g e 9,7 mg/g nas suas análises. Com base nas informações apresentadas, NÃO se pode concluir que

- a) em ambos os conjuntos de resultado se tem erro aleatório.
- b) o segundo analista cometeu um erro grosseiro.
- c) o erro sistemático do primeiro analista foi de 2,5 mg/g.
- d) os dois analistas obtiveram resultados com erro sistemático.
- e) os desvios padrões são semelhantes nos dois conjuntos de dados.

Comentários

Letra A: correta. Conforme estudamos, em todas medidas há erro aleatório, podendo ser maior ou menor, a depender da precisão do equipamento ou do método.

Letra B: correta. Note que os valores 9,9; 9,8 e 9,7 mg/g são muito distantes do valor ou faixa certificada $2,0 \pm 0,1$ mg/g, distanciamento esse que é característico de erro grosseiro.



Letra C: incorreta. O erro sistemático do primeiro analista é 0,5 mg/g pode ser obtido pela diferença entre o valor médio por ele obtido, 2,5 mg/g, e o valor certificado, 2,0 mg/g.

Letra D: correta. Os dois valores se distanciaram do valor certificado. Outra maneira de verificar a ocorrência de erro aleatório, é verificar se os valores estão fora do intervalo tolerável pela certificação, que nesse caso é $2,0 \pm 0,1$ mg/g, ou seja, de 1,9 a 2,1 mg/g. Qualquer valor fora dessa faixa poderá configurar um erro sistemático.

Letra E: correta. Aplicando os resultados na fórmula do desvio padrão amostral (s), obtemos 0,1 mg/g para ambos analistas.

Resposta: letra C

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2014) Na tabela abaixo são apresentados os resultados obtidos na determinação do valor de pH de amostras de água mineral.

Amostra	1	2	3	4	5
pH	6,0	6,4	6,6	6,8	7,2

Esse conjunto de dados analíticos

- a) é simétrico
- b) não apresenta desvio padrão.
- c) apresenta os valores da média aritmética e da mediana do conjunto diferentes.
- d) apresenta dispersão dos resultados igual a 2,0 unidades de pH.
- e) possui valores suficientes para afirmar que a distribuição da amostra é normal.

Comentários

Letra A: correta. Dizer que os dados são simétricos significa que os valores acima e abaixo da média estão dela equidistantes. Na sequência crescente apresentada no enunciado, 6,4 e 6,8 estão equidistantes da média 6,6, a distância de -0,2 (6,4-6,6) e de +0,2, respectivamente. O mesmo acontece para os valores 6,0 e 7,2 que se distanciam da média em -0,6 e +0,6, respectivamente (caso a ordem fornecida não fosse crescente, deveríamos ordená-las).

Letra B: incorreta. Qualquer série de valores, em que pelo menos 1 valor é diferente de outro valor, apresentará desvio padrão e variância diferente de ZERO. Conforme estudamos, o desvio padrão é uma medida de dispersão e só será nulo caso todos os valores da série analisada sejam iguais entre si.

Letra C: incorreta. Para esse conjunto de valores, média e a mediana (valor central) são iguais a 6,6.

Letra D: incorreta. O avaliador utilizou variação como sinônimo de desvio padrão. Aplicando os valores individuais na equação do desvio padrão amostral, obtemos 0,45.



Letra E: incorreta. Não será foco de nossa aula, mas, em geral, é necessário um levantamento amostral (tamanho de amostra) bem mais abrangente (maior) para inferir se o comportamento de uma população é normal.

Resposta: letra A

(FGV - Tecnologista em Saúde/Estatística - FIOCRUZ - 2010) Os níveis de glicose foram medido em dois grupos de indivíduos, sendo o grupo 1 formado por indivíduos sedentários e o grupo 2 por indivíduos não sedentários. O nível médio de glicemia para o grupo 1 foi de 98 mg/dL e para o grupo 2 foi de 110 mg/dL. Para determinar se a diferença entre essas medidas é significativa, o teste estatístico mais apropriado é:

- a) teste Normal
- b) teste t
- c) teste chi-quadrado
- d) teste F (ANOVA)
- e) teste log-rank

Comentários

Vimos que o teste t-Student é utilizado para comparação de médias.

Resposta: letra B

(FCC - Especialista em Regulação/Estatística - ANS - 2007) Um órgão municipal quer analisar o número de acidentes de trânsito em dois bairros do município de São Paulo no período de férias e acompanhou as ocorrências com vítimas no mês de janeiro (31 dias), obtendo a seguinte tabela:

Bairro	Número médio de acidentes	Desvio padrão do número de acidentes
A	2,13	1,31
B	2,54	1,15

Supondo que o número de acidentes tenha distribuição normal, o teste de hipótese adequado para comparar as médias do número de acidentes destes dois bairros é

- a) Teste t-Student.
- b) Teste de qui-quadrado.
- c) Teste MacNemar.



d) Teste F.

e) Teste de Bartlett.

Comentários

Conforme estudamos, o teste t-Student é utilizado para comparação de médias.

Resposta: letra A



LISTA DE QUESTÕES DA AULA

Texto referente às próximas duas questões:

elemento	concentração em $\mu\text{g/g}$		desvio-padrão
	casaco	vidraça	s
As	132	122	9,7
Co	0,54	0,61	0,026
La	4,01	3,60	0,20
Sb	2,81	2,77	0,26
Th	0,62	0,75	0,044

Um perito criminal recebeu em seu laboratório, como principal evidência em um caso criminal, pequenos fragmentos de vidro encontrados incrustados no casaco de um suspeito de assassinato. Esses fragmentos são idênticos em composição a uma rara vidraça belga de vidro manchado quebrada durante o crime. O perito decidiu então determinar os elementos As, Co, La, Sb e Th no vidro incrustado no casaco do suspeito para verificar se este era do mesmo material da vidraça belga. A técnica escolhida para essas determinações foi a espectroscopia de absorção atômica. As médias e os desvios-padrão das análises em triplicata desses cinco elementos nas amostras de vidro retiradas do casaco, bem como os valores conhecidos para a vidraça belga são mostrados na tabela acima.

Considerando essa situação hipotética, que $\sqrt{3} = 1,73$ e que o parâmetro t de *Student* para 2 graus de liberdade e 95% de confiança é igual a 4,303, julgue os itens a seguir:

1. (CESPE - Perito Criminal Federal/Química - PF - 2004) O desvio-padrão s para qualquer um dos elementos analisados pode ser corretamente calculado por meio da equação abaixo, em que x_1 , x_2 e x_3 são os valores individuais de concentração obtidos para cada alíquota analisada.

$$s = \sqrt{\frac{\left(x_1 - \frac{x_1 + x_2 + x_3}{3}\right)^2 + \left(x_2 - \frac{x_1 + x_2 + x_3}{3}\right)^2 + \left(x_3 - \frac{x_1 + x_2 + x_3}{3}\right)^2}{2}}$$

(CESPE - Perito Criminal Federal/Química - PF - 2004) Para um nível de confiança de 95% como critério de dúvida, com base nos dados obtidos, o perito tem argumentos suficientes para duvidar da culpabilidade do acusado, pois pelo menos dois dos elementos analisados possuem concentrações significativamente diferentes nos fragmentos de vidro incrustados no casaco, em comparação com a vidraça de vidro belga.

(FUNCAB - Perito Criminal/Farmácia - POLITEC-MT - 2013) Uma análise química é o conjunto de técnicas de laboratório utilizado na identificação das espécies químicas e na determinação de informações químicas ou físicas sobre uma amostra. No que diz respeito à preparação de uma amostra, analise as afirmativas a seguir:



- I. A amostragem é o processo usado para coleta de uma amostra representativa para análise.
- II. Materiais heterogêneos irão sempre produzir amostras heterogêneas para análise também, tendo, obrigatoriamente, a mesma composição do material.
- III. A preparação de uma amostra é o processo pelo qual se converte uma amostra bruta em uma amostra de laboratório homogênea.
- IV. A amostra bruta tem que ser representativa de um lote, e sua escolha deve ser crítica para que a análise seja válida.

Estão corretas apenas as afirmativas:

- a) I, II e III.
- b) I, III e IV.
- c) II, III e IV.
- d) I, II e IV.
- e) I e IV.

CESGRANRIO - Téc. de Laboratório/Química - CEFET-RJ - 2014) Os resultados (cinco réplicas) obtidos a partir de uma determinação da concentração de um analito em uma amostra são mostrados na Tabela abaixo.

Réplica	1	2	3	4	5
Resultado (em $\mu\text{g} \cdot \text{g}^{-1}$)	10,5	10,4	10,5	10,5	10,6

Se o resultado verdadeiro para o analito na amostra é $10,0 \mu\text{g} \cdot \text{g}^{-1}$, é possível concluir que o(a)

- a) erro sistemático é $0,5 \mu\text{g} \cdot \text{g}^{-1}$.
- b) desvio padrão é menor que $0,005 \mu\text{g} \cdot \text{g}^{-1}$.
- c) resultado da análise é exato.
- d) valor médio é diferente da mediana.
- e) mediana é $10,4 \mu\text{g} \cdot \text{g}^{-1}$.

(FUNCAB - Perito Oficial Criminal-Biologia/Química/Farmácia POLITEC-MT - 2013) Calcule o desvio-padrão do conjunto de valores (1, 2, 1, 2, 3, 1, 2, 3, 5, 10) e, em seguida, assinale a alternativa correta.

- a) $2,0 < d < 2,1$



- b) $2,1 < d < 2,2$
- c) $2,4 < d < 2,5$
- d) $2,5 < d < 2,6$
- e) $2,6 < d < 2,7$

(UECE - Perito Criminal/Eng. Química - PCCE - 2003) Na análise volumétrica, um dos problemas é a ocorrência de erro na titulação. Se 30,0 mL de HCl 0,100 M são titulados com NaOH 0,100 M e uma solução de vermelho de metila é usada como indicador, o erro de titulação, admitindo-se $\text{pH} = 4$ no ponto final, é:

Obs.: O ponto final ocorre antes do ponto de equivalência.

- a) +0,02%
- b) -0,10%
- c) -0,02%
- d) -0,20%

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2018) Um procedimento gravimétrico foi realizado de acordo com as instruções da literatura. Um precipitado foi obtido após a mistura da solução de amostra com a solução do reagente precipitante (esta última adicionada em quantidade para garantir excesso estequiométrico). Por conta das características do precipitado, uma fração ficou suspensa na solução (suspensão coloidal) e foi perdida na filtração para recolher o sólido. Após secagem, a massa final do precipitado foi medida em balança calibrada. O procedimento foi repetido três vezes para se obter um valor médio de massa de analito e seu desvio padrão.

Com relação aos erros associados ao resultado, tem-se

- a) apenas erro sistemático
- b) apenas erro aleatório
- c) apenas erro instrumental (da balança)
- d) tanto erro aleatório quanto erro sistemático
- e) que não há erro



(CESPE - Perito Criminal/Química - SGA/AC - 2008) Se uma grandeza x qualquer é calculada pela seguinte equação: $x = \frac{a \cdot b}{c}$, em que a , b e c são grandezas medidas com o fim de se obter o valor de x , então o desvio-padrão de x , s_x , pode ser corretamente calculado pela seguinte expressão:

$$s_x = x \sqrt{\left(\frac{s_a}{a}\right)^2 + \left(\frac{s_b}{b}\right)^2 + \left(\frac{s_c}{c}\right)^2},$$

em que s_a , s_b e s_c representam, respectivamente, os desvios-padrão das grandezas a , b e c .

Texto referente aos próximos dois exercícios:

indicador	pH de viragem	número de medidas	t
fenoftaleína	8,2 a 10,0	3	4,30
vermelho de metila	4,4 a 6,2	4	3,18
azul de bromotimol	6,0 a 7,6	5	2,78

Tabela 1

Tabela 2

Considere que uma solução de NaOH foi padronizada por meio da titulação, em quadruplicata, com 20,00 mL de uma solução padrão de biftalato de potássio 0,1000 mol/L e que a concentração média encontrada para a solução de NaOH foi de 0,080 mol/L, com uma estimativa do desvio padrão de $\pm 0,002$ mol/L. Considere, ainda, a tabela 1, em que é apresentado o pH de viragem de alguns indicadores ácido-base, e a tabela 2, em que são apresentados os valores do parâmetro t de *student* para uma probabilidade de 95%.

Com base nessas informações, julgue os itens seguintes:

(CESPE - Químico - FUB - 2011) Considerando-se a inexistência de erros sistemáticos, é correto afirmar que a probabilidade de que o valor real da concentração de NaOH esteja no intervalo compreendido entre 0,078 e 0,082 será maior que 95%.

(CESPE - Químico - FUB - 2011) Entre os indicadores apresentados na tabela 1, o mais adequado à titulação em questão é o vermelho de metila.

(FUNIVERSA - Papiloscopista - PCGO - 2015) Um estudo realizado com 200 policiais civis, após a realização de testes em que cada profissional deveria tomar decisões importantes referentes a diversos fatores do cotidiano, avaliou características psicológicas atenuantes. Como a variável idade se apresenta como um fator determinante na atitude do profissional, fez-se necessária a elaboração de uma tabela de distribuição de frequência das idades do grupo em questão.

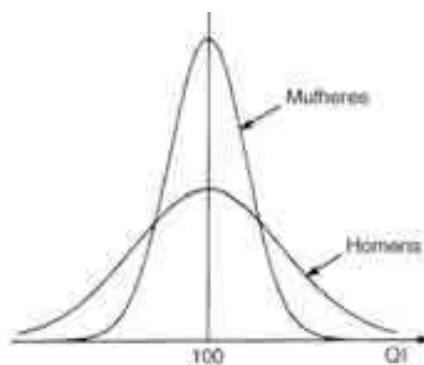


Classe das idades	N.º de policiais
21 - 26	56
26 - 31	48
31 - 36	45
36 - 41	23
41 - 46	15
46 - 51	13
Total	200

Com base nesse caso hipotético, assinale a alternativa que apresenta o ponto médio da classe mediana.

- a) 23,5
- b) 28,5
- c) 33,5
- d) 38,5
- e) 43,5

(CESPE - Perito Criminal/Química - SGA/AC - 2008) As curvas de distribuição normal mostradas abaixo têm aproximadamente o mesmo desvio-padrão e médias significativamente distintas.



(CESPE - Perito Criminal/Química - SGA/AC - 2008) O intervalo de confiança μ de uma medida pode ser corretamente definido pela seguinte expressão:

$$\mu = \bar{x} \pm \frac{ts}{n},$$

em que $\bar{x}$ representa a média, t , o parâmetro de Student, s , o desvio-padrão e n , o número de medidas.

(FUMARC - Analista da Polícia Civil/Estatística - PC-MG - 2013) Sejam $n_1 = 8$ e $n_2 = 13$ os tamanhos de duas amostras coletadas com o objetivo de testar a igualdade entre suas médias. Os desvios-padrão obtidos foram, respectivamente, $s_1 = 10$ e $s_2 = 15$. Para testar a igualdade das variâncias, foi realizado



um teste F. Os valores dos graus de liberdade da distribuição F e do valor calculado da estatística de teste F são, respectivamente:

- a) 7; 12; 1,50
- b) 12; 7; 1,50
- c) 12; 7; 2,25
- d) 13; 8; 0,44

(FUNCAB - Perito Criminalístico/Eng. Química ou Química ou Química Industrial - POLITEC-SE - 2014)
O teste de *Student* é uma ferramenta estatística. Um conjunto de dados com 9 determinações, cuja média é igual a 5,0, o valor verdadeiro 4,0 e o desvio padrão 2,0 apresenta o valor de igual a:

- a) 5,0
- b) 1,5
- c) 3,0
- d) 0,5
- e) 1,0

(CESGRANRIO - Especialista Técnico em Química - ANP - 2016) Considere que o teor de vitamina C foi determinado por dois métodos, sendo a titulação coulométrica usada como referência para testar um método espectrofotométrico.

Usaram-se quatro réplicas autênticas ($n = 4$) para cada método e limite de confiança de 95%, onde $t_{\text{tabelado}} = 3,18$.

Os intervalos de confiança obtidos com os métodos foram:

- Coulometria: $3,4 \pm 0,2 \text{ mg L}^{-1}$
- Espectrofotometria: $5,8 \pm 1,8 \text{ mg L}^{-1}$

Analisando os resultados para o método espectrofotométrico verifica-se que.

- a) houve interferência, caso não tenha ocorrido erro grosseiro do analista.
- b) não existiu erro sistemático, caso não tenha ocorrido erro grosseiro do analista.
- c) não existe possibilidade de ter havido erro grosseiro do analista.



- d) a precisão foi similar ao do método de referência.
- e) o resultado foi estatisticamente igual ao do método de referência.

(NEC - Químico - UFMA - 2015) Sobre o tratamento estatístico de dados e suas aplicações na análise química, considere as seguintes afirmações:

I - O teste de Fisher (teste F) serve para representar intervalos de confiança e verificar se o desvio padrão de dois resultados são similares.

II - O teste t de Student serve para comparar um grupo de resultados com outro, se são ou não diferentes.

III - O desvio padrão de uma amostra é a raiz quadrada da variância, podendo, então, ser um valor positivo ou negativo.

As afirmações corretas são:

- a) II e III
- b) I e III
- c) somente a III
- d) somente a II
- e) I, II e III

(CESPE - Especialista em Regulação - ANP - 2013) A amostragem é o processo de coletar uma pequena massa de um material cuja composição represente exatamente o todo do material que está sendo amostrado, podendo ser do tipo representativa ou aleatória.

(FUNIVERSA - Papiloscopista Policial - PCDF - 2015) A distribuição normal é uma das mais importantes distribuições de probabilidade da estatística. Também conhecida como distribuição *gaussiana*, apresenta propriedades importantes para a realização de modelagens e inferências estatísticas sobre diversas variáveis estudadas. Em relação à distribuição normal e a suas propriedades, assinale a alternativa correta.

- a) A distribuição normal é específica para variáveis aleatórias quantitativas discretas.
- b) Para uma variável que possui distribuição normal, a média e a mediana são iguais, mas a moda é diferente.
- c) A média, a mediana e a moda para uma variável que possui esta distribuição são iguais.
- d) A distribuição normal é assimétrica à direita.



e) A distribuição normal é assimétrica à esquerda.

(COMPERVE - Téc. de Laboratório/Química - UFRN - 2015) O valor verdadeiro da concentração de uma solução é $0,1085 \text{ mol L}^{-1}$. Mas o valor encontrado por um técnico, na sua determinação, foi de $0,1065 \text{ mol L}^{-1}$. Nesse caso, os valores para o erro absoluto e o erro relativo são, respectivamente:

- a) $+0,0018 \text{ mol L}^{-1}$ e 2,0%.
- b) $+0,0020 \text{ mol L}^{-1}$ e 1,84 %.
- c) $-0,0018 \text{ mol L}^{-1}$ e 2,0 %.
- d) $-0,0020 \text{ mol L}^{-1}$ e 1,84 %.

(PUCPR - Analista em Desenvolvimento Tecnológico - TECPAR - 2014) As ferramentas estatísticas são importantes para a validação de métodos de ensaio. Assinale a alternativa que indica uma ferramenta estatística que possibilita determinar se dois desvios-padrão possuem diferença significativa entre si.

- a) Teste de Stwart.
- b) Teste de Peterson.
- c) Teste N.
- d) Teste F.
- e) Teste de Chapman.

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2012) Sobre os erros associados aos resultados produzidos por um método analítico e as maneiras de avaliá-los, é correto

- a) avaliar o erro sistemático avaliando a variância associada ao resultado.
- b) corrigir um erro sistemático pela calibração do instrumento.
- c) comparar as variâncias de dois resultados usando um teste de rejeição como o teste Q.
- d) eliminar o erro aleatório com a repetição da análise.
- e) eliminar o erro sistemático aplicando um teste t de *Student*.

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2014) Um método analítico foi desenvolvido e usado por dois técnicos diferentes. Cada um deles realizou três análises independentes para a determinação de um analito (cujo valor certificado é $2,0 \pm 0,1 \text{ mg/g}$) num material de referência. O primeiro técnico obteve três resultados cujos valores foram 2,5 mg/g, 2,4 mg/g e 2,6 mg/g. O segundo



analista obteve 9,9 mg/g, 9,8 mg/g e 9,7 mg/g nas suas análises. Com base nas informações apresentadas, **NÃO** se pode concluir que

- a) em ambos os conjuntos de resultado se tem erro aleatório.
- b) o segundo analista cometeu um erro grosseiro.
- c) o erro sistemático do primeiro analista foi de 2,5 mg/g.
- d) os dois analistas obtiveram resultados com erro sistemático.
- e) os desvios padrões são semelhantes nos dois conjuntos de dados.

(CESGRANRIO - Téc. Químico de Petróleo Jr. - Petrobras - 2014) Na tabela abaixo são apresentados os resultados obtidos na determinação do valor de pH de amostras de água mineral.

Amostra	1	2	3	4	5
pH	6,0	6,4	6,6	6,8	7,2

Esse conjunto de dados analíticos

- a) é simétrico
- b) não apresenta desvio padrão.
- c) apresenta os valores da média aritmética e da mediana do conjunto diferentes.
- d) apresenta dispersão dos resultados igual a 2,0 unidades de pH.
- e) possui valores suficientes para afirmar que a distribuição da amostra é normal.

(FGV - Tecnologista em Saúde/Estatística - FIOCRUZ - 2010) Os níveis de glicose foram medido em dois grupos de indivíduos, sendo o grupo 1 formado por indivíduos sedentários e o grupo 2 por indivíduos não sedentários. O nível médio de glicemia para o grupo 1 foi de 98 mg/dL e para o grupo 2 foi de 110 mg/dL. Para determinar se a diferença entre essas medidas é significativa, o teste estatístico mais apropriado é:

- a) teste Normal
- b) teste t
- c) teste chi-quadrado
- d) teste F (ANOVA)
- e) teste log-rank



(FCC - Especialista em Regulação/Estatística - ANS - 2007) Um órgão municipal quer analisar o número de acidentes de trânsito em dois bairros do município de São Paulo no período de férias e acompanhou as ocorrências com vítimas no mês de janeiro (31 dias), obtendo a seguinte tabela:

Bairro	Número médio de acidentes	Desvio padrão do número de acidentes
A	2,13	1,31
B	2,54	1,15

Supondo que o número de acidentes tenha distribuição normal, o teste de hipótese adequado para comparar as médias do número de acidentes destes dois bairros é

- a) Teste t-Student.
- b) Teste de qui-quadrado.
- c) Teste MacNemar.
- d) Teste F.
- e) Teste de Bartlett.



GABARITO



1	Certo	14	C
2	Certo	15	B
3	B	16	A
4	A	17	D
5	E	18	Errado
6	D	19	C
7	D	20	D
8	Certo	21	D
9	Errado	22	B
10	Errado	23	C
11	B	24	A
12	Errado	25	B
13	Errado	26	A



PRINCIPAIS PONTOS DA AULA

Introdução à estatística

Exemplos de medidas experimentais: pesquisas eleitorais, pesagens repetitivas de um peso padrão, replicatas de uma amostra cujo analito foi determinado por titulação, replicatas da obtenção da absorbância de uma amostra por espectrofotometria.

Estatística descritiva consiste na etapa inicial de qualquer análise. Nesta etapa é realizada a sumarização dos dados, deixando-os mais sintéticos e entendíveis de forma mais imediata. Para tanto, se utiliza de gráficos, tabelas e medidas descritivas de posição e dispersão.

População: qualquer seleção de indivíduos ou valores, finita ou infinita.

Amostra: uma parte da população, normalmente selecionada com objetivo de se fazer inferências* sobre a população.

**inferência estatística é o ato de fazer afirmações sobre a população ou tomar decisões a partir dos valores obtidos de uma amostra representativa. Para se fazer uma inferência, admite-se que a amostra e a população guardam similaridades fortes entre si quanto às suas características e proporções.*

Amostra representativa: apresenta as características relevantes da população na mesma proporção em que elas ocorrem na própria população.

Amostra aleatória: amostra de N valores ou indivíduos obtidos de tal forma que todos os possíveis conjuntos de N valores da população tenham a mesma chance de ser conhecidos.

Tipos de amostragem

Amostragem aleatória ou casual: cada indivíduo da população possui a mesma probabilidade de ser selecionado.

Amostragem sistemática: retira-se indivíduos em intervalos constantes (k), ou seja, a cada k elementos.

Amostragem estratificada: de início, divide-se a população em estratos ou grupos e, na sequência, realiza-se a amostragem em cada grupo separadamente.



Amostragem por conglomerados: de início, divide-se a população em seções ou estratos, seleciona-se algumas delas e coleta dados de todos os indivíduos das seções selecionadas.

Amostragem por conveniência: é definida pela facilidade de coleta de dados

Medidas de posição ou tendência central

Medidas de posição: são variáveis quantitativas que indicam a posição do conjunto de observações (da amostra).

Medidas de dispersão: são variáveis quantitativas que indicam a dispersão, ou espalhamento, do conjunto de valores da amostra.

Medidas de posição ou de tendência central

Média aritmética (média, $\bar{x}$): é a soma das observações (x_i), dividida pelo número (n) delas. Usa-se símbolo μ quando a média aritmética for relativa aos dados populacionais

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

Caso tenhamos repetições dos valores das observações de diferentes indivíduos, do tipo x_1 aparece n_1 , x_2 aparece n_2 , então, a equação acima pode ser reescrita como:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_k x_k}{n} = \frac{1}{n} \sum_{i=1}^n n_i x_i$$

A equação acima também é conhecida como média aritmética ponderada (média ponderada). Em uma prova de concurso, por exemplo, em que questões de conhecimentos específicos apresentam pesos diferentes das questões de conhecimentos gerais, você pode calcular seu desempenho médio substituindo n pelo peso (p) de cada tipo de questão, como segue:

$$\bar{x} = \frac{p_1 x_1 + p_2 x_2 + \cdots + p_k x_k}{p_1 + p_2 + \cdots + p_k} = \frac{1}{p} \sum_{i=1}^n p_i x_i$$



Por fim, se $f_i = n_i/n$ representa a frequência relativa da observação x_i , então podemos obter uma equação para a média que se relacione com a frequência (note que aqui não se divide o somatório por n como nos casos anteriores):

$$\bar{x} = f_1x_1 + f_2x_2 + \cdots + f_kx_k = \sum_{i=1}^n f_i x_i$$

Mediana (md): é o valor que ocupa a posição central da série de observações quando estão ordenadas em ordem crescente. Quando o número de observações for par, é necessário calcular a média aritmética das duas observações centrais.

$$\text{Se } n \text{ ímpar, } md = x_{\left(\frac{n+1}{2}\right)}$$

$$\text{Se } n \text{ par, } md = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

Moda: é o valor mais frequente do conjunto de valores observados. Uma forma fácil de lembrar do conceito é relacioná-lo a uma roupa que "todo mundo" está usando porque está na moda.

Medidas de posição

Três tipos de erros: grosseiro, sistemático e aleatório. Em caso de dúvida, revise cada um pela explicação no próprio texto.

Erro (e): é a distância (diferença) entre um dado valor e a média ou entre um valor e o valor real ou verdadeiro (por ex: valor obtido por determinação analítica; valor certificado), conforme demonstrado abaixo.

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

$$e_i = x_i - x_{(\text{real})}$$

Em estatística, o erro também é conhecido como resíduo.

O erro também pode ser expresso em termos percentuais (erro relativo), o qual pode ser calculado pela expressão a seguir:

$$e_i(\%) = \frac{e_i}{x_{\text{real}}} \cdot 100$$



Amplitude (A): é a distância (diferença) entre o valor máximo e o valor mínimo para um conjunto de observações (amostra).

$$A = x_{\text{máx}} - x_{\text{mín}}$$

Variância: medida resumo e média da variabilidade dos dados obtida a partir dos quadrados dos resíduos.

Variância amostral (s^2):

$$s^2 = \frac{e_1^2 + e_2^2 + \dots + e_n^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Variância populacional (σ^2):

$$\sigma^2 = \frac{e_1^2 + e_2^2 + \dots + e_n^2}{n} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

CUIDADO: Note que o símbolo da variância e o denominador do lado direito são diferentes quando se trata de dados amostrais ou populacionais.

Desvio padrão: raiz quadrada da variância.

Desvio padrão amostral (s):

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

Desvio padrão populacional (σ):

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



Coeficiente de variação (CV): é uma medida de dispersão padronizada em relação à média (populacional ou amostral, a depender dos dados fornecidos), expressa em porcentagem. Também conhecido como desvio padrão relativo.

Coeficiente de variação amostral:

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

Coeficiente de variação populacional:

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

NOTA: Considerando que a padronização é realizada em relação à média, então um CV 20% significa que os dados variam, em média, 20% do valor da média em torno dela própria.

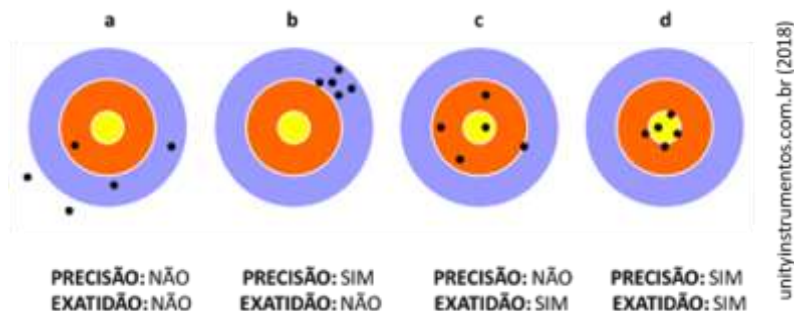
Diferença entre posição e exatidão

Precisão: está relacionada à reprodutibilidade de um método analítico ou de medição de um instrumento ou vidraria. Reprodutibilidade, por sua vez, de maneira bem objetiva, corresponde ao quão próximos se encontram as repetições. Por exemplo, estaríamos diante de uma determinação titulométrica precisa (reprodutível) caso fossem gastos 4,92; 4,93 e 4,92 mL de volume do titulante. O **desvio padrão** é um parâmetro que indica a **precisão**, sendo maior a precisão quanto menor for o desvio padrão.

Exatidão: está relacionada com a diferença entre o valor medido (experimental) e o valor real ou verdadeiro. Desta forma, a **exatidão** se relaciona com o erro. Maior será a exatidão, quanto menor for o erro. Valores certificados por organizações regulatórias como o NIST (*National Institute of Standards and Technology*) são exemplos de valores considerados verdadeiros. Suponhamos que nosso laboratório desenvolva um novo método de determinação de nitrogênio total em feijão. Podemos, então, comprar do NIST uma amostra certificada de feijão cujo teor de nitrogênio total é conhecido. Analisando esse material certificado em nosso laboratório, podemos calcular o erro (diferença entre os teores determinado e verdadeiro), que indicará a exatidão do novo método. Outro clássico exemplo de valor verdadeiro é a concentração teórica de um padrão primário de alta pureza.



Ilustração, em alvos, dos conceitos de precisão e exatidão, seguida de observações pertinentes



- a) Grande dispersão de resultados. Erros aleatórios elevados. Existência de erros sistemáticos. Resultado não preciso e não exato.
- b) Baixa dispersão de resultados. Erros aleatórios pequenos. Existência de erros sistemáticos. Resultado preciso, mas não exato.
- c) Grande dispersão de resultados. Erros aleatórios elevados. Não existência de erros sistemáticos. Resultado não preciso, mas exato.
- d) Baixa dispersão de resultados. Erros aleatórios pequenos. Não existência de erros sistemáticos. Resultado preciso e exato.

Testes de hipótese

Padronização:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

Precisamos entender quando aplicar cada um dos seguintes testes: **testes de hipótese: testes Z-score, t-student e F**. Confira abaixo:

- iii. **Para comparação de valor médio:** utiliza-se **teste Z-score** ou **t-student**. Quando for conhecido o comportamento populacional, ou seja, quando for conhecido μ e σ (média e desvio padrão populacionais), então deve-se utilizar a **distribuição normal** e, portanto, o **teste Z-score**, ou simplesmente **teste Z**. Como nem sempre é possível conhecer o comportamento populacional, então é possível fazer inferências (tomar decisões) a partir de dados amostrais ($\bar{x}$ e s), uma dessas maneiras é utilizar o **teste t**.

Observação: a **distribuição t de Student** tem formato de sino como a **distribuição normal**. A diferença entre elas é que a **função de densidade da distribuição t** é mais alargada que a da **distribuição z** porque é esperado maior variabilidade para dados amostrais, sendo que a largura da função está diretamente relacionada ao desvio padrão.

- iv. **Para comparação de variância ou de precisão de métodos:** utiliza-se o **teste F** que utiliza a **distribuição F de Fisher-Snedecor**. O comportamento da função densidade de probabilidade para essa distribuição é diferente de um sino, veja na figura abaixo.



Aplicando o teste Z

Hipótese nula (H_0): em que $\bar{x}_1 = \bar{x}_2$. Se passarmos $\bar{x}_2$ para o primeiro lado da igualdade, obteremos $\bar{x}_1 - \bar{x}_2 = 0$ e, por isso, o nome de **hipótese nula**.

Hipótese alternativa (H_1): em que $\bar{x}_1 \neq \bar{x}_2$.

O Z crítico é obtido da tabela, enquanto Z calculado é obtido pela equação abaixo:

$$Z_{\text{calculado}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Se $Z_{\text{calculado}} > Z_{\text{crítico}}$, então rejeitaríamos H_0 e afirmariamos que há evidências estatísticas da diferença dos valores comparados. Por outro lado, se $Z_{\text{calculado}} < Z_{\text{crítico}}$, então aceitaríamos H_0 e dizemos que não há evidências estatísticas que os valores comparados são diferentes.

Aplicando o teste t-student

4 possibilidades para o teste t-student. Quando usar cada uma delas?

Antes de discutirmos as 4 possibilidades, ressalto que é muito improvável que o examinador cobre o conhecimento memorizado das diferentes fórmulas do teste t-student. Entendo que o mais importante é saber identificar cada uma das 4 possibilidades (CUIDADO: como só há uma fórmula mais simples no teste Z e uma no teste F, então essas poderão ser exigidas. Sugiro que as memorize).

(1) Comparação de médias entre uma amostra ($\bar{x}$) e uma população (μ) de variância desconhecida. Nessa situação, o $t_{\text{calculado}}$ e grau de liberdade (gl) são calculados como segue:

$$t_{\text{calculado}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$gl = n - 1$$

Note que não conhecemos a variância populacional (σ^2). Caso fosse conhecido o valor σ^2 , então utilizaríamos o teste-Z.

(2) Comparação de médias de populações com variâncias diferentes e desconhecidas, a partir de médias amostrais $\bar{x}_1$ e $\bar{x}_2$.



$$t_{calculado} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1\bar{x}_2}}$$

onde $s_{\bar{x}_1\bar{x}_2}$ é o desvio padrão agrupado e é dado por:

$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$gl = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2+1}} - 2$$

(3) Comparação de médias de populações com variâncias desconhecidas e iguais, e tamanhos de amostras iguais ($n_1 = n_2$).

$$t_{calculado} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1\bar{x}_2} \sqrt{\frac{2}{n}}}$$

$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$

$$gl = 2n - 2$$

(4) Comparação de médias de populações com variâncias desconhecidas e iguais, mas tamanhos de amostras diferentes ($n_1 \neq n_2$).

$$t_{calculado} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1\bar{x}_2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$gl = n_1 + n_2 - 2$$



Aplicando o teste F

$$F_{\text{calculado}} = \frac{s_{\text{maior}}^2}{s_{\text{menor}}^2}$$

Intervalo de confiança

Intervalo de confiança para μ , a partir de uma observação x_i :

$$x_i - z\sigma < \mu < x_i + z\sigma$$

em que z é o valor da tabela que corresponde ao nível de confiança desejado para o intervalo.

Intervalo de confiança para μ , a partir de uma amostra (conjunto de indivíduos), utilizando a distribuição t-Student:

$$\bar{x} - t_{n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1} \frac{s}{\sqrt{n}}$$



ESSA LEI TODO MUNDO CONHECE: PIRATARIA É CRIME.

Mas é sempre bom revisar o porquê e como você pode ser prejudicado com essa prática.



1 Professor investe seu tempo para elaborar os cursos e o site os coloca à venda.



2 Pirata divulga ilicitamente (grupos de rateio), utilizando-se do anonimato, nomes falsos ou laranjas (geralmente o pirata se anuncia como formador de "grupos solidários" de rateio que não visam lucro).



3 Pirata cria alunos fake praticando falsidade ideológica, comprando cursos do site em nome de pessoas aleatórias (usando nome, CPF, endereço e telefone de terceiros sem autorização).



4 Pirata compra, muitas vezes, clonando cartões de crédito (por vezes o sistema anti-fraude não consegue identificar o golpe a tempo).



5 Pirata fere os Termos de Uso, adultera as aulas e retira a identificação dos arquivos PDF (justamente porque a atividade é ilegal e ele não quer que seus fakes sejam identificados).



6 Pirata revende as aulas protegidas por direitos autorais, praticando concorrência desleal e em flagrante desrespeito à Lei de Direitos Autorais (Lei 9.610/98).



7 Concurseiro(a) desinformado participa de rateio, achando que nada disso está acontecendo e esperando se tornar servidor público para exigir o cumprimento das leis.



8 O professor que elaborou o curso não ganha nada, o site não recebe nada, e a pessoa que praticou todos os ilícitos anteriores (pirata) fica com o lucro.



Deixando de lado esse mar de sujeira, aproveitamos para agradecer a todos que adquirem os cursos honestamente e permitem que o site continue existindo.