

Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Instituto de Química



Bruna Castello Longo Estrella

**ESTUDO DE CASOS DA APLICABILIDADE DO TESTE T-PAREADO E DA
REGRESSÃO LINEAR PELO MÉTODO DOS MÍNIMOS QUADRADOS: UMA
ABORDAGEM ESTATÍSTICA APLICADA À QUÍMICA ANALÍTICA**

Rio de Janeiro

2023

Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Instituto de Química



Bruna Castello Longo Estrella

**ESTUDO DE CASOS DA APLICABILIDADE DO TESTE T-PAREADO E DA
REGRESSÃO LINEAR PELO MÉTODO DOS MÍNIMOS QUADRADOS: UMA
ABORDAGEM ESTATÍSTICA APLICADA À QUÍMICA ANALÍTICA**

Trabalho de conclusão do curso submetido ao corpo docente do curso de Química com Atribuições Tecnológicas do Instituto de Química da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para a obtenção do grau de Química - Atribuições Tecnológicas.

Orientadoras: Profa. Dra. Débora França de Andrade (DQA-IQ/UFRJ)

Profa. Dra. Paula Fernandes de Aguiar (DQA-IQ/UFRJ)

Rio de Janeiro

2023

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Estrella, Bruna Castello Longo

E82e Estudo de casos da aplicabilidade do teste t-pareado e da regressão linear pelo método dos mínimos quadrados: uma abordagem estatística aplicada à Química Analítica / Bruna Castello Longo Estrella. --

Rio de Janeiro, 2023.

56 f.

Orientadora: Débora França de Andrade. Coorientadora: Paula Fernandes de Aguiar.

Trabalho de conclusão de curso (graduação) -

Universidade Federal do Rio de Janeiro, Instituto de Química, Bacharel em Química, 2023.


1. Estatística básica. 2. Teste t-pareado. 3. Regressão linear. I. de Andrade, Débora França, orient. II. de Aguiar, Paula Fernandes, coorient. III. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.


Bruna Castello Longo Estrella

**ESTUDO DE CASOS DA APLICABILIDADE DO TESTE T-PAREADO E DA
REGRESSÃO LINEAR PELO MÉTODO DOS MÍNIMOS QUADRADOS: UMA
ABORDAGEM ESTATÍSTICA APLICADA À QUÍMICA ANALÍTICA**


Trabalho de conclusão do curso
submetido ao corpo docente do curso
de Química com Atribuições
Tecnológicas do Instituto de Química
da Universidade Federal do Rio de
Janeiro, como parte dos requisitos
necessários para a obtenção do grau de
Química - Atribuições Tecnológicas.

Documento assinado digitalmente
 DEBORA FRANCA DE ANDRADE
Data: 14/08/2023 22:18:50-0300
Verifique em <https://validar.iti.gov.br>


Profa. Dra. Débora França de Andrade, DQA-IQ/UFRJ

Documento assinado digitalmente
 PAULA FERNANDES DE AGUIAR
Data: 11/08/2023 12:15:28-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Paula Fernandes de Aguiar, DQA-IQ/UFRJ

Documento assinado digitalmente
 ALINE DOMINGOS GONCALVES
Data: 11/08/2023 15:49:31-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Aline Domingos Gonçalves, DQA-IQ/UFRJ

Documento assinado digitalmente
 DANIELLA LOPEZ VALE
Data: 12/08/2023 15:00:48-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Daniella Lopez Vale, DQA-IQ/UFRJ

Rio de Janeiro, RJ - Brasil
Agosto 2023

AGRADECIMENTOS

Aos meus pais e meus avós, por terem me apoiado durante todos esses anos, sempre interessados com a minha faculdade e com o meu futuro, me permitindo colocar os estudos em primeiro lugar.

À Isabelle, minha irmã de alma, obrigada por estar presente todos os dias desde que entramos na faculdade. Em todos os momentos de desespero você estava sempre lá para me ouvir e me estender a mão. Te agradeço muito, também, por toda a ajuda com as disciplinas, mesmo quando você mudou de curso, continuou me ajudando com matérias que você nem fazia. Serei eternamente grata por tudo.

Ao Humilde, por terem transformado essa trajetória tão difícil em algo mais leve. Ter vocês como amigos foi essencial para que eu pudesse chegar até aqui. Obrigada por todos os risos, todos os encontros, todos os presentes e, principalmente, por todo o apoio e ajuda durante todos esses anos.

Ao Sérgio, meu orientador acadêmico, por todo o ensinamento passado durante as disciplinas que fiz com você, além de toda a ajuda em relação à grade curricular. Foi muito importante para mim.

Às minhas orientadoras de TCC, Paula e Débora, por terem aceitado embarcar nessa trajetória comigo e terem feito dela algo muito tranquilo, sempre muito dispostas a me ajudar e tornar o trabalho ainda melhor. Aprendi muito com vocês.

ESTRELLA, Bruna Castello Longo. Estudos de caso da aplicabilidade do teste t-pareado e da regressão linear pelo método dos mínimos quadrados: uma abordagem estatística aplicada à Química Analítica. Rio de Janeiro, 2023. Trabalho de Conclusão de Curso - Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2023.

RESUMO

Os testes estatísticos são de extrema importância em muitas áreas, principalmente quando se trata de pesquisa. O químico deve, portanto, não apenas entender e escolher qual teste usar, mas também saber aplicá-los. Sabe-se, no entanto, que há uma diversidade muito grande em relação a tais testes. Nesse sentido, no presente trabalho, o teste t-pareado foi analisado sob um novo olhar de forma a poder antecipar seu resultado, assim como a regressão linear realizada pelo método dos mínimos quadrados, assuntos que são pouco entendidos, mesmo entre pessoas que os utilizam rotineiramente. Os objetivos deste trabalho foram, portanto, tentar deixar mais claro o uso do teste t-pareado, a partir de uma relação matemática para antecipar e facilitar a aplicação deste teste e, também, analisar curvas analíticas construídas utilizando todos os valores de replicata de cada concentração, utilizando suas respectivas médias, além de números diferentes de replicata, a fim de avaliar o impacto das diferentes formas de construí-las. Os resultados obtidos a partir da utilização da fórmula do teste t-pareado mostram que foi possível criar uma avaliação preliminar para este teste, de modo que, sem mesmo chegar até o final de seu cálculo, já seja possível saber qual será seu resultado, além de terem simulados estudos de caso para a comprovação dessa análise. Já em relação à regressão linear, foram avaliados conjuntos de dados para a construção de curvas analíticas, tanto para concentrações em triplicata quanto para concentrações em quintuplicata, além de suas respectivas médias e de diferentes números de replicatas para algumas concentrações. Seus resultados mostram como as formas de construção da curva podem influenciar a precisão da previsão (e, conseqüentemente, a exatidão) realizada através do modelo estimado, permitindo, assim, que o analista possa definir uma estratégia para a sua construção. Este entendimento pode facilitar o uso da estatística e da matemática para todas e quaisquer pessoas. A partir da análise de parâmetros estatísticos relacionados aos resultados das curvas, concluiu-se que o modelo de regressão linear mais eficiente para a obtenção de resultados dependerá de quanto o analista poderá assumir como incerteza na

previsão dos resultados, levando também em consideração o custo e o tempo disponível para a realização das análises.

Palavras-chave: testes estatísticos, teste t-pareado, regressão linear.

ESTRELLA, Bruna Castello Longo. Estudos de caso da aplicabilidade do teste t-pareado e da regressão linear pelo método dos mínimos quadrados: uma abordagem estatística aplicada à Química Analítica. Rio de Janeiro, 2023. Trabalho de Conclusão de Curso - Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2023.

ABSTRACT

Statistical tests are extremely important in many areas, especially when it comes to research. Therefore, chemists must not only understand and choose which tests to use, but also know how to apply them. It is known, however, that there is a great diversity in relation to such tests. In this sense, in the present work, the paired t-test was analyzed under a new perspective in order to be able to anticipate its result, as well as the linear regression performed by the least squares method, subjects that are little understood, even among people who use them routinely. Therefore, the objectives of this work were to try to make clearer the use of the paired t-test, from a mathematical relation, to anticipate and facilitate the application of this test and, also, to analyze analytical curves built using all the replicate values of each concentration, using their respective averages, in addition to different numbers of replicates, in order to evaluate the impact of the different ways of building them. The results obtained from the use of the paired t-test formula show that it was possible to create a preliminary evaluation for this test, so that, without even reaching the end of its calculation, it is already possible to know what its result will be and simulated case studies were evaluated to corroborate this analysis. Regarding linear regression, datasets were evaluated for the construction of analytical curves, both for concentrations in triplicate and for concentrations in quintuplicate, in addition to their respective means and different numbers of replicates for some concentrations. Its results show how the ways of building the curve can influence the accuracy of the prediction (and, consequently, the exactness) made through the estimated model, thus allowing the analyst to define a strategy for its construction. This understanding can facilitate the use of statistics and mathematics for anyone and everyone. From the analysis of statistical parameters related to the results of the curves, it was concluded that the most efficient linear regression model for obtaining results will depend on how much the analyzer can assume as uncertainty in predicting the results, also taking into account the cost and the time available for carrying out the analyses.

Key-words: statistical tests, paired t-test, linear regression.

LISTA DE FIGURAS

Figura 1 - Teste bilateral (a), teste unilateral à esquerda (b) e teste unilateral à direita (c).....	19
Figura 2 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) de 5 concentrações diferentes (C1, C2, C3, C4 e C5) em triplicata.....	38
Figura 3 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) da média das triplicatas das 5 concentrações diferentes (C1, C2, C3, C4 e C5).....	38
Figura 4 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) com triplicatas das concentrações C1, C3 e C5 e duplicatas das concentrações C2 e C4.....	39
Figura 5 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) de 5 concentrações diferentes (C1, C2, C3, C4 e C5) em quintuplicata.....	42
Figura 6 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) da média das triplicatas das 5 concentrações diferentes (C1, C2, C3, C4 e C5).....	43
Figura 7 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) com triplicatas das concentrações C1, C3 e C5 e duplicatas das concentrações C2 e C4.....	44

LISTA DE TABELAS

Tabela 1 - Conjunto de dados do estudo de caso 1.....	27
Tabela 2 - Conjunto de dados do estudo de caso 2.....	28
Tabela 3 – Valores das concentrações de C1 até C5 (x) e das leituras (y) em triplicata.....	29
Tabela 4 - Valores das concentrações de C1 até C5 (x) e das leituras (y) para a média das triplicatas.....	29
Tabela 5 - Valores das concentrações de C1 até C5 (x) e das leituras (y) para as triplicatas C1, C3 e C5 e duplicatas C2 e C4.....	30
Tabela 6 - Valores das concentrações de C1 até C5 (x) e das leituras (y) em quintuplicata.....	30
Tabela 7 - Valores das concentrações de C1 até C5 (x) e das leituras (y) para a média das quintuplicatas.....	31
Tabela 8 - Valores das concentrações de C1 até C5 (x) e das leituras (y) para as quintuplicatas C1, C3 e C5 e duplicatas C2 e C4.....	31
Tabela 9 – Valores utilizados para que a hipótese nula seja aceita, no teste t-pareado.....	32
Tabela 10 – Conjunto de dados do estudo de caso 1 com o valor das diferenças (d) dos resultados obtidos para cada método.....	33
Tabela 11 – Resultados da aplicação do teste t-pareado, obtidos pelo Excel a partir do conjunto de dados do estudo de caso 1.....	34
Tabela 12 - Conjunto de dados do estudo de caso 2 com o valor das diferenças (d) dos resultados obtidos para cada método.....	34
Tabela 13 – Resultados da aplicação do teste t-pareado, obtidos pelo Excel, a partir do conjunto de dados do estudo de caso 2.....	35
Tabela 14 – Valores de G calculado para cada concentração em triplicata.....	36
Tabela 15 – Valores de G calculado para cada concentração em quintuplicata.....	37

Tabela 16 - Comparação dos resultados de equação da reta, coeficiente de determinação e variância residual para cada curva analítica construída com 5 concentrações diferentes, C1, C2, C3, C4 e C5 (a) todas as 5 concentrações em triplicata, (b) média das triplicatas de cada concentração e (c) triplicata das concentrações C1, C3 e C5 e duplicata das concentrações C2 e C4.....40

Tabela 17 – Comparação dos resultados de equação da reta, coeficiente de determinação e variância residual para cada curva analítica construída com 5 concentrações diferentes, C1, C2, C3, C4 e C5 (a) todas as 5 concentrações em quintuplicata, (b) média das quintuplicatas de cada concentração e (c) quintuplicata das concentrações C1, C3 e C5 e duplicata das concentrações C2 e C4.....44

LISTA DE SÍMBOLOS

α	coeficiente linear
a	estimativa do coeficiente linear
β	coeficiente angular
b	estimativa do coeficiente angular
C	valor do teste de Cochran
C_1, C_2, C_3, C_4, C_5	valores de concentração 1, 2, 3, 4 e 5
d	valor da diferença entre os dados
G	valor do teste de Grubb's
H_0	hipótese nula
H_1	hipótese alternativa
m	número de replicatas
n	número de observações
r	coeficiente de correlação
R^2	coeficiente de determinação
s	desvio padrão
s^2_e	variância residual
s^2_i	variâncias
s^2_{\max}	variância máxima
s^2_r	variância em relação à repetibilidade
s_d	desvio padrão da diferença dos dados
s_e	desvio padrão residual
t	valor de t de Student
VS	valor suspeito

x variável independente

x_i valores para as concentrações (i)

$|\bar{X}_d|$ módulo da média da diferença dos dados

\bar{x} valor médio de x

y variável dependente

y_i valores mensurados para as concentrações (i)

\hat{y} estimativa do valor de y previsto pelo método

\bar{y} valor médio de y

\hat{y}_i estimativa de y_i

ε erro

Σ somatório

ν grau de liberdade

SUMÁRIO

1 INTRODUÇÃO.....	16
2 OBJETIVOS.....	18
2.1 Objetivo geral.....	18
2.2 Objetivos específicos.....	18
3 FUNDAMENTAÇÃO TEÓRICA.....	19
3.1 Teste de hipótese.....	19
3.2 A regressão linear.....	21
3.3 Variância residual	23
3.4 Erro da previsão.....	23
3.5 Teste de Grubb's.....	24
3.6 Teste de Cochran.....	24
4 METODOLOGIA.....	26
4.1 Seleção dos testes matemáticos e estatísticos a serem estudados.....	26
4.2 Estudo do teste t-pareado.....	26
4.3 Estudo da influência do uso das replicatas e da média das replicatas na construção da curva analítica obtida através da regressão linear.....	28
4.3.1 Concentrações em triplicata, com a média das triplicatas e com diferentes números de replicata para algumas concentrações.....	29
4.3.2 Concentrações em quintuplicata, com a média das quintuplicatas e com diferentes números de replicata para algumas concentrações.....	30
5 RESULTADOS E DISCUSSÃO.....	32
5.1 Estudo do teste t-pareado.....	32
5.2 Estudo da influência do uso das replicatas e da média das replicatas na construção da curva analítica obtida através da regressão linear.....	35

5.2.1	Concentrações em triplicata, com a média das triplicatas e com diferentes números de replicata para algumas concentrações.....	37
5.2.2	Concentrações em quintuplicata, com a média das quintuplicatas e com diferentes números de replicata para algumas concentrações.....	42
5.2.3	Comparação entre o uso de triplicata e quintuplicata.....	46
6	CONCLUSÃO.....	48
	REFERÊNCIAS BIBLIOGRÁFICAS.....	49
	APÊNDICE A - TABELA COM AS CONDIÇÕES DO TESTE T-PAREADO ATÉ $n = 101$	51
	ANEXO A - TABELA DA DISTRIBUIÇÃO t-STUDENT.....	54
	ANEXO B - TABELA PARA TESTE DE GRUBB'S.....	55
	ANEXO C – TABELA PARA TESTE DE COCHRAN.....	56

1 INTRODUÇÃO

Cada vez mais, são exigidas aplicações de testes estatísticos e métodos matemáticos nos resultados obtidos de qualquer pesquisa ou tratamento de dados. Isto força o estudo da Estatística e da Matemática em todo e qualquer curso de formação. Os resultados de análises e processos químicos não são exceções (VIEIRA, 2008; MASSART *et al.*, 1998; MILLER & MILLER, 1993).

Porém, conhecer a Estatística e a Matemática pode ser bastante laborioso. O químico ou estudante do curso de Química, precisa aprender, pelo menos, o básico para que tenha autonomia na hora de decidir por um teste ou por um método matemático, e aplicá-lo corretamente. Cada teste estatístico, por exemplo, tem uma aplicação definida, seja na estatística paramétrica ou não paramétrica (MILLER & MILLER, 1993). Parece simples pensar que se pode, então, conhecer todos os principais testes estatísticos para poder usá-los sem grandes problemas ou restrições. A realidade, no entanto, é outra. A aplicação de métodos matemáticos e testes estatísticos usam manipulações numéricas e, por causa disso, estão sujeitos a artefatos, causados não só pelas características dos conjuntos de dados, mas também pela teoria que está por trás da construção dos testes e métodos. Assim, além de saber aplicar cada um deles, é fundamental entender esta teoria para identificar e eliminar qualquer possível artefato quanto à aplicação de cada um deles (MASSART *et al.*, 1998).

Neste trabalho, portanto, resolveu-se deixar mais claro como dois dos diversos tópicos presentes na Estatística e na Matemática interferem no dia a dia de seus usuários quando se trata, por exemplo, da estimativa de modelos. Os dois tópicos escolhidos são o teste de hipótese, mais especificamente aquele no qual o teste t-pareado é usado, e a regressão linear pelo método dos mínimos quadrados.

Quando se trata da Inferência Estatística, é extremamente importante buscar respostas sobre características específicas da população que se está analisando. Considerando estas características, portanto, deve-se definir regras e decidir sobre qualquer formulação de hipótese se referindo à população. Desse modo, selecionada uma variável x e coletada uma amostra aleatória da população, há o interesse em inferir sobre alguns parâmetros, como a média, a variância e a proporção, por exemplo, de modo a definir o comportamento da variável. A execução dos testes de hipótese oferece meios para que seja possível, com um certo nível de confiança, constatar se os valores dos

parâmetros ou até a distribuição relacionada à população em estudo, pode retratá-la de maneira satisfatória (GUIMARÃES, 2008).

Para a comparação de duas médias, o teste estatístico mais utilizado é o teste t de Student e este pode ser empregado em duas ocasiões distintas: quando os dados são pareados e quando os grupos são independentes. No presente trabalho, o caso criado para o estudo do teste t se utiliza de grupos de dados pareados, o que significa que será aplicado o teste t-pareado. Para este teste, é importante salientar que o pareamento precisa apresentar algum tipo de relação, de modo que uma mesma amostra seja analisada por dois métodos diferentes, por exemplo, como é o caso do trabalho em questão (VIEIRA, 2008).

Outro ponto importante na análise química e estatística é a construção de curvas analíticas e, para isso, utiliza-se a regressão linear, ferramenta que relaciona duas ou mais variáveis, a fim de determinar a concentração de uma substância em uma dada amostra. A regressão linear utiliza o método dos mínimos quadrados para estimar os coeficientes do modelo (LOPES, 2003).

Ambos os tópicos citados necessitam de conhecimento básico de Estatística para que possam ser aplicados. Porém, para que sua aplicação seja feita da forma mais correta possível, é preciso compreender muito bem toda a matemática que está por detrás de cada um deles. Este conhecimento é pouco observado pelos usuários da Estatística e pouco discutido nos cursos de Estatística, de forma geral.

Assim, é importante que estes tópicos sejam claramente entendidos para que possam ser aplicados de forma correta e o resultado de suas aplicações, melhor aproveitados.

2 OBJETIVOS

2.1. Objetivo geral

Encontrar uma relação matemática para prever o resultado da aplicação do teste t-pareado, a fim de se obter um resultado qualitativo de forma mais simples e rápida. Avaliar, estatisticamente, curvas analíticas construídas mediante o emprego de todos os valores das replicatas de cada concentração, além da média dessas replicatas e de números de replicatas diferentes por nível de concentração a partir de dados simulados, para que seja analisado o impacto que estes modelos causam no poder de previsão.

2.2. Objetivos específicos

Dentre os objetivos específicos do presente trabalho, podemos destacar:

- Determinação de uma relação matemática entre os valores mensurados de tal forma que eles possam prever o resultado da aplicação do teste t-pareado;
- Simulação de conjuntos de dados como estudos de caso para a demonstração da relação matemática encontrada;
- Construção, análise e comparação de algumas curvas analíticas, obtidas com todos os valores das replicatas, com as médias dessas replicatas e com números diferentes de replicatas para conjuntos de dados com dispersões diferentes em cada nível de concentração;
- Comparação dos parâmetros estatísticos diretamente relacionados à construção da reta de forma a avaliar o impacto que estas formas de construção da curva têm na precisão dos valores previstos pelos modelos estimados.

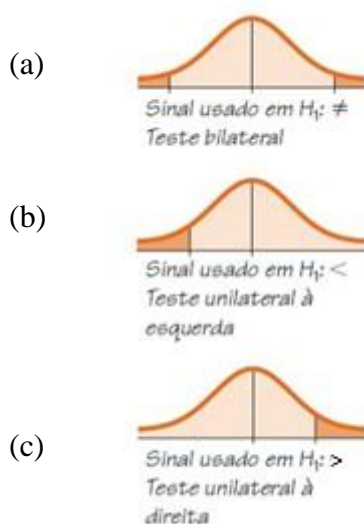
3 FUNDAMENTAÇÃO TEÓRICA

3.1 Teste de hipótese

Chama-se de hipótese estatística qualquer afirmação que diz respeito à distribuição de probabilidades de uma ou mais variáveis aleatórias (BOLFARINE; SANDOVAL, 2010). Para se testar uma hipótese estatística, faz-se necessário determinar um par de hipóteses, de modo que uma delas retrata uma afirmativa e a outra, portanto, a complementa. A hipótese que é caracterizada pela afirmativa de igualdade é a hipótese nula (dada por H_0) e o complemento desta hipótese nula é a hipótese alternativa (dada por H_1 , de forma a serem usados os seguintes símbolos: “ \neq ” e “ $<$ ” ou “ $>$ ”, por exemplo). A hipótese nula e a hipótese alternativa são responsáveis por descreverem dois possíveis cenários mutuamente excludentes, já que as duas hipóteses não podem ser aceitas ou rejeitadas ao mesmo tempo (FERREIRA, 2015).

Os testes de hipótese podem ser divididos em teste de hipótese em uma cauda (unilateral à esquerda ou unilateral à direita) e nas duas caudas (bilateral). O primeiro tipo é um teste no qual a hipótese alternativa H_1 determina uma mudança em alguma direção da hipótese nula H_0 , de modo a incluir em sua especificação um dos símbolos “ $<$ ” (caso seja unilateral à esquerda) ou “ $>$ ” (caso seja unilateral à direita). Já o segundo tipo é um teste no qual a hipótese alternativa H_1 determina uma mudança da hipótese nula H_0 sem especificar nenhuma direção, de modo a incluir em sua especificação o símbolo “ \neq ” (FERREIRA, 2015). Estes conceitos são mais bem visualizados na Figura 1:

Figura 1 – Teste bilateral (a), teste unilateral à esquerda (b) e teste unilateral à direita (c). Fonte: Adaptado de TRIOLA, 2017.



Há três tipos de teste de hipótese, são eles: comparação de um valor medido com um valor de referência (caso 1), comparação de médias de duas amostras independentes (caso 2) e comparação de médias de duas amostras pareadas (caso 3) (MISHRA *et al.*, 2019).

No presente trabalho, será detalhado somente este último caso por ser um dos focos do estudo.

O teste t para amostras pareadas, o qual é objeto de estudo neste trabalho, também chamado de teste t-pareado ou teste t de amostras dependentes, é utilizado para determinar se a mudança nas médias entre duas observações pareadas é estatisticamente significativa. Neste teste, as mesmas respostas são medidas para amostras diferentes, observadas, por exemplo, por dois métodos distintos. Para aplicar este teste, variáveis pareadas são utilizadas, sendo que as variáveis pareadas devem ser contínuas e normalmente distribuídas (MISHRA *et al.*, 2019).

Neste teste, devem ser realizados números iguais de medições em ambos os métodos a serem analisados, visto que cada amostra é medida nos dois métodos. Posteriormente, calcula-se o módulo da média ($|\bar{X}_d|$) e o desvio padrão (s_d) da diferença entre os dois métodos para cada uma das amostras (ALFASSI *et al.*, 2005), sendo zero o valor de referência entre estes dados. Além destes dois dados, para chegar a um resultado, também é necessário utilizar o número de observações (n). Dito isso, pode-se calcular o valor de t pela equação 1:

$$t = \frac{|\bar{X}_d|\sqrt{n}}{s_d} \quad (1)$$

Porém, o valor de t calculado não significa nada sozinho. É necessário, portanto, compará-lo com o valor do t crítico ou tabelado (Anexo A, página 54) já previamente estabelecido, para o mesmo nível de significância. Caso o valor de t tabelado seja maior que o valor de t calculado para um determinado número de graus de liberdade, não há diferença entre os métodos testados, por exemplo, ou seja, a hipótese nula é aceita; caso o valor de t calculado seja maior, a hipótese nula será rejeitada, de modo que, neste exemplo, os métodos testados não seriam estatisticamente equivalentes, visto que há diferença entre os métodos testados.

Deve-se salientar que, neste método de teste t-pareado, é assumido que os erros aleatórios são independentes da concentração. Ao longo de uma ampla faixa de concentração, essa suposição pode se tornar falsa e, dessa forma, esse método não deve ser utilizado em uma faixa muito extensa de concentrações (ALFASSI *et al.*, 2005).

3.2 A regressão linear

Os modelos de regressão são extremamente usados em várias áreas do conhecimento, como, por exemplo: ciências, computação, administração, engenharias, entre outros. Esta técnica tem como intuito a obtenção de uma equação que seja capaz de explicar a relação entre as respectivas variáveis, o que permite estimar valores da variável de interesse, com base em valores conhecidos da outra variável (GUIMARÃES, 2008).

Por meio da correlação, pode-se verificar se existe relação entre duas variáveis quantitativas: variável y (variável aleatória ou dependente) e variável x (variável controlada ou independente). Identificando uma relação linear entre as variáveis x e y , é possível determinar a equação da reta que melhor descreve o comportamento dos dados, a qual é conhecida como equação de regressão. Por meio dessa equação, prevê-se o valor da variável dependente relacionada com um valor fixo da variável independente. A fim de encontrar esta equação de regressão, utiliza-se a técnica de regressão linear simples (FERREIRA, 2015). A equação da reta é dada pela equação 2:

$$y = \alpha + \beta x \quad (2)$$

Em que y é a variável dependente, α é o coeficiente linear, β é o coeficiente angular e x é a variável independente.

Ao fazer o gráfico a partir da reta obtida, é pouco provável que os pontos gerados se apresentem em uma linha perfeita, por mais que seja possível observar um certo alinhamento. Os desalinhamentos são vistos como desvios do comportamento dos dados. Por isso, é necessário realizar um ajuste de uma linha reta a pontos que não estão em um completo alinhamento. Propõe-se, então, a adição de um novo item à equação, ε , responsável por representar a diferença entre o valor observado e o valor esperado, gerando a equação 3 (CALLEGARI-JACQUES, 2007):

$$y = \alpha + \beta x + \varepsilon \quad (3)$$

A reta de regressão verdadeira seria apenas gerada se se conhecessem os valores de x e de y de toda a população. Se isto acontecesse, seriam conhecidos o coeficiente linear α e a inclinação real da reta, dada pelo coeficiente angular β . Porém, comumente se é estudada somente a regressão entre x e y usando uma certa amostra da população. Assim, calcula-se a e b , sendo estes as estimativas dos parâmetros α e β . Estes valores são gerados pelo método dos mínimos quadrados, responsável por garantir que a reta gerada seja aquela para a qual se observa as menores distâncias (ao quadrado) entre os valores observados e a própria reta. O coeficiente b e o coeficiente a são calculados da seguinte maneira (CALLEGARI-JACQUES, 2007):

$$b = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (4)$$

$$a = \frac{\sum y_i - b \sum x_i}{n} = \bar{y} - b\bar{x} \quad (5)$$

Em que x_i são os valores para cada concentração (i), y_i são os valores mensurados para cada concentração (i), n é o número de observações, \bar{y} é a média para y e \bar{x} é a média para x . Portanto, a reta estimada de regressão será:

$$\hat{y} = a + bx \quad (6)$$

Se tratando da curva obtida pelo ajuste dos dados, quando esta se assemelha a uma reta e os dados estão próximos a ela, é provável que há uma correlação linear entre as variáveis em questão. Quando isto ocorre, determina-se o coeficiente de correlação representado pela letra r , e seu valor pode variar de -1 a $+1$, indicando, assim, se a correlação é positiva ou negativa. O coeficiente de correlação, portanto, tem como função medir o ajuste da reta aos pontos que a geraram, ou seja, qual é a proximidade destes pontos em relação à reta. Dito isso, a fórmula para o cálculo do r é dada por (DORIA FILHO, 1999):

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (7)$$

O método mais comum para testar a correlação linear é por meio do coeficiente de determinação, também conhecido como R^2 , o qual se dá a partir do quadrado do r ,

mencionado acima. Para linearidade absoluta, R^2 é igual a 1. Portanto, quanto menor o valor de R^2 , maior é o desvio da linearidade (ALFASSI *et al.*, 2005).

3.3 Variância residual

Outro parâmetro estatístico de extrema importância para a análise de regressão linear é a variância residual, dada por s_e^2 . Esta se trata de uma medida da dispersão dos resultados em torno da reta de regressão. Assim, ela é responsável por representar a variância na resposta que não pode ser contabilizada ou explicada pela reta de regressão. Caso o modelo linear proposto esteja correto, s_e^2 é uma estimativa da variância das medições e pode ser calculada a partir da seguinte equação (TRIOLA, 2017):

$$s_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} \quad (8)$$

Utilizando a variância residual calculada, é possível calcular outra variância, s_r^2 , que representa a repetibilidade da medição de cada amostra. Essa variância é utilizada quando se deseja comparar séries de medidas com números de replicatas diferentes e sua fórmula é dada por (HARTMANN, 1994):

$$s_r^2 = \frac{s_e^2}{b^2} \left(\frac{1}{m} \right) \quad (9)$$

Em que s_e^2 é a variância residual, b é a estimativa da inclinação da reta e m é o número de replicatas por concentração.

3.4 Erro da previsão

O erro da previsão pode ser calculado da seguinte forma:

$$\hat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (10)$$

Em que \hat{y}_0 é o valor estimado em questão, $t_{0,025;n-2}$ é o valor de t crítico, s_e é o desvio padrão residual, m é número de replicatas por concentração, n é o número de observações, x_i são os valores para as concentrações (i) e \bar{x} é valor de x médio. A partir

dessa equação, é possível calcular o intervalo de confiança para a previsão feita através da regressão linear (MASSART *et al.*, 1998).

3.5 Teste de Grubb's

Dito tudo isto, antes da realização de qualquer teste estatístico para obter a resposta desejada, é necessário verificar se há algum valor aberrante ou *outlier*, em inglês, no conjunto de dados. Um dos testes mais comuns para este fim é o teste de Grubb's simples, utilizado para se detectar um único *outlier*. A equação geral para o teste de Grubb's é dada por (OLIVEIRA *et al.*, 2014):

$$G = \frac{|(VS - \bar{x})|}{s} \quad (11)$$

Em que VS é o valor suspeito de ser aberrante, \bar{x} é a média e s é o desvio padrão dos valores mensurados. O valor calculado de G é comparado com o valor crítico para o teste de Grubb's, para um certo valor de n (observação). Quando o valor calculado é maior do que o valor crítico, significa que o valor suspeito é aberrante. Caso o valor calculado seja menor, o valor suspeito não é aberrante.

3.6 Teste de Cochran

O teste de Cochran é responsável por avaliar o desvio bilateral das variâncias dos dados a um nível de significância de 5%, ou seja, testa a homo e heterocedasticidade dos dados. A equação geral para o teste de Cochran é dada por (PINTO, 2019):

$$C = \frac{s_{max}^2}{\sum_{i=1}^m s_i^2} \quad (12)$$

Em que s_{max}^2 é o maior valor de variância (variância máxima) dentre as variâncias das concentrações usadas na construção da curva analítica e $\sum_{i=1}^m s_i^2$ é o somatório de todas as variâncias. O valor calculado de C é comparado com o valor crítico para o teste de Cochran para um certo valor de m (replicata para cada nível de concentração). Quando o valor calculado é maior do que o valor crítico, significa que as variâncias não são iguais, ou seja, há um comportamento heterocedástico, necessitando uma transformação para homocedástico, para que a regressão linear possa ser realizada e o modelo estimado neste processo tenha um melhor poder de previsão. Caso o valor calculado seja menor, as

variâncias são significativamente iguais, demonstrando um comportamento homocedástico.

4 METODOLOGIA

4.1 Seleção dos testes matemáticos e estatísticos a serem estudados

O teste de hipótese para amostras pareadas ou emparelhadas foi escolhido para ser melhor entendido porque, ao utilizá-lo em dados experimentais, foi observado que, em conjuntos de dados com resultados muito semelhantes, o resultado do teste t-pareado era contrário ao resultado de igualdade esperado, algo que não é intuitivo e, nem sempre, verdadeiro. O teste é aplicado em situações muito específicas, nas quais as amostras são diferentes, o que é menos comum no uso em química analítica e, conseqüentemente, como é menos usado, é menos entendido e menos estudado.

Já o uso da regressão linear, pelo método dos mínimos quadrados, na construção de curvas analíticas foi selecionado porque na construção destas curvas, é comum que sejam usadas as médias das replicatas de cada concentração, em vez de todas as replicatas, o que pode levar a resultados de previsão imprecisos e inexatos. Assim, o entendimento de como o uso da média das replicatas pode impactar a precisão e exatidão dos resultados torna-se um assunto interessante para demonstrar a melhor forma de construir uma curva analítica.

4.2 Estudo do teste t-pareado

Para entender como os valores dos dados usados em um teste t-pareado interferem no seu resultado, foi gerada uma tabela (Apêndice A, página 51), em Excel, relacionando o número de observações (n) começando por n = 2 até n = 101, o grau de liberdade (v), os valores de t crítico bilateral para nível de confiança de 95% (Anexo A, página 54) e os valores da divisão do valor do módulo da média da diferença dos dados ($|\bar{X}_d|$) pelo valor do desvio padrão da diferença dos dados (s_d).

A fim de encontrar estes últimos valores mencionados, foi utilizada a equação 1 (página 20) para o t calculado. Sabendo que, para a hipótese nula do teste t-pareado ser aceita e, portanto, o resultado de igualdade ser aceito, o valor de t crítico deve ser maior do que o valor de t calculado, ou seja:

$$t \text{ crítico} > |t|_{\text{calculado}} = \frac{|\bar{X}_d|\sqrt{n}}{s_d} \quad (13)$$

Logo:

$$\frac{t_{\text{crítico}}}{\sqrt{n}} > \frac{|\bar{X}_d|}{S_d} \quad (14)$$

Portanto, como tanto os valores de t crítico quanto os de n sempre são conhecidos, é possível encontrar um valor para qual $\frac{|\bar{X}_d|}{S_d}$ deva ser menor, indicando, assim, que t crítico é maior que t calculado e a hipótese nula será aceita.

Em seguida, foram simulados conjuntos de dados com valores semelhantes para serem utilizados como estudos de caso (1 e 2), para que a relação matemática obtida anteriormente seja demonstrada para as condições em que o teste t-pareado forneça resultados concordantes e discordantes entre os grupos de dados testados. Em seguida, estes dados foram analisados no Excel, selecionando a função do Teste-t: duas amostras em par para médias na Análise de Dados, a fim de demonstrar que a relação matemática encontrada cumpre seu papel. Estes dados não possuem valores aberrantes.

- Estudo de caso 1: neste caso, simula-se a comparação de dois métodos analíticos (método 1 e método 2), com uma análise de oito amostras, ligeiramente diferentes entre cada método, de modo a determinar se ambos os métodos fornecem a mesma resposta estatística. Na Tabela 1 é apresentado o conjunto de dados utilizado no estudo de caso 1. Todos os valores são fictícios e não representam um exemplo real.

Tabela 1 - Conjunto de dados do estudo de caso 1.

Amostra	Método 1	Método 2
1	0,20	0,18
2	0,14	0,15
3	0,23	0,25
4	0,12	0,10
5	0,21	0,20
6	0,15	0,13
7	0,27	0,23
8	0,20	0,21

- Estudo de caso 2: analogamente, é simulada uma nova comparação entre os mesmos dois métodos analíticos (método 1 e método 2) utilizados no estudo de caso 1, com a análise das mesmas oito amostras, obtendo resultados bastante semelhantes aos valores

apresentados no estudo de caso 1. Da mesma forma, busca-se determinar se ambos os métodos fornecem a mesma resposta estatística. Na Tabela 2 é apresentado o conjunto de dados utilizado no estudo de caso 2. Todos os valores são fictícios e não representam um exemplo real.

Tabela 2 - Conjunto de dados do estudo de caso 2.

Amostra	Método 1	Método 2
1	0,22	0,18
2	0,14	0,15
3	0,24	0,25
4	0,12	0,07
5	0,21	0,15
6	0,15	0,11
7	0,27	0,22
8	0,20	0,21

4.3 Estudo da influência do uso das replicatas e da média das replicatas na construção da curva analítica obtida através da regressão linear

Para compreender como o uso de concentrações em replicata e com a média das replicatas influencia na regressão linear pelo método dos mínimos quadrados, foram construídas, em Excel, três tipos de curva analítica: 1) com todas as replicatas das concentrações, 2) com a média das replicatas das concentrações e 3) com números diferentes de replicatas para as concentrações C1, C3 e C5. Em primeiro lugar, foram simulados dados para x (concentração) e para y (leitura do método) e, para uma melhor análise do efeito do número de replicatas e de sua média, estes dados foram analisados em triplicatas (número de replicatas igual a 3) e em quintuplicatas (número de replicatas igual a 5). Após a construção das curvas analíticas, estas foram comparadas em relação ao coeficiente de determinação, equação da reta e variância residual.

4.3.1 Concentrações em triplicata, com a média das triplicatas e com diferentes números de replicata para algumas concentrações

A primeira curva analítica foi construída utilizando todos os valores das triplicatas. Estes dados podem ser vistos na Tabela 3:

Tabela 3 – Valores das concentrações de C1 até C5 (x) e das leituras (y) em triplicata.

	Concentração		Leitura	
	x		y	
C1	10,0	20,7	21,6	21,3
C2	20,0	44,4	44,8	44,6
C3	30,0	62,1	61,6	61,7
C4	40,0	77,7	78,2	78,1
C5	50,0	105,4	104,9	105,3

Já para a segunda curva analítica construída, foi utilizada a média das triplicatas, e os dados estão apresentados na Tabela 4.

Tabela 4 – Valores das concentrações de C1 até C5 (x) e das leituras (y) para a média das triplicatas.

	Concentração		Leitura
	x		y
C1	10,0		21,2
C2	20,0		44,6
C3	30,0		61,8
C4	40,0		78,0
C5	50,0		105,2

Também foi construída uma curva analítica a partir da combinação de valores de triplicatas (C1, C3 e C5), em negrito, e duplicatas (C2 e C4), como pode ser visto na Tabela 5. Para as duplicatas, foram utilizados os dois primeiros valores medidos de concentração, retirados da Tabela 3, ignorando os três últimos valores.

Tabela 5 – Valores das concentrações de C1 até C5 (x) e das leituras (y) para as triplicatas C1, C3 e C5 e duplicatas C2 e C4.

	Concentração		Leitura	
	x		y	
C1	10,0	20,7	21,6	21,3
C2	20,0	44,4	44,8	-
C3	30,0	62,1	61,6	61,7
C4	40,0	77,7	78,2	-
C5	50,0	105,4	104,9	105,3

A partir dos dados apresentados nas Tabelas 3 a 5 foram construídas suas respectivas curvas analíticas, obtendo-se suas respectivas equações da reta (página 21), coeficientes de determinação (R^2) (página 22) e variância residual (página 23), e os resultados estão apresentados no item 5.2.1 (página 37).

Antes de construir as curvas analíticas para cada uma dessas três situações, foi aplicado o teste de Grubb's e o teste de Cochran, como descrito nos itens 3.5 e 3.6, respectivamente.

4.3.2 Concentrações em quintuplicata, com a média das quintuplicatas e com diferentes números de replicata para algumas concentrações

Neste tópico, a primeira curva analítica foi construída utilizando todos os valores das quintuplicatas. Estes dados podem ser vistos na Tabela 6.

Tabela 6 – Valores das concentrações de C1 até C5 (x) e das leituras (y) em quintuplicata.

	Concentração		Leitura			
	x		y			
C1	10,0	20,7	21,6	21,3	20,9	21,4
C2	20,0	44,4	44,8	44,6	43,9	44,9
C3	30,0	62,1	61,6	61,7	61,9	62,3
C4	40,0	77,7	78,2	78,1	77,9	78,4
C5	50,0	105,4	104,9	105,3	104,7	105,6

Já para a segunda curva analítica construída, foi utilizada a média das quintuplicatas. e os dados estão apresentados na Tabela 7:

Tabela 7 – Valores das concentrações de C1 até C5 (x) e das leituras (y) para a média das quintuplicatas.

	Concentração		Leitura	
	x		y	
C1	10,0		21,2	
C2	20,0		44,5	
C3	30,0		61,9	
C4	40,0		78,1	
C5	50,0		105,2	

Também foi construída uma curva analítica a partir da combinação de valores de quintuplicatas (C1, C3 e C5), em negrito, e duplicatas (C2 e C4), conforme pode ser visto na Tabela 8. Para as duplicatas, foram utilizados os dois primeiros valores de concentração, retirados da Tabela 6, ignorando os três últimos valores,

Tabela 8 – Valores das concentrações de C1 até C5 (x) e das leituras (y) para as quintuplicatas C1, C3 e C5 e duplicatas C2 e C4.

	Concentração			Leitura		
	x			y		
C1	10,0	20,7	21,6	21,3	20,9	21,4
C2	20,0	44,4	44,8	-	-	-
C3	30,0	62,1	61,6	61,7	61,9	62,3
C4	40,0	77,7	78,2	-	-	-
C5	50,0	105,4	104,9	105,3	104,7	105,6

A partir destas tabelas apresentadas foram construídas suas respectivas curvas analíticas e foram determinadas suas respectivas equações da reta (página 21), coeficientes de determinação (R^2) (página 22) e variância residual (página 23), e os resultados estão apresentados no item 5.2.2 (página 42).

Antes de construir as curvas analíticas para cada uma dessas três situações, foi aplicado o teste de Grubb's e o teste de Cochran, como descrito nos itens 3.5 e 3.6, respectivamente.

5 RESULTADOS E DISCUSSÃO

5.1 Estudo do teste t-pareado

A partir dos cálculos realizados no Excel, utilizando os dados mencionados no item 4.2 (página 26), foi possível realizar a determinação de valores para os quais a hipótese nula para o teste t-pareado é aceita, ou seja, quando o valor de t crítico é maior que o valor de t calculado. Este resultado pode ser visto por completo (n = 2 até n = 101) no Apêndice A (página 51) e parte dele (n = 2 até n = 20) pode ser observado na Tabela 9.

Tabela 9 – Valores utilizados para que a hipótese nula seja aceita, no teste t-pareado.

n	v	t(crítico) bi 95%	Para t(crítico) > t(calculado):
2	1	12,706	$8,984 > \bar{X}_d /S_d$
3	2	4,303	$2,484 > \bar{X}_d /S_d$
4	3	3,182	$1,591 > \bar{X}_d /S_d$
5	4	2,776	$1,241 > \bar{X}_d /S_d$
6	5	2,571	$1,050 > \bar{X}_d /S_d$
7	6	2,447	$0,925 > \bar{X}_d /S_d$
8	7	2,365	$0,836 > \bar{X}_d /S_d$
9	8	2,306	$0,769 > \bar{X}_d /S_d$
10	9	2,262	$0,715 > \bar{X}_d /S_d$
11	10	2,228	$0,672 > \bar{X}_d /S_d$
12	11	2,201	$0,635 > \bar{X}_d /S_d$
13	12	2,179	$0,604 > \bar{X}_d /S_d$
14	13	2,160	$0,577 > \bar{X}_d /S_d$
15	14	2,145	$0,554 > \bar{X}_d /S_d$
16	15	2,131	$0,533 > \bar{X}_d /S_d$
17	16	2,120	$0,514 > \bar{X}_d /S_d$
18	17	2,110	$0,497 > \bar{X}_d /S_d$
19	18	2,101	$0,482 > \bar{X}_d /S_d$
20	19	2,093	$0,468 > \bar{X}_d /S_d$

n: número de observações; v: grau de liberdade

A partir das equações 13 e 14 apresentadas no item 4.2 (página 26) e utilizando os valores de n e t crítico bilateral demonstrados na Tabela 9, pode-se encontrar os números apresentados na última coluna desta mesma tabela, sendo estes o verdadeiro resultado, visto que, por meio deles, é possível analisar se a hipótese nula será aceita ou rejeitada. Tomando o $n = 2$ como exemplo, caso a divisão do módulo da média da diferença dos valores obtidos pelo analista pelo desvio padrão dessa mesma diferença seja menor que 8,984, quer dizer que o valor de t crítico para este valor de n é maior que o valor de t calculado, ou seja, a hipótese nula é aceita. Caso a divisão gere um valor maior que 8,984, quer dizer que o valor de t crítico é menor que o valor de t calculado, o que significa que a hipótese nula é rejeitada. A seguir estão apresentados dois estudos de caso para que seja feita a demonstração da funcionalidade da relação matemática encontrada.

- Estudo de caso 1: para a criação deste estudo de caso, foram usados os dados apresentados na Tabela 1 (item 4.2, página 26). Entretanto, para a realização do teste t -pareado, é necessário saber o valor da diferença (d) dos resultados simulados para cada método presentes na Tabela 1 e, portanto, para isso, foi criada a Tabela 10.

Tabela 10 – Conjunto de dados do estudo de caso 1 com o valor das diferenças (d) dos resultados obtidos para cada método.

Amostra	Método 1	Método 2	d
1	0,20	0,18	0,02
2	0,14	0,15	-0,01
3	0,23	0,25	-0,02
4	0,12	0,10	0,02
5	0,21	0,20	0,01
6	0,15	0,13	0,02
7	0,27	0,23	0,04
8	0,20	0,21	-0,01

Estes valores de diferença foram utilizados para que pudessem ser encontrados os valores do módulo da média das diferenças e o desvio padrão das diferenças, uma vez que estes valores são necessários para o cálculo do valor de t calculado, como se pode ver na equação 1 (página 20). Portanto, os valores obtidos pelo Excel foram:

- Módulo da média das diferenças: 0,01
- Desvio padrão das diferenças: 0,020

O próximo passo para saber se este estudo de caso tem como resultado a hipótese nula aceita ou rejeitada é fazer a divisão entre os números mencionados acima, o que resulta no valor de 0,5. Utilizando a Tabela 9, com $n = 8$, observa-se que, para que t crítico seja maior que t calculado, ou seja, para que a hipótese nula seja aceita, o valor da divisão deve ser menor 0,836, o que de fato acontece. Este resultado pode ser corroborado pela análise feita pelo Excel, a qual foi gerada pela Análise de Dados, selecionando a opção do Teste-t: duas amostras em par para médias, obtendo a Tabela 11.

Tabela 11 – Resultados da aplicação do teste t-pareado, obtidos pelo Excel a partir do conjunto de dados do estudo de caso 1.

	Variável 1	Variável 2
Média	0,19	0,18
Variância	$2,51 \times 10^{-3}$	$2,64 \times 10^{-3}$
Observações	8	8
ν	7	
Stat t	1,218	
t crítico bi-caudal ($\alpha = 0,05$)	2,365	

A partir dos resultados apresentados na Tabela 11, observa-se, destacado em negrito e em verde, que o valor de Stat t (t calculado) é menor do que o valor de t crítico bicaudal, ou seja, a hipótese nula é aceita e os métodos são estatisticamente equivalentes, o que ratifica o resultado encontrado anteriormente.

- Estudo de caso 2: analogamente ao estudo de caso 1, a Tabela 12 acrescenta a coluna dos valores da diferença (d) dos resultados simulados para cada método na Tabela 2:

Tabela 12 - Conjunto de dados do estudo de caso 2 com o valor das diferenças (d) dos resultados obtidos para cada método.

Amostra	Método 1	Método 2	d
1	0,22	0,18	0,04
2	0,14	0,15	-0,01
3	0,24	0,25	-0,01
4	0,12	0,07	0,05
5	0,21	0,15	0,06
6	0,15	0,11	0,04
7	0,27	0,22	0,05
8	0,20	0,21	-0,01

Já para este estudo de caso, os valores encontrados a partir da diferença foram:

- Módulo da média das diferenças: 0,03
- Desvio padrão das diferenças: 0,031

Fazendo a divisão destes valores, tem-se 0,968 como resultado. Como foi visto anteriormente, o valor de t crítico para $n = 8$ é 0,836, ou seja, neste caso, o valor encontrado na divisão é maior que o valor de t crítico, o que quer dizer que a hipótese nula não será aceita, obtendo um resultado contrário ao encontrado no estudo de caso 1, apesar de seus valores serem semelhantes entre si.

Este resultado pode ser corroborado pela análise feita pelo Excel, o qual foi obtido pela Análise de Dados, selecionando a opção do Teste-t: duas amostras em par para médias, gerando a Tabela 13:

Tabela 13 – Resultados da aplicação do teste t-pareado, obtidos pelo Excel, a partir do conjunto de dados do estudo de caso 2.

	Variável 1	Variável 2
Média	0,19	0,17
Variância	$2,74 \times 10^{-3}$	$3,57 \times 10^{-3}$
Observações	8	8
ν	7	
Stat t	2,421	
t crítico bi-caudal ($\alpha = 0,05$)	2,365	

A partir dos resultados apresentados na Tabela 13, observa-se, destacado em negrito e em vermelho, que o valor de Stat t (t calculado) é maior do que o valor de t crítico bicaudal, ou seja, a hipótese nula é rejeitada e os métodos não são estatisticamente equivalentes, o que ratifica o resultado encontrado anteriormente.

5.2 Estudo da influência do uso das replicatas e da média das replicatas na construção da curva analítica obtida através da regressão linear

A partir dos dados apresentados nas Tabelas 3 a 8 no item 4.3 (páginas 29-31), foi possível construir uma curva analítica para cada um destes conjuntos de dados, com suas

respectivas particularidades em relação às replicatas e à média das replicatas, através da regressão linear. Em todas as curvas, o eixo y representa a leitura do método e o eixo x representa as concentrações, porém, como os valores foram simulados, não há uma unidade específica para as concentrações neste estudo. Em seguida, foram determinados o coeficiente de determinação (R^2), a equação da reta e a variância residual de cada curva, a título de comparação. Entretanto, em primeiro lugar, foram realizados o teste de Grubb's e o teste de Cochran para a curva analítica com todas as triplicatas e com todas as quintuplicatas das concentrações e não foram detectados valores aberrantes, assim como observou-se comportamentos homocedásticos em ambas as curvas. Os resultados da aplicação do teste de Grubb's e do teste de Cochran estão apresentados a seguir.

- Teste de Grubb's para todas as triplicatas:

Tabela 14 – Valores de G calculado para cada concentração em triplicata.

	Concentração	G calc para o menor y	G calc para o maior y
C1	10,0	1,09	0,873
C2	20,0	1,00	1,00
C3	30,0	1,13	0,756
C4	40,0	1,13	0,756
C5	50,0	0,756	1,13

O valor tabelado de G para 3 observações e nível de significância 5% é 1,153, como se pode ver no Anexo B (página 55). Como todos os valores presentes na tabela são menores que 1,153, pode-se concluir que não há valores aberrantes.

- Teste de Cochran para todas as triplicatas:

Ao calcular a variância máxima das concentrações com todas as triplicatas, encontrou-se 0,140 e, somando todas as variâncias, obteve-se 0,307. Dividindo a variância máxima pelo somatório de todas as variâncias, encontrou-se, assim, o valor de C calculado, o qual é 0,456. Como o valor de C tabelado (Anexo C, página 56) para 5 observações e 3 replicatas é 0,6838, conclui-se que o valor de C tabelado é maior que o valor de C calculado, ou seja, o comportamento da curva é homocedástico.

- Teste de Grubb's para todas as quintuplicatas:

Tabela 15 – Valores de G calculado para cada concentração em quintuplicata.

	Concentração	G calc para o menor y	G calc para o maior y
C1	10,0	1,30	1,13
C2	20,0	0,303	0,959
C3	30,0	1,12	1,33
C4	40,0	1,33	1,26
C5	50,0	1,30	1,13

O valor tabelado de G para 5 observações e nível de significância 5% é 1,671, como se pode ver no Anexo B (página 55). Como todos os valores presentes na tabela são menores que 1,671, pode-se concluir que não há valores aberrantes.

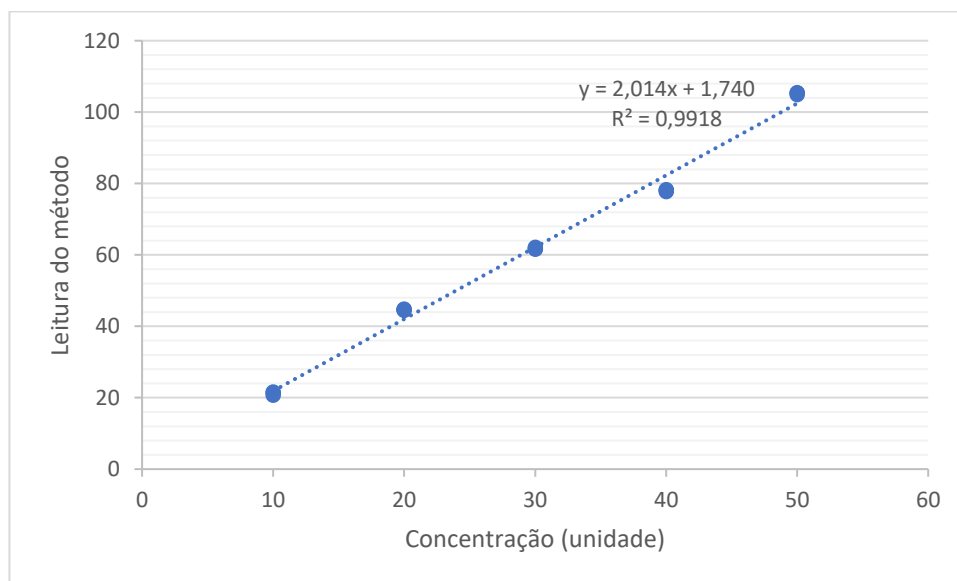
- Teste de Cochran para todas as quintuplicatas:

Ao calcular a variância máxima das concentrações com todas as quintuplicatas, encontrou-se 0,126 e, somando todas as variâncias, obteve-se 0,469. Dividindo a variância máxima pelo somatório das variâncias, encontrou-se, assim, o valor de C calculado, o qual é 0,268. Como o valor de C tabelado (Anexo C, página 56) para 5 observações e 5 replicatas é 0,5441, conclui-se que o valor de C tabelado é maior que o valor de C calculado, ou seja, o comportamento da curva é homocedástico.

5.2.1 Concentrações em triplicata, com a média das triplicatas e com diferentes números de replicata para algumas concentrações

Foi construída uma curva analítica utilizando todos os valores de concentração (C1, C2, C3, C4 e C5) e suas leituras em triplicatas, com seus respectivos valores de y (ver Tabela 3, página 29), o que gerou a curva analítica apresentada na Figura 2.

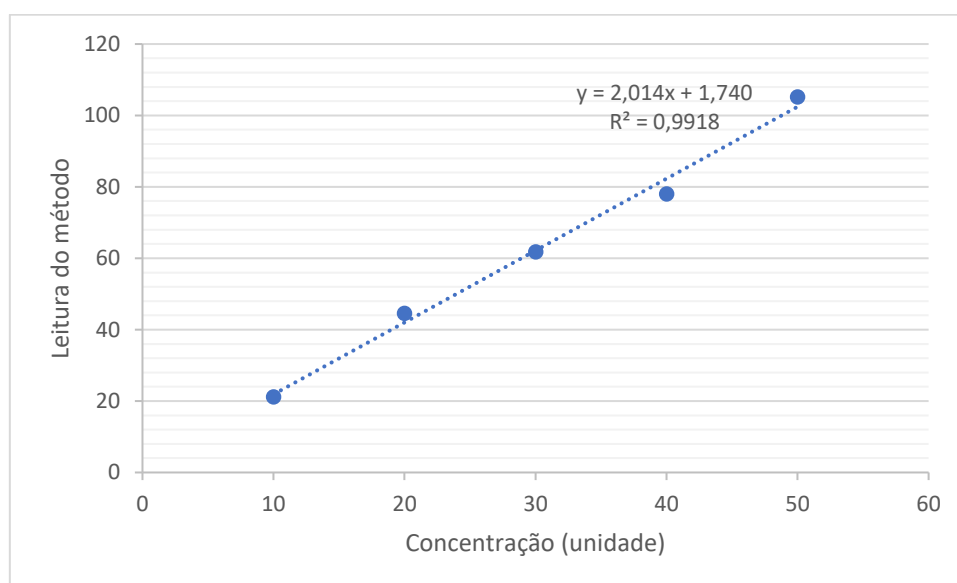
Figura 2 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) de 5 concentrações diferentes (C1, C2, C3, C4 e C5) em triplicata.



Na Figura 2, pode ser observado o valor calculado para o coeficiente de determinação (R^2) de 0,9918 e a equação da reta é dada por $y = 2,014x + 1,740$. Além disso, foi determinada sua variância residual, obtendo um valor de 7,768.

A segunda curva analítica foi construída utilizando a média das triplicatas de cada uma das 5 concentrações (C1, C2, C3, C4 e C5) (ver Tabela 4, página 29), gerando a curva analítica apresentada na Figura 3.

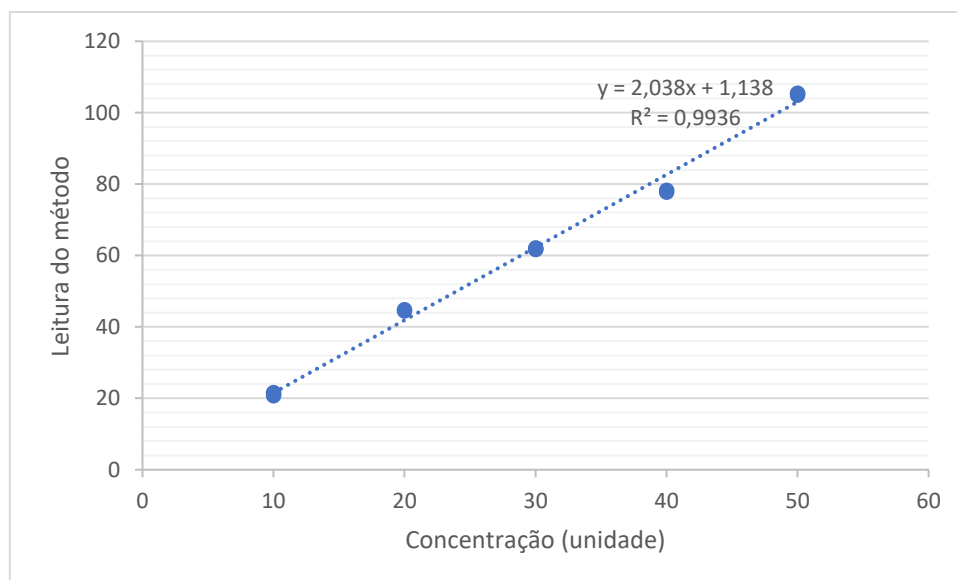
Figura 3 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) da média das triplicatas das 5 concentrações diferentes (C1, C2, C3, C4 e C5).



Na Figura 3, pode ser observado o valor calculado para o coeficiente de determinação (R^2) de 0,9918 e a equação da reta é dada por $y = 2,014x + 1,740$, assim como na curva analítica com todas as triplicatas. Entretanto, a variância residual utilizando a média das triplicatas foi de 11,119, valor superior ao valor da variância residual encontrada quando se utiliza todas as triplicatas (7,768). Dessa forma, por mais que os valores do coeficiente de determinação e da equação da reta sejam iguais nos dois casos, o mesmo não ocorre quando se trata da variância residual, sendo o modelo mais favorável, nesse caso, quando se utiliza todas as triplicatas, devido à obtenção de uma variância residual com menor valor. Isso se dá porque a variância residual e o erro da previsão são parâmetros diretamente proporcionais, ou seja, é primordial que o valor da variância residual seja mais baixo, pois leva a um valor de erro da previsão também mais baixo e, por conseguinte, um resultado mais exato (MASSART *et al.*, 1998).

A terceira curva analítica foi construída utilizando valores de triplicata das concentrações C1, C3 e C5 e duplicatas das concentrações C2 e C4 (ver Tabela 5, página 30), gerando a curva analítica apresentada na Figura 4.

Figura 4 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) com triplicatas das concentrações C1, C3 e C5 e duplicatas das concentrações C2 e C4.



Na Figura 4, pode ser observado o valor calculado para o coeficiente de determinação (R^2) de 0,9936 e a equação da reta é dada por $y = 2,038x + 1,138$. Além disso, foi determinado sua variância residual, através do Excel, obtendo um valor de 6,804. Observa-se que este valor de variância residual é inferior ao valor obtido quando se utiliza todas as concentrações em triplicata (7,768), e o seu valor do coeficiente de

determinação é maior, o que demonstra que, neste caso, é gerado um resultado ainda mais satisfatório se comparado ao uso de todas as concentrações em triplicata. Isso se deve porque as concentrações principais e mais importantes de uma curva analítica são as concentrações inicial (C1), central (C3) e final (C5), que são justamente as concentrações que estão com mais peso neste caso.

A fim de melhor analisar os resultados mencionados anteriormente, os resultados obtidos foram organizados na Tabela 16, com os valores obtidos de equação da reta, coeficiente de determinação (R^2) e variância residual:

Tabela 16 – Comparação dos resultados de equação da reta, coeficiente de determinação e variância residual para cada curva analítica construída com 5 concentrações diferentes, C1, C2, C3, C4 e C5 (a) todas as 5 concentrações em triplicata, (b) média das triplicatas de cada concentração e (c) triplicata das concentrações C1, C3 e C5 e duplicata das concentrações C2 e C4.

Curva analítica	Equação	R^2	Variância residual
a)Todas as triplicatas	$y = 2,014x + 1,740$	0,9918	7,768
b)Média das triplicatas	$y = 2,014x + 1,740$	0,9918	11,119
c)Triplicata C1, C3, C5	$y = 2,038x + 1,138$	0,9936	6,804

Na Tabela 16, observa-se que as equações da reta são exatamente iguais quando são utilizadas todas as triplicatas e a média das triplicatas, assim como seus respectivos valores de R^2 . O mesmo não acontece com a equação quando são usados valores das triplicatas das concentrações C1, C3 e C5, além do R^2 obtido ser ligeiramente maior. Em relação à variância residual, o seu maior valor é visto quando se utiliza a média das triplicatas, tornando este modelo possivelmente menos favorável porque, quanto maior a variância residual em torno de uma medida, menor é a precisão e, conseqüentemente, a exatidão do resultado. Entre a utilização de todas as triplicatas das concentrações e apenas a triplicata das concentrações C1, C3 e C5, observa-se que este último possui uma menor variância residual, podendo ser, então, mais adequado para a obtenção de um resultado estatístico mais preciso.

Em relação ao cálculo do erro da previsão para o melhor modelo (triplicata das concentrações C1, C3 e C5 e duplicata das concentrações C2 e C4), utilizando valores de x presentes na faixa de concentração utilizada, como 15,0 e 45,0, encontra-se como valores de y 31,7 e 92,9, respectivamente. A partir destes valores, pode-se calcular os intervalos de confiança para cada um deles através da equação 10 (item 3.4, página 23):

Cálculo para as triplicatas C1, C3 e C5:

→ $x = 15,0$; $y = 31,7$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,608$; $m = 3$; $n = 5$; $x_0 = 15,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 31,7 \pm 9,49$$

Portanto, o intervalo de confiança é de 22,21 a 41,19.

→ $x = 45,0$; $y = 92,9$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,608$; $m = 3$; $n = 5$; $x_0 = 45,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 92,9 \pm 9,49$$

Portanto, o intervalo de confiança é de 83,41 a 102,39.

Cálculo para as duplicatas C2 e C4:

→ $x = 15,0$; $y = 31,7$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,608$; $m = 2$; $n = 5$; $x_0 = 15,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 31,7 \pm 10,5$$

Portanto, o intervalo de confiança é de 21,2 a 42,2.

→ $x = 45,0$; $y = 92,9$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,608$; $m = 2$; $n = 5$; $x_0 = 45,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 92,9 \pm 10,5$$

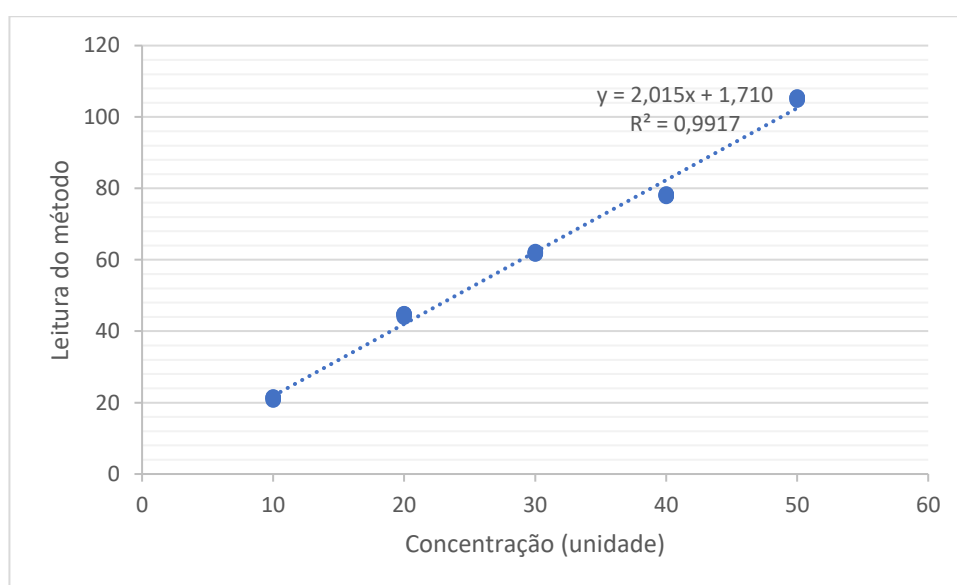
Portanto, o intervalo de confiança é de 82,4 a 103,4.

Como no modelo da curva analítica analisada utilizou-se triplicatas das concentrações C1, C3 e C5 e duplicatas em C2 e C4, o intervalo de confiança desta curva está entre os valores de intervalo de confiança calculados acima, ou seja, entre 9,49 e 10,5, valores relativamente grandes. Para $x = 15,0$, os intervalos de confiança correspondem a 30% e 33% do valor de y , respectivamente. Já para $x = 45,0$, estes intervalos correspondem a 10% e 11% do valor de y , respectivamente. Estes resultados serão comparados com os valores calculados para as quintuplicatas, na página 45.

5.2.2 Concentrações em quintuplicata, com a média das quintuplicatas e com diferentes números de replicata para algumas concentrações

Para ratificar o estudo realizado sobre triplicatas, foi feito, analogamente, o mesmo estudo, porém, dessa vez, envolvendo quintuplicatas. Em primeiro lugar, foi construída uma curva analítica utilizando todos os valores de concentração (C1, C2, C3, C4 e C5) e suas leituras em quintuplicatas, com seus respectivos valores de y (ver Tabela 6, página 30), o que gerou a curva analítica apresentada na Figura 5:

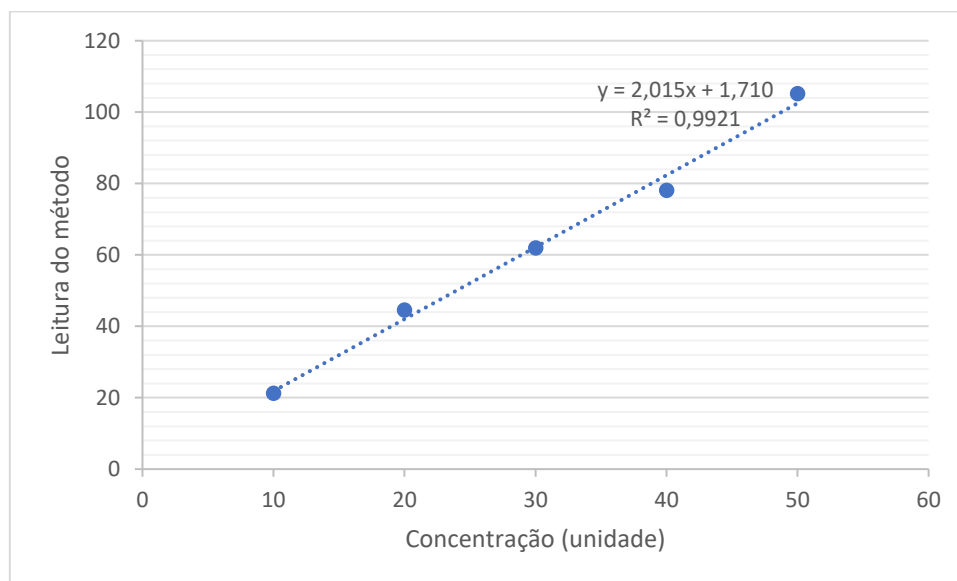
Figura 5 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) de 5 concentrações diferentes (C1, C2, C3, C4 e C5) em quintuplicata.



Na Figura 5, pode ser observado o valor calculado para o coeficiente de determinação (R^2) de 0,9917 e a equação da reta é dada por $y = 2,015x + 1,710$. Além disso, foi determinado sua variância residual, obtendo um valor de 7,119.

A segunda curva analítica foi construída utilizando a média das quintuplicatas de cada uma das 5 concentrações (C1, C2, C3, C4 e C5) (ver Tabela 7, página 31), gerando a curva analítica apresentada na Figura 6:

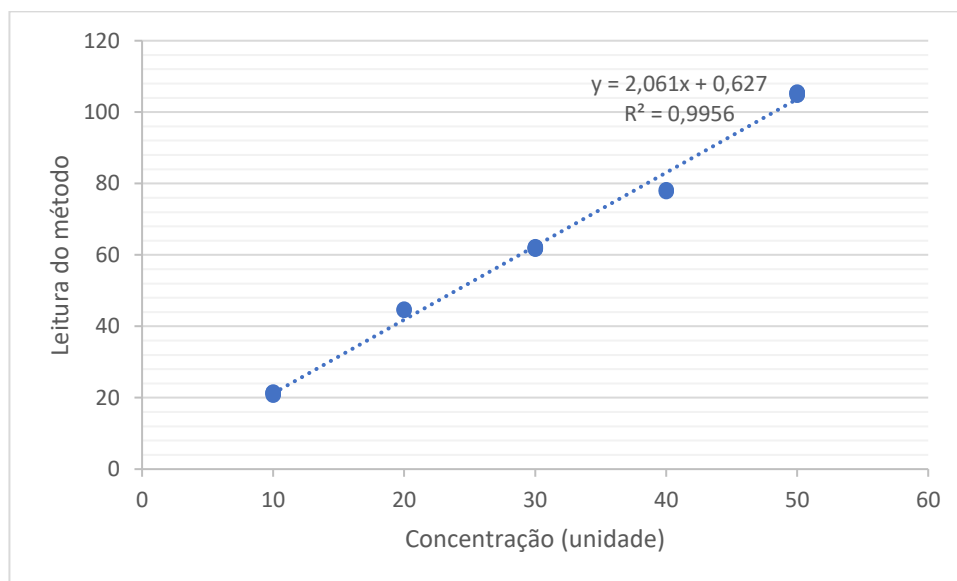
Figura 6 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) da média das triplicatas das 5 concentrações diferentes (C1, C2, C3, C4 e C5).



Na Figura 6, pode ser observado o valor calculado para o coeficiente de determinação (R^2) de 0,9921 e a equação da reta é dada por $y = 2,015x + 1,710$, assim como a equação da reta utilizando todas as quintuplicatas. Entretanto, a variância residual utilizando a média das quintuplicatas foi de 10,760, valor superior ao valor da variância residual encontrada quando se utiliza todas as quintuplicatas (7,119). Dessa forma, por mais que os valores do coeficiente de determinação sejam semelhantes e as equações da reta sejam iguais nos dois casos, o mesmo não ocorre quando se trata da variância residual, sendo mais favorável, neste caso, a utilização de todas as quintuplicatas, devido à obtenção de uma variância residual com menor valor, conforme foi analisado no item 5.2.1, na página 39.

A terceira curva analítica foi construída utilizando valores de quintuplicata das concentrações C1, C3 e C5 e duplicatas das concentrações C2 e C4 (ver Tabela 8, página 31), gerando a curva analítica apresentada na Figura 7:

Figura 7 - Curva analítica obtida com os valores de x (concentração) versus y (leitura do método) com triplicatas das concentrações C1, C3 e C5 e duplicatas das concentrações C2 e C4.



Na Figura 7, pode ser observado o valor calculado para o coeficiente de determinação (R^2) de 0,9956 e sua equação da reta é dada por $y = 2,061x + 0,627$. Além disso, foi determinado sua variância residual, obtendo um valor de 4,818. Observa-se que este valor de variância residual é inferior ao valor obtido quando se utiliza todas as concentrações em quintuplicata (7,119), e o seu valor do coeficiente de determinação é maior, o que demonstra que esta forma de análise possivelmente gera um resultado ainda mais satisfatório se comparado ao uso de todas as concentrações em quintuplicata. Isso se deve porque as concentrações principais e mais importantes de uma curva analítica são as concentrações inicial (C1), central (C3) e final (C5), que são justamente as concentrações que estão com mais peso neste caso.

A fim de melhor analisar os resultados mencionados anteriormente, os resultados obtidos foram organizados na Tabela 17, com os valores obtidos de equação da reta, coeficiente de determinação (R^2) e variância residual:

Tabela 17 – Comparação dos resultados de equação da reta, coeficiente de determinação e variância residual para cada curva analítica construída com 5 concentrações diferentes, C1, C2, C3, C4 e C5 (a) todas as 5 concentrações em quintuplicata, (b) média das quintuplicatas de cada concentração e (c) quintuplicata das concentrações C1, C3 e C5 e duplicata das concentrações C2 e C4.

Curva analítica	Equação	R^2	Variância residual
a) Todas as quintuplicatas	$y = 2,015x + 1,710$	0,9917	7,119
b) Média das quintuplicatas	$y = 2,015x + 1,710$	0,9921	10,760
c) Quintuplicata C1, C3, C5	$y = 2,061x + 0,627$	0,9956	4,818

Na Tabela 17, observa-se que as equações da reta são exatamente iguais quando são utilizadas todas as quintuplicatas e a média das quintuplicatas. O mesmo não acontece com a equação quando são usadas as quintuplicatas das concentrações C1, C3 e C5, além do valor do R^2 obtido ser maior. Em relação à variância residual, o seu maior valor é visto quando se utiliza a média das quintuplicatas, sendo este caso, então, menos favorável porque, quanto maior a variância residual em torno de uma medida, menor é a precisão e, conseqüentemente, a exatidão do resultado. Entre a utilização de todas as quintuplicatas e apenas das concentrações C1, C3 e C5, observa-se que este último possui uma menor variância residual, possivelmente sendo, então, mais adequado para a obtenção de um resultado estatístico mais preciso.

Em relação ao cálculo do erro da previsão para o melhor modelo (quintuplicata das concentrações C1, C3 e C5 e duplicata das concentrações C2 e C4), utilizando valores de x presentes na faixa de concentração utilizada, como 15,0 e 45,0, encontra-se como valores de y 31,5 e 93,4, respectivamente. A partir destes valores, pode-se calcular o intervalo de confiança para cada um deles através da equação 10 (item 3.4, página 23):

Cálculo para as quintuplicatas C1, C3 e C5:

→ $x = 15,0$; $y = 31,5$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,195$; $m = 5$; $n = 5$; $x_0 = 15,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 31,5 \pm 7,25$$

Portanto, o intervalo de confiança é de 24,25 a 38,75.

→ $x = 45,0$; $y = 93,4$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,195$; $m = 5$; $n = 5$; $x_0 = 45,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 93,4 \pm 7,25$$

Portanto, o intervalo de confiança é de 86,15 a 100,65.

Cálculo para as duplicatas C2 e C4:

→ $x = 15,0$; $y = 31,5$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,195$; $m = 2$; $n = 5$; $x_0 = 15,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 31,5 \pm 8,82$$

Portanto, o intervalo de confiança é de 22,68 a 40,32.

→ $x = 45,0$; $y = 93,4$; $t_{0,025;n-2} = 4,1765$; $s_e = 2,195$; $m = 2$; $n = 5$; $x_0 = 45,0$; $\bar{x} = 30,0$; $x_i = 10,0$; $20,0$; $30,0$; $40,0$ e $50,0$:

$$\widehat{y}_0 \pm t_{0,025;n-2} s_e \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 93,4 \pm 8,82$$

Portanto, o intervalo de confiança é de 84,58 a 102,22.

A partir dos cálculos realizados, observa-se que os intervalos de confiança encontrados no uso do modelo com quintuplicatas são menores que os intervalos da curva com as triplicatas. Para $x = 15,0$, os intervalos de confiança correspondem 23% e 28% do valor de y , respectivamente, resultado menor que o encontrado para as triplicatas, o qual está entre 30% e 33%. Já para $x = 45,0$, estes intervalos correspondem a 8% e 9% do valor de y , respectivamente, também menores do que para as triplicatas, o qual está 10% e 11%, gerando um resultado mais satisfatório quando são utilizadas concentrações em quintuplicata, como esperado, visto que há um maior número de repetições, o que leva a um desvio padrão menor e, conseqüentemente, um intervalo de confiança menor.

5.2.3 Comparação entre o uso de triplicata e quintuplicata

Além do cálculo do intervalo de confiança, a fim de saber qual modelo é mais satisfatório, também é necessário calcular a variância relacionada à repetibilidade, a qual é dada pela equação 9 (item 3.3, página 23). À critério de comparação, foram feitos os cálculos para o melhor modelo das triplicatas e para o melhor modelo das quintuplicatas:

- Triplicatas das concentrações C1, C3 e C5:

$$s_r^2 = \frac{s_e^2}{b^2} \left(\frac{1}{m}\right) = \frac{46,29442}{4,154317} \left(\frac{1}{3}\right) = 3,72$$

- Quintuplicatas das concentrações C1, C3 e C5:

$$s_r^2 = \frac{s_e^2}{b^2} \left(\frac{1}{m}\right) = \frac{23,21312}{2,060682} \left(\frac{1}{5}\right) = 2,25$$

A partir da equação, é possível observar que a variância em questão é indiretamente proporcional ao número de replicatas, ou seja, uma maior quantidade de replicatas gera um valor menor de variância e, portanto, o uso de quintuplicatas fornece um resultado menor e, dessa forma, mais satisfatório estatisticamente falando, como esperado, desconsiderando aspectos como tempo, quantidade e custo.

6 CONCLUSÃO

A partir do estudo feito sobre do teste t-pareado, conclui-se que, apesar de certos métodos de análise terem obtido como respostas valores semelhantes, não significa que ambos os casos terão o mesmo resultado, sendo este que a hipótese nula é aceita ou rejeitada, o que não é intuitivo. Independente desta semelhança, o que realmente é necessário para fazer este julgamento é o módulo da média da diferença e o desvio padrão dos valores resultados em cada amostra. Para isso, foi criada Tabela 9 (completa no Apêndice A, página 51), com os valores utilizados para que a hipótese nula seja aceita, como uma avaliação preliminar do teste t-pareado. Esta proposta realizada no trabalho mostrou-se adequada aos exemplos mostrados, sendo útil para antecipar e facilitar o uso do teste t-pareado.

Em relação à influência das replicatas e da média das replicatas na construção das curvas analíticas obtidas por meio da regressão linear, conclui-se que, utilizar a média das replicatas para fazer qualquer tipo de análise pode não ser considerado analiticamente correto, o que pode ser corroborado pelo valor da variância residual ser maior neste caso, gerando um resultado potencialmente mais impreciso, o que não é desejável.

No que se refere a utilizar todas as replicatas em todas as concentrações estudadas (C1, C2, C3, C4 e C5) ou apenas triplicatas/quintuplicatas nas concentrações C1, C3 e C5 e duplicatas nas concentrações C2 e C4, fica claro que este último caso possivelmente gera um resultado mais favorável, visto que o valor de sua variância residual é menor, o que pode não ser intuitivo. Isso se dá pelo fato dos pontos representados pelas concentrações C1, C3 e C5 permitirem definir de forma mais significativa a inclinação da reta.

Já em relação ao número de replicatas por nível de concentração, ficou claro que quanto maior o número de replicatas, menor será o valor da variância residual e, conseqüentemente, maior a exatidão ou poder de previsão do modelo. Deve-se, entretanto, analisar outros aspectos referentes ao projeto, como tempo, quantidade e custo para que seja feita a escolha correta nas condições apresentadas.

REFERÊNCIAS BIBLIOGRÁFICAS

ALFASSI, Z. B.; BOGER, Z.; RONEN, Y. **Statistical Treatment of Analytical Data**. Oxford: Blackwell Publishing, 2005.

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à Inferência Estatística**. 2ª ed. São Paulo: SBM, 2010.

CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artmed, 2007.

DORIA FILHO, U. **Introdução à Bioestatística: para simples mortais**. 1ª ed. Elsevier, 1999.

FERREIRA, V. **Estatística Básica**. 1ª ed. Rio de Janeiro: SESES, 2015.

GUIMARÃES, P. R. B., **Métodos Quantitativos Estatísticos**. 1ª ed. Curitiba: IESDE Brasil S.A., 2008.

HARTMANN, C. Increasing precision by replicate measurements, **Analisis Magazine**, v. 22, n. 5, M19-M21, 1994.

LOPES, L. F. D., **Apostila Estatística**. Departamento de Estatística, UFSM, 2003.

MASSART, D.L.; VANDEGINSTE, B. G. M.; BUYDENS, L. M. C.; DE JONG, S.; LEWI, P. J.; SMEYERS-VERBEKE, J. **Handbook of chemometrics and qualimetrics. Data handling in science and technology**. Vol. 20A-20B. Amsterdam, New York: Elsevier Science, 1998.

MILLER, J. C.; MILLER, J. N.; **Estatística para química analítica**. 2ª ed. Addison-Wesley: Iberoamerican S.A. Wilmington, EUA, 1993.

MISHRA, P.; SINGH, U.; PANDEY, C. M.; MISHRA, P.; PANDEY, G. Application of Student's t-test, Analysis of Variance and Covariance. **Annals of Cardiac Anaesthesia**. EUA, 2019.

OLIVEIRA, C. D.; DE CAROLI, A. A.; AMARAL, A. S.; VILCA, O. L. Detecção de Fraudes, Anomalias e Erros em Análise de Dados Contábeis: Um Estudo com Base em Outliers. **Redeca**, v. 1, n. 1. p. 102-127, 2014.

PINTO, G. M. F. Testes estatísticos aplicados na validação de métodos. São Paulo. 2019. Disponível em: < <https://cromvallab.com/2019/07/15/testes-estatisticos-aplicados-na-validacao-de-metodos-rdc-166-2017/>> . Acesso em: 01 de junho de 2023.

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

VIEIRA, S. **Introdução à bioestatística**. 4^a ed. Rio de Janeiro: Elsevier, 2008.

**APÊNDICE A – TABELA COM AS CONDIÇÕES DO TESTE T-
PAREADO ATÉ n = 101**

n	v	t(tab) bi 95%	Para t(tab) > t(calc):
2	1	12,706	8,984 > $ \bar{X}_d /S_d$
3	2	4,303	2,484 > $ \bar{X}_d /S_d$
4	3	3,182	1,591 > $ \bar{X}_d /S_d$
5	4	2,776	1,241 > $ \bar{X}_d /S_d$
6	5	2,571	1,050 > $ \bar{X}_d /S_d$
7	6	2,447	0,925 > $ \bar{X}_d /S_d$
8	7	2,365	0,836 > $ \bar{X}_d /S_d$
9	8	2,306	0,769 > $ \bar{X}_d /S_d$
10	9	2,262	0,715 > $ \bar{X}_d /S_d$
11	10	2,228	0,672 > $ \bar{X}_d /S_d$
12	11	2,201	0,635 > $ \bar{X}_d /S_d$
13	12	2,179	0,604 > $ \bar{X}_d /S_d$
14	13	2,160	0,577 > $ \bar{X}_d /S_d$
15	14	2,145	0,554 > $ \bar{X}_d /S_d$
16	15	2,131	0,533 > $ \bar{X}_d /S_d$
17	16	2,120	0,514 > $ \bar{X}_d /S_d$
18	17	2,110	0,497 > $ \bar{X}_d /S_d$
19	18	2,101	0,482 > $ \bar{X}_d /S_d$
20	19	2,093	0,468 > $ \bar{X}_d /S_d$
21	20	2,086	0,455 > $ \bar{X}_d /S_d$
22	21	2,080	0,443 > $ \bar{X}_d /S_d$
23	22	2,074	0,432 > $ \bar{X}_d /S_d$
24	23	2,069	0,422 > $ \bar{X}_d /S_d$
25	24	2,064	0,413 > $ \bar{X}_d /S_d$
26	25	2,060	0,404 > $ \bar{X}_d /S_d$
27	26	2,056	0,396 > $ \bar{X}_d /S_d$
28	27	2,052	0,388 > $ \bar{X}_d /S_d$
29	28	2,048	0,380 > $ \bar{X}_d /S_d$
30	29	2,045	0,373 > $ \bar{X}_d /S_d$
31	30	2,042	0,367 > $ \bar{X}_d /S_d$
32	31	2,039	0,360 > $ \bar{X}_d /S_d$
33	32	2,037	0,355 > $ \bar{X}_d /S_d$
34	33	2,034	0,349 > $ \bar{X}_d /S_d$
35	34	2,032	0,343 > $ \bar{X}_d /S_d$
36	35	2,030	0,338 > $ \bar{X}_d /S_d$
37	36	2,028	0,333 > $ \bar{X}_d /S_d$
38	37	2,026	0,329 > $ \bar{X}_d /S_d$
39	38	2,024	0,324 > $ \bar{X}_d /S_d$
40	39	2,022	0,320 > $ \bar{X}_d /S_d$
41	40	2,021	0,316 > $ \bar{X}_d /S_d$
42	41	2,019	0,312 > $ \bar{X}_d /S_d$
43	42	2,017	0,308 > $ \bar{X}_d /S_d$
44	43	2,016	0,304 > $ \bar{X}_d /S_d$

(continuação)				
n	v	t(tab) bi 95%	Para t(tab) > t(calc):	
45	44	2,015	0,300	> $ \bar{X}_d/S_d$
46	45	2,014	0,297	> $ \bar{X}_d/S_d$
47	46	2,013	0,294	> $ \bar{X}_d/S_d$
48	47	2,012	0,290	> $ \bar{X}_d/S_d$
49	48	2,011	0,287	> $ \bar{X}_d/S_d$
50	49	2,010	0,284	> $ \bar{X}_d/S_d$
51	50	2,009	0,281	> $ \bar{X}_d/S_d$
52	51	2,008	0,278	> $ \bar{X}_d/S_d$
53	52	2,007	0,276	> $ \bar{X}_d/S_d$
54	53	2,006	0,273	> $ \bar{X}_d/S_d$
55	54	2,005	0,270	> $ \bar{X}_d/S_d$
56	55	2,004	0,268	> $ \bar{X}_d/S_d$
57	56	2,003	0,265	> $ \bar{X}_d/S_d$
58	57	2,002	0,263	> $ \bar{X}_d/S_d$
59	58	2,001	0,261	> $ \bar{X}_d/S_d$
60	59	2,000	0,258	> $ \bar{X}_d/S_d$
61	60	2,000	0,256	> $ \bar{X}_d/S_d$
62	61	1,999	0,254	> $ \bar{X}_d/S_d$
63	62	1,999	0,252	> $ \bar{X}_d/S_d$
64	63	1,998	0,250	> $ \bar{X}_d/S_d$
65	64	1,998	0,248	> $ \bar{X}_d/S_d$
66	65	1,997	0,246	> $ \bar{X}_d/S_d$
67	66	1,997	0,244	> $ \bar{X}_d/S_d$
68	67	1,996	0,242	> $ \bar{X}_d/S_d$
69	68	1,996	0,240	> $ \bar{X}_d/S_d$
70	69	1,995	0,238	> $ \bar{X}_d/S_d$
71	70	1,995	0,237	> $ \bar{X}_d/S_d$
72	71	1,994	0,235	> $ \bar{X}_d/S_d$
73	72	1,994	0,233	> $ \bar{X}_d/S_d$
74	73	1,993	0,232	> $ \bar{X}_d/S_d$
75	74	1,993	0,230	> $ \bar{X}_d/S_d$
76	75	1,992	0,228	> $ \bar{X}_d/S_d$
77	76	1,992	0,227	> $ \bar{X}_d/S_d$
78	77	1,992	0,226	> $ \bar{X}_d/S_d$
79	78	1,991	0,224	> $ \bar{X}_d/S_d$
80	79	1,991	0,223	> $ \bar{X}_d/S_d$
81	80	1,991	0,221	> $ \bar{X}_d/S_d$
82	81	1,990	0,220	> $ \bar{X}_d/S_d$
83	82	1,990	0,218	> $ \bar{X}_d/S_d$
84	83	1,990	0,217	> $ \bar{X}_d/S_d$
85	84	1,989	0,216	> $ \bar{X}_d/S_d$
86	85	1,989	0,214	> $ \bar{X}_d/S_d$
87	86	1,988	0,213	> $ \bar{X}_d/S_d$
88	87	1,988	0,212	> $ \bar{X}_d/S_d$
89	88	1,988	0,211	> $ \bar{X}_d/S_d$
90	89	1,988	0,210	> $ \bar{X}_d/S_d$

(continuação)				
n	v	t(tab) bi 95%	Para t(tab) > t(calc):	
91	90	1,987	0,208	$> \bar{X}_d/S_d$
92	91	1,987	0,207	$> \bar{X}_d/S_d$
93	92	1,987	0,206	$> \bar{X}_d/S_d$
94	93	1,986	0,205	$> \bar{X}_d/S_d$
95	94	1,986	0,204	$> \bar{X}_d/S_d$
96	95	1,986	0,203	$> \bar{X}_d/S_d$
97	96	1,986	0,202	$> \bar{X}_d/S_d$
98	97	1,985	0,201	$> \bar{X}_d/S_d$
99	98	1,985	0,200	$> \bar{X}_d/S_d$
100	99	1,985	0,199	$> \bar{X}_d/S_d$
101	100	1,985	0,198	$> \bar{X}_d/S_d$

n: número de observações; v: grau de liberdade. Fonte: autoria própria

ANEXO A – TABELA DA DISTRIBUIÇÃO t-STUDENT

<hr/>											
Unicaudal	75%	80%	85%	90%	95%	97,5%	99%	99,5%	99,75%	99,9%	99,95%
<hr/>											
Bicaudal	50%	60%	70%	80%	90%	95%	98%	99%	99,5%	99,8%	99,9%
<hr/>											
1	1,000	1,376	1,963	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587

Fonte: https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_t_de_Student

ANEXO B – TABELA PARA TESTE DE GRUBB’S

α n	0,1	0,05	0,025	0,01	0,005
3	1,148	1,153	1,154	1,155	1,155
4	1,425	1,462	1,481	1,492	1,496
5	1,602	1,671	1,715	1,749	1,764
6	1,729	1,822	1,887	1,944	1,973
7	1,828	1,938	2,020	2,097	2,139
8	1,909	2,032	2,127	2,221	2,274
9	1,977	2,110	2,215	2,323	2,387
10	2,036	2,176	2,290	2,410	2,482

n: número de observações; α : nível de significância. Fonte:

<https://www.ufsm.br/app/uploads/sites/522/2020/11/06Renius.pdf>

ANEXO C – TABELA PARA TESTE DE COCHRAN

n k	2	3	4	5	6	7	8	9	10
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010
3	0,9669	0,8709	0,7977	0,7457	0,7071	0,6771	0,6530	0,6333	0,6167
4	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017
5	0,8413	0,6838	0,5931	0,5441	0,5065	0,4783	0,4564	0,4387	0,4241
6	0,7808	0,6161	0,5321	0,4803	0,4447	0,4194	0,3980	0,3817	0,3682
7	0,7271	0,5612	0,4800	0,4307	0,3974	0,3726	0,3535	0,3384	0,3299
8	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926
9	0,6385	0,4775	0,4027	0,3584	0,3286	0,3067	0,2901	0,2768	0,2659
10	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,0666	0,2541	0,2439

n: número de replicatas; k: número de concentrações. Fonte:

http://www.de.ufpb.br/~ulisses/disciplinas/tabela_cochran_5.pdf