

## **Aula 00 - Profº Diego Souza**

*ITEP-RN (Perito Criminal - Toxicologia)  
Conhecimentos Específicos - 2021  
(Pós-Edital)*

Autor:

**Ana Cristina dos Santos Lopes,  
Diego Souza, Juliana da Silva  
Xavier, Rafaela Gomes**

12 de Abril de 2021

## Sumário

Estatística aplicada à química .....	2
1 – Estatística descritiva.....	2
1.1 - Definições iniciais da estatística (não precisa decorá-las, apenas compreendê-las) .....	3
1.2 - Medidas de posição ou de tendência central .....	6
1.2 - Medidas de dispersão.....	7
1.2 – Erros em análises laboratoriais (três tipos de erro) .....	9
1.3 – Diferença entre precisão e exatidão.....	11
2 – Estatística inferencial (testes de hipótese).....	15
2.1 – Histograma de frequência .....	15
2.2 – Distribuição normal.....	15
2.3 – Como usar a tabela da área da cauda da distribuição normal padronizada?.....	17
2.4 – Testes de hipótese: teste Z-score, t-student e teste F.....	18
2.5 – Intervalos de confiança para média populacional .....	23
Questões Comentadas .....	26
Gabarito .....	40



# ESTATÍSTICA APLICADA À QUÍMICA

## PDF SIMPLIFICADO

O assunto de hoje é estatística, com ênfase [é claro] na sua aplicação em dados analíticos (relativo às análises laboratoriais). 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).

Lembro que esta é a **versão simplificada da aula**, na qual destacarei os principais e mais importantes tópicos para sua prova, de maneira mais objetiva e concisa. É uma ótima opção para quem precisa empregar um ritmo de estudo mais acelerado ou para quem já tem um bom nível nos tópicos aqui ensinados e quer reforçá-los e revisá-los de maneira mais rápida.

Instagram: Prof.DiegoSouza

Telegram: t.me/prof.diegosouza

YouTube: Prof. Diego Souza

## 1 – Estatística descritiva

Exemplos de como as estatísticas (não só a descritiva, mas também a inferencial e a multivariada) podem ajudar em dados analíticos laboratoriais? (justificando sua importância)

- 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.



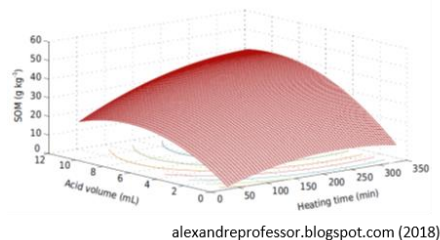
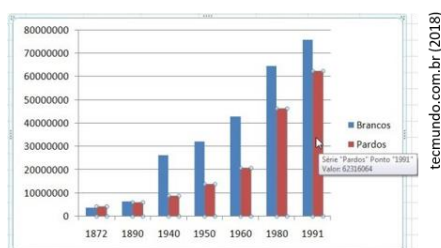
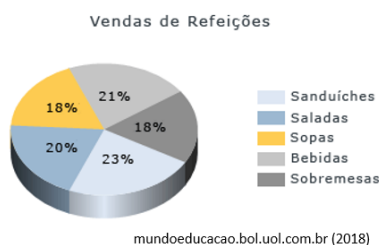
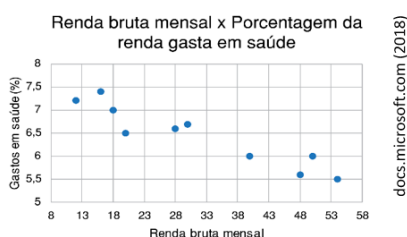
## 1.1 - Definições iniciais da estatística (não precisa decorá-las, apenas compreendê-las)

**Estatística:** 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.<sup>1</sup>

**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.

### Exemplos de tipos de gráfico



**População:** qualquer seleção de indivíduos ou valores, finita ou infinita. É muito comum encontrarmos os termos “universo” e “universo estatístico” como sinônimo de **população**. 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**.

**Amostra:** uma parte da população, normalmente selecionada com objetivo de se fazer inferências\* sobre a população. Por que ela é importante? Porque em muitos casos ficaria muito oneroso conhecer os dados da população inteira. Para saber a idade média do goianiense, uma alternativa mais viável é selecionarmos uma amostra de moradores de Goiânia e, a partir dessa amostra, estimar a idade média da população dessa

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



cidade. E é exatamente isso que acontece em pesquisas eleitorais, não é mesmo? Nesse contexto, já é muito importante o conceito de inferência estatística que veremos a seguir.

**Inferência estatística:** 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.

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.



### 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 escurificada (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**

Ainda na estatística descritiva, utilizamos dois tipos de medidas para sumarizar os dados. São elas:

**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.

## 1.2 - 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  $\mu$  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:

$$\bar{x} = \frac{p_1x_1 + p_2x_2 + \dots + p_kx_k}{p_1 + p_2 + \dots + p_k} = \frac{1}{p} \sum_{i=1}^n p_i x_i$$

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):

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

**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.

## 1.2 - 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:

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

Vamos considerar que se pesou 7 caroços de feijão, retirados de um saco de feijão, obtendo **valor máximo** (0,2673 g) e **valor mínimo** (0,1188 g). Calculando a diferença entre esses dois extremos, 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).

$$A = x_{m\acute{a}x} - x_{m\acute{i}n}$$

Voltando à análise do erro: 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 0,0 caso façamos o somatório dos erros** de uma série de medidas. Por isso, 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. Uma alternativa é realizarmos o cálculo do **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) do quadrado do resíduo 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 ( $\sigma^2$ ):**

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

**CUIDADO:** Note que o símbolo da variância e o denominador ( $n$  ou  $n-1$ ) do lado direito são diferentes quando se trata de dados amostrais ou populacionais.

Um pequeno inconveniente da variância é que sua unidade passa a ser o quadrado da unidade original. Por exemplo, se os valores são dimensões em metros, então a variância desses valores será expressa em metro<sup>2</sup> (m<sup>2</sup>). Uma alternativa para contornar inconveniente é calcular a raiz quadrada da variância e assim a unidade



voltará para dimensão original da variável medida. Exemplo: 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 ( $\sigma$ ):**

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

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.

## 1.2 – Erros em análises laboratoriais (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 para o fim aqui pretendido. 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. 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 sempre 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)<sup>2</sup>, 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;

---

<sup>2</sup> 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.

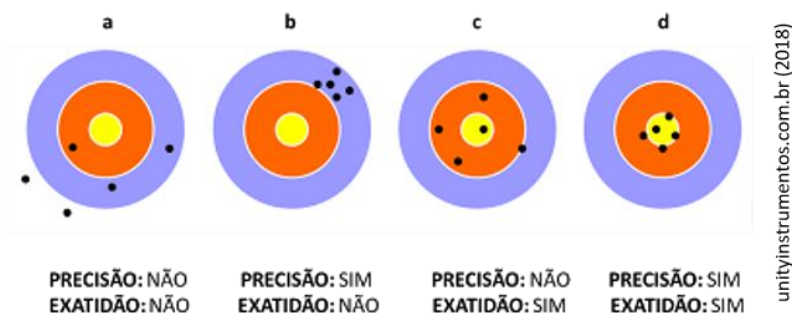
### 1.3 – Diferença entre precisão e exatidão

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 (ou melhor) 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 (ou melhor) 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. Para esse tipo de material, pode aparecer também o termo **amostra de referência**.

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.





(Adaptado de UEPI - Perito Criminal/Química - PCPI) 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**

(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**

(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 é 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**





## 2 – Estatística inferencial (testes de hipótese)

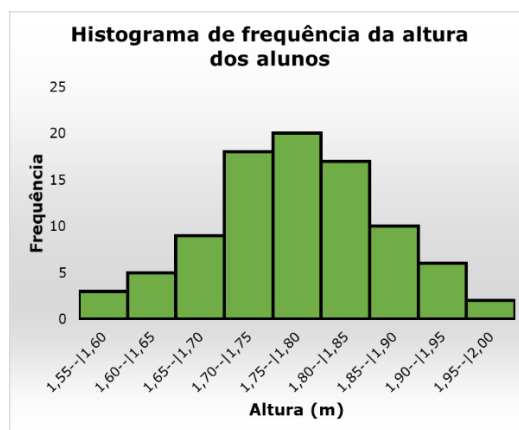
### 2.1 – Histograma de frequência

Chegou a hora de aprendermos a tomar decisões apropriadas baseada em probabilidades. Antes, porém, vamos explorar rapidamente um gráfico específico, histograma de frequência, que nos ajudará no entendimento da distribuição normal e dos testes de hipótese.

Correlacione a tabela abaixo (etapa prévia à construção do gráfico) e o histograma de frequência à direita. Os valores correspondem às alturas de alunos de uma determinada turma.

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
<b>Total</b>	<b>90</b>	<b>1,00</b>	

(\*) 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. No gráfico, 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.

### 2.2 – Distribuição normal

*O que seriam as distribuições estatísticas ou distribuições de frequência?*

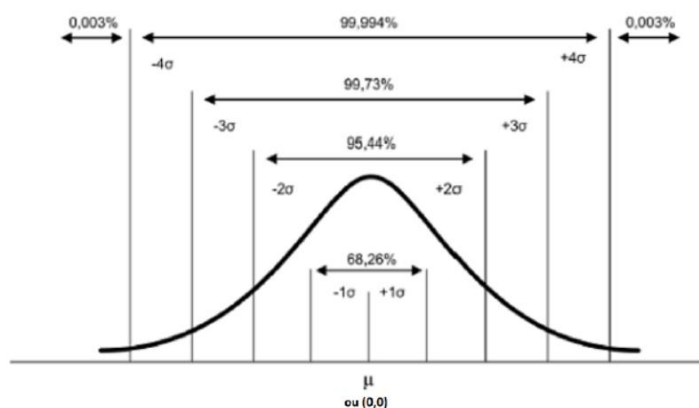
É um modelo (equação matemática) descritivo sobre a chance que uma variável aleatória pode apresentar ao longo de um intervalo de valores, análogo ao comportamento observado em histogramas de frequência. Já sabemos, por exemplo, que a chance de encontrarmos valores próximos à média é maior. A distribuição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

**Distribuição normal ou gaussiana:** a mais utilizada. 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. 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. É descrita pela equação abaixo (não precisa decorá-la) e ilustrada pelo gráfico abaixo.

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



portalaction.com.br (2018)

Distribuição de uma variável  $x$  em torno da média populacional  $\mu$  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. Você se lembra? Por isso, é plausível utilizarmos a **distribuição normal** para prever probabilidades de ocorrências para as alturas para todos os alunos da turma, 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**.

**Padronizar uma variável aleatória:** é subtrair de cada valor a **média populacional** ( $\mu$ ) e dividir pelo **desvio padrão** ( $\sigma$ ), obtendo uma nova variável aleatória adimensional chamada **Z** ou **Z<sub>score</sub>**. O **Z<sub>score</sub>** 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<sub>score</sub>** é -2, então esse valor está abaixo da média numa distância de 2 desvios padrão ( $2\sigma$ ). O cálculo do **Z<sub>score</sub>** 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. Analise a discussão a seguir correlacionando com a figura logo acima. Se deslocarmos apenas 1 desvio padrão ( $\sigma$ ) para mais e para menos em torno da média, engloba-se 68,26% da população. Caso façamos o mesmo com  $2\sigma$ , teremos mais que 95%. Já com  $3\sigma$  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  $3\sigma$  é um **outlier** ou **amostra anômala**.

*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  $(-3\sigma$  e  $+3\sigma)$ :

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

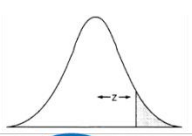
Isso significa também que a chance é de apenas 0,27% de selecionarmos um valor fora desse intervalo, ou seja, desse valor estar nas caudas da curva normal (distribuição normal). Se dividirmos essa possibilidade por 2, então será 0,135% para cada cauda (uma cauda para valores acima de  $+3\sigma$  e outra cauda para valores inferiores a  $-3\sigma$ ). 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. **Felizmente não precisamos utilizar integral para esses cálculos**, pois temos as tabelas que trazem essa informação para nós.

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

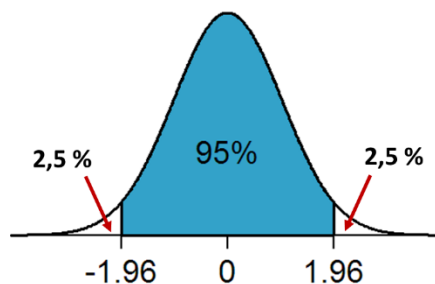
A Tabela contém o que se chama de área da cauda (à direita) da distribuição normal padrão para valores de  $Z$  que vão de 0,00 a 3,99. Acompanhe a explicação que se segue correlacionando com a tabela  $Z$  (trecho dela abaixo e completa no **final da apostila**). A primeira coluna, à esquerda, dá o 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  $Z$ , devemos 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** (para obter em porcentagem, basta multiplicar por 100:  $0,025 \times 100 = 2,5\%$ ), é a fração da área total sob a curva que está localizada à direita de  $Z = 1,96$ , vide figura ao lado. 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 ao lado. 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  $Z = 1,96$ , conforme ilustrado na imagem abaixo. 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.

Tabela da área de distribuição normal padronizada



$z$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068



Poderíamos estar interessados em utilizar a tabela no caminho inverso. Poderia ser perguntado, por exemplo: qual o intervalo de Z-score que compreende 95% dos dados? Se estamos interessados em compreender 95%, a área não compreendida será de 5%. Como essa área se divide em duas caudas, cada uma terá 2,5%, o que, em fração, corresponde a 0,025. Agora, basta buscarmos 0,025 nos valores do meio da tabela e identificarmos (avaliando valores correspondentes da primeira linha e da primeira coluna) o Z-score correspondente, 1,96. Portanto, o intervalo entre -1,96 e +1,96 compreende 95% dos dados da população normal.

Os exemplos acima foram simétricos 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  $[-0,62; 1,31]$ .

## 2.4 – Testes de hipótese: teste Z-score, t-student e teste F

Com o embasamento das seções anteriores, finalmente iremos aprender a tomar decisões e fazer inferências apropriadas com grandes chances de estarem corretas, a partir de dados experimentais. De antemão explico que os **testes de hipótese** (também conhecidos como **teste de significância**), em geral, servem para comparar valores (médias e variâncias) que, ao ser aplicado às análises laboratoriais, estão relacionados à exatidão e à precisão dos métodos. Em suma, são ferramentas que podem neste contexto ajudar a definir se um método é mais exato e/ou mais preciso que o outro.

### Qual método aplicar em cada situação?

- i. **Para comparação de valores médios (médias):** utiliza-se **teste Z-score** ou **t-student**. Quando for conhecido o comportamento populacional, ou seja, quando for conhecido  $\mu$  e  $\sigma$  (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.

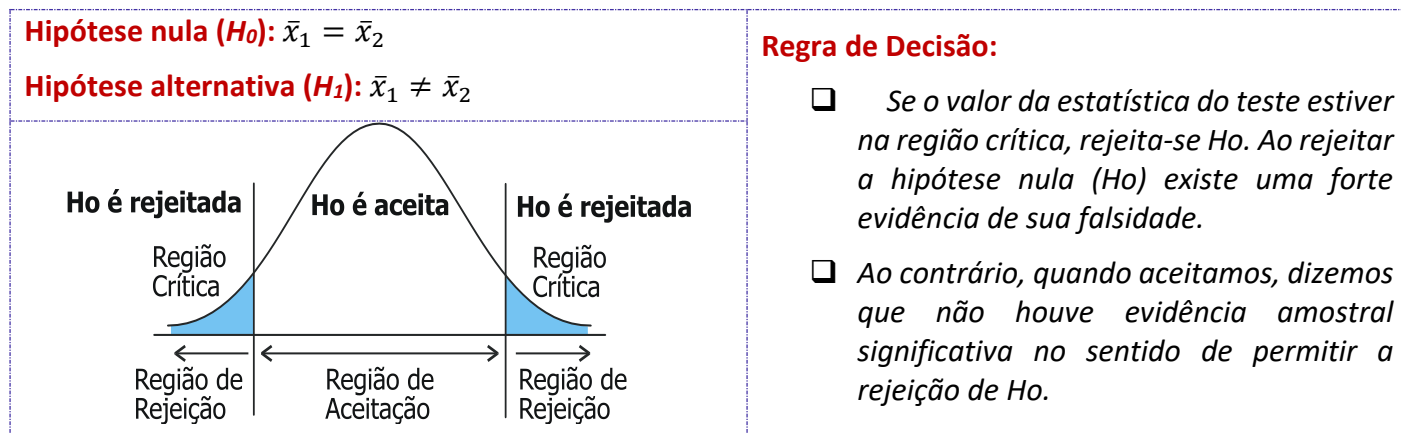
- ii. **Para comparação de variância ou de precisão de métodos:** utiliza-se o **teste F** que utiliza a **distribuição F de Fisher-Snedecor**. O comportamento da função densidade de probabilidade para essa distribuição é diferente de um sino.

### Qual a lógica por de trás dos testes de hipótese?

Em geral, o valor calculado do teste irá medir uma diferença entre dois valores comparados. Se essa diferença for grande é porque há diferença estatística entre eles, caso contrário, não há. Para definir o que é grande ou pequeno, precisamos de um valor limiar que é justamente o **valor crítico** (daqui a pouco falaremos sobre como encontrá-lo) também obtido pelo teste. Ou seja, **valor calculado** superior ao valor crítico indicará uma



diferença estatisticamente significativa entre os valores comparados. Se for menor, não haverá diferença estatística entre eles. Para tomada de decisão, são definidas duas **hipóteses** (**nula**,  $H_0$ , e **alternativa**,  $H_1$ ) conforme mostrado abaixo. No gráfico, estão identificadas as **regiões de rejeição** de  $H_0$ , nas extremidades, e a de sua **aceitação** mais ao centro. Por fim, conclui-se sobre o resultado por meio da **regra de decisão**, também mostrado abaixo.



### Como identificar se um teste de hipótese é unicaudal ou bicaudal?

Você precisa entender quando estará diante de um teste de hipótese unilateral ou bilateral, pois isso mudará o valor do  $Z_{crítico}$ . Essa identificação é relativamente simples. Se o nosso objetivo é comparar dois valores (saber se eles são diferentes ou não), então nosso teste será **bicaudal** porque a diferença pode ser à esquerda (valor inferior) ou à direita (valor superior). Por outro lado, estaremos diante de um **teste unicaudal** se o objetivo for verificar se um valor é superior a outro valor (**unicaudal à direita**) ou se é inferior (**unicaudal à esquerda**). Tranquilo, não é mesmo? O número de caudas de um teste influencia naquele valor que procuramos na tabela (ex: tabela Z-score). Se estivermos interessados em uma região crítica de 5% (ou seja 0,05) e o teste for bicaudal, teremos que dividir 0,05 por 2 (= 0,025) para identificar o valor de Z na tabela. Se o teste for unicaudal, não realizamos essa divisão e o identificaremos o valor de Z correspondente a 0,05 de área de cauda.

### Aplicando o teste Z

(comparação de médias quando se conhece dados populacionais)

Situação hipotética: Um consumidor pagou 1,90 em pedaço de bolo. Sendo conhecidos o preço médio populacional ( $\mu$ ), 1,30 reais, e o desvio padrão ( $\sigma$ ), 0,69 reais, relacionados ao preço do bolo historicamente praticado por essa lanchonete. Pergunta-se: ao nível de 95% de confiança, o consumidor estaria pagando um valor diferente pelo bolo?

#### Comentários:

Como são conhecidos média ( $\mu$ ) e desvio padrão ( $\sigma$ ) populacionais, então **devemos utilizar o teste Z**. Como estamos interessados em definir se há diferença, então o **teste é bicaudal**. Hipóteses a serem testadas:

$H_0$ : 1,90 (valor pago) = 1,30 ( $\mu$ ), equivalência estatística

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





Descobrimos o valor crítico: na tabela! Relembrando: no nosso caso, 95% de confiança ou  $\alpha$  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,  $1 - 0,95 = 0,05$  ( $=\alpha$ ). Como o teste é bicaudal, devemos dividir 0,05 por 2, 0,025. Pronto! Vamos à **Tabela-Z** (final da apostila) localizar o  $Z_{score}$  correspondente a essa área de 0,025....  $Z_{score}$  (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 (que eu sugiro decorar).

Fórmula a ser decorada	Aplicando os valores do problema
$Z_{calculado} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$	$Z_{calculado} = \frac{1,90 - 1,30}{\frac{0,69}{\sqrt{1}}} = 0,87$
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.

Posso estar sendo redundante, mas, para uma compreensão mais intuitiva deste e dos demais testes estatísticos, observe que, no numerador da fórmula acima, temos uma diferença ( $\bar{x} - \mu$ ) entre um valor testado (a média) e um valor populacional. Se ela for grande (superior ao valor crítico), é porque há evidências da diferença entre os valores comparados. Se for menor, não há evidências dessa diferença.



Considere **média e desvio padrão históricos** (dados coletados no decorrer de um longo período) como sendo dados populacionais. Por exemplo: peso médio de um produto e desvio padrão obtidos a partir de várias amostras ao longo de um ano (valores que podem ser considerados como média e desvio padrão populacionais)

**Amostras de referência** com **valores** de média e desvio-padrão **certificados** também devem ser considerados dados populacionais. Por exemplo: uma amostra de solo estabilizada com teores de determinados minerais certificados.

## Aplicando o teste t-Student

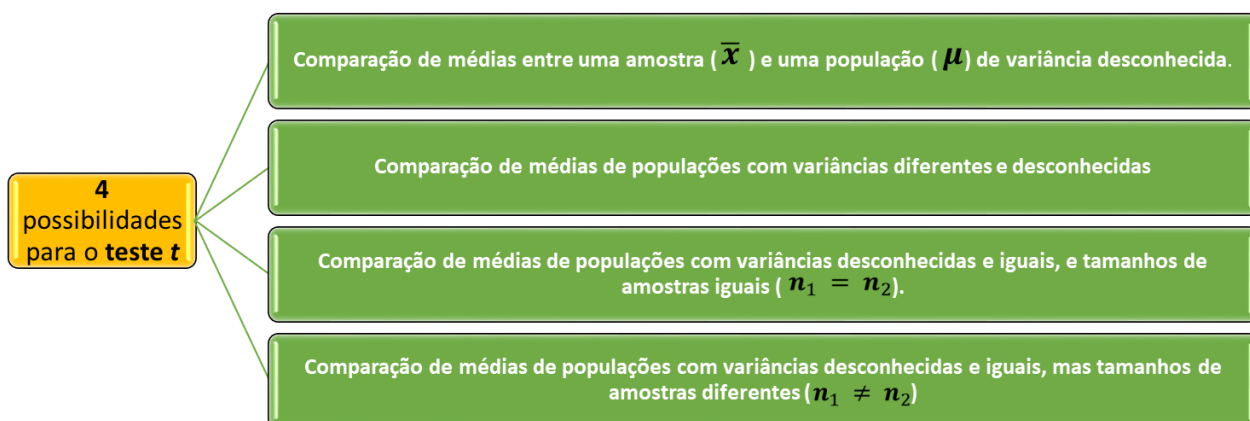
(comparação de média quando **não** se conhece dados populacionais)

Situação hipotética: Um laboratório está avaliando a substituição de um método 1, de determinação de amilose em arroz, por um método 2. 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, a equipe do laboratório está interessada em saber: o resultado médio obtido pelos dois métodos é estatisticamente equivalente ao nível de 95% de confiança ou alpha ( $\alpha$ ) 0,05?

### Comentários

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:



Cada possibilidade do teste t apresenta fórmulas diferentes, resalto 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).

No caso hipotético analisado, 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 respectivas fórmulas, 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** (final da apostila) 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 ( $\alpha$ ) 0,05. Como nosso teste é bilateral, devemos dividir  $\alpha$  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 é que não há evidências estatísticas que as médias, 24,5 g.kg<sup>-1</sup> e 25,5 g.kg<sup>-1</sup>, 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.

### Aplicando o teste F (comparação de variâncias ou de desvios-padrão)

Situação hipotética anterior, mas com pergunta diferente: Um laboratório está avaliando a substituição de um método 1, de determinação de amilose em arroz, por um método 2. Para tanto, analisou-se uma mesma amostra de arroz, 5 vezes, pelos dois métodos, obtendo os seguintes resultados para teor de amilose: 24,5±1,5 g.kg<sup>-1</sup> (método 1); 25,5±2,0 g.kg<sup>-1</sup> (método 2). Com base nesses resultados, a equipe do laboratório está interessada em saber: **a variância apresentada entre os dois métodos é estatisticamente equivalente ao nível de 95% de confiança ou alpha ( $\alpha$ ) 0,05?**

#### Comentários

A pergunta a ser respondida é sinônima a "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). Hipóteses a serem testadas:

$$H_0: s_1^2 = s_2^2$$

$$H_1: s_1^2 \neq s_2^2$$

Descobrimo o valor crítico: na tabela! 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<sup>2</sup> e 1,5<sup>2</sup> na equação abaixo, 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 a equipe do laboratório:** não há evidências estatísticas que as **variâncias** ou a **precisão** dos dois métodos (1 e 2) sejam diferentes.



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

## 2.5 – Intervalos de confiança para 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 ( $\mu$ ) estar contida. Aqui vou dispensar a discussão conceitual e de interpretação desses intervalos, pois o que você realmente precisa saber é como calcular os intervalos de confiança. Há duas possibilidades: na primeira, o desvio padrão populacional ( $\sigma$ ) 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  $\mu$ , 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  $\mu$ , a partir de uma amostra (conjunto de indivíduos), utilizando a distribuição t-Student:

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



(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:  $470 \pm 15 \mu\text{g kg}^{-1}$

Analisando 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

Analisando 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 **teste 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:  $91,0 \text{ mg.dL}^{-1}$ ;  $95,0 \text{ mg.dL}^{-1}$ ;  $96,0 \text{ mg.dL}^{-1}$  e  $98,0 \text{ mg.dL}^{-1}$ , com desvio padrão  $2,9 \text{ mg.dL}^{-1}$ . 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  $\text{mg.dL}^{-1}$ , 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:  $\bar{x}=95,0 \text{ mg.dL}^{-1}$  (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<sup>-1</sup>.

**Resposta: letra D**



## QUESTÕES COMENTADAS

Texto referente às questões 1 e 2:

	concentração em µg/g		desvio-padrão
elemento	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.

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

### 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 triplicada ( $n=3$ ). 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**

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.

### 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 ( $\mu$ ) 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	$t_{calculado}$	$t_{critico}$
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  $|t_{calculado}| > t_{critico}$ , 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**

3. (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**

**4. (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**

**5. 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 $\mu\text{g.g}^{-1}$ )	10,5	10,4	10,5	10,5	10,6

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

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

### 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:



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

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

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

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

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

Portanto, a alternativa é verdadeira.

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

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**

**6. (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**

7. (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. Concluimos, portanto, que estão presentes tanto erro aleatório quanto erro sistemático no experimento.

**Resposta: letra D**

8. (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:

9. (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:

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



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

$$n=4$$

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

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

$$\begin{aligned}\bar{x} - t_{n-1} \frac{s}{\sqrt{n}} &< \mu < \bar{x} + t_{n-1} \frac{s}{\sqrt{n}} \\ 0,080 - 3,18 \frac{0,002}{\sqrt{4}} &< \mu < 0,080 + 3,18 \frac{0,002}{\sqrt{4}} \\ 0,077 &< \mu < 0,083\end{aligned}$$

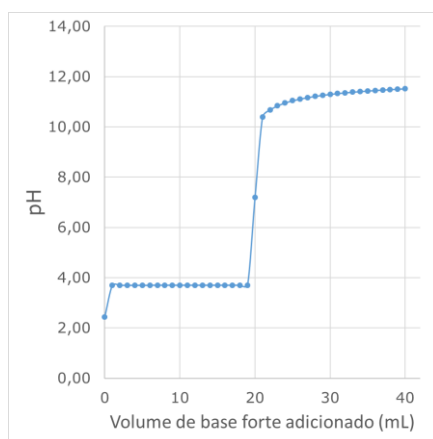
A probabilidade de a média estar contida no intervalo de 0,077 a 0,083 é de 95%. Portanto, podemos concluir que a probabilidade associada ao intervalo apresentado no enunciado (de 0,078 a 0,082) é menor que 95%, já que ele é mais estreito do que o calculado.

**Resposta: errado**

**10. (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**

11. (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:

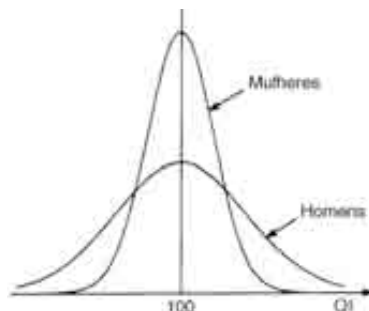
$$\bar{x} = \frac{26 + 31}{2} = 28,5$$

**Resposta: letra B**





12. (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**

13. (CESPE - Perito Criminal/Química - SGA/AC - 2008) O intervalo de confiança  $\mu$  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**

14. (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**

**15. (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 ( $\mu$ ) 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**

**16. (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**

17. (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**



## GABARITO

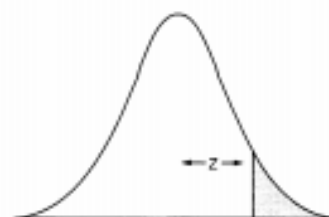
GABARITO



1	Certo	10	Errado
2	Certo	11	B
3	E	12	Errado
4	B	13	Errado
5	A	14	C
6	E	15	B
7	D	16	A
8	Certo	17	D
9	Errado		



**Tabela A.1** Área da cauda da distribuição normal padronizada.

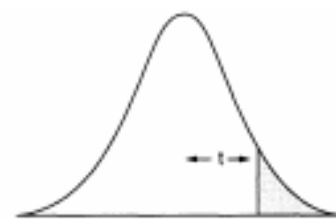


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
3,5	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
3,6	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
3,9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Fonte: Box, G. E. P.; Hunter, W. G., e Hunter, J. S., *Statistics for experimenters. An introduction to design, data analysis and model building*. Nova York, Wiley, 1978. Reproduzida com autorização da John Wiley & Sons, Inc.



**Tabela A.2** Pontos de probabilidade da distribuição  $t$  com  $v$  graus de liberdade.



Área de probabilidade										
$v$	0,4	0,25	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	0,325	1,000	3,078	6,314	12,706	31,821	63,657	127,32	318,31	636,62
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925	14,089	22,326	31,598
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,213	12,924
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,263	0,711	1,415	1,895	2,365	2,998	3,449	4,029	4,785	5,408
8	0,262	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,261	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,260	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,259	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,259	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,258	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,258	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,258	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,257	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,257	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,257	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,257	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,257	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,256	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,256	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,767
24	0,256	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,256	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,256	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,256	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690
28	0,256	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,256	0,683	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659
30	0,256	0,683	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
40	0,255	0,681	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
60	0,254	0,679	1,296	1,671	2,000	2,390	2,660	2,915	3,232	3,460
120	0,254	0,677	1,289	1,658	1,980	2,358	2,617	2,860	3,160	3,373
$\infty$	0,253	0,674	1,282	1,645	1,960	2,326	2,576	2,807	3,090	3,291

Fonte: Box, G. E. P.; Hunter, W. G., e Hunter, J. S., *Statistics for experimenters. An introduction to design, data analysis and model building*. Nova York, Wiley, 1978. Pearson, E. S. e Hartley, H. O. (eds.), *Biometrika tables for statisticians*, vol. I, Cambridge University Press (1958). Fisher, R. A. e Yates, F., *Statistical tables for Biological, Agricultural and Medical Research*, Longman Group Ltd. (1974) - Tabela III. Reproduzida com autorização da John Wiley & Sons, Inc; Biometrika Trustees e Longman Group UK Ltd.





**Tabela A.4** Pontos de percentagem da distribuição F, 5% (continuação).

$v_1 \backslash v_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	$\infty$
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
$\infty$	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00



# 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.