UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO
MELHORAMENTO GENÉTICO DE PLANTAS

**JOSÉ LUCAS DE ARAÚJO**

# ANÁLISE DO USO DE TESTES DE COMPARAÇÕES MÚLTIPLAS EM ESTUDOS PUBLICADOS EM PERIÓDICOS DE ALTO IMPACTO - MELHORAMENTO GENÉTICO DE PLANTAS

Orientador: José Wilson da Silva
Mestrando: José Lucas de Araújo

**RECIFE-PE**

**2022**

JOSÉ LUCAS DE ARAÚJO

# ANÁLISE DO USO DE TESTES DE COMPARAÇÕES MÚLTIPLAS EM ESTUDOS PUBLICADOS EM PERIÓDICOS DE ALTO IMPACTO - MELHORAMENTO GENÉTICO DE PLANTAS

Dissertação apresentada ao Programa de Pós-Graduação em Agronomia, na área de concentração em Melhoramento Genético de Plantas (PPGAMGP) da Universidade Federal Rural de Pernambuco como exigência à obtenção do título de Mestre.

**Orientador:**

_____

Prof. Dr. José Wilson da Silva
Universidade Federal Rural de Pernambuco (UFRPE)

**Examinadores:**

_____

Prof. Dr. Paulo Ricardo dos Santos
Instituto Federal de Educação Ciência e Tecnologia do Amapá (IFAP)

_____

Dr. Maxwel Rodrigues Nascimento
Universidade Estadual do Norte Fluminense Darcy Ribeiro

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

A663a Araújo, José de Lucas de Araújo
ANÁLISE DO USO DE TESTES DE COMPARAÇÕES MÚLTIPLAS EM ESTUDOS PUBLICADOS EM PERIÓDICOS DE ALTO IMPACTO - MELHORAMENTO GENÉTICO DE PLANTAS / José de Lucas de Araújo Araújo. - 2022.
41 f.

Orientador: Jose Wilson da Silva.
Coorientador: Paulo Ricardo dos Santos.
Inclui referências.

Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Agronomia - Melhoramento Genético de Plantas, Recife, 2022.

1. Teste de comparação de médias. 2. Melhoramento Genético de Plantas. 3. Dados estatísticos. I. Silva, Jose Wilson da, orient. II. Santos, Paulo Ricardo dos, coorient. III. Título

CDD 581.15

## AGRADECIMENTOS

À **Universidade Federal Rural de Pernambuco (UFRPE)**, pela oportunidade do conhecimento, pela rede de apoio e ampliação do aprendizado;

Ao **Programa de Pós-Graduação em Agronomia - Melhoramento Genético de Plantas** pelo acolhimento e por me permitir vivenciar a grande oportunidade de fazer parte do programa, agradeço ao corpo docente por participarem da minha formação acadêmica nessa jornada de extrema importância na minha vida pessoal;

**À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)**, pela concessão da bolsa de mestrado;

Ao meu orientador **Prof. Dr. José Wilson da Silva** pelos ensinamentos e dedicação à esta pesquisa, agradeço imensamente pelas contribuições repassadas. Minha admiração, respeito e gratidão sempre;

Às amizades valiosas que fiz ao longo do mestrado, amigos estes que levarei para o resto da vida, as reuniões no NEMPE para estudos, distrações precisas e conversas importantes, jamais serão esquecidas, em especial: **Islan, Fabian, Jackson, Robson, Jordana, Roberta, Suzanny e Jamile.**

Aos meus pais, **Gilvan Celso de Araújo e Ana Patrícia da Silva**, minha gratidão eterna pelos incentivos diários. Aos meus avós, **Severino Jerônimo (*in memorian*)** e **Maria das Graças**, sem vocês eu nada seria.

À minha esposa, **Jakeline Moreira** que me apoia e me incentiva diariamente. Nunca me deixa desistir.

À todos que contribuíram de forma direta e indiretamente, os meus mais sinceros agradecimentos.

*Ao meu filho, José Miguel, minha força e melhor presente de Deus, a ti dedico.*

## RESUMO

A estatística aplicada a melhoramento genético e experimentos agrícolas configura-se como uma ferramenta primordial para o alcance de resultados efetivos, possibilitando a estimativa do erro experimental, além de validar a importância dos contrastes analíticos. Com isso, teve-se como objetivo identificar a correta aplicação de testes de comparação de médias em estudos voltados para o melhoramento genético de plantas depositados em periódicos de alto impacto, favorecendo resultados confiáveis em pesquisas que utilizam esse tipo de análise. Para o alcance dos objetivos, a metodologia foi dada por uma pesquisa bibliográfica de abordagem transversal por meio de publicações de artigos em revistas de alto impacto entre os anos de 2016 e 2021, que utilizaram em suas metodologias o teste de comparação de médias, sendo utilizado como descritores: estatística, melhoramento genético e interpretação de dados. Após a coleta, foram selecionados 60 artigos, sendo submetidos a uma primeira classificação, de acordo com a utilização ou não de algum tipo de teste de comparação de médias. Aqueles que se utilizaram deste procedimento foram, posteriormente, agrupados conforme a sua aplicação, sendo classificados, conforme critérios empregados por Bertoldo *et al.* (2008), em: i) adequado; ii) parcialmente adequado e iii) inadequado. Com isso, pode-se concluir que os principais aspectos que afetam diretamente a confiabilidade da interpretação dos resultados do pesquisador são: i) premissas que não levam em conta diferentes testes estatísticos, ii) conhecimento preliminar sobre o tipo de fatores e iii) Inconsistências na seleção de testes estatísticos que podem levar a conclusões incompletas e/ou inadequadas.



**Palavras-chave:** Teste de comparação de médias, Melhoramento Genético de Plantas, Dados estatísticos.

## ABSTRACT

The statistics applied to genetic improvement and agricultural experiments is configured as a key tool to achieve effective results, enabling the estimation of experimental error, in addition to validating the importance of analytical contrasts. With that, the objective was to identify the correct application of tests of comparison of averages in studies focused on the genetic improvement of plants deposited in high impact journals, favoring reliable results in research that uses this type of analysis. To achieve the objectives, the methodology was given by a bibliographic research of a transversal approach through publications of articles in high impact journals between the years 2016 and 2021, which used in their methodologies the test of comparison of means, being used as descriptors: statistics, genetic improvement and data interpretation. After collection, 60 articles were selected, being submitted to a first classification, according to the use or not of some type of test to compare means. Those who used this procedure were later grouped according to their application, being classified according to the criteria employed by Bertoldo et al. (2008), in: i) adequate; ii) partially adequate and iii) inadequate. With this, it can be concluded that the main aspects that directly affect the reliability of the researcher's interpretation of results are: i) assumptions that do not take into account different statistical tests, ii) preliminary knowledge about the type of factors and iii) Inconsistencies in the selection of statistical tests that may lead to incomplete and/or inappropriate conclusions.



**Keywords:** Average comparison test, Plant Genetic Improvement, Statistical data

## LISTA DE GRÁFICOS

## LISTA DE TABELAS

# SUMÁRIO

## 1 INTRODUÇÃO

A estatística aplicada a melhoramento genético e experimentos agrícolas configura-se como uma ferramenta primordial para o alcance de resultados efetivos, possibilitando a estimativa do erro experimental, além de validar a importância dos contrastes analíticos. No entanto, as explicações biológicas dos fenômenos analisados devem revelar sincronia, em que as inferências feitas devem ser consistentes com a quantidade de informações obtidas (WANG *et al*., 2016).

Eventualmente, de acordo com Zoanetti (2013), a escolha do tipo de procedimento estatístico a ser realizado após a Análise de Variância (ANOVA) não é o mais recomendado para o tipo de experimento realizado, principalmente quando os testes de comparação de médias são aplicados a fatores quantitativos como produto, concentração e densidade. Desse modo, a escolha inadequada do tipo de análise pode levar a dificuldades na interpretação dos dados experimentais, por limitações na inferência, ou mesmo por generalizações feitas de forma errada.

Assim, a escolha de qual teste de comparação de médias usar depende da precisão desejada e do nível de discriminação de tratamento fornecido pelo programa estatístico, tendo em vista que os testes de comparação múltipla "a posterior" são testes não planejados antes da realização de um experimento, que complementam o teste F de uma análise de variância projetado para detectar diferenças (ortogonais ou não positivas) entre duas médias (ou um conjunto de médias) de tratamentos (RALSTON; DUNDAS; LEYLAND, 2014).

Em vista disso, os processamentos de estatísticas experimentais têm sido difundidos em diversas áreas, incluindo estudos científicos, para realização de planejamentos e estudos experimentais, interpretação de dados e apresentação de resultados de forma a auxiliar pesquisadores. Em particular, a sua aplicação no campo do melhoramento genético de plantas procura determinado genótipo ou grupo de genótipos que sejam mais produtivos e ao mesmo tempo tenham o menor custo.

Considerando este fator, percebe-se a necessidade de se explorar o ambiente de cultivo dos genótipos para uma maior produção com menores custos. Assim, enfatizar a interação genótipo e ambiente tornou-se um

importante tema de estudos em melhoramento genético de plantas, integrando a estatística como ferramenta-chave para o entendimento e comparação de dados em um contexto mais amplo e eficiente.

Por isso, a estatística é usada como ferramenta para conclusões que tenham base em dados experimentais. No entanto, essas conclusões dependem de como o experimento foi conduzido, onde estatísticos exigirão uma descrição detalhada do experimento e seus objetivos.

Em alguns casos, um estatístico confirma, depois de descrever um experimento, que não pode tirar nenhuma conclusão efetiva, pois o pesquisador não usou um projeto apropriado ou não atendeu aos pressupostos básicos necessários para a validade da análise estatística, configurando o planejamento experimental fator essencial para alcance de resultados pertinentes.

Partindo desse pressuposto, é primordial destacar que a estatística é uma área da ciência que utiliza a análise dos dados para verificar as hipóteses, observando o poder da evidência e, se há correlações entre grupos ou a autenticidade de fenômenos de interesse. Com isso, o cientista deve levantar teorias, perceber os fenômenos biológicos que advêm da população e obter a partir disto uma amostra para confirmá-las. Dessa forma, a afinidade de uma amostra com o meio populacional que a concebeu possibilita que os resultados da analogia dos dados sejam mais autênticos (WINDISH; DIENER-WEST, 2006).

Com isso, pode-se justificar que a utilização da análise estatística propicia ao leitor e aos pesquisadores em geral, compreender a informação oriunda dos dados apurados durante a realização do estudo, determinando contribuições importantes para a confiabilidade dos resultados. Considerando que frequentemente, pesquisadores acabam fazendo uso inadequado dos testes estatísticos em razão do não emprego das pressuposições necessárias e, consequentemente, apresentam dificuldades para análise dos dados e compreensão dos resultados (BERTOLDO *et al*., 2007).

Nesse viés, a presente pesquisa tem como intuito a investigação de pesquisas que abordam o melhoramento genético de plantas com a utilização de testes de comparação de média, a fim de se obter resultados mais precisos em termo estatísticos. Em vista disso, Salman e Giachetto (2014) apresentam os princípios básicos da experimentação, que se configuram como a base dos

delineamentos experimentais clássicos, representando parte importante em termos de análise, são eles: repetição, casualização e controle local.

A repetição diz respeito à quantidade de parcelas que receberão um mesmo tratamento. Com isso, é necessário destacar que os tratamentos devem ter repetições para que se possa estimar o erro experimental, o qual é de fundamental importância nos testes de hipóteses, pois deve-se obter uma quantidade de repetições adequadas para gerar uma boa estimativa experimental, melhorando assim, as análises de interesse e a efetividade dos resultados.

No entanto, deve-se observar que o número de repetições pode ser limitado devido a alguns fatores, tais como os tratamentos que serão comparados, a disponibilidade de material, de área experimental, dentre outros. Assim, Salman e Giachetto (2014) comentam que a confiabilidade das estimativas depende do número de aplicações repetidas de tratamento, e o número adequado de repetições varia de acordo com as variáveis em questão.

No caso da casualização, ela refere-se à distribuição aleatória dos tratamentos às parcelas de modo que todas tenham a mesma oportunidade de receber qualquer um dos tratamentos, garantindo, sobretudo, que os erros sejam independentes.

Com isso, essa ferramenta busca garantir que os tratamentos não sejam consistentemente favorecidos ou desfavorecidos em sucessivas repetições devido a variáveis de origem conhecida ou desconhecida. Neste caso, os tratamentos devem ser distribuídos aleatoriamente entre as unidades experimentais, onde variações que causam erro experimental são transformadas em variáveis aleatórias. Portanto, cada grupo experimental deve ter a mesma chance de receber indivíduos com alterações semelhantes (DA SILVA *et al*., 2016).

Além disso, Da Silva *et* al. (2016) comentam que o princípio de controle local visa eliminar o erro experimental entre os grupos por meio da padronização dentro da parcela, aplicação de tratamentos e padronização de onde os experimentos são realizados. Os grupos usados para testar os tratamentos devem ser homogêneos em relação à idade, classe, peso, sexo, raça, ascendência, ou outras características importantes que devem ser consideradas para cada estudo específico.

Para realização da pesquisa, a seleção do delineamento experimental apropriado se caracteriza como fator importantes, bem como a aplicação correta da medida estatística, no intento da estimativa certa do erro experimental (BERTOLDO *et al.,* 2007). Para isso, é necessário realizar uma análise coerente dos dados, compreendendo a natureza dos fatores que serão estudados.

Dessa forma, é importante entender como a natureza dos dados estudados em determinada pesquisa podem ser analisados, compreendendo estudos quantitativos ou qualitativos. Proetti (2018) comenta que, de forma geral, dados quantitativos fornecem números que demonstram os objetivos gerais do estudo, enquanto os dados qualitativos ajudam a compreender a complexidade e o detalhamento das informações obtidas.

A presente pesquisa está associada as características qualitativas, empregando os testes de comparações múltiplas, sendo considerado o mais adequado quando se trata de comparar os tratamentos dentro de um grupo de amostras (ALVAREZ; ALVAREZ, 2006).

Dessa forma, teve-se como objetivo identificar a correta aplicação de testes de comparação de médias em estudos voltados para o melhoramento genético de plantas depositados em periódicos de alto impacto, favorecendo resultados confiáveis em pesquisas que utilizam esse tipo de análise.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Análise de variância e teste de média

A análise de variância ou ANOVA é um processo usado para comparar as distribuições de três ou mais grupos em amostras independentes. ANOVA também é um método de resumir um modelo de regressão linear decompondo a soma dos quadrados para cada fonte de variação, usando o teste F para testar a hipótese de que qualquer fonte de variação é igual a zero (DA ROCHA; JÚNIOR, 2018).

Nesse cenário, um problema muito comum em experimentos agrícolas e de melhoramento genético é a comparação das médias de tratamentos de interesse, e descobrir se essa diferença é significativa, além de entender qual o melhor tratamento dentro dos objetivos propostos.

Segundo os autores Girardi, Filho e Storck (2009), a maneira mais comum de lidar com esse problema é a análise de variância de dados experimentais, um procedimento estatístico que compara a variação induzida pelo tratamento com a variação induzida pelo acaso. A hipótese testada na ANOVA, também conhecida como hipótese da homogeneidade, é que as médias populacionais dos tratamentos não diferem entre si em um determinado nível de significância, que pode ser expresso da seguinte forma:

$$H0: \mu 1 = \mu 2 = \cdots = \mu a,$$

Onde $a$ representa o número total de tratamentos no ensaio, $\mu i$ representa a média populacional do i-ésimo tratamento, $i = 1, \ldots, a$.

O teste básico para comparação de tratamentos em ANOVA foi proposto por Ronald Aylmer Fisher e é conhecido como teste z de Fisher. Quando o teste F é significativo, entende-se que o teste de comparação de médias pode ser aplicado para investigar a diferença final entre um par específico de médias ou uma combinação linear desses métodos (PAGOTTO *et al*. 2021).

Segundo Pagotto *et* al. (2021), ao realizar uma ANOVA para experimentos com apenas dois tratamentos, pode-se visualizar qual é o melhor por meio da média. No entanto, quando há mais de dois tratamentos, apenas fazendo um teste "F" (testando se há diferença entre as médias dos tratamentos) não é possível identificar qual é o mais apropriado. Assim, é necessário aplicar testes para comparar as médias dos tratamentos, onde a comparação pode ser utilizada como complemento aos estudos com o uso da ANOVA, como teste de Tukey, e teste de Duncan.

A análise de variância de fator único concentra-se na comparação de dois ou mais tratamentos ou medias. Onde *I* é igual ao número de tratamentos que serão comparados; e μ1, μ2, μ3, ..., μi as médias populacionais ou médias dos tratamentos.

Sendo assim, as hipóteses de interesse são:

*H0: μ1 = μ2 = μ3 = ... = μi*

H1: Pelo menos duas médias diferentes

Quando se pretende obter a estatística do teste, é importante saber:

A soma dos quadrados dos tratamentos *(SQT r)* = $\sum_{i=1}^{I} \sum_{j=1}^{I} (xi - \bar{x})^2$

A soma dos quadrados totais *(SQT)* = $\sum_{i=1}^{I} \sum_{j=1}^{I}$ *(xij - x̄)²*

A soma dos quadrados dos resíduos *(SQR)* = $\sum_{i=1}^{I} \sum_{j=1}^{I}$ *(xij - x̄)² - (xi - x̄)²*

$a$ = número de tratamentos; b = número de repetições

Quadrados médios dos tratamentos *QMT* = $\frac{SQTr}{a-1}$

Quadrados médios dentro dos tratamentos (resíduos) *QMR* = $\frac{SQR}{a(b-1)}$

Para uma melhor visualização, a tabela a seguir (Tabela 1) demonstra de forma prática a análise da variância (ANOVA) e o teste *F.*

**Tabela 1.** Análise da variância

| CAUSA DE VARIAÇÃO | g.1. | SQ | QM | F |
|---|---|---|---|---|
| Tratamento | $a - 1$ | *(SQT)* = $\sum_{i=1}^{I} \sum_{j=1}^{I}$ *(xij - x̄)²* | *QMTr* = $\frac{SQTr}{a-1}$ | $\frac{SQTr}{QMR}$ |
| Resíduo | *a (b-1)* | *(SQR)* = $\sum_{i=1}^{I} \sum_{j=1}^{I}$ *(xij - x̄)² - (xi - x̄)²* | *QMR* = $\frac{SQR}{a(b-1)}$ | com a − 1 e a(b − 1) g.l. |
| **Total** | *ab-1* | *(SQT)* = $\sum_{i=1}^{I} \sum_{j=1}^{I}$ *(xij - x̄)²* | | |

Fonte: adaptado de Scudino (2008)

Dessa forma, para Tavares (2020) a ANOVA é essencialmente um procedimento aritmético para decompor a variação total entre unidades experimentais (variância total ou soma dos quadrados dos desvios totais ou soma dos quadrados totais) em correlações com fontes ou causas de variação previstas ou identificáveis. Mais especificamente, a variação total é dividida em três grupos de causas ou fontes de variação:

(I) Variação relacionada com os tratamentos.

(II) Variação relacionada com causas controladas pelo delineamento experimental.

(III) Variação relacionada com o erro experimental.

Com isso, a ANOVA é uma técnica estatística comumente utilizada na análise de dados experimentais quantitativos, em diferentes áreas de estudo, ou mesmo para comparação de dois tratamentos.

### 2.1.1 Independência

Ao declarar um modelo estatístico para uma determinada variável resposta, assume-se que o termo de erro é uma variável aleatória normal, independente e identicamente distribuída. Isso significa que se, em um determinado experimento, tratamentos idênticos forem colocados próximos uns dos outros ou em uma ordem lógica, pode-se esperar que eles apresentem rendimentos semelhantes devido à sua disposição. Portanto, é importante não agrupar todas as parcelas contendo o mesmo tratamento em uma série adjacente de parcelas, mas distribuir os tratamentos aleatoriamente entre as parcelas experimentais (SCUDINO, 2008). Assim, em um experimento bem projetado, à randomização adequada é uma boa proteção contra violações dessa suposição.

A falta de independência dos erros também pode ser resultado de correlações temporais e não espaciais, onde uma simples medição de peso em uma balança pode levar a erros associados devido à falta de ajuste, resultando em subestimação contínua. Experimentos realizados em estações frequentementes também revelam medidas correlacionadas (PHAM *et al*., 2020). Não há ajuste ou transformação simples para superar o erro da falta de independência (SCOTT *et al*., 2020). Outra maneira é mudar o projeto básico do experimento ou a forma como o experimento é conduzido. Se os erros não forem independentes, a validade do teste F de significância pode ser seriamente comprometida.

### 2.1.2 Homogeneidade de variâncias

A ausência de homogeneidade faz com que alguns grupos apresentem maior variação do que outros. Ao realizar experimentos, uma amostra pode ter sido obtida em condições menos padronizadas que outras e, portanto, apresentar maior variância. Esses casos são conhecidos por "variações que não se relacionam funcionalmente com a média" (PHAM *et al*., 2020).

Naturalmente, não se pode esperar que as variáveis de resposta dessas fontes satisfaçam a suposição de homogeneidade de variâncias. Por exemplo, no melhoramento, pode-se esperar que a variância do genótipo da geração F2 seja maior que a do genótipo F1, porque a variabilidade genética de F2 é muito

maior que a de F1. Em experimentos envolvendo tratamentos e controles, espera-se que os tratamentos também mostrem maior variabilidade do que os controles.

Em outros casos, a heterogeneidade na variância pode ser causada pela distribuição dos dados. Em algumas distribuições, a variância pode variar com a média (SCHOBER; VETTER, 2020). Por exemplo: uma variável que segue uma distribuição de Poisson tem uma variância igual à média ($s^2 = m$). Então, uma população que mostra uma média alta tem uma variância maior, onde esses casos de heterogeneidade de variância são bem conhecidos quando "as variâncias estão funcionalmente relacionadas à média" (PHAM *et al*., 2020).

### 2.1.3 Normalidade dos erros

Das quatro hipóteses, esta é a menos provável de se sustentar, pois se a variável resposta for discreta certamente não obedece a essa suposição. Variáveis que representam contagens e variáveis contínuas que representam peso ou altura individual, restritas a valores positivos por definição, também não satisfazem essa suposição, sendo válida apenas de forma aproximada (MARINI *et al*., 2015; SCHOBER; VETTER, 2020).

Alguns argumentos protegem as inferências da falta de normalidade através da robustez do teste F e do teorema do limite central, sendo que a falta de normalidade só pode desempenhar um papel significativo ao nível do teste F se a distribuição for muito assimétrica (MARINI *et al*., 2015).

### 2.1.4 Aditividade

A suposição de aditividade significa que não há interação entre os dois efeitos no modelo. Em uma análise de variância bidirecional, a ausência dessa interação pode ser chamada de aditividade de efeitos principais. Isso significa que qualquer variável observada pode ser decomposta em componentes adicionais (PIMENTEL *et al*., 2014).

Pimentel *et al*. (2014) afirma ainda que as interações entre os fatores podem surgir por vários motivos: i) mais comumente, a interação resulta em uma determinada combinação de tratamentos (por exemplo, quando o fator A nível 2 é combinado com o fator B nível 3); ii) quando uma determinada combinação de

tratamentos com efeitos semelhantes ocorrer quando a repetição se torna muito anômala; iii) se os efeitos dos dois fatores A e B sobre a variável resposta Y forem multiplicativos ao invés de aditivos, há um efeito de interação.

### 2.1.5 Teste de F

O teste de ANOVA usa o teste F para determinar se a variabilidade entre as médias do grupo é maior que a variabilidade das observações dentro dos grupos. Se essa proporção for suficientemente grande, pode-se concluir que nem todas as médias são iguais (EMERSON, 2017).

O teste de F aponta as diferenças entre as médias dos tratamentos, mas não de forma aprofundada. A sua aplicação pode ser universal, porém é interessante o uso de outros testes para complementar a análise como Tukey ou Duncan. Para Mahbobi e Tiemann (2015), faz-se necessário uma comparação entre o valor apontado pela ANOVA e pelo valor tabelado de F para que as hipóteses sejam construídas, onde o valor do F indica a distribuição da probabilidade da união entre as funções.

Para cada F há um nível de significância, o que culmina na criação da tabela de F, guiando o valor de para as comparações. Quando não há diferenças significativas entre os tratamentos, considera-se que eles são semelhantes, e quando há diferença significativa, pelos menos dois tratamentos evidenciam suas diferenças. Portanto, o teste de F indica quando há diferenças, mas de modo não aprofundado, precisando de outro teste para indicar onde estas diferenças se localizam.

O valor de F é calculado através da fórmula:

$$\mathrm{F_{cal}} = \frac{\mathrm{SQtrat}}{\mathrm{SQres}}$$

Onde:

- SQtrat é igual a soma dos quadrados dos tratamentos
- SQres é igual a soma dos quadrados dos resíduos

### 2.1.6 Teste t de Student

O teste de Student foi criado por William Sealy Gosset, um químico que trabalhava em uma cervejaria e utilizava pequenas amostras para comparar a

qualidade delas. Com o objetivo de achar o ponto chave para a qualidade da cerveja utilizando apenas poucas amostras, ele criou o método de teste de média conhecido hoje como o teste de Student, entretanto, seu nome não foi adicionado ao teste devido a cervejaria não querer ter os seus dados divulgados, para que a concorrência não se sobressaísse (ZABELL, 2008).

O teste t é um teste paramétrico usado para comparar duas populações $P_1$ e $P_2$, analisando os dados obtidos a partir de uma amostra de cada um deles. Para testar a igualdade ou equivalência de duas populações, é necessário estimar parâmetros para cada um deles, como a média e o desvio padrão. Assumindo que a população é normalmente distribuída (TAVARES *et al.*, 2020).

O teste que foi publicado por Student (1908), faz comparações entre os grupos de médias, de modo a respeitar os contrastes e a ortogonalidade das amostras. Além disso, possui algumas características como as médias que fazem parte dos contrastes serem escolhidas antes de iniciar a análise, a quantidade de constastes e tratamentos, e a aplicação do teste ser vinculada ao valor de F de Fisher.

Neste teste, o contraste não tem que necessariamente englobar todas as médias, porém deve descrever de forma clara as respostas apresentadas em cada um dos contrastes, definindo a média que será estudada e evitando novos contrastes após o início da análise (MAHBOBI; TIEMANN, 2015). Assim, a ortogonalidade é garantida por causa da independência das comparações, não alterando o comportamento da variação, sendo a variância é a média do contraste.

O valor dos contrastes é calculado utilizando a seguinte fórmula:

$$Y_1 = a_1\mu_1 + \cdots + a_n\mu_n$$
$$\vdots$$
$$Y_{n-1} = j_1\mu_1 + \cdots + j_n\mu_n$$

Em que $\mu_1$ representa a média populacional do i-ésimo tratamento e $J_i$ o seu respectivo coeficiente.

### 2.1.7 Teste de Tukey

Baseia-se na amplitude total estudentizada, podendo ser utilizado para comparar todo e qualquer contraste entre dois tratamentos. O teste é preciso e fácil de usar quando todos os tratamentos são repetidos o mesmo número de

vezes. Se o número de repetições for diferente, o teste de Tukey ainda pode ser utilizado, mas é apenas uma aproximação (KIM, 2017).

O teste de Tukey (1949) é feito através de um algoritmo que obtém apenas um valor que deve ser calculado e só pode ser aplicado na análise de variância (ANOVA), que irá detectar a existência de diferenças entre as médias dos tratamentos, previamente testados por F. Para Montgomery *et al.* (2003), o algoritmo utilizado na realização do teste de Tukey não é de uso universal e necessita de uma diferença pré observada entre as médias no contraste. O valor do Tukey é utilizada para as combinações de pares de médias, pois o seu valor limita-se a apenas ser utilizado nos pares de médias. O teste segue a seguinte fórmula:

$$q_{tabelado} = \frac{D_{Tukey}}{\sqrt{\frac{QMR}{r}}}$$

Onde:

- QRM é igual ao quadrado médio dos resíduos;
- R significa número de repetições do experimento.

A diferença mínima encontrada na análise será o parâmetro para determinar as diferenças dentro do contraste. Tukey (1949) aconselha que seja construída uma ordem entre os valores das médias dos tratamentos antes da análise, para que assim possa ser formulada uma função, obedecendo a seguinte fórmula:

$$Y_k = |\mu_n - \mu_m|$$

### 2.1.8 Teste de Dunnett

O teste de Dunnett (1964) se configura como uma modificação do teste de t, e deve ser aplicado com a finalidade de comparar a média do controle com as demais médias dos outros tratamentos. O teste foi criado em um experimento para o desenvolvimento de uma dieta para frangos, com o objetivo de analisar o teor de gordura nos músculos do peito dos animais. O experimento teve um controle sem gordura, três grupos com dietas ricas em gordura e com 4 repetições, e todos da mesma raça (HOTHORN, 2016).

Dado que as amostras são aleatórias e independentes, a partir de variáveis com distribuição normal, o teste de Dunnett é um teste que compara simultaneamente a média do tratamento testado com a média do tratamento controle. A limitação do uso deste teste é dada pela difícil obtenção de valores para os quantis da probabilidade e estatística da distribuição t multivariada, pois o teste pode ser aplicado tanto em situações balanceadas quanto desbalanceadas, unilaterais ou bilaterais, com infinitas possibilidades de correlação entre comparações (OLIVEIRA, 2008).

Segundo Hothorn (2016), os contrastes são formados em pares de médias e devem ser comparados entre a testemunha e um tratamento, reduzindo para uma unidade, em relação ao teste de Tukey, o teste segue a seguinte fórmula:

$$\begin{gathered} Y_1 = |\mu_1 - \mu_1| \\ Y_2 = |\mu_2 - \mu_2| \\ Y_3 = |\mu_3 - \mu_3| \\ \vdots \\ Y_{n-1} = |\mu_1 - \mu_{n+1}| \end{gathered}$$

Desta forma, pode-se compreender que o teste de Dunnett (1955) segue a mesma lógica do teste de Tukey (1949), porém possui valores diferentes, já que é uma modificação do teste de t. Além disso, as fragilidades dos testes também são semelhantes, uma vez que a resposta pode ser ambígua, encontrando diferenças onde não existem, ou vice-versa.

Pode-se observar também que o contraste não se relaciona de modo comparativo com outros tratamentos, o que é um ponto negativo do teste, onde para que haja uma relação de todos contra todos, deve-se transformar cada tratamento em uma testemunha, e coletar as respostas. A melhor forma de utilização deste teste é para finalidades agrícolas, ou seja, no desenvolvimento de cultivares melhoradas geneticamente, com o intuito de fazer comparações entre a cultivar desenvolvida e os padrões das cultivares disponíveis no mercado (MINAMOTO *et al,* 2016).

### 2.1.9 Teste de Duncan

O teste de Duncan é um teste que considera o número de médias envolvidas no contraste, os seus tratamentos podem ou não ser balanceados e

as médias são determinadas em valores decrescentes. Além disso, seu valor pode variar de acordo com as médias de cada contraste. Para dados balanceados, o teste segue o algoritmo a seguir:

$$D_{Duncan} = z_{Duncan}\sqrt{\frac{QMR}{r}}$$

Onde:

- QMR significa o Quadrado médio do resíduo;
- R significa número de repetições do experimento.

A principal característica deste teste é que o experimento será balanceado, tendo o mesmo número de repetições em todos os tratamentos. (SILVA; DE AZEVEDO, 2016) comentam que para que o teste seja efetuado em experimentos não balanceados, o mesmo deve ser adaptado para que possa ser aplicado de maneira harmônica.

### 2.1.10 Teste de Scott-Knott

Scott e Knott (1974) apresentaram um teste com uma comparação distinta em relação aos tradicionalmente encontrados com frequência nos artigos e revistas científicas, tais como Tukey, Dunnettt, Duncan e outros dentro desta linha analítica. Assim, como nos demais, o teste de Scott e Knott aborda uma organização crescente ou decrescente das médias a serem realizadas, de acordo com cada contraste inicial (1 e 2) e contendo grupos de médias, o cálculo do valor $B_0$ para os que podem ser formados.

Nesta sequência, o valor $B_0$ é calculado distintas vezes até ser aproveitado o maior valor $B_0$ dentre todos os calculados. A maneira para obtenção do valor de $B_0$ é apontada por Hartigan (1972) um trabalho anterior ao de Scott e Knott (1974) definindo a função para determinação desta variável. Sendo:

$$B_0 = \frac{\left(\sum_{i=1}^{j} T_{i1}\right)^2}{j} + \frac{\left(\sum_{i=1}^{k} T_{k1}\right)^2}{1} + \frac{\left[\sum_{i=1}^{j+1}(T_{i2} + T_{k2})\right]^2}{j+1}$$

Em que:

$T_{i1}$ = médias contidas no contraste 1;

$T_{i2}$ = médias contidas no contraste 2;

j e 1 = quantidade de médias por contraste.

Em que:

**$T_{i1}$** – médias contidas no contraste 1;

**$T_{i2}$** – médias contidas no contraste 2;

**j e l** – quantidade de médias por contraste

Costa (2003) mostra uma maneira simples para a obtenção do valor de $B_0$ através da formação de partições com a média maior isolada das demais em um contraste e a próxima com a maior e a segunda maior média e assim sucessivamente. Isto vai isolar, automaticamente, o maior valor $B_0$ e a análise será realizada com mais rapidez, determinando, por fim, a existência ou não das diferenças significativas.

Já para Scott e Knott (1974) o valor $B_0$ é a variância entre grupos de tratamentos da partição e a partida para realizar as análises, uma vez que o autor agregou ao valor $B_0$ outras variâncias para chegar em uma distribuição estatística conhecida como forma de garantia da qualidade do teste.

Os autores explicam que o teste foi refeito considerando vários trabalhos publicados, dentre eles: Tukey (1949) e Duncan (1955) separando os grupos por quantidade de médias iguais nos dois contrastes iniciais e estudando as respostas aplicadas pelo trabalho em relação aos resultados dos demais.

Calculado o valor de B0 determina-se o estimador máximo de verossimilhança de σ² o qual partirá da suposição de que todas as médias serão semelhantes. Esse estimador será necessário posteriormente, para comparar a hipótese H0 em aceitar ou rejeitar a mesma. O estimador vem do resultado da razão de verossimilhança do parâmetro σ² desenvolvida e obtida da seguinte maneira:

$$\hat{\sigma}_0^2 = \frac{\sum_{i=1}^{t}(\bar{y}_i - \bar{y}) + vs_{\bar{y}}^2}{t + v}$$

Em que:

**$Y_i$** – média de cada tratamentos;

**Y** – média geral dos tratamentos;

**8 2/Y**– variância do erro ou QMR;

**v** – grau de liberdade do resíduo;
**t** – médias de tratamentos envolvidos no grupo ou partição.

Em suma, a classificação e agrupamento entre médias são definidas na aplicação do teste, tornando-se um problema para os casos de comparações menos complexas do tipo: testemunha e tratamento. Nesse sentiso, o trabalho de Silva (2007) aponta que o teste de Scott-Knott aplicado a uma variável comparada com as demais apresentando rigidez em excesso e um número reduzido na formação de grupamento entre médias. Por outro lado, Ferreira *et al.* (1999) incentivam mais trabalhos em torno do teste de Scott e Knott para delineamento experimentais com um número expressivo de k tratamentos, tendo em vista que ele é recente.

### 2.1.11 Erro tipo I e tipo II

O erro do Tipo I ocorre quando a hipótese nula é verdadeira e o pesquisador a rejeita. A probabilidade de cometer um erro do Tipo I é α, que é o nível de significância que o pesquisador estabelece para seu teste de hipótese. Um alfa de 0,05 significa que se deseja aceitar uma probabilidade de erro de 5% para rejeitar a hipótese nula. Para reduzir esse risco, deve-se usar valores alfa mais baixos. No entanto, usar um valor alfa mais baixo significa que é menos provável a determinação de diferenças reais, se existirem (GIRARDI; CARGNELUTTI FILHO; STORCK, 2009).

O erro tipo II ocorre quando a hipótese nula é falsa e não é rejeita. A probabilidade de cometer um erro do Tipo II é β, que depende do poder do teste. Pode-se reduzir o risco de cometer erros do Tipo II garantindo que seus testes sejam adequadamente competentes, onde o pesquisador pode fazer isso garantindo que o tamanho da amostra seja grande o suficiente para detectar diferenças reais, quando realmente houver diferenças (GIRARDI; CARGNELUTTI FILHO; STORCK, 2009).

A tabela a seguir simplifica como podem ocorrer erros associados a testes de hipótese (Tabela 2).

**Tabela 2.** Erros possíveis associados a teste de hipóteses

| SITUAÇÃO | CONCLUSÃO DO TESTE | |
|---|---|---|
| Real | Rejeitar H0 | Não rejeitar H0 |
| H0 verdadeira | Erro tipo I | Decisão correta |
| H0 falsa | Decisão correta | Erro tipo II |

Fonte: adaptado de Delacre, Lakens e Leys (2017).

### 2.1.12 Diferença mínima significativa (DMS)

A diferença mínima significativa determina quais tratamentos não são estatisticamente iguais, para que seja possível a comparação dos tratamentos de forma mais eficiente. Nesse sentido, para que seja determinado quais tratamentos são iguais ou não se utiliza testes de média, como: teste “t” de Student, teste de Tukey e teste de Dunnett (Tabela 3).

**Tabela 3**. Testes que utilizam Diferença Mínima Significativa (DMS) e suas respectivas fórmulas

| TESTES | FÓRMULAS |
|---|---|
| Teste T | $dms = {}^{t\delta.\alpha.}\sqrt{\left(\frac{1}{ri}+\frac{1}{rj}\right)QMR}$ |
| Teste de Tukey | $dms = q\alpha,(\delta,k)\sqrt{\left(\frac{1}{ri}+\frac{1}{rj}\right)\frac{QMR}{2}}$ |
| Teste de Dunnett | $dms = d\alpha,(\delta,t)\sqrt{\left(\frac{1}{rt}+\frac{1}{rc}\right)\frac{QMR}{2}}$ |

Fonte: Adaptado de Frankenberg (2019).

Onde ri e rj são o número de repetições de cada tratamento. O teste de Dunnett é usado para comparar grupos de tratamento com grupos controle, enquanto rt e rc representam o número de repetições para cada grupo.

## 2.2 Melhoramento de plantas

O melhoramento vegetal é a ciência de detectar genótipos superiores aos previamente inseridos ou utilizados em benefício da humanidade. É considerada uma ciência porque usa os princípios da genética, experimentação e estatística, dependendo da sensibilidade dos criadores para distinguir variação genética de variação casual (CECCARELLI, 2015). Com o início do cultivo agrícola, as plantas sofreram adaptações e mudanças, como maior retenção de sementes,

crescimento mais firme e aumento no tamanho e número de inflorescências que as tornaram superiores (CECCARELLI, 2015; BAKER, 2020).

As origens do melhoramento de plantas foram impulsionadas pelo trabalho dos biólogos Charles Darwin (1809 - 1882) e Gregor Mendel (1822 - 1884), o que levou a inúmeros debates que continuaram até o início do século XX (BETRÁN *et al*., 2009). Darwin tentou demonstrar como a evolução das espécies acontece usando a teoria da seleção natural, que muda gradualmente ao longo do tempo.

Darwin também desenvolveu uma teoria plausível da herança do caráter das pequenas mudanças que ocorrem na prole como resultado do cruzamento parental. Mendel descreveu a lei de separação de acordo com as características das espécies de ervilha. As características do estudo são determinadas por diversos fatores que estarão localizados nos gametas de cada genitor. Tem havido muito debate sobre se a mudança contínua proposta por Darwin está relacionada a fatores de dispersão mendelianos. No início do século 20, o debate foi resolvido, resultando na teoria da evolução sintética (RICHARDS *et al*., 2015).

Desde a definição e consolidação da teoria das origens do melhoramento vegetal, da ciência , aplicada a projetos de pesquisa, os principais objetivos são (CALIGARI, 2001): a) Rendimento útil: Além do rendimento total, as melhorias visam a diferença entre tempo de armazenamento, desperdício e aspectos relacionados à aceitação do consumidor; b) Estabilidade do rendimento: garantir que as variedades mantenham a estabilidade da produção em diversos ambientes, mitigando estresses bióticos e abióticos; c) Qualidade do produto: inclui qualidade nutricional e sabor, valor nutricional, valor calórico, teor de proteína e gordura; d) Impacto ambiental: Isso deu origem a muito debate em todo o mundo, pois a agricultura afeta qualquer área onde é praticada. Assim, o melhoramento de plantas visa utilizar métodos ecologicamente sustentáveis, como produzir variedades resistentes a pragas e doenças e com alta capacidade de assimilação e absorção de nutrientes; e) alta adaptabilidade: necessidade de produzir variedades que respondam positivamente a melhores condições ambientais, levando em consideração o clima de um determinado local e tipo de prática agrícola; f) capacidade preditiva: os melhoristas de plantas devem ter a capacidade de “prever o futuro”, por exemplo, como as mudanças climáticas

afetarão os padrões de crescimento? Qual é a gama de pragas e doenças? Quais são as necessidades do consumidor final? (CALIGARI, 2001).

Assim, o uso da estatística está previsto em todas as etapas dos experimentos de melhoramento de plantas, desde a obtenção de genitores para formar blocos híbridos e seleção de populações segregantes até estágios avançados de registro e recomendação de variedades.

## 2.3 A aplicabilidade dos testes de média em pesquisas para melhoramento genético de plantas

A estatística consiste em uma área da ciência que utiliza a análise dos dados para verificar hipóteses estatísticas, observando o poder da evidência e as correlações entre grupos ou a autenticidade de fenômenos de interesse. Os cientistas devem elencar as possibilidades e observar os fenômenos biológicos que advêm da população, obtendo assim, as suas amostras com a finalidade de averiguar que os resultados da analogia dos dados sejam mais autênticos para a explanação das conjecturas (WINDISH; DIENER-WEST, 2006).

Além disso, a análise estatística propicia aos leitores e aos pesquisadores em geral a compreensão da informação oriunda dos dados apurados durante a realização de um estudo e assim, utilizá-la em benefício da sociedade (WINDISH; DIENER-WEST, 2006). Contudo, corriqueiramente, os cientistas fazem uso inadequado dos testes estatísticos em razão do não emprego das pressuposições necessárias. Consequentemente, pesquisadores apresentam dificuldades tanto na analogia dos dados apurados, quanto na compreensão de resultados alcançados, podendo assim, proceder com conclusões incertas (BERTOLDO et al., 2007).

Nesta perspectiva, é importante realizar a seleção do delineamento experimental apropriado atrelado ao melhoramento genético, bem como a aplicação correta da medida estatística, no intento da estimativa certa do erro experimental (BERTOLDO et al., 2007).

Para que se obtenha a aplicação correta destes, é de suma importância que os pesquisadores tenham conhecimento acerca dos tipos de variáveis existentes, de tratamentos, de fatores e do delineamento experimental que irão edificar a sua pesquisa. Entretanto, embora uma variedade de testes estatísticos

esteja à disposição dos cientistas, uma quantidade elevada ainda faz uso de forma incorreta (BERTOLDO *et al*., 2008).

Bertoldo et al. (2008), desenvolveram um estudo que buscou identificar quais os principais erros e acertos na aplicação de testes de comparação de médias em trabalhos científicos. Os autores constataram que a maior parte das pesquisas que obtinham mais de um fator foram classificadas como errôneas (72%) em razão do excesso dos testes de correlações de médias. De outro modo, 4% e 24% foram classificadas como relativamente corretas e corretas, respectivamente.

No estudo científico de várias esferas do conhecimento, as análises estatísticas são instituídas como instrumentos para avaliar suas hipóteses e à comparação multivariada entre as médias de tratamentos experimentais, antecedida pela análise de variância, que é, certamente, uma das mais utilizadas (COUTO *et al.*, 2009). Desta forma, embora haja uma certa preocupação dos cientistas com a análise dos dados, diversas vezes pode-se perceber uma negligência com a utilização destes testes e a má interpretação dos seus resultados.

Os testes de comparações de médias em tratamentos são de enorme relevância no experimento realizado (CONAGIN *et al*., 2008), quando o objetivo da pesquisa é confrontar tratamentos qualitativos. A aplicação de um teste é efetuada quando a análise de variância constata existência de efeito significante dos tratamentos à um estipulado grau de significância, de maneira em que haja a recursa da hipótese de nulidade (ao menos um contraste ortogonal sobre tratamentos desiguais de zero).

Em meio aos testes mais usuais para constatação dos resultados experimentais, estão o Teste F, aplicado à fim de obter a possibilidade de diferença significativa acerca de contrastes ortogonais dos tratamentos, e os testes de Duncan, Dunnet, Tukey, além do LSD, habitualmente utilizado para detalhar esta informação, possibilitando demonstrar estatisticamente, de maneira específica, quais tratamentos irão divergir ou não.

Para a realização da seleção do teste que irá ser aplicado, o processo dependerá de suas qualificações estatísticas, tendo em consideração os tipos de erros existentes, ou seja, observando se os mesmos estão de maneira controlada (MACHADO et al., 2005).

## 3 METODOLOGIA

A coleta de dados foi realizada, inicialmente, através da extração de fontes secundárias que representam uma revisão bibliográfica da literatura, essa retificação reúne publicações desenvolvidas por meio de diversas metodologias e que permite aos revisores a síntese dos resultados sem alterar os efeitos dos estudos desenvolvidos, fornecendo sobre o problema informações de forma mais ampla para a construção e entendimento do assunto em questão (SOUZA, 2021).

Dessa forma, foi realizada uma pesquisa bibliográfica de abordagem transversal por meio de publicações de artigos em revistas de alto impacto entre os anos de 2016 e 2021, que utilizaram em suas metodologias o teste de comparação de médias, sendo utilizado como descritores: estatística, melhoramento genético e interpretação de dados.

Após a coleta, foram selecionados 60 artigos, sendo submetidos a uma primeira classificação, de acordo com a utilização ou não de algum tipo de teste de comparação de médias. Aqueles que se utilizaram deste procedimento foram, posteriormente, agrupados conforme a sua aplicação, sendo classificados, conforme critérios empregados por Bertoldo *et al.* (2008), em: i) adequado; ii) parcialmente adequado e iii) inadequado.

A categoria adequada foi aplicada (comparação de médias aos tratamentos não relacionados e de cunho qualitativo e regressão com relação a fatores quantitativos) quando o teste de comparação de médias foi aplicado em razão dos objetivos da pesquisa e da ordem de fatores. E foi considerado inadequado para os fatores quantitativos quando não foi efetuado algum teste de comparação de médias e quando não foi averiguado o efeito da correlação em experimentos fatoriais (tendo mais de um único fator) (BERTOLDO *et al.*, 2008).

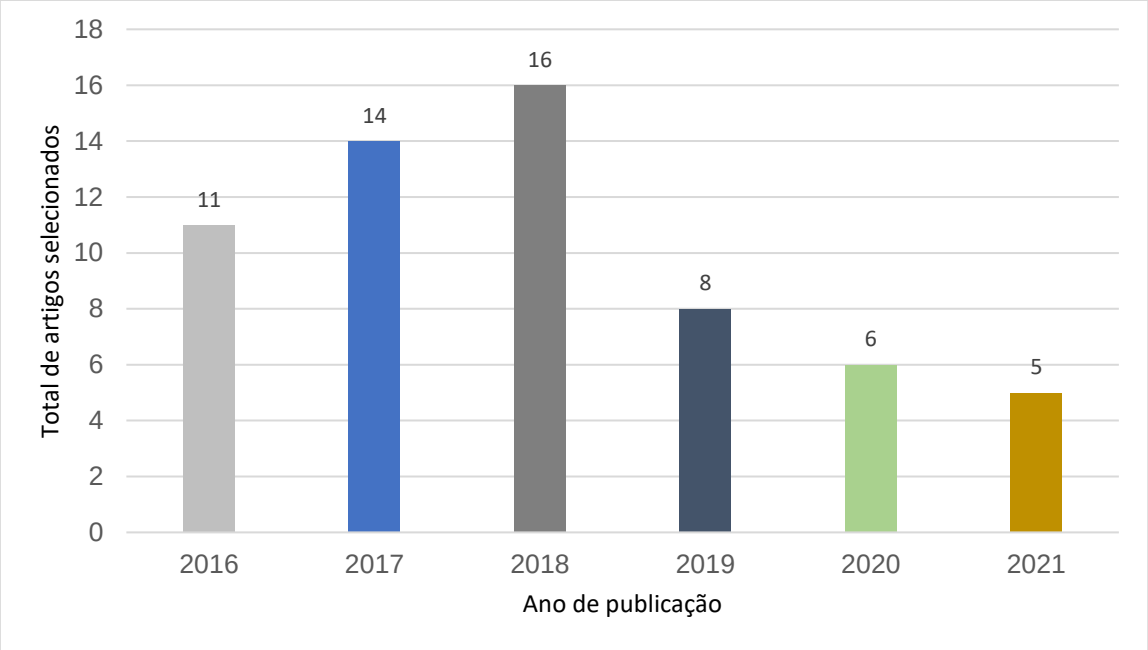
Já a designação de parcialmente adequado foi aplicada aos testes de comparações de médias (uns contra os outros) quando o apropriado seria testar contrastes preliminarmente articulados (elementos qualitativos estruturados, em que o recomendado seria o uso de contrastes ortogonais). Os estudos foram separados, com o intuito de exemplificação de qual teste foi utilizado em cada um dos trabalhos.

## 4 RESULTADOS E DISCUSSÕES

Testes estatísticos permitem aos pesquisadores extrapolar resultados experimentais. No entanto, para a correta aplicação das estatísticas deve-se compreender os tipos de fatores, variáveis de resposta, tratamentos e desenhos experimentais que compõem o estudo. Dessa forma, GRANATO; DE ARAÚJO CALADO; JARVIS, 2014), destacam que um dos objetivos da estatística é tomar decisões sobre uma população com base em observações de uma amostra, ou seja, obter conclusões válidas sobre todos os parâmetros populacionais de uma amostra extraída dessa população, onde para tentar tomar uma decisão é conveniente fazer suposições ou conjecturas sobre a população de interesse.

Em vista disso, com o intuito de identificar os principais erros e acertos na aplicação de testes de comparação de médias em estudos voltados para o melhoramento genético de plantas, depositados em periódicos de alto impacto, foi realizada a seleção de 60 estudos para análise da qualidade das estatísticas utilizadas, determinando a confiabilidade desses testes com base nas características fundamentais de cada objetivo das pesquisas selecionadas. Em relação ao quantitativo das pesquisas escolhidas e os seus respectivos anos, foi construido o gráfico abaixo para melhor entendimento (Gráfico 1).

**Gráfico 1.** Relação dos trabalhos utilizados para a análise e os seus respectivos anos de publicações




Fonte: Elaborado pelo autor (2022).

Uma análise de variância preliminar permite ao pesquisador testar a hipótese de que H é rejeitado ou não (hipótese nula). Portanto, permite verificar se existem diferenças entre os parâmetros estudados (CRO; KENWARD; CARPINTEIRO, 2015). ANOVA é uma técnica que envolve a decomposição da variância total e graus de liberdade em partes que são atribuíveis a causas conhecidas e independentes (fatores controlados). Dependendo da estrutura do experimento, os graus de liberdade devem ser expandidos para verificar os efeitos das interações entre os fatores, quando significativos. Assim, em experimentos de fatores, o primeiro passo é analisar o efeito da interação.

Dessa forma, um experimento pode apresentar um ou mais fatores, chamados de unifatorial ou fatorial, podendo ser classificados como qualitativos e/ou quantitativos. Os fatores qualitativos podem ser divididos em dois subgrupos: i) estruturados e ii) não estruturados.

Os fatores qualitativos estruturados são aqueles cujos níveis podem ser classificados em grupos, cuja comparação constitui o objetivo do trabalho, em que pesquisas sobre fatores qualitativos não estruturados visa comparar fatores experimentais em todos os níveis. Os fatores qualitativos podem ser representados por variáveis nominais, chamadas de variáveis específicas, enquanto os processos quantitativos são expressos em termos de intervalos ou variáveis racionais (BEWICK; CHEEK; BALL, 2004).

Em testes relacionados ao melhoramento genético, os experimentos são conduzidos por meio de análises fatoriais onde o nível de um fator é combinado com o nível de outro fator, de forma que as permutações de fatores permitem o estudo das interações, além de uma estimativa mais precisa da variância do erro experimental, aumentando o poder dos testes estatísticos.

Diante disso, pode-se observar (Tabela 4) que o percentual de estudos classificados como apropriados quanto ao uso de testes de comparação de médias ao tipo de fator em pesquisas para a categoria unifatorial foi de 45% (27 artigos) e fatorial de 15% (9 artigos). Enquanto o percentual de trabalhos classificados como parcialmente apropriado para as categorias unifatorial e fatorial foi de 5% (3 artigos) e 20% (12 artigos), respectivamente. Já para a classificação de inapropriados para unifatorial e fatorial foi de 5% (3 artigos) e 10% (6 artigos) respectivamente.

**Tabela 4.** Estudos selecionados e revisados quanto aos fatores (um fator ou mais de um fator) quanto ao tipo de fator ((qualitativo não estruturado (Qlne), qualitativo estruturado (Qle), quantitativo (Qt)) e quanto ao uso (apropriado, parcialmente apropriado e inapropriado), publicados em periódicos de alto impacto entre os anos de 2016 e 2021.

| **Unifatorial** | | | | | |
|---|---|---|---|---|---|
| **Classificação** | | **Qlne** | **Qle** | **Qt** | **Total** | **%** |
| Apropriado | | 15 | 2 | 10 | 27 | 45 |
| Parcialmente apropriado | | 0 | 2 | 1 | 3 | 5 |
| Inapropriado | | 2 | 1 | 0 | 3 | 5 |
| **Total** | | **17** | **5** | **11** | **33** | |
| **Porcentagem** | | **52** | **15** | **33** | | **100** |
| **Fatorial** | | | | | | |
| **Classificação** | **Adequado** | **Parcialmente Adequado** | **Inadequado** | **Total** | | |
| **[1]Ql x Ql** | 4 | 5 | 3 | 12 | | |
| **Ql x Qt** | 2 | 6 | 2 | 10 | | |
| **[2]Qt x Qt** | 1 | 0 | 0 | 1 | | |
| **Ql x Ql x Ql** | 0 | 1 | 1 | 2 | | |
| **Ql x Ql x Qt** | 1 | 0 | 0 | 1 | | |
| **Ql x Qt x Qt** | 1 | 0 | 0 | 1 | | |
| **Total** | 9 | 12 | 6 | 27 | | |
| **%** | **33** | **44** | **23** | | | |

[1]Ql = qualitativo. [2]Qt = quantitativo.

Os resultados mostram as dificuldades dos pesquisadores na análise dos dados, principalmente em experimentos de fator único e quantitativos. Em experimentos com fatores revisados e classificados como inapropriados, os autores não consideraram o efeito de interações significativas e realizaram testes de comparação de médias separados, como o teste de Tukey, para cada fator, caracterizando o mau uso dessa ferramenta.

Em vista disso, autores como Judd, Westfall e Kenny (2017) comentam que em experimentos de fatores o primeiro passo é analisar os efeitos das interações entre eles, em que para detectar interações, uma análise preliminar de variância é fundamental. Em estatística experimental, especialmente ao analisar a variância, o teste de hipóteses tem sido amplamente utilizado para tirar conclusões sobre as fontes de variação consideradas nos modelos estatísticos. Nesse caso, como o teste F é significativo para efeitos de interação,

onde os graus de liberdade devem ser rearranjados para comparar os níveis de um fator dentro dos níveis de outro fator.

Dessa forma, fixando um fator e variando o nível de outro, podem ser realizados os chamados testes de efeitos simples. Este procedimento de mudar um fator de cada vez se aplica quando o objetivo é estabelecer uma lei fundamental, o que leva a uma compreensão detalhada dos efeitos de um fator quando os outros permanecem constantes (FARZTDINOV; MCDYER, 2012).

Nessa perspectiva, a Tabela 5 mostra a utilização do teste de comparação de médias em experimentos unidirecionais e fatoriais entre os artigos revisados. Os trabalhos de fator único, 45%, 30% e 25% foram classificados de acordo com apropriados, parcialmente apropriados e inadequados, respectivamente. No experimento fatorial, 27% dos trabalhos revisados foram considerados inadequados para testes de comparação de médias, enquanto 46% foram adequados e 27% foram parcialmente adequados.

**Tabela 5.** Estudos unifatoriais e fatorial publicados em periódicos de alto impacto (2016 a 2021), quanto ao uso de testes de comparação de médias.

| **Unifatorial** | | | |
|---|---|---|---|
| **Teste** | **Adequado** | **Parcialmente adequado** | **Inadequado** |
| **T de Student** | 2 | 2 | 0 |
| **Tukey** | 5 | 3 | 1 |
| **Dunnett** | 3 | 1 | 0 |
| **Duncan** | 1 | 4 | 7 |
| **Scott-Knott** | 4 | 0 | 0 |
| **Total** | 15 | 10 | 8 |
| **Fatorial** | | | |
| **Teste** | **Adequado** | **Parcialmente adequado** | **Inadequado** |
| **T de Student** | 2 | 2 | 0 |
| **Tukey** | 9 | 1 | 3 |
| **Dunnett** | 0 | 0 | 0 |
| **Duncan** | 2 | 2 | 4 |
| **Scott-Knott** | 0 | 2 | 0 |
| **Total** | **13** | **7** | **7** |

Esses resultados se assemelham aos assumidos em pesquisa realizada por Lúcio *et al.* (2003) e Bezerra *et al.* (2002), que revisaram trabalhos publicados em periódicos de auto impacto, onde 63% e 65,6% destes foram classificados como apropriados quanto ao uso dos testes de comparação de médias, respectivamente. Por outro lado, autores como Cardellino e Siewerdt (1992) e Santos *et al*. (1998), classificaram como incorreta a maioria das pesquisas analisadas em relação aos testes de comparação múltipla utilizados na Revista da Sociedade Brasileira de Zootecnia (Qualis A Internacional) e Pesquisa Agropecuária Brasileira (Qualis A Internacional), respectivamente.

Com base nos resultados, verificou-se que a maior dificuldade para os pesquisadores está relacionada ao manejo e entendimento da interação quando significativa. A falta de compreensão do processo de aplicação e a incapacidade de interpretar os resultados são os principais motivos para a aplicação inadequada do teste de comparação de médias. Nesse viés, dependendo do tipo de fatores abrangidos pelo trabalho, a figura 1 mostra alguns procedimentos para os pesquisadores extrapolarem os resultados.

**Figura 1.** Opção para escolha do teste de média em relação ao tipo de fator


Fonte: Adaptado de Bertoldo *et al.* (2008).

Exemplos:

**Classe I – Apropriado**

Um dos trabalhas analisados e tidos como apropriado em relação ao uso de testes de média foi publicado na Revista Functional Plant Breeding Journal no ano de 2017, com o objetivo de avaliar as respostas das plantas às características fenotípicas de crisântemos em diferentes populações em duas épocas de plantio, realizada em hastes simples e cultivadas em estufas plásticas. Os fatores neste experimento foram duas épocas de plantio e oito populações de plantas (2x8). Os autores não encontraram interação significativa entre esses fatores. Como a interação não foi significativa, eles analisaram esses fatores separadamente, e para época de plantio aplicaram o teste de Duncan por ser um fator qualitativo, e para população de plantas, por ser um fator quantitativo, ajustaram as curvas de regressão para as variáveis.

Em uma mesma perspectiva, o estudo publicado na revista Proceedings of the Royal Society B-Biological Sciences no ano de 2019 com o objetivo de avaliar a qualidade fisiológica das sementes de *Delphinium* comercializadas por diferentes indústrias. Como apresentou um único fator, qualitativo específico não estruturado (quatro lotes de três diferentes empresas de importação de sementes de *Delphinium*) foi aplicado teste de Duncan para a comparação das médias de maneira correta, cujo objetivo era comparar todos contra todos. Apesar da correta utilização de teste de comparação de média nesse caso, a escolha do teste de Duncan não é adequada em estudos com três ou mais médias sendo comparadas, a teoria do teste de Duncan é inadequada, pois o nível de significância global não é mantido. Como alternativa, os autores poderiam utilizar teste de Tukey, Scheffé ou ainda Bonferroni.

O terceiro exemplo é dado por uma pesquisa publicada no ano de 2016, que teve por objetivo a avaliação do desempenho de plantas enxertadas em seis porta-enxertos. Como os fatores do estudo foram qualitativos (porta-enxertos) e o objetivo foi comparar todos os tratamentos entre si, a comparação entre médias foi um procedimento adequado. Portanto, os autores utilizaram o teste de Duncan para extrapolar corretamente os resultados obtidos.

**Classe II – Parcialmente Apropriado**

Dos estudos classificados como parcialmente apropriado, tem-se como exemplo a pesquisa publicada na revista The FASEB Journal, que teve como objetivo principal caracterizar química e fisicamente determinados substratos e misturas (combinações), e comparar os resultados de formulações de substratos utilizados para o cultivo de mudas e flores em recipientes, utilizando o teste de Duncan. Os tratamentos avaliados no estudo foram: solo + areia; solo + areia + RDCA; solo + areia + CAC; turfa SCv + CAC e turfa SCv + RDCA.

Como o objetivo foi igualar um tratamento específico, comparações por contrastes não ortogonais (como comparações de teste de médias) não são os procedimentos mais adequados, pois para obter dados mais informativos é necessário comparar tratamentos com os quais fatores realmente fazem a diferença. Por exemplo, comparando os tratamentos solo+areia (1:1) e solo+areia+RDCA, onde os pesquisadores podem identificar se o RDCA é um fator na avaliação final.

Como os dois tratamentos são significativamente diferentes fisicamente, ao aplicar contrastes ortogonais, os resultados podem ser melhor explorados, identificando corretamente qual fator está afetando positivamente (aumentando a média do tratamento) ou negativamente (diminuindo a média), testado pelo teste F ou teste t. Nesse caso, o teste de Duncan não pode determinar se o RDCA oferece uma vantagem porque envolve apenas dois meios.

Portanto, como alternativa, algumas comparações ortogonais entre os tratamentos podem ser realizadas para testar o efeito de diferentes combinações.

**Classe III – Inapropriado**

Em relação aos estudos analisados e classificados como inapropriado, tem-se como exemplo um estudo que avaliou o efeito do ácido indolbutírico no enraizamento de estacas da variedade *Hayward* (Kiwi). Neste experimento, os autores tiveram apenas um fator: cinco doses de AIB, onde de acordo com o objetivo, os fatores estudados foram quantitativos, pois os pesquisadores estavam interessados em validar o intervalo entre os níveis de AIB, ou seja, validar a concentração ideal no enraizamento de estacas.

Em busca de conclusões, os autores utilizaram um teste de comparação de médias (teste de Duncan a 5% de significância), enquanto as concentrações de 4.000, 6.000 e 8.000 não diferiram, eles concluíram que a concentração de 6.000 ppm proporcionou o melhor resultado para estacas enraizadas. Como o objetivo foi a análise de intervalo de dose, a aplicação de testes comparativos foi incorreta. Portanto, é suficiente ajustar a equação de regressão para diferentes concentrações de AIB, o que pode indicar que concentrações entre 4.000 ppm e 6.000 ppm são mais eficazes.

Dessa forma, ao comparar tratamentos usando experimentos fatoriais ou níveis de fatores quantitativos, graus de liberdade ou somas de quadrados devem ser divididos para corresponder aos efeitos ou interações principais, caso em que a regressão é a técnica apropriada, onde se a regressão for significativa, não são necessárias comparações múltiplas.

## CONCLUSÃO

Diante do que foi exposto, pode-se concluir que os principais aspectos que afetam diretamente a confiabilidade da interpretação dos resultados do pesquisador são: i) premissas que não levam em conta diferentes testes estatísticos, ii) conhecimento preliminar sobre o tipo de fatores e iii) Inconsistências na seleção de testes estatísticos que podem levar a conclusões incompletas e/ou inadequadas.

Dessa maneira, devido à variedade de testes de comparação de médias disponíveis e à dificuldade dos pesquisadores em descobrir o melhor procedimento a ser utilizado, métodos baseados no tipo de dados disponíveis e nos objetivos são frequentemente propostos para minimizar erros causados por escolhas erradas, buscando favorecer aos cientistas conclusões confiáveis sobre os tratamentos que estão sendo estudados, construindo inferências mais efetivas.

## REFERÊNCIAS

ALVAREZ, V. H.; ALVAREZ, G.A.M. Comparação de médias ou teste de hipóteses? Contrastes! **Sociedade Brasileira de Ciências do Solo**. Viçosa, v. 3, n. 1, p. 24-33, 2006.

BAKER, R. Jacob. **Selection indices in plant breeding**. CRC Press, 2020.

BERTOLDO, J. G. *et al*. Teste de comparação de médias: dificuldades e acertos em artigos científicos. **Current Agricultural Science and Technology**, v. 13, n. 4, 2007.

BERTOLDO, J. G. *et al*. Teste de comparação de médias: dificuldades e acertos em artigos científicos. **Current Agricultural Science and Technology**, v. 13, n. 4, 2007.

BERTOLDO, J. G. *et al*. Uso ou abuso em testes de comparações de média: ¿conhecimento científico ou empírico? **Ciência Rural**, v. 38, p. 1145-1148, 2008.

BETRÁN, Javier et al. Theory and application of plant breeding for quantitative traits. **and farmer participation**, p. 27, 2009.

BEWICK, Viv; CHEEK, Liz; BALL, Jonathan. Statistics review 9: one-way analysis of variance. **Critical care**, v. 8, n. 2, p. 1-7, 2004.

BEZERRA NETO, Francisco; NUNES, Glauber Henrique S.; NEGREIROS, Maria Zuleide de. Avaliação de procedimentos de comparações múltiplas em trabalhos publicados na revista Horticultura Brasileira de 1.983 a 2.000. **Horticultura brasileira**, v. 20, p. 5-9, 2002.

CALIGARI, Peter DS. Plant breeding and crop improvement. **e LS**, 2001.

CARDELLINO, R. A.; SIEWERDT, F. Utilização adequada e inadequada dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, v. 21, n. 6, p. 985-995, 1992.

CECCARELLI, Salvatore. Efficiency of plant breeding. **Crop Science**, v. 55, n. 1, p. 87-97, 2015.

CONAGIN, A.; BARBIN, D.; DEMÉTRIO, C. G. B. Modifications for the Tukey test procedure and evaluation of the power and efficiency of multiple comparison procedures. **Scientia Agricola**, v. 65, 2008.

COSTA, J. R. **Técnicas experimentais aplicadas às ciências agrárias**. Seropédica: Embrapa Agroecologia, 2003. 102p.

COUTO, M. R. M. *et al*. Transformações de dados em experimentos com abobrinha italiana em ambiente protegido. **Ciência Rural** [online], v. 39, n. 6, 2009.

CRO, Suzie; KENWARD, Michael; CARPENTER, James. Variance estimation in reference based sensitivity analysis for longitudinal trials with protocol deviation. **Trials**, v. 16, n. 2, p. 1-1, 2015.

DA ROCHA, Keslei Rosendo; JÚNIOR, Arilton Januário Bacelar. ANOVA medidas repetidas e seus pressupostos: análise passo a passo de um experimento. **Revista Eletrônica Perspectivas da Ciência e Tecnologia-ISSN: 1984-5693**, v. 10, p. 29, 2018.

DA SILVA, Gonçalo Mesquita et al. Planejamento de experimento a Pasto. **PUBVET**, v. 10, p. 356-447, 2016.

DUNNETT, C. W. A multiple comparision procedure for comparing several treatments with a control. Journal Aimer. Statist. **Assosc., Washington**, v. 50, p. 1096-1121, 1955.

EMERSON, Robert Wall. ANOVA e testes t. **Journal of Visual Impairment & Blindness** , v. 111, n. 2, pág. 193-196, 2017.

FARZTDINOV, Vadim; MCDYER, Fionnuala. Distributional fold change test–a statistical approach for detecting differential expression in microarray experiments. **Algorithms for Molecular Biology**, v. 7, n. 1, p. 1-16, 2012.

FERREIRA, D. F.; MUNIZ, J. A.; AQUINO, L. H. Comparações múltiplas em experimentos com grande número de tratamentos – utilização do teste de Scott Knott. **Ciência Agrotecnologia**, Lavras, v. 23, p. 745-752, 1999.

GIRARDI, Luís Henrique; CARGNELUTTI FILHO, Alberto; STORCK, Lindolfo. Erro tipo ie poder de cinco testes de comparação múltipla de médias. **Rev. Bras. Biom**, v. 27, n. 1, p. 23-36, 2009.

GRANATO, Daniel; DE ARAÚJO CALADO, Verônica Maria; JARVIS, Basil. Observations on the use of statistical methods in food science and technology. **Food Research International**, v. 55, p. 137-149, 2014.

HOTHORN, Ludwig A. The two-step approach—a significant ANOVA F-test before Dunnett's comparisons against a control—is not recommended. **Communications in Statistics-Theory and Methods**, v. 45, n. 11, p. 3332-3343, 2016.

JUDD, Charles M.; WESTFALL, Jacob; KENNY, David A. Experiments with more than one random factor: Designs, analytic models, and statistical power. **Annual review of psychology**, v. 68, n. 1, p. 601-625, 2017.

KIM, Tae Kyun. Understanding one-way ANOVA using conceptual figures. **Korean journal of anesthesiology**, v. 70, n. 1, p. 22-26, 2017.

LÚCIO, Alessandro Dal'Col et al. Características experimentais das publicações 9 da Ciência Rural de 171 a 2000. **Ciência Rural** , v. 33, p. 161-164, 2003.

MACHADO, A. A.; DEMÉTRIO, C. G. B.; FERREIRA, D. F.; SILVA, J. G. C. Estatística experimental: uma abordagem fundamental no planejamento e no uso de recursos computacionais. *In*: Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, Londrina. **Anais** [...] **Reunião Brasileira da Sociedade Internacional de Biometria,** 2005.

MAHBOBI, Mohammad; TIEMANN, Thomas K. F-Test and One-Way ANOVA. **Introductory Business Statistics with Interactive Spreadsheets-1st Canadian Edition**, 2015.

MARINI, Federico et al. Analysis of variance of designed chromatographic data sets: The analysis of variance-target projection approach. **Journal of Chromatography A**, v. 1405, p. 94-102, 2015.

MINAMOTO, Toshifumi et al. Techniques for the practical collection of environmental DNA: filter selection, preservation, and extraction. **Limnology**, v. 17, n. 1, p. 23-32, 2016.

MONTGOMERY, D. C; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros.** 2° edição. Rio de Janeiro: LTC, 2003.

OLIVEIRA, Andréia Fróes Galuci. Testes estatísticos para comparação de médias. **Revista Eletrônica Nutritime**, v. 5, n. 6, p. 777-788, 2008.

PAGOTTO, Lyvia Gonzalez et al. Análise de variância e testes de médias: um estudo aplicado em experimentos com variedades de algodoeiro e seleções de citrumelo. **Brazilian Applied Science Review**, v. 5, n. 3, p. 1287-1296, 2021.

PHAM, Hung Viet et al. Problems and opportunities in training deep learning software systems: An analysis of variance. In: **Proceedings of the 35th IEEE/ACM international conference on automated software engineering**. 2020. p. 771-783.

PIMENTEL, Adérico Júnior Badaró et al. Estimação de parâmetros genéticos e predição de valor genético aditivo de trigo utilizando modelos mistos. **Pesquisa Agropecuária Brasileira**, v. 49, p. 882-890, 2014.

PROETTI, Sidney. As pesquisas qualitativa e quantitativa como métodos de investigação científica: Um estudo comparativo e objetivo. **Revista Lumen-ISSN: 2447-8717**, v. 2, n. 4, 2018.

RALSTON, Kevin; DUNDAS, Ruth; LEYLAND, Alastair H. Comparação do Índice Escocês de Privação Múltipla (SIMD) 2004 com o SIMD 2009+ 1: a escolha da medida afeta a interpretação da desigualdade na mortalidade?. **International Journal of Health Geographics** , v. 13, n. 1, pág. 1-6, 2014.

RICHARDS, Robert J. Darwin's theory of natural selection and its moral purpose. In: **Debates in Nineteenth-Century European Philosophy**. Routledge, 2015. p. 211-225.

SALMAN, Ana Karina Dias; GIACHETTO, Poliana Fernanda. Conceitos estatísticos aplicados à experimentação zootécnica. **Embrapa Rondônia-Artigo em periódico indexado (ALICE)**, 2014.

SANTOS, JOSÉ WELLINGTON; MOREIRA, JOSÉ DE ALENCAR NUNES. Avaliação do emprego dos testes de comparação de médias na revista pesquisa agropecuária brasileira (pab) de 1980 a 19941. **Pesq. agropec. bras., Brasília**, v. 33, n. 3, p. 225-230, 1998.

SCHOBER, Patrick; VETTER, Thomas R. Analysis of variance in medical research. **Anesthesia & Analgesia**, v. 131, n. 2, p. 508-509, 2020.

SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**, New Haven, v.30, n. 2, p. 507-512, 1974.

SCOTT, Stacey B. et al. A coordinated analysis of variance in affect in daily life. **Assessment**, v. 27, n. 8, p. 1683-1698, 2020.

SCUDINO, Patrıcia Araújo. A Utilizaçao de Alguns Testes Estatısticos para Análise da Variabilidade do Preço do Mel nos Municıpios de Angra dos Reis e Mangaratiba, Estado do Rio de Janeiro. **Repositório ufrrj**, 2008.

SILVA, C. M. R. **Uso do teste de Scott-Knott e da análise de agrupamentos, na obtenção de grupos de locais para experimentos com cana-de-açúcar**. 2007. 48p. Dissertação (Mestrado em Agronomia: Estatística e Experimentação) – Escola Superior de Agricultura Luiz de Queiroz, São Paulo, 2007.

SILVA, Francisco de Assis Santos; AZEVEDO, Carlos Alberto Vieira. Comparison of means of agricultural experimentation data through different tests using the software Assistat. **African Journal of Agricultural Research**, v. 11, n. 37, p. 3527-3531, 2016.

SOUZA, Elza Maria de; SILVA, Daiane Pereira Pires; BARROS, Alexandre Soares de. Educação popular, promoção da saúde e envelhecimento ativo: uma revisão bibliográfica integrativa. **Ciência & Saúde Coletiva**, v. 26, p. 1355-1368, 2021.

STUDENT, The problable error of a mean. Biometrika. **Oxfird**, v. 06, p. 01-25, 1908.

TAVARES, Helen Hana Fernandes et al. Factors associated with Burnout Syndrome in medical students. **Mundo da Saúde**, v. 44, n. 1, p. 280-289, 2020.

TUKEY, J. W. **Comparing individual means in the analysis of variance.** Biometrics, New Haven, v. 05, n. 02, p. 99-114, 1949.

WANG, Shuai et al. Comparison of multiple single-nucleotide variant association tests in a meta-analysis of Genetic Analysis Workshop 19 family and unrelated data. In: **BMC proceedings**. BioMed Central, 2016. p. 187-191.

WINDISH, D. M.; DIENER-WEST, M. A clinician-educator's roadmap to choosing and interpreting statistical tests. **Journal of general internal medicine**, v. 21, n. 6, 2006.

ZABELL, S. L. **On student's 1908 article "the probable error of a mean".** Journal of the american statistical association, Washington, v. 103, n. 481, p. 01-07, 2008.

ZOANETTI, Nathan et al. Fixed or mixed: a comparison of three, four and mixed-option multiple-choice tests in a Fetal Surveillance Education Program. **BMC Medical Education**, v. 13, n. 1, p. 1-11, 2013.