

Usando o Software R na Química Analítica

Genilson Pereira Santana

Ebook Clube da Química

Genilson Pereira Santana

Usando o Software R na Química Analítica

*Manaus
Edição do Autor
2020*

Todos os direitos reservados

Nenhuma parte desta publicação poderá ser reproduzida por qualquer meio ou forma sem a prévia autorização do editor. A violação dos direitos autorais é crime estabelecido na lei no 9.610/98 e punido pelo artigo 184 do Código Penal

Usando o Software R na Química Analítica

Copyright 2020 clubedaquimica.com
Todos os direitos reservados

Nenhuma parte dessa publicação será produzida, guardada pelos sistema "retrieval" ou transmitida de qualquer modo ou por qualquer outro meio seja este eletrônico, mecânico, de fotocópia, ou outros, sem prévia autorização por escrito da editora.

Arte de Capa: Genilson Pereira Santana
Revisão: Franklin Monteiro Brasil
Diagramação: Genilson Pereira Santana

Santana, Genilson Pereira

Usando o Software R na Química Analítica / Genilson Pereira Santana
Manaus, 2020

Formato: Livro Digital
Veiculação: Digital
ISBN: 978-65-00-02726-6



O Autor



Professor Titular da Universidade Federal do Amazonas (UFAM) tem Bacharelado em Química pela Universidade Federal de Viçosa, mestre em Química Analítica e doutor em Físico-Química pela Universidade Federal de Minas Gerais. Atualmente sou Diretor Departamento de Inovação Tecnológica/UFAM Criador e Editor Chefe - Scientia Amazonia, Criador da Revista e editor Chefe - Divirta-se com o Clube da Química, mantém o site clubedaquimica.com, Blog do Clube da Química e Escola de Química. Em termos de pesquisa coordena do Grupo de Pesquisa - GQA.

Sumário

Sumário	5
1 Introdução	7
2 Introdução ao R	9
2.1 Informações gerais	9
2.2 Instalação de R	10
2.3 Iniciando o R	10
2.4 Comandos Básicos	10
2.4.1 Entrada de Dados Externos	10
2.4.2 Alguns Parâmetros importantes do Gráfico	11
2.4.3 Gráficos estatísticos	12
3 O uso da estatística na Química Analítica	13
4 Média	27
5 Desvio Padrão	31
6 Mediana	37
7 Propagação de erros	41
7.1 Adição e subtração	41
7.2 Multiplicação e divisão	42
8 Arredondamento de dados	47
9 Medidas da Precisão	51
9.1 Coeficiente de Variação	51
10 Amostra	55
10.1 Amostragem	55
11 Teste de significância	63
11.1 Teste de F	63
12 Teste de t	69
12.1 Teste de <i>t</i> para múltiplas amostras	72
13 Rejeição de resultados	79
14 Boxplot	83
15 Calibração	87
15.1 Comparação com padrões	87
15.1.1 Comparação direta	88

15.1.2	Titulações	88
15.1.3	Calibração com padrões externos	88
15.1.4	Método dos Mínimos Quadrados ou Quadrados Mínimos Ordinários	89
15.2	Curva de calibração no R	91
15.2.1	Usos da curva de calibração	98
16	Tabelas	113
	Referências	115

1 Introdução

O livro tem como objetivo ensinar as ferramentas de estatística usadas no cotidiano da química com o software Open Source R. A ideia deste ebook é mostrar o R para o aluno de Química como uma ferramenta estatística com alto potencial para o tratamento de dados normalmente obtidos no laboratório. Pelo fato de ser uma ferramenta aberta e com muitas possibilidades a ideia é mostrar ao aluno de química uma literatura de Química Analítica diferente das ferramentas apresentadas nos livros didáticos. Geralmente, os livros didáticos de Química Analítica sugerem como pacote para o tratamento estatístico as planilhas eletrônicas, principalmente o Excel. A problemática do uso de planilhas eletrônicas está na obtenção de licenças, que nem sempre o aluno de química tem poder aquisitivo para adquiri-las legalmente. Outro problema no uso de planilhas eletrônicas é o questionamento dos resultados. Por causa disso, neste material fizemos a opção pelo software R, um programa de computador com muitas aplicações na área de Química.

Além disso, ao elaborar esse e-book procuramos colocar o uso da Ferramenta R de uma forma que nós químicos possamos usá-la rapidamente. Para isso, procurei usar exemplos contidos nos livros de Química Analítica. A ideia do livro é que o Químico aprenda a gostar do R como eu passei a admirá-lo e resolver meus tratamentos de dados. Confesso que gostaria de colocar também a análise exploratória de PCA e HCA, espero que numa segunda edição isso seja possível.

2 Introdução ao R

2.1 Informações gerais

O R começou a ser desenvolvido por Robert Gentleman e Ross Ihaka do Departamento de Estatística da Universidade de Auckland em Nova Zelândia, mais conhecidos por “R & R”, apelido do qual originou-se o nome R do programa. O objetivo inicial de “R & R”, em 1991, era produzir um software para as suas aulas de laboratório baseado na já revolucionária linguagem S, utilizada pelo software comercial S-Plus criado por Jonh M. Chambers da AT&T que atualmente vem contribuindo para o aperfeiçoamento e ampliação das análises estatísticas do R.

O primeiro relato da distribuição do R foi em 1993, quando algumas cópias foram disponibilizadas no StatLib, um sistema de distribuição de softwares estatísticos. Com o incentivo de um dos primeiros usuários deste programa, Martin Mächler do ETH Zürich (Instituto Federal de Tecnologia Zuriqne da Suíça), “R & R”, em 1995, lançaram o código fonte do R, disponível por ftp (uma forma de se transferir dados pela internet), sobre os termos de Free Software Foundations GNU general license, que seria um tipo de “licença para softwares livres“. O R é distribuído livremente no site <http://www.r-project.org>.

Em 1997 foi formado um grupo de profissionais que têm acesso ao código fonte do R, possibilitando assim a atualização mais rápida do software. Desde então o R vem ganhando cada vez mais adeptos em todo o mundo, em parte devido ao fato de ser totalmente gratuito e também por ser um programa que exige do usuário o conhecimento das análises que está fazendo, diminuindo assim as chances de uma interpretação errada dos resultados.

Outro fato importante para a difusão do R é a sua compatibilidade com quase todos os sistemas operacionais. O R está disponível para a maior parte dos MacOS, Windows a partir do 95 e para UNIX e sistemas similares como Linux e FreeBSD.

2.2 Instalação de R

O R pode ser instalado via o CRAN (*Comprehensive R Archive Network*) ou a distribuição pode ser obtida para os sistemas operacionais Windows, Linux e Mac. No caso do Ubuntu é possível instalar diretamente do seu sistema de pacotes existente em sua biblioteca.

2.3 Iniciando o R

Para iniciar o R no windows basta clicar sobre o ícone ou na parte iniciar programas, no caso do Linux basta digitar R em um terminal ou console.

2.4 Comandos Básicos

Os comandos são realizados com auxílio de operadores e de funções da seguinte forma: nome da função(argumento 1, argumento 2,...,argumento N). Tenha sempre o cuidado de não utilizar acentuação nem espaços no nome da função.

2.4.1 Entrada de Dados Externos

Os dados contidos em planilha eletrônica irá convertê-la para ponto ou ii) alterar a configuração do editor de planilha para que aceite o ponto como separador decimal.

De acordo com a Tabela 1, nota-se que os dados estão com o ponto como separador decimal. **Uma forma fácil de saber se o valor digitado foi reconhecido como numérico ou não, é observar o seu posicionamento na coluna. Os caracteres são alinhados à esquerda e os numéricos à direita.** Deve-se, ainda, evitar fazer formatações na planilha de dados e, todas as fórmulas executadas, devem ser convertidas a valores. Como o Programa R irá reconhecer células em branco que já tenham sido trabalhadas como sendo valor perdido, para evitar possíveis

problemas é aconselhável copiar os valores de entrada para uma nova planilha para depois convertê-la para o formato .csv.

Tabela 1 – Exemplo de dados externos

	A	B	C
1	x	y	
2	A	12.3	
3	A	11.8	
4	A	13.6	
5	A	13	
6	A	12.5	
7	B	11.4	
8	B	10.9	
9	B	12.1	
10	B	11.2	
11	B	10.5	
12			

Para salvar a planilha de dados no formato .csv, deve-se clicar em "Arquivo", e em "Salvar como ...", e selecionar a pasta onde o arquivo será salvo. Em seguida, deve-se escolher o nome do arquivo que e o formato a ser salvo. No Excel, será "CSV (separado por vírgula) (*.csv)" e clicar em "Salvar", aparecerão duas mensagens. Clicar "OK" na primeira e em "Sim" na segunda. No Calc, será "Texto CSV (.csv)" e após clicar em "Salvar", serão exibidas três mensagens. Clicar em "Sim", alterar a vírgula para ponto e virgula no item "Delimitador de campos" e clicar em "OK". A terceira será um aviso, em que se deve clicar em "OK".

2.4.2 Alguns Parâmetros importantes do Gráfico

- main="título" Adiciona um título ao gráfico
- xlab,ylab Adiciona títulos aos eixos respectivamente
- pch Coloca um símbolo correspondente a um número no gráfico
em geral utiliza-se pch=3
- col Altera a cor dos símbolos

2.4.3 Gráficos estatísticos

<code>hist(x)</code>	Constroi um histograma de frequência x	
<code>plot(density(x))</code>	Constroi uma função de densidade x	_____
<code>boxplot(x)</code>	Constroi um boxplot de x	
<code>qqnorm(x)</code>	Constroi um gráfico QQ de x	_____

3 O uso da estatística na Química Analítica

A química avançou consideravelmente suas metodologias de análises químicas. Até bem pouco tempo atrás a química quantitativa era feita basicamente por titulações, gravimetrias e fotometrias. A quantidade de dados obtidos com essas técnicas eram relativamente pequenas, consequentemente a forma de tratar os dados estatisticamente se resumia a comparar médias e variâncias. A evolução da informática e dos materiais levaram as análises químicas a produzir grandes quantidades de dados. Os equipamentos modernos são capazes de produzir milhares e milhares de números (Figura 1). Por exemplo, em volumes menores que 1 mL é possível determinar com o ICP-OES a tabela periódica toda, e se for ICP-MS os respectivos isótopos também.

Bruce R. Kowalski da Universidade de Washington, introduziu na química um novo termo, chamado de Quimiometria. Mais tarde foi criada a Sociedade Internacional de Quimiometria tendo como principal objetivo estabelecer metodologias para o tratamento de dados na área de Química. A



Figura 1 – Representação da capacidade de um laboratório moderno.

Quimiometria é, portanto, uma disciplina que usa a matemática e métodos estatísticos para a obtenção de informações relevantes de diversas medidas realizadas na área de química.

A quimiometria é a parte da química que utiliza métodos matemáticos e estatísticos para:

1. Definir ou selecionar as condições ótimas de medidas e experiências e
2. Permitir a obtenção do máximo de informações a partir da análise de dados químicos.

A obtenção de dados confiáveis requer a execução de um procedimento experimental bem definido, cujos detalhes são determinados pela finalidade do experimento. Digamos, por exemplo, que o problema seja descobrir a concentração de ácido acético numa amostra de vinagre. Isso geralmente é feito utilizando a volumetria de neutralização tendo como reagentes o hidróxido de sódio e a fenolftaleína. Inicialmente, é necessário o uso de um padrão primário (bifitalato de potássio) para padronizar a solução de hidróxido de sódio. Ressalta-se que a confiabilidade da análise está diretamente relacionada à padronização do hidróxido de sódio. Outras variáveis importantes na análise do ácido acético são as etapas necessárias para a determinação da concentração, como pesagens, diluições, titulação e leituras das medidas.

Quais são as maneiras de você detectar um erro determinado?

As maneiras de detectar um erro determinado são:

1. Analisar a amostra conhecida conforme o Material Padrão de Referência;
2. Analise a amostra em "branco" que não contém o analito. Se for observado um resultado diferente de zero, o método responde a mais do que o pretendido por você.
3. Utilize diferentes métodos analíticos para o mesmo analito. Se os resultados não concordarem, existe erro em um (ou mais) dos seus métodos;

4. Arredondamento a partir de uma série de experimentos. Analise amostras idênticas, em diferentes laboratórios, manipuladas por diferentes métodos. As discordâncias além do erro aleatório esperado são um erro sistemático.

Naturalmente, os resultados obtidos nunca serão todos iguais, mas logicamente se o vinagre foi fabricado de acordo com a legislação vigente (4,00% de ácido acético) podemos considerar o produto como de qualidade. Suponhamos que, ao titular duas amostras de procedências diferentes, um químico tenha obtido 3,80 e 4,20% de ácido acético, respectivamente. Isso quer dizer que ele deve aceitar a segunda amostra e rejeitar a primeira por conter menos ácido que o mínimo determinado por lei?

É impossível responder esse questionamento sem ter uma estimativa da incerteza associada a esses resultados. Por exemplo, cada uma das operações de laboratório necessárias para a titulação está sujeita a erros, que influenciam diretamente no resultado final.

Uma medida pode ser afetada por diversos fatores, como, distração do químico durante a titulação, vidrarias com escalas que não condizem com o valor medida (descalibradas), entre outras. Os erros que acompanham uma medida podem ser classificados em duas categorias:

1. ***Erros determinados ou sistemáticos*** - Possuem um valor definido e, pelo menos em princípio, podem ser medidos (se detectados) e computados no resultado final.
2. ***Erros indeterminados*** - Não possuem valor definido, não são mensuráveis e flutuam de um modo aleatório.

Os erros determinados, por exemplo quando se trabalha com um indicador inadequado, não observância da solubilidade dos precipitados, processo mal executado de co e pós-precipitação, denominado de *erro de método*. As manipulações feitas durante a análise química ocorrem erros, a saber, deixar o béquer destampado permitindo a introdução de poeira na solução; deixar um líquido contido em um frasco sob forte aquecimento; entre outras. Neste caso, o químico está cometendo o chamado *erro ope-*

racional. A incapacidade do químico em executar determinada operação, por problema de inaptidão, é conhecida como *erros pessoais*.

Tal erro aparece com a dificuldade, por exemplo, em observar corretamente a mudança de cor de indicadores. Agora, sem sombras de dúvidas, o erro pessoal mais grave é o químico pré-julgar ou mesmo ter preconceito em relação a análise química. As imperfeições dos instrumentos, aparelhos volumétricos e reagentes é outra fonte importante de erros determinados, sendo conhecida como *erros devidos a instrumentos e reagentes*. Classicamente, estão incluídas o uso de equipamentos não calibrados, as impurezas e uso fora do prazo de validade dos reagentes.

Evidentemente, nem as análises química e nem a estatística trata dos erros determinados. O químico deve durante sua trajetória como analista evitar a todo custo cometer esses erros. Conhecer e entender todos os passos de uma análise química é regra fundamental para o sucesso dos resultados no final do processo. Mesmo assim, a ausência de erros determinados não conduz a resultados que sejam numericamente iguais, sempre haverá variações. Isso é consequência direta dos chamados *erros indeterminados*, que não podem ser localizados e nem corrigidos.

É natural, por exemplo, que duas titulações de uma mesma amostra produzam resultados numéricos não idênticos. Como diferenciar os resultados obtidos do determinado para o indeterminado? A Figura 2 mostra que:

1. Os valores obtidos flutuam, mas também tendem a concentrar-se em torno de um certo valor intermediário.
2. A flutuação em torno desse valor ocorre aparentemente ao acaso. Se determinada titulação resultou num valor abaixo da média, por exemplo, isso não permite prever em que direção flutuará o valor da próxima titulação, nem de quanto será o seu desvio.
3. A aparência do gráfico sugere que a concentração da amostra está mesmo fora da especificação, já que a maioria dos valores determinados é inferior a 4%.

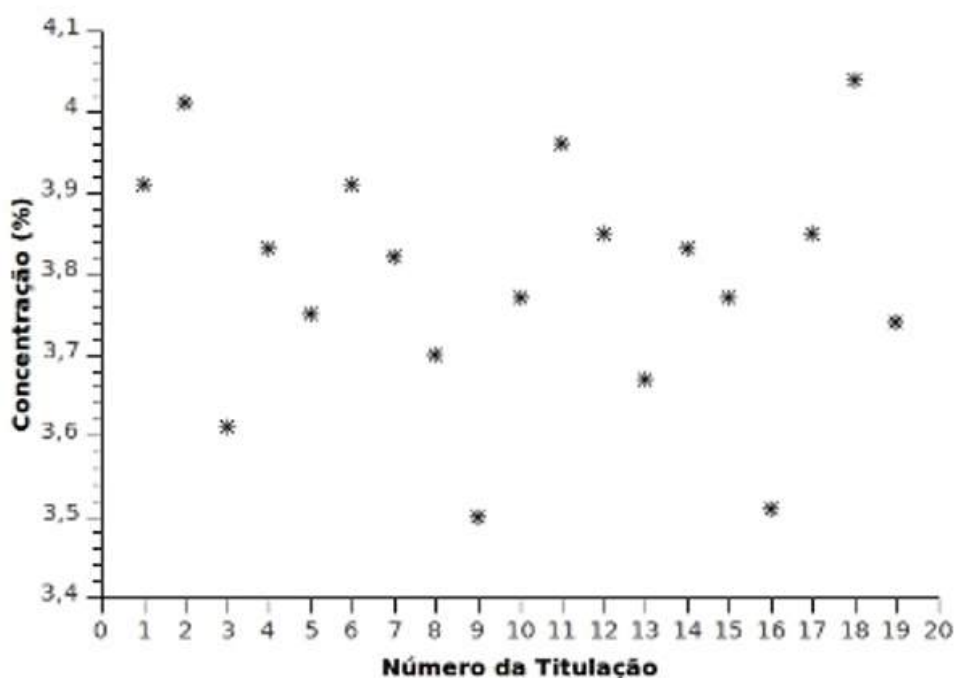


Figura 2 – Resultados de titulação feitas na mesma amostra de vinagre.

Essas situações são corriqueiras nas mais variadas determinações experimentais e indicam que, por mais que se tente controlar o processo, algumas fontes de erro sempre terminam permanecendo. Geralmente, os resultados oscilam para mais ou para menos, tendo como influência o acaso. Durante uma determinação um químico pode realizar infinitas medidas, o que lhe dá condições de ter um conjunto relativamente grande de valores.

Esses valores, na estatística são chamados de população que é qualquer coleção de indivíduos, finita ou infinita. Em cada medida realizada existem erros indeterminados¹ e, portanto, o resultado obtido é pura e simplesmente uma probabilidade de ser correto ou não.

Por causa do grande número de medida, uma parte da população é, normalmente, selecionada com o objetivo de se fazer inferências sobre a população. Essa parte da população denomina-se amostra cuja escolha é deve ser rigorosamente aleatória, portanto, não tendenciosa. Esta condição é muito importante na prática, porque as inferências estatísticas sempre supõem que as amostras são representativas da população. Por isso, ao realizar um experimento, devemos sempre tomar cuidado para coletar

¹ Espera-se que o analista não tenha cometido qualquer erro determinado

os dados de modo que a hipótese de aleatoriedade seja, não rigorosamente, pelo menos aproximadamente obedecida.

A Figura 3 mostra um gráfico construído com os resultados obtidos de uma série de resultados de uma análise química de soda caustica. Nota-se que a concentração variou de 0,10 a 0,30 mol L^{-1} distribuídas em intervalos segundo uma frequência de ocorrência de concentração. No centro do histograma tem-se a maior frequência o que significa geralmente que a concentração da soda caustica mais provável está entre 0,18 e 0,20 mol L^{-1} .

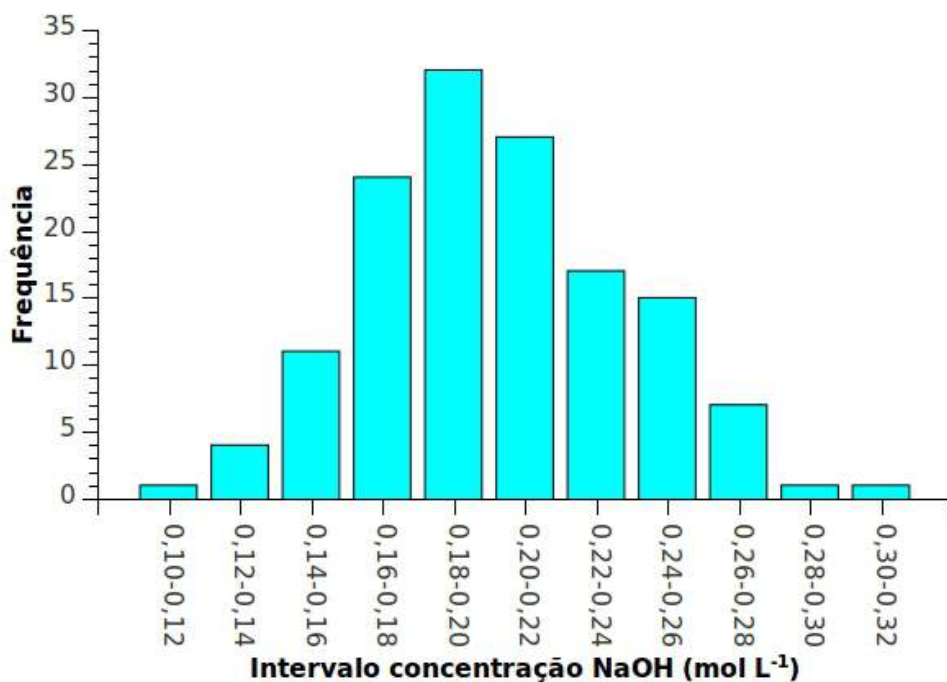


Figura 3 – Distribuição dos resultados de uma análise química de NaOH

Esse comportamento é mais comum quando se faz medições de concentração nas análises químicas. O gráfico construído mostra na realidade uma distribuição dos erros que é denominada gaussiana, sendo mais conhecida como distribuição normal. Quando o gráfico é construído com um número infinito de medida no centro do gráfico se localiza o valor verdadeiro, por exemplo, de uma determinada concentração. Em torno do valor verdadeiro estão os erros. Matematicamente a distribuição normal

pode descrita com a seguinte equação:

$$y = \frac{\exp\{-(x - \mu)^2/2\sigma^2\}}{\sigma\sqrt{2\pi}} \quad (3.1)$$

O gráfico da distribuição Normal é apresentado na Figura 4, em que pode ser verificada a simetria a partir μ . Interessante afirmar que os resultados das análises químicas são influenciados aleatoriamente que produzem valores normalmente distribuídos. Assim, resultados de dois ou mais séries de medidas representam curvas de normais, contendo cada uma μ e desvio padrão absoluto σ ; ou seja, existem famílias de distribuições normais. O ponto mais alto de cada curva de distribuição se concentra a média, mediana e a moda da distribuição. O desvio padrão determina a largura da curva; valores maiores resultam em curvas mais largas e mais planas (maior variabilidade nos dados).

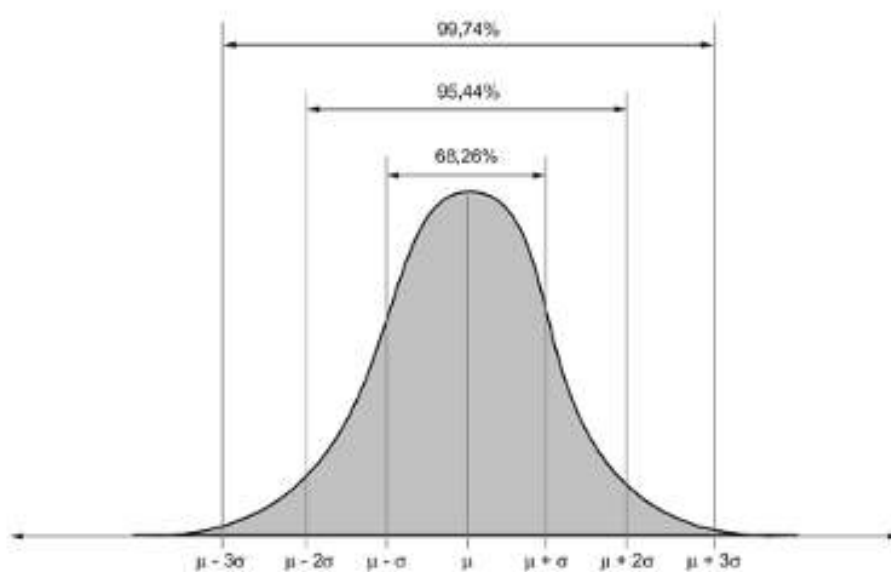


Figura 4 – Distribuição normal

Na prática quem trabalha com tratamento estatístico utiliza uma estimativa do valor mais provável de ser o verdadeiro, sendo conhecida como média aritmética (\bar{x}). A estimativa do erro, que indica a dispersão das medidas, também é expressa na forma de desvio padrão (dp).

Exemplo 3.1 *Um químico, usando a potenciometria com íons seletivos, fez 50 medida para a determinação de íons nitratos ($\mu\text{g L}^{-1}$), cujos resultados são apresentados na Tabela 2. Construa uma tabela de frequência*

com os resultados e estabeleça qual é o valor mais provável de íons nitratos presentes na amostra.

Tabela 2 – Resultados das determinações de íons nitratos ($\mu\text{g L}^{-1}$)

0,51	0,51	0,50	0,51	0,49	0,52	0,53	0,50	0,50	0,47
0,51	0,52	0,53	0,48	0,49	0,50	0,52	0,49	0,49	0,50
0,49	0,48	0,46	0,49	0,49	0,48	0,49	0,49	0,51	0,47
0,51	0,51	0,51	0,48	0,50	0,47	0,50	0,51	0,49	0,48
0,51	0,50	0,50	0,53	0,52	0,52	0,50	0,50	0,51	0,51

Ao observamos os valores da Tabela 2, nota-se que o menor valor é 0,46 e o maior 0,53, variando de 0,01 unidade. Ao contar os valores de concentração de nitrato nos intervalos 0,01 unidade obtém-se uma frequência dos dados, conforme mostrado na Tabela 3

Tabela 3 – Frequência para a medições obtidas de íons nitratos

Concentra de íons nitratos ($\mu\text{g L}^{-1}$)	Frequência
0,46	1
0,47	3
0,48	5
0,49	10
0,50	10
0,51	13
0,52	5
0,53	3

Exercício 3.1 Construa um histograma para os dados da Tabela 4. A literatura em geral recomenda que o número de barras seja aproximadamente igual à raiz quadrada do número total de observações. Como a Tabela tem 20 valores, seu histograma deve ter 4 ou 5 barras. Prefira cinco, que é um número ímpar, e lhe permitirá enxergar melhor possíveis simetrias.

Exemplo 3.2 Como forma de estabelecer uma distribuição dos comportamento das massas de caroços de feijão, faça um gráfico de frequência usando a Tabela 5 no R.

Tabela 4 – Resultados de 20 titulações feitas no mesmo lote de vinagre

Titulação	Concentração (%)	Titulação	Concentração (%)
1	3,91	11	3,96
2	4,01	12	3,85
3	3,61	13	3,67
4	3,83	14	3,83
5	3,75	15	3,77
6	3,91	16	3,51
7	3,82	17	3,85
8	3,70	18	4,04
9	3,50	19	3,74
10	3,77	20	3,97

Vamos inicialmente entra com os dados no R

```
> feijao <- c(0.1188, 0.2673, 0.1795, 0.2369, 0.1826, 0.1860, 0.2045,
0.1795, 0.1910, 0.1409, 0.1733, 0.2146, 0.1965, 0.2326, 0.2382, 0.2091,
0.2660, 0.2126, 0.2048, 0.2058, 0.1666, 0.2505, 0.1823, 0.1590, 0.1722,
0.1462, 0.1985, 0.1769, 0.1810, 0.2126, 0.1596, 0.2504, 0.2285, 0.3043,
0.1683, 0.2833, 0.2380, 0.1930, 0.1980, 0.1402, 0.2060, 0.2097, 0.2309,
0.2458, 0.1496, 0.1865, 0.2087, 0.2335, 0.2173, 0.1746, 0.1677, 0.2456,
0.1828, 0.1663, 0.1971, 0.2341, 0.2327, 0.2137, 0.1793, 0.2424, 0.2012,
0.1968, 0.2433, 0.2311, 0.1902, 0.1970, 0.1644, 0.1935, 0.1421, 0.1202,
0.2459, 0.2098, 0.1817, 0.1736, 0.2296, 0.2200, 0.2025, 0.1814, 0.1340,
0.2051, 0.2455, 0.2008, 0.1740, 0.2089, 0.2595, 0.1470, 0.2670, 0.1701,
0.2055, 0.2215, 0.2080, 0.1848, 0.2184, 0.2254, 0.1573, 0.1696, 0.1463,
0.1917, 0.2593, 0.1799, 0.2585, 0.2153, 0.2365, 0.1629, 0.1875, 0.2657,
0.2666, 0.2535, 0.1874, 0.1869, 0.2266, 0.2143, 0.1399, 0.2790, 0.1988,
0.1904, 0.1911, 0.2186, 0.1606)
```

Digite

```
> hist(feijao, main =, xlab = "massa de feijão (g)", ylab = "Fre-quência")
```

em que feijão são os dados, main título do gráfico, xlab e ylab são as legendas dos eixos x e y, respectivamente. O resultado é o gráfico mostrado na Figura 5

Tabela 5 – Pesos de caroços de feijão extraídos aleatoriamente de um pacote de 1 kg de feijão preto (g)

0,1188	0,2673	0,1795	0,2369	0,1826	0,1860	0,2045
0,1795	0,1910	0,1409	0,1733	0,2146	0,1965	0,2326
0,2382	0,2091	0,2660	0,2126	0,2048	0,2058	0,1666
0,2505	0,1823	0,1590	0,1722	0,1462	0,1985	0,1769
0,1810	0,2126	0,1596	0,2504	0,2285	0,3043	0,1683
0,2833	0,2380	0,1930	0,1980	0,1402	0,2060	0,2097
0,2309	0,2458	0,1496	0,1865	0,2087	0,2335	0,2173
0,1746	0,1677	0,2456	0,1828	0,1663	0,1971	0,2341
0,2327	0,2137	0,1793	0,2424	0,2012	0,1968	0,2433
0,2311	0,1902	0,1970	0,1644	0,1935	0,1421	0,1202
0,2459	0,2098	0,1817	0,1736	0,2296	0,2200	0,2025
0,1814	0,1340	0,2051	0,2455	0,2008	0,1740	0,2089
0,2595	0,1470	0,2670	0,1701	0,2055	0,2215	0,2080
0,1848	0,2184	0,2254	0,1573	0,1696	0,1463	0,1917
0,2593	0,1799	0,2585	0,2153	0,2365	0,1629	0,1875
0,2657	0,2666	0,2535	0,1874	0,1869	0,2266	0,2143
0,1399	0,2790	0,1988	0,1904	0,1911	0,2186	0,1606

Na prática ao invés de usar a distribuição normal na maioria dos resultados analíticos utiliza-se a distribuição de t de Student, que é uma família de distribuições de probabilidade contínuas usadas para estimar o μ de uma distribuição normal. A grande vantagem dessa distribuição é o pequeno tamanho da amostra usado para encontrar a média, além de não precisar do desvio padrão, geralmente desconhecido. Ela desempenha papel em várias análises estatísticas, incluindo a avaliação da significância estatística entre duas amostras meios, na construção de intervalos de confiança entre duas médias da população, e em linear análise de regressão.

A distribuição de t é uma distribuição de probabilidade teórica. É simétrica, campaniforme, e semelhante à curva normal padrão, porém com caudas mais largas, ou seja, uma simulação da t de Student pode gerar valores mais extremos que uma simulação da distribuição normal. O único parâmetro que a define e caracteriza a sua forma é o número de graus de liberdade. Quanto maior for esse parâmetro, mais próxima da

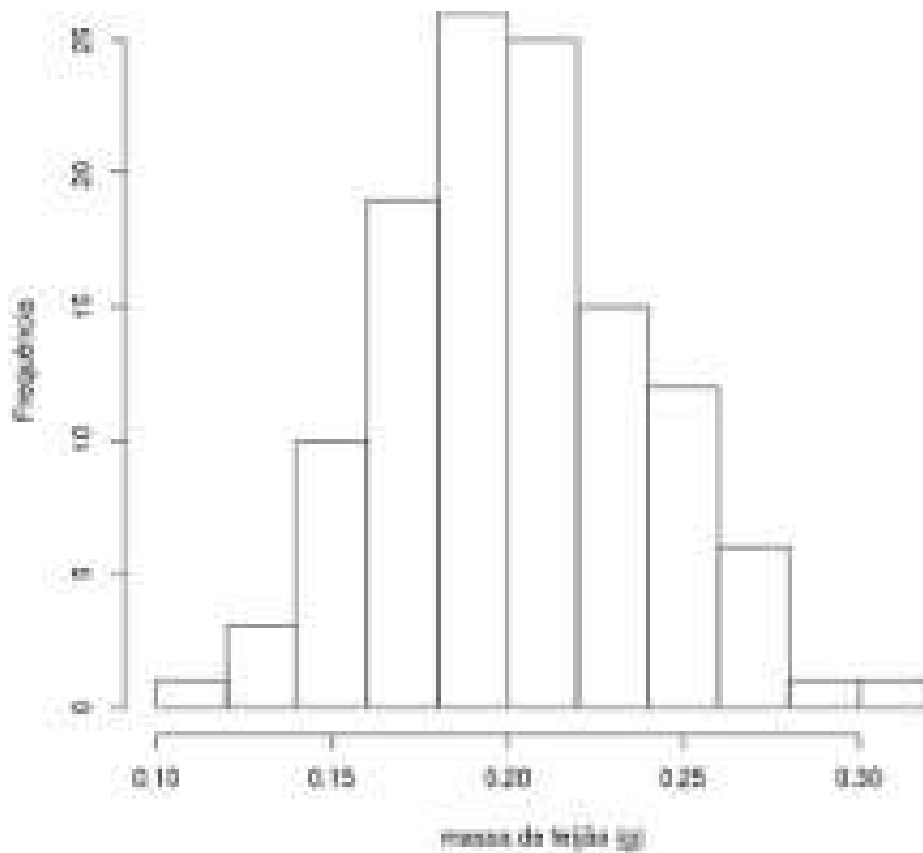


Figura 5 – Distribuição de frequência relativa à massa de feijão

normal ela será. A Figura 6 mostra uma comparação entre a curva da distribuição normal e de t . Nota-se que para um grau de liberdade de 5 as duas curvas são praticamente iguais. A semelhança de comportamento tornam a distribuição de t muito utilizada na estatística.

Exercício 3.2 *Utilize o R para verificar se as medidas de volumes de uma bureta, apresentadas na Tabela, abaixo seguem a distribuição normal.*

9,988	9,973	9,986	9,980	9,975	9,982	9,986	9,982
9,990	9,980	9,989	9,978	9,971	9,982	9,983	9,988
9,980	9,994	9,992	9,984	9,981	9,987	9,978	9,983
9,991	9,981	9,969	9,985	9,977	9,976	9,983	9,976
9,988	9,971	9,986	9,978	9,986	9,982	9,977	9,986
9,983	9,980	9,984	9,979				

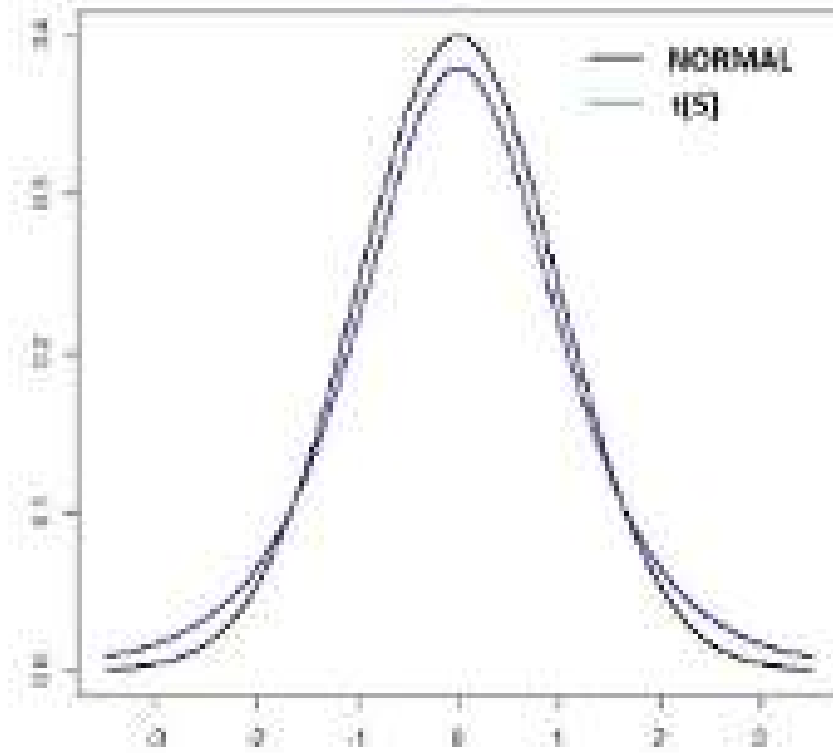


Figura 6 – Comparação da distribuição normal com t

Exercício 3.3 *Um químico estava com dúvidas se a massa de aproximadamente 3,000 g de uma moeda correspondia ao valor mais provável de uma moeda. A importância dessa informação está no aumento ou na redução do erro de uma medida. Use a tabela abaixo para estabelecer qual é provável massa da moeda.*

3,126	3,140	3,092	3,095	3,080	3,065	3,117	3,034
3,053	3,099	3,065	3,059	3,068	3,060	3,078	3,125
3,055	3,105	3,065	3,083	3,065	3,073	3,084	3,148
3,116	3,005	3,115	3,103	3,086	3,103	3,049	2,998
3,181	3,108	3,114	3,121	3,105	3,078	3,147	3,104
3,101	3,049	3,082	3,142	3,082	3,066	3,128	3,112
3,084	3,104	3,093	3,126	3,138	3,131	3,120	3,100
3,091	3,077	3,178	3,054	3,086	3,123	3,115	3,055
3,066	3,113	3,102	3,033	3,112	3,103	3,126	3,111
3,113	3,085	3,117	3,142	3,031	3,083	3,104	

Exercício 3.4 *Um levantamento de três meses dos resultados das análises de densidade (kg m^{-3} num pó de polímero foi realizado como forma de saber*

qual era o comportamento dos resultados. Faça um gráfico de distribuição para verificar o comportamento da densidade.

328	338	334	330	335	330	335	342	339	331	328
332	337	336	333	331	337	316	338	335	332	335
335	335	353	338	320	331	331	334	343	342	352
338	334	340	323	340	331	353	333	335	335	338
340	337	345	332	348	332	342	343	336	335	334
339	330	328	349	337	340	348	332	338	340	336
328	319	336	327	328	339	333	338	332	332	

Exercício 3.5 Uma solução padrão estava sendo usada como referência de um método antigo do laboratório. O químico anotou as 60 últimas determinações para avaliação. Como os dados construa um histograma.

21,0	16,5	18,7	24,5	16,0	21,2	25,4	20,2	14,0	26,0
20,0	18,2	22,2	21,1	20,2	23,4	19,2	16,5	19,0	19,7
17,0	24,7	20,3	17,4	20,5	21,5	22,5	14,5	20,8	21,2
17,7	20,5	19,5	20,9	21,5	19,8	16,2	19,5	20,0	18,2
22,9	21,6	21,8	23,0	18,9	24,0	22,5	20,8	23,8	19,5

4 Média

Por definição média ou média aritmética como a soma de todos os valores dos elementos dividido pelo número de variáveis. Geralmente, a média é uma medida de posição central, e as variáveis se distribuem simetricamente ao seu redor.

O porquê da utilização de triplicatas nas aulas de química analítica?

Esse procedimento está relacionado com a estimativa da concentração da amostra no analito. No $n = 3$, o intervalo de confiança do valor médio está no limite crítico da distribuição de t de Student com $N-1$ graus de liberdade. Aumentando o valor de N , o intervalo de confiança irá diminuir por causa da redução do valor de t_{-1} e do fator $1/\sqrt{N}$. A análise química em triplicata representa um melhoramento considerável da precisão de $t_2 = 12,71$ para $t_3 = 3,04$. No caso de uma quadruplicata observa-se que a precisão permanece praticamente a mesma (Figura 7).

Por ser uma das medidas mais úteis e seguras, a média é o resultado mais usados na estatística para fazer inferências. Um dos seus maiores problemas é o falseamento do resultado final por causa da desproporcionalidade observada entre as medidas. Assim, haverá um valor de média muito alto ou baixo o que leva a interpretações erradas.

A média aritmética é calculada da seguinte forma:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{\sum_{i=1}^n x_i}{N} \quad (4.1)$$

Exemplo 4.1 *Os valores de cinzas em $\mu\text{g g}^{-1}$ de uma amostra foram: 100, 120, 115, 110 e 115. Calcule a média dos resultados obtidos, Depois use o R para calcular a média.*

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{N} = \frac{100 + 120 + 115 + 110 + 115}{5}$$
$$\bar{x} = \frac{170}{5} = 114 \mu\text{g g}^{-1}$$

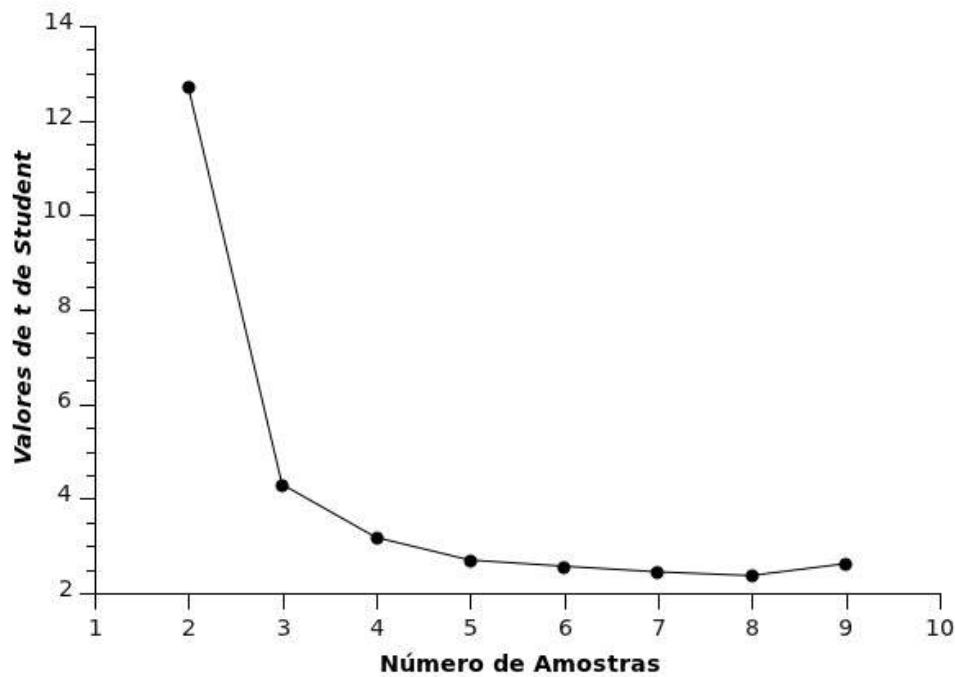


Figura 7 – Gráfico de $1/\sqrt{N}$ em função do número de réplicas N, ilustrando como a precisão diminui quando se aumenta o número de réplicas.

Entrada de dados no R digite

> x<-c(coloque cada número aqui)

Exemplo:

> x<-c(100,120,115,110,115)

Aqui x recebeu os números 100,120,115,110,115

Tirar a média no R

digite mean(dados)

Exemplo

Média x digite > mean(x)

Usando o R

Primeiramente entre com o conjunto de dados:

```
> x<-c(100,120,115,110,115)
```

Bastar digitar:

```
> mean(x)
```

No R o resultado será apresentado da seguinte forma:

```
[1] 114
```

Exercício 4.1 *Cinco amostras de solo foram pesadas antes da análise. As massas registradas são as seguintes: 23,67, 34,53, 31,56, 26,34 e 42,19 g. Calcule a média para a massa dessas cinco amostras.*

Exercício 4.2 *Um estudante de química realizou como parte de seu trabalho de conclusão de curso seis análises de ferro por espectrometria de absorção atômica de água de um poço artesiano. Sabendo que os resultados foram 19,4, 19,5; 19,6, 19,8, 20,1 e 20,3 $\mu\text{g L}^{-1}$, calcule a média destes valores.*

5 Desvio Padrão

A média é sempre acompanhada por uma diferença entre os valores medidos, cujo nome é desvio; ou seja

$$\text{desvio} = \sum_{i=1}^n x_i - \bar{x} \quad (5.1)$$

A diferença $(x_i - \bar{x})$ é utilizada para calcular a medida de dispersão em relação a média, sendo mais conhecida como desvio padrão (dp). A dispersão mostra o distanciamento que ocorre entre o valor verdadeiro e o valor medido, estabelecendo nesse sentido uma grandeza denominada de *Erro Absoluto*:

$$E = x - x_v \quad (5.2)$$

em que E é o erro absoluto, x o valor medido e x_v o valor verdadeiro.

Essa grandeza pode ser utilizada para calcular, por exemplo, em termos relativos o erro cometido por um analista durante a determinação de uma concentração qualquer, sendo expressa da seguinte forma:

$$E_r = \frac{E}{x_v} \quad (5.3)$$

O que é exatidão e precisão?

A exatidão é a proximidade de um valor medido em relação ao valor verdadeiro ou aceito. A precisão descreve a reproduzibilidade dos resultados; ou seja, quão próximas as medidas se encontram umas das outras. Geralmente, a precisão é considerada a dispersão de um conjunto de medidas ou leituras.

Exemplo 5.1 *O resultado de uma análise de cloro foi de 32,90%, enquanto o valor verdadeiro deveria ser de 33,30%. Determine o erro absoluto e relativo cometido pelo analista ao efetuar essa análise química.*

1. *O erro absoluto cometido pelo analista pode ser calculado da seguinte forma:*

$$E = x - x_v = 32,90 - 33,3 = -0,40\%$$

2. De posse do valor do erro absoluto podemos calcular o erro relativo. Para isso devemos utilizar a seguinte equação:

$$E_r = \frac{-0,40}{33,30} * 100 = -1,2\%$$

O que é erro relativo?

O erro relativo de uma medida é o erro absoluto dividido pelo valor verdadeiro, podendo expressos em percentagem, partes por mil ou partes por milhão.

Exercício 5.1 Calcule o erro relativo em percentagem para a análise de uma amostra de água extraída de um rio, cujo valor registrado é de $15,7 \text{ mgL}^{-1}$ de Cu, quando o valor verdadeiro é de fato $18,0 \text{ mgL}^{-1}$.

Exercício 5.2 Uma titulação ácido-base indicou que a concentração de uma solução de HCl de $0,104 \text{ molL}^{-1}$ em vez de $0,110 \text{ mol L}^{-1}$ (valor verdadeiro). Expresse o erro relativo dessa análise química.

Exercício 5.3 Calcule os erros relativos para os seguintes resultados

	Valor medido	Valor aceito
a	22,62 g	22,57 g
b	45,02 mL	45,31 mL
c	2,68%	2,71%
d	85,6 cm	85,0 cm

Devemos chamar a atenção para os valores negativos encontrados nos dois erros, cuja explicação se dar em virtude da técnica utilizada pelo analista que não foi suficiente para alcançar o valor verdadeiro. A Figura 8 representa muito bem o que ocorre geralmente em uma análise química, que é a distância observada entre o valor encontrado e o verdadeiro. Evidentemente que todo químico analítico que chegar o mais próximo possível do valor verdadeiro, daí o termo exatidão.

Como em química analítica geralmente a quantidade de dados obtidas nas análises químicas é baixo, o dp é calculado por sua estimativa:

$$dp = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (5.4)$$

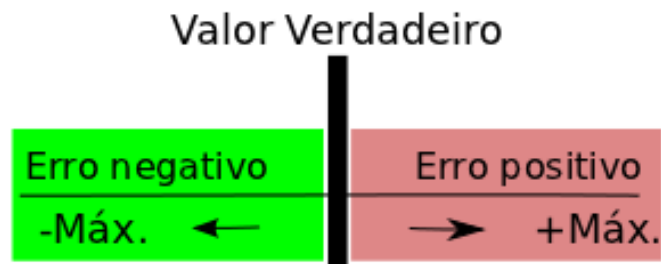


Figura 8 – Variação do erro em relação ao valor verdadeiro

Exemplo 5.2 *Determine o valor de desvio padrão usando o R para o conjunto de dados do Exemplo 4.1.*

Tirar o desvio padrão no R

Digite `sd(dados)`

Usando o R

Primeiramente entre com o conjunto de dados:

> `x<-c(100,120,115,110,115)`

Bastar digitar:

> `sd(x)`

No R o resultado será apresentado da seguinte forma:

[1] 9.354143

Exercício 5.4 *Para a realização da análise química do sódio em água mineral um químico utilizou a fotometria de chama. Ele gostaria de saber qual é o desvio padrão do conjunto de dados obtido pelo químico: 19,4, 20,6, 18,7, 19,2, 21,6, 18,9 e 19,9. Utilize o R para calcular o desvio padrão.*

Exercício 5.5 *Uma amostra de água foi coletada em um igarapé contaminado para a determinação de Cd. Todos os valores encontrados estão acima da resolução estabelecida pelo CONAMA e o químico precisa apresentar os resultados ao órgão ambiental. Uma das maneira é apresentar os resultados na forma de média e desvio padrão. Com os seus resultados calcule os dois parâmetros utilizando o R. Resultados encontrados em mgL^{-1} : 20,2, 18,5, 21,4, 19,2, 21,8 e 18,8.*

Exercício 5.6 *Seis garrafas de vinho da mesma variedade foram analisadas para se determinar o conteúdo de açúcar residual, os resultados são apresentados na Tabela 6*

Tabela 6 – Resultados da análise de açúcar residual das seis garrafas de vinho

Garrafa	Açúcar residual (mv^{-1})				
1	0,99	0,84	1,02		
2	1,02	1,13	1,17	1,02	
3	1,25	1,32	1,13	1,20	1,12
4	0,72	0,77	0,61	0,58	
5	0,90	0,92	0,73		
6	0,70	0,88	0,72	0,73	

1. *Avalie o desvio padrão para cada conjunto de dados*
2. *Combine os dados para obter um desvio padrão absoluto para o método*

Exercício 5.7 *Uma amostra padrão de sódio de $42,0 g L^{-1}$ foi distribuída para cinco laboratórios diferentes. Depois de feitas as análises de sódio os cinco laboratório forneceram os seguintes resultados:*

A	42,5	41,6	42,1	41,9	41,1	42,2
B	39,8	43,6	42,1	40,1	43,9	41,9
C	43,5	42,8	43,8	43,1	42,7	43,3
D	35,0	43,0	37,1	40,5	36,8	42,2
E	42,2	41,6	42,0	41,8	42,6	39,0

Calcule a média e o desvio padrão. Com os resultados obtidos faça um comentário a respeito da precisão e exatidão dos laboratórios.

Exercício 5.8 *Um analista utilizou uma bureta automática para realizar uma série de titulações, obtendo os seguintes valores em mL: 10,002, 9,993, 9,984, 9,996, 9,989, 9,983, 9,991, 9,990, 9,988 e 9,999. Utilize o R para calcular o erro absoluto (Dado o volume teórico das titulações é 10,000 mL) e o erro relativo.*

Tabela 7 – valores obtidos por quatro estudantes de titulações realizadas em sala de aula

Estudante	resultado (mL)	Comentário
A	10,08	Preciso e inexato
	10,11	
	10,09	
	10,10	
	10,12	
B	9,88	Exato e impreciso
	10,14	
	10,02	
	9,80	
	10,21	
C	10,19	Inexato e impreciso
	9,79	
	9,69	
	10,05	
	9,78	
	10,04	Exato e preciso
	9,98	
	10,02	
	9,97	
	10,04	

Exercício 5.9 *A Tabela 7 mostra uma série de resultados de volumes gastos na titulação e comentários a respeito da precisão e exatidão do método. Sabendo que o volume teórico para a titulação utilize o R para calcular a média e o desvio padrão e confirme se os comentários estão corretos ou não.*

6 Mediana

A mediana é uma medida de tendência central, um número que caracteriza as observações de um grupo de dados ordenados separa a metade inferior da amostra, população ou distribuição de probabilidade, da metade superior. Dessa forma, 1/2 da população terá valores inferiores ou iguais à mediana e 1/2 da população terá valores superiores ou iguais à mediana.

No caso de dados ordenados de amostras de tamanho n , se for ímpar, a mediana será o elemento central (EC)

$$EC = \frac{(n + 1)}{2} \quad (6.1)$$

Se n for par, a mediana será o resultado da média simples entre os elementos

$$\frac{n}{2} \text{ e } \frac{n}{2} + 1. \quad (6.2)$$

Exemplo 6.1 Calcule a mediana para o seguinte conjunto de resultados de uma análise química de uma amostra de calcário: 1,0, 3,0, 5,0, 7,0 e 9,0%.

Como a sequência é um número ímpar, a posição central será

$$\frac{n + 1}{2} = \frac{5 + 1}{2} = 3$$

Logo, a mediana é o terceiro elemento que é 5.

Usando o R

Primeiramente entre com o conjunto de dados:

`> x <- c(1,3,5,7,9)`

Bastar digitar:

`> median(x)`

No R o resultado será apresentado da seguinte forma:

`[1] 5`

Exercício 6.1 Determine a mediana para o seguinte conjunto de análises químicas realizadas para o Pb^{2+} : 20,1, 19,5, 20,3, 19,7, 20,0, 19,4, e 19,6.

Exemplo 6.2 *Um analista após analisar amostras de águas por espectrometria de absorção atômica encontrou os seguintes valores para a concentração de ferro: 1,0, 2,0, 4,0, 8,0, 9,0, 10,0 mg L⁻¹. Determine a mediana para esses resultados.*

Como os resultados formam um conjunto par, portanto, não existe um elemento central. Neste caso, a mediana será calculada com os dois valores centrais (o terceiro e o quarto valores). Assim

$$\text{mediana} = \frac{4 + 8}{2} = 6$$

Usando o R

Primeiramente entre com o conjunto de dados:
> x<-c(1,2,4,8,9,10)
Bastar digitar:
> median(x)
No R o resultado será apresentado da seguinte forma:
[1] 6

Exercício 6.2 *O químico que realizou o Pb²⁺ do exercicios 6.1 esqueceu de colocar um resultado. Calcule a mediana para o novo conjunto de dados: 20,1, 19,5, 20,3, 19,7, 20,0, 19,4, 19,6 e 19,9.*

Exercício 6.3 *As medida da concentração de potássio em mg L⁻¹ uma amostra foram 10,4, 10,4, 10,6, 10,3, 10,5 e 10,6. Calcule a média, mediana e desvio padrão para essa série de resultados.*

Exercício 6.4 *Para investigar a reprodubibilidade de um método para a determinação de selênio em arroz, um analista executou uma série de análises químicas, cujos resultados foram:*

<i>Amostra</i>	<i>Conc. de selênio ($\mu\text{g g}^{-1}$)</i>
<i>1</i>	<i>0,07</i>
<i>2</i>	<i>0,07</i>
<i>3</i>	<i>0,08</i>
<i>4</i>	<i>0,07</i>
<i>5</i>	<i>0,07</i>
<i>6</i>	<i>0,08</i>
<i>7</i>	<i>0,08</i>
<i>8</i>	<i>0,09</i>
<i>9</i>	<i>0,08</i>

Calcule a média, mediana e desvio padrão dos resultados obtidos

7 Propagação de erros

A partir do conhecimento da incerteza dos equipamentos utilizados durante uma análise química pode-se estimar a incerteza final em uma medida. Os erros em números individuais irá se propagar por uma série de cálculos, que depende diretamente do tipo de operação realizada: adição, subtrações, multiplicação ou divisão.

7.1 Adição e subtração

Considere a adição e subtração da seguinte sequência de números:

$$(65,00 \pm 0,07) + (16,13 \pm 0,01) - (22,68 \pm 0,02) = 58,51(\pm?)$$

A incerteza absoluta¹ na adição e subtração é aditiva, será necessário representada pelos desvios padrão representa erros indeterminados associados com cada número. Se esses desvios fossem somados, então a incerteza final poderia ser $\pm 0,10$. No entanto, a incerteza não é aditiva, sendo necessário levar em conta a variância absoluta (s_a^2):

$$s_a^2 = s_b^2 + s_c^2 + s_d^2 \quad (7.1)$$

$$s_a = \sqrt{s_b^2 + s_c^2 + s_d^2} \quad (7.2)$$

No caso do exemplo acima tem-se

$$s_a = \sqrt{(\pm 0,07)^2 + (\pm 0,01)^2 + (\pm 0,02)^2} = \pm 7,3 \times 10^{-2}$$

Exemplo 7.1 *Um aluno está em dúvida para expressar o resultado de uma análise química. A análise de fósforo (mg L^{-1}) teve os seguintes resultados $3,978 \pm 0,004$, $2,536 \pm 0,003$ e $3,680 \pm 0,003$. Como o aluno deve expressar os seus dados?*

A média dos dados pode ser calculada da seguinte forma:

$$\bar{x} = \frac{(3,978 \pm 0,004\%) + (2,536 \pm 0,003\%) + (3,680 \pm 0,003\%)}{3}$$

¹ Valor máximo dos desvios de uma medida em relação à média

A incerteza na soma é

$$s_a = \sqrt{(\pm 0,004)^2 + (\pm 0,003)^2 + (\pm 0,003)^2} = \pm 0,006 \text{ mg L}^{-1}$$

A média pode ser calculada da seguinte forma

$$\bar{x} = \frac{3,978 + 2,536 + 3,680}{3} = 3,398$$

Portanto, o resultado final será

$$\bar{x} = 3,398 \pm 0,006 \text{ mg L}^{-1}$$

7.2 Multiplicação e divisão

Considere a seguinte operação de um cálculo arbitrário realizado durante uma análise química

$$\frac{(13,67 \pm 0,02)(120,4 \pm 0,2)}{4,623 \pm 0,006} = 356,0(\pm?)$$

Neste caso, as incertezas relativas² são aditivas, sendo que o erro mais provável é representado pelo quadrado das somas das variâncias relativas:

$$(s_a)_{rel} = \sqrt{(s_b)_{rel}^2 + (s_c)_{rel}^2 + (s_d)_{rel}^2} \quad (7.3)$$

O cálculo da incerteza relativa dos dados acima é feito da seguinte forma:

$$(s_b)_{rel} = \frac{\pm 0,02}{13,67} = \pm 0,0015$$

$$(s_c)_{rel} = \frac{\pm 0,2}{120,4} = \pm 0,0017$$

$$(s_d)_{rel} = \frac{\pm 0,006}{4,623} = \pm 0,0013$$

$$(s_a)_{rel} = \sqrt{(\pm 0,0015)^2 + (\pm 0,0017)^2 + (\pm 0,0013)^2} = \pm 2,6 \times 10^{-3}$$

A incerteza absoluta será

$$s_a = 356,0x(\pm 2,6 \times 10^{-3}) = \pm 0,93$$

A expressão correta do resultado será $356,0 \pm 0,9$

² Valor do quociente entre a incerteza absoluta e o valor mais provável da medida. A incerteza relativa exprime-se por vezes em termos de percentagem e define então a chamada percentagem de erro, ou incerteza percentual

Exemplo 7.2 Calcule a incerteza no número de milimols de cloreto contido em 250 mL. Para isso, o químico utilizou-se de três alíquotas iguais de 25,00 mL para a titulação com nitrato de prata. Os resultados das titulações foram 36,78, 36,82 e 36,75 mL. A concentração do nitrato utilizado era $0,1167 \pm 0,0002 \text{ mol L}^{-1}$

Para a expressar o resultado corretamente é necessário primeiramente calcular a média e desvio padrão dos volumes obtidos na titulação.

Usando o R

Primeiramente entre com o conjunto de dados:

```
> x<-c(36,78,36,82,36,75)
```

Para o cálculo da média basta digitar:

```
> mean(x)
```

No R o resultado será apresentado da seguinte forma:

```
[1] 36.78333
```

O cálculo do desvio padrão

```
> sd(x)
```

No R o resultado será apresentado da seguinte forma:

```
[1] 0.03511885
```

Neste ponto, o resultado seria expresso da seguinte forma:

$$\text{mmol Cl}^- = (0,1167 \pm 0,0002 \text{ mmol mL}^{-1})(36,78 \pm 0,04 \text{ mL}) = 4,292(\pm?)$$

No entanto, é necessário calcular as incertezas relativas

$$(s_b)_{rel} = \frac{\pm 0,0002}{0,1167} = \pm 0,0017$$

$$(s_c)_{rel} = \frac{\pm 0,035}{36,78} = \pm 0,00095$$

$$(s_a)_{rel} = \sqrt{(\pm 0,0017)^2 + (\pm 0,00095)^2} = \pm 1,9 \times 10^{-3}$$

Agora temos condições de calcular a incerteza absoluta em milimols de Cl^- é

$$4,292 \times (\pm 0,0019) = \pm 0,0082 \text{ mmol}$$

mmol de Cl^- em 25 mL = $4,292 \pm 0,0082$ mmol

Para 250 mL tem-se mmol $\text{Cl}^- = 10(4,292 \pm 0,0082) = 42,92 \pm 0,08$ mmol

Exercício 7.1 Calcule a incerteza das seguintes somas:

1. $(128 \pm 2) + (1025 \pm 8) - (636 \pm 4)$

2. $(16,25 \pm 0,06) - (9,43 \pm 0,03)$

3. $(46,1 \pm 0,4) + (935 \pm 1)$

Exercício 7.2 Calcule a incerteza das seguintes equações:

1. $(2,78 \pm 0,04)(0,000506 \pm 0,000006)$

2. $(36,2 \pm 0,4) / (27,1 \pm 0,6)$

3. $(50,23 \pm 0,07)(27,86 \pm 0,05) / (0,1167 \pm 0,0003)$

8 Arredondamento de dados

É especialmente importante postergar o arredondamento até que o cálculo seja completado. Pelos menos um dígito extra, depois dos algarismos significativos, deve ser mantido durante todos os cálculos de maneira que se evitem os erros no arredondamento. Algumas vezes esse dígito extra é chamado dígito "guarda". As calculadoras modernas geralmente mantêm vários dígitos extras que não são significativos e o usuário precisa ser cuidadoso no arredondamento apropriado de resultados finais para que apenas os algarismos significativos sejam incluídos.

Exemplo 8.1 *Uma amostra de 3,4842 g de uma mistura sólida contendo ácido benzóico (C_6H_5COOH , HBz), ($122,123 \text{ g mol}^{-1}$) foi dissolvida e titulada com base em bureta de vidro até o ponto final na presença de fenolftaleína. O ácido consumiu 41,36 mL de NaOH $0,2328 \text{ mol L}^{-1}$. Calcule a porcentagem de HBz na amostra.*

A reação do HBz com o NaOH é 1:1, então

$$N^{\circ} \text{ mol milimol HBz} = N^{\circ} \text{ mol milimol NaOH}$$

$$N^{\circ} \text{ mol milimol HBz} = 41,36 \cdot 0,2328 = 9,628608 \text{ milimol}$$

Em termos de massa de HBz temos

$$\text{massa de HBz} = \frac{N^{\circ} \text{ milimol HBz} \cdot 122,123 \text{ g}}{1.000 \text{ milimol}}$$

$$\text{massa de HBz} = \frac{9,628608 \cdot 122,123}{1.000} = 1,17587449478$$

Agora temos condições de calcular a % de HBz

$$\% \text{ de HBz} = \frac{1,17587449478}{3,4842} \cdot 100 = 33,74877657075$$

Neste caso, para expressar o resultado obtido deve ser feito um arredondamento. Para que seja feito, o arredondamento deve obedecer alguns princípios, como a precisão do equipamento utilizado, operações matemáticas necessárias para encontrar o resultado final. Vejamos o valor de

% encontrado para HBz de 33,74877657075 é um valor alto para as quatro casas decimais da balança usada para pesa a amostra e muito maior que a bureta usada para realizar a titulação ($\pm 0,02 mL$). Ao verificar as incertezas nesse processo verifica-se:

1. A posição do nível de líquido na bureta estimada como $\pm 0,02 mL$. No entanto, as leituras iniciais e finais precisam ser feitas, assim, o dp do volume pode ser calculado levando em conta a propagação de erro:

$$dp = \sqrt{0,02^2 + 0,02^2} = 0,028 mL \quad (8.1)$$

A incerteza relativa no volume será:

$$\frac{dp}{V \text{ (gasto na bureta)}} = \frac{0,028}{41,36} * 100 = 0,068\%$$

2. Geralmente a incerteza absoluta para uma massa obtida em uma balança analítica será da ordem de $\pm 0,0001 g$ (s_D). Dessa forma, a incerteza relativa no caso da balança será

$$\frac{s_D}{\text{massa pesada}} = \frac{0,0001}{3,4842} * 100 = 0,0029\%$$

3. A incerteza absoluta associada com a concentração molar de uma solução de um reagente é $\pm 0,0001 mol L^{-1}$ (s_M) e assim a incerteza relativa na concentração molar do NaOH será

$$\frac{s_M}{\text{Conc.}} = \frac{0,0001}{0,2328} * 100 = 0,043 \%$$

4. A incerteza relativa na massa molar do HBz é várias ordens de grandeza menor que qualquer incerteza associada com os três dados experimentais e, portanto, sem consequência. Observe, contudo, que devemos manter dígitos suficientes no cálculo para que a massa molar seja dada, pelo menos, com um dígito a mais (o dígito guarda) que qualquer um dos dados experimentais. Assim, 122,123 no cálculo da massa molar (estamos mantendo dois dígitos extras).

5. Para estabelecer a incerteza do valor de concentração molar do NaOH (s_f), temos que levar em conta as três incertezas relativas

$$\frac{s_f}{medida} = \sqrt{\left(\frac{0,028}{41,36}\right)^2 + \left(\frac{0,0001}{3,4842}\right)^2 + \left(\frac{0,0001}{0,2328}\right)^2}$$

$$\frac{s_f}{medida} = 8,02 \times 10^{-4}$$

$$s_f = 8,02 \times 10^{-4} \cdot medida = 8,02 \times 10^{-4} \cdot 33,74877657075 = 0,0270757998961$$

Assim, a incerteza no resultado calculado é 0,027% de HBz e devemos relatar o resultado como 33,75($\pm 0,03$)% de HBz.

Os valores de média e desvio padrão encontrados nos exemplos 4.1 e 5.2 são usados geralmente para expressar os resultados de uma análise química. Assim, o teor de cinzas é geralmente expresso da seguinte forma $114 \pm 9 \mu g g^{-1}$.

9 Medidas da Precisão

Os químicos usam o desvio padrão da amostra para relatar a precisão dos seus dados. Muitas vezes são encontrados outros três termos nos trabalhos analítico.

9.1 Coeficiente de Variação

O coeficiente de variação (também conhecido como erro relativo) é uma medida relativa de variabilidade, sendo usada para expressar a relação percentual do desvio padrão com a média dos valores obtidos. O CV, como é conhecido o coeficiente de variação, é expresso da seguinte forma:

$$CV = \frac{dp.100\%}{\bar{x}} \quad (9.1)$$

O CV é uma variável que pode ser usada para avaliar a qualidade de laboratórios e métodos. Por exemplo, Horwitz realizou um estudo com resultados de 150 laboratórios estabeleceu o CV aceitável para métodos (Figura 9). Nota-se que o CV aumenta à medida que a quantidade a ser medida diminui; ou seja, quanto maior a capacidade de medida mais alta é a variação dos erros. No estudo também ficou proposta uma equação para encontrar o CV em função da concentração (C em massa analito/massa amostra):

$$CV(\%) \approx 2^{(1-0,5\log C)} \quad (9.2)$$

Exemplo 9.1 *Um químico para avaliar um novo método analítico, preparou uma solução conhecida de cálcio de 50 mg L^{-1} . Esta solução foi analisada seis vezes por absorção atômica, sendo obtido os seguintes resultados em mg L^{-1} : 48,2, 51,0, 46,6, 51,5, 43,8 e 46,9. Diante desses resultados calcule a média, desvio padrão e CV usando o R.*

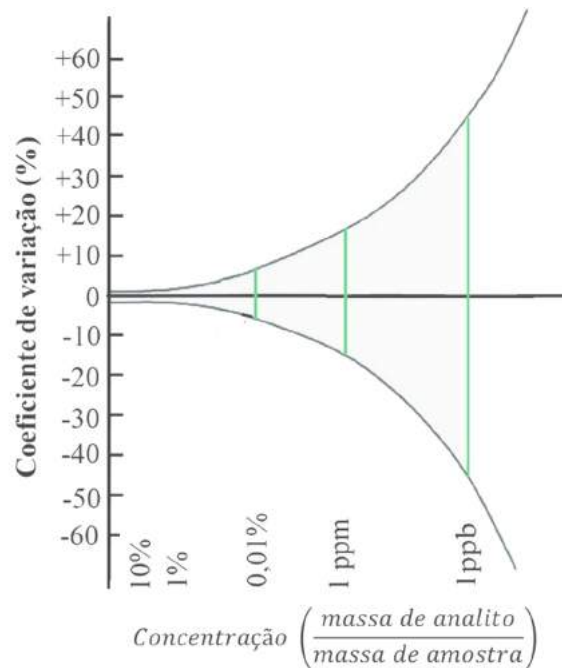


Figura 9 – Coeficiente de variação com função da concentração da amostra.

Como interpretar o CV?

Quando o CV for inferior a 10% dizemos que o experimento apresentou uma ótima precisão experimental; quando variar de 10 a 15% dizemos que o experimento apresentou boa precisão experimental; quando estiver no intervalo $> 15\% \leq 20\%$ dizemos que o experimento apresentou uma precisão experimental regular ou aceitável; quando estiver no intervalo $> 20\% \leq 30\%$ dizemos que o experimento apresentou uma péssima precisão experimental e quando for superior a 30% dizemos que o experimento apresentou uma precisão experimental muito péssima.

Usando o R

Primeiramente entre com o conjunto de dados:

```
> x<-c(48.2,51.0,46.6,51.5, 43.8,46.9)
```

Para calcular a média bastar digitar:

```
> mean(x)
```

No R o resultado será:

```
[1] 48
```

Para calcular o dp

```
> sd(x)
```

O resultado será:

```
[1] 2.901724
```

Para calcular o CV

```
> sd(x)/mean(x)*100
```

O resultado será:

```
[1] 6.045258
```

Exercício 9.1 *Os resultados da Tabela 8 são referentes a dois métodos usados para determinar a quantidade de cálcio no sangue. O primeiro foi por AAS e o segundo por colorimetria. Calcule no R o CV de cada método e discuta os resultados.*

Tabela 8 – Resultados de dois métodos usados na determinação de cálcio no sangue

AAS (mg L ⁻¹)	Colorimetria (mg L ⁻¹)
10,9	9,2
10,1	10,5
10,6	9,7
11,2	11,5
9,7	11,6
10,0	9,3
	10,1
	11,2

Exercício 9.2 O estanho e o zinco contidos em uma solda foram analisados por espectrometria de absorção atômica. a) Sn: 0,022, 0,025 e 0,026% e b) Zn: 33,27, 33,37 e 33,34%. Utilizando o R calcule a Média, Desvio padrão e CV.

Exercício 9.3 Os resultados das análises químicas de uma liga de prata apresentaram os seguintes percentuais: 95,67, 95,61, 95,71 e 95,60% de Ag. Utilizando o R calcule: Média, Desvio padrão e CV.

Exercício 9.4 Uma série de medidas da quantidade de etanol em efluente tratado resultou nos seguintes dados (%): 0,77, 0,67, 0,71, 0,90 e 0,78. Calcule o CV desses resultados.

Exercício 9.5 As porcentagens de um constituinte A num composto AB foram determinadas encontrando-se 48,32, 48,36, 48,23, 48,11 e 48,38%. Calcule a média, desvio-padrão e CV para esse resultados.

Exercício 9.6 Calcule o coeficiente de variação para as seguintes medidas de ferro (mgL^{-1}) feita em amostra de água contaminada: 34,6, 29,5, 32,2, 33,7, 34,6, 32,4 e 35,1.

10 Amostra

Antes de definir o conceito de amostra devemos saber antes o que seja população. População é o conjunto de todos os elementos¹ existentes ou que serão obtidos em um processo qualquer ou que se deseja estudar sob certas condições. Exemplos: um lote de peças, habitantes do Estado do Amazonas, águas de um igarapé, etc.

Ao conjunto representativo de todos os elementos extraídos de uma população aleatoriamente é chamado de amostra. Exemplos: uma amostra retirada de um lote de comprimidos, uma amostra de rotina durante o processo de um lote de latas de alumínio, etc.

O que são réplicas?

As réplicas são amostras com aproximadamente o mesmo tamanho das que são submetidas a análises exatamente da mesma forma.

O número de elementos (ou quantidade) existente na amostra, geralmente é indicado pela letra **n**. Quando realizamos um estudo estatístico devemos considerar o número de amostras. O número ideal de análises ou de medidas (**n** ou **k**) sobre uma amostra, para realizar um estudo estatístico, é algo que deve ser determinado em comum acordo entre o grupo de trabalho do laboratório e deve estar de acordo com a característica da análise, quantidade de amostra, dificuldade analítica e equipamentos disponíveis e confiáveis.

10.1 Amostragem

Uma amostra de n unidades tomada de um universo de N unidades será uma possibilidade entre $c = \frac{N!}{n!(N-n)!}$ diferentes amostras. Como selecionar a amostra? No caso de estudos ambientais, o uso de amostragem sistemática, estratificada ou não, ou mesmo preferencial, é a mais utilizada. Pela dificuldade em localizar os pontos de amostragem previamente

¹ Unidade considerada para um estudo, por exemplo, um objeto, um indivíduo, uma quantidade de matéria

escolhidos ao acaso sobre mapas da área, a amostragem aleatória irrestrita tem sido considerada pouco prática nos estudos de ecossistemas naturais.

A amostragem sistemática é feita quando apenas o primeiro membro da amostra ou do estrato for selecionado ao acaso, sendo os demais tomados a intervalos regulares. A amostragem sistemática tem como princípio a divisão da área de coleta em estratos ou segmentos, sendo feita de forma subjetiva.

Os guias de estatística mostram que para amostras heterogêneas, temos inicialmente que estabelecer qual é a quantidade mínima de amostra a ser coletada. Ingamells e colaboradores determinaram a constante de amostragem de Ingamells, k_s :

$$mCV^2 = k_s \quad (10.1)$$

O m é o peso da amostra analisada e CV é a percentagem relativa do desvio padrão da composição da amostra. O k_s representa a massa de amostra para 1% de incerteza na amostragem para um limite de confiança de 65%.

Exemplo 10.1 *A constante de Ingamells para a análise da quantidade de nitrogênio em trigo é de 0,50 g. Qual massa deve ser pesada para obter uma precisão de 0,2% (CV) na determinação de nitrogênio em trigo?*

Sabemos que

$$mCV^2 = k_s$$

Substituindo a constante de Ingamells k_s de 0,50 g e o CV^2 de 0,2% teremos:

$$\begin{aligned} m(0,2)^2 &= 0,5 \\ m &= \frac{0,5}{(0,2)^2} = 12,5 \text{ g} \end{aligned}$$

Na prática, os 12,5 g não é totalmente determinado pois o processo de trituração da amostra não consegue deixá-la totalmente homogênea.

Exemplo 10.2 *Para conhecer a constante de Ingamells é necessário inicialmente determinar a média de um analito a partir de uma massa conhecida de amostra. Assim, vamos imaginar a análise de fósforo em amostras*

de sedimentos pesando 0,500 g. A média obtida para 10 determinações de fósforo foi de 23,14 mg kg⁻¹ e o desvio padrão de 2,20 mg kg⁻¹. Então vamos calcular k_s para essa análise.

Cálculo do erro relativo (%)

$$CV = \frac{2,20}{23,14} 100 = 9,50\%$$

Agora temos condições de calcular o k_s

$$k_s = mCV^2 = 0,500 \cdot 9,50^2 = 45,19 \text{ g}$$

Exercício 10.1 A constante de amostragem (k_s) para o radioisotopo ²⁴Na em amostras homogêneas de fígado humano é de aproximadamente 35 g. Qual erro relativo seria cometido por você se em sua análise química fossem utilizados 1,0 g?

Exercício 10.2 Para encontrar o k_s de K₂O de solo, um analista realizou 12 determinações em ICP. Os resultados estão apresentados abaixo:

0,247 0,300 0,236 0,258 0,304 0,330

0,247 0,275 0,212 0,311 0,258 0,187

Sabendo que para todas as análises químicas de K₂O foram usadas amostras pesando 0,100 g determine o valor de k_s . Dica calcule a média e o desvio padrão para determinar o CV.

Até o momento aprendemos a calcular a massa de amostra necessária para fazer uma análise química. Vamos aprender qual o número de amostras é necessário para fazer uma análise química. O número de amostras a serem coletadas pode ser estimado a partir do erro padrão (desvio padrão relativo - s_m) em que o desvio padrão de uma observação dividido pela raiz quadrada do número de observações (N) é dado por

$$s_m = \frac{dp}{\sqrt{N}} \quad (10.2)$$

Observa-se que o s_m diminui à medida que aumenta o número de observações, mais precisamente, ele é inversamente proporcional à raiz quadrada do número de observações.

Exemplo 10.3 *Um químico está interessado em analisar a quantidade de cálcio presente no leite em caixa de um lote. Ele está em dúvida qual a quantidade mínima tem que ser coletada para que os resultados de cálcio no leite sejam confiáveis. Como ele não sabia a quantidade mínima necessária para analisar resolveu fazer três titulações para determinar a quantidade e encontrou o desvio padrão 0,50. Sabendo que o erro relativo admitido para análise é de 0,20, calcule o número mínimo de coletas a serem feitas.*

Sabemos que

$$0,20 = \frac{0,50}{\sqrt{N}}$$

$$\sqrt{N} = \frac{0,5}{0,2}$$

$$N = 2,5^2 = 6,25$$

Conforme podemos perceber bastar coletar 6 amostras que os resultados obtidos corresponderam ao admitido para análise química do leite

Agora, qual o número ideal amostra deve ser coletado para se ter uma boa representatividade? Idealmente, várias amostras deveriam ser coletadas e analisadas para que os resultados pudesse gerar o valor mais provável. Para responder a essa pergunta devemos primeiramente definir o que seja intervalo de confiança (colocar um fator probabilístico), neste ponto fazer uma estimativa de μ :

$$\mu = \bar{x} \pm \frac{t.dp}{\sqrt{N}} \quad (10.3)$$

em que N número de amostras e t um fator estatístico obtido da distribuição de *student*, sendo determinado pelo número de amostras e nível de confiança desejado (Tabela 18, pág. 114). Primeiramente, vamos trabalhar um pouco com essa equação em alguns exemplo e exercícios.

Exemplo 10.4 *A quantidade de íons de sódio presente na urina foi determinada por potenciometria com um eletrodo de íon seletivo. Os seguintes valores foram obtidos: 102, 97, 99, 98, 101, 106 mmol L⁻¹. Quais são as concentração média e o limite de confiança de 95% e 99%?*

A média e o desvio padrão dos valores acima é respectivamente 100,5 e 3,27 mmol L⁻¹. Existem seis medidas e portanto o grau de liberdade é

cinco. Sendo assim, os valores tabelados de t (Tabela 18, pág. 114) para 95% e 99% são respectivamente 2,57 e 4,03. Com esses dados podemos encontrar os limites de confiança usando a equação 10.3. Para 95% tem-se

$$100,5 \pm \frac{2,57 \cdot 3,27}{\sqrt{6}} = 100,5 \pm 3,4 \text{ mmol L}^{-1}$$

Para 99% tem-se

$$100,5 \pm \frac{4,03 \cdot 3,27}{\sqrt{6}} = 100,5 \pm 5,4 \text{ mmol L}^{-1}$$

Exercício 10.3 A quantidade de carboidrato de uma glicoproteína do açúcar foi determinada em cinco amostras. Os valores encontrados pelo analista em 100 g de amostra foram 12,6, 11,9, 13,0, 12,7 e 12,5 g. Determine o intervalo de confiança para 90% e 95% dos valores de concentração encontrados.

Exercício 10.4 Durante um experimento um químico obteve os seguintes valores de pH para um tampão

5,12 5,20 5,15 5,17 5,16 5,19 5,15

Calcule o limite de confiança para 95% desses valores.

Exercício 10.5 Dez replicatas referente à análise de mercúrio (ng mL^{-1}) em amostras de solos contaminados foram os seguintes:

23,3 22,5 21,9 21,5 19,9 21,3 21,7 23,8 22,6 24,7

Calcule a média, desvio padrão, CV e limite de confiança de 95%.

Retornando ao processo de estimativa do número de amostras ideal para uma análise química, temos que reescrever a equação 10.3:

$$\mu - \bar{x} = \pm \frac{tdp}{\sqrt{N}} \quad (10.4)$$

Desconsiderando o \pm temos

$$\mu - \bar{x} = \frac{tdp}{\sqrt{N}} \quad (10.5)$$

Se observamos a diferença $\mu - \bar{x}$, chega-se a um erro, que na equação 10.2 é o s_m ; ou seja, temos aqui uma forma de expressar um erro cometido

pelo uso de amostras pequenas em relação ao grande número de medidas. Nesse caso necessariamente o t de student sugere uma proximidade probabilística de \bar{x} está próximo de μ .

Vamos estabelecer uma equação que permite saber o número ideal para a realização de uma análise química:

$$s_m = \frac{tdp}{\sqrt{N}} \quad (10.6)$$

$$S_m^2 = \frac{t^2 dp^2}{N} \quad (10.7)$$

Então, o número ideal para uma análise química representativa é

$$N = \frac{t^2 dp^2}{s_m^2} \quad (10.8)$$

Exemplo 10.5 *A quantidade de ferro permitida no ouro é 5% (massa/massa), para um desvio padrão de 0,021 (CV 2,1%). Quantas amostras devem ser tomadas para que seja tenha um desvio padrão relativo de 0,016. Faça o seu cálculo para um nível de confiança de 95%.*

Inicialmente, devemos calcular o número de amostra baseado no limite de confiança para infinitas medidas, cujo valor de t é 1,96. Dessa forma, teremos um N de

$$N = \frac{(1,96)^2(0,021)^2}{(0,016)^2} = 6,6$$

A partir do resultado 6,6 começaremos a fazer interações para obter um número finito de amostras. Dessa forma tem-se para $N = 7$, $t = 2,37$

$$N = \frac{(2,37)^2(0,021)^2}{(0,016)^2} = 9,7$$

Na nova interação $N = 10$, $t=2,23$

$$N = \frac{(2,23)^2(0,021)^2}{(0,016)^2} = 8,6 = 9$$

Percebe-se que o N está entre 9 e 10, ou seja, para se ter um nível de confiança de 95%, 10 amostras de ouro devem ser analisadas.

Exercício 10.6 *A análise da acidez da aspirina revelou que o método empregado de titulação tinha um desvio padrão de 0,25 para uma quantidade de 1,00 g em comprimidos triturados. Quantas amostras são necessárias para se obter um erro relativo de 0,05 no limite de confiança de 95%?*

Exercício 10.7 *Na busca da melhor condição de medida um aluno de iniciação de pesquisa resolveu fazer uma série de medidas (Tabela 9). Calcule a média, desvio padrão e o número ideal para um limite de confiança de 95%. Considere como ideal um s_m de 0,1.*

Tabela 9 – Série de medidas obtidas pelo aluno de iniciação científica

48,4	50,8	51,3	47,9
48,6	50,6	49,3	49,7
48,2	51,0	49,3	50,3
54,8	54,6	50,7	53,9
49,6	54,2	48,3	50,5
54,8	54,8	52,3	52,5

Exercício 10.8 *O s_m da determinação das cinzas do arroz de 1,5 g é de $\pm 2,0\%$. Use essa informação para calcular quantas amostras pesando 1,5 g serão necessárias para uma análise química com limite de confiança de 95%. Dado o desvio-padrão é de 0,80%*

Exercício 10.9 *Um químico estava precisando determinar a quantidade de cálcio de um lago Amazônico. Então, ele levou em seu barco um fotômetro de chama e os reagentes para a realização da análise. Tradicionalmente, são realizadas determinações em triplicatas e os resultados expressos em termos de média e desvio padrão. Após ter realizado as três análises, o químico resolveu aprimorar seus resultados para isso ele não sabia quantas amostras são necessárias para se a obtenção de dados mais confiáveis. Sabendo que para amostras de 100 mL a quantidade de cálcio é de $4,10 \pm 0,12 \text{ mg L}^{-1}$, calcule a quantidade de amostras que devem ser coletadas para um erro s_m aceitável de 0,01.*

Exercício 10.10 *Geralmente nas aulas práticas de química analítica são feitas três titulações e os resultados expressos em termos de média e desvio padrão. Um estudante realizou uma dessas titulações e encontrou os seguintes volumes: 3,91, 4,01 e 3,97 mL. Com esses valores determine qual é o s_m para um limite de confiança de 95%.*

11 Teste de significância

A comparação dos valores de um conjunto de resultados com o valor verdadeiro ou com os valores de outros conjuntos de resultados permite verificar a acurácia e a precisão de um método analítico, por exemplo. O desenvolvimento de um novo método analítico passa necessariamente por um teste estatístico para verificar a significância entre o método clássico e o novo. Entre os métodos geralmente usados se destacam o teste de F, *t* student e *t* pareado.

Esses métodos utilizam o número de graus de liberdade, em termos estatísticos, o número de valores independentes necessários para determinar a quantidade estatística. Assim, um conjunto de n valores tem n graus de liberdade enquanto a soma $\sum(x - \bar{x})^2$ tem $n-1$ graus de liberdade. Isso ocorre porque qualquer que seja o valor definido de \bar{x} , somente $n - 1$ valores podem ser escolhidos independentemente definido pelos outros valores.

11.1 Teste de F

Este é um teste indicado para comparar se a de precisão entre dois métodos diferença entre dois métodos baseado sobre seus desvios padrão. O teste de Fisher (F), que é a comparação de duas variâncias¹, serve para determinar se duas populações possuem variabilidades semelhantes ou se as variabilidades podem ser consideradas diferentes. No primeiro caso, as populações são chamadas de Homocedásticas, enquanto no segundo caso as populações são designadas Heterocedásticas. O teste de F é definido em termos as variâncias de dois métodos, em que a variância é a divisão entre o quadrado dos desvios padrão:

$$F = \frac{s_1^2}{s_2^2} \quad (11.1)$$

¹ A variância é o quadrado do desvio padrão possuindo as mesmas unidades dos dados só que ao quadrado. Apesar de a vantagem de ser aditiva em muitas situações, os químicos preferem usar o dp como uma medida de precisão.

em que $s_1^2 > s_2^2$. Existem dois diferentes graus de liberdade v_1 e v_2 , cada um é calculado da seguinte forma $v = N - 1$.

Se o F calculado na equação 11.1 for maior que o F tabelado (Tabela 17, pág. 113), o resultado indica que existe uma diferença significativa entre as duas medidas usadas para comparar, por exemplo, dois métodos analítico.

Exemplo 11.1 *A qualidade de uma análise química de cálcio executada por um analista principiante e um experiente foi avaliada pela realização de 6 e 5 determinações complexométricas por EDTA, respectivamente. Os resultados obtidos pelos dois analistas foram: para o iniciante de $32,25 \pm 0,34\%$ e o mais experiente de $35,35 \pm 0,25\%$.*

Para estabelecer se o analista principiante, inicialmente devemos calcular o valor de F_{calc} :

$$F_{calc} = \frac{0,34^2}{0,25^2} = 1,85$$

Ao consultar os valores respectivos dos graus de liberdade 5 e 4 na Tabela 17 (pág. 113) encontramos o valor de $F = 6,26$. Como o valor calculado é menor que o valor tabelado, podemos afirmar que no nível de confiança de 95% os resultados obtidos pelo analista principiante não se diferencia daqueles obtidos pelo analista experiente.

Exemplo 11.2 *Um analista A encontrou no 1º dia, numa análise de uma solução de concentração conhecida, um desvio de 2,5 ppm de Cl; foram realizadas sete determinações. No 7º dia, o mesmo analista encontrou um desvio de 3,3 ppm de Cl em 10 determinações. Existe alguma variação nas determinações do analista A para um limite de confiança de 95%.*

Inicialmente vamos calcular o valor de F

$$F_{calc} = \frac{3,3^2}{2,5^2} = 1,74$$

Ao consultar os valores respectivos dos graus de liberdade 4 e 10 na Tabela 17 (pág. 113) encontramos o valor de $F = 4,10$. Como o valor calculado é menor que o valor tabelado, podemos afirmar que no nível de confiança de 95% os resultados obtidos pelo analista principiante não se diferencia daqueles obtidos pelo analista experiente.

Exercício 11.1 Um químico trainee de um famoso laboratório de análises químicas de ambientais determinou a quantidade um antibiótico como sua prova de emprego. Para verificar a qualidade do trainee, um químico experiente do laboratório também fez a mesma análise química. Com os dados dos dois químicos abaixo verifique se o químico trainee terá ou não o emprego. Critério para empregar o químico trainee é que seu resultado (variância) seja estatisticamente igual ao do químico experiente. Abaixo se encontram os resultados das análises:

$$\text{Trainee } \bar{X}=14,57 \text{ mgdL}^{-1} \quad dp=0,53 \text{ mgdL}^{-1} \quad n = 6$$

$$\text{Experiente } \bar{X}=13,95 \text{ mgdL}^{-1} \quad dp=0,42 \text{ mgdL}^{-1} \quad n = 5$$

Exercício 11.2 A qualidade de dois métodos estava precisando ser avaliada. Caso fossem iguais, o laboratório iria economizar cerca de 60%, representando uma redução de custo considerável. Com base nos resultados abaixo decida se suas variâncias são iguais. Para isso, utilize o teste de F.

$$\text{Método 1 } \bar{X}=35,25\% \quad dp=0,34\% \quad n = 5$$

$$\text{Método 2 } \bar{X}=35,35\% \quad dp=0,25\% \quad n = 5$$

Exercício 11.3 O valor aceito para o teor de sulfato de uma amostra padrão obtida de cinco análises químicas prévias é de 54,20% com um desvio padrão de 0,15%. Cinco análises da mesma amostra foram feitas por um novo procedimento instrumental, obtendo-se os seguintes valores, 54,01, 54,24, 54,05, 54,27 e 54,11%. Está este novo método produzindo resultados consistentes com o valor aceito.

Exercício 11.4 Catarino determinaram a quantidade de ferro em dois vinho branco por espectrometria de absorção atômica de chama. Eles encontraram que o vinho branco continha $1,10 \pm 0,02$ em três amostras analisadas. Para verificar se o resultado encontrado era aceitável ele obtiveram o resultados de 30 laboratórios credenciados ($1,12 \pm 0,20$). Verifique se o resultado da análise de ferro obtido por Catarino possui a mesma variância que os laboratórios credenciados.

Exercício 11.5 As fábricas de água mineral compram uma quantidade muito grande de rótulos e por isso a quantidade de um determinado íon é

sempre a mesma. Preocupado com esse fato, um químico resolveu verificar se as informações contidas no rótulo eram aceitáveis. No seu teste ele usou o íon bicarbonato, realizando diversas análises químicas de uma marca de água mineral. Ele obteve $18,97 \pm 0,10 \text{ mg L}^{-1}$ após 10 titulações seguidas. Para realizar o teste de F ele conseguiu saber que a informação do rótulo era $18,96 \pm 0,02 \text{ mg L}^{-1}$, média e desvio padrão de 6 amostras. Verifique se a informação do rótulo é ou não aceitável.

Exemplo 11.3 Um novo método colorimétrico foi desenvolvido a fim de substituir o método para a determinação de glicose conhecido como Folin-Wu. Para verificar a qualidade do novo método foram feitas uma série de medidas, sendo obtidos os seguintes resultados em mg dL^{-1} : **Novo método** = 127, 125, 123, 130, 131, 126 e 129, **Método Folin-Wu** = 130, 128, 131, 129, 127 e 125. Utilize o teste de F para avaliar se os dois métodos são significativamente diferentes ou não.

Usando o R

Sempre coloque os dados com maior número de variáveis

Primeiramente entre com o conjunto de dados:

```
> Novo <-c(127, 125, 123, 130, 131, 126, 129)
```

```
> Folin <-c(130, 128, 131, 129, 127, 125)
```

Para realizar o teste de F basta digitar:

```
> var.test(Novo, Folin)
```

No R o resultado será:

F test to compare two variances

data: novo and Folin

$F = 1.7653$, num df = 6, denom df = 5, p-value = 0.5494

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.2529925 10.5698854

sample estimates:

ratio of variances

1.765306

Os resultados fornecidos pelo R mostram que o F calculado é de 1,7653, tendo como graus de liberdade no numerador num $df = 6$ (relativos aos dados do método desenvolvido) e denominado denom $df = 5$. Ao verificar na tabela os valores de F teórico para os respectivos graus de liberdade nota-se que o F calculado é menor, portanto podemos afirmar que os dois métodos para um nível de significância de 95% os dois métodos são iguais. Outra forma de avaliar o resultado é pelo valor da probabilidade $p > 0,05$ ($p\text{-value} = 0,5494$) indicando que a possibilidade da hipótese que os dois métodos sejam iguais é verdadeira.

Exercício 11.6 A Tabela 10 mostra os resultados da análise de Ti (%) obtida de amostra de minérios de ferro por dois métodos distintos. Utilizando o R verifique se a variância dos dois métodos são significativamente iguais ou diferentes.

Tabela 10 – Resultados do teor (%) de titânio obtidos a partir de dois métodos diferentes em amostras de minério de ferro

Ordem	Método 1	Método 2
1	0,0134	0,0135
2	0,0144	0,0136
3	0,1266	0,0137
4	0,0111	0,0137
5	0,0137	0,0136

Exercício 11.7 O lítio é um metal utilizado em importantes áreas do conhecimentos, como medicina, geologia, astrofísica e química nuclear. A razão $6^Li/7^Li$ em material de referência foi utilizada para testar duas metodologias e os dados são apresentados na Tabela 11. Pede-se verificar com o R se as variâncias são iguais ou diferentes.

Exercício 11.8 Amostras ambientais de um rio foram coletadas em dois locais diferentes, montante e jusante a fim de verificar se uma fábrica estava contaminando o percurso de água. Como essa fábrica produz níquel, os químicos ambientais resolveram analisar este metal, considerado, potencialmente tóxicos e os resultados obtidos foram: montante 0,01, 0,02 e

Tabela 11 – Resultados da razão $6^L i / 7^L i$ em material de referência obtidos a partir de dois métodos diferentes

Ordem	Método 1	Método 2
1	0,082601	0,08183
2	0,082621	0,08186
3	0,082589	0,08205
4	0,082617	0,08206
5	0,082598	0,08215
6		0,08208

0,03 mg L⁻¹ e jusante 0,15, 0,20 e 0,25 mg L⁻¹. Usando o R verifique se a variância da montante e jusante podem ser consideradas semelhantes.

12 Teste de t

Este teste é usado para amostras pequenas com o propósito de comparar a média de uma série de resultados com um valor de referência. É também usado para testar a diferença entre as médias de dois conjuntos de resultados, \bar{x}_1 , \bar{x}_2 , por exemplo, o primeiro um método novo e segundo de referência.

O valor de t quando se comparam duas médias é dado por

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{1/n_1 + 1/n_2}} \quad (12.1)$$

em que s_p , o desvio padrão agrupado, que é calculado da seguinte forma:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad (12.2)$$

Os valores calculados são comparados com os tabelados (Tabela 18, pág. 114).

Exemplo 12.1 *Anarez trabalhando com uma série de amostras de plantas obteve diversos resultados relativos à determinação do boro por espectrofotometria e fluorimetria. Dentre os resultados obtidos foram escolhidos: o primeiro da análise espectrofotométrica cujo valor foi $28,00 \pm 0,30$ e o segundo fluorimétrico com média de $26,25 \pm 0,23$. Sabendo que cada média represente 10 medidas, avalie se as médias diferem entre si.*

Primeiramente, vamos calcular o valor de s_p

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{9 \cdot 0,30^2 + 9 \cdot 0,23^2}{16}} = 0,284$$

Agora podemos calcular t

$$t = \frac{28,00 - 26,25}{0,284 \sqrt{1/9 + 1/9}} = 13,07$$

Ao comparar o valor calculado para grau liberdade 18 com a Tabela 18 (pág. 114), cujo $t = 2,10$, nota-se uma diferença entre as médias com nível de significância de 5%.

Exercício 12.1 O teor de cálcio na urina de um pessoa foi determinado em dois dias diferentes, sendo obtidos os seguintes resultados (mgL^{-1}): 238 ± 8 ($n = 4$) e 255 ± 10 ($n = 2$). As médias são significativamente diferentes no nível de confiança de 95%?

Exemplo 12.2 Um novo método colorimétrico foi desenvolvido a fim de substituir o método o método de Folin-Wu, sendo obtidos os seguintes resultados em mg dL^{-1} : **Novo método** = 127, 125, 123, 130, 131, 126 e 129, **Método Folin-Wu** = 130, 128, 131, 129, 127 e 125. Utilize o teste de t para avaliar se as duas médias são significativamente diferente ou não.

Usando o R

Primeiramente entre com o conjunto de dados:

```
> Novo <- c(127, 125, 123, 130, 131, 126, 129)
```

```
> Folin <- c(130, 128, 131, 129, 127, 125)
```

Para realizar o teste de F basta digitar:

```
> var.test(Novo, Folin)
```

No R o resultado será:

Welch Two Sample t-test

data: novo and Folin $t = -0.7493$, $df = 10.86$, $p\text{-value} = 0.4696$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

```
-4.129653 2.034414
```

sample estimates:

ratio of x mean of y

```
127.2857 128.3333
```

Nota-se que o p -value se encontra dentro do intervalo de confiança de $-4,13$ a $2,03$, portanto a hipótese que as duas médias são iguais é aceitável dentro de um limite de confiança de 95%. Observação o valor negativo de t não deve ser considerado no julgamento da hipótese. Na realidade esse valor deve ser considerado em termos absolutos.

Exercício 12.2 As razões isotópicas do lítio são importantes em medicina, geologia, astrofísica e química nuclear. A razão ${}^8\text{Li}/{}^7\text{Li}$ em um material de referência padrão foi medida por dois métodos:

Método 1: 0,082601, 0,082621, 0,082589, 0,082617, 0,082598

Método 2: 0,082604, 0,082542, 0,082599, 0,082550, 0,082583, 0,082561

Os dois métodos produzem resultados estatisticamente equivalentes?

Exercício 12.3 A % (massa/massa) de Na_2CO_3 presente na soda cáustica foi determinada por volumetria de neutralização usando dois indicadores diferentes. Verifique usando o *R* se os dois indicadores possuem resultados equivalentes:

Indicador A: 36,82 37,04 36,93 37,01 36,20 37,00

Indicador B: 31,01 36,15 31,73 33,19 30,27 33,94

Exercício 12.4 Os dados mostrados na Tabela 12 foram obtidos durante um estudo da concentração de Zn em um lago da Amazônia. Utilizando o

Tabela 12 – Dados da concentração de um lago na Amazônia

Ponto de coleta	[Zn ²]	[Zn ²]
	interface ar-água	interface água-sedimento
1	0,430	0,415
2	0,266	0,238
3	0,567	0,390
4	0,531	0,410
5	0,707	0,605
6	0,716	0,609

teste de *t* pareado determine se existe diferença significativa entre a concentração de Zn² entre a interface com ar-água e sedimento e água. Nível de confiança de 95%.

Exercício 12.5 Um laboratório de uma empresa desenvolveu um método baseado no efeito catalítico sobre a oxidação do ácido sulfanílico pelo KIO_4 com meta a melhorar a determinação de Fe^{3+} de soro humano. Para verificar os seus resultados forma feitas oito determinações, juntamente com um método padrão. Os resultados em micromols/L é apresentado na Tabela 13.

Verifique se o Método proposto pode substituir o método padrão com nível de confiança de 95%.

Tabela 13 – Resultado da determinação de amostras de soro humano

Amostra	Método proposto	Método padrão
1	8,25	8,06
2	9,75	8,84
3	9,75	8,36
4	9,75	8,73
5	10,75	13,13
6	11,25	13,65
7	13,88	13,85
8	14,25	13,43

12.1 Teste de t para múltiplas amostras

Geralmente em laboratórios de análises químicas é necessário a criação de novos métodos de análises. Muitas vezes esses métodos consistem de mudanças relativamente pequenas ou mesmo de alterações bastante consideráveis. A eficiência depende normalmente da análise de vários tipos de amostras, sendo assim, vamos supor uma determinação de diferentes amostras de ferro pelos métodos A e B. A quantidade de ferro encontrada nos dois métodos são apresentadas na Tabela 14

Tabela 14 – Resultados das análises de ferro usando os métodos A e B.

	1	2	3	4	5
Método A	17,6	6,8	14,2	20,5	9,7
Método B	17,9	7,1	13,6	20,3	10,2

Para esses casos, a realização do teste de t simples não resolve o problema, além de serem necessários uma quantidade de cálculos relativamente grande.

Uma alternativa é utilizar o teste de t múltiplo em que a diferença entre cada medida pareada sobre cada amostra é computada. Neste caso, a media da diferença \bar{D} calculada e os desvios padrão individual usados

para o cálculo de s_d são calculados da seguinte forma:

$$t = \frac{\bar{D}}{s_d} \sqrt{N} \quad (12.3)$$

$$s_d = \sqrt{\frac{\sum(D_i - \bar{D})^2}{N - 1}} \quad (12.4)$$

em que D_i é a diferença entre os dois métodos em cada amostra utilizada e \bar{D} a média de todas as amostras.

Exemplo 12.3 *Imagine que você desenvolveu um novo método para determinar a quantidade de Fe presente nas águas ricas em matéria orgânica do rio Negro. Para ter certeza da capacidade do seu método você realizou uma série de análises química conjuntamente com a metodologia recomendada pelo Standard Methods. Os resultados referentes a cada amostra são apresentados abaixo:*

Amostra	Método Novo	Standard Methods
1	10,2	10,5
2	12,7	11,9
3	8,6	8,7
4	17,5	16,9
5	11,2	10,9
6	11,5	11,1

Com um limite de confiança de 95% verifique se o método desenvolvido por você equivale ao Standard Methods.

Primeiramente devemos calcular os valores de \bar{D} e o Σ para isso podemos fazer os seguintes cálculos:

<i>Am.</i>	<i>Novo</i>	<i>SD</i>	D_i	$D_i - \bar{D}$	$(D_i - \bar{D})^2$
1	10,2	10,5	-0,3	-0,6	0,36
2	12,7	11,9	0,8	0,5	0,25
3	8,6	8,7	-0,1	-0,4	0,16
4	17,5	16,9	0,6	0,3	0,09
5	11,2	10,9	0,3	0,0	0,00
6	11,5	11,1	0,4	0,1	0,01
Σ			1,7	Σ	0,87
$\bar{D} =$			0,28		

Am = amostra; *Novo* = método novo e *SD* = Standard Methods

Uma vez cálculos vamos determinar o *t*:

$$s_d = \sqrt{\frac{0,87}{6-1}} = 0,42$$

$$t = \frac{0,28}{0,42} \cdot \sqrt{6} = 1,63$$

Ao comparar o valor de *t* calculado (1,63) com o tabelado para um grau de liberdade de 5 ($t=2,57$, Tabela 18, pág. 114), nota-se com o $t_{calc} < t_{tabelado}$. Este resultado permite afirmar que não diferença significativa entre os dois métodos.

Exercício 12.6 Dois métodos analíticos diferentes foram usados para determinar cloro residual em efluentes de esgoto. Os dois métodos foram usados as mesmas amostras, coletadas em locais diferentes com tempo de contato de cloro em períodos diferentes. A concentração de cloro, expressa em $mg L^{-1}$, tiveram os seguintes dados:

<i>Amostra</i>	<i>Método A</i>	<i>Método B</i>
1	0,39	0,36
2	0,84	1,35
3	1,76	2,56
4	3,35	3,92
5	4,69	5,35
6	7,70	8,33
7	10,52	10,70
8	10,92	10,91

Faça o teste de t pelo método pareado para verificar se os dois métodos são iguais ou diferentes.

Exemplo 12.4 *Dois cristalógrafos desenvolveram um método para a determinação da % de caulinita em um mistura complexa de minerais. Para testar a nova metodologia, os cristalógrafos mediram nove amostras sintéticas de de caulinita com quantidades desconhecidas de outros minerais com o método novo e um estabelecido (padrão) na literatura. Os resultados foram os seguintes:*

Novo: 5,0 10,0 20,0 40,0 50,0 60,0 80,0 90,0 95,0

Padrão: 6,8 11,7 19,8 40,5 53,6 61,7 78,9 91,7 94,7

Utilize R para determinar se o método desenvolvido pelos cristalógrafo são equivalente, e portanto, pode ser substituir o método padrão.

Usando o R

```
Primeiramente entre com o conjunto de dados:  
> Padrao <-c(6.8, 11.7, 19.8, 40.5, 53.6, 61.7, 78.9, 91.7, 94.7)  
> Novo <-c(5, 10, 20,40,50,60,80,90, 95)  
  
j t.test(Novo,Padrao,paired=T)  
Paired t-test  
data: Novo and Padrao  
t = -2.1724, df = 8, p-value = 0.06158  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-2.1531073 0.0642184  
sample estimates:  
mean of the differences  
-1.044444
```

Nota-se que o p -value se encontra dentro do intervalo de confiança de - 2.15 a 0.064, portanto a hipótese que as duas médias são iguais é aceitável dentro de um limite de confiança de 95%. Observação o valor negativo de t não deve ser considerado no julgamento da hipótese. Na realidade esse valor deve ser considerado em termos absolutos.

Exercício 12.7 *A homogeneidade dos níveis de cloreto presente em uma amostra de água de um lago foi testada por meio de análises de porções retiradas do topo e do fundo da coluna de água, tendo apresentado os seguintes resultados, em mg L^{-1} de Cl^- :*

Topo	Fundo
26,30	26,22
26,43	26,32
26,28	26,20
26,19	26,11
26,49	26,42

1. *Aplique o teste de t em um nível de confiança de 95% para determinar se as médias são diferentes.*
2. *Agora use o teste de t pareado e determine se há diferença significativa entre os valores para o topo e fundo em um nível de confiança de 95%.*
3. *Por que se chega a diferentes conclusões quando se usa o teste de t pareado e quando apenas se combina os dados e se usa o teste de t para diferenças nas médias?*

Exercício 12.8 *Dois laboratórios diferentes fizeram uma série de análise de magnésio e gostariam de saber se seus resultados são equivalentes em um nível de confiança de 95%. O primeiro laboratório obteve os seguintes resultados em %: 3,92, 3,28, 4,18, 3,53 e 3,35 e o segundo 2,99, 2,87, 2,17, 3,40 e 3,92. Com o teste de t pareado verifique se os dois laboratórios produzem resultados equivalentes.*

Exercício 12.9 *Um novo método eletroquímico foi desenvolvido para determinar rapidamente a quantidade da monensina sódico, um antibiótico usado largamente para manipulação do metabolismo de bovinos, por um aluno de pós-graduação. Por outro lado, o método padrão para análise é baseado na atividade microbiológica o que leva muito tempo para obter os resultados. Como parte para da eficiência do novo método, amostras foram analisada em diferentes tempo, sendo o mesmo procedimento para o*

método padrão. Com os resultados apresentados na Tabela 15, verifique se o novo método pode substituir o método padrão. Para a sua avaliação use o *R*.

Tabela 15 – Resultados obtidos com os métodos microbiológico e eletroquímico

Amostra	Método microbiológico	Método eletroquímico
1	129,5	132,2
2	89,6	91,0
3	76,6	73,6
4	52,2	58,2
5	110,8	104,2
6	50,4	49,9
7	72,4	82,1
8	141,4	154,1
9	75,0	73,4
10	34,1	38,1
11	60,3	60,1

Exercício 12.10 *Dois cientistas teimosos resolveram mostrar um para o outro que eram capazes de estudar uma amostra oriunda de Marte. Para ter acesso ao artefato marciano, os cientistas tinham que analisar níquel em 10 amostras de planetas desconhecidos fornecida pela NASA. O primeiro cientista usou a ativação neutrônica para determinar o chumbo e o segundo usou o ICP-MS. Os resultados foram:*

Cientista 1: 34,0 22,6 29,7 24,9 17,8 14,8 8,6 31,4 10,8 17,3 15,7 18,4

Cientista 2: 33,4 28,4 29,5 24,8 18,3 15,4 8,5 31,9 11,5 17,6 15,4 18,3

*Pergunta-se: Usando o teste de *t* pareado no *R* verifique se os resultados dos dois cientistas são comparáveis.*

13 Rejeição de resultados

Durante a realização de uma série de análises químicas é comum o analista se deparar com o problema de aceitar ou não determinado resultados. Geralmente, essa situação ocorre durante uma titulação cujos os valores da bureta conduzem, quando o volume é muito diferente, ao pensamento de rejeitar ou não aquele resultado. Os alunos de modo geral preferem fazer outra titulação de forma a obter volumes próximos, no pensamento de que o seu resultado anterior estava errado.

Por outro lado, a rejeição dos resultados deve obedecer a um critério científico aceitável de forma que o conjunto dos dados esteja fazendo parte de uma mesma distribuição. Normalmente, é recomendado o teste de Q (ou Teste de Dixon) para a rejeição de resultados. O teste de Q rejeita valores críticos com um nível de confiança, sendo que sua aplicação deve ser feita da seguinte maneira:

1. Colocar os valores obtidos em ordem crescente;
2. Determinar a diferença existente entre o maior e o menor valor da série (faixa);
3. Determinar a diferença entre o menor valor da série e o resultado mais próximo;
4. Dividir esta diferença (em módulo) pela faixa, obtendo um valor de Q;
5. Se $Q > Q_{\text{tabelado}}$ (Tabela 19, pág. 114), o menor valor é rejeitado
6. Se o menor valor é rejeitado, determinar a faixa para os valores restantes e testar o maior valor da série;
7. Repetir o processo até que o menor e o maior valores sejam aceitos;
8. Se o menor valor é aceito, então o maior valor é testado e o processo é repetido até que o maior e o menor valores sejam aceitos;

9. Quando a série de medidas é constituída por três valores, aparentemente um valor será duvidoso, de modo que somente um teste precisa ser feito.

Exemplo 13.1 Um químico obteve os seguintes resultados durante suas análises químicas os seguintes resultados de cloreto (mgL^{-1}): 15,42, 15,51, 15,52, 15,53, 15,68, 15,52, 15,56, 15,53, 15,54, 15,56.

Primeiramente vamos ordenar os valores assim temos:

1.passo: 15,42, 15,50, 15,51, 15,52, 15,52, 15,53, 15,54, 15,56, 15,58 e 15,68.

2.passo: Diferença entre o maior e menor valor

$$\Delta_1 = 15,68 - 15,42 = 0,26$$

3.passo: Diferença entre o menor e o próximo da série

$$\Delta_2 = |15,42 - 15,50| = 0,08$$

4.passo: Dividir esta diferença (em módulo) pela faixa, obtendo um valor de Q

$$Q_{calc} = \frac{0,08}{0,26} = 0,308$$

5.passo: Comparando o valor calculado com Q_{tab} de 2,23 para um n de 6 observa-se que o valor calculado é menor, portanto, o valor menor próximo não pode ser rejeitado.

Agora vamos testar o maior valor, repetindo o 3. passo para o próximo mais alto. Assim

3.passo: Cálculo da diferença

$$\Delta = 15,68 - 15,58 = 0,10$$

4.passo: Divisão da diferença

$$Q_{calc} = \frac{0,10}{0,26} = 0,38$$

5.passo: Comparando com Q_{tab} verifica-se novamente que o maior valor também não pode ser rejeitado.

Podemos resolver a questão da rejeição de dados facilmente no R. Para podemos utilizar dois testes: teste de Dixon ou de Grubb. Inicialmente, temos que instalar um pacote no R chamado *Outliers*. Então basta digitar no R:

```
>install.packages("outliers")
```

Alternativamente, você pode utilizar a opção Packages no R e procurar o pacote outliers. Em ambas as opções é necessário escolher um local do espelho (mirror). Como opção escolho sempre o da Universidade Federal do Paraná. Vamos usar o R para resolver o problema do exemplo 13.1

Exemplo 13.2 *Solução no R para o exemplo 13.1*

Primero passo vamos entrar com os dados no R:

```
>dadosj-c(15.42, 15.51, 15.52, 15.53, 15.68, 15.52, 15.56, 15.53, 15.54, 15.56)
```

Em seguida digitar

```
> dixon.test(dados,type=10,two.sided=TRUE)
```

Resposta do programa

Dixon test for outliers

data: dados

$Q = 0.4615$, $p\text{-value} = 0.05297$

alternative hypothesis: highest value 15.68 is an outlier

Como podemos observar o valor de $Q = 0,46$ com $p\text{-value} = 0,053$ sugerem a rejeição do valor maior de 15,68. Esse resultado é o mesmo sugerido acima.

Exemplo 13.3 *Aplique o teste de Dixon na sequência de resultados abaixo para verificar se existe algum número anômalo que merece ser desprezado.*

Conc. mol L^{-1} : 0,10 0,11 0,12 0,11 0,10

Teste de Dixon no R dos dados

```
> dados<-c(0.10, 0.11, 0.12, 0.11, 0.10)
```

```
> dixon.test(dados,type=10,two.sided=TRUE)
```

Dixon test for outliers

data: dados

$Q = 0.5$, $p\text{-value} = 0.2972$

alternative hypothesis: highest value 0.12 is an outlier

Observa-se que o valor $Q = 0,5$ recomenda a não rejeição dos dados, diferente do Exemplo 13.2.

O intervalo de Q entre 1,0 e 0,60 o teste de Dixon recomenda aceitar o valor sugerido como um outlier. Abaixo de 0,60 é recomendado rejeitar o valor sugerido de outileir. Ao contrário da forma feita manualmente, o próprio R sugere se é o valor maior ou menor que deve ser rejeitado.

Exercício 13.1 *Aplique o teste de Q aos conjuntos de dados abaixo para verificar que dado(s) referentes à análise de ferro em minério poder(em) ser desprezados a um nível de confiança de 95%.*

1. 85,10, 84,62, 84,70

2. 85,10, 84,62, 84,65, 84,70

Exercício 13.2 *Quatro valores de replicatas foram obtidas para a determinação de pesticida em um rio:*

$0,403, 0,410, 0,401, 0,380 \mu\text{g L}^{-1}$

Verifique a possibilidade de rejeitar o valor de $0,380 \mu\text{g L}^{-1}$.

Exercício 13.3 *A análise de uma amostra de calcita gerou percentagens de CaO de 55,95, 56,00, 56,04, 56,08 e 56,23. Utilize o R para verificar se existe algum resultado anômolo que merece ser rejeitado.*

Exercício 13.4 *aplique o teste Q (Dixon) aos conjuntos de dados que seguem para determinar se existem resultados anômolos.*

1. 41,27 41,61 41,84 41,70

2. 7,295 7,284 7,388 7,292

14 Boxplot

Em 1977, John Tukey publicou uma proposta que posteriormente foi reconhecida como sendo eficiente método para mostrar cinco número que resumizam qualquer conjunto de dados. O gráfico proposto é chamada de boxplot (Figura 10). Esse gráfico possibilita representar a distribuição de um conjunto de dados com base em alguns de seus parâmetros descritivos, quais sejam: a mediana (Q_2), o quartil inferior (Q_1), o quartil superior (Q_3) e do intervalo interquartil ($IQR = Q_3 - Q_1$).

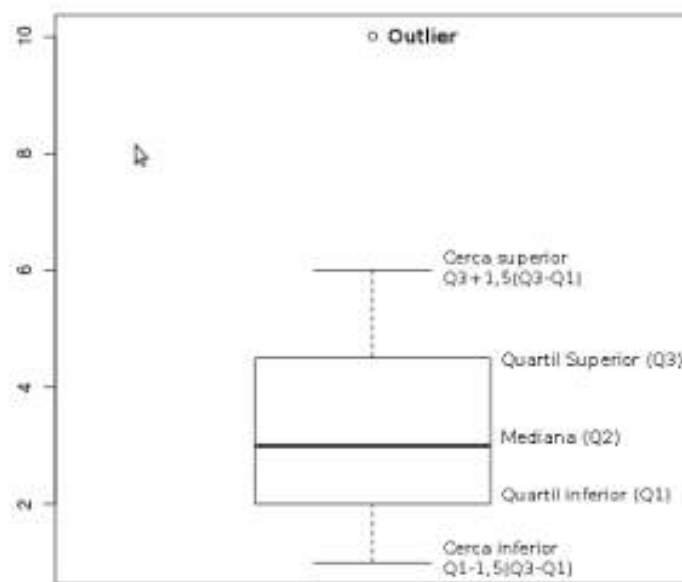


Figura 10 – Boxplot

A linha central da caixa marca a mediana do conjunto de dados. A parte inferior da caixa é delimitada pelo quartil inferior (Q_1) e a parte superior pelo quartil superior (Q_3). As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior a $Q_1 - 1,5(Q_3 - Q_1)$ e do quartil superior até o maior valor não superior a $Q_3 + 1,5(Q_3 - Q_1)$. Os pontos inferiores e superiores fora da caixa são representados individualmente no gráfico, sendo estes valores caracterizados como outliers.

O boxplot permite avaliar a simetria dos dados, sua dispersão e a

existência ou não de outliers nos mesmos, sendo especialmente adequado para a comparação de dois ou mais conjuntos de dados correspondentes às categorias de uma variável qualitativa.

Exemplo 14.1 *Construa no R o boxplot usando os seguintes dados que forma obtidos de análises químicas de cálcio usando quatro métodos diferentes. Os resultados são expresso em %.*

A	B	C	D
0,10	0,20	0,30	0,30
0,20	0,30	0,40	0,60
0,15	0,25	0,25	0,70
0,25	0,25	0,10	0,35
0,35	0,20	0,35	0,10
0,20	0,40	0,20	1,50

Usando o R

Primeiramente entre com o conjunto de dados:

```
> A <-c(0.10,0.20,0.15,0.25,0.35,0.20)
```

```
> B <-c(0.20,0.30,0.25,0.25,0.20,0.40)
```

```
> C <-c(0.30,0.40,0.25,0.10,0.35,0.20)
```

```
> D <-c(0.30,0.60,0.70,0.35,0.10,1.50)
```

Para realizar o teste de F basta digitar:

```
> boxplot(A,B,C,D,xlab="Método",ylab="Concentração (%)")
```

No final do processo o R irá desenhar o gráfico da Figura 11

Exercício 14.1 *Construa no R um boxplot com os dados:*

Método	1	2	3	4	5
A	11,08	11,19	11,17	11,50	11,14
B	11,35	10,76	10,63	11,13	
C	10,75	11,33	11,14	11,55	11,30
D	11,11	11,04	11,43	10,98	

Exercício 14.2 *Em um trabalho de química ambiental sobre a análise de pH de amostras em três locais diferentes de um igarapé contaminado. Os resultados do estudo são apresentados abaixo.*

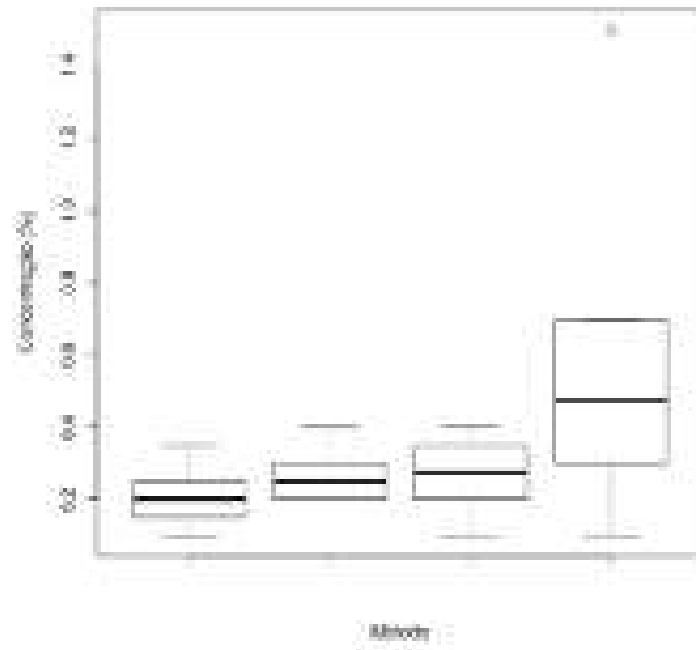


Figura 11 – Boxplot relativo ao exemplo 14.1

<i>Ordem</i>	<i>pH1</i>	<i>pH2</i>	<i>pH3</i>
1	6,64	4,57	7,31
2	6,54	6,62	7,27
3	6,56	6,66	7,15
4	6,44	6,62	7,07
5	6,45	6,43	6,98
6	6,41	6,74	7,02

Faça um gráfico de boxplot e discuta os resultados em termos de variação do valor pH.

15 Calibração

Experimentos envolvendo a análise espectrofotométrica quantitativa de amostras com muitos componentes cujos espectros estejam superpostos são bastante importantes em disciplinas de química analítica, sejam elas básicas ou mais avançadas. Em geral, as concentrações dos compostos de interesse numa amostra são determinadas usando um sistema de equações simultâneas obtido pela lei de Beer em tantos comprimentos de onda quantos forem os analitos.

O processo de calibração normalmente consiste em duas etapas. Na primeira, a etapa descritiva, medidas feitas numa série de padrões analíticos de concentrações conhecidas são usadas para construir um modelo que relacione a grandeza da medida com a concentração da espécie de interesse. Na segunda etapa, conhecida como preditiva, usa-se o modelo para prever concentrações de novas amostras, a partir dos sinais analíticos medidos para elas.

Curvas de calibração são construídas em cada sinal (por exemplo, comprimento de onda) a partir de soluções padrão de cada analito a fim de estabelecer constantes de proporcionalidade individuais entre concentração e intensidades de absorção. No caso de misturas binárias simples, muitas vezes obtemos bons resultados por este método.

Entretanto, quando se passa para amostras reais, podem surgir problemas devido a interferências espectrais e desconhecimento da real identidade dos compostos de interesse. Nessas situações, a resolução simultânea das equações já não fornece resultados precisos e por isso foram desenvolvidos nos métodos para resolver o problema. Aprenderemos os dois modos de realizar uma calibração.

15.1 Comparação com padrões

Basicamente existem dois tipos de métodos de comparação: a direta e o procedimento titulométrico.

15.1.1 Comparação direta

Alguns procedimentos analíticos envolvem a comparação de uma propriedade do analito (ou do produto de uma reação com o analito) com padrões, de forma que a propriedade que está sendo avaliada se iguale ou quase se iguale àquela do padrão. Por exemplo, nos colorímetros antigos, a cor resultante de uma reação química do analito era comparada com a cor produzida pela reação de padrões. Se a concentração do padrão era variada por diluição, por exemplo, era possível se obter uma igualdade de cor bastante exata. Então, a concentração do analito era igual à concentração do padrão após a diluição.

15.1.2 Titulações

As titulações estão entre os procedimentos analíticos mais exatos. Em uma titulação, o analito reage com um reagente (o titulante), que seja padronizado ou padrão primário, em uma reação de estequiometria conhecida. Geralmente, a quantidade de titulante é variada até que a equivalência química seja encontrada, sendo detectada por meio da mudança de cor de um indicador químico ou por meio de alteração da resposta instrumental. A quantidade de reagente padronizado necessário para se obter a equivalência química pode ser então relacionada à quantidade de analito presente. A titulação é, desta forma, um tipo de equiparação química.

Exemplo 15.1 *Um estudante realizou uma titulação entre o hidróxido de sódio padronizado com biftalato de potássio, um padrão primário, e o vinagre tendo como indicador a fenolftaleína.*

Verifica-se nesta análise química clássica o processo de comparação entre o hidróxido de sódio padronizado e o vinagre. Sem o estabelecimento da concentração de hidróxido de sódio é impossível afirmar qual é a possível concentração do ácido acético no vinagre.

15.1.3 Calibração com padrões externos

Um *padrão externo* é preparado separadamente da amostra. Padrões externos são empregados para se calibrar instrumentos e procedimentos

quando não há efeitos interferentes advindos dos componentes matriz presentes na solução do analito. Uma série de padrões externos contendo o analito em concentrações conhecidas é preparada. De forma ideal, três ou mais soluções são utilizadas em um processo de calibração.

A calibração é realizada pela obtenção do sinal de resposta (absorbância, altura do pico, área do pico/banda, etc.) como uma função da concentração conhecida do analito. Uma curva de calibração (ou curva analítica) é preparada construindo-se um gráfico a partir dos dados ou ajustando-os a uma equação matemática adequada, como aquela empregada no método dos mínimos quadrados.

$$y = a + bx \text{ ou } \textit{sinal} = a + b \cdot \textit{conc} \quad (15.1)$$

O próximo passo é a etapa de previsão, em que o sinal obtido da amostra é usado para encontrar a concentração do analito:

$$\textit{conc} = \frac{1}{b}(\textit{sinal} - a) \quad (15.2)$$

A qualidade da determinação da concentração de uma amostra desconhecida está condicionada à relação entre os pontos experimentais obtidos pela medida de concentrações conhecidas com o sinal do aparelho.²²²²

15.1.4 Método dos Mínimos Quadrados ou Quadrados Mínimos Ordinários

Uma curva de calibração típica é mostrada na Figura 12 para a determinação de uma amostra arbitrária. Inicialmente, vamos considerar que o ajuste dos dados colocados no gráfico seja uma função linear representados pela equação da reta:

$$y = a + bx \quad (15.3)$$

Os mínimos quadrados se refere ao valor residual obtido pela diferença vertical (distância) entre dados experimentais e o modelo teórico (geralmente representado por uma linha que passa na média dos pontos experimentais). A obtenção do modelo teórico tem como base a distribuição normal dos pontos que compõem a curva de calibração. A Figura

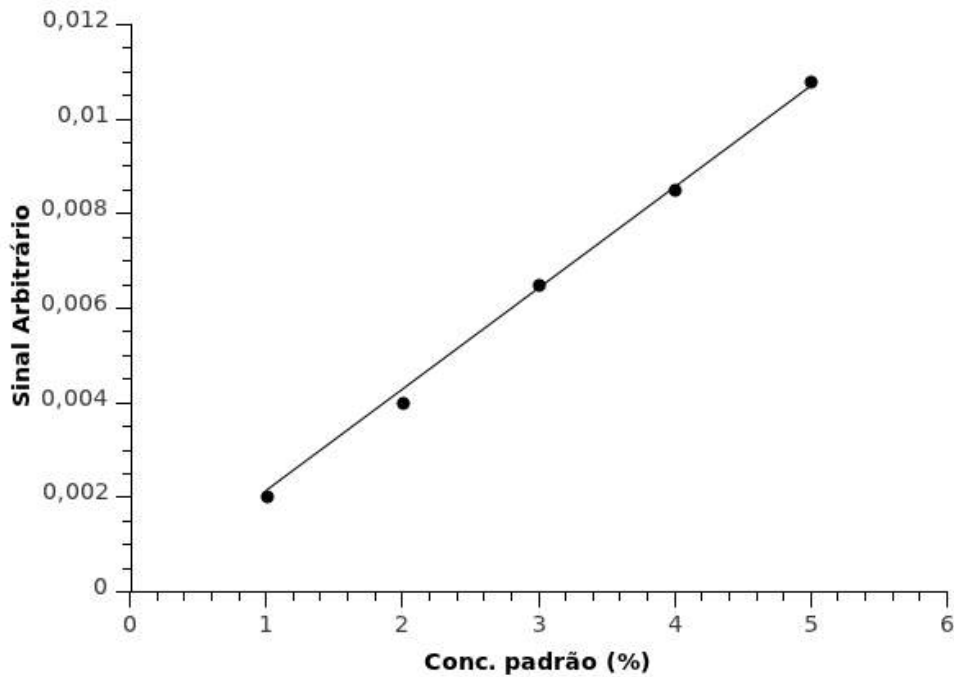


Figura 12 – Curva de calibração típica

13 mostra como seria os pontos médios dentro de uma série de distribuições normais numa curva de calibração.

Cada distância, d , é determinada da seguinte forma:

$$d = y - \hat{y} \quad (15.4)$$

em que \hat{y} é o valor previsto pelo modelo teórico, sendo encontrado da forma:

$$\hat{y} = a + bx \quad (15.5)$$

Substituindo a equação 15.4 na equação 15.5 temos

$$d = y - a - bx \quad (15.6)$$

Em termos globais tem-se

$$\sum d^2 = \sum (y - a - bx)^2 \quad (15.7)$$

Ao término de uma regressão linear tem-se a curva mostrada na Figura 14. Nota-se o aparecimento de um resíduo, cujo valor está diretamente relacionado ao processo de minimização da linha teórica e os pontos

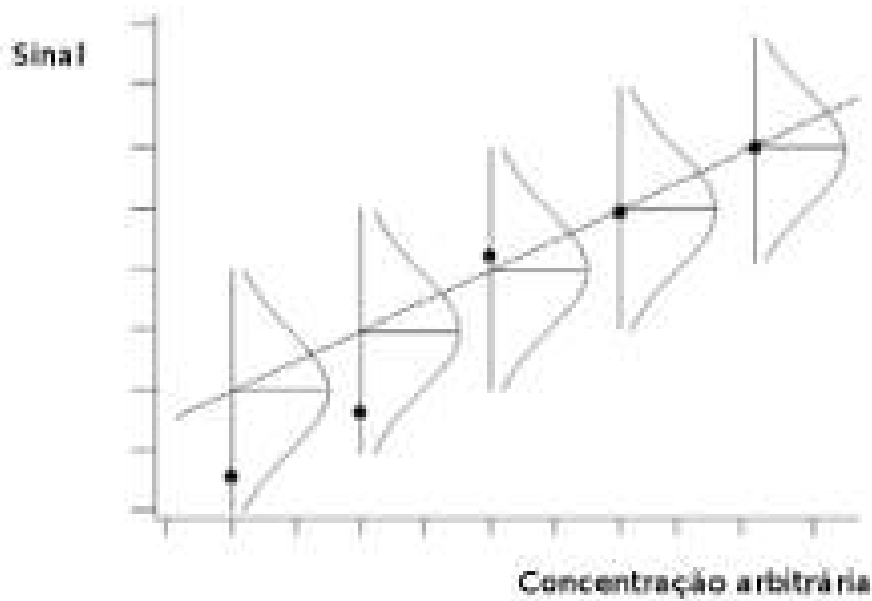


Figura 13 – Representação dos pontos de uma curva de calibração com as respectivas distribuições normais. Adaptada de Thompson

experimentais. O processo de minimização é obtido por

$$SS_{residual} = \sum_{i=1}^N [y_i - (a + bx_i)]^2 \quad (15.8)$$

N é o número de pontos utilizados na construção da curva de calibração.

Portanto, a equação da reta na curva de calibração é dada por

$$y = a + bx + erro \quad (15.9)$$

15.2 Curva de calibração no R

O método dos mínimos quadrados, ou Mínimos Quadrados é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas resíduos). Um requisito para o método mínimos quadrados é que o fator imprevisível (erro) seja distribuído aleatoriamente, essa distribuição

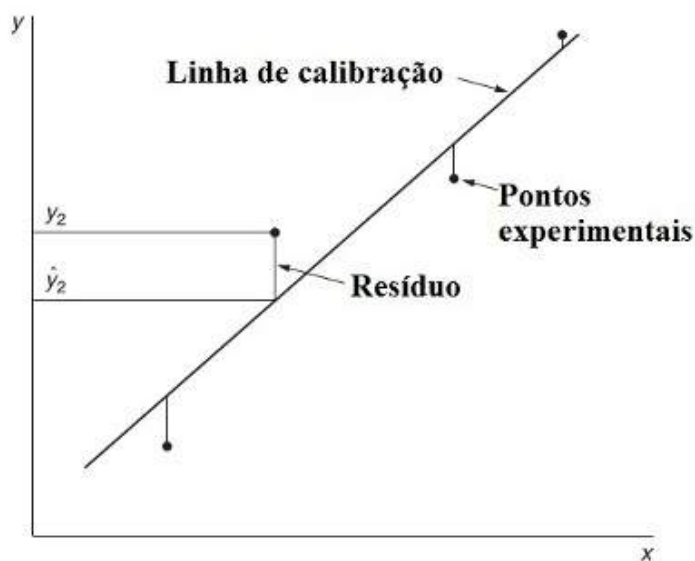


Figura 14 – Representação de uma curva de calibração típica após a realização de uma regressão linear.

seja normal e independente. Outro requisito é que o modelo seja linear nos parâmetros; ou seja, as variáveis apresentam uma relação linear entre si.

Inicialmente devemos estabelecer um critério para a obtenção da equação $y = ax + b$, para a obtenção dos coeficientes a e b . Os desvios individuais correspondem à diferença entre a medida real da medida experimental e a medida aproximada pela reta de regressão. O somatório dos quadrados dos desvios é mostrado abaixo:

$$\sum_{i=1}^n d_i^2 = \sum_{i=1}^n (y_i - y_c)^2 \quad (15.10)$$

y_i é o valor observado de y e y_c valor calculado de y na função linear.

Os coeficientes da reta são calculados pelas fórmulas abaixo:

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (15.11)$$

$$b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \quad (15.12)$$

A equação linear pode ser obtida no R por meio da função $lm()$ que serve para calcular a regressão linear simples.

Exemplo 15.2 Os dados na Tabela 16 foram obtidos durante uma determinação colorimétrica de glicose em soro sanguíneo. Considerando uma

relação linear, encontre as estimativas por mínimos quadrados da inclinação e do intercepto.

Tabela 16 – curva de calibração da determinação colorimétrica da glicose em soro sanguíneo

Conc. glicose (mmol L^{-1})	Absorbância, A
0,0	0,002
2,0	0,150
4,0	0,294
6,0	0,434
8,0	0,570
10,0	0,704

O primeiro passo para construir a curva de calibração no R a primeira coisa a ser feita é a entrada dos dados:

```
> x <- c(0,2,4,6,8,10)
```

```
> y <- c(0.002,0.150,0.294,0.434,0.570,0.704)
```

Vamos criar um data.frame

```
> dados=data.frame(x,y)
```

Verifique se o data.frame foi criado com o comando

```
> is.data.frame(dados) se aparece o resultado
```

```
[1] TRUE
```

significa que o data.frame foi criado.

Agora vamos fazer a regressão linear. Para isso, basta digitar

```
> regressao=lm(y ~ x, data=dados)
```

Ao digitar essa expressão o comando lm() irá atribuir à regressão os resultados da regressão linear.

Digitando agora:

```
> regressao
```

Teremos como resposta

Call:

```
lm(formula = y ~ x, data = dados)
```

Coefficients:

(Intercept) x

0.008286 0.070143

ou seja

$$y = 0,008286 + 0,070143x$$

Agora ao digitar

```
> summary(regressao)
```

Obteremos como resposta dados mais completos a respeito da regressão linear

Call:

```
lm(formula = y ~ x, data = dados)
```

Residuals:

	1	2	3	4	5	6
	-0.0062857	0.0014286	0.0051429	0.0048571	0.0005714	-0.0057143

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.008286	0.004039	2.051	0.11
x	0.070143	0.000667	105.161	4.9e-08 ***

— Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.005581 on 4 degrees of freedom

Multiple R-squared: 0.9996, Adjusted R-squared: 0.9995

F-statistic: 1.106e+04 on 1 and 4 DF, p-value: 4.903e-08

Os dados completos da regressão mostram i) os "Resíduos" que fornece a diferença entre os sinais experimentais e previstos; ii) As estimativas dos coeficientes do modelo juntamente com os desvios padrão; iii) valor de t e iv) probabilidade para a hipótese nula de que os coeficientes têm valores de zero. No nosso exemplo, verifica-se que a interceptação e inclinação são diferentes. Na parte inferior encontramos o desvio padrão sobre a regressão, o coeficiente de correlação e um resultado de teste F da hipótese nula de que a MS_{reg} / MS_{res} é 1.

Em uma regressão linear é necessário verificar suas tendências para isso basta digitar:

```
> layout(matrix(1:4,2,,2))
```

```
> plot(regressao)
```

A Figura 15 mostra a tendência da regressão linear. É possível observar que a esquerda um gráfico dos erros residuais plotados versus seus valores ajustados. Para ter uma boa curva de calibração os resíduos

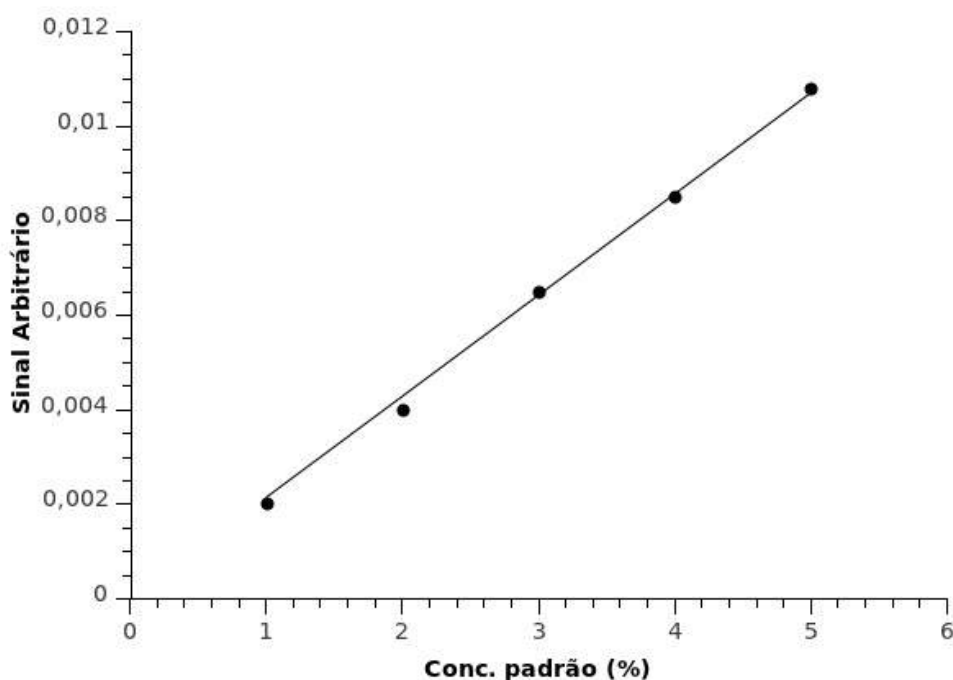


Figura 15 – Tendências observadas na curva de calibração

devem ser distribuídos aleatoriamente em torno da linha horizontal que representa um erro residual de zero, isto é, não deve haver uma tendência nítida da distribuição de pontos.

Na parte esquerda inferior da Figura 15 é um gráfico QQ padrão, que deve sugerir que os erros residuais são normalmente distribuídos. Já parte superior a direita é mostrado um gráfico da contendo a raiz quadrada dos resíduos padronizados (uma espécie de raiz quadrada do erro relativo) como uma função dos valores ajustados. Mais uma vez, não deve haver qualquer tendência evidente neste gráfico. Por fim, na parte inferior a direita mostra uma medida da importância na determinação do resultado de regressão. É possível observar nesse gráfico linhas de contorno para a distância de Cook, que é outra medida da importância de cada observação para a regressão. Distâncias menores significa que a remoção da observação tem pouco efeito sobre os resultados da regressão. Distâncias maiores que 1 sugerem a presença de possível outlier possível ou um modelo pobre.

Além disso, o Multiple R-squared e Adjusted R-squared são bastante similar indicando que a equação da reta encontrada pelo método

dos mínimos quadrados pode ser utilizada tranquilamente para encontrar a concentração de amostras desconhecidas de soro sanguíneo.

Exercício 15.1 A forma reduzida da nicotinamida adenina dinucleotídeo (NADH) é uma coenzima importante e altamente fluorescente. Apresenta uma absorção máxima a 340 nm. As soluções padrão de NADH forneceram as seguintes intensidades de fluorescência:

Conc. padrão NADH $\mu\text{mol L}^{-1}$	Intensidade relativa
0,100	2,24
0,200	4,52
0,300	6,63
0,400	9,01
0,500	10,94
0,600	13,71
0,700	15,49
0,800	17,91

Encontre a inclinação e o intercepto da curva de calibração. Faça uma avaliação completa da curva de calibração no R.

Exercício 15.2 Amostras de água de igarapé foram analisadas por ICP-MS em um estudo multielementar. O vanádio foi um dos elementos determinados e sua curva de calibração se encontra abaixo:

Conc. V (pg mL^{-1})	Intensidade (unidades arbitrárias)
0,00	2,1
2,00	5,0
4,00	9,2
6,00	12,5
8,00	17,4
10,00	20,9
12,00	24,7

Determine a equação da reta dessa curva de calibração. Encontre os desvios padrão da inclinação e do intercepto. Faça uma avaliação completa da curva de calibração no R

Exercício 15.3 Verifique se a curva de calibração abaixo pode ser usada para a determinação do analíto de sódio.

C_s (mol L ⁻¹)	Sinal
0,000	0,00
0,100	12,36
0,200	24,83
0,300	35,91
0,400	48,79
0,500	60,42

Exercício 15.4 A sequência de dados abaixo foi obtida de uma provável curva de calibração de Cd²⁺:

[Cd ²⁺] (mol L ⁻¹)	15,4	30,4	44,9	59,0	72,7	86,0
Absorbância	4,8	11,4	18,2	26,6	32,3	37,7

Construa uma curva de calibração para os dados acima e verifique a possibilidade de sua utilização em uma análise de Cd²⁺

Exercício 15.5 Em um laboratório de eletroanalítica foram feitos vários testes para o desenvolvimento de uma metodologia para analisar chumbo em amostras de ar. Os resultados abaixo é uma das curvas de calibração obtida por essa metodologia:

Alta do pico	1,4	3,8	7,5	10,2	11,7	15,0
Conc. Pb ng g ⁻¹	2,4	2,5	1,3	1,3	0,7	1,2

Utilizando a regressão linear do R avalie a qualidade da curva de calibração da metodologia desenvolvida para a determinação de chumbo.

Exercício 15.6 O mesmo laboratório do exercício 15.5 resolveu analisar a mesma amostra, só que desta vez, usando a espectrometria do ultravioleta-vísivel (Uv-vis). Os resultados da curva de calibração são apresentados abaixo:

Absorbância	0,05	0,17	0,32	0,60	1,07	1,40
Conc. Pb ng g ⁻¹	10	25	50	100	200	300

Utilizando a regressão linear do R avalie a qualidade da curva de calibração da espectrometria do Uv-vis. Compare os seus resultados com aqueles obtidos no exercício 15.5.

15.2.1 Usos da curva de calibração

Dentre os usos da curva de calibração, o mais utilizado é a determinação da concentração, dita anteriormente e expressa pela equação 15.2 (pág. 89). Vamos considerar alguns exemplos para reforçar o uso da curva de calibração na determinação da concentração.

Exemplo 15.3 *O manganês foi determinado em amostras de interesse geológico por fluorescência de raios X. A curva de calibração apresentou os seguintes valores:*

% padrão Mn	0,00	0,10	0,20	0,30	0,40
Intensidade pico Mn	80	106	129	154	167

Pergunta-se:

1. *Faça a regressão linear e estabeleça a qualidade da curva de calibração.*
2. *Calcule a concentração de Mn presente numa amostra cuja intensidade medida na fluorescência de raios X foi de 145.*

Primeiramente temos que encontrar a equação da reta da curva de calibração. Para isso, temos que utilizar o mesmo procedimento anterior; ou seja, a função `lm()`. Vamos lá:

Entrada de dados:

```
> Conc<-c(0, 0.1,0.2,0.3,0.4)
```

```
> Int<-c(80, 106, 129, 154, 167)
```

Regressão

```
> dados=data.frame(Int,Conc)
```

```
> is.data.frame(dados)
```

```
[1] TRUE
```

```
> regressao=lm(Int Conc,dados)
```

```
> summary(regressao)
```

Call:

```
lm(formula = int padrao, data = dados)
```

Residuals:

1	2	3	4	5
-2.8	1.0	1.8	4.6	-4.6

Coefficients:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(> t)</i>	
<i>(Intercept)</i>	82.800	3.298	25.10	0.000139	** —
<i>padrao</i>	222.000	13.466	16.49	0.000486	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 0.01908 on 3 degrees of freedom

Multiple R-squared: 0.9891, Adjusted R-squared: 0.9854

F-statistic: 271.8 on 1 and 3 DF, p-value: 0.0004857

Em princípio, os resultados da regressão linear mostram que tem uma curva de calibração com valores bastante satisfatórios. Caso tenha dúvida basta digitar:

```

> layout(matrix(1:4,2,,2))
> plot(regressao) l e/ou
> layout(matrix(1:1,1,,1))
> plot(padrao int,dados)
> abline(regressao)

```

A curva de calibração é apresentada na Figura 16. Nota-se claramente que a interseção da curva de calibração. A equação da reta dessa curva de calibração é

$$int = 222conc + 82,8$$

Portanto, a concentração de Mn pode ser calculada da seguinte forma:

$$[Mn] = \frac{Intensidade + 0,367}{0,00446}$$

Como a intensidade é 145 a concentração de Mn:

$$[Mn] = \frac{145 - 82,8}{222} = 0,28$$

Exercício 15.7 *Em uma análise usual de proteína, um corante se liga em uma proteína e a cor do corante muda de marrom para azul. A intensidade da cor azul é proporcional à quantidade de proteína presente. Com os dados*

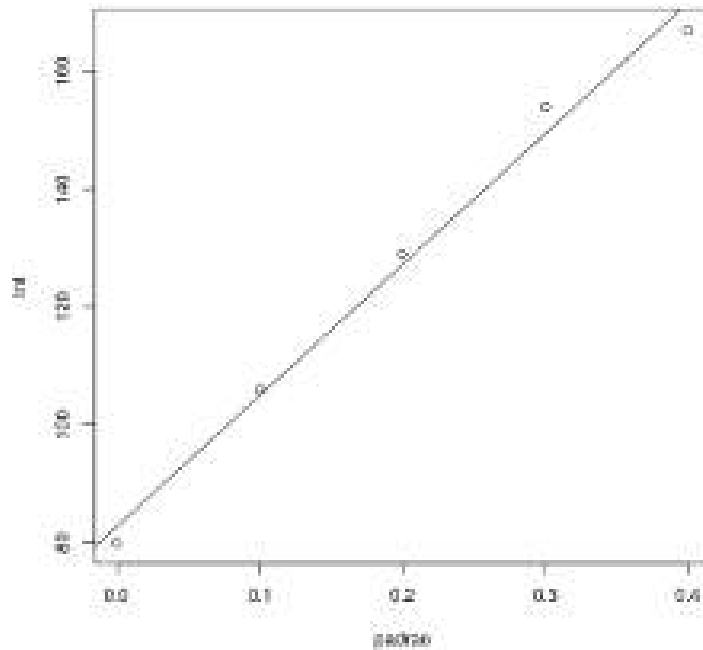


Figura 16 – Curva de calibração do exemplo 15.3

da curva de calibração abaixo, determine a concentração de uma amostra de proteína cuja absorbância foi de 0,973. Expresse seu valor em micrograma de proteína.

Proteína (μg)	0,00	9,36	18,72	28,08	37,44
Absorbância	0,466	0,676	0,883	1,086	1,280

Exercício 15.8 Em uma análise instrumental foram obtidos os seguintes resultados:

Conc.	0	1	2	3	4	5	6	7
Sinal	0,2	3,6	7,5	11,5	15,0	17,0	20,4	22,7

Faça a curva de calibração e encontre a concentração correspondente aos sinais 5,16 e 16 unidades.

Existe uma concordância na literatura de que a sensibilidade de um instrumento ou de um método é uma medida de sua habilidade em discriminar entre pequenas diferenças na concentração de um analito. Dois fatores limitam a sensibilidade:

1. A inclinação da curva de calibração;

2. A reprodutibilidade ou precisão do dispositivo de medida

Para dois métodos que tenham a mesma precisão, aquele que tem a curva de calibração mais inclinada será o mais sensível. A Figura 17 mostra o efeito da inclinação sobre a sensibilidade do método; ou seja, observa-se uma mudança para alguns casos significativa dependendo da curva de calibração. Nota-se que a inclinação só é uma forma de mostrar que a sensibilidade de uma curva de calibração é diferente de outra. O uso da inclinação falha como uma figura de mérito para considerar a precisão das medidas individuais em uma curva de calibração.

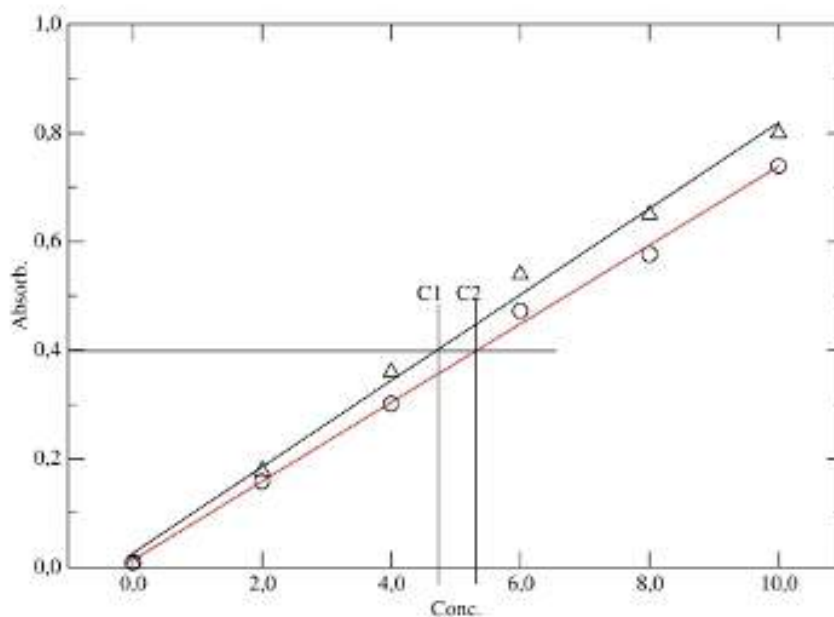


Figura 17 – Efeito da inclinação sobre a sensibilidade do método em uma Curva de calibração

O que é sensibilidade? A sensibilidade de um instrumento ou método é uma medida da sua habilidade em discriminar pequenas diferenças na concentração do analito.

Como uma curva de calibração não "fixa"; ou seja, o erro de uma regressão reflete na sensibilidade. Se os sinais de um instrumento analítico são referentes aos valores de concentração mais baixo a sensibilidade será diferente dos mais altos. Mandel e colaboradores reconheceram a necessi-

dade de incluir a precisão em uma definição matemática de sensibilidade, para isso foi proposta a seguinte definição de sensibilidade analítica, γ :

$$\gamma = \frac{\text{coeficiente angular}}{dp \text{ das medidas}} \quad (15.13)$$

A sensibilidade analítica oferece a vantagem de ser relativamente imune aos fatores de amplificação. Por exemplo, o aumento do ganho de um instrumento por um fator cinco produz um aumento de cinco vezes no valor do coeficiente angular. Normalmente, esse aumento será acompanhado por um aumento correspondente no desvio padrão da medida. Apesar da nova situação a sensibilidade permanece praticamente a mesma.

Exemplo 15.4 *Uma série de medidas de fluorescência foram realizadas para a determinação de soluções ácidas de quininas. Abaixo está apresentada os dados da curva de calibração:*

<i>Conc. ng mL⁻¹</i>	0	10	20	30	40	50
<i>Intensidade</i>	4	22	44	60	75	104

Após a construção da curva foram feitas três medidas referente a uma triplicata, cujos valores das intensidades foram 44, 46 e 47. Diante dessas informações, utilize o R para determinar a sensibilidade do métodos usado para fazer as medidas de quininas.

```
Primeramente - Realizar regressão linear no R > conc<-c(0, 10,20,30,40,50)
```

```
> int<-c(4,22,44,60,75,104)
```

```
> dados=data.frame(int,conc)
```

```
> is.data.frame(dados)
```

```
[1] TRUE
```

```
> regressao
```

```
Call:
```

```
lm(formula = int ~ conc, data = dados)
```

```
Coefficients:
```

<i>(Intercept)</i>	<i>conc</i>
<i>3.286</i>	<i>1.929</i>

```
> summary(regressao)
```

Call:

```
lm(formula = int conc, data = dados)
```

Residuals:

```
1      2      3      4      5      6
0.7143 -0.5714 2.1429 -1.1429 -5.4286 4.2857
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.28571	2.67325	1.229	0.286
conc	1.92857	0.08829	21.842	2.6e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.694 on 4 degrees of freedom

Multiple R-squared: 0.9917, Adjusted R-squared: 0.9896

F-statistic: 477.1 on 1 and 4 DF, p-value: 2.6e-05

```
> plot(conc,int)
```

```
> abline(regressao)
```

```
> layout(matrix(1:4,2,,2))
```

```
> plot(regressao)
```

Depois de toda essa avaliação vamos calcular a sensibilidade para as medidas 44, 46 e 47 na curva de calibração. Antes do cálculo propriamente dito temos que calcular o desvio padrão:

```
> medj-c(44,46,47)
```

```
> sd(med) [1] 1.527525
```

Agora temos condições de calcular a sensibilidade dos pontos pela equação 15.13

$$\gamma = \frac{\text{coeficiente angular}}{\text{dp das medidas}} = \frac{1.929}{1.527525} = 1,26$$

Exercício 15.9 A quantidade de chumbo presente na água de chuva foi determinada por espectrometria de absorção atômica. O químico que realizou as medidas forneceu a seguinte curva de calibração:

Conc. ng mL ⁻¹	10	25	100	200	300
Absorbância	0,05	0,17	0,22	0,60	1,40

Perde-se

1. Faça a curva de calibração;

2. Avalie a qualidade da curva de calibração;
3. Expresse a curva de calibração;
4. Sabendo que foram realizada 3 medidas, cujas absorvâncias foram 0,18, 0,19 e 0,20 calcule a sensibilidade do método.

Exercício 15.10 Os dados da curva de calibração abaixo foram obtidos de uma análise química de um metal raro. Determine a sensibilidade, sabendo que as medidas de concentração apresentam um desvio padrão de 0,5.

Conc. $mg\ mL^{-1}$	0	10	20	30	40	50
Sinal	1	263	531	801	1053	1333

Dentre as limitações de uma curva de calibração, a quantidade mínima a ser determinada é uma questão motivo de discussão. A quantidade mínima ou limite de detecção (L_D) mostra uma das limitações da lei de Beer. Segundo a lei de Beer

$$A = \epsilon CL \tag{15.14}$$

Na equação 15.14 é prevista entre a concentração zero e o primeiro ponto da curva de calibração uma linha reta. No entanto, podemos perceber até aqui, que esse comportamento nem sempre chegar ao zero após uma regressão linear. Por motivo óbvio, a oscilação devida ao erro indeterminado limita a chegada da linha reta da regressão ao zero, quando passa existe uma incerteza nas medidas. Outro fator que causa a oscilação é o próprio L_D , cuja definição é dada como a concentração ou massa mínima de analito que pode ser detectada em um nível confiável.

O que é limite de detecção?

O limite de detecção é a concentração mínima do analito que pode ser detectada em um dado nível de confiança.

Este limite depende da razão entre a magnitude do sinal do analito e o tamanho das flutuações estatísticas do sinal do branco. Isto é, a menos que o sinal analítico seja maior que o branco por um fator múltiplo de k da variação no branco devido aos erros aleatórios, é impossível de detectar o sinal analítico com certeza. Na verdade, o L_D é um conceito qualitativo, pois 50% das medidas efetuadas neste limiar são indistinguíveis do branco.

A Figura 18 mostra como seria o L_D ; ou seja, a sua localização se encontra entre o branco e o limite de quantificação.

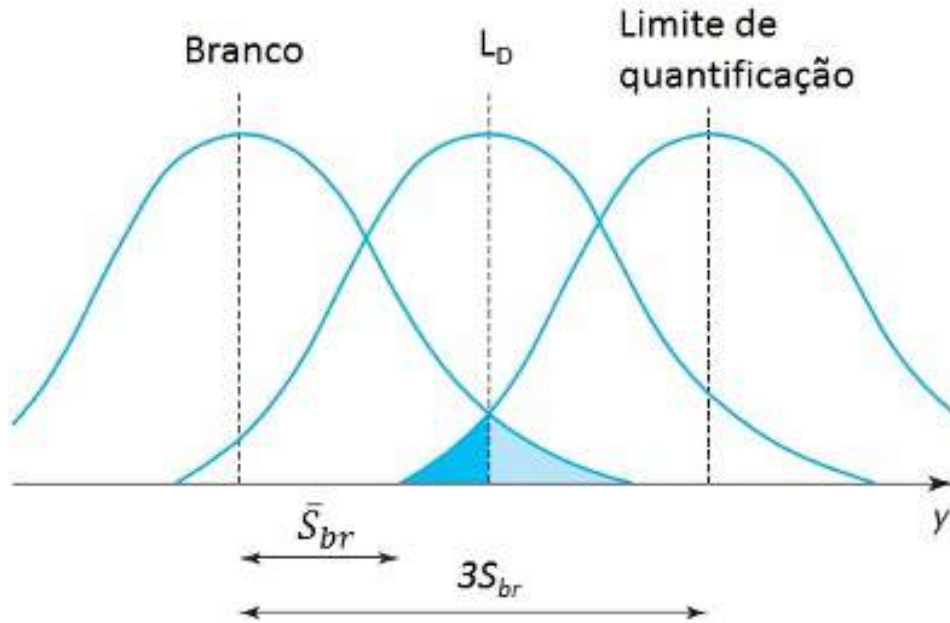


Figura 18 – Representação estatística do limite de detecção. Adaptada: Miler

Assim que o L_D for atingido, o sinal analítico e o seu desvio padrão se aproximam do sinal do branco (S_{br}) e de seu desvio padrão (dp_{br}). O sinal analítico mínimo distinguível S_m é então tomado como a soma do sinal médio do branco (\bar{S}_{br}) mais um múltiplo de k do desvio padrão (limite de confiança) do branco S_{br} . Assim

$$S_m = \bar{S}_{br} + kS_{br} \quad (15.15)$$

Sendo que \bar{S}_{br} e S_{br} calculado da seguinte forma:

$$\bar{S}_{br} = \frac{\sum_{i=1}^N br_i}{N} \quad (15.16)$$

$$S_{br} = \sqrt{\frac{\sum_{i=1}^N (br_i - \bar{S}_{br})^2}{N - 1}} \quad (15.17)$$

Experimentalmente, S_m pode ser determinado por medidas do branco br repetidas vezes, entre 20 e 30 vezes. Os dados resultantes são tratados estatisticamente para obter \bar{S}_{br} e S_{br} . A utilização de $k = 3$ permite uma confiança nível de 99,86% em que $S_m \geq (\bar{S}_{br} + 3S_{br})$ para uma medida

baseada no erro do sinal de um branco que segue a distribuição normal. Deve ser enfatizado que se \bar{S}_{br} não segue uma normal distribuição, então a probabilidade de $S_m \geq (\bar{S}_{br} + 3S_{br})$ deve ser $100(1 - \frac{1}{k^2})$, ou 89%. Assim, Os valores de $k < 3$ não devem ser utilizados para o cálculo de L_D . Na maioria das vezes $k=3$ tem sido usado para o cálculo do L_D . Geralmente a equação 15.15 é usada da seguinte forma:

$$S_m = \bar{S}_{br} + 3S_{br} \quad (15.18)$$

Finalmente, o coeficiente angular da curva de calibração e usada para encontrar a concentração do L_D :

$$L_D = \frac{S_m - \bar{S}_{br}}{\text{Coeficiente angular}} \quad (15.19)$$

Substituindo a equação 15.18 em 15.19 tem-se

$$L_D = \frac{3S_{br}}{\text{Coeficiente angular}} \quad (15.20)$$

Exemplo 15.5 *Um novo equipamento foi desenvolvido por um aluno de doutorado para a determinação de fósforo em água eutrofizada. Abaixo são apresentadas as medidas de valor do branco e uma curva de calibração. Usando o R encontre o L_D*

Absorbância do branco

0,012 0,056 0,004 0,022 0,015 0,011 0,042 0,094 0,089 0,055

Curva de calibração

<i>Conc. mg g⁻¹</i>	0,00	0,50	1,25	2,00	3,00	4,00
<i>Absorbância</i>	0.00	0,514	0,993	1,486	2,044	2,342

Para determinar o L_D inicialmente temos que encontrar duas variáveis: o desvio padrão do branco (S_{br}) e o coeficiente angular da curva de calibração.

Cálculo do desvio padrão do branco no R

Entrada dos dados do branco

>brancoj-c(0.012, 0.056, 0.004, 0.022, 0.015, 0.011, 0.042, 0.094, 0.089, 0.055)

Cálculo do S_{br}

```
> sd(branco)
```

```
[1] 0.03278211
```

Cálculo do coeficiente angular

Entrada dos dados

```
> conc<-c(0,0.50,1.25,2,3,4)
```

```
> abs<-c(0,0.514,0.993,1.486,2.044,2.342)
```

```
> dados=data.frame(abs,conc)
```

```
> is.data.frame(dados)
```

```
[1] TRUE
```

```
> regressao=lm(abs conc,dados)
```

```
> regressao
```

Call:

```
lm(formula = abs conc, data = dados)
```

Coefficients:

```
(Intercept) conc
```

```
0.1830      0.5843
```

Qualidade da curva de calibração

```
> summary(regressao)
```

Call:

```
lm(formula = abs conc, data = dados)
```

Residuals:

```
      1      2      3      4      5      6  
-0.18302  0.03885  0.07965  0.13444  0.10817 -0.17810
```

Coefficients:

	<i>Estimate</i>	<i>Std. Error</i>	<i>t value</i>	<i>Pr(> t)</i>	
(Intercept)	0.18302	0.10691	1.712	0.162084	
conc	0.58427	0.04718	12.385	0.000244	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Residual standard error: 0.1603 on 4 degrees of freedom

Multiple R-squared: 0.9746, Adjusted R-squared: 0.9682

F-statistic: 153.4 on 1 and 4 DF, p-value: 0.0002443

Os dados da curva de calibração mostram que sua qualidade é baixa, mas o que em princípio irá refletir no cálculo do L_D do método desenvolvido pelo aluno. Calculando teremos:

$$L_D = \frac{3.0,0328}{0,5843} = 0,1682 \text{ mg } g^{-1}$$

O resultado mostra que o L_D do método desenvolvido pelo aluno é relativamente alto para os padrões exigidos atualmente.

Exercício 15.11 Os resultados abaixo se referem à determinação de prata por espectrometria de absorção atômica.

Absorbância do branco

0,0052 0,0033 0,008 0,006 0,0031 0,0006 0,0035 0,0095

Curva de calibração

Conc. ng g^{-1}	0,00	5	10	15	20	25
Absorbância	0,003	0,127	0,251	0,390	0,498	0,625

Baseado nas informações acima determine o L_D da análise de prata.

Exercício 15.12 Uma análise pelo método dos mínimos quadrados de uma curva de calibração para a determinação de chumbo resultou na seguinte equação da reta:

$$S = 1,12c_{Pb} + 0,312$$

c_{Pb} representa a concentração de chumbo em partes por milhão e S é uma medida de intensidade relativa da linha de emissão do chumbo. A tabela abaixo mostra alguns dados derivados das medidas de chumbo.

Conc. Pb (ppm)	N	Média de S	desvio padrão
10,000	10	11,62	0,1500
1,000	10	1,12	0,0250
0,000	24	0,0296	0,0082

Baseado nos dados acima calcule o L_D

Exercício 15.13 Os dados abaixo são referente à determinação de ortofosfato (H_3PO_4) presente em água de Igarapé.

Conc. H_3PO_4	Sinal analítico	desvio padrão
0,00	0,031	0,0079 (n=25)
2,00	0,173	0,0094
6,00	0,422	0,0084
10,00	0,702	0,0084
14,00	0,956	0,0085
18,00	1,248	0,0110

1. Calcule a sensibilidade da curva de calibração
2. Diferente do branco, as outras medidas foram realizadas cinco vezes. Diante dessas informações, calcule a sensibilidade em cada medida.
3. calcule o coeficiente de variação das medidas
4. qual é o limite de detecção do método?

Exercício 15.14 A resposta de um teste colorimétrico para glucose foi avaliada com a ajuda de soluções padrão. Após 30 medidas o branco apresentou como desvio padrão o valor de 0,005. Utilize a concentração de calibração a seguir para determinar o L_D :

Curva de calibração						
Conc. $ng\ g^{-1}$	0,00	2	4	6	8	10
Absorbância	0.002	0,150	0,294	0,434	0,570	0,704

Exercício 15.15 Lucker usando amostras brutas mediu a concentração de Cd diretamente por ETAAS no comprimento de onda de 228,8 nm. Para determinar o valor de L_D os autores utilizaram 40 medidas de brancos encontrando o valor de 0,0013 $ng\ mg^{-1}$ para uma curva de calibração de $y = 0,42x - 0,0002$. Diante dessas informações, calcule o desvio padrão do branco.

Exercício 15.16 Um novo programa de computador foi desenvolvido para um espectrômetro de absorção atômica com forno de grafite pela equipe de programação de uma indústria de equipamentos químicos. Para verificar a qualidade dos resultados do novo programa Kuselman resolveram usar uma

solução padrão de chumbo. A Tabela abaixo apresenta valores de altura do pico referentes à análise de chumbo obtida do programa (Prog.) considerado ótima e do novo.

<i>Conc. Pb (ppb)</i>	<i>0,0</i>	<i>5</i>	<i>10</i>	<i>20</i>	<i>30</i>	<i>40</i>
<i>Prog. Padrão</i>	<i>0,011</i>	<i>0,056</i>	<i>0,104</i>	<i>0,188</i>	<i>0,267</i>	<i>0,337</i>
<i>Prog. Novo</i>	<i>0,015</i>	<i>0,060</i>	<i>0,113</i>	<i>0,188</i>	<i>0,281</i>	<i>0,370</i>

- 1. Determine a sensibilidade da curva nos dois métodos.*
- 2. Calcule a concentração nas duas curvas para as alturas de pico de 0,234 e 0,248.*
- 3. Discuta os valores de concentração obtidos anteriormente.*
- 4. Sabendo que o desvio padrão para o programa padrão foi de 0,003 e do programa novo de 0,02, calcule os L_D para os dois programas.*

Nem sempre uma curva de calibração contendo todos os parâmetros aceitáveis para uma medida analítica pode ser recomendada. Isso porque a resposta de uma medida analítica terá mais confiabilidade se estiver livre de interferências. Infelizmente, nenhum método analítico está totalmente livre de interferências de outras espécies, e frequentemente são necessárias diversas etapas para se minimizar os seus efeitos. A Figura 19 representa uma medida contendo o analito e uma interferência; ou seja o sinal analítico registrado no aparelho se refere a dois componentes (analito e interferência).

Seletividade se refere a extensão que o método pode ser usado para determinar analitos em misturas ou matrizes sem interferências de outros componentes de comportamento similar.

Para o entendimento da seletividade vamos utilizar a situação em que um analito é composto dentre outras espécies A, B e C. B e C são interferentes em potencial de A. Se c_A, c_B e c_C são as concentrações das três espécies e m_A, m_B e m_C são suas sensibilidades de calibração, então o sinal instrumental total deverá ser dado por

$$Sinal = m_{ACA} + m_{BCB} + m_{CCC} + intercepto \quad (15.21)$$

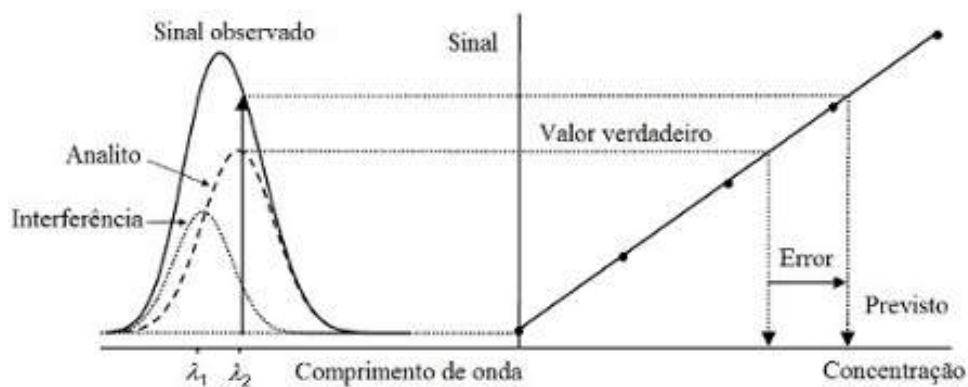


Figura 19 – Efeito da interferência sobre uma Curva de calibração. Fonte: Olivieri

Definindo coeficiente de seletividade para A com relação a B como

$$k_{B,A} = \frac{m_B}{m_A} \quad (15.22)$$

O coeficiente de seletividade representa a resposta relativa do método à espécie B quando comparado com A. Um coeficiente similar para A com relação a C seria:

$$k_{C,A} = \frac{m_C}{m_A} \quad (15.23)$$

substituindo as equações 15.22 e 15.23 em 15.21 tem-se:

$$Sinal = m_A(c_A + k_{B,A}c_B + k_{C,A}c_C) + intercepto \quad (15.24)$$

16 Tabelas

Tabela 17 – Alguns valores de F ao nível de 95% de confiança

	$v_1=2$	3	4	5	6	7	8
$v_2=2$	19,0	19,2	19,2	19,3	19,3	19,4	19,4
3	9,55	9,28	9,12	9,01	8,94	8,89	8,85
4	6,94	6,59	6,39	6,26	6,16	6,09	6,04
5	5,79	5,41	5,19	5,05	4,95	4,88	4,82
6	5,14	4,76	4,53	4,39	4,28	4,21	4,15
7	4,74	4,35	4,12	3,97	3,87	3,79	3,73
8	4,46	4,07	3,84	3,69	3,58	3,50	3,44

Tabela 18 – Valores para t de Student em função do grau de liberdade $n - 1$ para um limite de confiança de 95%

Grau de liberdade $n - 1$	Probabilidade
1	12,71
2	4,30
3	3,18
4	2,78
5	2,57
6	2,45
7	2,37
8	2,31
9	2,26
10	2,23
11	2,20
12	2,18
13	2,16
14	2,15
15	2,13
16	2,12
17	2,11
18	2,10
19	1,96

Tabela 19 – Valores para Q em função do número de observações a um limite de confiança de 95%

Número de observações (N)	Q
2	-
3	0,970
4	0,829
5	0,710
6	0,625
7	0,568
8	0,526
9	0,493
10	0,466

Referências

N. Baccan, J. C. Andrade, O. E. S. Godinho, and J. S. Barone. *Química Analítica Quantitativa Elementar*. Edgar Blücher, Campinas, 4 a revista, ampliada e reestruturada edition, 2001.

Freddy Cienfuegos. *Estatística Aplicada ao Laboratório*. Editora Interciência, 2005.

D. C. Harris. *Quantitative Chemical Analysis*. W. H. Freeman and Company, New York, 6a edition, 2007.

D. C. Harris. *Explorando a Química Analítica*. LTC Editora, Rio de Janeiro, 4 a edition, 2011.

S. Higson. *Química Analítica*. MacGraw Hill, São Paulo, 2009.

D. Kealey and P. J. Haines. *Instant Notes - Analytical Chemistry*. Bios Scientific Publishers Limited, Oxford, 2002.

G. L. Long and J. D. Winefordner. Limit of detection - a closer look at the iupac definition. *Analytical Chemistry*, 55(7):712A–718A, 1983.

J. Mendham, R. C. Denny, J. D. Barnes, and M. Thomas. *Vogel - Análise Química Quantitativa*. LTC Editora, Rio de Janeiro, 6a edition, 2002.

J. C. Miller and J. N. Miller. *Estatística para Química Analítica*. Addison-Wesley Iberoamericana, México, 2. a edition, 1993.

J. N. Miller and J. C. Miller. *Statistics and Chemometrics for analytical chemistry*. Pearson Education Limited, Edinburgh, 5 edition, 2005.

M. F. Pimentel and Barr. Calibração: Uma revisão para químicos analíticos. *Química Nova*, 19(3):268–277, 1996.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

D. A. Skoog, D. M. West, F. J. Holler, and S. R. Crouch. *Fundamentos de Química Analítica*. Pioneira Thomson Learning, São Paulo, 8 a edition, 2006.

J. Vessman, R. I. Stefan, J. F. Van Staden, K. Danzer, W. Lindner, D. T. Burns, A. Fajgelj, and H. Müller. Selectivity in analytical chemistry (iupac recommendations 2001). *Pure and Applied Chemistry*, 73(8):1381–1386, 2001.