



Considerações sobre planejamento experimental e métodos estatísticos em Ciências Agrárias

Considerations on experimental planning and statistical methods in Agricultural Sciences

Marcelo Cavalcante⁽¹⁾; João Gomes da Costa⁽²⁾

⁽¹⁾ORCID: <https://orcid.org/0000-0002-0722-0618>; Instituição Federal de Alagoas/Docente, pesquisador permanente do Programa de Mestrado Profissional em Tecnologias Ambientais, BRAZIL, E-mail: marcelo.cavalcante@ifal.edu.br;

⁽²⁾ORCID: <http://orcid.org/0000-0002-0761-0755>; Embrapa Alimentos e Territórios/Pesquisador, Bolsista de Produtividade em Pesquisa do CNPq, BRAZIL, E-mail: joao-gomes.costa@embrapa.br.

Todo o conteúdo expresso neste artigo é de inteira responsabilidade dos seus autores.

Recebido em: 8 de junho de 2021; Aceito em 19 de julho de 2021; publicado em 10 de outubro de 2021. Copyright© Autor, 2021.

RESUMO: Na pesquisa científica, o planejamento e execução de experimentos consistem em práticas comuns nas Ciências Agrárias. Apesar de os métodos experimentais serem amplamente conhecidos, percebe-se que, muitas vezes, suas pressuposições não são verificadas, refletindo em baixa precisão e qualidade dos resultados. Esta pesquisa objetiva, por meio de uma revisão de literatura: 1) discutir conceitos básicos da estatística experimental; e 2) apresentar os principais modelos e métodos estatísticos utilizados em Ciências Agrárias. Foi executado um levantamento bibliográfico, por meio de publicações de livros e artigos científicos obtidos nas bases de dados Periódicos Capes, SciELO, Scopus e Google Scholar. No planejamento experimental, a escolha do delineamento deverá considerar a heterogeneidade ambiental. Antes de testar a hipótese científica, deve-se testar as pressuposições da Anova. A transformação de dados é uma ferramenta eficiente, muitas vezes, necessária, mas pouco explorada. Ao se testar tratamentos qualitativos, utiliza-se teste de comparação/agrupamento de médias, sendo o teste Scott-Knott o mais robusto. Quando quantitativo, regressão linear. Quando avaliações sucessivas são realizadas na mesma unidade experimental, utiliza-se a análise de medidas repetidas. O coeficiente de variação indica a precisão experimental e varia com os tratamentos, variáveis e ambiente. O coeficiente de repetibilidade estima o número de avaliações e de experimentos a serem realizados, com alto grau de precisão. O planejamento experimental permite economia de tempo, custos e mão de obra, sem comprometer a precisão das conclusões.

PALAVRAS-CHAVE: Erro Experimental, Reprodutibilidade, Anova, Precisão.

ABSTRACT: In scientific research, planning and carrying out experiments are common practices in Agricultural Sciences. Although the experimental methods are widely known, it is clear that their assumptions are often not verified, reflecting the low precision and quality of the results. This research aims, through a literature review: 1) discuss basic concepts of experimental statistics; and 2) present the main statistical models and methods used in Agricultural Sciences. A bibliographic survey was carried out through publications of books and scientific articles obtained from the Capes Periodicals, SciELO, Scopus and Google Scholar databases. In planning, the choice of design should consider environmental heterogeneity. Before testing the scientific hypothesis, one must test the Anova assumptions. Data transformation is an efficient tool that is often necessary but little explored. When testing qualitative treatments, a comparison/grouping of means test is used, with the Scott-Knott test being the most robust. When quantitative, linear regression. When successive evaluations are carried out in the same experimental unit, repeated measures analysis is used. The coefficient of variation indicates experimental precision and varies with treatments, variables and environment. The repeatability coefficient estimates the number of evaluations and experiments to be carried out, with a high degree of precision. Experimental planning saves time, costs and labor without compromising the accuracy of conclusions.

KEYWORDS: Experimental Error, Reproductivity, Anova, Precision.

INTRODUÇÃO

A estatística experimental é uma ciência que usa a análise de dados para testar hipóteses científicas. Segundo Bhering & Teodoro (2021), o uso inadequado de modelos e métodos estatísticos gera erros que podem levar a resultados e discussões falsas e motivar conclusões equivocadas. Quando a pesquisa é planejada adequadamente, permite dirimir erros e aumentar a eficiência na execução dos projetos, com reflexos na redução de tempo, custos, infraestrutura e mão de obra.

Conceitualmente, um experimento deverá ter os princípios da repetição, da casualização e, em alguns casos, o princípio do controle local. As repetições permitem a obtenção do erro experimental, utilizadas nos testes de hipóteses e comparações múltiplas. A casualização garante a independência dos resíduos, pré-requisito para análise de variância. Já o controle local estratifica o ambiente heterogêneo em subunidades homogêneas, de modo que a variação dentro do bloco seja a menor possível, enquanto que, entre blocos, poderá ser grande ou pequena (FERREIRA, 2018).

A análise matemática dos diferentes métodos estatísticos, antes não compreensível à grande maioria dos pesquisadores, considerando sua complexidade, foi superada com a disseminação de conhecimento em linguagem mais acessível, nas diferentes áreas do conhecimento. A bioinformática gerou, ao longo dos anos, uma grande quantidade de softwares estatísticos, tornando ainda mais acessível às análises de dados. Porém, o planejamento, a execução do projeto, a obtenção dos dados, a atenção aos requisitos para aplicação de cada método estatístico e a interpretação dos resultados exige do pesquisador conhecimento em estatística experimental.

Bertoldo et al. (2007) elencaram os principais problemas que afetam a confiança na interpretação dos resultados por parte dos pesquisadores, que são: 1) a não consideração das pressuposições dos diferentes testes estatísticos; 2) conhecimento incipiente em relação aos tipos de tratamentos avaliados; 3) incoerência na seleção dos testes estatísticos que acarreta em conclusões incompletas e/ou inapropriadas.

Dentro das diferentes áreas do conhecimento, projetos com aplicação equivocada de conceitos básicos da estatística experimental ou a ausência de informações necessárias a reprodutibilidade da pesquisa, requisito essencial que garante transparência e credibilidade na ciência, conduzem a resultados imprecisos, de baixa confiabilidade científica. Nas Ciências Agrárias, ainda se observa, com frequência, planejamento e uso de modelos e métodos estatísticos sem garantias às pressuposições necessárias, demonstrando a falta de conhecimento para condução de pesquisa científica aplicada.

Ao revisar 292 artigos publicados em periódicos de Ciências Agrárias no período de 2000 a 2006, Bertoldo et al. (2008a) observaram que 75% foram classificados como inapropriado devido ao uso abusivo dos testes de comparação de médias, inclusive quando os tratamentos foram quantitativos. Bertoldo et al. (2008b) encontraram resultados semelhantes (72% incorretos) ao analisar 226 artigos publicados no periódico Ciência Rural. Do mesmo modo, Possatto Jr. et al. (2019), observaram que 60,8% dos 1.237 artigos publicados pelo periódico Acta Scientiarum. Agronomy entre 1998 e 2016, erraram na aplicação de testes de comparação de médias.

Pelas razões apresentadas, esta pesquisa objetiva, por meio de uma revisão de literatura: 1) discutir conceitos básicos do planejamento experimental; e 2) apresentar os principais modelos e métodos estatísticos utilizados em Ciências Agrárias.

PROCEDIMENTO METODOLÓGICO

Nesta pesquisa, foi executado um levantamento bibliográfico, por meio de publicações de livros e artigos científicos obtidos nas seguintes bases de dados: Periódicos Capes, SciELO, Scopus e Google Scholar. Adotaram-se as seguintes palavras-chave: delineamento inteiramente casualizado (*randomized complete design*), delineamento em blocos casualizados (*randomized complete block design*), análise de variância (*analysis of variance*), erro experimental (*experimental error*), distribuição normal (*normal distribution*), homocedasticidade (*homocedasticity*), testes Tukey, Duncan, Dunnett, Scott-Knott, regressão polinomial (*polynomial regression*), AIC, BIC, medidas repetidas no tempo (*repeated measures in time*), esfericidade (*sphericity*), coeficiente de variação (*coefficient of variation*), coeficiente de repetibilidade (*repeatability coefficient*), inferência Bayesiana (*Bayesian inference*).

Levou-se em consideração aspectos ligados: 1) ao planejamento experimental; 2) aos testes de hipóteses e tipos de erros envolvidos; 3) a análise de variância (Anova), seus pressupostos, transformação dos dados e testes envolvidos; 4) aos delineamentos experimentais básicos (DIC e DBC); 5) a análise estatística de tratamentos qualitativos (testes de comparação/agrupamento de médias) e quantitativos (regressão polinomial); 6) as medidas repetidas no tempo (métodos univariado, multivariado e modelos mistos); 7) ao coeficiente de variação na Anova; e 8) a análise de repetibilidade (métodos da Anova, componentes principais, análise estrutural e inferência Bayesiana).

Nesta pesquisa não foi abordada a descrição matemática de cada um dos métodos estatísticos, considerando que estas informações poderão ser obtidas na literatura, que é rica e de fácil aquisição.

RESULTADOS E DISCUSSÃO

Planejamento experimental

A principal etapa de uma pesquisa científica é a de planejamento experimental. Nesta, o pesquisador deverá descrever a justificativa/importância para execução do projeto, os objetivos e, principalmente, a metodologia, com detalhamento do ambiente experimental, do modelo estatístico, dos tratamentos, repetições, parcelas e unidades experimentais, do manejo a ser adotado, das variáveis e métodos de avaliação, e dos testes estatísticos que validarão os resultados, de forma precisa e acurada, de modo que as conclusões possam ser apresentadas à comunidade científica e extrapoladas para os setores de produção. O detalhamento metodológico permitirá ainda que a pesquisa tenha o caráter reprodutibilidade, podendo ser reproduzida pelos pesquisadores.

Teste de hipótese

Na pesquisa experimental, a hipótese científica corresponde a uma teoria que poderá ser validada/testada por um método científico, permitindo-se obter conclusões a partir de testes de hipóteses. Existem duas hipóteses possíveis: H_0 (nulidade), em que não há diferença estatística entre os tratamentos; H_1 (alternativa), indica diferença significativa entre os tratamentos.

Ao testar hipóteses, pode-se cometer dois tipos de erros: o tipo I, também chamado falso positivo, ocorre quando se rejeita H_0 , indicando diferença estatística quando, na verdade, as diferenças observadas são devidas ao acaso; o tipo II, também chamado falso negativo, ocorre quando o teste aceita H_0 , quando falsa. A escolha do método para testar hipótese deve considerar o nível de significância (α) e o poder do teste (β).

O nível de significância corresponde à probabilidade máxima aceitável de cometer o erro do tipo I. Em Ciências Agrárias, por convenção, adota-se 5% de probabilidade ($P\text{-valor} = P < 0,05$) na maioria das pesquisas, ou seja, 95% de confiança de

que o resultado é correto ($1 - \alpha$). Já o poder do teste é a probabilidade de rejeitar a hipótese de que não há diferença significativa entre os tratamentos (erro tipo II). O teste F é o mais utilizado para comparação de variâncias independentes.

Análise de variância

A variância (s^2) é uma medida de dispersão ou variabilidade dos dados, estudada na estatística descritiva. Porém, na experimentação, a variância é desdobrada em causas conhecidas e desconhecidas. Às causas conhecidas, denominam-se de variação premeditada (tratamentos) e variação externa (blocos) e, às causas desconhecidas, chamadas de variação acidental (erro experimental ou resíduo). A análise de variância (Anova), também denominada análise paramétrica, é o método que permite identificar ou não diferença estatística entre dois ou mais tratamentos, qualitativos ou quantitativos.

Tem-se como critérios para o uso da Anova: 1) a aditividade, 2) a independência dos resíduos, 3) a distribuição normal e 4) a homogeneidade de variâncias residuais. Se estas pressuposições não forem checadas poderão comprometer a validade dos resultados dos testes e as estimações realizadas (NOGUEIRA; PEREIRA, 2013).

Quando os parâmetros, fixos e/ou aleatórios, do modelo estatístico é aditivo, atende-se à pressuposição de aditividade. Cada delineamento estatístico tem seu modelo linear aditivo. O delineamento inteiramente casualizado (DIC) considera o modelo $Y_{ij} = m + t_i + e_{ij}$, em que Y_{ij} é o valor observado da variável estudada que recebeu o tratamento i na repetição j ; 'm', é a média geral do experimento; t_i corresponde ao efeito do tratamento i ; e_{ij} é o erro experimental na parcela que recebeu o tratamento i . Já o delineamento em blocos casualizados (DBC), adiciona-se ao modelo o efeito de blocos (b_j), tendo-se $Y_{ij} = m + b_j + t_i + e_{ij}$. Erros na instalação do experimento e na tabulação dos dados conduzem a não aditividade, a heterogeneidade de variâncias e maior probabilidade de erro tipo II. O teste Tukey é utilizado para avaliar a não aditividade ($P > 0,05$; aceita H_0 , o modelo é aditivo).

O erro de uma observação deverá ser independente dos demais, garantindo, com isso, a independência dos resíduos. A ausência de casualização e ainda, medidas repetidas no tempo em uma unidade experimental, promovem dependência dos resíduos, tendo-se, como consequência, a heterocedasticidade (ausência de homogeneidade de variância residuais), de modo que os testes de significância não serão válidos. Utiliza-se o teste de Durbin-Watson para verificação de independência dos resíduos ($P > 0,05$; aceita H_0).

A distribuição normal é o pressuposto mais estudado, considerando que a maioria dos fenômenos biológicos segue esta distribuição. Baseia-se na média e no desvio padrão e indica que os dados experimentais de variáveis contínuas refletem a população estudada. *Outliers* (dados discrepantes), problemas de amostragem (pequeno número de repetições), natureza da variável-resposta conduz a problemas na distribuição normal, tendo-se, como consequência, maior probabilidade do erro tipo I. A análise gráfica, a partir de histogramas de frequência, consiste em uma avaliação subjetiva e deverá ser evitada. Por outro lado, o teste de Shapiro-Wilk é um dos mais robustos para avaliação de normalidade ($P > 0,05$; aceita H_0).

A experimentação é baseada na premissa de que os testes de hipóteses são realizados pressupondo que os desvios, devido ao erro experimental associados aos tratamentos, devem possuir uma variância comum (NOGUEIRA; PEREIRA, 2013). Situações como *outliers*, violação de outras pressuposições da Anova, número reduzido de parcelas, desbalanceamento (tratamentos com número diferentes de repetições) são causadores de heterocedasticidade, tendo-se, como consequência, a perda de validade de testes estatísticos que utilizam o quadrado médio do resíduo (Q_{MR}) em suas equações, a exemplo do teste F. O teste de Levene ou Bartlett são os mais indicados para indicar a homocedasticidade ($P > 0,05$; aceita H_0).

Nos casos de violação dos pressupostos da Anova, utiliza-se, como estratégia, a transformação dos dados, que consiste em uma ferramenta eficiente, usada para normalização dos resíduos e estabilização de variâncias.

A natureza da variável-resposta, a exemplo de dados de contagem (N° frutos, folhas, ovos) e escala de notas, que seguem distribuição de Poisson, conduz a problemas na distribuição normal, indicando-se a transformação raiz quadrada. Dados de percentagem, que seguem distribuição binomial, utiliza-se a transformação angular. Dados provenientes de contagem de organismos, em que há proporcionalidade entre as médias e os desvios padrões dos tratamentos, utiliza-se a transformação logarítmica.

Mais recentemente, o método Box-Cox vem sendo utilizado na transformação dos dados em pesquisas das Ciências Agrárias por identificar a melhor transformação, com base no valor de lambda, permitindo que as pressuposições de homocedasticidade e normalidade sejam simultaneamente satisfeitas (AZEVEDO et al., 2016b).

Após transformados, o pesquisador deverá averiguar a sua efetividade, testando-se novamente as pressuposições para Anova, pois a seleção de uma transformação equivocada poderá agravar o problema (AZEVEDO et al., 2016b). Por outro lado, persistindo a violação, o pesquisador poderá optar pela análise não paramétrica ou pelos

modelos lineares generalizados (MLG). Atendidas as pressuposições, todas as análises estatísticas deverão ser realizadas na escala de transformação escolhida e, na apresentação dos resultados, utilizam-se as médias originais.

Na Anova, o teste F é utilizado para testar a hipótese H_0 . No DIC, a fonte de variação corresponde aos tratamentos, não existindo o princípio do controle local. Quando em DBC, além dos tratamentos, adicionam-se os blocos como fonte de variação. Uma análise equivocada é considerar o F de blocos, quando significativo, como indicador de ambiente heterogêneo. Os tratamentos têm repetições. Porém, cada um dos blocos, não. Portanto, variância de blocos (quadrado médio) indica que a heterogeneidade ambiental foi controlada, utilizando-se para estimar o QM_R . Teste F de blocos não deverá ter interpretação válida.

Nos arranjos fatoriais e em parcelas subdivididas, com dois ou mais fatores, bem como na interação Genótipo x Ambiente (G x E), o teste F também avalia a significância das interações. Quando não significativa ($P > 0,05$), a discussão dos resultados deverá abordar cada fator individualmente, pois são independentes. Por outro lado, sendo significativa ($P < 0,05$), indica que a resposta de um fator depende da presença ou ausência do outro. Portanto, a discussão deverá ser feita somente sobre a interação, mesmo que cada fator seja significativo.

Delineamentos experimentais

Os tratamentos poderão ser fixos, quando definidos pelo pesquisador, ou aleatórios, que corresponde a uma amostra representativa de uma população. O delineamento experimental é a forma com que os tratamentos e suas repetições são distribuídos nas 'n' parcelas. Em Ciências Agrárias, os delineamentos inteiramente casualizado (DIC) e em blocos casualizados (DBC) são amplamente utilizados.

O controle local consiste em um dos princípios da estatística experimental (FERREIRA, 2018), baseado na identificação e no isolamento do fator promotor de heterogeneidade do ambiente. Porém, a percepção desses fatores depende da experiência/vivência científica da equipe, ou seja, algumas fontes de variação externa poderão não ser identificadas, levando o pesquisador afirmar que o ambiente é homogêneo. Outro pesquisador, porém, poderá identificar variações no mesmo ambiente. Portanto, além da percepção da equipe, conhecer os tratamentos, o histórico da área e uma avaliação detalhada do local (análises físico-química-microbiológica em

experimentos de campo) onde será instalado o experimento promoverá a escolha mais assertiva do delineamento experimental, reduzindo sensivelmente os erros.

O DIC é utilizado quando há homogeneidade do ambiente, em que os tratamentos com suas respectivas repetições são distribuídos (sorteadas - casualização) nas parcelas. A Anova tem o máximo número de graus de liberdade do resíduo (GL_R) e, conseqüentemente, o menor QM_R , permitindo maior precisão experimental, considerando que tais parâmetros compõem o estimador do teste F (QM_T/QM_R), permitindo maior precisão, reduzindo os erros tipo I e II.

Quando o pesquisador observar fatores que promovem heterogeneidade no ambiente experimental, a exemplo de variação da temperatura e luminosidade em nível laboratorial/casa de vegetação, do tipo de solo e da fertilidade, da declividade do terreno, de práticas de manejo adotadas em nível de campo, ou no sexo, peso corporal, na idade em experimentação animal, são fontes de variação externa que deverão ser controladas. O DBC é utilizado para controlar as variações ambientais. Os blocos, como fonte de variação na Anova, reduzem os GL_R e aumentam o QM_R , reduzindo a precisão experimental.

Apesar de o critério está estabelecido, convencionou-se, erroneamente, vincular a adoção do delineamento segundo o critério 'local da pesquisa', ou seja, quando executada em laboratório ou em nível de campo, utilizam-se o DIC ou DBC, respectivamente, admitindo-se que estes ambientes são homogêneos e heterogêneos, respectivamente. Embora a premissa possa ser verdadeira, o pesquisador poderá observar heterogeneidade em nível de laboratório e homogeneidade em nível de campo. Cada ambiente experimental é único e precisa ser avaliado, *in loco*, pelo pesquisador.

Adotar o DIC em um ambiente heterogêneo poderá favorecer tratamentos, reduzindo a precisão experimental e aumentar os tipos de erros. Por outro lado, adotar DBC em um ambiente homogêneo reduzirá a precisão do teste F e dos testes de comparação de médias, devido à redução dos GL_R . Ainda, adotar o DBC sem conhecer a fonte de variação externa impossibilita a homogeneidade dentro do bloco e a igualdade de condições entre os tratamentos, causando a dependência dos erros. Estas situações, além de fornecer resultados de baixa precisão, poderão refletir negativamente na normalidade dos dados e na homocedasticidade, impossibilitando a análise paramétrica (Anova).

Nos casos em que o pesquisador, mesmo avaliando o ambiente, não consiga identificar fatores que promovam heterogeneidade ambiental, qualquer variação existente é considerada variação acidental, sendo computada no erro ou resíduo. Neste

caso, indica-se o uso do DIC, mesmo em condições de campo. Visando aumentar a precisão experimental, recomenda-se aumentar o número dos GL_R, aumentando-se o número de repetições, de modo que o número de parcelas experimentais seja superior a 20 unidades, obtidas pelo produto entre o N° de tratamentos x N° repetições (FERREIRA, 2018).

Recomenda-se também aumentar o número de indivíduos na unidade experimental, reduzindo o erro padrão da média/parcela. Segundo Lúcio et al. (2016), para a cultura da alface, recomendou-se utilizar quatro plantas/parcela para experimentos em estufa e túneis, e cinco plantas para experimentos em nível de campo. Em eucalipto com dois anos, Araújo et al. (2015) recomendaram área útil de seis plantas/parcela, obtendo-se, assim, acurácia de 90%. Em frangos de corte Cobb (sexo masculino, 21 dias de idade, peso médio 932 ± 45 g), Schone et al. (2017) utilizaram quatro aves como unidade experimental. Em suínos Durok (machos, castrados, peso médio $30,35 \pm 1,96$ kg), Palhares et al. (2020) utilizaram dois animais por unidade experimental.

O porte da planta ou animal influenciará diretamente a extensão da unidade experimental, pois, mesmo controlando a heterogeneidade do ambiente adotando o DBC, não há garantia de homogeneidade dentro de blocos. A revisão sistemática de literatura evidenciará o “*n*” mínimo para cada tipo de experimento.

Análise estatística de tratamentos qualitativos e quantitativos

Um tratamento é considerado qualitativo quando se diferencia por suas qualidades. Enquadram-se nesta categoria os testes de competição de cultivares, tipos de adubos, de rações. Já os tratamentos quantitativos podem ser ordenados segundo algum critério numérico, a exemplo dos experimentos que testam níveis de irrigação, adubos, níveis de inclusão de aditivos em rações.

Quando o pesquisador tem o objetivo de avaliar dois tratamentos, o teste F é suficiente para identificar diferenças, não sendo necessário a aplicação de testes adicionais, a exemplo dos de comparação de médias. Porém, quando existem mais de três tratamentos qualitativos, recomenda-se utilizar o teste de comparação/agrupamento de médias.

Métodos de comparação e agrupamento de médias

Os testes de comparação e de agrupamento de médias são aplicados quando o teste F rejeita H_0 , ou seja, indica diferença significativa entre os tratamentos ($P < 0,05$). Existem vários métodos e a escolha do mais adequado deverá considerar o nível de significância e o poder do teste. Entre os mais utilizados em Ciências Agrárias, destacam-se o de Tukey, Duncan e Dunnett.

O teste Tukey (Δ) consiste em um método de comparação múltipla de médias, que é robusto, controlando eficientemente o erro tipo I (Souza et al., 2012), ou seja, quando não houver diferenças estatísticas este teste irá apontar. Porém, permite o aparecimento do erro tipo II (aceitar H_0 quando falsa), principalmente nos casos em que o número de tratamentos e o CV (erro experimental) são elevados ($> 25\%$) e quando a média dos tratamentos apresenta diferentes números de repetições (experimento desbalanceado). Além desse aspecto, a ambiguidade entre os tratamentos poderá promover dificuldades na interpretação dos resultados.

O teste de Duncan (D) é um teste de comparação múltipla de médias que controla, eficientemente, os erros tipo I e II, sendo mais sensível quando os tratamentos apresentam médias próximas, indicado quando são testados mais de quatro tratamentos. Em relação ao teste Tukey, mostra-se mais indicado quando o CV é superior a 15% (Sampaio, 2010).

O teste Dunnett (d') é usado em condições particulares, quando a comparação múltipla entre todos os tratamentos não é interessante, mas se pretende comparar todos os tratamentos com uma referência (testemunha ou controle), quando o teste F for significativo ($P < 0,05$), que poderá ser um produto comercial (substrato, inseticida, ração, cultivar, tipo de queijo). Este método controla bem os erros tipo I e II.

O teste Scott-Knott é um teste de agrupamento de médias que não permite a ambiguidade/sobreposição/dualidade entre os tratamentos, comum aos testes de comparação de médias, ou seja, dentro do grupo há alta similaridade ($P > 0,05$) e entre grupos, máxima variação ($P < 0,05$). Por ser mais robusto que os demais testes, controla eficientemente os erros tipo I e II, sendo indicado também quando há violação da normalidade (BORGES; FERREIRA, 2003). Não há restrições a sua adoção, podendo ser usado em qualquer delineamento e esquemas experimentais, quando o número de tratamentos é pequeno ou grande. O teste Scott-Knott, apresenta ainda maior facilidade na interpretação dos resultados, sendo o mais indicado quando comparado aos demais testes.

Regressão linear na Anova

Na Anova, os tratamentos quantitativos são analisados a partir de uma equação de regressão, que possibilita modelar as relações de causa-efeito existentes entre uma variável-resposta dependente (Y) em função de alterações da variável explicativa independente (X), podendo-se, assim, estimar valores máximos e mínimos, obter funções para análises econômicas e estabelecer previsões dentro dos limites estudados (BHERING; TEODORO, 2021).

Dentro dos modelos lineares, os clássicos são os de regressão polinomial de 1º grau (linear), de 2º grau (quadrática) e de 3º grau (cúbica). A regressão de 1º grau é a mais simples, seguindo o modelo linear $\hat{Y}_i = a \pm bx_i + e_i$, em que \hat{Y}_i corresponde ao valor observado em cada unidade experimental; 'a', corresponde ao intercepto, local onde a curva de regressão toca o eixo Y, indicando o valor de Y quando $X = 0$; 'b' indica o coeficiente angular da reta ('b' positivo, reta inclinada para cima; 'b' negativo, reta inclinada para baixo; 'b' próximo a 0, reta paralela ao eixo X) e indica o aumento ou a redução de Y com a variação de X; x_i , as doses/níveis aplicados aos indivíduos; 'e_i', corresponde ao erro experimental. Nos demais tipos são acrescentados parâmetros, tornando o modelo mais complexo.

A adoção de testes de comparação de médias para tratamentos quantitativos pode levar a conclusões errôneas, pois se o coeficiente de regressão for significativo ($P < 0,05$), todos os efeitos de tratamentos, incluindo os níveis intermediários dentro dos limites estudados, são significativamente diferentes. Alguns autores, como Ferreira (2018) recomendam regressão polinomial com, no mínimo, três níveis de um fator. Porém, o ajuste do modelo poderá explicar, no máximo, o polinômio de 1º grau, prejudicando o uso dos demais modelos. Neste caso, admite-se o uso de testes de comparação/agrupamento de médias, a critério do pesquisador. Recomenda-se que, na fase de planejamento, o pesquisador adote, no mínimo, quatro níveis de um fator quantitativo, para se ter resultados mais abrangentes.

O coeficiente de determinação (R^2) é um parâmetro que indica a qualidade, a precisão, a eficiência do modelo de regressão. O R^2 varia de 0 a 1 e, quanto mais próximo a unidade, melhor a qualidade do ajuste, ou seja, o modelo de regressão consegue explicar a associação existente entre X e Y. O valor de R^2 aumenta com o número de parâmetros do modelo, ou seja, o modelo linear terá R^2 menor que o modelo quadrático que, por sua vez, será menor que o modelo cúbico. Portanto, o critério de seleção do modelo é a significância ($P < 0,05$), pelo teste F, dos coeficientes de regressão e não seu R^2 .

O modelo de regressão polinomial do 3º grau, muitas vezes, não tem explicação biológica, devendo-se ser evitado. Porém, quando a variável independente corresponde a períodos de avaliação, observa-se efeito cúbico sobre os ciclos de produção, explicado pelas variações climáticas, principalmente a precipitação pluvial em regiões tropicais (DAHER et al., 2017).

O uso do software Excel para elaboração das equações de regressão e seus respectivos R^2 considera as médias de cada tratamento. O Excel não testa a significância (P-valor) dos parâmetros de regressão, impossibilitando conhecer se as alterações em X promoverão efeitos significativos sobre Y. Recomenda-se, portanto, que o pesquisador realize a análise em software estatística confiável e quando significativo, substituir os parâmetros da equação fornecida pelo Excel, bem como o R^2 , pelos os da Anova, acrescentando-se os asteriscos ‘***’ e ‘*’, que representam, respectivamente, significativo a 1 e 5%.

Medidas repetidas no tempo

Entende-se por medidas repetidas no tempo às múltiplas avaliações tomadas em uma sequência de tempos em uma mesma unidade experimental (MALHEIROS, 2004). Este tipo de análise é comum em pesquisas com vegetais e na experimentação animal, em que as variáveis (altura da planta, produtividade, peso) são avaliadas na mesma unidade experimental, em intervalos de tempo pré-estabelecidos. No melhoramento genético, permite inferir sobre a interação G x E, fornecendo informações preliminares sobre a adaptabilidade e estabilidade (CRUZ; REGAZZI, 2001).

Dados de medidas repetidas no tempo, tanto podem ser analisados através de um modelo univariado (que impõe uma restrição rigorosa para a matriz de covariâncias), por meio de um modelo multivariado (que adota uma matriz de covariâncias sem restrições), ou ainda, através de um modelo misto, que possibilita a utilização de diferentes estruturas para a matriz de covariâncias (XAVIER; DIAS, 2001).

O modelo univariado é bastante difundido e consiste em considerar um esquema de parcelas subdivididas ou sub-subdivididas, colocando-se os tempos na subparcela ou sub-subparcela, o que nem sempre é correto, pois esse esquema pressupõe que a matriz de covariância (Σ) tenha uma estrutura homogênea, que nem sempre é verdadeiro (MALHEIROS, 2004), considerando que medidas repetidas conduzem a heterocedasticidade dos resíduos, violando um dos pressupostos da Anova.

Uma condição necessária e suficiente para que os testes F sejam exatos é que a matriz Σ satisfaça a condição da simetria composta, também chamada de condição Huynh-Feldt (H-F). Isto implica que as variâncias da diferença entre pares de erros sejam todas iguais (FIELD, 2006). Um problema com relação a validade dos testes surge quando se têm estruturas Σ diferentes das estruturas de simetria composta e erros independentes, levando a testes F não exatos.

O teste de esfericidade de Mauchly (W) verifica se a matriz de covariâncias atende a condição H-F. Caso a suposição da esfericidade seja violada, torna-se necessário que se faça a correção dos graus de liberdade do fator tempo e suas interações (MALHEIROS, 2004). Nos casos em que, mesmo com a correção dos graus de liberdade os dados não apresentem a condição H-F, Field (2006) recomendou o emprego de testes multivariados por não dependerem da assunção de esfericidade e todas as considerações correspondentes sobre o teste F apropriado e correções.

Uma técnica alternativa aos modelos uni e multivariados quando a suposição da matriz de covariâncias não é satisfeita é a análise com modelos mistos, que são uma extensão dos MLG. Este método permite selecionar a matriz de covariâncias (componente de variância, simetria composta, sem estrutura, auto-regressiva, toeplitz e várias outras) que melhor explica o comportamento das observações, tendo a vantagem de ajustar modelos que reduzem o número de parâmetros. A matriz que apresentar o menor valor do critério de informação de Akaike (AIC) ou do critério de Schwarz's Bayesian Criterion (BIC) é selecionada para todas as análises, para cada variável (XAVIER; DIAS, 2001).

Na análise de medidas repetidas de cana de açúcar, utilizando a variável toneladas de colmos/ha, Silva et al. (2015) observaram que a adoção do modelo univariado (parcelas subdivididas) figurou como um dos piores, não sendo recomendado. Por outro lado, a matriz de covariância não estruturada apresentou os melhores ajustes, utilizando-se o critério de AIC. Em capim elefante, para as variáveis massa de forragem, altura da planta, comprimento e largura foliar, diâmetro do colmo e comprimento do entrenó Cavalcante et al. (2013) observaram que a matriz componente de variância foi aquela que melhor se ajustou, segundo o menor valor de AIC.

Para características de crescimento e carcaça de novilhas de corte, Yokoo et al. (2014), utilizando o critério BIC, observaram que as variáveis altura do posterior, índice de musculosidade, profundidade do músculo Gluteus medius se ajustaram a matriz simetria composta; já as variáveis peso vivo, perímetro torácico e espessura da gordura

subcutânea se ajustaram as matrizes fator analítico de 1ª ordem, ante dependente de 1ª ordem, auto-regressiva e heterogênea simetria composta, podendo-se selecionar qualquer uma dessas para as análises estatísticas.

Para a análise de medidas repetidas no tempo, torna-se necessário que o pesquisador realize uma análise preliminar, para garantir que o método estatístico, univariado, multivariado ou modelos mistos está sendo usado adequadamente.

Coefficiente de variação na Anova

Na experimentação em Ciências Agrárias, o coeficiente de variação, que é uma medida de variabilidade dos dados, é utilizado como um indicador de qualidade de uma pesquisa. Ferreira (2018) estabeleceu que, quando o CV for inferior a 10%, considera-se que a pesquisa tem ótima precisão experimental; entre 10 e $\leq 15\%$, boa precisão experimental; entre >15 e $\leq 20\%$, regular precisão experimental; entre > 20 e $\leq 30\%$, péssima precisão; quando superior a 30%, muito péssima precisão experimental, recomendando-se sua repetição.

Oportuno ressaltar que este indicador varia de acordo com a variável analisada, do material genético avaliado e do local de experimentação. É comum elevados CV's em experimentação animal, principalmente os de grande porte em experimentos sob pastejo, quando se avalia sistema radicular ou quando se testa materiais de elevada variabilidade genética. Por outro lado, espera-se baixo CV quando a pesquisa é desenvolvida em condições homogêneas e quando o pesquisador controla a variação externa, adotando-se o DBC.

Visando aumentar a precisão, todos os equipamentos utilizados nas avaliações precisam ser calibrados previamente e a equipe de trabalho treinada quanto aos métodos utilizados. É importante dar condição de igualdade para todos os tratamentos. Portanto, os mesmos equipamentos precisam ser utilizados em todos os tratamentos, pelo mesmo avaliador, seguindo a mesma metodologia. Em experimentos sob o DBC, recomenda-se finalizar as avaliações dos tratamentos de cada bloco no mesmo dia, reiniciando um novo bloco no dia seguinte. Sempre que possível, a repetição do experimento melhora a precisão experimental, pois aumenta os GLR. Estes aspectos reduzem sensivelmente o CV, aumentando a precisão e a confiabilidade dos resultados.

Algumas pesquisas foram desenvolvidas no sentido de definir limites para o CV para cada variável analisada. Mahallem et al. (2008) estabeleceram os limites para frangos de corte, em que, para as variáveis rendimento de carcaça e rendimento de

gordura abdominal, CV superior a 3,5 e 27,6%, respectivamente, são considerados muito alto. Para a produtividade de grãos de milho, CV de 19,1% é considerado muito alto (FRITSCHÉ-NETO et al., 2012).

Em alface, CV superior a 10,9% para a altura do hipocótilo em experimentos de geminação é considerado muito alto (SCHMILDT et al., 2017). Amaral et al. (1997) estabeleceram que o CV acima de 4,3% para o diâmetro do fruto de citros é considerado muito alto e acima de 66,8% muito alto para o número de frutos/planta. Em cultura de tecidos, a variável percentagem de formação de calo, é considerado baixo e muito alto os CV de 25,9 e 96,7%, respectivamente. Portanto, a classificação proposta por Ferreira (2018) deverá ser considerada quando não existam informações específicas para a variável estudada.

Análise de repetibilidade

Em experimentos com espécies perenes ou com animais, dependendo da natureza da pesquisa, muitas vezes, um único ciclo de avaliação poderá não ser suficiente para validar os resultados. Em programas de melhoramento genético vegetal ou animal, ao selecionar um genótipo, espera-se que sua superioridade inicial perdure ao longo de vários períodos de tempo (CAVALCANTE et al., 2012). Em pesquisas com espécies anuais, como os de competição de cultivares, recomenda-se repetir o experimento 'n' vezes para que o desempenho dos genótipos leve em consideração as variações ambientais.

Por definição, o coeficiente de repetibilidade mede a correlação média entre dois fenótipos sucessivos de um mesmo indivíduo, cujas avaliações foram repetidas no tempo (CRUZ; REGAZZI, 2001). A repetibilidade permite determinar quantas avaliações ou experimentos deverão ser conduzidos para que as conclusões tenham um determinado grau de precisão. Com isso, além de resultados precisos e acurados, é possível se obter economia de tempo, dos recursos físicos, financeiros e de mão de obra.

O coeficiente de repetibilidade, que varia de 0 a 1, poderá ser estimado pelos métodos da Anova, componentes principais (via covariância ou correlação), análise estrutural (via covariância ou correlação). Mais recentemente, tem-se utilizado a inferência Bayesiana, tendo-se como vantagens, o uso de dados de experimentos desbalanceados e quando o modelo estatístico é complexo (AZEVEDO et al., 2021).

Para se determinar o nível de precisão ou previsibilidade da repetibilidade é utilizado o coeficiente de determinação (R^2), em que, quando alto, indica que o

acrescimento no número de medições ou no número de experimentos refletirá em pouco acréscimo na precisão (CRUZ; REGAZZI, 2001). A repetibilidade varia com o tipo de variável estudada, com as propriedades genéticas dos genótipos (acessos x linhagens) e com as condições ambientais, ou seja, quanto maior as variações externa e acidental, menor o valor de repetibilidade e, conseqüentemente, maior o número de medições.

Em morangueiro, Andrade Jr. et al. (2020) observaram que seis medições para o peso médio dos frutos e 40 medições para o número de frutos serão necessárias para atingir R^2 de 80%. Em couve, Azevedo et al. (2016a) fizeram 10 medições para o número total de folhas de couve e observaram que duas seriam suficientes para predizer o valor real e selecionar os melhores genótipos, com 99% de confiabilidade, independentemente do método utilizado. A partir da abordagem Bayesiana, Azevedo et al. (2021) observaram que, com oito colheitas de couve, é possível avaliar todas as variáveis com altas estimativas de R^2 ($> 85\%$).

Em feijão, espécie anual, Cargnelutti Filho et al. (2006) observaram que a execução de sete experimentos possibilitará a indicação de cultivares superiores, com 85% de exatidão, independentemente do método utilizado para estimativa do coeficiente de repetibilidade.

As informações fornecidas pela análise de repetibilidade poderão ser úteis aos pesquisadores na ocasião do planejamento experimental para definir o número de avaliações e de experimentos que serão realizados, com reflexos nos cronogramas de execução e de desembolso dos projetos.

CONCLUSÕES

A pesquisa em Ciências Agrárias adota diferentes modelos e métodos estatísticos para validação dos resultados. Neste aspecto, a fase de planejamento experimental é a mais importante, pois definirá todo arcabouço metodológico necessário para obtenção e análise dos dados.

Na escolha do delineamento experimental, deve-se avaliar a heterogeneidade ambiental. Para aplicação da Anova, deve-se checar suas pressuposições. Quando os tratamentos são qualitativos, deve-se empregar testes de comparação/agrupamento de médias, com destaque ao teste Scott-Knott como o mais robusto. Quando quantitativos, utiliza-se regressão polinomial com, no mínimo, quatro níveis do fator.

Em experimentos com medidas repetidas no tempo, torna-se necessário averiguar as condições para adequação metodológica. O coeficiente de variação indica a precisão experimental, que varia com o tipo de tratamento, variáveis e ambiente. A

análise de repetibilidade estima o número de medições e de experimentos que deverão ser executados, com alta precisão.

O uso adequado de modelos e métodos estatísticos torna o processo de produção de conhecimento mais eficiente, reduzindo tempo, recursos físicos, econômicos e humanos, sem prejuízo à precisão e acurácia dos resultados.

REFERÊNCIAS

1. ANDRADE, Valter Carvalho de. et al. 2020. Repeatability and heritability of production characters in strawberry fruits. *Horticultura Brasileira*, v. 38, n. 1, p. 89-93. <http://dx.doi.org/10.1590/S0102-053620200114>.
2. AMARAL, Alexandre Morais do. et al. 1997. Avaliação do coeficiente de variação como medida de precisão na experimentação com citros. *Pesquisa Agropecuária Brasileira*, v. 32, n. 12, p. 1221-1225.
3. ARAÚJO, Márcio José da. et al. 2015. Número de repetições, de plantas por parcela e de avaliações para testes clonais de eucalipto. *Pesquisa Agropecuária Brasileira*, v. 50, n. 10, p. 923-931. <http://dx.doi.org/10.1590/S0100-204X2015001000008>.
4. AZEVEDO, Alcinei Místico. et al. 2021. Determination of the optimal number of evaluations in half-sib progenies of kale by Bayesian approach. *Horticultura Brasileira*, v. 39, n. 1, p. 20-25. <https://doi.org/10.1590/s0102-0536-20210103>.
5. AZEVEDO, Alcinei Místico. et al. 2016^a. Estudo da repetibilidade genética em clones de couve. *Horticultura Brasileira*, v. 34, n. 1, p. 54-58. <https://doi.org/10.1590/S0102-053620160000100008>.
6. AZEVEDO, Alcinei Místico. et al. 2016^b. Transformação Box-Cox na homocedasticidade e normalidade uni e multivariada em experimentos de batata-doce. *Horticultura Brasileira*, v. 34, n. 1, p. 93-101. <https://doi.org/10.1590/S0102-053620160000100014>.
7. BERTOLDO, Juliano Garcia. et al. 2007. Teste de comparação de médias: dificuldades e acertos em artigos científicos. *Revista Brasileira de Agrociências*, v. 13, n. 4, p. 441-447. <https://doi.org/10.18539/cast.v13i4.1409>.
8. BERTOLDO, Juliano Garcia. et al. 2008^a. Problemas relacionados com o uso de testes de comparação de médias em artigos científicos. *Biotemas*, v. 21, n. 2, p. 145-153.

9. BERTOLDO, Juliano Garcia. et al. 2008b. Uso ou abuso em testes de comparações de média: conhecimento científico ou empírico? *Ciência Rural*, v. 38, n. 4, p. 1145-1148. <https://doi.org/10.1590/S0103-84782008000400039>.
10. BHERING, Leonardo Lopes; TEODORO, Paulo Eduardo. 2021. *Estatística experimental no RBio*. Curitiba: Brazil Publishing. 478p.
11. BORGES, Lívia Costa; FERREIRA, Daniel Furtado. 2003. Poder e taxas de erro tipo I dos testes Scott-Knott, Tukey e Student-Newmankeuls sob distribuições normal e não normais dos resíduos. *Revista de Matemática e Estatística*, v. 21, n. 1, p. 67-83.
12. CARGNELUTTI FILHO, Alberto. et al. 2006. Número necessário de experimentos para a comparação de cultivares de feijão. *Ciência Rural*, v. 36, n. 6, p. 1701-1709. <https://doi.org/10.1590/S0103-84782006000600006>.
13. CAVALCANTE, Marcelo. et al. 2012. Coeficiente de repetibilidade e parâmetros genéticos em capim-elefante. *Pesquisa Agropecuária Brasileira*, v. 47, n. 4, p. 569-575. <https://doi.org/10.1590/S0100-204X2012000400013>.
14. CRUZ, Cosme Damião; REGAZZI, Adair José. 2001. *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: UFV. 390p.
15. DAHER, Rogério Figueiredo. et al. 2017. Variação sazonal na produção de forragem de clones intra e interespecíficos de capim elefante. *Revista Agrarian*, v. 10, n. 38, p. 294-303. <http://dx.doi.org/10.30612/agrarian.v10i38.4072>.
16. FERREIRA, Paulo Vanderlei. 2018. *Estatística Experimental Aplicada às Ciências Agrárias*. Viçosa: Editora UFV. 588p.
17. FIELD, Andy. 2006. *Discovering statistics using SPSS*. London: SAGE Publications. 816p.
18. FRITSCHÉ-NETO, Ricardo. et al. 2012. Updating the ranking of the coefficients of variation from maize experiments. *Acta Scientiarum. Agronomy*, v. 34, n. 1, p. 99-101. <https://doi.org/10.1590/S1807-86212012000100014>.
19. LÚCIO, Alessandro Dal'Col. et al. 2011. Método de Papadakis e tamanho de parcela em experimentos com a cultura da alface. *Horticultura Brasileira*, v. 34, n. 1, p. 66-73. <http://dx.doi.org/10.1590/S0102-053620160000100010>.
20. MALHEIROS, Euclides Braga. 2004. Precisão de testes F univariados usados em experimentos com medidas repetidas no tempo, quando a condição de esfericidade da matriz de covariâncias não é verificada. *Revista de Matemática e Estatística*, v. 22, n. 2, p. 23-29.

21. MOHALLEM, Daniel Fernandes. et al. 2008. Avaliação do coeficiente de variação como medida da precisão em experimentos com frangos de corte. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v. 60, n. 2, p. 449-453. <https://doi.org/10.1590/S0102-09352008000200026>.
22. NOGUEIRA, Denismar Alves; PEREIRA, Giselle Machione. 2013. Desempenho de testes para homogeneidade de variâncias em delineamentos inteiramente casualizados. *Sigmae*, v. 2, n. 1, p. 7-22.
23. PALHARES, Liliane Olímpio. et al. 2020. Efeito dos níveis de lisina/proteína ideal sobre desempenho, características de carcaça, digestibilidade e parâmetros sanguíneos de suínos machos, castrados, da raça Duroc, de 30 kg a 50 kg. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v. 72, n. 4, p. 1305-1312. <https://doi.org/10.1590/1678-4162-11240>.
24. POSSATTO JR., Omar. et al. 2019. Survey of statistical methods applied in articles published in Acta Scientiarum. Agronomy from 1998 to 2016. *Acta Scientiarum. Agronomy*, v. 41, e42641. <https://doi.org/10.4025/actasciagron.v41i1.42641>.
25. SAMPAIO, Ivan Barbosa Machado. 2010. *Estatística Aplicada à Experimentação Animal*. 3ª Ed. Belo Horizonte: Fundação de Ensino e Pesquisa em Medicina Veterinária e Zootecnia. 264p.
26. SCHONE, Rodrigo André. et al. 2017. Resíduo seco de destilaria com solúveis (DDGS) na alimentação de frangos de corte (22-42 dias). *Revista Ciência Agronômica*, v. 48, n. 3, p. 548-557. <https://doi.org/10.5935/1806-6690.20170064>.
27. SCHMILDT, Edson Romais. et al. 2017. Coeficiente de variação como medida da precisão em experimentos de alface. *Revista Agro@ambiente*, v. 11, n. 4, p. 290-295. <https://doi.org/10.18227/1982-8470ragro.v11i4.4412>.
28. SILVA, Emerson Noleto. et al. 2015. Seleção da matriz de variância-covariância residual na análise de ensaios varietais com medidas repetidas em cana-de-açúcar. *Ciência Rural*, v. 5, n. 6, p. 993-999. <https://doi.org/10.1590/0103-8478cr20141531>.
29. SOUZA, Clayton Albuquerque de. et al. 2012. Avaliação de testes estatísticos de comparações múltiplas de médias. *Revista Ceres*, v. 59, n. 3, p. 350-354. <https://doi.org/10.1590/S0034-737X2012000300008>.
30. XAVIER, Lara Hoffmann; DIAS, Carlos Tadeu dos Santos. 2001. Acurácia do modelo univariado para análise de medidas repetidas por simulação

multidimensional. *Scientia Agricola*, v. 58, n. 2, p. 241-250.

<https://doi.org/10.1590/S0103-90162001000200005>.

31. YOKOO, Marcos Jun-Iti. et al. 2014. Medidas repetidas no estudo de características de crescimento e carcaça avaliadas por ultrassom em novilhas de corte cruzadas.

Boletim de Indústria Animal, v. 71, n. 3, p. 200-210.

<https://doi.org/10.17523/bia.v71n3p200>.