



UNIVERSIDADE
ESTADUAL DE LONDRINA

GABRIEL DANILO SHIMIZU

**PACOTES ESTATÍSTICOS EM LINGUAGEM R APLICADOS
NA PESQUISA E ENSINO NA ÁREA DE
CIÊNCIAS AGRÁRIAS**

Londrina
2024

GABRIEL DANILO SHIMIZU

**PACOTES ESTATÍSTICOS EM LINGUAGEM R APLICADOS
NA PESQUISA E ENSINO NA ÁREA DE
CIÊNCIAS AGRÁRIAS**

Tese de doutorado apresentado à Universidade Estadual de Londrina - UEL, como requisito parcial para a obtenção do título de doutor em Agronomia.

Orientador: Prof. Dr. Leandro Simões Azeredo Gonçalves

Londrina
2024

SHIMIZU, GABRIEL DANILO.

Pacotes estatísticos em linguagem R aplicados na pesquisa e ensino na área de ciências agrárias / GABRIEL DANILO SHIMIZU. - Londrina, 2024.
120 f.

Orientador: Leandro Simões Azeredo Gonçalves.

Tese (Doutorado em Agronomia) - Universidade Estadual de Londrina, Centro de Ciências Agrárias, Programa de Pós-Graduação em Agronomia, 2024.
Inclui bibliografia.

1. Estatística Experimental - Tese. 2. AgroR - Tese. 3. AgroReg - Tese. 4. seedreg - Tese. I. Gonçalves, Leandro Simões Azeredo. II. Universidade Estadual de Londrina. Centro de Ciências Agrárias. Programa de Pós-Graduação em Agronomia. III. Título.

CDU 63

GABRIEL DANILO SHIMIZU

**PACOTES ESTATÍSTICOS EM LINGUAGEM R APLICADOS NA
PESQUISA E ENSINO NA ÁREA DE CIÊNCIAS AGRÁRIAS**

Tese de doutorado apresentado à Universidade Estadual de Londrina - UEL, como requisito parcial para a obtenção do título de doutor em Agronomia.

BANCA EXAMINADORA

Prof. Orientador Dr. Leandro Simões Azeredo
Gonçalves
Universidade Estadual de Londrina - UEL

Prof. Dr. Guilherme Biz
Universidade Estadual de Londrina - UEL

Profa. Dra. Inês Cristina de Batista Fonseca
Universidade Estadual de Londrina - UEL

Prof. Dr. Lucas Santana da Cunha
Universidade Estadual de Londrina - UEL

Prof. Dr. Luiz Alexandre Peternelli
Universidade Federal de Viçosa - UFV

Londrina, 26 de fevereiro de 2024.

AGRADECIMENTOS

Agradeço a Deus por me guiar, iluminar e me dar tranquilidade para seguir em frente com os meus objetivos e não desanimar com as dificuldades. Agradeço a ele também por manter os meus familiares ao meu lado.

Ao meu irmão Diogo Kazuiti Shimizu e principalmente aos meus pais Cláudio Onzo Shimizu e Regina Shizuko Koga Shimizu, que sempre me motivaram, entenderam as minhas faltas e momentos de afastamento e reclusão e me mostraram o quanto era importante estudar.

A todos os familiares, em especial aos meus tios Pedro Katsumi Koga (*in memoriam*) e Yedo Mitsuo Koga (*in memoriam*).

A Universidade Estadual de Londrina (UEL) pelo acolhimento, infraestrutura e pelo excelente corpo docente, contribuindo de forma significativa para meu desenvolvimento pessoal;

Ao Programa de Pós-Graduação em Agronomia da UEL, em especial ao meu orientador Prof. Dr. Leandro Simões Azeredo Gonçalves pelos ensinamentos, orientações, paciência, conselhos e pela dedicação em ajudar em meu trabalho;

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela aprovação e suporte financeiro;

Aos colegas da Universidade Estadual de Londrina, em especial a Rodrigo Marubayashi; Jean Paula; Walter Junior; Hugo Guariz; Ana Couto; Laura Santos; Bruna Baixo; Lara Guanais; Osvaldo Matsuo (*in memoriam*); e a todos os demais amigos que a pós-graduação me proporcionou.

Aos funcionários e técnicos da UEL, em especial ao Leonardo Pinheiro da ATI da UEL e Diego Luduvério da AINTEC;

Ao departamento de estatística da Universidade Estadual de Londrina pela oportunidade em colaborar como professor temporário, o que possibilitou agregar maior conhecimento na área estatística, em especial aos professores Silvano Cesar da Costa, Vanderli Melen e Guilherme Biz, e aos técnicos Fernando e Adilson.

Aos colegas da Agência Estadual de Defesa Sanitária Animal e Vegetal (IAGRO), em especial à Giuliano Caseiro e Alexandro Santana.

A minha namorada, Vanessa Silva Sousa, pela companhia, amizade, carinho, por me proporcionar diversos momentos felizes e por tudo mais que vivemos e viveremos.

Aos que direta ou indiretamente contribuíram para a realização deste trabalho

Tente uma, duas, três vezes e se possível tente a quarta, a quinta e quantas vezes for necessário. Só não desista nas primeiras tentativas, a persistência é amiga da conquista. Se você quer chegar à onde a maioria não chega, faça aquilo que a maioria não faz.

Bill Gates

SHIMIZU, Gabriel Danilo. **Pacotes estatísticos em linguagem R aplicados na pesquisa e ensino na área de ciências agrárias**. 2024, 120f. Tese de doutorado (Agronomia) – Centro de ciências agrárias, Departamento de Agronomia, Universidade Estadual de Londrina, Londrina, 2024.

RESUMO

A análise estatística é uma etapa primordial na pesquisa, pois é a partir dela que se conclui sobre os objetivos de um estudo. Para isso, normalmente os pesquisadores têm utilizado um *software* estatístico para realizar tais análises e, dentre estes, há de se destacar o *software* R. O R é uma linguagem de programação que apresenta vantagens frente aos demais programas por ser de código fonte livre e gratuito, além de conter uma grande comunidade que disponibiliza códigos em fóruns e tutoriais em livros ou na internet. As análises no R podem ser acessadas ainda em sua base, com pacotes que são nativos, ou também pacotes que são instalados conforme a necessidade do usuário. Contudo, embora a riqueza de materiais disponíveis, além de muitos pacotes estatísticos, ainda carece de pacotes que apresentam uma estrutura de funções simples, com muitas informações e que explore a questão da representação gráfica de resultados. Isso é particularmente verdadeiro, quando os experimentos são voltados para algumas áreas específicas que apresentam certa particularidade, como a tecnologia de sementes. Por outro lado, há também a demanda da academia pela criação de *softwares* ou aplicativos que apresentam a mesma ideia dos pacotes R já citados, mas com uma estrutura mais intuitiva e dessa forma, sem a necessidade de escrever códigos. Partindo dessas ideias, o objetivo do trabalho foi desenvolver os pacotes *AgroR*, *AgroReg* e *seedreg*, além de construir os respectivos aplicativos shiny. O *AgroR* aborda os principais projetos experimentais e oferece diversas funções para análise exploratória de dados, análises inferenciais, representações gráficas e algumas funções auxiliares, bem como um vasto conjunto de dados. O *AgroReg* apresenta diversas funções e modelos de regressão, como por exemplo, polinomial, não paramétrico (loess), segmentado, logístico, exponencial e logarítmico, entre outros. As funções fornecem o coeficiente de determinação (R^2), coeficientes do modelo e os respectivos p-valores do teste t, raiz do erro quadrático médio (RMSE), critério de informação de Akaike (AIC), critério de informação bayesiano (BIC), máximo e mínimo valores preditos e o gráfico de regressão. Além disso, outras medidas de qualidade do modelo e análise gráfica de resíduos também estão incluídas. Por sua vez, o *seedreg* aborda cálculos tradicionalmente usados na área de tecnologia

de sementes e outras abordagens potenciais, bem como testes estatísticos com aplicações na área. O pacote apresenta 18 funções, realizando o cálculo de IVG ou IVE tradicionais ou baseados em regressão logística, tempo médio de germinação ou emergência tradicionais e baseados em T50 (TME ou TMG) e velocidade de emergência ou germinação (v). Todos os pacotes desenvolvidos podem ser instalados através do repositório CRAN, enquanto os aplicativos shiny podem ser acessados no website <https://fisher.uel.br>. Espera-se que os pacotes e aplicativos facilitem a análise estatística de experimentos e atenda a demanda pela comunidade científica.

Palavras-chave: Estatística experimental. AgroR. AgroReg. Seedreg.

SHIMIZU, Gabriel Danilo. **Statistical packages in R language applied in research and teaching in the field of agricultural sciences.** 2024, 120f. Tese de doutorado (Agronomia) – Centro de ciências agrárias, Departamento de Agronomia, Universidade Estadual de Londrina, Londrina, 2024.

ABSTRACT

Statistical analysis is a fundamental step in research, as it is from this analysis that conclusions are drawn regarding the objectives of a study. For this purpose, researchers typically use statistical software to perform such analyses, among which R software stands out. R is a programming language that offers advantages over other programs due to being open source and free, in addition to having a large community that shares codes in forums and tutorials in books or online. Analyses in R can be accessed in its base, with native packages, or with packages that are installed according to the user's needs. However, despite the wealth of materials available and many statistical packages, there is still a lack of packages that present a simple function structure, with lots of information and that explore the issue of graphical representation of results. This is particularly true when experiments are focused on certain specific areas that present certain particularities, such as seed technology. On the other hand, there is also a demand from academia for the creation of software or applications that present the same idea as the mentioned R packages, but with a more intuitive structure and thus, without the need for writing codes. Based on these ideas, the aim of the work was to develop the *AgroR*, *AgroReg*, and *seedreg* packages, as well as to build the respective shiny applications. *AgroR* addresses the main experimental designs and offers various functions for exploratory data analysis, inferential analyses, graphical representations, and some auxiliary functions, as well as a vast dataset. *AgroReg* features various functions and regression models, such as polynomial, non-parametric (loess), segmented, logistic, exponential, and logarithmic, among others. The functions provide the coefficient of determination (R^2), model coefficients and their respective t-test p-values, root mean squared error (RMSE), Akaike information criterion (AIC), Bayesian information criterion (BIC), maximum and minimum predicted values, and the regression graph. Additionally, other model quality measures and graphical analysis of residuals are also included. In turn, *seedreg* addresses calculations traditionally used in the field of seed technology and other potential approaches, as well as statistical tests with applications in the area. The package features 18 functions, performing the calculation of traditional GGI or

GGE based on logistic regression, traditional average germination or emergence time and based on T50 (TME or TMG), and emergence or germination speed (v). All developed packages can be installed through the CRAN repository, while the shiny applications can be accessed on the website <https://fisher.uel.br>. It is hoped that the packages and applications will facilitate the statistical analysis of experiments and meet the demand by the scientific community.

Keywords: Experimental statistics. AgroR. AgroReg. seedreg

LISTA DE FIGURAS

- Figura 1** – Representação dos princípios da experimentação agrícola: repetição (A); Casualização (B) e Controle local (C). **Fonte:** o próprio autor.20
- Figura 2** – Representação de um croqui experimental de um delineamento inteiramente casualizados (A) e blocos casualizados (B). **Fonte:** o próprio autor..... 23
- Figura 3** – Representação de um projeto experimental que foi coletado mais de uma amostra por parcela (A) e representação de um pseudo-experimento que não seguiu os princípios da experimentação agrícola (B). Cor laranja indica a amostra útil da parcela. **Fonte:** o próprio autor..... 28
- Figura 4** – Análise de regressão polinomial de segundo grau com tratamentos espaçados em escala log (Vermelho) e tratamentos espaçados de forma equidistante (Azul). **Fonte:** o próprio autor..... 31
- Figura 5** – Análise de regressão linear simples (A) e regressão logística (B). **Fonte:** o próprio autor. **Fonte:** o próprio autor. 32
- Figura 6** – Representações gráficas para fatores com níveis qualitativos: colunas (A); caixas (B); pontos (C) e barras (D). **Fonte:** o próprio autor..... 41
- Figura 7** – Representação gráfica de colunas compostas de uma interação dupla significativa com fatores qualitativos (A); representação gráfico de interação significativa com um fator qualitativo e um quantitativo (B). IN: inoculado; NI: Não inoculado. **Fonte:** o próprio autor. 42
- Figura 8** – Representação gráfica dos efeitos de um projeto experimental em esquema fatorial 2×2 : Efeito isolado de Fator 2 (A); efeito isolado de Fator 1 (B); efeito de interação simples (C) e efeito de interação complexo (D). **Fonte:** o próprio autor. 43
- Figura 9** – Número de downloads dos principais pacotes usados na estatística experimental desde o primeiro lançamento até 03/01/2024. **Fonte:** o próprio autor. 47
- Figura 10** – Interface gráfica do *software* Sisvar..... 50

Figura 11 – Interface gráfica do <i>software</i> SASM-Agri.....	51
Figura 12 – Interface gráfica do <i>software</i> Assistat 7.7.....	51
Figura 13 – Organograma de criação de um projeto de pacote R usando o IDE Rstudio. Fonte: o próprio autor.	53
Figura 14 – Estrutura básica de um pacote R. Fonte: o próprio autor.	54
Figura 15 – Exemplo de estrutura básica da função de um pacote R. Fonte: o próprio autor.	55
Figura 16 – Instalação e checagem de um pacote R. Fonte: o próprio autor.	56
Figura 17 – Processo de criação de um projeto de aplicativo <i>shiny</i> usando o IDE Rstudio e o pacote <i>shiny</i> . Fonte: o próprio autor.	57
Figura 18 – principais tipos de entrada de controle de um aplicativo shiny. Fonte: https://shiny.posit.co/	58
Figura 19 – Exemplo de croquis experimentais usando o pacote e/ou aplicativo AgroR.....	74
Figura 20 – Diagrama dos projetos experimentais e suas respectivas funções do pacote <i>AgroR</i>	75
Figura 21 – Exemplo de gráficos gerados pelas funções de análise e pertencentes a classe “graphics”: <i>bar_graph</i> (A); <i>sk_graph</i> (B); <i>bar_graph2</i> (C) <i>barfacet</i> (D).	76
Figura 22 – Processo de análise de sumarização usando a função <i>summarise_anova</i> e sua associação com o pacote <i>knitr</i> e <i>rmarkdown</i> para geração de tabelas automáticas.....	77
Figura 23 – Visão geral da estrutura do aplicativo shiny AgroR App e suas principais funções implementadas.	78
Figura 24 – Exemplificação da saída de uma função linear (A) e logística de Brain-Cousens (B) e união das curvas em um gráfico (C) usando as funções do pacote <i>AgroReg</i>	94

Figura 25 – Visualização de todas as observações por gráfico de dispersão (A) e por meio de média e desvio-padrão (B).....	96
Figura 26 – Gráfico de regressão com os dez modelos de regressão usados para exemplificação dos comandos do pacote <i>AgroReg</i>	97
Figura 27 – gráfico <i>QQ-plot</i> do pacote <i>AgroReg</i> de resíduos dos modelos: assintótico negativo (A), biexponencial (B), log-logístico de três parâmetros (C), Brain-Counses de quatro parâmetros (D), Cedergreen-Ritz de cinco parâmetros (E), segmentado linear-linear (F), segmentado linear-platô (G), segmentado quadrático-platô (H), Mistcherlich (I) e Michaelis-Menten de dois parâmetros (J).....	98
Figura 28 – Representação gráfica das comparações entre modelos de regressão pelos critérios de AIC, BIC, R^2 e RMSE.	98
Figura 29 – Modelo de gráfico do pacote <i>seedreg</i> com os modelos: linear, quadrático, cúbico (A); LL.3 e LL.4 (B); BC.4 e BC.5 (C), CRS.4 e CRS.5 (D), normal (E), <i>piecewise_model</i> (F); <i>loess</i> (G) e teste de comparação múltipla (G).....	114

LISTA DE TABELAS

Tabela 1 – Modelo de tabela para dados transformados.....	38
Tabela 2 – Principais funções de saída, tipo de saída e função de renderizar de um aplicativo shiny.	58
Tabela 3 – Funções, descrições, modelo matemático e aplicações dos modelos implementados no pacote <i>AgroReg</i>	89
Tabela 4 – Modelo de regressão, função do pacote <i>AgroReg</i> , coeficiente de determinação (R^2), critério de inferência de Akaike (AIC), critério de informação bayesiana (BIC), raiz quadrado do erro quadrático (RMSE) e coeficientes do modelo do conjunto de dados exemplo “granada”.	99
Tabela 5 – Critério de inferência de Akaike (AIC); critério de inferência bayesiana; coeficiente de determinação (R^2); raiz quadrado do erro quadrático (RMSE); temperatura ótima estimada (TOE); temperatura mínima estimada (TME) dos modelos linear, quadrático, cúbico, logístico de três (LL.3) ou quatro parâmetros (LL.4), modelo de Brain-Cousens de quatro (BC.4) ou cinco parâmetros (BC.5), Cedergreen-Ritz-Streibig de quatro (CRS.4) ou cinco parâmetros (CRS.5), segmentado linear-linear (Piecewise) loess e teste de comparação múltipla (TCM).	112

SUMÁRIO

1 INTRODUÇÃO	16
2 HIPÓTESES	17
3 OBJETIVOS	17
3.1 OBJETIVO GERAIS	17
3.2 OBJETIVOS ESPECÍFICOS	17
4 REVISÃO BIBLIOGRÁFICA	18
4.1 EXPERIMENTAÇÃO AGRÍCOLA	18
4.1.1 Croquis Experimentais e os Princípios da Estatística Experimental	19
4.1.2 Delineamento Experimental	22
4.1.3 Ensaios Fatoriais	24
4.2 ANÁLISE DE VARIÂNCIA	25
4.2.1 Estruturas de Tratamento e Testes Complementares	29
4.2.1.1 Fatores de nível qualitativo	29
4.2.1.2 Fatores de nível quantitativo	30
4.2.1.3 Combinações entre fatores	34
4.2.2 Transformação de Dados	35
4.3 A REPRESENTAÇÃO GRÁFICA DE RESULTADOS	38
4.3.1 Tipo de Variável	40
4.3.2 Número de Fatores	41
4.4 DESAFIOS NA ANÁLISE DE EXPERIMENTOS AGRÍCOLAS	43
4.4.1 Área de Tecnologia de Sementes	43
4.5 <i>SOFTWARES</i> COM APLICAÇÕES EM EXPERIMENTAÇÃO AGRÍCOLA	46
4.5.1 <i>Software R</i>	46
4.5.2 <i>Sisvar</i>	49
4.5.3 <i>SASM-Agri</i>	50
4.5.4 <i>Assistat</i>	51
4.5.5 <i>SAS</i>	52
4.5.6 Outros <i>softwares</i>	52
4.6 CONSTRUÇÃO DE UM PACOTE R	53
4.7 APLICATIVO SHINY	56
5 REFERÊNCIAS BIBLIOGRÁFICAS	59
ARTIGO A	70
AGROR: UM PACOTE R E UMA INTERFACE SHINY PARA ANÁLISE DE EXPERIMENTOS AGRÍCOLAS	70
RESUMO	70

INTRODUÇÃO	71
MATERIAL E MÉTODOS	72
DESENVOLVIMENTO DO PACOTE	72
DESENVOLVIMENTO DO APLICATIVO SHINY	73
RESULTADOS E DISCUSSÃO	73
PACOTE <i>AGRO</i> R	73
FUNÇÕES PARA PLANEJAMENTO EXPERIMENTAL	73
FUNÇÕES PARA ANÁLISE DE EXPERIMENTOS	74
FUNÇÕES GRÁFICAS	76
CONJUNTO DE DADOS	77
OUTRAS FUNÇÕES	77
APLICATIVO SHINY	78
PACOTE <i>AGRO</i> R E APLICATIVO <i>AGRO</i> R SHINY EM NÚMEROS	79
CONCLUSÕES.....	80
REFERÊNCIAS BIBLIOGRÁFICAS	80
ARTIGO B	83
AGROREG: PRINCIPAIS MODELOS DE REGRESSÃO NAS CIÊNCIAS AGRÁRIAS IMPLEMENTADAS NA FORMA DE UM PACOTE R	83
RESUMO.....	83
MATERIAL E MÉTODOS	85
criação do pacote <i>AGRO</i> REG	85
instalação.....	86
conjunto de dados.....	86
“ <i>aristolochia</i> ”	86
“ <i>granada</i> ”	87
modelos de regressão	87
informações e parâmetros estatísticos	91
RESULTADOS E DISCUSSÃO	93
informações gerais	93
exemplo aplicado	95
REFERÊNCIAS BIBLIOGRÁFICAS	100
ARTIGO C	106
<i>SEED</i>REG: UM PACOTE R PARA ÁREA DE TECNOLOGIA DE SEMENTES.	106
RESUMO.....	106
INTRODUÇÃO	106
MATERIAIS E MÉTODOS	108
criação do pacote <i>SEED</i> REG	108
exemplo com conjunto de dados	110
RESULTADOS E DISCUSSÃO	111
CONCLUSÃO.....	115
REFERÊNCIAS BIBLIOGRÁFICAS	115
CONCLUSÕES GERAIS	119

1 INTRODUÇÃO

A estatística é a ciência que permite obter conclusões a partir de dados e está vinculada a todas as áreas da ciência. Dentre áreas mais utilizadas, a estatística experimental se destaca e tem por objetivo o estudo dos experimentos, desde o planejamento, execução, análise dos dados e interpretação dos resultados. A estatística experimental é relativamente antiga e bastante conhecido no âmbito científico, cuja origem é datada em 1926, pelo estatístico, biólogo evolutivo e geneticista inglês Ronald Fisher.

Normalmente, para se analisar dados provenientes de experimentos agrícolas, prevalece a utilização da análise de variância e de testes complementares como o teste de comparação e/ou método de agrupamento de médias ou regressão. Para realizar essas análises, existem diversos *softwares* que podem auxiliar (NUNES *et al.*, 2015), como o Sisvar (FERREIRA, 2000; FERREIRA, 2008; FERREIRA, 2011; FERREIRA, 2014; FERREIRA, 2019), SAS (<https://www.sas.com>), SASM-agri (CANTERI *et al.*, 2001), Assistat (SILVA; AZEVEDO, 2016), statistica (Statsoft., 2004), Genes (CRUZ, 2006a, 2006b, 2008, 2013). No entanto, cada vez mais tem prevalecido a utilização da linguagem R (R CORE TEAM, 2024), uma vez que apresenta código fonte livre, gratuito e aberto, com um número relativamente grande de usuários e desenvolvedores. Dentro da linguagem R, as análises podem ser realizadas ainda com funções nativas ou por meio de implementações na forma de pacote, como o dae (BRIAN, 2021), agricolae (MENDIBURU, 2021), ExpDes (FERREIRA; CAVALCANTI; NOGUEIRA, 2014) e easyanova (ARNHOLD, 2013), sendo esses os principais pacotes desenvolvidos para área das ciências agrárias.

Os experimentos agrícolas, desde que bem planejados, já possuem testes estatísticos apropriados para cada tipo de projeto experimental, embora muitos ainda tenham dúvidas em quais testes utilizar. Isso ocorre, pois, algumas pressuposições devem ser atendidas, sendo: *i*) a normalidade dos erros, *ii*) homogeneidade das variâncias, *iii*) independência dos erros e *iv*) aditividade dos fatores (KNIEF; FORSTMEIER, 2021; PIMENTEL-GOMES, 2009). Assim, muitas vezes o pesquisador ignora tais pressupostos, em virtude da falta de conhecimento ou mesmo por acreditar não ser importante (BAILAR, 1986), o que pode comprometer a confiabilidade de todos os testes paramétricos, podendo levar a falsas conclusões a respeito dos efeitos de tratamento (LÚCIO *et al.*, 2012; MARTIN; STORCK, 2008; MELO *et al.*, 2013; XU, LI; SONG, 2013). Contudo, sabemos que na prática, problemas podem ser recorrentes, como a perda de parcela, experimentos que são economicamente inviáveis de montar com número grande

de repetições, dados inflados de zeros, entre outros fatores. Nesses casos, pode-se usar alternativas aos testes tradicionais, como por exemplo a utilização de modelos lineares generalizados ou, estatísticas não-paramétricas, entre outras.

Até o momento, embora existam *softwares* e pacotes que realizam essas análises, a grande maioria é pouco intuitiva e didática, não auxiliando na tomada de decisão em relação à validade dos testes, sobretudo quando determinada pressuposição não é atendida. Além disso, poucos pacotes exibem representação gráfica dos resultados.

2 HIPÓTESES

Os novos pacotes em linguagem R poderão contribuir com a pesquisa, uma vez que abordam uma vasta gama de modelos estatísticos para experimentos de forma a conduzir ao teste estatístico mais adequado, gerando assim resultados com maior nível de confiabilidade. Dessa forma, promovendo o avanço na pesquisa e conseqüentemente refletindo no progresso na agricultura e da sociedade.

3 OBJETIVOS

3.1 OBJETIVO GERAIS

Criar pacotes e aplicativos mais “amigáveis” em linguagem R, com foco na análise estatística de experimentos agronômicos.

3.2 OBJETIVOS ESPECÍFICOS

- Desenvolver o pacote e aplicativo shiny *AgroR* que possibilita a análise de experimentos em diversos delineamentos experimentais, encaminhando-os testes estatísticos mais adequados, tendo em vista as pressuposições do modelo;
- Desenvolver o pacote *AgroReg*, com ênfase em análise de regressão linear e não-linear, com foco nos principais modelos utilizados nas ciências agrárias;
- Desenvolver o pacote *seedreg*, com ênfase na área de tecnologia de sementes.

4 REVISÃO BIBLIOGRÁFICA

4.1 EXPERIMENTAÇÃO AGRÍCOLA

A experimentação agrícola é o campo da ciência dedicada ao estudo de experimentos, estando no domínio da estatística probabilística e atuando no planejamento, implementação, coleta de dados, análise e interpretação de resultados (BANZATTO; KRONKA, 2013; COCHRAN; COX, 1986; FEERREIRA, 2018; PIMENTEL-GOMES, 2009; STEEL *et al.*, 1996; STORCK *et al.*, 2016; TARTAGLIA *et al.*, 2021). Conhecer a experimentação é importante para todo profissional, pesquisador e/ou usuário dos resultados da pesquisa, pois é a partir da experimentação que há suporte probabilístico, permitindo inferir sobre o comportamento de diferentes fenômenos da natureza com grau de incerteza e margem de erro conhecidos (LÚCIO; SARI, 2017).

Os experimentos agrônômicos são geralmente laboriosos, caros e muitas das vezes podem levar anos para serem concluídos. Além disso, são difíceis e complexos tanto no planejamento quanto na execução e, dependem de muitos fatores que afetam tanto a eficiência quanto a confiabilidade dos resultados, devido à variabilidade natural dos sistemas biológicos e agrícolas (PIEPHO; EDMONDSON, 2018). Por sua vez, os usuários dos resultados da pesquisa devem conhecer a experimentação agrícola e seus princípios, para interpretar seus resultados, permitindo a troca de ideias com os pesquisadores por meio de uma linguagem técnica adequada e universal. Assim, torna-se importante para todos os profissionais envolvidos direta ou indiretamente com a pesquisa agrícola (LÚCIO; SARI, 2017).

A experimentação é um procedimento planejado a partir de uma hipótese para desencadear a ocorrência de fenômenos sob condições controladas com o objetivo de observar e analisar seus resultados e/ou efeitos (BANZATTO; KRONKA, 2006; COCHRAN; COX, 1986; PIMENTEL-GOMES, 2009; STEEL *et al.*, 1997; STORCK *et al.*, 2016). Assim, a introdução de um novo tratamento (e/ou um produto) por exemplo deve ser somente divulgada aos agricultores somente após ser submetida a testes experimentais, quando as respostas às hipóteses levantadas terão embasamento científico e confiabilidade.

As etapas para planejar, instalar e implementar experimentos, bem como a análise estatística dos dados experimentais, deve ser realizada com rigor técnico-científico para minimizar a interferência de efeitos externos não controlados. Desta forma, tais efeitos originarão e contribuirão unicamente para a variância residual ou erro experimental. O erro

experimental é a variação atribuída a efeitos aleatórios. Assim, quanto maior o erro experimental, maior a diferença entre os tratamentos precisa ser para discriminá-los (BANZATTO; KRONKA, 2006; COCHRAN; COX, 1986; PIMENTEL-GOMES, 2009; STEEL *et al.*, 1997; STORCK *et al.*, 2016).

Frequentemente, o erro experimental é inflacionado devido a problemas no planejamento e implementação do experimento, geralmente atrelado ao baixo número de repetições e amostra coletadas. Nesses casos, os experimentos têm baixa precisão e os resultados experimentais, pouca confiabilidade. Portanto, a confiabilidade dos resultados está diretamente associada à precisão experimental. Outro ponto que merece atenção é a estrutura de medição dos fatores de tratamento (PIEPHO; EDMONDSON, 2018), que podem ser qualitativos ou quantitativos, mas isso será abordado nas seções 4.2.1.

4.1.1 Croquis Experimentais e os Princípios da Estatística Experimental

A elaboração do croqui experimental, etapa primordial que antecede a implantação do experimento é vital para maior confiabilidade em relação aos resultados obtidos e, basicamente é nessa etapa que é realizado a definição do número de repetições e da casualização das parcelas (ou aleatorização) (STORCK *et al.*, 2016; ZIMMERMANN, 2004). O princípio da repetição (Figura 1 A) tem como objetivo estimar o erro experimental (LÚCIO; SARI, 2017). Dessa forma, quando as unidades experimentais são repetidas nos tratamentos, a variância do erro pode ser calculada e comparada com os efeitos dos tratamentos avaliados. Sem o princípio da repetição é impossível saber se as diferenças entre os tratamentos são reais ou apenas uma manifestação aleatória das unidades experimentais utilizadas no estudo (STORCK *et al.*, 2016).

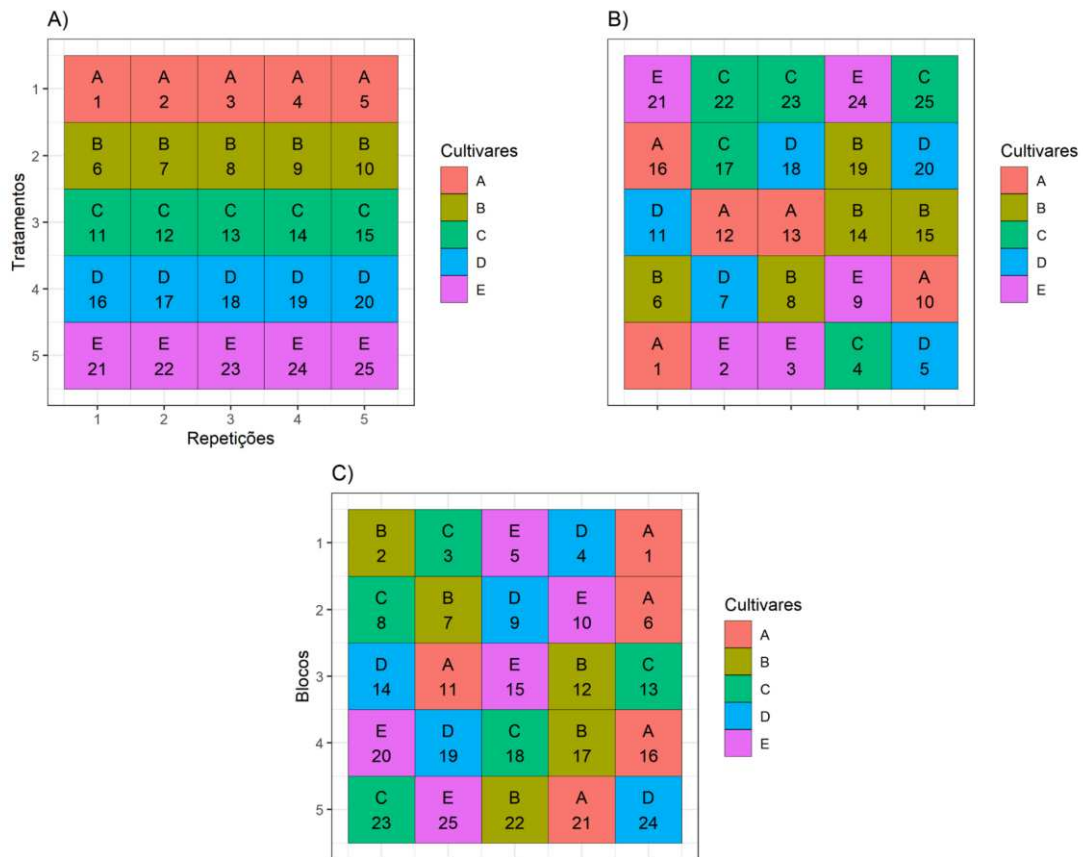


Figura 1 – Representação dos princípios da experimentação agrícola: repetição (A); Casualização (B) e Controle local (C). **Fonte:** o próprio autor.

O uso de repetições aproxima o valor estimado do valor real, reduzindo tanto o erro experimental quanto a variação dos valores das variáveis respostas (LÚCIO; SARI, 2017; TARTAGLIA *et al.*, 2021). Portanto, as principais funções da repetição são estimar os efeitos aleatórios e favorecer a detecção de diferenças significativas entre os tratamentos. Por isso, quanto maior for o número de repetições, mais precisão terão as estimativas de diferença entre os tratamentos nele testados, ou seja, maior será a confiabilidade (TARTAGLIA *et al.*, 2021). Isso significa que se deve usar o máximo de repetições possível. Esse máximo possível, porém, é aquele que permita não só a confiabilidade dos resultados como assegure boa condução do experimento, visto que número de repetições, a boa condução do experimento e confiabilidade dos resultados estão intimamente associados.

O princípio da repetição também é acompanhado de um outro importante componente do experimento, o número de amostras por repetição. Coletar não apenas uma amostra por repetição, mas um conjunto de amostras é essencial para a representatividade da

parcela e aumentar a precisão da inferência na maioria dos experimentos (CARGNELUTTI FILHO *et al.*, 2021). O princípio de coletar várias amostras por repetição é fundamentada em um importante teorema da estatística, o teorema do limite central, que estabelece que qualquer que seja a distribuição de probabilidade natural da variável, à medida que se aumenta o número de amostras das repetições, qualquer que seja a distribuição original da variável, ainda que desconhecida, a média dessas amostras tende a seguir a distribuição normal (ALMEIDA, 2019). Baseado nesse teorema, pode-se exemplificá-lo utilizando à área da tecnologia de sementes; se houvesse um experimento de teste de germinação em que o pesquisador decide usar apenas uma semente por repetição, cada repetição poderia assumir dois possíveis valores de média, 0 ou 100%; se caso adotasse duas sementes, os possíveis cenários seriam 0, 50 ou 100%; ou seja, a medida que aumentamos o números de amostras (sementes) aumentamos o número de possibilidades que nossas repetições podem assumir e esses valores de média passam não mais a se comportarem como sua distribuição natural, no caso binomial, mas como uma distribuição normal.

Coletar amostras por repetição é uma etapa essencial em diversos tipos de experimentos, contudo, estimar o número mínimo de amostras por repetição é objeto de muitos estudos na literatura (BUBANS *et al.*, 2023; FARIA *et al.*, 2020; STORCK *et al.*, 2014). No geral, as metodologias são baseadas na relação entre número de amostras e uma medida de variabilidade, como desvio-padrão ou coeficiente de variação. Essa relação pode ser explicada matematicamente por um modelo de regressão, normalmente um modelo exponencial, segmentado ou logístico, sendo que o número de amostras é definido como aquele em que a variabilidade da parcela tende a se estabilizar (CARGNELUTTI FILHO *et al.*, 2021).

O princípio da casualização (Figura 1 B) consiste em conferir a todos os tratamentos a mesma probabilidade de serem designados a qualquer uma das unidades experimentais (LÚCIO; SARI, 2017; STORCK *et al.*, 2016). Dessa forma, a casualização tem como objetivo validar o erro experimental. O princípio da casualização pode ser considerado como um seguro ao pesquisador, ou seja, uma precaução contra possíveis vieses causados por infinitas variáveis aleatórias, evitando a interferência de fatores externos, como o fator humano, assegurando a imparcialidade na distribuição dos tratamentos. Sua função, por conseguinte, é propiciar que todos os tratamentos tenham a mesma probabilidade de serem designados em qualquer unidade experimental (parcela) e dar suporte à validade dos testes aplicados para comparar os tratamentos. Quando a área experimental não é homogênea, pode-se aplicar o princípio do controle local (Figura 1 C) e dividir em subáreas homogêneas.

O controle local em experimentação agrícola refere-se a técnicas e estratégias utilizadas para minimizar os efeitos de variações indesejadas no ambiente ou nas condições experimentais, que podem influenciar os resultados dos experimentos (FERREIRA, 2011; PIMENTEL-GOMES, 2009). Essas variações podem ser devidas a fatores como heterogeneidade do solo, variações climáticas, diferenças na exposição à luz, entre outros. O objetivo do controle local é garantir que as diferenças observadas nos resultados dos experimentos sejam atribuídas ao tratamento testado, e não a essas variações ambientais ou experimentais indesejadas. O princípio do controle local pode também ser utilizado como fonte de variação na análise, como por exemplo a utilização de blocos homogêneos que dividem a área experimental, que resulta em um projeto do tipo delineamento em blocos casualizados (BARBIN, 2013). Outro exemplo de controle por meio de fonte de variação é o projeto do tipo delineamento em quadrado latino, que usa duas fontes de variação, a linha e a coluna. Esses tipos de projetos, diminuem o efeito residual da análise e conseqüentemente a variação experimental.

O princípio da casualização das parcelas na área experimental pode ser realizado por meio do sorteio manual, nomeando os tratamentos e sorteando de forma aleatória, de acordo com o projeto experimental. Esse processo pode também ser realizado com o auxílio de *softwares*, como por exemplo o Sisvar (FERREIRA, 2011), que fornece uma planilha com as parcelas e os respectivos tratamentos/repetição. Outro exemplo de *software* que permite isso é o *software* R, mais precisamente o pacote agricolae (MENDIBURU, 2021). Este por sua vez, apresenta vantagens em relação ao *software* Sisvar por apresentar diversos projetos experimentais e disposições. Contudo, poucos pacotes ou *softwares* estatísticos realizam a construção de um croqui em termos gráficos, se limitando apenas na geração de uma planilha dos sorteios.

4.1.2 Delineamento Experimental

O delineamento experimental pode ser entendido como a forma que os tratamentos são arrançados dentro da área experimental e é composto de duas estruturas: a “estrutura de tratamentos”, que consiste em um conjunto de tratamentos, combinações de fatores que se deseja estudar ou comparar; e uma “estrutura de delineamento” que consiste no modo de agrupamentos das unidades na área experimental (YASSIN, 2001). Os experimentos agrícolas podem apresentar homogeneidade na área experimental e, dessa forma, as condições

ambientais influenciam de forma igual entre tratamentos e repetições (BANZATTO; KRONKA, 2013). Assim, a estrutura de delineamento é completamente casualizado, sendo chamada de delineamento inteiramente casualizado (Figura 2 A). Normalmente, estes tipos de estruturas são mais comuns com experimentos realizados em ambientes controlados, como em estufas, casa de vegetação e experimentos de laboratórios (YASSIN, 2001). Quando a área experimental apresenta heterogeneidade, mas pode ser dividida em mais do que um grupo de unidades experimentais, de modo que as unidades dentro de cada grupo são mais homogêneas, então podemos separar em grupos homogêneos (LÚCIO; SARI, 2017) e sortear os tratamentos dentro do fator controlado, sendo, portanto, a representação mais simples de controle local, e caracterizando como um projeto de delineamento em blocos casualizados (Figura 2 B) (PIMENTEL-GOMES, 2009). Tais projetos são mais comuns em experimentos de campo (YASSIN, 2001).

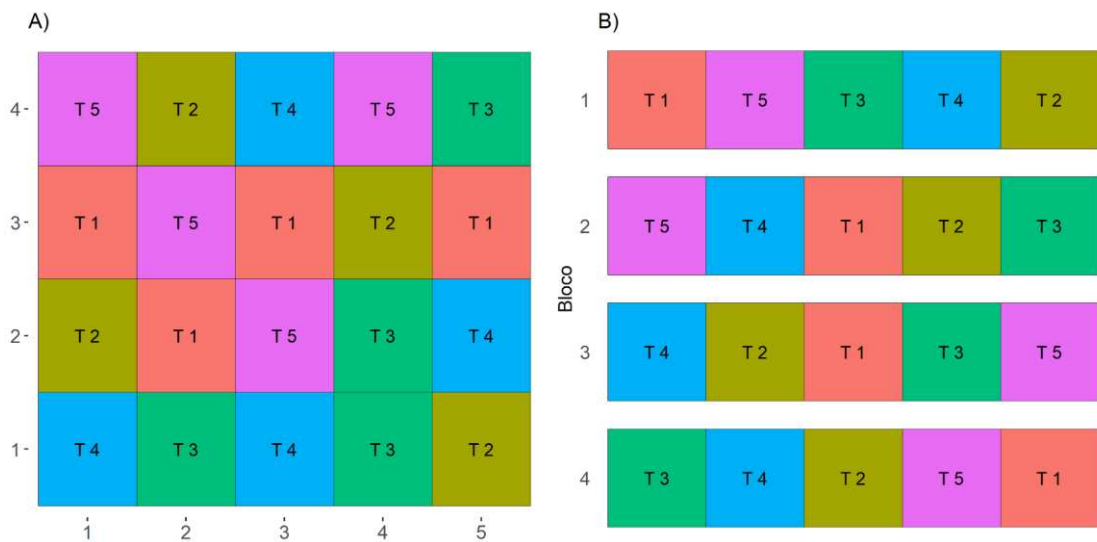


Figura 2 – Representação de um croqui experimental de um delineamento inteiramente casualizados (A) e blocos casualizados (B). **Fonte:** o próprio autor.

Nos experimentos mais simples comparam-se tratamentos de apenas um tipo ou fator, sendo que os demais fatores permanecem constantes e controlados. Desse modo, por exemplo, quando se deseja estudar adubações, todos os demais fatores, como variedades e tratamentos culturais, por exemplo, devem ser os mesmos para todas as adubações a serem experimentadas. Conforme Banzatto e Kronka (2013), existem casos em que vários fatores devem ser estudados simultaneamente para que se possa chegar a resultados mais abrangentes e mais conclusivos.

Nesses casos, segundo os autores, é comum utilizar os experimentos fatoriais, que são aqueles nos quais são estudados, ao mesmo tempo, dois ou mais fatores ou tipos diferentes de tratamentos. Cada subdivisão de um fator é denominada de nível do fator e os tratamentos nos experimentos fatoriais completos consistem em todas as combinações possíveis entre os diferentes níveis dos diversos fatores.

4.1.3 Ensaios Fatoriais

Os experimentos que apresentam sua estrutura de tratamentos composta pela combinação de dois ou mais fatores, chamadas de esquema fatorial, tiveram seu início com Yates (1937) e são bastante discutidos em diversos trabalhos ao longo dos anos, como Cochran e Cox (1957), Hinkelmann e Kempthorne (1994), Kirk (1995), Pimentel-Gomes (2009), Steel e Torrie (1996). Esses projetos são utilizados em experimentos agrícolas visando estudar os efeitos simples dos fatores e as suas interações. O efeito principal é o efeito isolado de cada fator, não levando em conta os níveis dos demais fatores, isto é, estuda isoladamente o efeito de cada fator em separado, enquanto o efeito de interação estuda o desempenho de cada fator, levando em consideração os níveis dos demais fatores e, ela sendo significativa, deve-se ignorar os efeitos isolado.

Os experimentos fatoriais não constituem um delineamento experimental, mas um esquema de desdobramentos dos graus de liberdade e soma de quadrados de tratamentos em um determinado delineamento, como inteiramente casualizado, blocos casualizados ou quadrado latino. Como mencionado por Yassin (2001), os experimentos em esquema fatorial embora têm sido utilizados no estudo de dados originados de pesquisas agrícolas, também são úteis nas áreas industriais e biológicas, em situações como por exemplo: i) experimentos exploratórios para determinar os efeitos de cada um de um número relativamente elevado de fatores sobre características de interesse dentro de um sistema industrial, cada um com dois níveis, chamados de fatoriais 2^k , para seleção de fatores mais importantes; ii) Experimentos com propósito de recomendação para ampla faixa de condições. Em experimentos com um único fator principal pode ser conveniente incluir fatores adicionais de forma a ampliar a base para as inferências que serão derivadas; e iii) Estudo das interações entre os efeitos de vários fatores, uma vez que a resposta de um fator pode alterar dependendo dos níveis dos outros fatores do experimento.

4.2 ANÁLISE DE VARIÂNCIA

A análise de variância foi introduzida por Ronald A. Fisher e é essencialmente um processo aritmético para particionar uma soma total de quadrados em componentes associados a fontes conhecidas de variação. Tem sido usado para aproveitar em todos os campos de pesquisa em que os dados são medidos quantitativamente (STELL; TORRIE, 1996). De acordo com Quinn e Keough (2002), os dois principais objetivos da análise de variância clássica são:

- Examinar a contribuição relativa de diferentes fontes de variação (fatores ou combinação de fatores, ou seja, as variáveis preditoras) à quantidade total da variabilidade na variável resposta, e;
- Para testar a hipótese nula (H_0) que o grupo populacional ou os meios de tratamento são iguais.

De acordo com Casella e Berger (2002), dado um certo número de tratamentos (digamos que i tratamentos) e denotando seus efeitos dos tratamentos por $\tau_1, \tau_2, \dots, \tau_i$, a hipótese é:

$$H_0: \tau_1 = \tau_2 = \dots = \tau_i = 0$$

$$H_1: \tau_i \neq 0, \text{ pelo menos um tratamento}$$

No entanto, para que o teste F da análise de variância seja válido, são necessárias quatro premissas básicas não sejam rejeitadas (PIMENTEL-GOMES, 2009):

- 1) Os erros são independentes;
- 2) Os erros são normalmente distribuídos;
- 3) Erros apresentam homogeneidade de variância;
- 4) Aditividade dos termos do modelo.

Quando as pressuposições são violadas, a confiabilidade de todos os testes paramétricos fica comprometida, pois ocorrem alterações na probabilidade de ocorrência do erro tipo I e tipo II, podendo levar a falsas conclusões a respeito dos efeitos de tratamento (LÚCIO *et al.*, 2012; MARTIN; STORCK, 2008, XU; LI; SONG, 2013).

As pressuposições normalmente são testadas por meio de testes de hipóteses, como Shapiro-Wilk (SHAPIRO; WILK, 1965), Lilliefors (LILLIERFOS; HUBERT, 1967), Anderson-Darling (ANDERSON; DARLING, 1952), Cramer-von Mises (CRAMER, 1928), Shapiro-Francia (SHAPIRO; FRANCA, 1972) para normalidade dos erros; Bartlett (BARTLETT, 1937), Hartley (HARTLEY, 1950) ou Levene (LEVENE, 1960) para homogeneidade das variâncias e Durbin-Watson (DURBIN; WATSON, 1950) no caso de independência dos erros. Para aditividade dos termos, pode-se utilizar o teste de não-aditividade de Tukey, embora poucos trabalhos cite a utilização dele. Em todos os casos, o valor de p deve ser superior ao nível de significância estabelecido pelo pesquisador (geralmente 1 ou 5%). Por outro lado, a visualização gráfica das pressuposições, seja por gráfico de resíduos padronizados, seja para um gráfico do tipo *halfnormalplot* tem sido bastante utilizado para o diagnóstico dos resíduos e geralmente é associado essas duas abordagens.

Basicamente, a ANOVA verifica, por meio do teste F, se a variância entre os efeitos do tratamento se deve ou não ao acaso (BANZATTO; KRONKA, 2013; LÚCIO; SARI, 2017). O valor de F calculado corresponde à razão entre o quadrado médio dos tratamentos (QM_{trat}), que representa a variância entre os efeitos dos tratamentos, e o quadrado médio do erro (QM_{res}), que representa a variação aleatória. Entende-se assim que, se a variação entre os tratamentos for devido ao acaso, o valor de F será igual a 1. Valores altos de F calculados são uma indicação de que a variação entre os tratamentos é maior do que as variações devido ao acaso, portanto, ao menos dois tratamentos estão diferindo entre si. O valor de p também pode ser usado como critério de decisão. Quando o valor de p é inferior ao nível de significância, a hipótese H_0 é rejeitada e conclui-se que os tratamentos diferem entre si (PIMENTEL-GOMES, 2009).

O valor de p , ou valor de probabilidade, representa a probabilidade de observar um resultado tão extremo quanto o obtido no teste, assumindo que a hipótese nula é verdadeira (SCHERVISH, 1996). A hipótese nula, neste contexto, geralmente afirma que todas as médias dos grupos são iguais, ou seja, não há diferença significativa entre os grupos. Se o valor de p é menor que um limiar pré-estabelecido (geralmente 0,05 ou 5%), então conclui-se que há evidências estatísticas suficientes para rejeitar a hipótese nula. Isso significa que é improvável que as diferenças observadas entre as médias dos grupos tenham ocorrido por acaso, e, portanto, as diferenças são consideradas estatisticamente significativas. Por outro lado, se o valor de p é maior que o limiar, não se pode rejeitar a hipótese nula, o que sugere que as diferenças observadas entre as médias dos grupos podem ser atribuídas ao acaso, e não são

consideradas estatisticamente significativas (FERREIRA, 2018; LÚCIO; SARI, 2017; PIMENTEL-GOMES, 2009).

Quando se realiza a análise de variância, duas questões devem ser respondidas: se o desenho experimental foi eficiente e se há diferença significativa entre os efeitos dos tratamentos (LÚCIO; SARI, 2017). Contudo, falhas no planejamento e na implementação do experimento podem inflacionar o erro experimental e, conseqüentemente, elevam o valor do quadrado médio do resíduo, diminuindo o valor calculado de F e tornando a diferenciação entre os tratamentos mais difícil, aumentando a probabilidade de não rejeitar H_0 , mesmo quando ela é falsa.

Um outro ponto em particular que, embora simples, parece se tornar frequente na atualidade é questão da confusão entre uma repetição genuína de uma pseudo-repetição. Pode se dizer que, uma repetição genuína é aquela que absorve todo o erro experimental, enquanto as pseudo-repetições absorvem apenas parte da variabilidade total e, dessa forma, pode estar subestimando o valor do quadrado médio do resíduo (HURLBERT, 1984). Existem basicamente duas situações em que ocorre frequentemente esse tipo problema, quando o pesquisador considera uma amostra como sendo uma repetição e situações em que o planejamento experimental é realizado sem seguir os princípios da experimentação agrícola. Por exemplo, supondo que um pesquisador tenha realizado um experimento com cinco tratamentos e cinco repetições e, em cada repetição foi medido a altura de nove plantas úteis (Figura 3 A). Nesse caso, o pesquisador deveria calcular a média das nove plantas e representar como sendo o respectivo tratamento/repetição, totalizando assim 25 parcelas. Essa média pode ser substituída por outras estatísticas como a mediana, bastante usado na área de fitossanidade, ou a moda. Contudo, não é incomum ver situações em que ao invés de calcular a média das nove plantas, o pesquisador considera como sendo 45 repetições por tratamento, ou seja, como sendo 225 parcelas no total. Nesse caso, essas pseudo-repetições são medidas redundantes, o que acaba reduzindo o quadrado médio do resíduo e aumentando a chance de acusar diferença, quando na verdade H_0 é verdadeiro.

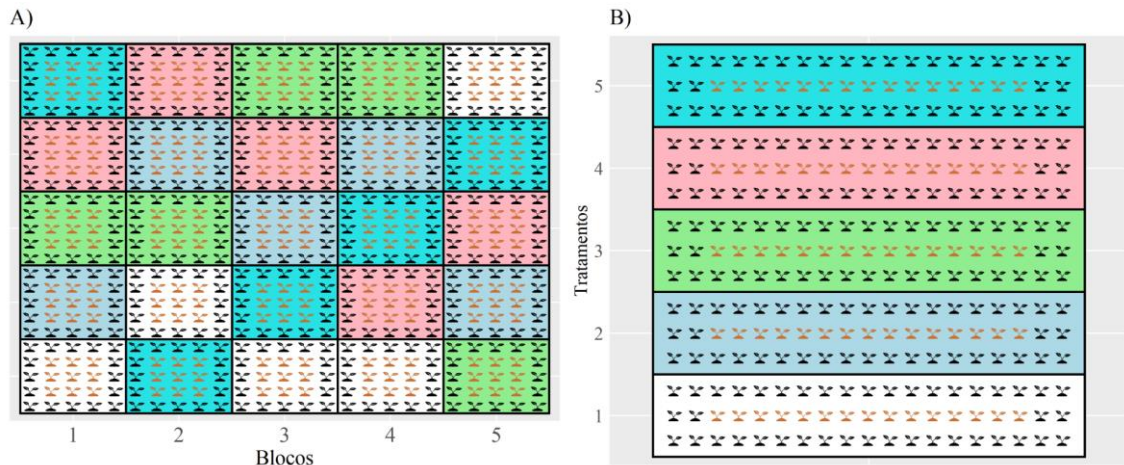


Figura 3 – Representação de um projeto experimental que foi coletado mais de uma amostra por parcela (A) e representação de um pseudo-experimento que não seguiu os princípios da experimentação agrícola (B). Cor laranja indica a amostra útil da parcela. **Fonte:** o próprio autor.

Na Figura 3 B há uma representação de pseudo-experimento ou de um experimento sistêmico, conforme relatado por Ferreira (2018) que, nesse caso, não houve casualização das repetições na área experimental. Também é possível afirmar que não houve repetições genuínas nesse experimento, uma vez que as repetições são dependentes espacialmente. Dessa forma, o pesquisador tem apenas uma repetição para cada tratamento com 15 amostras úteis cada. Em contrapartida, vários casos desse tipo de situação ocorrem na prática e, o pesquisador acaba por considerar essas amostras como sendo a representação da repetição. Uma das principais consequências é a possibilidade de o efeito do tratamento estar sendo mascarado por qualquer outro efeito não mensurável. Por exemplo, supondo que a situação da Figura 3 B foi realizada com o intuito de avaliar o rendimento de uma dada cultura em função da quantidade de lâmina de água da irrigação. Assim, o pesquisador julgando estar certo no planejamento do experimento, instala uma irrigação automática para cada tratamento, sendo que irá avaliar as 15 plantas centrais e considerar como sendo a repetição. Contudo, por acidente, uma das mangueiras do tratamento 3 por exemplo acaba entupindo, aplicando menos do que a quantidade definida para o tratamento, prejudicando o desenvolvimento da cultura. Dessa forma, todas as quinze unidades amostradas são afetadas pelo mesmo problema, devido a dependência e nesse caso, se o pesquisador não estiver atento, pode estar julgando o tratamento como inferior aos demais, quando na verdade há um outro efeito não controlado mascarando o resultado.

4.2.1 Estruturas de Tratamento e Testes Complementares

A análise dos experimentos agrícolas depende criticamente de suposições sobre as estruturas de medição dos fatores de tratamento, e a compreensão dessas estruturas é essencial para uma análise correta dos dados experimentais (PIEPHO; EDMONDSON, 2018). Neste tópico, serão abordadas as principais categorias de medição que ocorrem frequentemente em experimentos agrícolas e o direcionamento em relação a quais análises devem ser realizadas.

4.2.1.1 Fatores de nível qualitativo

Os fatores de tratamento são ditos fatores qualitativos (nominais ou categóricos) quando os níveis não são ordenáveis em uma escala numérica (PIEPHO; EDMONDSON, 2018). Por exemplo, as cultivares de uma espécie em um estudo de competição de cultivares é um fator de nível qualitativo porque os níveis de fatores não são ordenáveis em uma escala numérica. Não há relação entre os níveis de fatores individuais, portanto não há ordenação especial dos níveis e nenhuma relação quantitativa entre os níveis de fator. Nesse exemplo, supondo que o experimento seja adequadamente planejado, as variâncias das diferenças médias estimadas podem ser calculadas e os níveis de fator comparados descritivamente por um erro padrão de uma diferença (SED) ou analiticamente por um teste de múltiplas comparações (BRETZ; HOTHORN; WESTFALL, 2011; HSU, 1996).

Quando a hipótese nula do efeito do tratamento for rejeitada, o pesquisador ainda se depara com o problema de decidir qual das médias não é igual. Um teste F significativo indica que algo aconteceu em um experimento que tem uma pequena probabilidade de acontecer por acaso (KIRK, 1995). Assim, de acordo com Bretz, Hothorn e Westfall (2011), Borges e Ferreira (2003), Kirk (1995), Stell e Torrie (1996), Quinn e Keough (2002), quando os tratamentos são qualitativos e há mais de dois níveis de um fator qualitativo, várias propostas de contrastes entre as médias estão disponíveis na literatura, como o teste de Tukey (TUKEY, 1949), Duncan (DUNCAN, 1955), LSD (FISHER, 1935), Student-Newman-Keuls (NEWMAN, 1939; KEULS, 1952), Dunn (DUNN, 1961), Dunnett (DUNNETT, 1964); Scheffé (SCHEFFÉ, 1959), entre outros.

De forma geral, dois testes têm sido mais utilizados em detrimento aos demais, em virtude do maior rigor na separação entre os níveis dos fatores, são eles os testes de Tukey e o LSD. Por outro lado, quando há um número relativamente grande de níveis dos

tratamentos, ambos os testes podem gerar pouca discriminação, devido a ambiguidade, o que pode tornar em muitos casos, inconclusivo a pesquisa (BORGES; FERREIRA, 2003). Contudo, para esses casos, pode-se utilizar uma alternativa que é o método de agrupamento Scott-Knott (1974). Este teste visa a separação de médias de tratamentos em grupos distintos, através da minimização da variação dentro e maximização da variação entre grupos. Os resultados são facilmente interpretados, devido à ausência de ambiguidade. Desta forma este procedimento resulta em maior objetividade e clareza. Contudo, Borges e Ferreira (2003) concluíram em seu estudo que o método de agrupamento de Scott-Knott apresentou maior poder de acerto que outros testes como SNK e Tukey, além de ser menos sensível a ausência de normalidade dos erros, exibindo resultados muito semelhantes.

Quando há apenas dois níveis do fator qualitativo, embora possa ser realizado, não há a necessidade de realizar a análise por meio da análise de variância, podendo o mesmo ser realizado por um teste t para amostras independentes com variância desconhecida ou um teste t pareado para variâncias desconhecidas no caso de um experimento em delineamento em blocos casualizados.

4.2.1.2 Fatores de nível quantitativo

Os fatores de tratamento são fatores de nível quantitativos (numéricos ou métricos) se tiverem uma ordem dos níveis de fator com diferenças quantitativas mensuráveis entre os níveis (PIEPHO; EDMONDSON, 2018). Por exemplo, um fator com níveis representados por diferentes quantidades ou proporções ou taxas do mesmo fertilizante, ou mesmo a quantidade/dose de um mesmo produto seria um fator de nível quantitativo.

Nesses casos, os modelos de regressão utilizam as informações de distância na escala das variáveis preditoras quantitativas e são essenciais para a captura de informações quantitativas sobre os níveis de um fator de nível quantitativo. Em comparação, uma simples análise fatorial qualitativa ignoraria qualquer informação sobre a relação distância e ordem entre os níveis de fator (PIEPHO; EDMONDSON, 2018).

Um modelo comum para um fator de nível quantitativo com não muitos níveis (geralmente não mais do que 7 ou 8) é um modelo de regressão polinomial de baixa ordem instalado nos níveis de tratamento quantitativos separados de forma equidistante (PIEPHO; EDMONDSON, 2018). Esse ponto é particularmente importante, pois tratamentos quantitativos com níveis distanciados de forma não equidistantes não podem ser analisados por

método de polinômios ortogonais normalmente usados no desdobramento dos polinômios na análise de variância, além disso, podem gerar ajustes de modelo de regressão que resulte em predições de valores superestimados ou subestimados (Figura 4). Nesse exemplo, quando o experimento foi conduzido com tratamentos em escala não equidistante, nesse caso em escala log e embora apresente um bom ajuste ($R^2 = 0,94$), está resultando em uma equação completamente diferente se tivesse adotado a utilização de tratamentos separados de forma equidistante. A dose de máxima resposta nos tratamentos em escala log está resultando em uma resposta de 174, enquanto para a outra, seria 120.

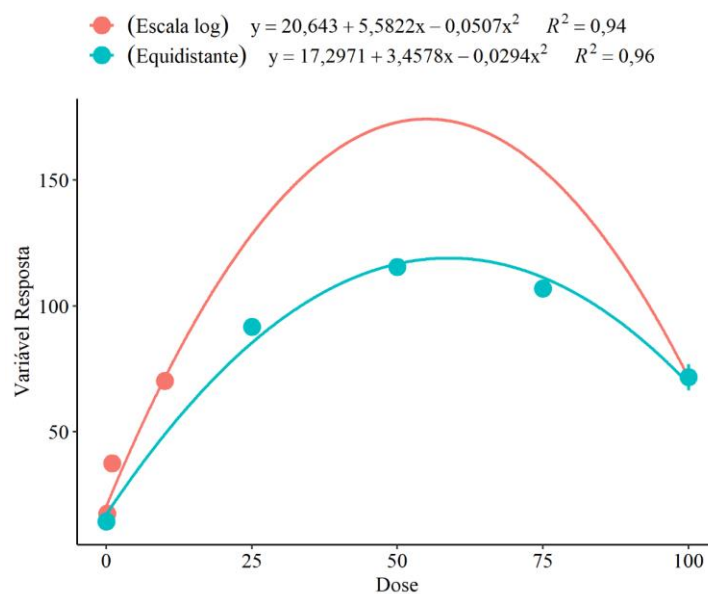


Figura 4 – Análise de regressão polinomial de segundo grau com tratamentos espaçados em escala log (Vermelho) e tratamentos espaçados de forma equidistante (Azul). **Fonte:** o próprio autor.

O modelo de regressão ajustado deve explicar os dados observados da forma mais parcimoniosa possível e onde uma análise de regressão polinomial é utilizada normalmente deve incluir apenas termos polinomiais até o termo de grau mais alto que é estatisticamente significativo no modelo construído (PIEPHO; EDMONDSON, 2018). Nesse sentido, um polinomial significativo requer pelo menos três níveis de fator quantitativo, embora seja recomendado ao menos cinco. Portanto, um fator de tratamento com apenas dois níveis é normalmente considerado qualitativo, independentemente dos níveis reais de tratamento (FERREIRA, 2018; PIEPHO; EDMONDSON, 2018).

Um outro ponto a se destacar na análise de fatores quantitativos é a explicação biológica sobre o fenômeno. Por exemplo, não é incomum ver trabalhos em que foi ajustado um modelo de regressão linear simples para explicar um dado fenômeno, como uma curva de crescimento de uma planta, ou mesmo o acúmulo de uma determinada substância no fruto. Nesses casos, é simplesmente lógico que o crescimento e acúmulo só podem assumir respostas positivas, ou seja, é impossível ter valores estimados abaixo de zero. Contudo, o que acontece é que muitos modelos de regressão podem apresentar valores preditos negativos, o que biologicamente não existe explicação. Assim, repare por exemplo na Figura 5, para um mesmo conjunto de dados é apresentado dois modelos de regressão diferentes. Na primeira (Figura 5 A), o modelo de regressão para T2 está estimando valores negativos até próximo de dois dias, enquanto para o modelo logístico (Figura 5 B) isso não ocorre. Resumindo, a escolha de um modelo de regressão não se deve apenas pelos parâmetros estatísticos, como R^2 , RMSE, AIC, BIC, entre outros, mas também deve ser levado em consideração se o dado fenômeno é biologicamente explicável usando o ajuste em questão.

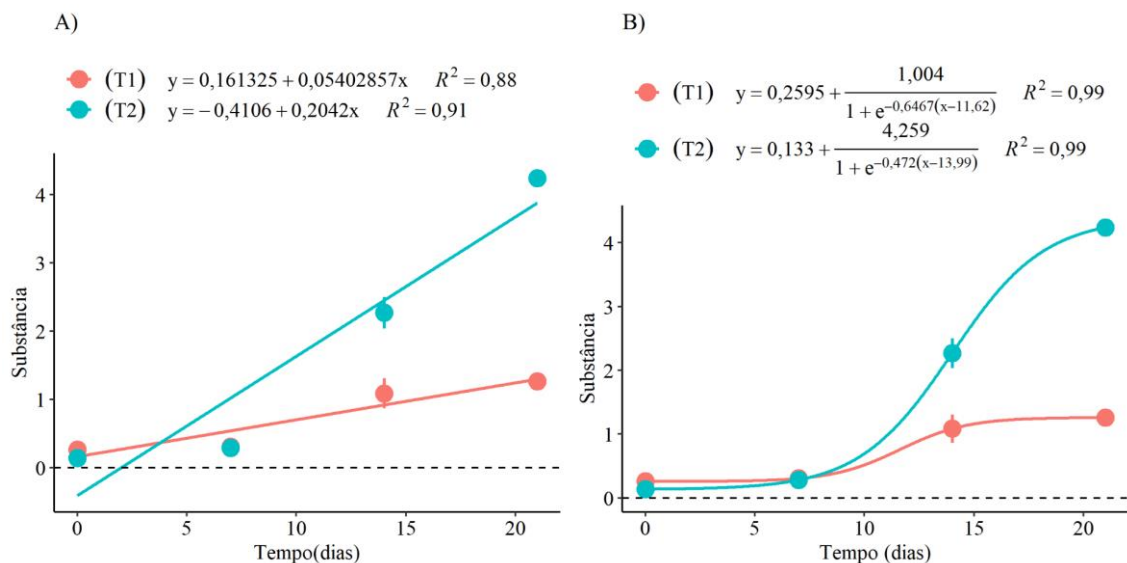


Figura 5 – Análise de regressão linear simples (A) e regressão logística (B). Fonte: o próprio autor. **Fonte:** o próprio autor.

Como relatado acima, algumas medidas estatísticas podem nortear sobre a escolha de um modelo em detrimento aos demais. Por exemplo, o R^2 , conhecido como coeficiente de determinação é uma medida estatística que avalia a proporção da variabilidade em uma variável dependente que pode ser explicada pelas variáveis independentes em um modelo estatístico. Em outras palavras, indica o quão bem os valores observados de uma

variável podem ser preditos pelo modelo estatístico (CHICCO; WARRENS; JURMAN, 2021). É uma medida que varia de 0 a 1, sendo que valores mais próximos de 0 indicam que o modelo não explica nenhuma variação na variável dependente, enquanto um de 1 significa que o modelo explica toda a variação na variável dependente.

Outra importante medida é a raiz quadrada do erro quadrático, conhecida como RMSE, dado pela fórmula:

$$MSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}$$

em que: \hat{Y}_i é resposta predita pelo modelo, Y_i é a resposta observada, n é o tamanho amostral.

O RMSE é uma métrica estatística usada para avaliar a precisão de um modelo de regressão. Ele mede a média das diferenças entre os valores observados e os valores preditos ao quadrado, e então extrai a raiz quadrada desse resultado. Fornece uma medida da dispersão dos erros de predição. Quanto menor for o valor do RMSE, melhor será a precisão do modelo, indicando que as predições estão mais próximas dos valores reais.

Outras duas importantes medidas são o AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion), que são critérios de informação utilizados na seleção de modelos estatísticos (ARCHONTOULIS; MIGUEZ, 2015). Ambos os critérios visam equilibrar a capacidade de ajuste do modelo com a penalização por complexidade, evitando assim o ajuste excessivo ou a escolha de modelos muito simples (BURNHAM; ANDERSON, 2004).

$$\begin{aligned} AIC_i &= -2\log l + 2p \\ BIC_i &= -2\log l + p\log n \end{aligned}$$

em que: l e p são a função de verossimilhança e o número de parâmetros para cada modelo, e n é o número de observações.

Em ambos os critérios, o termo “log-likelihood” representa a função de verossimilhança logarítmica do modelo, que mede quão bem o modelo descreve os dados observados. A ideia subjacente é escolher o modelo que fornece um bom ajuste aos dados, mas penaliza modelos mais complexos para evitar sobreajuste. Ao comparar modelos, é comum

selecionar aquele que tem o valor mais baixo de AIC ou BIC (ARCHONTOULIS; MIGUEZ, 2015).

4.2.1.3 Combinações entre fatores

Em experimentos fatoriais ou no caso do esquema de parcelas subdivididas, a análise complementar depende dos resultados da análise de variância. Para interações significativas, os níveis dos fatores devem ser analisados estatisticamente um dentro do outro (LÚCIO; SARI, 2017). Por exemplo, em um projeto fatorial completo com dois fatores (A e B) e se uma análise de variância não mostra evidência de interação entre os fatores, então a interpretação dos efeitos do tratamento pode ser realizada apenas para os efeitos principais separados para A e B sem perda de informações de tratamento. Se, por outro lado, houver uma interação significativa entre os fatores A e B, então a tabela fatorial para $A \times B$ contém informações de tratamento fatorial não aditivo e não pode ser marginalizada sem uma perda significativa de informações de tratamento (PIEPHO; EDMONDSON, 2018).

Para ilustrar os efeitos testados em um esquema fatorial, vamos supor um projeto experimental com dois fatores (A e B), conforme exemplificado por Banzatto e Kronka (2013) e Storck *et al.* (2016). Nesse caso, há várias possibilidades de resultados que o pesquisador pode encontrar para realizar testes complementares, conforme a seguir:

Se os fatores A e B forem qualitativos:

- a) Interação não significativa e fatores com mais de dois níveis: Se há efeito dos fatores isolados, o procedimento complementar seria um teste de médias para cada fator separadamente;
- b) Interação não significativa e fator com dois níveis: se há efeito dos fatores isolados e sendo que o fator apresenta apenas dois níveis, não são necessários testes complementares. O próprio F da ANOVA já será informativo;
- c) Ausência de efeitos significativos: Não são necessário testes complementares;
- d) Interação significativa: O procedimento complementar seria um teste de médias para cada fator discriminado dentro dos níveis do outro fator e vice-versa.

Se um dos fatores for quantitativo e com mais de três níveis, por exemplo A qualitativo e B quantitativo:

- a) Interação não significativa: Se há efeito isolado de A e sendo o número de níveis maior que dois, realizar teste de médias; se há efeito isolado de B, realizar análise de regressão;
- b) Interação significativa: teste complementar seria um teste de média para o fator A dentro de cada nível de B e análise de regressão para o fator B para cada nível de A. Nesse caso, as análises de regressão devem ser realizadas para aqueles níveis que apresentaram diferença significativa quando desdobrado efeito de B dentro de cada nível de A;
- c) Ausência de efeitos significativos: Não são necessários testes complementares;

Se ambos os fatores forem quantitativos e com mais de três níveis:

- a) Interação não significativa: Se há efeito dos fatores isolados, o procedimento complementar seria uma análise de regressão para cada fator separadamente;
- b) Interação significativa: o teste complementar seria uma análise de regressão múltipla (PIEPHO; EDMONDSON, 2018);
- c) Ausência de efeitos significativos: Não são necessários testes complementares;

4.2.2 Transformação de Dados

No geral, os dados de experimentos agrônômicos se aproximam da distribuição normal à medida que se aumenta o número de observações (repetições e amostras), conforme mencionado no teorema central do limite. Todavia, é comum que devido a limitação financeira ou mão-de-obra, não seja possível utilizar-se de um número de repetições elevado, o que pode gerar possíveis problemas com as pressuposições dos testes paramétricos.

A maioria das técnicas estatísticas paramétricas (ANOVA, teste de comparação de médias, regressão) se fundamentam na pressuposição de normalidade dos erros. Quando cada variável resposta é analisada individualmente, a normalidade univariada é necessária para que os níveis de significância reais dos testes de hipóteses sejam próximos aos estabelecidos pelo pesquisador. Além da normalidade, a homogeneidade das variâncias, também denominada homocedasticidade das variâncias, é muito importante para a inferência estatística (COUTO *et al.*, 2009), pois as hipóteses para os efeitos testáveis dos modelos

estatísticos são testadas considerando-se um erro em comum. Segundo Resende (2007), a violação de qualquer uma das outras pressuposições da análise de variância (aditividade, independência e normalidade dos erros) pode conduzir à heterocedasticidade.

A heterocedasticidade pode ser classificado em irregular, quando alguns tratamentos apresentam maior variabilidade do que os outros, sem existir, entretanto, relação entre a média e a variância; ou regular, quando existe alguma associação entre as médias dos tratamentos e a variância. Caso as pressuposições da normalidade dos erros e homogeneidade das variâncias não sejam atendidas, as informações obtidas pelas análises estatísticas podem gerar sérios equívocos (XU; LI; SONG, 2013). A confiabilidade de todos os testes paramétricos fica comprometida, pois ocorrem alterações na probabilidade de ocorrência do erro tipo I (rejeição da hipótese de nulidade quando ela é verdadeira) e tipo II (não rejeição da hipótese de nulidade quando ela é falsa), podendo levar a falsas conclusões a respeito dos efeitos de tratamento (MARTIN; STORCK, 2008; STEEL *et al.*, 1997; STORCK *et al.*, 2016).

Na prática, os problemas podem estar associados ao reduzido número de repetições, à assimetria externa, presença de erros grosseiros (*outliers*), comportamento anormal de alguns tratamentos ou de parte do experimento. Em certos casos, quando não é possível satisfazer as pressuposições para análise de variância na própria escala observada e quando a heterocedasticidade é regular, transformações têm sido realizadas para uma escala mais apropriada com o objetivo de se conseguir homogeneidade de variâncias e uma distribuição aproximadamente normal (COUTO *et al.*, 2009; CUSTÓDIO; BARBIN, 2009; LÚCIO *et al.*, 2010, 2011).

A transformação de dados consiste na utilização de um artifício matemático para tornar os testes paramétricos válidos. As transformações diretas dos dados mais comumente utilizadas são: a logarítmica, a raiz quadrada e a transformação angular. Um ponto importante a se destacar é que nem sempre a transformação consegue solucionar os problemas para a validade dos testes paramétricos (FERREIRA, 2018). Além disso, a transformação não deve ser usada para reduzir um coeficiente de variação alto, mas sim para os casos em que uma das pressuposições não é atendida (RIBEIRO-OLIVEIRA *et al.*, 2018; SOUZA *et al.*, 2008).

A transformação raiz quadrada é normalmente usada para dados provenientes de contagens, e que por natureza seguem a distribuição de Poisson, em que a média e variância são constantes (BARTLETT, 1936, FERREIRA, 2018). Alguns exemplos de variáveis são: número de bactérias em uma placa, número de plantas ou insetos em uma dada área, por exemplo. Neste caso, a transformação raiz quadrada dos dados estabiliza a variância, além de

torná-la independente da média. Quando se tem observações em que a resposta assume valores zero, pode-se somar uma constante, geralmente 1 ou 0,5. A transformação \sqrt{x} possui a característica de sempre reduzir o coeficiente de variação para aproximadamente metade.

Outra transformação relatada na literatura é a logarítmica, que estabiliza a variância quando o desvio-padrão na escala original varia diretamente com a média, ou seja, o coeficiente de variação é constante de tratamento para tratamento (BARTLETT, 1936; BERRY, 1987, FERREIRA, 2018). Esse tipo de relação entre média e desvio-padrão é encontrado geralmente quando os efeitos são multiplicativos em lugar de aditivos. Nessa situação, tal transformação, além de estabilizar a variância, produz aditividade nos efeitos e tende a aproximar a distribuição dos erros à normal. Essa transformação é usada para números inteiros positivos que cobrem uma grande amplitude, sendo que não pode ser usada diretamente quando ocorrem zeros ou quando alguns dos valores são menores que 10. Neste caso, é necessário ter-se uma transformação que equivale à transformação x para valores pequenos e $\log(x)$ para valores grandes de x . A transformação $\log(x+1)$ é a que mais se aproxima da desejada.

A transformação angular por sua vez, é utilizada para homogeneizar a variância residual dos dados binomiais de proporção $\frac{x}{n}$, ou porcentagens $100 \times \frac{x}{n}$ (FERREIRA, 2018; WARTON; HUI, 2011; ZUBIN, 1935), principalmente na área de sementes, em que as respostas são por exemplo, germinou ou não germinou.

$$Y' = \arcsin \sqrt{y} \text{ (em termos decimais)}$$

$$Y' = \arcsin \sqrt{\frac{y}{100}} \text{ (em termos percentuais)}$$

Normalmente, quando todos os dados estão entre 30 e 70%, não é necessário efetuar a transformação, conforme mencionado por Ferreira (2018).

Por fim, outro importante método para definir a transformação mais adequada, foi aquela proposto por Box e Cox (1964), que mostraram como o parâmetro de transformação λ em $y^* = y^\lambda$ pode ser estimado simultaneamente com outros parâmetros do modelo (média geral e efeitos de tratamentos) usando o método de máxima verossimilhança. O procedimento consiste em realizar, para vários valores de λ , uma análise de variância padrão. A estimativa de máxima verossimilhança de λ é o valor para o qual a soma de quadrado do resíduo, $SQRes(\lambda)$, é mínima. Este valor de λ é encontrado através do gráfico de $SQRes(\lambda)$ vs

λ , sendo que λ é o valor que minimiza a $SQRes(\lambda)$. Ou, ainda, o valor de λ que maximiza a função de log-verossimilhança.

Se o valor de λ for igual a 1, isto quer dizer que não é necessário transformar os dados. O valor de λ pode ser aproximada, sendo os mais comuns: 0,5 (\sqrt{y}); 0 ($\log(y)$); -0,5 ($\frac{1}{\sqrt{y}}$); -1($\frac{1}{y}$).

Outra forma de determinar o valor de λ para transformação de dados é relatado por Barbin (2013) e estabelece uma regressão linear entre o log de S^2 (variável dependente) e o log de \bar{x} (variável independente). Assim temos que: $\lambda = 1 - \frac{\hat{b}}{2}$, em que \hat{b} é o coeficiente angular obtido da regressão linear.

Após efetuar as análises utilizando os dados transformados é comum no contexto da ciência agrárias que, na hora de apresentar os resultados em artigos, teses, dissertações, por exemplo, apresente a média dos dados originais com o teste paramétrico dos dados transformados (FERREIRA, 2018). Assim, na legenda, deve-se especificar qual transformação de dados foi utilizada, veja exemplo da Tabela 1. Esse mesmo procedimento também pode ser usado graficamente.

Tabela 1 – Modelo de tabela para dados transformados.

Tratamento	Perda de massa (%)	SS (°Brix) ^t
T1	1,84 b	3,77 b
T2	1,68 b	3,78 ab
T3	2,62 a	3,87 a
T4	2,62 a	3,75 b
T5	2,64 a	3,79 ab
T6	2,16 ab	3,83 ab
F	12,31	4,55
CV (%)	10,84	1,11

Médias seguidas de mesma letra na coluna não diferem pelo teste de Tukey a 5% de significância. ^tDados transformados para \sqrt{y} .

4.3 A REPRESENTAÇÃO GRÁFICA DE RESULTADOS

Nossa capacidade de absorver e processar informações quando apresentadas na forma de uma imagem é tão boa que é natural explorar esse talento ao analisar dados obtido de um experimento. Quando os dados são apresentados de forma gráfica, podem ser detectadas tendências ou características nos dados que dificilmente reconheceríamos se os dados foram apenas apresentados em forma de tabela (KIRKUP, 2019). Isto é especialmente verdadeiro em situações em que um conjunto de dados consistem em centenas ou milhares de valores, o que é uma ocorrência comum quando um *software* é usado para auxiliar na coleta de dados. Além disso, a representação de dados na forma de um gráfico é uma excelente maneira de resumir muitas das características importantes de um experimento. Um gráfico pode indicar:

- (i) as quantidades que estão sendo estudadas
- (ii) a faixa de valores obtidos por meio de medição
- (iii) a incerteza em cada valor
- (iv) a existência ou ausência de uma tendência nos dados coletados (por exemplo, os pontos plotados podem estar em uma linha reta ou curva, ou podem parecer dispersos aleatoriamente no gráfico)
- (v) quais pontos traçados não seguem a tendência geral exibida pela maioria dos dados.

A representação gráfica, no caso particular das ciências agrárias é bastante utilizada quando se trabalha com fatores de níveis quantitativos, o que gera gráficos de dispersão com linha de tendência do estudo em questão (ARCHONTOULIS; MIGUEZ, 2015). Contudo, vários trabalhos utilizam a representação gráfica de estudos em que o objetivo foi analisar fatores de níveis qualitativos, sendo que, nesses casos em questão, a adoção de gráficos de colunas, de barras ou de caixas tem sido prevaletido frente às demais formas.

Essas representações gráficas ajudam o leitor e o escritor a interpretar os resultados, visto que, é mais fácil ver possíveis tendências quando na forma de gráfico em relação a uma tabela. Não que essa última seja limitada, mas a percepção é facilitada no caso daquela primeira. Além disso, a visualização gráfica também auxilia na identificação de possíveis erros na análise. Por exemplo, quando se trabalha com dados de produtividade de soja, que normalmente é próximo de 3.500 kg ha^{-1} , em média, espera-se que os valores dessa variável sejam próximo a isso. Contudo, erros podem acontecer no momento de calcular as estimativas de produtividade, com valores fora dos esperados.

A escolha da forma como será representado graficamente é talvez um dos pontos de maior preocupação e, embora já se tenha modelos pré-estabelecidos para cada variável ou combinação de variáveis, ainda gera discórdia na comunidade científica. De forma geral, independentemente do tipo de gráfico, todos devem apresentar alguns elementos essenciais e características, como apresentar apenas informações relevantes e de forma mais simples possível, para facilitar o entendimento pelo leitor, além de conter informações básicas, como por exemplo, os nomes dos eixos, a presença das escalas e de legendas (HARRIS, 1999). Aliado a isso, todo gráfico deve ter boa resolução gráfica e dimensões de fonte adequadas.

4.3.1 Tipo de Variável

O tipo de variável é um dos pontos primordiais para definir a forma de representação gráfica. A definição dos tipos de variáveis, sobretudo em relação as classificações em relação aos níveis dos fatores em um experimento agrícola já foram abordadas na seção 4.2.1.

De forma geral, quando os fatores apresentam níveis quantitativos, os dados são apresentados graficamente na forma de um diagrama de dispersão com a linha de tendência e informações dos modelos ajustados bem como alguns parâmetros (Figura 4 e Figura 5). Quando é trabalhado fatores de níveis qualitativo, tem-se prevalecido a utilização de gráficos de colunas (Figura 6 A) ou barras (Figura 6 D). Outras formas de representação comuns incluem o gráfico de caixas (Figura 6 B) e o de pontos (Figura 6 C).

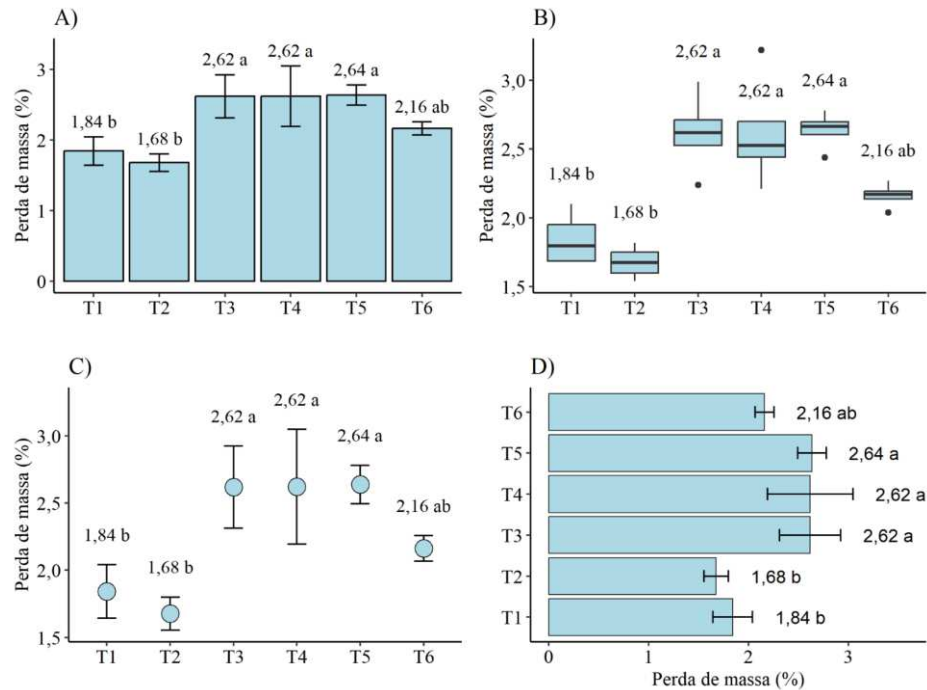


Figura 6 – Representações gráficas para fatores com níveis qualitativos: colunas (A); caixas (B); pontos (C) e barras (D). **Fonte:** o próprio autor.

Um gráfico de dispersão apresenta como característica a presença de duas variáveis quantitativas, uma chamada de independente e a outra chamada de dependente ou variável resposta. Um gráfico de coluna apresenta uma variável independente qualitativa que é representada no eixo X e uma variável dependente quantitativa no eixo Y, sendo que as colunas devem ser representadas de mesma cor ou marcação. Essa pode ser invertida, com as barras na horizontal, sendo então chamada de gráfico de barras. O gráfico de caixas e pontos para esse propósito possuem a mesma característica, embora essa primeira apresente várias medidas descritivas.

4.3.2 Número de Fatores

O número de fatores também influencia a forma como será representada graficamente os resultados. Por exemplo, quando o experimento é analisado com dois fatores, um de nível qualitativo e um de nível quantitativo e, se a interação é significativa e com todos os níveis do fator qualitativo em relação aos níveis dos fatores quantitativos também significativos, representamos como um gráfico de dispersão com uma curva de tendência para cada nível do fator qualitativo (Figura 7 B). Por outro lado, quando há um experimento que foi

analisado dois fatores qualitativos, pode-se representar isso na forma de colunas compostas (Figura 7 A) ou também como um gráfico de interação.

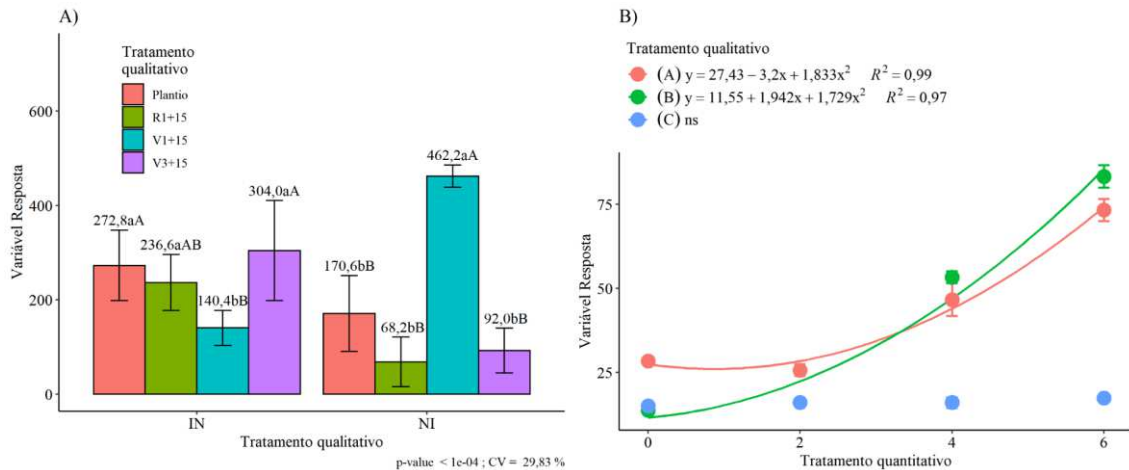


Figura 7 – Representação gráfica de colunas compostas de uma interação dupla significativa com fatores qualitativos (A); representação gráfico de interação significativa com um fator qualitativo e um quantitativo (B). IN: inoculado; NI: Não inoculado. **Fonte:** o próprio autor.

O gráfico de interação é talvez uma das formas de representação mais subjogadas na comunidade científica, principalmente por quem não entende de estatística. Contudo, é a forma mais correta de representação de uma interação significativa de dois fatores qualitativos. Por exemplo, supondo um projeto fatorial com dois fatores e dois níveis para cada um (Fator 1: A e B; Fator 2: T1 e T2), quando não há interação significativa, a alteração de um nível de um fator não irá refletir na resposta de outro fator (Figura 8 A e Figura 8 B), que demonstram o efeito isolado de Fator 2 e 1, respectivamente. Na Figura 8 A, sempre o T2 apresenta resposta melhor que T1, independente se o Fator 1 é A ou B. Por outro lado, quando há interação significativa, isso basicamente indica a mudança na resposta dos níveis de um fator em relação ao outro. Por exemplo, na Figura 8 C, quando estamos no nível A do Fator 1, o nível T2 do Fator 2 parece ter resposta inferior a T1, por outro lado, quando estamos no nível B do Fator 1, temos que ambos os níveis do Fator 2 tem resposta similar. Um outro exemplo de interação significativa ocorre na Figura 8 D, contudo, nesse caso a alteração dos níveis de um fator afeta fortemente os níveis de outro fator, exibindo respostas completamente divergentes, nesse caso, chamamos de interação do tipo complexo.

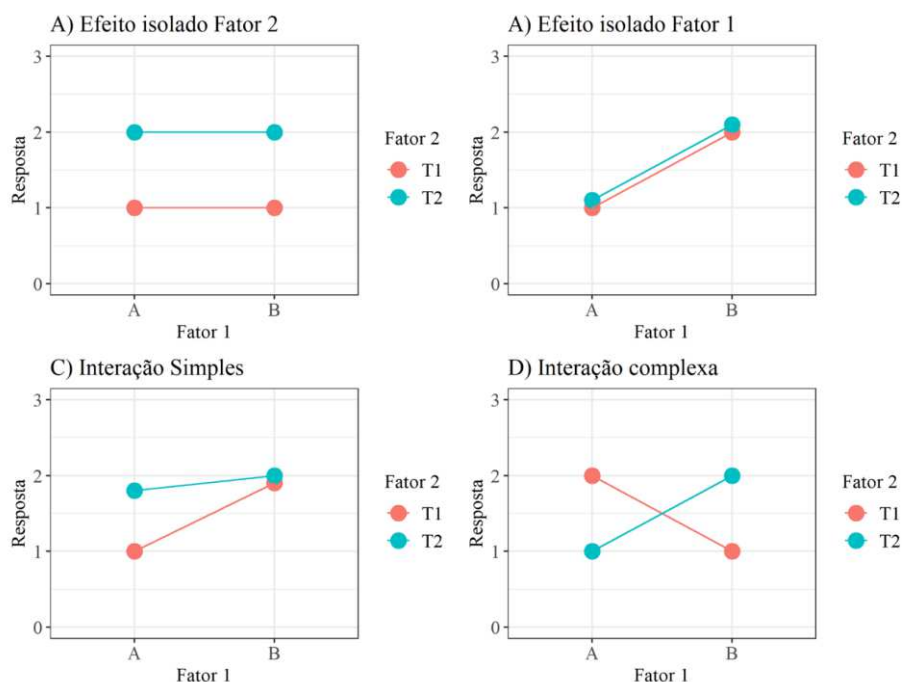


Figura 8 – Representação gráfica dos efeitos de um projeto experimental em esquema fatorial 2×2 : Efeito isolado de Fator 2 (A); efeito isolado de Fator 1 (B); efeito de interação simples (C) e efeito de interação complexo (D). **Fonte:** o próprio autor.

4.4 DESAFIOS NA ANÁLISE DE EXPERIMENTOS AGRÍCOLAS

4.4.1 Área de Tecnologia de Sementes

Os dados de germinação de sementes possuem várias características que os diferenciam de outros dados frequentemente coletados em pesquisas agrícolas (SCOTT; JONES; WILLIAMS, 1984). A germinação é considerada como uma variável qualitativa, pois os valores assumem respostas binomiais, ou seja, plantas germinadas e plantas não germinadas. Normalmente, os pesquisadores estão interessados em avaliar os efeitos específicos de um ou mais tratamentos no tempo até a germinação. No entanto, os indivíduos dentro de um tratamento podem responder em momentos diferentes, de modo que uma distribuição de respostas é observada para cada tratamento ou unidade experimental. Essa distribuição é frequentemente viesada para a direita ou a esquerda, gerando inflações de zeros ou 100%. Além disso, para uma dada população de sementes, a germinação completa raramente ocorre. Dentre as causas da não germinação incluem 1) inviabilidade inicial; 2) perda de viabilidade pelo efeito do próprio

tratamento; 3) censura aleatória, por exemplo, perda ou contaminação acidental; e 4) censura à direita devido ao término prematuro do teste, por exemplo, dormência.

Testes para avaliação da qualidade das sementes são rotineiramente utilizados em experimentação. Um dos mais destacados desses testes de rotina é o teste de germinação, quase sempre associado a algum teste indicativo de vigor, como o teste de crescimento de plântulas (NAKAGAWA, 1999). No teste de germinação, é comum a contagem diária do número de sementes germinadas, e esses dados são frequentemente utilizados para cálculo do índice de velocidade de germinação (MAGUIRE, 1962), tempo médio de germinação e velocidade de germinação (EDMOND; DRAPALA, 1958). Este mesmo procedimento é aplicado aos dados de emergência de plântulas, além de ser usado analogamente a outros campos da pesquisa, como na brotação de estacas de videira (RIBEIRO *et al.*, 2021).

Em relação aos dados de crescimento de plântulas, variáveis como parte aérea, raiz e comprimento total são as mais utilizadas. Assim, vários índices podem ser calculados a partir desses dados, o que pode ser uma tarefa complicada ao lidar com muitos lotes/tratamentos, como ocorre, por exemplo, em estudos de fenotipagem (JOOSEN *et al.*, 2010).

Nas últimas duas décadas, o advento do processamento computacional permitiu a quantificação de várias características de germinação de sementes, como os já citados anteriormente. De fato, uma variedade de técnicas computacionais tem sido proposta para melhorar a visualização de dados de germinação das sementes, como a captura de imagens digitais de sementes (DELL'AQUILA, 2009; HOFFMASTER *et al.*, 2005; WAGNER *et al.*, 2011) e seu processamento (SAKO *et al.*, 2001; SINÍCIO *et al.*, 2009).

Embora muitos programas de *software* de baixo custo já estejam disponíveis para as mais diversas áreas, há poucos que permitem cálculos voltados para a área de tecnologia de sementes e que também seja de fácil uso. No entanto, há algumas planilhas gratuitas que não são realmente um *software*, mas um conjunto de planilhas desenvolvidas no pacote Microsoft® Excel. Nesta categoria estão incluídas as planilhas desenvolvidas por Joosen *et al.* (2010), Miranda *et al.* (2014) e Ranal *et al.* (2009). A principal limitação das planilhas propostas por Miranda *et al.* (2014) e Ranal *et al.* (2009) é a impossibilidade de trabalhar com diferentes tratamentos no mesmo arquivo ao mesmo tempo; cada tratamento ou situação deve ser interpretado em diferentes planilhas que consomem tempo. As planilhas propostas por Joosen *et al.* (2010) seriam as únicas, permitindo, portanto, o cálculo de dois ou mais tratamentos ao mesmo tempo. No entanto, essa não é uma ferramenta de domínio universal. Ela exige que o

usuário entenda profundamente o funcionamento dessas planilhas para melhor explorar a ferramenta e suas funcionalidades.

Implementações desses cálculos podem ser encontradas também em pacotes estatísticos do *software* R, como no caso do SeedCalc (SILVA; MEDEIROS; OLIVEIRA, 2019), que utiliza dados de contagem diária de germinação/emergência e medidas de comprimento de plântulas para gerar automaticamente uma série de variáveis relacionadas à qualidade fisiológica das sementes, incluindo porcentagem de germinação/emergência, tempo necessário para germinação ou emergência (T_{10} , T_{50} , T_{90} e tempo médio de germinação), velocidade (índice de velocidade de germinação, taxa média de germinação), variabilidade ou heterogeneidade (coeficiente de variação do tempo de germinação, variação de germinação), incerteza e sincronia.

Outro exemplo é o *software* GerminaR/GerminaQuant (BENITES-ALFARO; POMPELLI, 2019; LOZANO-ISLA; MARQUES *et al.*, 2015), que foi desenhado para facilitar o entendimento dos processos de germinação e a análise dos dados. O *package* GerminaR fornece um aplicativo web interativo para análise estatística e gráfica da germinação. Além disso, os pesquisadores podem usar o console R para análise de dados reprodutíveis. Foi o primeiro aplicativo de análise de germinação que forneceu tanta facilidade e flexibilidade de interação para uso (JOOSEN *et al.*, 2010; MIRANDA *et al.*, 2014; RANAL *et al.*, 2009). O pacote e o aplicativo *web* funcionam independentemente do sistema operacional, funcionando facilmente no Windows, Linux, Macintosh, Android, iOS e outros dispositivos móveis.

Vale ressaltar que dentre esses pacotes e *softwares* citados, a maioria, não necessariamente, apresenta todas as aplicações no cotidiano dos pesquisadores ligados à área de sementes, parte pela baixa divulgação em ambiente acadêmico, parte por não apresentar a realidade da maioria dos experimentos dessa área. Isso é notoriamente verdadeiro, quando analisamos a infinidade de possíveis experimentos, a exemplo no caso do estudo da germinação de sementes sob diferentes temperaturas. Nesses casos, a adoção de metodologias tradicionais de análise (Anova/Teste de médias) não é uma boa alternativa, devendo-se aplicar métodos baseados em regressão, sobretudo modelos não-lineares, como ocorre em alguns trabalhos (ANDRADE; CADIMA; ABREU, 2018; ATABAKI *et al.*, 2021, CABRERA-SANTOS *et al.*, 2022). Visto isso, há a necessidade de explorar os diversos ramos da área de semente que ainda carecem de ferramentas para análise.

4.5 SOFTWARES COM APLICAÇÕES EM EXPERIMENTAÇÃO AGRÍCOLA

4.5.1 *Software* R

No ano de 1992, os pesquisadores Ross Ihaka e Robert Gentleman, ambos da Universidade de Auckland, na Nova Zelândia, desenvolveram um *software* para análise estatística e atribuíram o nome R, o qual foi baseado na linguagem de programação S inicialmente desenvolvido nos AT&T Bell Laboratories em meados dos anos 1970 (CALENGE, 2006). Quando Ihaka e Gentleman lançaram o *software* R, atribuíram a este a funcionalidade de ser de código aberto (IHAKA; GENTLEMAN, 1996), lançando, então a plataforma “*Comprehensive R Archive Network (CRAN)*” que é o núcleo onde se hospeda a maioria dos pacotes do *software* R. O CRAN é uma comunidade cada vez mais crescente e o principal repositório de materiais relacionados ao R (HORNİK; LEISCH, 2002).

Desde sua criação, o R passou a ser uma linguagem de programação e de ambiente para análises estatísticas computacionais e geração de gráficos. O *software* R fornece ampla variedade de análises estatísticas, tais como modelagem linear e não linear, testes estatísticos clássicos, análise de séries temporais, classificação, cluster e técnicas gráficas; todas de alta reprodutibilidade. O R foi projetado dentro de uma linguagem de programação legítima, que permite aos usuários adicionar funções dentro das suas necessidades. As funcionalidades do R podem ser estendidas via pacotes para desenvolver determinadas funções. Atualmente existem cerca de oito pacotes básicos fornecidos com a distribuição de R e outros tantos estão disponíveis através do CRAN, com servidores na internet que cobrem uma gama muito ampla de ferramentas estatísticas modernas (R CORE TEAM, 2023).

Atualmente, mais de 20.000 pacotes estão disponíveis no repositório CRAN (<https://cran.r-project.org/web/packages/>) nas diversas áreas do conhecimento. Os pacotes podem ser de uso geral, como os implementados na sua base, como o *stats*, *graphics*, *base*, entre outros, ou pacotes de uso mais específico, como por exemplo o *pedigreemm* (VAZQUEZ *et al.*, 2010), que é utilizado na área de animais, com ênfase na matriz de relacionamento utilizando modelos lineares generalizados mistos. Os pacotes podem também ser de uso específico, mas generalizado para diferentes situações, como exemplo, o pacote *multcomp* (HOTHORN; BRETZ; WESTFALL, 2008), com funções globais para testar contrastes em modelos lineares. Assim, R tornou-se uma importante ferramenta tecnológica na análise e manipulação de dados.

São inúmeros os pacotes com aplicação na experimentação agrícola, embora um número restrito de pacotes tem sido usado efetivamente pelos pesquisadores, como: *dae* (BRIEN, 2021), *easyanova* (ARNHOLD, 2013), *ExpDes* (*ExpDes.pt*) (FERREIRA *et al.*, 2014) e o *agricolae* (MENDIBURU, 2021) (Figura 9).

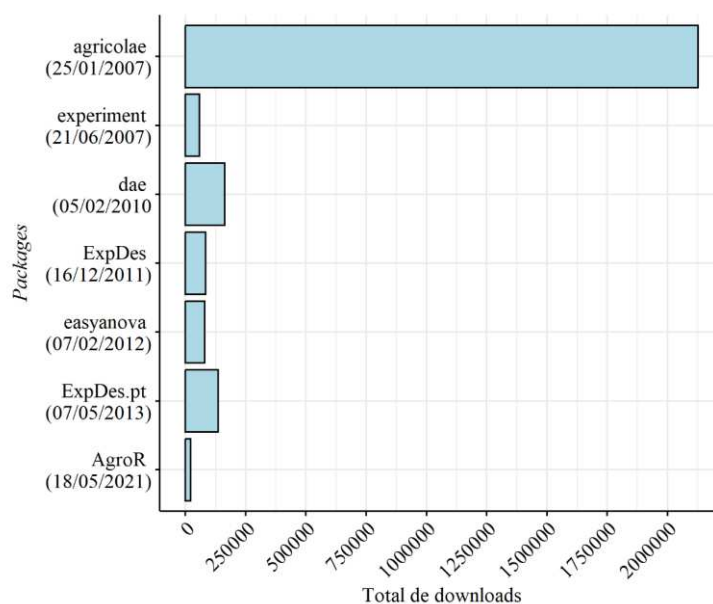


Figura 9 – Número de downloads dos principais pacotes usados na estatística experimental desde o primeiro lançamento até 03/01/2024. **Fonte:** o próprio autor.

O pacote *dae* (BRIEN, 2021) fornece várias ferramentas sobre desenho experimental e arranjos de fatores. Inclui um grupo de funções que auxiliam na geração de projetos experimentais, como *fac.nested()* que pode lidar com várias estruturas e funções aninhadas para combinar vários fatores em um (função *fac.combine()*) ou dividindo um fator em vários fatores (função *fac.divide()*). Também inclui funções que facilitam a verificação diagnóstica após um função base *aov()* (Analysis of Variance Model), como função *fitted.aovlist()* que extrai os valores instalados para todos os termos do modelo e *residuals.aovlist()* que extrai os resíduos na análise.

O pacote *ExpDes* (FERREIRA *et al.*, 2014) e sua versão em português, o *ExpDes.pt*, tem como objetivo principal a análise de experimentos equilibrados para modelos fixos. Uma característica especial deste pacote é permitir a análise completa em uma única execução. Sua saída resume o que o pesquisador precisará para interpretar a hipótese estudada. O principal objetivo do pacote é analisar experimentos simples em delineamento inteiramente casualizados (*crd()* ou *dic()*), blocos casualizados (*cbd()* ou *dbc()*) e quadrado latino (*latsd()* ou

dql()). Também permite a análise de tratamentos em um desenho fatorial com dois ou três fatores (*fat2.crd()*, *fat3.crd()*, *fat2.rbd()*, *fat3.rbd()*) e também a análise de esquema de parcelas subdivididas (*split2.crd()*, *split2.rbd()*). Outra funcionalidade é analisar experimentos com um tratamento adicional em delineamento inteiramente casualizado ou blocos casualizados com dois ou três fatores (*fat2.ad.crd()*, *fat2.ad.rbd()*, *fat3.ad.crd()* e *fat3.ad.rbd()*).

O pacote *easyanova* (ARNHOLD, 2013) possui funções para realizar análise de variância de tratamentos qualitativos, com várias opções de testes de contrastes de médias e análise de resíduos. Diferente do *ExpDes*, o pacote retorna vários testes de médias, em contrapartida, mesmo não ocorrendo diferença significativa dos fatores estudados é retornado o teste de médias, o que muitas vezes pode confundir o usuário. Outra desvantagem está no fato do pacote não realizar análise de regressão. O pacote dispõe das funções *ea1()*, *ea2()* e *ec()* e 19 conjuntos de dados. A função *ea1()* realiza análises em 13 delineamentos e os testes não-paramétricos de Kruskal-Wallis e Friedman. A função *ea2()* realiza análises em esquemas experimentais com interação (13 esquemas). A função *ec()* testa contrastes de grupos de médias. Os dados e exemplos disponíveis no pacote são voltados principalmente à experimentação animal. Entre delineamentos possíveis de serem analisados têm-se alguns muito utilizados na experimentação animal, como os quadrados latinos e parcelas subdivididas no tempo. As funções do pacote também são de fácil utilização e fazem ajustes necessários nos casos de experimentos desbalanceados.

O pacote *agricolae* (MENDIBURU, 2021) oferece uma ampla funcionalidade no desenho de experimentos, principalmente para experimentos em agricultura e melhoramento de plantas, que também podem ser utilizados para outros fins. Contém os seguintes delineamentos: desenhos de treliça, alfa, cíclicos, blocos incompletos balanceados, blocos aleatorizados completos, latim, greco-latino, desenhos de blocos aumentados, parcelas divididas e parcelas em tiras. Também tem vários procedimentos de análise de dados experimentais, como as comparações de tratamentos de Waller Duncan, Bonferroni, Duncan, Student-Newman-Keuls, Scheffe, ou o clássico LSD e Tukey; e comparações não paramétricas, como Kruskal-Wallis, Friedman, análise de estabilidade e outros procedimentos aplicados em genética, bem como procedimentos em biodiversidade e estatísticas descritivas.

Existem diversos outros pacotes que podem ser utilizados para analisar experimentos agrícolas, embora os principais tenham já sido mencionados. Contudo, há também usuários que preferem realizar suas análises de forma mais manual, ou seja, escrevendo seus códigos manualmente, sem a necessidade de pacotes específicos, ou apenas usando aqueles

pacotes de uso mais geral. Por exemplo, uma análise de variância no R pode ser realizada usando a própria base, como por exemplo, através das funções *aov* ou *lm*, após isso, o quadro é retornado usando a função *anova* ou *summary*, ou através do pacote *car*, usando a função *Anova*. Uma vez constatado diferença significativa, o usuário pode usar funções, como por exemplo o *TukeyHSD*, ou através de pacotes como o *multcomp* e o *emmeans*, quando o fator é qualitativo. Resumindo, são infinitas as possibilidades de analisar os dados usando o R, cabe a cada pesquisador escolher a forma que julgar mais fácil e confiável.

4.5.2 Sisvar

O Sisvar (Figura 10) é um sistema de análise estatística desenvolvido em linguagem Pascal por Daniel Furtado Ferreira, que foi lançado em 1996, embora o seu desenvolvimento tenha sido iniciado em 1994. Os principais objetivos foram de obter um *software* que pudesse ser usado diretamente no curso de estatística experimental do Departamento de Ciências Exatas da Universidade Federal de Lavras e iniciar o desenvolvimento de um *software* genuinamente brasileiro, gratuito que atendesse às demandas e peculiaridades das pesquisas realizadas no país, além de permitir que os resultados da pesquisa pudessem ser analisados de forma eficiente e confiável.

O Sisvar foi bastante aceito pela comunidade científica, sobretudo no Brasil, e isso é decorrente do fato de que ele é capaz de permitir uma maior interatividade com o usuário e produzir análises confiáveis, pelo fato delas serem exatas, precisas, simples e robustas. Contudo, não é possível saber o número exato de trabalhos que utilizaram o *software*, uma vez que vários trabalhos sobre ele surgiram ao longo dos anos (FERREIRA, 2019; FERREIRA, 2014; FERREIRA, 2011; FERREIRA, 2008; FERREIRA, 2000).

Embora o Sisvar seja de fato um dos principais *softwares* estatísticos, há de se destacar algumas limitações. Por exemplo, para o teste de normalidade dos erros de Shapiro-Wilk, é necessário calcular os resíduos separadamente do *software* e após isso realizar o teste, o que consome tempo e torna passível de erros. Possui também a desvantagem de necessitar de uma entrada de dados em formato específico, no caso, a extensão *.dbf*, além de exigir ampla noção de modelos de análise de variância para a correta análise, o que facilmente pode gerar equívocos. E por fim, a saída final, gerada em extensão *.rtf*, apresenta uma estrutura pouco amigável de interpretação, o que pode confundir o usuário, sobretudo em projetos mais complexos.

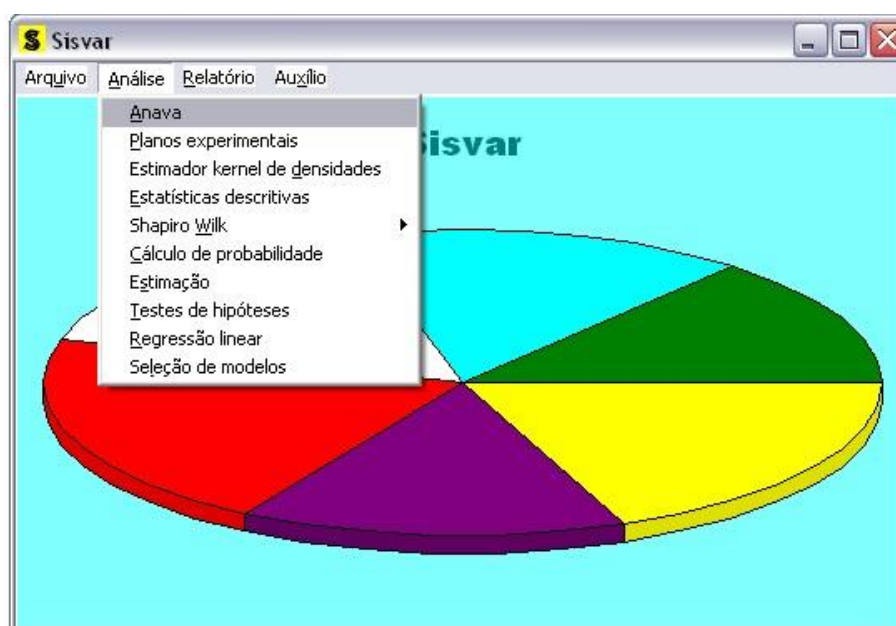


Figura 10 – Interface gráfica do *software* Sisvar

4.5.3 SASM-Agri

O SASM-Agri (Figura 11) (CANTERI *et al.*, 2001) foi desenvolvido utilizando-se o ambiente Borland Delphi, compatível com Windows 98 ou superior, o que permite importar dados de planilhas e exportar resultados para outros aplicativos. Os dados também podem ser transformados antes da análise. O sistema usa funções recursivas que agilizam e aumentam a precisão da apresentação dos resultados, e foi sendo um dos pioneiros na implementação do método de agrupamento de Scott-Knott. Contudo, apresenta poucos projetos experimentais implementados, sobretudo em experimentos mais complexos como esquemas fatoriais e parcelas subdivididas.

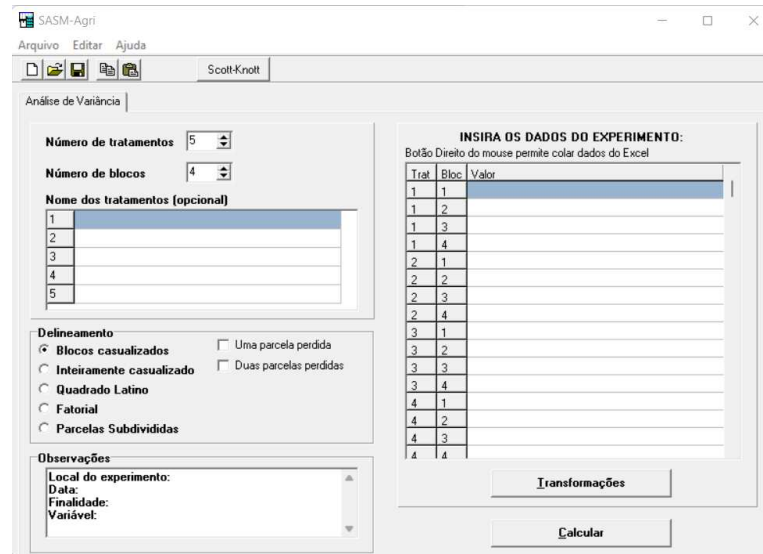


Figura 11 – Interface gráfica do *software* SASM-Agri

4.5.4 Assistat

O *software* Assistat 7.7 (Figura 12) é um aplicativo para Windows e desenvolvido por Silva e Azevedo (2016) e funciona com a geração de arquivos em extensão .txt. A entrada de dados pode ser realizada por meio de digitação, leitura de dados em arquivo .txt e importação de dados de uma planilha do Excel. É distribuído gratuitamente, sendo útil na análise de dados experimentais de diversos projetos: análise de variância e seus pressupostos; transformação de dados; teste de médias e análise de regressão, dentre outras funções. Apresenta uma interface amigável e de fácil uso. Contudo, não está mais disponível para uso.

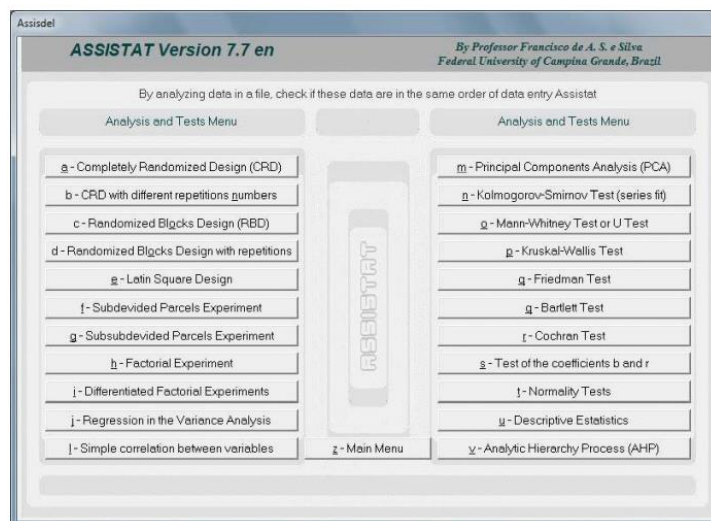


Figura 12 – Interface gráfica do *software* Assistat 7.7

4.5.5 SAS

O SAS (www.sas.com) é um dos mais antigos *softwares* estatísticos, cuja origem é datada em 1976 e compreende um sistema integrado de aplicações para o processamento e análise estatística de dados, consistindo em módulos de acesso e recuperação de dados, gerenciamento de arquivos, geração de gráficos e relatórios. Pode ser instalado em diversos sistemas operacionais e trabalha com quatro ações básicas sobre os dados: acessar, manipular, analisar e apresentar.

O SAS é um *software* pago, dessa forma apresenta suporte pela empresa, diferente da linguagem R, assim, qualquer inovação ou nova técnica estatística deve ser examinada e aceita, o que aumenta a segurança e a estabilidade da plataforma. Esse *software* se destaca no manuseio de conjuntos de Big Data. As edições do servidor do SAS de fato podem analisar uma grande quantidade de dados sem problemas, enquanto a linguagem R é mais provável que gere erros ou trave o ambiente de desenvolvimento. A sintaxe de programação SAS pode ser considerada como uma linguagem de alto nível que é intuitiva e fácil de aprender, além disso, foi projetada como DML (Linguagem de Manipulação de Dados).

Para a análise de um experimento, as implementações já estão no próprio *software*, diferente do R, que em muitos casos necessita instalar os pacotes complementares. Comandos como o PROC Anova, realiza à análise de variância de efeitos fixos, enquanto por exemplo, o PROC glm realiza à análise para modelo linear generalizado ou PROC MIXED realiza por meio de modelos mistos. Diversos testes de média são implementados no SAS, além de análise de regressão e técnicas multivariadas. Resumindo, quase todas as análises são implementadas nesse *software*, contudo, a maior limitação é ser um *software* pago, embora uma versão gratuita esteja disponível virtualmente para acadêmicos (https://www.sas.com/na_us/software/on-demand-for-academics.html).

4.5.6 Outros *softwares*

Existe vários outros *softwares* relatados na literatura, alguns pagos e outros gratuitos, cada um com suas vantagens e desvantagens. A exemplo de *software* pago, existe o Action Stat Pro (<http://loja.portalaction.com.br/produto/action-stat-pro/>), que realiza análise de diversos projetos experimentais, além de análise de regressão, teste não-paramétricos e análise multivariada. Outros exemplos incluem o *software* Prism

(<https://www.graphpad.com/scientific-software/prism/>),

Statistica

(<https://www.statistica.com/na/>), Genes (CRUZ, 2006a, 2006b, 2008, 2013), entre outros.

4.6 CONSTRUÇÃO DE UM PACOTE R

Um pacote R é um compilado de funções simples e complexas destinadas a realização de determinada atividade, seja uma análise estatística, a construção de um gráfico, entre outras finalidades (VENABLES; SMITH; R CORE TEAM, 2024). É um processo detalhado que permite aos usuários e desenvolvedores criar, compartilhar e implementar conjuntos de funções, dados e documentação de maneira estruturada e acessível.

Para a construção de um pacote R, é necessário ter a linguagem R instalada no computador, bem como seus complementos (pacotes e *softwares*). Recomenda-se a utilização de um IDE (Integrated Development Environment), como o Rstudio (<https://posit.co/>), pois facilita a criação de pacotes.

Atualmente, dois pacotes do R têm sido utilizados para facilitar a construção de outros pacotes, são eles o *devtools* (WICKHAM *et al.*, 2022a) e o *roxygen2* (WICKHAM *et al.*, 2022b). Recomenda-se no IDE Rstudio, após a instalação desses pacotes, a criação de um projeto de pacote R (ir em: Project > New Project... > New Directory > R package) (Figura 13).

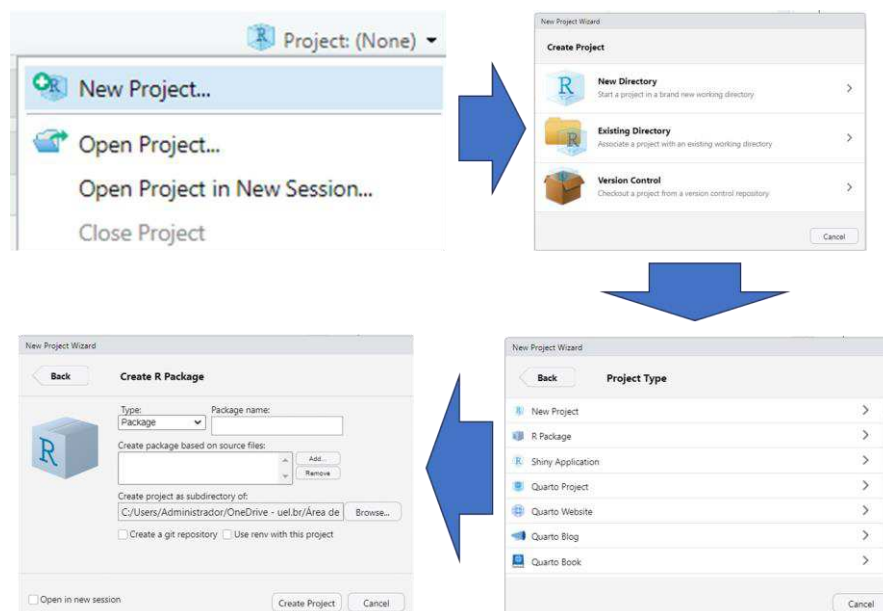


Figura 13 – Organograma de criação de um projeto de pacote R usando o IDE Rstudio. **Fonte:** o próprio autor.

A estrutura básica de um pacote R é composto de uma subpasta “R” que contém as funções criadas, uma subpasta “man” que contém a documentação do pacote em extensão .Rd, um arquivo “DESCRIPTION” que contém diversas informações a respeito do pacote (Figura 14).

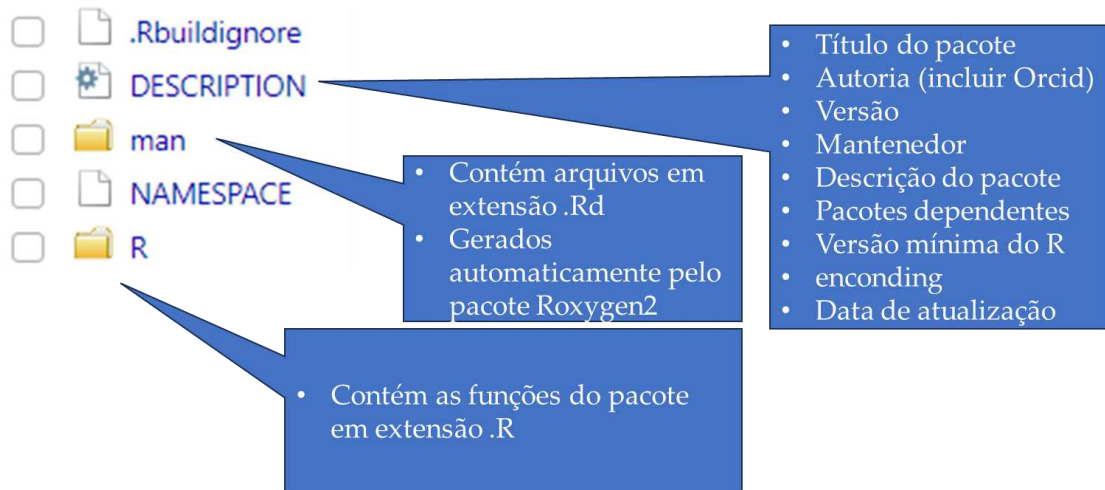


Figura 14 – Estrutura básica de um pacote R. **Fonte:** o próprio autor.

A estrutura pode ser composta de outros arquivos, como a subpasta “data”, que inclui o conjunto de dados do pacote; o arquivo “NEWS.md” que é usado para informar modificações no pacote das versões disponibilizadas, normalmente atrelada a correções de erros e adição de novas funções; o arquivo “README.md” que é usada como um arquivo geral para apresentar o pacote, informações úteis, restrições, entre outros.

A função de um pacote R deve ser construída em um arquivo .R, e contém o título da função, a descrição da finalidade, a autoria, os argumentos da função e suas descrições, as informações retornadas da função, exemplos de utilização e a função construída (Figura 15) (MONTGOMERY; MOORE, 2014; VIDONI; CODABUX, 2023). Para automatizar a geração da documentação em arquivo .Rd, no Rstudio, deve-se configurar em: Tools > Project Options ... > Build Tools > Marcar Generate documentation with Roxygen > Configure...> marcar todas as caixas.

Título da função	#' Graph: Bar graph for one factor
Descrição da função	#' @description This is a function of the bar graph for one factor
Autoria	#' @author Gabriel Danilo Shimizu, \email{shimizu@uel.br}
Argumentos da função	#' @author Leandro Simoes Azeredo Goncalves #' @author Rodrigo Yudi Palhaci Marubayashi
O que a função retorna	#' @param model DIC, DBC or DQL object #' @param fill fill bars #' @param horiz Horizontal Column (\emph{default} is TRUE) #' @param width.col Width Column
Exemplo de utilização	#' @export #' @return Returns a bar chart for one factor #' @seealso \link{radargraph}, \link{barplot_positive}, \link{plot_TH}, \link{plot_TH} #' @examples #' data("laranja") #' a=with(laranja, DBC(trat, bloco, resp, #' mcomp = "sk", angle=45, #' ylab = "Number of fruits/plants")) #' bar_graph(a, horiz = FALSE)
Função do pacote	bar_graph=function(model, fill="lightblue", horiz=TRUE, width.col=0.9){ requireNamespace("ggplot2") data=model[[1]]\$data
No caso de importar pacote, usar requireNamespace ao invés de require/library	

Figura 15 – Exemplo de estrutura básica da função de um pacote R. **Fonte:** o próprio autor.

Após a conclusão do pacote, deve-se checar o mesmo usando o pacote *devtools* (Figura 16). Para isso, o desenvolvedor deve clicar em “Build” e em seguida “Install”. Após esse processo, o pacote será instalado localmente. Uma vez instalado, deve-se clicar em “Check” e quando concluído, será retornado à ocorrência de erros, avisos e notas. Caso o desenvolvedor queira disponibilizar no repositório CRAN, erros e avisos no geral não são tolerados, por outro lado, no caso de notas, o mesmo pode ser aceito, mediante a justificativa. Vale ressaltar que embora a checagem pelo pacote *devtools* seja eficiente, alguns erros podem não ser detectados, visto que a checagem é realizada no computador local, sob um sistema operacional específico e com configurações específicas, o que não ocorre no repositório CRAN, que possui diversos computadores de diferentes configurações e em diferentes versões do R.

Para submissão ao CRAN, deve-se gerar um arquivo tar.gz usando a função *build()* do pacote *devtools*. A submissão ao CRAN deve ser encaminhada no url: <https://cran.r-project.org/submit.html>.

```

Environment History Connections Build Tutorial
Install Test Check More
==> Rcmd.exe INSTALL --preclean --no-multiarch --with-keep.source Teste
* installing to library 'C:/Program Files/R/R-4.3.1/library'
* installing *source* package 'Teste' ...
** using staged installation
** R
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded from temporary location
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (Teste)

Environment History Connections Build Tutorial
Install Test Check More
checking for dependencies in the compilation directory

See
'C:/Users/Administrador/OneDrive - uel.br/Área de Trabalho/FISHER_UEL_BR/Teste.Rcheck/00check.1
for details.

— R CMD check results — Teste C
Duration: 31.1s

> checking DESCRIPTION meta-information ... WARNING
Non-standard license specification:
  What license is it under?
Standardizable: FALSE

0 errors ✓ | 1 warning ✗ | 0 notes ✓
Erro: R CMD check found WARNINGS
Execução interrompida

Exited with status 1.

```

Figura 16 – Instalação e checagem de um pacote R. **Fonte:** o próprio autor.

4.7 APLICATIVO SHINY

O *shiny* é um pacote R que ajuda os usuários a construir um aplicativo web com funções em R. O aplicativo *shiny* gerado, desde que instalado sob um servidor e disponibilizado via web não requer a instalação do R ou qualquer conhecimento prévio de programação. É particularmente útil para usuários que precisam executar funções R que não estar disponíveis em outras linguagens de programação (JIANG *et al.*, 2024).

Um aplicativo *shiny* pode ser construído de diversas formas e depende essencialmente do conhecimento do desenvolvedor. Nessa seção, será abordado brevemente a construção de um aplicativo *shiny* simples, usando o próprio pacote *shiny*. Contudo, vale ressaltar que existem pacotes baseados no pacote *shiny*, a exemplo do pacote *golem*. Além disso, a estrutura gráfica do aplicativo pode ser facilmente alterada usando outros pacotes como *shinydashboard*, *shinythemes*, *shinyWidgets*, entre outros.

Uma vez instalado o pacote *shiny*, o desenvolvedor poderá construir o aplicativo criando um projeto do tipo “Shiny application” (Figura 17). Após a criação do projeto, será gerado um arquivo “app.R”, que contém um exemplo básico de estrutura de um aplicativo.

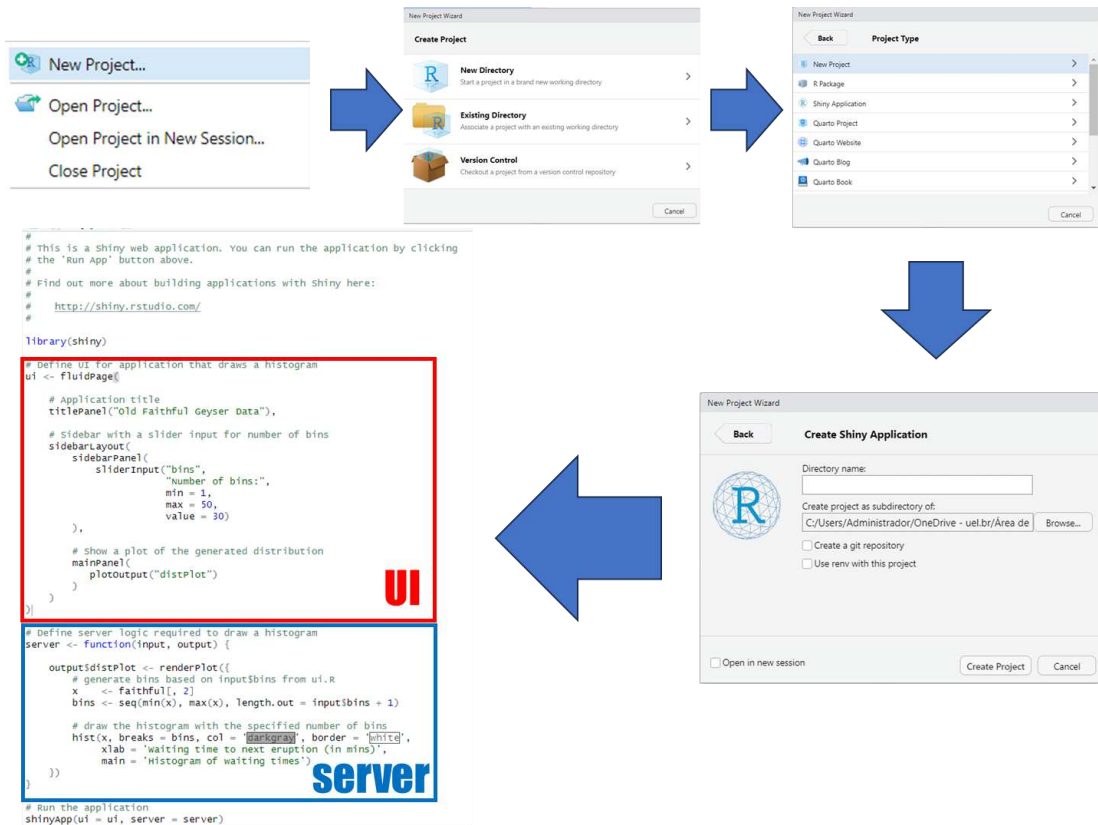


Figura 17 – Processo de criação de um projeto de aplicativo *shiny* usando o IDE Rstudio e o pacote *shiny*. Fonte: o próprio autor.

Um aplicativo shiny é composto basicamente de uma interface de usuário (*ui*) e uma função que realiza a tarefa solicitada e retorna ao usuário (*server*). Compreender as informações de entrada e saída são importantes no desenvolvimento do aplicativo *shiny*. As informações de entradas controladas (*widgets*) podem ser variadas, como informações textuais (*textInput*), numéricas (*numericInput*), selecionáveis (*selectInput*), entre outros (Figura 18). As informações de saída também devem ser especificadas na interface de usuário, pois é necessário informar o tipo de saída desejado: gráfico (*plotOutput*), texto (*verbatimTextOutput*), tabela (*dataTableOutput*), entre outros. Tanto em ambientes de entrada (argumento *inputId*), quanto de saídas (argumento *outputId*), é necessário adicionar um identificador (Tabela 2).

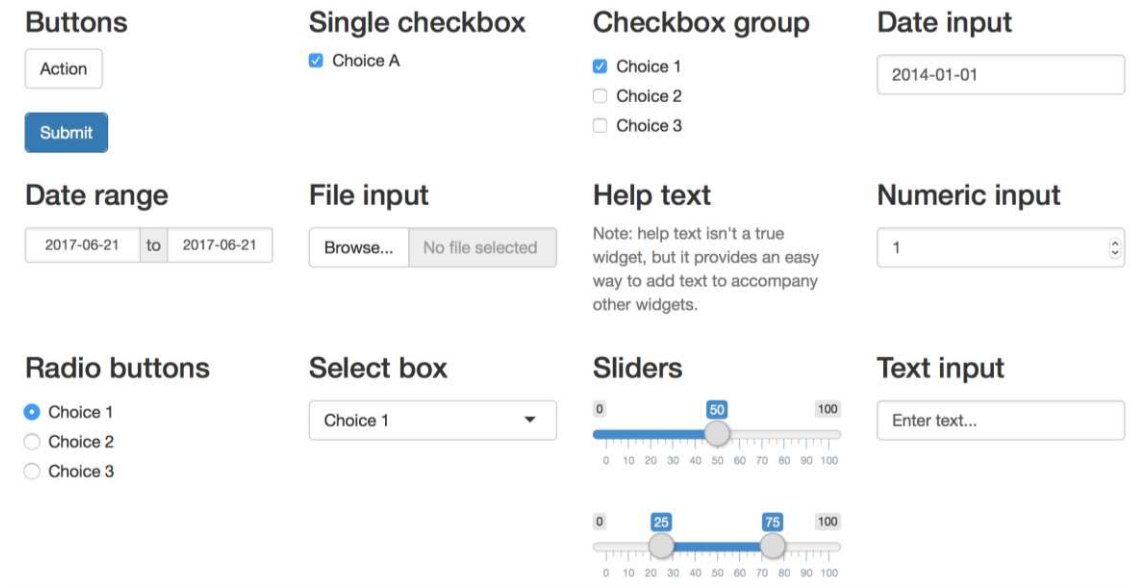


Figura 18 – principais tipos de entrada de controle de um aplicativo shiny. **Fonte:** <https://shiny.posit.co/>

Tabela 2 – Principais funções de saída, tipo de saída e função de renderizar de um aplicativo shiny.

Função de saída	Tipo de saída	Função de renderizar
<i>dataTableOutput</i>	DataTable	<i>renderDataTable</i>
<i>htmlOutput</i>	Converte HTML	<i>renderUI</i>
<i>imageOutput</i>	Imagem	<i>renderImage</i>
<i>plotOutput</i>	Gráfico	<i>renderPlot</i>
<i>tableOutput</i>	Tabela	<i>renderTable</i>
<i>textOutput</i>	Texto	<i>renderText</i>
<i>uiOutput</i>	Converte HTML	<i>renderUI</i>
<i>verbatimTextOutput</i>	Texto	<i>renderPrint</i>

Uma vez construído o aplicativo *shiny*, para sua execução, poderá executar em “Run App” e será retornado localmente ao usuário. O aplicativo também pode ser executado localmente em um navegador. Contudo, caso o desenvolvedor queira disponibilizar o aplicativo via *web*, poderá seguir três caminhos: i) criar um pacote R com as funções do aplicativo *shiny*; ii) disponibilizar o aplicativo através do servidor do próprio *shiny* (<https://www.shinyapps.io/>); iii) instalar um servidor particular (necessário instalar o shiny server).

5 REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, C. O. **Ensaio: da amostra ao teorema do limite central: Um pouco dos fundamentos e uma aplicação prática**. Embrapa Mandioca e Fruticultura-Documentos (INFOTECA-E), 2019.

ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. **Ann Math Statist**, v. 23, n. 8, p. 193-212, 1952. doi: 10.1214/aoms/1177729437

ANDRADE, J. A.; CADIMA, J.; ABREU, F. M. Modeling germination rate and cardinal temperatures of seven mediterranean crops. **Journal of Crop Improvement**, v. 32, n. 6, p. 878-902, 2018. doi: 10.1080/15427528.2018.1542362

ARCHONTOULIS, S. V.; MIGUEZ, F. E. Nonlinear regression models and applications in agricultural research. **Agronomy Journal**, v. 107, n. 2, p. 786-798, 2015. doi: 10.2134/agronj2012.0506

ARNHOLD, E. Package in the R environment for analysis of variance and complementary analyses. **Brazilian Journal of Veterinary Research and Animal Science**, v. 50, n. 6, p. 488-492, 2013. doi: 10.11606/issn.1678-4456.v50i6p488-492

ATABAKI, Z. M.; GHEREKHLOO, J.; GHADERI-FAR, F.; ANSARI, O.; HASSANPOUR-BOURKHEILI, S. Investigating the effects of temperature on seed germination of cutleaf geranium (*Geranium dissectum* L.) and determination of its cardinal temperatures. **Phytoparasitica**, v. 49, n. 1, p. 143-152, 2021. doi: 10.1007/s12600-020-00865-w

BAILAR, J. Science, statistics, deception. **Annals of Internal Medicine**, v. 104, s.1, p. 259-260, 1986. doi: 10.7326/0003-4819-104-2-259

BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola**. 4ed. Jaboticabal: FUNEP. 237p, 2013.

BARBIN, D. **Planejamento e análise estatística de experimentos agrônômicos**, 2ed., Editora Mecenas. 2013.

BARTLETT, M. S. Properties of sufficiency and statistical tests. **Proceedings of the Royal Statistical Society, Series A**, v. 160, p. 268-282, 1937.

- BERRY, D. A. Logarithmic transformations in ANOVA. **Biometrics**, v. 43, n. 2, p. 439-456, 1987. doi: 10.2307/2531826
- BORGES, L. C.; FERREIRA, D. F. Poder e taxas de erro tipo I dos testes Scott-Knott, Tukey e Student-Newman-Keuls sob distribuições normal e não normais dos resíduos. **Revista de matemática e estatística**, v. 21, n. 1, p. 67-83, 2003.
- BOX, G. P.; COX, D. R. J. Na Analysis of Transformations. **Journal of the Royal Statistical Society**, 26, 211-243, 1964. doi: 10.1111/j.2517-6161.1964.tb00553.x
- BRETZ, F.; HOTHORN, T.; WESTFALL, P. (2011) **Multiple comparisons using R**. Boca Raton, FL: CRC Press. doi: 10.1201/9781420010909
- BRIEN, C. (2021) **Dae: Functions Useful in the Design and ANOVA of Experiments**. R Package Version 3.2-13. Disponível em: <http://CRAN.R-project.org/package=dae> Acesso em: 14 mar. 2024.
- BUBANS, V. E.; CARGNELUTTI FILHO, A.; NEU, I. M.; SILVEIRA, D. L.; SOMAVILLA, F. M.; OSMARI, B. F. Plot size and number of repetitions in forage wheat. **Revista Caatinga**, v. 36, p. 234-242, 2023. doi: 10.1590/1983-21252023v36n125rc
- BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: understanding AIC and BIC in model selection. **Sociological methods & research**, v. 33, n. 2, P. 261-304, 2004. doi: 10.1177/0049124104268644
- CABRERA-SANTOS, D.; ORDOÑEZ-SALANUEVA, C. A.; SAMPAYO-MALDONADO, S.; CAMPOS, J. E.; OROZCO-SEGOVIA, A.; FLORES-ORTIZ, C. M. Quantifying Cardinal Temperatures of Chia (*Salvia hispanica* L.) Using Non-Linear Regression Models. **Plants**, v. 11, n. 9, p. 1-18, 2022. doi: 10.3390/plants11091142
- CALENGE, C. The package “adehabitat” for the R software: a tool for the analysis of space and habitat use by animals. **Ecological modelling**, v. 197, n. 3-4, p. 516-519, 2006. doi: 10.1016/j.ecolmodel.2006.03.017
- CANTERI, M. G.; ALTHAUS, R. A.; VIRGENS FILHO, J. S.; GIGLIOTI, E. A.; GODOY, C. V. **SASM-AGRI-Sistema para análise e separação de médias em experimentos agrícolas pelos métodos Scott-Knott, Tukey e Duncan**. Embrapa Soja-Artigo em periódico indexado (ALICE), 2001.

- CARGNELUTTI FILHO, A.; LOREGIAN, M. V.; BUBANS, V. E.; SOMAVILLA, F.; COSTA, S. L. D. Comparison of methods for estimating the optimum plot size for pearl millet, slender leaf rattlebox, and showy rattlebox. **Revista Caatinga**, v. 34, p. 249-256, 2021. doi: 10.1590/1983-21252021v34n201rc
- CASELLA, G.; BERGER, R. L. Asymptotic evaluations. **Statistical inference**, v. 2, p. 467-515, 2002.
- CHICCO, D.; WARRENS, M. J.; JURMAN, G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. **PeerJ Computer Science**, v. 7, p. e623, 2021. doi: 10.7717/peerj-cs.623
- COCHRAN, W. G.; COX, G. M. **Experimental design**. 2ed. New York: John Wiley. 611p, 1986.
- COUTO, M. R. M.; LÚCIO, A. D.; LOPES, S. J.; CARPES, R. H. Transformação de dados em experimentos com abobrinha italiana em ambiente protegido. **Ciência Rural**, v. 39, p. 1701-1707, 2009. doi: 10.1590/S0103-84782009005000110
- CRAMÉR, H. On the Composition of Elementary Errors. **Scandinavian Actuarial Journal**, n. 1, p. 13-74, 1928. doi: 10.1080/03461238.1928.10416862
- CRUZ, C. D. GENES: *software* para análise de dados em estatística experimental e em genética quantitativa. **Acta Scientiarum-Agronomy**, v. 35, p. 271-276, 2013. doi: 10.4025/actasciagron.v35i3.21251
- CRUZ, C. D. **Programa Genes - Análise multivariada e simulação**. 1st ed. Viçosa: UFV, 2006c.
- CRUZ, C. D. **Programa Genes - Biometria**. 1st ed. Viçosa: UFV, 2006b.
- CRUZ, C. D. **Programa Genes - Diversidade Genética**. 1st ed. Viçosa: UFV, 2008.
- CRUZ, C. D. **Programa Genes - Estatística Experimental e Matrizes**. 1st ed. Viçosa: UFV, 2006a.
- CUSTÓDIO, T. N.; BARBIN, D. Modelos de predição para sobrevivência de plantas de *Eucalyptus grandis*. **Ciência e Agrotecnologia**, v. 33, p. 1948-1952, 2009. doi: 10.1590/S1413-70542009000700043

- DELL'AQUILA, A. New perspectives for seed germination testing through digital imaging technology. **The Open Agriculture Journal**, v. 3, n. 3, p. 37-42, 2009. doi: 10.2174/1874331500903010037
- DUNCAN, D. B. Multiple range and multiple F tests. **Biometrics**, v. 11, n. 1, p. 1-42, 1955. doi: 10.2307/3001478
- DUNN, O. J. Multiple comparisons among means. **Journal of the American Statistical Association**, v. 56, p. 52-64, 1961. doi: 10.2307/2282330
- DUNNETT C. W. New tables for multiple comparisons with a control, **Biometrics**, v. 20, p. 482-491, 1964. doi: 10.2307/2528490
- DURBIN, J.; WATSON, G. S. Testing for Serial Correlation in Least Squares Regression, I. **Biometrika**, v. 37, n. 3-4, p. 409-428, 1950. doi: 10.1093/biomet/37.3-4.409
- EDMOND, J. B.; DRAPALA, W. J. The effects of temperature, sand and soil, and acetone on germination of okra seed. **Proceedings of the American Society for Horticultural Science**, v. 71, p. 428-434, 1958.
- FARIA, G. A.; LOPES, B. G.; PEIXOTO, A. P. B.; FERREIRA, A. F. A.; MALTONI, K. L.; PIGARI, L. B. Experimental plot size of passion fruit. **Revista Brasileira de Fruticultura**, v. 42, p. e-125, 2020. doi: 10.1590/0100-29452020125
- FERREIRA, D. F. Análises estatísticas por meio do Sisvar para Windows versão 4.0. **Reunião anual da região brasileira da sociedade internacional de biometria**, v. 45, n. 2000, p. 235, 2000.
- FERREIRA, D. F. SISVAR: A computer analysis system to fixed effects split plot type designs: Sisvar. **Revista brasileira de biometria**, v. 37, n. 4, p. 529-535, 2019. doi: 10.28951/rbb.v37i4.450
- FERREIRA, D. F. Sisvar: a computer statistical analysis system. **Ciência e agrotecnologia**, v. 35, p. 1039-1042, 2011. doi: 10.1590/S1413-70542011000600001
- FERREIRA, D. F. Sisvar: a Guide for its Bootstrap procedures in multiple comparisons. **Ciência e agrotecnologia**, v. 38, p. 109-112, 2014. doi: 10.1590/S1413-70542014000200001
- FERREIRA, D. F. SISVAR: um programa para análises e ensino de estatística. In: **Revista symposium**, p. 36-41, 2008.

FERREIRA, E. B.; CAVALCANTI, P. P.; NOGUEIRA, D. A. ExpDes: an R package for ANOVA and experimental designs. **Applied Mathematics**, v. 5, n. 19, p. 2952-2958, 2014. doi: 10.4236/am.2014.519280

FERREIRA, P. V. **Estatística experimental aplicada às ciências agrárias**. Viçosa, MG, UFV, 588p, 2018.

FISHER, R.A. **The Design of Experiments**. Oliver & Boyd, Edinburgh, 1935.

HARRIS, R. L. **Information graphics: A comprehensive illustrated reference**. Oxford University Press, USA, 1999.

HARTLEY, H. O. The maximum F-ratio as a short cut test for homogeneity of variance, **Biometrika**, v. 37, p. 308-312, 1950.

HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments**. New York: John Wiley. v. 1, 495p, 1994.

HOFFMASTER, A. L.; XU, L.; FUJIMURA, K.; MCDONALD, M. B.; BENNET, M. A.; EVANS, A. F. The OSU seed vigor imaging system (SVIS) for soybean and corn seedlings. **Seed Technology**, v. 27, n. 1, p. 7-24, 2005

HORNIK, K.; LEISCH, F. **Vienna and R: Love, marriage and the future**. na, 2002.

HOTHORN, T.; BRETZ, F.; WESTFALL, P. Simultaneous Inference in General Parametric Models. **Biometrical Journal**, v. 50, n. 3, p. 346-363, 2008. doi: 10.1002/bimj.200810425

HSU, J. C. **Multiple comparisons**. Theory and methods. London, UK: Chapman & Hall, 1996. doi: 10.1007/978-1-4899-7180-7

HURLBERT, S. H. Pseudoreplication and the design of ecological field experiments. **Ecological monographs**, v. 54, n. 2, p. 187-211, 1984. doi: 10.2307/1942661

IHAKA, R.; GENTLEMAN, R. R: a language for data analysis and graphics. **Journal of computational and graphical statistics**, v. 5, n. 3, p. 299-314, 1996. doi: 10.1080/10618600.1996.10474713

JIANG, Z.; CAPPELLERI, J. C.; GAMALO, M.; CHEN, Y.; THOMAS, N.; CHU, H. A comprehensive review and shiny application on the matching-adjusted indirect comparison. **Research Synthesis Methods**, n.1, p. 1-16, 2024. doi: 10.1002/jrsm.1709

JOOSEN, R. V. L.; KODDE, J.; WILLEMS, L. A. J.; LIGTERINK, W.; VAN DER PLAS, L. H. W.; HILHORST, H. W. M. Germinator: A software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. **The Plant Journal**, v. 62, p. 148-159, 2010. doi: 10.1111/j.1365-313x.2009.04116.x

KEULS, M. The use of the "studentized range" in connection with an analysis of variance. **Euphytica**, v. 1, n. 2, p. 112-122, 1952. doi:10.1007/bf01908269.

KIRK, R. E. **Experimental Design: Procedures for the Behavioral Sciences**. 3rd Edition, Brooks/Cole Publishing Company, New York, 1995.

KIRKUP, L. **Experimental Methods for Science and Engineering Students An Introduction to the Analysis and Presentation of Data**, p. 24-58, 2019. doi: 10.1017/9781108290104.005

KNIEF, U.; FORSTMEIER, W. Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, v. 53, n. 6, p. 2576-2590, 2021. doi: 10.3758/s13428-021-01587-5

LEVENE, H. **Robust tests for equality of variances**. In: Ingram Olkin; Harold Hotelling. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. [S.l.]: Stanford University Press. pp. 278-292, 1960.

LILLIEFORS, H. W. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. **Journal of the American Statistical Association**, v. 62, n. 318, p. 399-402, 1967. doi: 10.1080/01621459.1967.10482916

LOZANO-ISLA, F.; BENITES-ALFARO, O. E.; POMPELLI, M. F. GerminaR: An R package for germination analysis with the interactive web application "GerminaQuant for R." **Ecological Research**, v. 34, n. 2, p. 339-346, 2019. doi: 0.1111/1440-1703.1275

LÚCIO, A. D.; COUTO, M. R. M.; LOPES, S. J.; STORCK L. Transformação Box-Cox em experimentos com pimentão em ambiente protegido. **Horticultura Brasileira**, v. 29, n. 1, p. 38-42, 2011. doi: 10.1590/S0102-05362011000100007

LÚCIO, A. D.; COUTO, M. R. M.; TREVISAN, J. N.; MARTINS, G. A. K.; LOPES, S. J. Excesso de zeros nas variáveis observadas: estudo de caso em experimento com brócolis. **Bragantia**, v. 69, p. 1035-1046, 2010. 10.1590/S0006-87052010000400032

LUCIO, A. D.; SARI, B. G. Planning and implementing experiments and analyzing experimental data in vegetable crops: problems and solutions. **Horticultura Brasileira**, v. 35, p. 316-327, 2017. doi: 10.1590/S0102-053620170302

LÚCIO, A. D.; SCHWERTNER, D. V.; HAESBAERT, F. M.; SANTOS, D.; BRUNES, R. R.; RIBEIRO, A. L.; LOPES, S. J. Violação dos pressupostos do modelo matemático e transformação de dados. **Horticultura Brasileira**, v. 30, n. 3, p. 415-423, 2012. doi: 10.1590/S0102-05362012000300010

MAGUIRE, J. D. Speed of germination-aid in selection and evaluation for seedling emergence and vigor. **Crop Science**, v. 2, n. 1, p. 176-177, 1962. doi: 10.2135/cropsci1962.0011183X000200020033x

MARQUES, F. R. F.; MEIADO, M. V.; CASTRO, N. M. C. R. D.; CAMPOS, M. L. D. O.; MENDES, K. R.; SANTOS, O. D. O. D.; POMPELLI, M. F. GerminaQuant: a new tool for germination measurements. **Journal of Seed Science**, v. 37, p. 248-255, 2015. doi: 10.1590/2317-1545v37n3145605

MARTIN, T. N.; STORCK, L. **Análise das pressuposições do modelo matemático em experimentos agrícolas no delineamento blocos ao acaso**. In: MARTIN, T. N.; ZIECH, M. F. (org). Sistemas de Produção Agropecuária. Curitiba: UTFPR, p. 177-196, 2008.

MELO, R. C.; TREVISANI, N.; SANTOS, M.; GUIDOLIN, A. F.; COIMBRA, J. L. M. Statistical model assumptions achieved by linear models: classics and generalized mixed. **Revista Ciência Agronômica**, v. 51, n. 1, p. 1-9, 2020. doi: 10.5935/1806-6690.20200015

MENDIBURU, F. 2021. **agricolae: Statistical procedures for agricultural research**. R package Version 1.2-5. R Found. Stat. Comput., Vienna. Disponível em: <http://CRAN.R-project.org/package=agricolae>. Acesso em: 14 mar. 2024.

MIRANDA, R. Q.; CORREIA, R. M.; ALMEIDA-CORTEZ, J. S.; POMPELLI, M. F. Germinação de *Prosopis juliflora* (Sw.) D.C sementes em diferentes potenciais e temperaturas osmóticas. **Biologia de Espécies Vegetais**, v. 29, p. e9-e20, 2014. doi: 10.1111/1442-1984.12025

MONTGOMERY, J.; MOORE, R. T. Building and Maintaining R Packages with devtools and roxygen2. **Political Methodologist**, v. 22, n. 1, p. 1-27, 2014.

NAKAGAWA, J. Testes de vigor baseados no desempenho das plântulas. In: KRZYZANOSKI, F. C.; VIEIRA, R. D.; FRANÇA NETO, J. B. (Ed.). **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, p. 2.1-2.24, 1999.

NEWMAN, D. (1939). "The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation". **Biometrika**, v. 31, n. 1, p. 20-30. doi: 10.1093/biomet/31.1-2.20

NUNES, C. A.; ALVARENGA, V. O.; SOUZA SANT'ANA, A.; SANTOS, J. S.; GRANATO, D. The use of statistical software in food science and technology: Advantages, limitations and misuses. **Food Research International**, v. 75, p. 270-280, 2015. doi: 10.1016/j.foodres.2015.06.011

PIEPHO, H. P.; EDMONDSON, R. N. A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels. **Journal of Agronomy and Crop Science**, v. 204, n. 5, p. 429-455, 2018. doi: 10.1111/jac.12267

PIMENTEL-GOMES, F. **Curso de estatística experimental**. 15ed. Piracicaba: Nobel. 451p, 2009.

POWELL, S. G.; BAKER, K. R.; LAWSON, B. Errors in operational spreadsheets. **Journal of Organizational and End User Computing**, v. 21, n. 3, p. 24-36, 2009. doi: 10.4018/joeuc.2009070102.

QUINN, G. P.; KEOUGH, M. J. **Experimental Design and Data Analysis for Biologists**. Cambridge University Press, Cambridge, 2002. doi: 10.1017/CBO9780511806384

R CORE TEAM (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.R-project.org/>. Acesso em: 14 mar. 2024.

RANAL, M. A.; SANTANA, D. G.; FERREIRA, W. R.; MENDES-RODRIGUES, C. Calcular medidas de germinação e organizar planilhas. **Brazilian Journal of Botany**, v. 32, p. 849-855, 2009. doi: 10.1590/s0100-84042009000400022

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético**, Colombo: Embrapa Florestas. 561p, 2007.

- RIBEIRO-OLIVEIRA, J. P.; SANTANA, D. G. D.; PEREIRA, V. J.; SANTOS, C. M. D. Data transformation: an underestimated tool by inappropriate use. **Acta Scientiarum Agronomy**, v. 40, e35300, 2018. doi: 10.4025/actasciagron.v40i1.35300
- SAKO, Y.; MCDONALD, M. B.; FUJIMURA, K.; EVANS, A. F.; BENNET, M. A. A system of automated seed vigour assessment. **Seed Science & Technology**, v. 29, n. 3, p. 625-636, 2001.
- SCHEFFE, H. **The Analysis of Variance**. Wiley, New York, v.72, 1959.
- SCHERVISH, M. J. P values: what they are and what they are not. **The American Statistician**, v. 50, n. 3, p. 203-206, 1996. doi: 10.1080/00031305.1996.10474380
- SCOTT, A.; KNOTT, M. Cluster-analysis method for grouping means in analysis of variance. **Biometrics**, v. 30, n. 3, p. 507-512, 1974. doi: 10.2307/2529204
- SCOTT, S. J.; JONES, R. A.; WILLIAMS, W. Review of data analysis methods for seed germination. **Crop science**, v. 24, n. 6, p. 1192-1199, 1984. doi: 10.2135/cropsci1984.0011183X002400060043x
- SHAPIRO, S. S.; FRANCA, R. S. An Approximate Analysis of Variance Test for Normality. **Journal of the American Statistical Association**, v. 67, n. 337, p. 215-216, doi: 1972. 10.2307/2284728
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3-4, p. 591-611, 1965. doi: 10.1093/biomet/52.3-4.591
- SILVA, F. D. A. S.; AZEVEDO, C. A. V. The Assistat Software Version 7.7 and its use in the analysis of experimental data. **African Journal of Agricultural Research**, v. 11, n. 39, p. 3733-3740, 2016. doi: 10.5897/AJAR2016.11522
- SILVA, L. J. D.; MEDEIROS, A. D. D.; OLIVEIRA, A. M. S. SeedCalc, a new automated R software tool for germination and seedling length data processing. **Journal of Seed Science**, v. 41, p. 250-257, 2019. doi: 10.1590/2317-1545v42n2217267
- SINÍCIO, R.; BHERING, M. C.; VIDIGAL, D. S.; DIAS, D. C. F. S. Validação do aplicativo computacional SeedSolve para previsão das perdas de germinação e vigor de sementes armazenadas. **Revista Brasileira de Sementes**, v. 31, n. 2, p. 9-18, 2009. doi: 10.1590/S0101-31222009000200001

SOUZA, R. A.; HUNGRIA, M.; FRANCHINI, J.; CHUEIRE, L. M. O.; BARCELLOS, F. G.; CAMPO, R. J. Avaliação qualitativa e quantitativa da microbiota do solo e da fixação biológica do nitrogênio pela soja. **Pesquisa Agropecuária Brasileira**, v. 43, n. 1, p. 71-82, 2008. doi: 10.1590/S0100-204X2008000100010

STATSOFT, Inc. (2004). **STATISTICA (data analysis software system)**, version 7. www.statsoft.com

STELL, R. G. D.; TORRIE, J. H. **Principles and Procedures of Statistics: A Biometrical Approach**. McGraw-Hill, New York, 633p, 1996.

STORCK, L.; LOPES, S. J.; ESTEFANEL, V.; GARCIA, D.C. **Experimentação Vegetal**. 3Ed. Santa Maria: Editora UFSM. 198p., 2016.

STORCK, L.; LÚCIO, A. D. C.; KRAUSE, W.; ARAÚJO, D. V. D.; SILVA, C. A. Scaling the number of plants per plot and number of plots per genotype of yellow passion fruit plants. **Acta Scientiarum. Agronomy**, v. 36, p. 73-78, 2014. doi: 10.4025/actasciagron.v36i1.17697

TARTAGLIA, F. D. L.; LUCIO, A. D. C.; DIEHL, M. I.; TISCHLER, A. L.; KRYSCZUN, D. K.; ZEMOLIN, J. A.; MARQUES, L. E. Experimental plan for tests with pea. **Agronomy Journal**, v. 113, n. 2, p. 1394-1406, 2021. doi: 10.1002/agj2.20575

TUKEY, J. Comparing Individual Means in the Analysis of Variance. **Biometrics**, v. 5, n. 2, p. 99-114, 1949.

VAZQUEZ, A. I.; BATES, D. M.; ROSA, G. J. M.; GIANOLA, D.; WEIGEL, K. A. An R package for fitting generalized linear mixed models in animal breeding. **Journal of animal science**, v. 88, n. 2, p. 497-504, 2010. doi: 10.2527/jas.2009-1952

VENABLES, W. N., SMITH, D.M.; R CORE TEAM. **An Introduction to R**, Notes on R: A Programming Environment for Data Analysis and Graphics Version 4.3.3. Disponível em: <https://cran.r-project.org/doc/manuals/R-intro.pdf>. Acesso em: 19 mar. 2024.

VIDONI, M.; CODABUX, Z. Towards a taxonomy of Roxygen documentation in R packages. **Empirical Software Engineering**, v. 28, n. 4, p. 106, 2023. doi: 10.1007/s10664-023-10345-4

WAGNER, M. H.; DEMILLY, D.; DUCOURNAU, S.; DÜRR, C.; LÉCHAPPÉ, J. Computer vision for monitoring seed germination from dry state to young seedlings. **Seed Science**, v. 142, n. 1, p. 49-51, 2011.

WARTON, D. I.; HUI, F. K. C. The arcsine is asinine: the analysis of proportions in ecology. *Ecology*, v. 92, n. 1, p. 3-10, 2011.

WICKHAM, H.; DANENBERG, P.; CSÁRDI, G.; EUGSTER, M. (2022b). **roxygen2: In-Line Documentation for R**. R package version 7.2.3, <<https://CRAN.R-project.org/package=roxygen2>>. Acesso em: 12 de março de 2024.

WICKHAM, H.; HESTER, J.; CHANG, W.; BRYAN, J. (2022a). **devtools: Tools to Make Developing R Packages Easier**. R package version 2.4.5. Disponível em: <<https://CRAN.R-project.org/package=devtools>>. Acesso em: 12 de março de 2024.

XU, W.; LI, W.; SONG, D. Testing normality in mixed models using a transformation method. *Stat Papers*, v. 84, p. 54-71, 2013. doi: 10.1007/s00362-011-0411-4

YASSIN, N. **Análise de experimentos fatoriais de dois fatores com tratamentos adicionais**. 2001, 118p, Dissertação de Mestrado, Universidade Federal de Lavras, 2010.

YATES, F. **The design and analysis of factorial experiments**. Imperial Bureau of Soil Science Technical Communication, 35p, 1937.

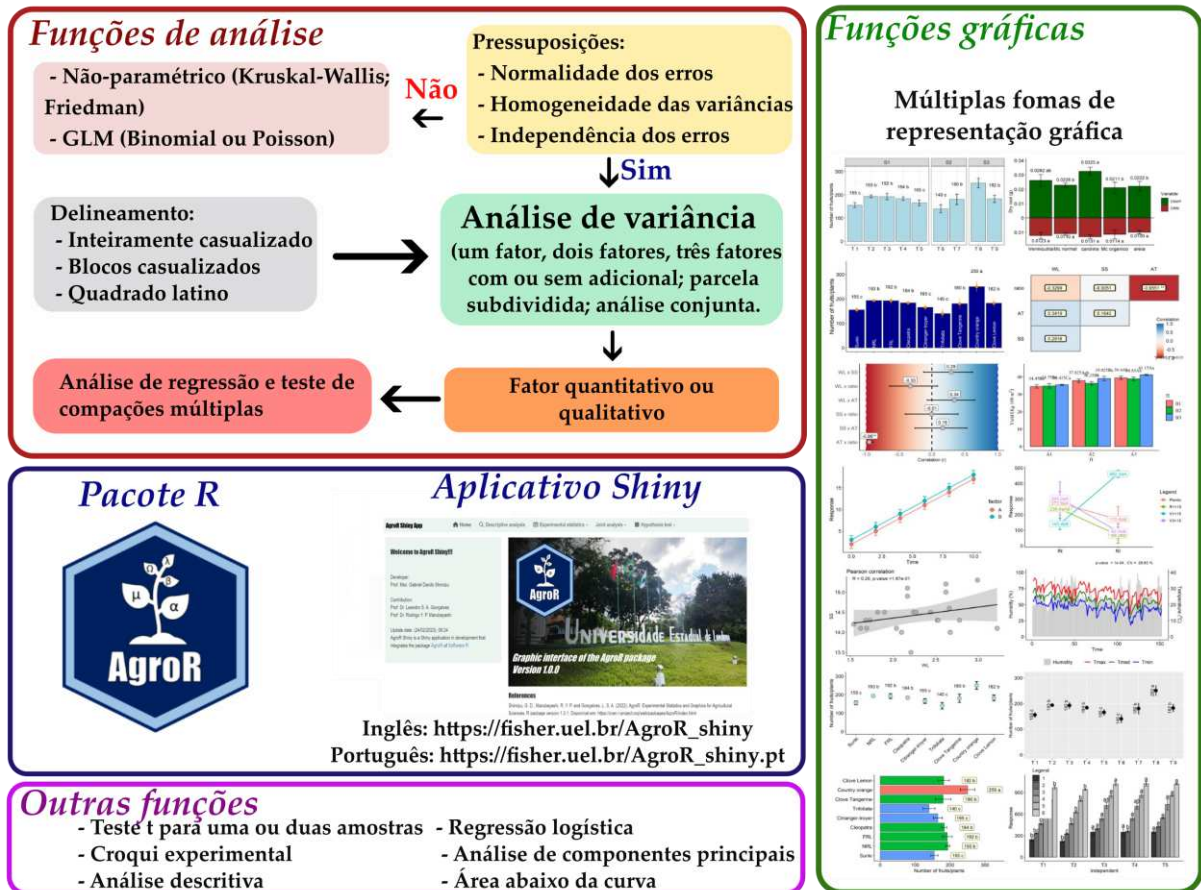
ZIMMERMANN, F. J. P. **Estatística aplicada à pesquisa agrícola**. Santo Antônio de Goiás: Embrapa. 402p, 2004.

ZUBIN, J. Note on a transformation function for proportions and percentages. *Journal of Applied Psychology*, v. 19, n. 2, p. 213-220, 1935.

Artigo A

AgroR: Um pacote R e uma interface Shiny para análise de experimentos agrícolas

Título curto: Um pacote e aplicativo R para experimentos agrícolas



Resumo gráfico: Resumo das principais funções e resultados que o pacote e aplicativo shiny AgroR realizam para dados de experimentos agrícolas.

RESUMO

O *software* R tem sido amplamente usado na pesquisa. Contudo, por ser uma linguagem de programação, a maioria dos pesquisadores tem dificuldade em estruturar os comandos necessários para realizar as análises. Com o objetivo de superar essas dificuldades, desenvolvemos o *AgroR*, um pacote em linguagem R de fácil uso e sua versão em forma de aplicativo *shiny* para análise de dados oriundos de experimentos agrícolas. O *AgroR* aborda os principais projetos experimentais e oferece diversas funções para análise exploratória de dados, análises inferenciais, representações gráficas e algumas funções auxiliares, bem como um vasto

conjunto de dados. O pacote *AgroR* 1.3.5 está disponível no repositório CRAN e pode ser instalado pelo comando: `install_packages("AgroR")`, enquanto o aplicativo *shiny* pode ser acessado pelo url: https://uel.br/fisher.uel.br/AgroR_shiny. Espera-se que o *AgroR* atenda à demanda por um pacote de fácil uso e que apresente diversas informações ao usuário.

Palavras-chave: Estatística experimental; *software* R; agricultura.

INTRODUÇÃO

A estatística experimental é amplamente utilizada nas ciências agrárias e tem por objetivo o estudo dos experimentos, desde o planejamento, execução, análise dos dados e interpretação dos resultados. Apesar de ser relativamente antigo e bastante conhecido, datado a origem com Ronald A. Fisher (1926), existe atualmente uma lacuna entre o conhecimento estatístico e sua aplicabilidade na rotina dos pesquisadores. A grande maioria das análises da estatística experimental é realizada por meio da análise da variância, o qual realiza os desdobramentos das variâncias em função de fontes de variação conhecidos ou não. No entanto, para que seja válido tal análise, é necessária que quatro pressuposições sejam atendidas: *i*) a normalidade dos erros, *ii*) homogeneidade das variâncias, *iii*) independência dos erros e *iv*) aditividade dos fatores (KNIEF; FORSTMEIER, 2021). Nesse sentido, muitas vezes o pesquisador ignora tais pressupostos, em virtude da falta de conhecimento ou mesmo por acreditar não ser importante (BAILAR, 1986). Quando as pressuposições são violadas, a confiabilidade de todos os testes paramétricos fica comprometidas, pois ocorrem alterações na probabilidade de ocorrência do erro tipo I e/ou tipo II, podendo levar a falsas conclusões a respeito dos efeitos de tratamento (LÚCIO *et al.*, 2012; MARTIN; STORCK, 2008; MELO *et al.*, 2013; XU, LI; SONG, 2013).

Existem inúmeros *softwares* que podem ser utilizados como ferramenta para auxiliar nessas análises (NUNES *et al.*, 2015) e dentre esses, a linguagem R (R CORE TEAM, 2023) tem ganhado força e sendo bastante requerido por revistas de alto impacto, uma vez que apresenta código fonte livre, gratuito e aberto, com um número relativamente grande de usuários e desenvolvedores. Todavia, há grande dificuldade entre os usuários, principalmente para aqueles que utilizam esporadicamente, pois, compreender a estatística, ao mesmo tempo que compreende a linguagem R acaba sendo oneroso, o que desincentiva a comunidade científica.

Aliar a interpretação estatística e o aprendizado do R é um trabalho árduo, ainda que se utilize dos vários tutoriais e vídeos disponíveis na internet. Por outro lado, o próprio R

possui uma série de complementos, denominados *packages*, que podem ser instalados e utilizados em conjunto com a linguagem (ARNHOLD, 2014; KORMANN *et al.*, 2019), facilitando o seu uso, aprimorando processos e simplificando em funções simples e intuitivas. Na versão básica e em pacotes adicionais, o ambiente R fornece funções para realizar análises de variância de tratamentos qualitativos e quantitativos, análises de resíduos e testes de contraste, entre outros. No entanto, as funções disponíveis que executam essas análises são complicadas para usuários inexperientes no ambiente R e em estatística. Isso é particularmente verdadeiro quando as análises envolvem projetos mais complexos (ARNHOLD, 2013). Por outro lado, a partir de 2012, houve a introdução dos chamados aplicativos *shiny* (CHANG *et al.*, 2022), que é um pacote que permite a criação de interface gráficas que facilitam usuários em suas análises. Além disso, por ser possível implementar em um servidor, pode ser acessado por qualquer pessoa por meio de um navegador, desde que tenha acesso a internet, não exigindo, portanto, a instalação em um computador pessoal.

Diante do exposto, o trabalho tem por objetivo apresentar e sumarizar o pacote *AgroR* e seu aplicativo shiny e suas diversas funções para análise de experimentos agrícolas.

MATERIAL E MÉTODOS

DESENVOLVIMENTO DO PACOTE

O pacote foi desenvolvido com *software* R versão 4.2.1 (R CORE TEAM, 2022) e a interface *RStudio* versão 1.2.5001 com uso dos pacotes *roxygen2* (WICKHAM; DANENBERG; EUGSTER, 2017) e *devtools* versão 1.3.2 (WICKHAM; CHANG, 2017). O pacote engloba um compilado de funções oriundas de outros pacotes e da base do R, simplificando-os de forma a facilitar a análise e o entendimento pelo usuário. Para a realização das análises de variância de efeitos fixos, foi utilizado o procedimento *aov()* da base do R (*stats*), enquanto a transformação de dados foi realizada pelo método de Box e Cox (1964), utilizado o procedimento do pacote *MASS* (RIPLEY *et al.*, 2013). O teste de comparação múltipla de Tukey, Duncan e LSD, assim como os testes não-paramétricos de Kruskal-Wallis e Friedman foram implementados tendo como base o pacote *agricolae* (MENDIBURU; SIMON, 2015). O método de agrupamento de Scott-Knott também foi implementado. Os pacotes *lmtest* (ZEILEIS; HOTHORN, 2002), *nortest* (GROSS; LIGGES, 2015) e *car* (FOX *et al.*, 2012) foram usados em alguns testes para pressuposição da anova. Todos os gráficos são gerados a partir do pacote *ggplot2* (WICKHAM; CHANG, 2017).

DESENVOLVIMENTO DO APLICATIVO SHINY

O AgroR App é uma interface gráfica construído em linguagem R 4.2.1 (R CORE TEAM, 2022) que utiliza o pacote *shiny* (CHANG *et al.*, 2022) para a criação de uma aplicação web que realiza o planejamento e a análise de diversos projetos experimentais. Foi construído um ambiente de entrada (ui – user interface) e um ambiente de saída (server) usando o pacote *shiny*.

O aplicativo pode ser acessado através do servidor da Universidade Estadual de Londrina, sob o link (https://fisher.uel.br/AgroR_shiny) ou em sua versão em português (https://fisher.uel.br /AgroR_shiny.pt), cujo registro de proteção intelectual é BR512023003032-5.

RESULTADOS E DISCUSSÃO

PACOTE AGRO R

FUNÇÕES PARA PLANEJAMENTO EXPERIMENTAL

A função *sketch* tem a finalidade de desenvolver um desenho experimental conforme o projeto experimental (Figura 19). Essa função permite a elaboração de croquis de experimentos em delineamento inteiramente casualizado (DIC), blocos casualizados (DBC), quadrado latino (DQL), fatorial duplo e triplo, experimentos em faixas, parcelas subdivididas em DIC e DBC e parcelas subdivididas em DBC. Os argumentos *trat* e *r* são obrigatórios, devendo especificar também *trat1* e *trat2* dependendo do projeto experimental. Defina o argumento *design* para alterar o projeto experimental, sendo o padrão *DIC*. O usuário pode adicionar corredores por meio dos argumentos *add.streets.x* e *add.streets.y*, especificando os vetores com as parcelas que compõem cada rua (Figura 19). Também é possível fornecer o croqui na forma de identificação (ID) e gerar um arquivo em extensão .csv baseado no croqui experimental. Para mais detalhes consulte: *?sketch*.

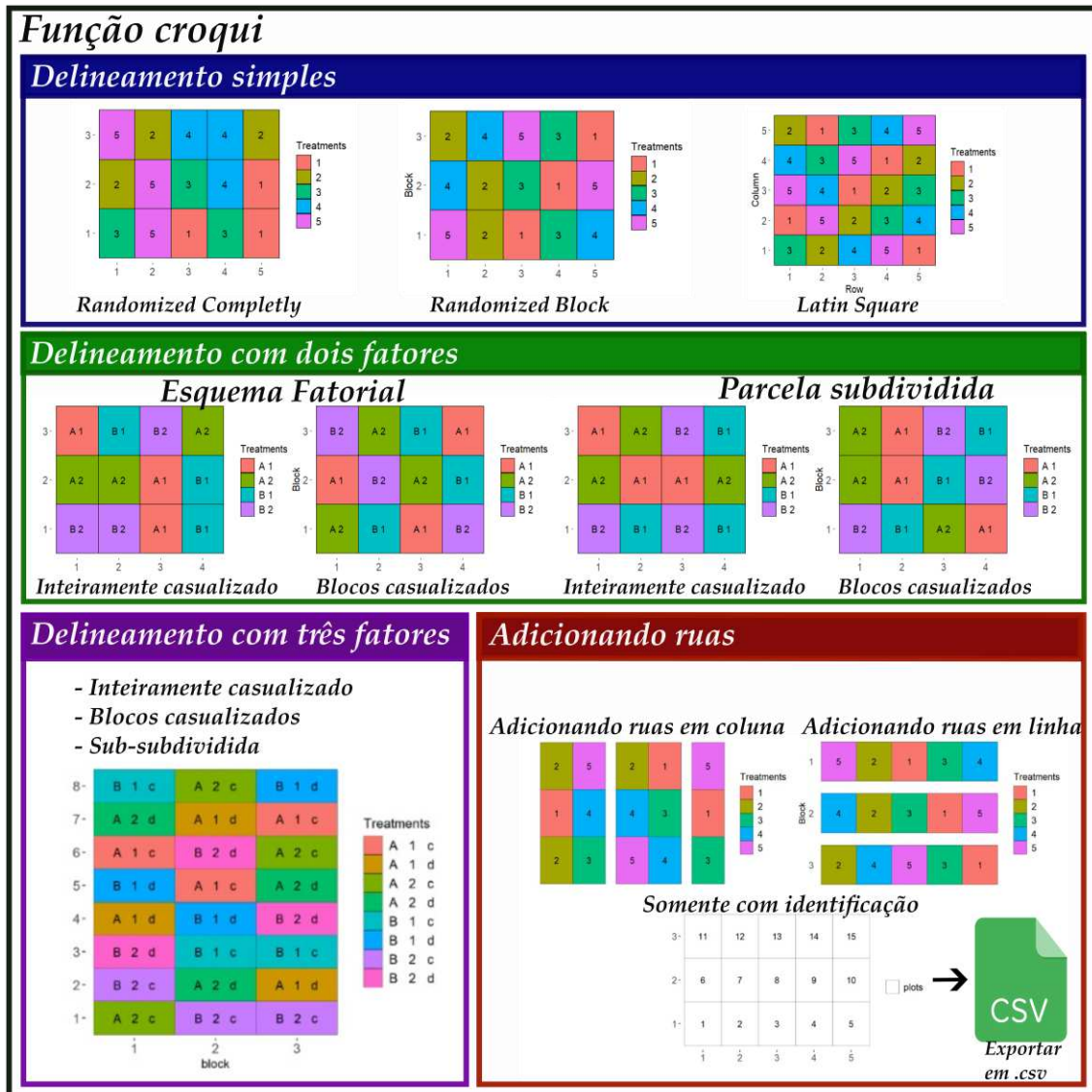


Figura 19 – Exemplo de croquis experimentais usando o pacote e/ou aplicativo *AgroR*.

FUNÇÕES PARA ANÁLISE DE EXPERIMENTOS

Para as análises, foram implementados os experimentos em delineamentos experimentais inteiramente casualizado, blocos casualizados e quadrado latino, com separação por um fator ou analisado por meio de ensaios fatoriais de dois e três fatores com ou sem tratamento adicional, parcelas subdivididas, parcela subsubdividida, parcela subdividida com subparcela em esquema fatorial e, análise conjunta de experimentos (Figura 20). O pacote também apresenta teste de hipóteses para uma média ou duas médias populacionais. Os fatores podem ser quantitativos ou qualitativos, sendo isso definido pelo usuário. As pressuposições da ANOVA podem ser definidos quanto ao teste, sendo testado normalidade dos erros (Shapiro-Wilk [padrão], Lilliefors, Cramer-Von Mises, Shapiro-Francia e Anderson-Darling),

homogeneidade das variâncias (Bartlett [padrão], Hartley e Levene) e independência dos erros (Durbin-Watson).

Sempre que uma pressuposição não é atendida, o pacote retorna um aviso alertando a invalidade do modelo ao mesmo tempo que auxilia nas possíveis soluções. Ainda dentro da função, o usuário pode definir como teste não-paramétrico no caso das funções *DIC* e *DBC*, ou transformação de dados (BOX; COX, 1964) nos demais projetos experimentais. Deve-se ressaltar que quando os dados forem transformados, os gráficos finais são com as médias dos dados originais. Os métodos estatísticos não-paramétricos implementados são o de Kruskal-Wallis no caso de *DIC* e as correções por diferentes métodos (holm, hommell, hochberg, bonferroni, BH, BY, fdr) ou teste de Friedman, no caso do *DBC* pelo método LSD. Também é implementado os testes de Dunn e Dunnett. Alternativamente, para *DIC* e *DBC*, desde que conhecido a natureza da variável, é possível analisar por modelo linear generalizado nas famílias binomial, quasibinomial, poisson ou quasipoisson, sendo os contrastes realizados pelos pacotes *multcomp* (HOTHORN *et al.*, 2016) e *emmeans* (LENTH, 2023).

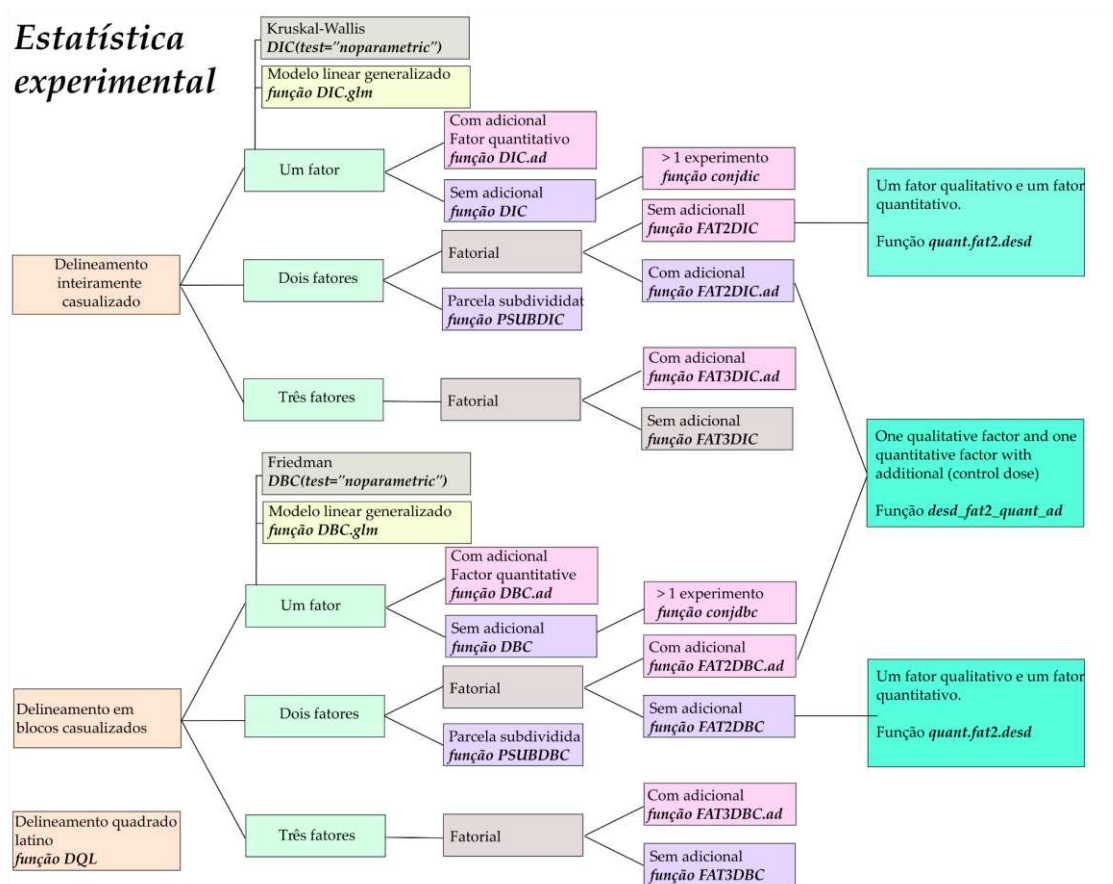


Figura 20 – Diagrama dos projetos experimentais e suas respectivas funções do pacote *AgroR*.

FUNÇÕES GRÁFICAS

A maioria das funções de análise retornam gráficos. Todavia, o pacote apresenta outras opções de representação gráfica, sendo que o usuário deve definir a saída da respectiva análise como um objeto e alocar na função gráfica desejada (Figura 21). O argumento dos parâmetros gráficos em sua maioria, são importados da saída original da respectiva função. Dentre as funções gráficas implementadas inclui, desde gráfico simples, como por exemplo o gráfico de colunas, similar ao retornado pelas funções de análises, até gráficos mais elaborados, como no caso das funções *bar_graph2*, *sk_graph*, *barfacet*, *bargraph_onefactor*, *bargraph_twofactor*, *barplot_positive*, entre outros. O pacote também apresenta funções para análise de correlação entre variáveis, como as funções *plot_cor*, *corgraph* e *cor_ic*.

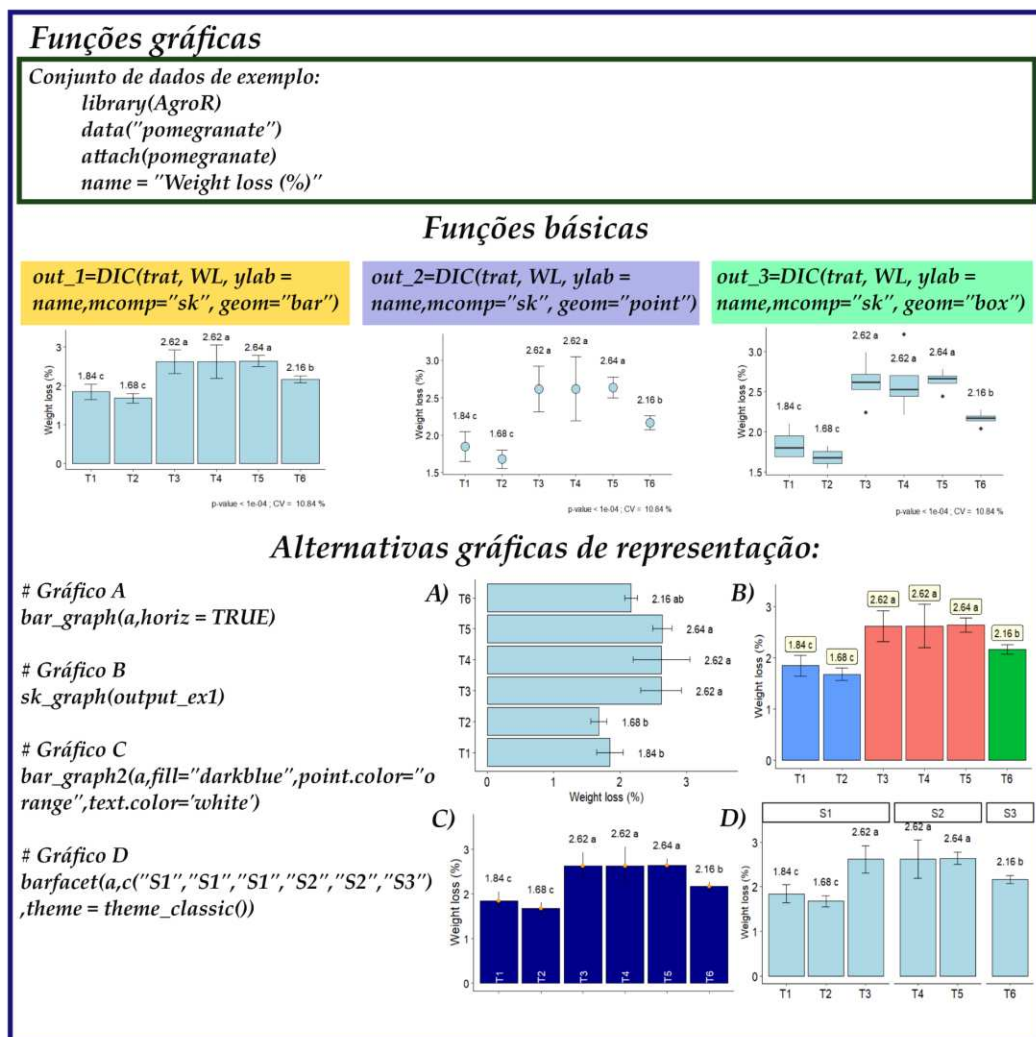


Figura 21 – Exemplo de gráficos gerados pelas funções de análise e pertencentes a classe “graphics”: *bar_graph* (A); *sk_graph* (B); *bar_graph2* (C) *barfacet* (D).

CONJUNTO DE DADOS

O pacote *AgroR* é dedicado a área de ciência agrárias, especialmente para a agronomia. Os 22 conjuntos de dados são oriundos de experimentos da Universidade Estadual de Londrina, de exemplos da literatura (BARBIN, 2013) ou dados fictícios. Esses conjuntos de dados são usadas para exemplificar a utilização de todas as funções implementadas no pacote. Os dados podem ser chamadas utilizando a função *data()*.

OUTRAS FUNÇÕES

Por fim, temos as funções úteis que, embora apresente variadas finalidades, podemos destacar a função de croqui (*sketch*) já citada anteriormente, a função para cálculo da área abaixo da curva de progresso, usada principalmente na área de doenças de plantas (SHANER; FINNEY, 1977) e a função para cálculo de intervalo de confiança. Além dessas funções, de forma a facilitar a construção de tabelas com os resultados, foi criada as funções *summarise_anova* (Figura 22) e *summarise_dunnett*, que permite simplificar os resultados de múltiplas análises em uma única saída para fatores qualitativos e com projetos que englobam até dois fatores de interesse.

Análise individual por variável resposta.

Alocando as saídas das variáveis (Exemplo: *out1*, *out2*, *out3*, *out4*)

	WL	SS	AT	ratio
T1	1.842 b	14.225 b	0.9 a	16.233 a
T2	1.678 b	14.275 b	0.9 a	16.732 a
T3	2.618 a	15 a	0.975 a	15.484 a
T4	2.62 a	14.05 b	1.05 a	13.667 a
T5	2.638 a	14.35 ab	1.075 a	13.508 a
T6	2.163 ab	14.65 ab	1.025 a	14.477 a
CV(%)	10.838	2.219	16.407	18.56
p-value	p<0.001	0.007	0.522	0.485
Transformation	No transf	No transf	No transf	No transf

```
library(AgroR)
data(pomegranate); attach(pomegranate)
out1 = DIC(trat, WL, ylab = "WL")
out2 = DIC(trat, SS, ylab = "SS")
out3 = DIC(trat, AT, ylab = "AT")
out4 = DIC(trat, ratio, ylab = "ratio")
```

Saída do console em R

```
summarise_anova(list(out1, out2, out3, out4), design = "DIC", inf = "p")
```

Relatório com os pacotes Rmarkdown + knitr

summarise_anova(list(*out1*, *out2*, *out3*, *out4*), design = "DIC", divisor = FALSE, inf = "p")

documento word (.docx)

	WL	SS	AT	ratio
T1	1.842 b	14.225 b	0.9 a	16.233 a
T2	1.678 b	14.275 b	0.9 a	16.732 a
T3	2.618 a	15 a	0.975 a	15.484 a
T4	2.62 a	14.05 b	1.05 a	13.667 a
T5	2.638 a	14.35 ab	1.075 a	13.508 a
T6	2.163 ab	14.65 ab	1.025 a	14.477 a
CV(%)	10.838	2.219	16.407	18.56
p-value	p<0.001	0.007	0.522	0.485
Transformation	No transf	No transf	No transf	No transf

documento pdf (.pdf)

	WL	SS	AT	ratio
T1	1.842 b	14.225 b	0.9 a	16.233 a
T2	1.678 b	14.275 b	0.9 a	16.732 a
T3	2.618 a	15 a	0.975 a	15.484 a
T4	2.62 a	14.05 b	1.05 a	13.667 a
T5	2.638 a	14.35 ab	1.075 a	13.508 a
T6	2.163 ab	14.65 ab	1.025 a	14.477 a
CV(%)	10.838	2.219	16.407	18.56
p-value	p<0.001	0.007	0.522	0.485
Transformation	No transf	No transf	No transf	No transf

documento html (.html)

	WL	SS	AT	ratio
T1	1.842 b	14.225 b	0.9 a	16.233 a
T2	1.678 b	14.275 b	0.9 a	16.732 a
T3	2.618 a	15 a	0.975 a	15.484 a
T4	2.62 a	14.05 b	1.05 a	13.667 a
T5	2.638 a	14.35 ab	1.075 a	13.508 a
T6	2.163 ab	14.65 ab	1.025 a	14.477 a
CV(%)	10.838	2.219	16.407	18.56
p-value	p<0.001	0.007	0.522	0.485
Transformation	No transf	No transf	No transf	No transf

Figura 22 – Processo de análise de sumarização usando a função *summarise_anova* e sua associação com o pacote *knitr* e *rmarkdown* para geração de tabelas automáticas.

APLICATIVO SHINY

As funções de análise foram implementadas a partir do pacote AgroR, sendo que os principais projetos experimentais estão incluídos no aplicativo, em que o mesmo é subdividido em quatro seções: Análise descritiva, Estatística experimental, Análise conjunta, teste de hipóteses (Figura 23). Dentro de “Estatística experimental”, há o “Croqui experimental”, “Análise de variância”, “Dunnett”, “Exemplos de tabulação”, “Análise de variância (+1 variável simultaneamente)”, “Não-paramétrico”, enquanto para “Análise conjunta” há “Similaridade de experimentos” e “Análise conjunta” e por fim, para “Teste de hipóteses” há para “Uma média” e “Duas médias”. Todas as análises foram planejadas de tal forma a englobar o maior número possível de situações em que o pesquisador pode encontrar nas ciências agrárias, além de seguir um passo a passo de forma a encaminhar para a interpretação dados.

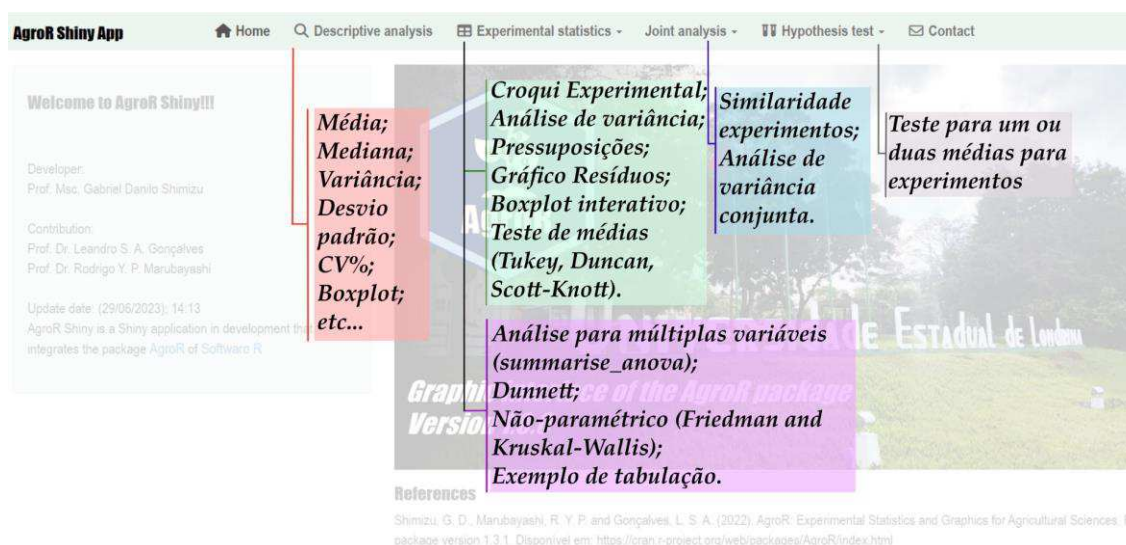


Figura 23 – Visão geral da estrutura do aplicativo shiny AgroR App e suas principais funções implementadas.

Na aba de análise descritiva é possível calcular medidas exploratórias de posição e dispersão, como média, mediana, máximo, mínimo, variância amostral, desvio-padrão amostral e coeficiente de variação do conjunto de dados e separados por tratamento de um experimento em DIC, DBC ou DQL, além de arranjos de fatores, como esquema fatorial e esquema de parcelas subdivididas. Também é retornado um gráfico de caixas interativo, visando a identificação de possíveis pontos discrepantes.

A aba de “croqui experimental” tem finalidade de auxiliar no planejamento do desenho experimental. Essa aba é baseada na função *sketch* do pacote AgroR, em que diversos projetos experimentais podem ser desenhados, como experimento com um fator em delineamento inteiramente casualizado, blocos casualizados e quadrado latino (DIC, DBC e DQL), além de arranjos de fatores, como esquema fatorial duplo e triplo, parcelas subdivididas e parcelas subsubdivididas. É possível inserir ruas dividindo as parcelas, gerar apenas a identificação das parcelas, as posições das repetições, dimensões de figuras, tamanhos de rótulos, entre outros parâmetros.

Na subseção “Analysis of variance” o usuário deve definir o projeto experimental em relação ao delineamento, número de fatores, se há um fator de separação, nível de significância da análise de variância e dos testes complementares, necessidade de transformação de dados. Na sub-aba de pressuposições, o usuário pode definir o teste de normalidade dos erros e homogeneidade das variâncias desejados. Em “Teste complementares” é definido o tipo de tratamento, o teste de comparações múltiplas no caso de fatores qualitativos e/ou o grau do polinômio no caso de fatores de níveis quantitativos. Na última sub-aba “Graphics”, pode-se definir os parâmetros gráficos. Dúvidas em relação a utilização podem ser acessados em “Help”. Ainda dentro da aba de “Experimental statistics”, o usuário pode analisar os dados por meio de testes não-paramétricos, ou por meio do teste de Dunnett para comparar um controle com os demais tratamentos.

Na aba de “Análise conjunta” é possível efetuar a análise de conjunta de experimentos conforme procedimento descrito por Barbin (2003) e Ferreira (2016). Assim, desde que, os experimentos apresentem homogeneidade das variâncias entre os experimentos e desde que não constatados interações significativas dos tratamentos com o experimento, o aplicativo retorna uma única análise, do contrário, o teste retorna os testes individualizados. Para identificar quais experimentos apresentam mesma variabilidade, pode-se utilizar a aba de “Similaridade de experimentos”, devendo adicionar o valor de cada quadrado médio do resíduo.

Na aba “Teste de hipóteses” há a opção de efetuar o teste de hipótese para uma ou duas médias populacionais por meio de testes paramétricos e não paramétricos, unilaterais ou bilaterais. Esses testes são amplamente utilizados por exemplo, para testar se as respostas dos tratamentos estão dentro de um limite pré-estabelecido na literatura.

O pacote *AgroR* possui até a data 26/03/2024, 23.114 downloads através do repositório CRAN, sendo que a média mensal foi de 862 downloads. O pacote possui sete citações de acordo com o google acadêmico (<https://scholar.google.com/>). Contudo, em buscas na web, foram computadas citações em 13 artigos científicos, uma tese de doutorado, uma dissertação de mestrado, seis trabalhos de conclusão de curso, três publicações em eventos e uma publicação técnica. O aplicativo *AgroR shiny* em sua versão em inglês possui um total de 1.021 usuários em 40 países e 139 cidades, que acessaram 2.457 vezes e com tempo médio de utilização de 9 minutos e 16 segundos. O aplicativo *AgroR shiny* em sua versão em português possui um total de 397 usuários em 22 países e 130 cidades, que acessaram 800 vezes e com tempo médio de utilização de 9 minutos.

CONCLUSÕES

- Espera-se que o pacote *AgroR* e seu aplicativo *AgroR Shiny* atenda à demanda por um pacote/aplicativo de fácil uso e que apresente diversas informações ao usuário.

REFERÊNCIAS BIBLIOGRÁFICAS

- ARNHOLD, E. Pacote em ambiente R para análise de variância e análises complementares. **Brazilian Journal of Veterinary Research and Animal Science**, v. 50, n. 6, p. 488-492, 2013. doi: 10.11606/issn.1678-4456.v50i6p488-492
- ARNHOLD, E. Pacote em ambiente R para automatizar estatísticas descritivas. **Sigmae**, v. 3, n. 1, p. 36-42, 2014.
- BAILAR, J. Science, statistics, deception. **Annals of Internal Medicine**, v. 104, s.1, p. 259-260, 1986. doi: 10.7326/0003-4819-104-2-259
- BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 26, n. 2, p. 211-243, 1964. doi: 10.1111/j.2517-6161.1964.tb00553.x
- CHANG, W.; CHENG, J.; ALLAIRE, J.; XIE, Y.; MCPHERSON, J. (2022) **Shiny: Web application framework for R**, Disponível em: <https://cran.r-project.org/web/packages/shiny/index.html>. Acesso em: 14 mar. 2024.
- FOX, J.; WEISBERG, S.; ADLER, D.; BATES, D.; BAUD-BOVY, G.; ELLISON, S.; HEIBERGER, R. Package ‘car’. *Vienna: R Foundation for Statistical Computing*, 2012.

- GROSS, J.; LIGGES, U. nortest: Tests for Normality. R package version 1.0-4, 2015. Disponível em: <https://CRAN.R-project.org/package=nortest>. Acesso em: 14 mar. 2024.
- HOTHORN, T.; BRETZ, F.; WESTFALL, P.; HEIBERGER, R. M.; SCHUETZENMEISTER, A.; SCHEIBE, S.; HOTHORN, M. T. (2016). **Package ‘multcomp’. Simultaneous inference in general parametric models.** *Project for Statistical Computing*.
- KNIEF, U.; FORSTMEIER, W. Violating the normality assumption may be the lesser of two evils. **Behavior Research Methods**, v. 53, n. 6, p. 2576-2590, 2021. doi: 10.3758/s13428-021-01587-5
- KORMANN, R.; ROSA, E. N.; PAIXÃO, C. A.; FERREIRA, E. B.; NOGUEIRA, D. A. GExpDes: Interface Gráfica para o ExpDes. **Sigmae**, v. 8, n. 2, p. 170-179, 2019.
- KÜHN, D. D. **Pesquisa e Análise de Dados: problematizando o rural e a agricultura numa perspectiva científica** (DERAD604). PLAGEDER, 96p, 2017.
- LENTH, R. (2023). **emmeans: Estimated Marginal Means, aka Least-Squares Means.** R package version 1.8.5. Disponível em: <https://CRAN.R-project.org/package=emmeans> Acesso em: 14 mar. 2024.
- LÚCIO, A. D.; SCHWERTNER, D. V.; HAESBAERT, F. M.; SANTOS, D.; BRUNES, R. R.; RIBEIRO, A. L.; LOPES, S. J. Violação dos pressupostos do modelo matemático e transformação de dados. **Horticultura Brasileira**, v. 30, n. 3, p. 415-423, 2012. doi: 10.1590/S0102-05362012000300010
- MARTIN, T. N.; STORCK L. **Análise das pressuposições do modelo matemático em experimentos agrícolas no delineamento blocos ao acaso.** In: MARTIN, T.N.; ZIECH, M.F. (org). *Sistemas de Produção Agropecuária*. Curitiba: UTFPR. p. 177-196, 2008.
- MELO, R. C.; TREVISANI, N.; SANTOS, M.; GUIDOLIN, A. F.; COIMBRA, J. L. M. Statistical model assumptions achieved by linear models: classics and generalized mixed. **Revista Ciência Agronômica**, v. 51, n. 1, p. 1-9, 2020. doi: 10.5935/1806-6690.20200015
- MENDIBURU, F.; SIMON, R. *Agricolae - Ten years of an open source statistical tool for experiments in breeding, agriculture and biology.* **PeerJ**, v. 1, p. 1-18. doi: 10.7287/peerj.preprints.1404v1
- NUNES, C. A.; ALVARENGA, V. O.; SOUZA SANT'ANA, A.; SANTOS, J. S.; GRANATO, D. The use of statistical software in food science and technology: Advantages, limitations and

misuses. **Food Research International**, v. 75, p. 270-280, 2015. doi: 10.1016/j.foodres.2015.06.011

PIMENTEL-GOMES, F. **Curso de estatística experimental**. 15. ed. Piracicaba: FEALQ, 451p, 2009.

R CORE TEAM (2023). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <http://www.R-project.org/>. Acesso em: 14 mar. 2024.

RIPLEY, B.; VENABLES, B.; BATES, D. M.; HORNIK, K.; GEBHARDT, A.; FIRTH, D.; RIPLEY, M. B. **Package ‘MASS’**. CRAN R. 2013.

SHANER, G.; FINNEY, R. The effect of nitrogen fertilization on the expression of slow-mildewing resistance in Knox wheat. **Phytopathology**, v. 67, n. 8, p. 1051-1056, 1977. doi: 10.1094/Phyto-67-1051

WICKHAM, H.; CHANG, W. **devtools: tools to make developing R packages easier**. R package version 1.13. 4. 2017.

WICKHAM, H.; CHANG, W. **ggplot2: Create elegant data visualizations using the grammar of graphics**. R package version, v. 2, n. 1, 2009. doi: 10.1007/978-0-387-98141-35

WICKHAM, H.; DANENBERG, P.; EUGSTER, M. **roxygen2: In-Line Documentation for R**, R package version, v. 6, n. 1, 2017.

XU, W.; LI, W.; SONG, D. Testing normality in mixed models using a transformation method. **Statistical Papers**, v. 54, p. 71-84, 2013. doi: 10.1007/s00362-011-0411-4

ZEILEIS, A.; HOTHORN, T. Diagnostic Checking in Regression Relationships. **R News**, v. 2, n. 3, p. 7-10, 2002. Disponível em: <https://CRAN.R-project.org/package=lmtest>. Acesso em: 14 mar. 2024

ARTIGO B

AgroReg: Principais modelos de regressão nas ciências agrárias implementadas na forma de um pacote R

Título curto: Pacote AgroReg para análise de regressão.

Manuscript category: Biometry, Modeling, and Statistics

Artigo Publicado na Revista Scientia Agrícola (*Qualis-Capes 2017-2020: A2; JCR 2021: 2,60*)

<http://doi.org/10.1590/1678-992X-2022-0041>

RESUMO

A análise de regressão representa uma das metodologias de elevada importância nas ciências agrárias, visto que muitos fatores estudados são quantitativos. Geralmente, os pesquisadores tem utilizado modelos polinomiais para explicar os resultados, sobretudo devido a maioria dos *softwares* existentes realizarem essa análise e falta de conhecimento de outros modelos. Por outro lado, sabe-se que muito dos fenômenos naturais não apresentam tal comportamento, contudo, a utilização de modelos não-lineares é onerosa e requer conhecimento avançado em linguagem de programação, como o R. Assim, o trabalho tem a finalidade de apresentar alguns modelos de regressão encontrados em trabalhos científicos e implementa-los na forma de um R *package* denominado *AgroReg*. O pacote engloba 44 funções de análise com 66 modelos de regressão, como polinomiais, não-paramétrico (loess), segmentados, logísticos, exponenciais, logarítmicos, entre outros. As funções retornam o coeficiente de determinação (R^2), coeficientes do modelo e respectivos valores de p do teste t , raiz quadrada do erro quadrático (RMSE), critério de informação de Akaike (AIC) e critério de inferência bayesiana (BIC), valores máximos e mínimos preditos e o gráfico de regressão. Outras medidas de qualidade do modelo e análise gráfica de resíduos também estão implementadas. O pacote pode ser baixado por meio do repositório CRAN, pelo comando: `install.packages(AgroReg)`. O *AgroReg* constitui uma ferramenta de análise promissora no âmbito da pesquisa agrícola pois apresenta funções simples

e amigáveis, que permitem um processamento rápido e eficiente dos dados, com o objetivo de oferecer maior confiabilidade e informação.

Palavras-chave: experimentos agrícolas; análise de regressão; seleção de modelos.

INTRODUÇÃO

Os experimentos agrônômicos são geralmente laboriosos, caros e muito das vezes podem levar anos para serem realizados. Além disso, muitas vezes são difíceis e complexos tanto no planejamento, quanto na execução e dependem de muitos fatores que afetam tanto a eficiência quanto a confiabilidade dos resultados, devido à variabilidade natural dos sistemas biológicos e agrícolas (PIEPHO; EDMONDSON, 2018). O bom planejamento experimental e a análise estatística adequada, bem como o controle efetivo da variabilidade, pelo uso de fatores controlados, são vitais para o sucesso dos experimentos (MONTGOMERY, 2017).

O desenho e a análise dos experimentos dependem das estruturas de medição dos fatores de tratamento, e a sua compreensão é essencial para uma análise correta (PIEPHO; EDMONDSON, 2018). Esses fatores podem ser classificados em qualitativos ou quantitativos dependendo do número de níveis do fator. No caso de fatores com níveis qualitativos, o mesmo não apresenta ordem específica dos níveis, como por exemplo, avaliação de cultivares (MONTGOMERY, 2017). Nesses casos, desde que o experimento tenha número adequado de repetições, os níveis do fator podem ser comparados descritivamente por um erro-padrão da diferença ou por um teste de comparações múltiplas (BRETZ *et al.* 2011; HSU, 1996; PIEPHO; EDMONDSON, 2018).

Os fatores de tratamento são ditos quantitativos, quando os níveis do fator são ordenáveis em uma escala numérica (MONTGOMERY, 2017), como por exemplo, doses de um mesmo fertilizante. Nesse caso, é recomendado a utilização da análise de regressão, que utilizam as informações de distância na escala das variáveis preditoras quantitativas e são essenciais para a captura de informações, permitindo estimar valores ainda que o mesmo não tenha sido observado no estudo (BANZATTO; KRONKA, 2013; COCHRAN; COX, 1986; PIMENTEL GOMES, 2009; STORCK *et al.*, 2016). Em comparação, uma simples análise qualitativa ignoraria qualquer informação sobre a relação distância e ordem entre os níveis de fator (PIEPHO; EDMONDSON, 2018). Contudo, um experimento com tal propósito requer

pelo menos três níveis do fator quantitativo, embora seja desejável ao menos cinco níveis, uma vez que um fator com apenas dois níveis quantitativos é normalmente considerado qualitativo.

Experimentos que estudam fatores quantitativos nas ciências agrárias são reportados em diversos artigos na literatura, por exemplo, estudo de densidade ou população de plantas (VAN ROEKEL; COULTER, 2011; VAN ROEKEL; COULTER, 2012; WILLIAMS *et al.*, 2021); estudos de pós-colheita de frutos (MARODIN *et al.*, 2016); na influência da temperatura na germinação de sementes (MOTSA *et al.*, 2015), em doses de herbicida no controle de plantas daninhas (NOEL *et al.*, 2018) ou em estudos de curva de crescimento de plantas (LÚCIO; SARI, 2017). Na maioria desses estudos, há uma predominância da utilização de modelos polinomiais para explicar os resultados e, embora não seja incorreto, é sabido que muito dos fenômenos naturais não apresentam tal comportamento, e sim modelos específicos, ditos não-lineares, como logísticos, exponencias, entre outros (ARCHONTOULIS; MIGUEZ, 2015).

A utilização de modelos não-lineares geralmente é realizada por meio de uma linguagem de programação, tal como o SAS (<https://www.sas.com/>) ou R (<https://www.r-project.org/>). No caso do *software* R, o mesmo pode ser acessado ainda na própria base da linguagem, com funções como o *lm*, *nls* ou *glm*. Em contrapartida, os modelos não-lineares, realizados pela função *nls*, requerem as especificações de valores *a priori* para obter as estimativas dos coeficientes do modelo, o que acaba sendo um processo demorado. Dessa forma, a implementação do chamado *R package* pode ser útil na realização dessas análises, por ser uma forma mais simples e fácil para os usuários. Assim, o presente trabalho tem a finalidade de apresentar alguns modelos de regressão encontrados em trabalhos científicos e implementá-los na forma de um *R package* denominado *AgroReg*.

MATERIAL E MÉTODOS

CRIAÇÃO DO PACOTE *AGROREG*

O pacote foi construído utilizando a linguagem R na versão 4.1.0 (R CORE TEAM, 2021), sendo a documentação e checagem gerada pelos pacotes *devtools* (WICKHAM *et al.*, 2021) e *roxygen2* (WICKHAM *et al.*, 2021), a fim de facilitar a construção e facilitar a adequação da política do CRAN. Os pacotes *drc* (RITZ *et al.*, 2016) e *ggplot2* (WICKHAM, 2016) foram importados e adicionados como dependência, sendo o primeiro para análise de regressão logística e o segundo para representação gráfica. Outros pacotes (*boot*, *minpack.lm*,

dplyr, *rcompanion* e *broom*) também foram adicionados como dependência para facilitar a construção das funções.

INSTALAÇÃO

Todas as funções desenvolvidas foram escritas na linguagem de programação R e, portanto, podem ser executadas no ambiente R ou em qualquer GUI (Interface Gráfica do Usuário) que utilize essa linguagem, como o *RStudio* (<https://www.rstudio.com/>). O R pode ser instalado nos sistemas Windows, Linux ou Mac. Assim, este pacote pode ser utilizado livremente pela comunidade científica, independentemente do sistema operacional. O pacote *AgroReg* pode ser instalado através do repositório CRAN, conforme o seguinte comando:

```
install.packages("AgroReg")
```

O seguinte comando deve ser executado para carregar o pacote:

```
library(AgroReg)
```

A documentação do pacote pode ser acessada pelo link: <https://cran.r-project.org/web/packages/AgroReg/AgroReg.pdf>

CONJUNTO DE DADOS

A coleção de funções disponíveis no pacote *AgroReg* implementa vários métodos para descrever muito dos fenômenos observados em estudos quantitativos nas ciências agrárias, conforme mencionado no Tabela 3. Assim, os dados obtidos em experimentos reais foram implementados para melhor exemplificação das aplicações do pacote, pois, o uso das funções contidas no pacote e a interpretação de seus resultados são melhor apresentados na forma de um exemplo aplicado com dados reais.

“aristolochia”

Os dados para exemplificação das funções *LM*, *LM_i*, *LM13*, *LM13_i*, *LM23*, *LM23_i*, *LM2_i3*, *logistic*, *LL*, *CD*, *BC*, *GP*, *SH*, *gaussianreg*, *loessreg*, *newton*, *valcam* e *VG*, são

oriundos de um experimento conduzido no Laboratório de Análises de Sementes do Centro de Ciências Agrárias da Universidade Estadual de Londrina (UEL), no qual foram avaliadas cinco temperaturas (15, 20, 25, 30 e 35°C) na germinação de *Aristolochia elegans*. O experimento foi conduzido em delineamento inteiramente casualizado com doze repetições de 25 sementes cada. Os dados podem ser acessados pelo comando `data("aristolochia")`.

“*granada*”

O conjunto de dados "granada" foi obtido de dados parciais de um experimento conduzido na UEL, com o intuito de avaliar a cinética de secagem da casca de romã ao longo do tempo. A perda de massa foi avaliada aos 60, 210, 390, 720, 930, 1410, 1890, 2.370 minutos após o início do experimento. Esse conjunto de dados foi usado para exemplificar as funções *AM*, *asymptotic*, *asymptotic_neg*, *asymptotic_i*, *asymptotic_ineg*, *biexponential*, *hill*, *MM*, *GP*, *weibull*, *GP*, *valcam*, *linear.linear*, *linear.plateau*, *quadratic.plateau*, *plateau.linear*, *plateau.quadratic*, *midilli*, *midillim*, *PAGE*, *peleg*, *potential*, *yieldloss*, *lorenz* e *mitscherlich* (Tabela 3). Os dados podem ser acessados pelo comando `data("granada")`.

MODELOS DE REGRESSÃO

Todos os modelos de regressão implementados no pacote estão apresentados na Tabela 3, além das funções e descrições, bem como as aplicações em artigos da literatura na área das ciências agrárias. Os modelos foram agrupados em modelos não-paramétricos (*loess*), polinomiais, logísticos ou em forma de S, logarítmicos, em forma de sino, segmentados e exponenciais. Os modelos foram em grande parte, extraídos de periódicos científicos de trabalhos inéditos, como Sadegui *et al.* (2019) ou artigos de revisão, como o de Archontoulis e Miguez (2015), visando abranger o maior número possível de modelos de regressão. Além disso, modificações de alguns modelos também foram implementados.

Os modelos polinomiais, também chamados de modelos lineares, foram implementadas a partir da função *lm* da base do R. Esse mesmo procedimento foi usado para obtenção das curvas logarítmicas, o modelo de Valcam e alguns modelos exponenciais. Por outro lado, a regressão não-paramétrica *loess*, conhecida também como regressão local, foi realizada através da função *loess*, também da base do R.

Modelos de equação logística, também chamado de curva sigmoide, que em muitos casos são usados para descrever a curva de crescimento de plantas, germinação de

sementes ao longo do tempo ou para dose-resposta de herbicidas (ARCHONTOULIS; MIGUEZ, 2015), apresentam como característica o formato de S, sendo essas implementados nos pacotes *drc* (RITZ *et al.*, 2016) e *aomisc* (<https://github.com/OnofriAndreaPG/aomisc>). Assim, no *AgroReg*, estas funções foram importadas e apenas sumarizadas em uma função mais simples e com maiores informações.

Por fim, para as demais funções, foi utilizado a função *nls* do pacote *stats*, que utiliza a metodologia de mínimos quadrados ordinários. Entretanto, nesses casos, foi utilizado algoritmos pré-estabelecidos para automatizar as informações iniciais. Assim, a maioria das funções não exige o fornecimento de informações iniciais para a obtenção dos modelos, embora em alguns casos, possa ocorrer o problema da convergência dos coeficientes, por não estimar bons valores iniciais. Nesses casos, o usuário poderá especificar informações a priori pelo argumento *initial*, conforme cada modelo de regressão.

Tabela 3 – Funções, descrições, modelo matemático e aplicações dos modelos implementados no pacote *AgroReg*.

Função	Descrição	Modelo	Aplicações
Descritivo			
Nreg	Gráfico descritivo	-	-
Não-paramétrico			
loess_model	Não-paramétrica local loess de grau 0, 1 ou 2	-	-
Modelos polinomiais ou lineares			
LM	Linear simples, quadrático, quadrático inverso, cúbico ou quártica.	$y = \beta_0 + \beta_1 X$ $y = \beta_0 + \beta_1 X + \beta_2 X^2$ $y = \beta_0 + \beta_1 X + \beta_2 X^{0,5}$ $y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$ $y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_3 X^4$	Doses de silício e sua influência na pós-colheita de tomate (MARODIN <i>et al.</i> , 2016), Curva de embebição de sementes de <i>Chrysanthemum leucanthemum</i> (PEGÔ; GROSSI; BARBOSA, 2012), rendimento de batata em função de doses de K ₂ O (MAINGI; MBUVI, 2020)
LM_i	Linear simples, quadrático, quadrático inverso, cúbico ou quártica sem intercepto.	$y = \beta_1 X$ $y = \beta_1 X + \beta_2 X^2$ $y = \beta_1 X + \beta_2 X^{0,5}$ $y = \beta_1 X + \beta_2 X^2 + \beta_3 X^3$ $y = \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_3 X^4$	Cinética de secagem de <i>Cydonia oblonga</i> (TZEMPELIKOS <i>et al.</i> , 2015)
LM13	Cúbico sem β_2	$y = \beta_0 + \beta_1 X + \beta_3 X^3$	Dose de palhada de <i>Merremia aegyptia</i> em caupi-hortaliça (SILVA <i>et al.</i> , 2020).
LM13i	Cúbico inverso sem β_2	$y = \beta_0 + \beta_1 X + \beta_3 X^{\frac{1}{3}}$	-
LM23	Cúbico sem β_1	$y = \beta_0 + \beta_2 X^2 + \beta_3 X^3$	Dose de palhada de <i>Merremia aegyptia</i> em caupi-hortaliça (SILVA <i>et al.</i> , 2020).
LM23i	Cúbico inverso sem β_1	$y = \beta_0 + \beta_2 X^{\frac{1}{2}} + \beta_3 X^{\frac{1}{3}}$	-
LM2i3	Cúbico sem β_1 e com β_3 inverso	$y = \beta_0 + \beta_2 X^2 + \beta_3 X^{\frac{1}{3}}$	-
valcam	Valcam	$y = \beta_0 + \beta_1 X + \beta_2 X^{1,5} + \beta_3 X^2$	Cinética de secagem de <i>Bauhinia forficata</i> (SILVA <i>et al.</i> , 2017)
Modelos logísticos, sigmoides ou em forma de S			
logistic	Logístico com três (npar="L.3"), quatro (npar="L.4") ou cinco (npar="L.5") parâmetros.	$y = \frac{d}{1 + e^{b(x-e)}}$ $y = c + \frac{d - c}{1 + e^{b(x-e)}}$ $y = c + \frac{d - c}{1 + e^{b(x-e)^f}}$	Curva de germinação de <i>Eleusine indica</i> (KERR <i>et al.</i> , 2018) e curva de crescimento de tangerina satsuma (<i>Citrus unshiu</i> Marc.) (YANO <i>et al.</i> , 2018) e curva de produção de morango (DIEL <i>et al.</i> , 2018)
LL	Log-logístico com três (npar="LL.3"), quatro (npar="LL.4") ou cinco (npar="LL.5") parâmetros.	$y = \frac{d}{1 + e^{b(\log(x) - \log(e))}}$ $y = c + \frac{d - c}{1 + e^{b(\log(x) - \log(e))}}$ $y = c + \frac{d - c}{1 + e^{b(\log(x) - \log(e))^f}}$	Dose-resposta de fungicidas (NOEL <i>et al.</i> , 2018), progresso do vírus Tomato brown rugose fruit virus (ToBRFV) em cultivares de tomate (GONZÁLEZ-CONCHA <i>et al.</i> , 2021)
BC	Brain-Cousens de quatro (npar="BC.4") ou cinco parâmetros (npar="BC.5").	$y = \frac{d + fx}{1 + e^{b(\log(x) - \log(e))}}$ $y = c + \frac{d - c + fx}{1 + e^{b(\log(x) - \log(e))}}$	Dose-resposta de Herbicida (SCHABENBERGER <i>et al.</i> , 1999), doses de <i>Trichoderma asperellum</i> em sementes de trigo (COUTO <i>et al.</i> , 2021)

Tabela 3 – continuação...

Função	Descrição	Modelo	Aplicações
CD	Cedergreen-Ritz-Streibig de quatro (npar="CRS.4") ou cinco (npar="CRS.5") parâmetros	$y = \frac{d + f e^{-\frac{1}{x}}}{1 + e^{b(\log(x) - \log(e))}}$ $y = c + \frac{d - c + f e^{-\frac{1}{x}}}{1 + e^{b(\log(x) - \log(e))}}$	Cinética de secagem de frutos de Kiwi (SADEGUI <i>et al.</i> , 2019) e dose-resposta de flufenaceta e pendimethalina no controle de <i>A. myosuroides</i> (METCALFE <i>et al.</i> , 2017)
Weibull	Weibull de três (npar="w3") ou quatro (npar="w4") parâmetros	$y = d e^{-e^{b(\log(x) - \log(e))}}$ $y = c + (d - c)(e^{-e^{b(\log(x) - \log(e))}})$	Curva de retenção de água de substratos (BATEMAN <i>et al.</i> , 2019)
GP	Gompertz de dois (npar="g2"), três (npar="g3") ou quatro (npar="g4") parâmetros	$y = e^{-e^{b(x-e)}}$ $y = d e^{-e^{b(x-e)}}$ $y = c + (d - c)(e^{-e^{b(x-e)}})$	Curva de crescimento (MELLO <i>et al.</i> , 2018) e dose resposta (MENDES <i>et al.</i> , 2019)
VB	Von Bertalanffy	$y = L(1 - e^{-k(x-t_0)})$	Curva de produção de morango (DIEL <i>et al.</i> , 2018)
lorentz	Lorentz de três ("lo3") ou quatro ("lo4") parâmetros.	$y = \frac{d}{1 + b(x - e)^2}$ $y = c \frac{d - c}{1 + b(x - e)^2}$	-
Modelos em forma de sino			
beta_reg	Regressão Beta, desenvolvido por Yin <i>et al.</i> (1995)	$y = d \left\{ \left(\frac{X - X_b}{X_o X_b} \right) \left(\frac{X_c - X}{X_c X_o} \right)^{\frac{X_c - X_o}{X_o - X_b}} \right\}^b$	Estimativa da temperatura cardinal para cultivares de Quinoa (MAMEDI <i>et al.</i> , 2017) e em <i>Alyssum homolocarpum</i> (ZAFERANIEH <i>et al.</i> , 2020)
gaussianreg	Função análoga a distribuição gaussiana ou modelo Bragg	$y = d e^{-b(x-e)^2}$ $y = c + (d - c) e^{-b(x-e)^2}$	-
Modelos segmentados			
linear.linear	Segmentado linear-linear	$y = \beta_0 + \beta_1 x (x < bp)$ $y = \beta_0 + \beta_1 bp + wx (x > bp)$	Estimativa de temperatura cardinal e ótima na germinação de sementes (MOTSA <i>et al.</i> , 2015); relação do teor de fósforo com o rendimento de trigo e milho (BIN <i>et al.</i> , 2016)
linear.plateau	Segmentado linear-platô	$y = \beta_0 + \beta_1 x (x < bp)$ $y = \beta_0 + \beta_1 bp (x > bp)$	Estimativa de densidade mínima ótima de plantas (FERREIRA <i>et al.</i> , 2020), rendimento de bulbo de cebola em função de doses de nitrogênio (GONÇALVES <i>et al.</i> , 2019).
plateau.linear	Segmentado platô-linear	$y = \beta_0 + \beta_1 bp (x < bp)$ $y = \beta_0 + \beta_1 x (x > bp)$	Relação entre a temperatura média diária e o peso e teor de óleo de girassol (ANGELONI <i>et al.</i> , 2021).
quadratic.plateau	Segmentado quadrática-platô	$y = \beta_0 + \beta_1 x + \beta_2 x^2 (x < bp)$ $y = \beta_0 + \beta_1 bp + \beta_2 bp^2 (x > bp)$	Densidade (VAN ROEKEL; COULTER, 2011) ou população de plantas de milho (WILLIAMS <i>et al.</i> , 2021)
plateau.quadratic	Segmentado platô-quadrático	$y = \beta_0 + \beta_1 x + \beta_2 x^2 (x > bp)$ $y = \beta_0 + \beta_1 bp + \beta_2 bp^2 (x < bp)$	-
Modelos logarítmicos			
LOG	Logarítmico	$f(x) = \beta_0 + \beta_1 \ln(x)$	Cinética de secagem de frutos de Kiwi (SADEGUI <i>et al.</i> , 2019)
LOG2	Logarítmico quadrático	$f(x) = \beta_0 + \beta_1 \ln(x) + \beta_2 \ln(x)^2$	-

Tabela 3 – continuação ...

Função	Descrição	Modelo	Aplicações
thompson	Thompson ou logarítmico quadrático sem intercepto	$f(x) = \beta_1 \ln(x) + \beta_2 \ln(x)^2$	Cinética de secagem de frutos de Kiwi (SADEGUI <i>et al.</i> , 2019)
Modelos exponenciais ou equações que apresentam características de crescimento exponencial			
asymptotic	Logarítmica ou Assintótico	$y = \alpha e^{-\beta x} + c$	Cinética de secagem de frutos de Kiwi (SADEGUI <i>et al.</i> , 2019)
asymptotic_i	Henderson e Pabis ou modelo bi-paramétrico exponencial	$y = \alpha e^{-\beta x}$	Secagem de sementes de laranja (ROSA <i>et al.</i> , 2015)
asymptotic_ineg	Assintótico negativo sem intercepto	$y = -\alpha e^{-\beta x}$	-
biexponential	Biexponencial	$y = A1 e^{-e^{lrc1x}} + A2 e^{-e^{lrc2x}}$	Degradação de atrazina no solo (SWARCEWICZ; GREGORCZYK, 2013)
mitscherlich	Lei de Mitscherlich	$y = A (1 - 10^{-eb-ex})$	Avaliação do efeito do pH na disponibilidade de fosfato (BARROW <i>et al.</i> , 2020)
yieldloss	Yield loss	$y = \frac{ix}{1 + \frac{i}{A}x}$	Relação entre a densidade de plantas daninhas e a perda de rendimento (COUSENS, 1985)
hill	Hill	$y = \frac{a \times x^c}{b + x^c}$	-
MM	Michaelis-Menten hipérbole retangular	ou $y = \frac{Vm \times x}{k + x}$ $y = c + \frac{Vm \times x}{k + x}$	Cinética de secagem da casca de romã (SHIMIZU <i>et al.</i> , 2020)
SH	Steinhart-Hart	$y = \frac{1}{A + B \ln(x) + C [\ln(x)]^3}$	Relação de densidade de plantas de milho com massa seca de grãos (ZHAI <i>et al.</i> , 2020)
PAGE	Page	$y = e^{-kx^n}$	Cinética de secagem de <i>Cydonia oblonga</i> (TZEMPELIKOS <i>et al.</i> , 2015), Kiwi (SADEGUI <i>et al.</i> , 2019)
newton	Newton ou Lewis	$y = e^{-kx}$	Cinética de secagem de Kiwi (SADEGUI <i>et al.</i> , 2019) e cebola (SHARMA <i>et al.</i> , 2005)
potential	Potencial	$y = \beta_1 X^n$	Cinética de secagem de frutos de Kiwi (SADEGUI <i>et al.</i> , 2019)
midilli	Midilli	$y = \alpha e^{kx^n} + bx$	Cinética de secagem de frutos de Kiwi (SADEGUI <i>et al.</i> , 2019), efeitos das condições de secagem na qualidade funcional e física da azeitona seca de outono (ZANNOU <i>et al.</i> , 2021)
midillim	Midilli modificado	$y = \alpha e^{kx} + bx$	Cinética de secagem de frutos de Kiwi (SADEGUI <i>et al.</i> , 2019)
AM	Avhad e Marchetti	$y = \alpha e^{kx^n}$	Cinética de secagem de sementes de abacate Hass (AVHAD; MARCHETTI, 2016)
peleg	Peleg	$y = \frac{1 - x}{a + bx}$	Efeitos das condições de secagem na qualidade funcional e física da azeitona seca de outono (ZANNOU <i>et al.</i> , 2021)

INFORMAÇÕES E PARÂMETROS ESTATÍSTICOS

As funções foram desenvolvidas de forma a fornecer estimativas de valores máximos e mínimos obtidos na curva dentro do intervalo estudado. Além disso, parâmetros estatísticos, como AIC (critério de informação de Akaike), BIC (critério de inferência Bayesiana), R^2 (coeficiente de determinação) ou Pseudo- R^2 (Correlação entre observados pelos

preditos), RMSE (raiz quadrada do erro-quadrático) e p -valor do teste t dos coeficientes também são retornados. No caso de modelos polinomiais é retornado também o fator de inflação de variância (VIF).

A raiz quadrada do erro-quadrático é calculada pela seguinte fórmula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}$$

em que: \hat{Y}_i é resposta predita pelo modelo, Y_i é a resposta observada, n é o tamanho amostral.

Os critérios de informação de Akaike (AIC) e inferência bayesiana (BIC) são calculados pela fórmula:

$$\begin{aligned} AIC_i &= -2\log L_i + 2p_i \\ BIC_i &= -2\log L_i + p_i \log n \end{aligned}$$

em que L_i e p_i são a probabilidade e o número de parâmetros para cada modelo, e n é o número de observações.

O fator de inflação de variância (VIF) é calculado pela fórmula:

$$VIF_j = \frac{1}{1-R_j^2} \quad j = 1, 2, \dots, p$$

em que: p é o número das variáveis predictoras; R_j^2 é o coeficiente de correlação múltipla, resultante da regressão X_j nos outros $p-1$ regressores.

Outros parâmetros estatísticos de bondade de ajuste, como MBE - erro médio (BIAS), MBER - erro médio relativo (BIAS), MAE - erro médio absoluto, RMAE - erro médio absoluto relativo, SE - erro médio, MSE - erro quadrático médio, rMSE - erro quadrático médio relativo, EF - eficiência de modelagem, SD – desvio-padrão das diferenças e CRM - coeficiente de massa residual são fornecidos separadamente das análises, através da função *stat_param*. Análises gráficas de resíduos podem ser realizadas pelo comando *extract.model*.

$$\begin{aligned} MBE &= \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i) \\ RMBE &= \frac{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)}{\bar{Y}_o} \end{aligned}$$

$$\begin{aligned}
 MAE &= \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i| \\
 RMAE &= \frac{\frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|}{\bar{Y}_i} \\
 SE &= \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \\
 MSE &= \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n} \\
 rMSE &= \frac{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}{\bar{Y}_o} \\
 EF &= 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \\
 SD &= \frac{\sum_{i=1}^n (\epsilon - \bar{\epsilon})^2}{n-1} \\
 CRM &= \frac{\bar{\epsilon}}{\bar{Y}_i}
 \end{aligned}$$

em que: \hat{Y}_i é a resposta predita pelo modelo, Y_i é a resposta observada, \bar{Y}_i é a média da resposta observada, $\bar{\epsilon}$ é a média da diferença entre a resposta predita pela observada e n é o tamanho amostral.

RESULTADOS E DISCUSSÃO

INFORMAÇÕES GERAIS

O pacote possui 44 funções de análise de regressão, podendo ser acessados também usando a função *regression* e definindo o argumento *model* conforme o modelo de regressão solicitado (Tabela 3). Essa função possui como default o modelo linear simples (*model="LMI"*), conforme a seguir:

```

> data("aristolochia")
> with(aristolochia, regression(trat, resp, model="LMI"))

```

Para mais informações, consulte a documentação da função (*?regression*).

Em todos as funções de análise, os dois primeiros argumentos são obrigatórios, sendo a variável independente e a variável dependente, respectivamente. No caso de funções polinomiais (*LM* e *LM_i*), o argumento *degree* define o grau do polinômio, enquanto

para funções logísticas e algumas exponenciais, como o *logistic*, *LL*, *BC*, *CD*, *GP*, *weibull*, *lorentz* e *MM*, o argumento *npar* define o número de parâmetros.

Na Figura 24 A e 24 B é representado os gráficos da análise de regressão linear simples e logístico de Brain-Cousens de quatro parâmetros, respectivamente. A junção de curvas pode ser acessada pela função *plot_arrange* (Figura 24 C), sendo o único argumento obrigatório uma lista com as saídas de cada análise, conforme a seguir:

```
> data("aristolochia")
> reg1 = with(aristolochia, LM(trat, resp))
> reg2 = with(aristolochia, BC(trat, resp))
> plot_arrange(list(reg1, reg2))
```

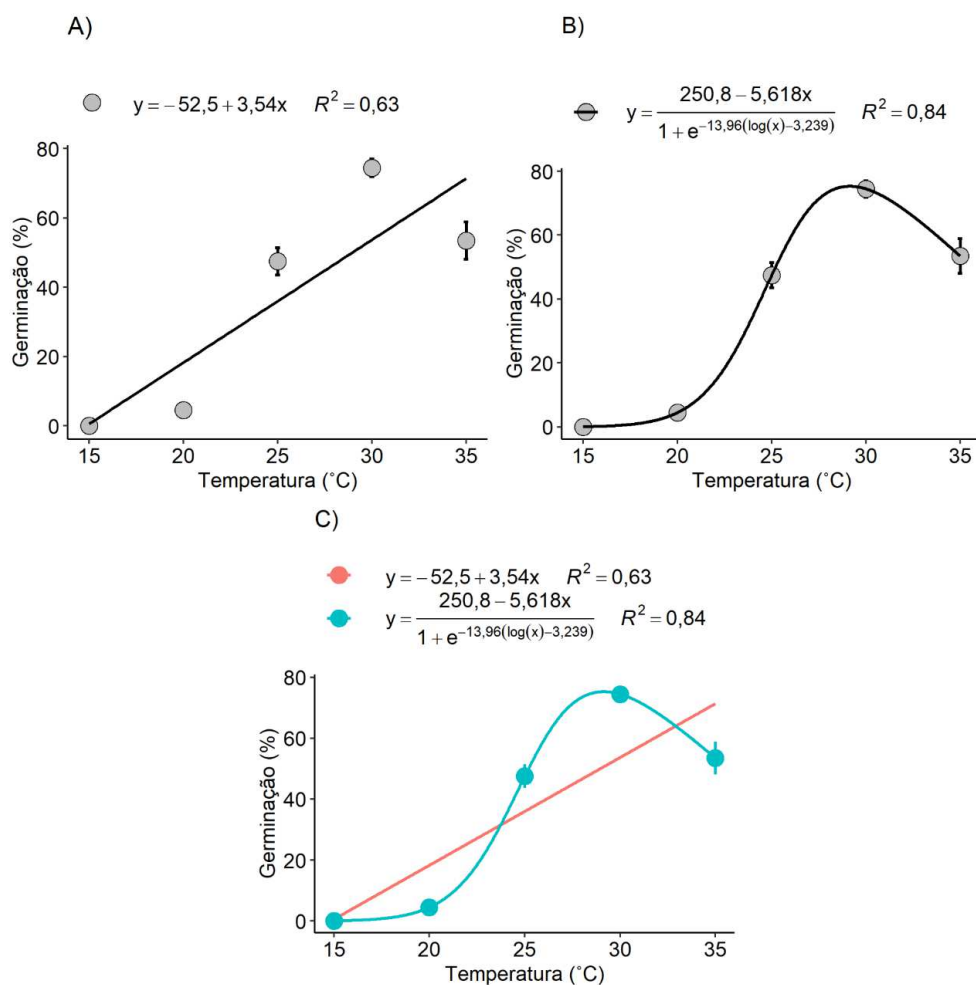


Figura 24 – Exemplificação da saída de uma função linear (A) e logística de Brain-Cousens (B) e união das curvas em um gráfico (C) usando as funções do pacote *AgroReg*.

Essa etapa de síntese da representação gráfica é uma etapa trabalhosa e passível de erro, sobretudo quando o usuário tem pouco experiência com a linguagem R. Assim, fornecer os gráficos de forma automática, como as equações e o coeficiente de determinação

(R^2 ou Pseudo- R^2) evita erros pelo pesquisador e otimiza tempo nesse processo dentro da análise de regressão. Além disso, o usuário pode também alterar diversos parâmetros gráficos, como formato, tamanho e cor de marcação, títulos e tamanho de eixos, plotagem de barras de erro-padrão, desvio-padrão ou sem barras, posição das equações, entre outros argumentos (consulte a ajuda do pacote *AgroReg*).

Por fim, o pacote também fornece informações importantes para seleção de modelos de regressão, como o AIC e BIC. Para ambos os critérios estatísticos, um valor menor indica um modelo preferível. O BIC difere do AIC apenas no segundo termo da equação, o que depende de n . Assim, à medida que n aumenta, o BIC favorece os modelos mais simples (menos parâmetros), motivo pelo qual, às vezes, os índices AIC e BIC discordam (ARCHONTOULIS; MIGUEZ, 2015). Além disso, também é retornado o coeficiente de determinação (R^2), em que valores próximos a 1 são desejáveis, embora no caso de modelos lineares (função *LM*), deva-se atentar ao problema da multicolinearidade, que é avaliado como “fator de inflação de variância” (VIF) na função, e que é recomendado por Myers e Montgomery (2002) ser inferior a 5 e por Petrini *et al.* (2012) se inferior a 10. Todas essas informações podem ser sumarizadas em uma tabela usando a função *comparative_model* e inserindo uma lista com as variáveis retornadas em cada função de análise.

EXEMPLO APLICADO

Para exemplificar e facilitar a compreensão do pacote *AgroReg*, bem como as interpretações, foi inserido o exemplo aplicado com o conjunto de dados "granada". A primeira etapa de qualquer análise estatística consiste em estudar informações exploratórias descritivas, obtendo por exemplo, medidas de posição, como média, mediana, máximos e mínimos; e medidas de dispersão, como variância e desvio-padrão. Por outro lado, no caso das análises de regressão, um aspecto que deve ser realizado a priori é a visualização gráfica dos resultados (ARCHONTOULIS; MIGUEZ, 2015), pois, com tal informação é possível identificar padrões e assim, direcionar a modelos específicos, o que evita processos desnecessários e encaminha para explicar de forma biologicamente aceitável. Essa importante etapa, passou a ser mais explorada a partir do conjunto de dados conhecido como quarteto de Anscombe, proposto pelo estatístico Francis Anscombe em 1973, que observou que em quatro conjuntos de dados, produziu-se ajuste e coeficientes de regressão idênticos, entretanto, quando visualizados graficamente revelaram padrões surpreendentemente diferentes de covariação entre x e y .

No caso do conjunto de dados “granada”, gráficos exploratórios usando a função *Nreg* (Figura 25). Visualmente é perceptível que há baixa variabilidade no conjunto de dados e que há crescimento acentuado até 1.000 minutos com posterior estabilidade. Esse comportamento é semelhante a modelos exponenciais ou que se comportam de forma similar, como por exemplo, Michaelis-Menten, modelos logísticos ou Mitscherlich. Modelos segmentados também são usados para explicar esse comportamento, como os modelos linear-linear, linear-platô e quadrático-platô. A seguir, a rotina dessas respectivas funções e o aspecto da curva (Figura 26).

```
> with(granada, Nreg(time, WL))
> models = c("asymptotic_neg", "biexponential", "LL3", "BC4", "CD5", "linear.linear",
"linear.plateau", "quadratic.plateau", "mitscherlich", "MM2")
> m = lapply(models, function(x) {
m = with(granada, regression(time, WL, model = x))})
> plot_arrange(m, trat = paste("(",models,")"))
```

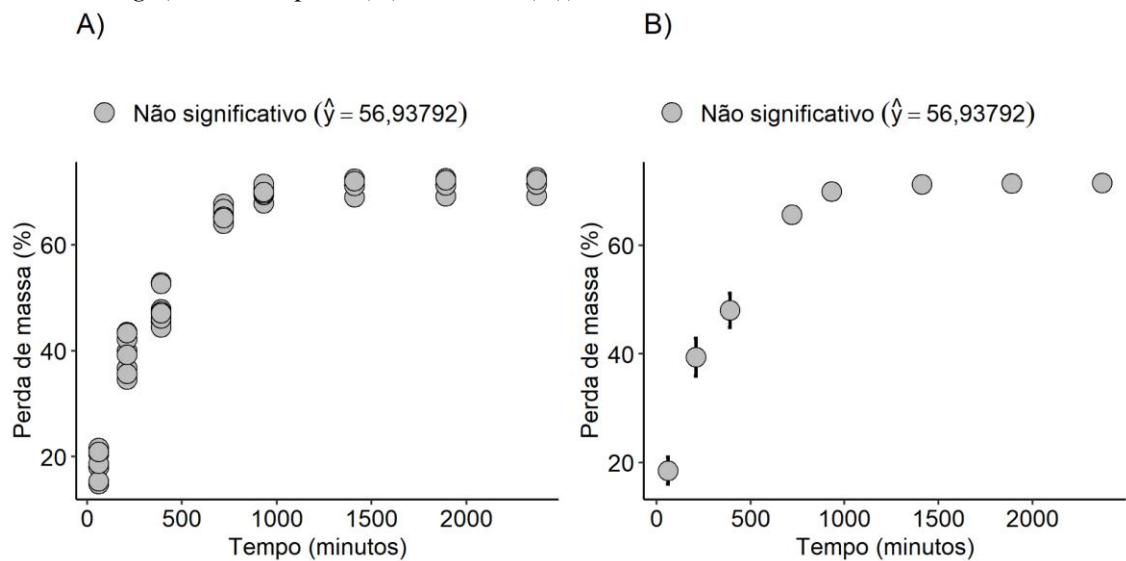


Figura 25 – Visualização de todas as observações por gráfico de dispersão (A) e por meio de média e desvio-padrão (B).

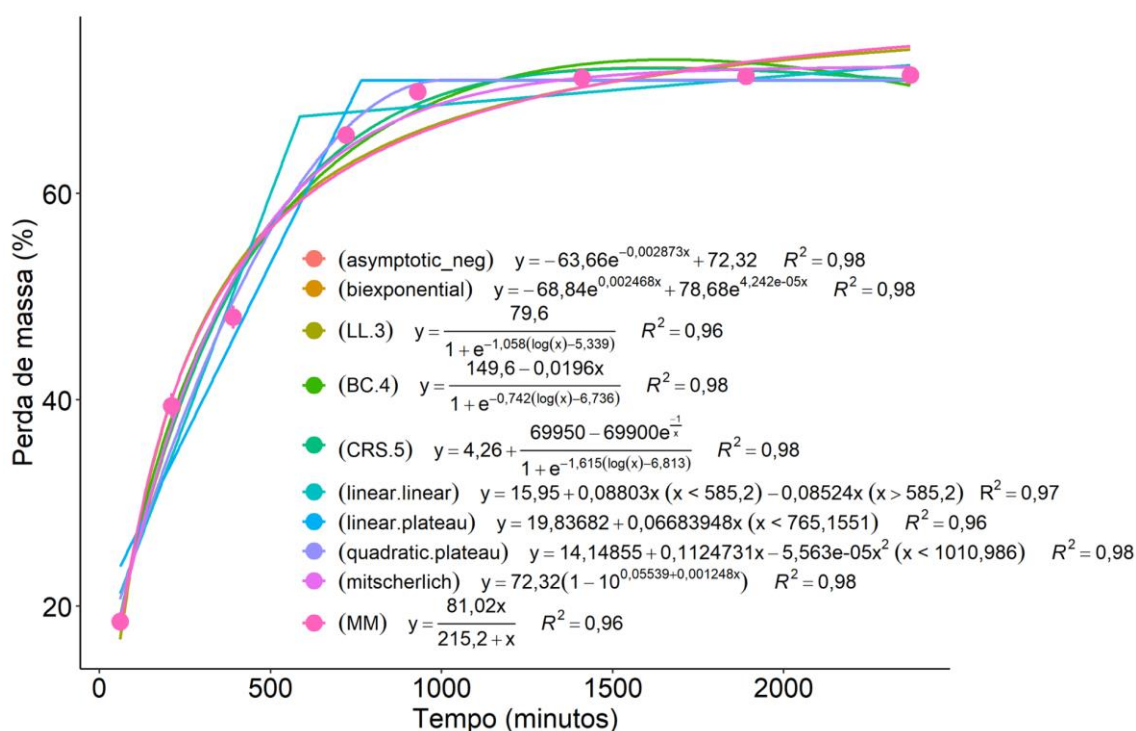


Figura 26 – Gráfico de regressão com os dez modelos de regressão usados para exemplificação dos comandos do pacote *AgroReg*.

Após a obtenção do modelo, a próxima etapa é a análise dos resíduos. No *AgroReg*, essa análise pode ser realizada de forma gráfica, conforme a seguir:

```
> a = with(granada, asymptotic_neg(time, WL))
> extract.model(a, type = "qqplot")
```

Com base no gráfico de quantis teóricos (Figura 27), todos os modelos apresentam pontos próximos a curva da distribuição normal. Embora tenha modelos que foram melhores ajustados, como log-logístico de três parâmetros e Brain-Cousens de quatro parâmetros. Na Tabela 4 são apresentados os parâmetros estatísticos de cada um dos modelos usados na Figura 26. Além disso, o pacote também implementa gráficos de barras que facilitam a visualização dos parâmetros de escolha do modelo (Figura 28). Dessa forma, no caso do exemplo, o modelo biexponencial foi o que apresentou os menores valores de AIC (307,46) e BIC (318,44) e também foi um dos modelos que apresentou menor valor de raiz quadrada do erro quadrático (RMSE) e maior coeficiente de determinação (R^2), além de apresentar todos os coeficientes significativos pelo teste t ($p < 0,05$). Entretanto, nota-se que, embora exista modelos

que são estatisticamente mais adequados, quase todos os modelos usados poderiam ser usados para explicar o comportamento desse estudo.

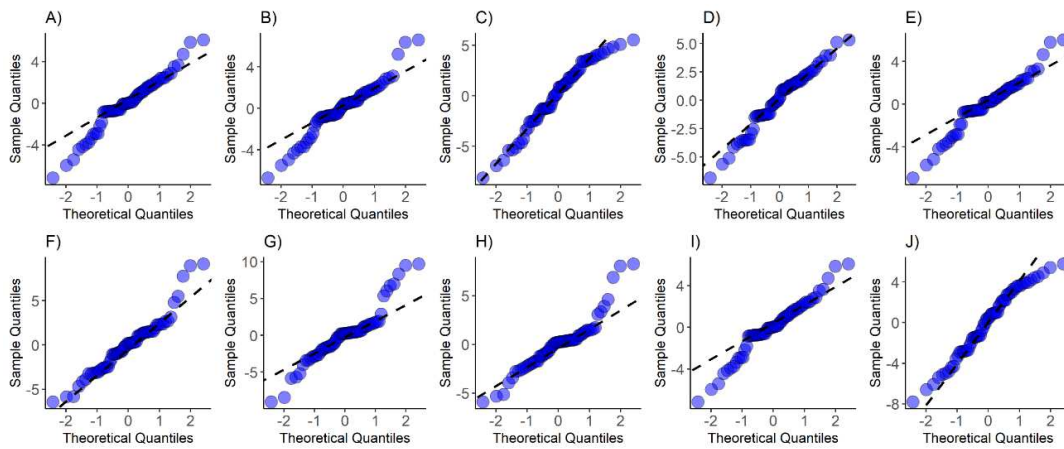


Figura 27 – gráfico *QQ-plot* do pacote *AgroReg* de resíduos dos modelos: assintótico negativo (A), biexponencial (B), log-logístico de três parâmetros (C), Brain-Counses de quatro parâmetros (D), Cedergreen-Ritz de cinco parâmetros (E), segmentado linear-linear (F), segmentado linear-platô (G), segmentado quadrático-platô (H), Mistcherlich (I) e Michaelis-Menten de dois parâmetros (J).

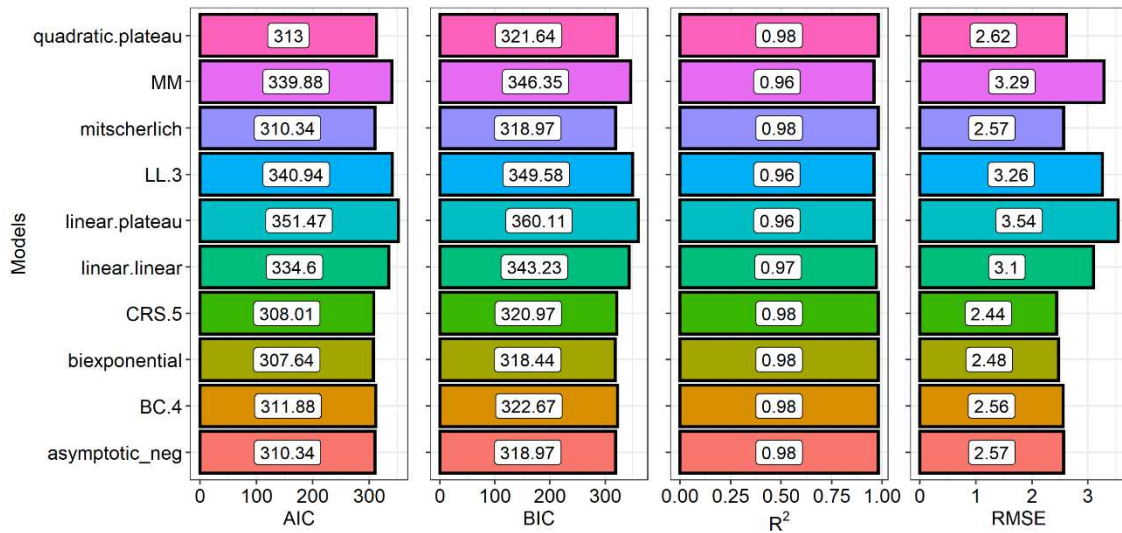


Figura 28 – Representação gráfica das comparações entre modelos de regressão pelos critérios de AIC, BIC, R² e RMSE.

Tabela 4 – Modelo de regressão, função do pacote *AgroReg*, coeficiente de determinação (R^2), critério de inferência de Akaike (AIC), critério de informação bayesiana (BIC), raiz quadrado do erro quadrático (RMSE) e coeficientes do modelo do conjunto de dados exemplo “granada”.

Modelo	Função	R^2	AIC	BIC	RMSE	Parâmetros
Assintótico negativo	asymptotic_neg	0,98	310,34	318,97	2,568	Alpha = -63,66** Beta = -0,00287** Theta = 72,32** A1 = -68,845**
Biexponencial	biexponential	0,98	307,46	318,44	2,475	lrc1 = -6,004** A2 = 78,6817** lrc2 = -10,06778**
Log-logístico de 3 parâmetros	LL	0,96	340,94	349,58	3,26	b = -1,058** d = 79,600** e = 208,408**
Brain-Cousens	BC	0,98	311,88	322,67	2,559	b = -0,742** d = 149,638** e = 842,598*
Cedergreen-Ritz	CD	0,98	308,01	320,97	2,444	f = -0,0196017** b = -1,613** c = 4,182 ^{ns} d = 71590** e = 921,4** f = -71540**
Linear-linear	linear.linear	0,97	334,60	343,23	3,1036	B ₀ = 15,947** X = 0,088** W = -0,085**
Linear-platô	linear.plateau	0,96	351,47	360,11	3,54	a = 19,84** b = 0,0668** c = 765**
Quadrático-platô	quadratic.plateau	0,98	313,00	321,64	2,62	B ₀ = 14,1** B ₁ = 0,112** B ₂ = 0,0000556** Jp = 1011**
Mitscherlich	mitscherlich	0,98	310,34	318,97	2,567	a = 72,32** b = 0,00125** e = 44,38**
Michaelis-Menten	MM	0,96	339,88	346,55	3,285	Vm = 81,02** K = 215,21**

*, **, ^{ns} significativo a 5%, 1% e não significativo pelo teste *t*, respectivamente.

O *AgroReg* constitui uma ferramenta de análise promissora no âmbito da pesquisa agrícola pois apresenta funções simples, que permitem um processamento rápido e eficiente dos dados, com o objetivo de oferecer maior confiabilidade e informação. Além disso, novas funções e atualizações serão realizadas para melhoria a fim de atender a demanda da comunidade científica.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANGELONI, P.; AGUIRREZÁBAL, L.; ECHARTE, M. M. Assessing the mechanisms underlying sunflower grain weight and oil content responses to temperature during grain filling. **Field Crops Research**, v. 262, n. 108040, 2021. doi: 10.1016/j.fcr.2020.108040
- ANSCOMBE, F. J. Graphs in Statistical Analysis. **The American Statistician**, v. 27, p. 17-21, 1973.
- ARCHONTOULIS, S. V.; MIGUEZ, F. E. Nonlinear regression models and applications in agricultural research. **Agronomy Journal**, v. 107, p. 786-798, 2015. doi: 10.2134/agronj2012.0506
- AVHAD, M. R.; MARCHETTI, J. M. Mathematical modelling of the drying kinetics of Hass avocado seeds. **Industrial Crops and Products**, v. 91, p. 76-87, 2016. doi: 10.1016/j.indcrop.2016.06.035
- BANZATTO, D. A.; KRONKA, S. N. 2013. 4ed. **Experimentação agrícola**. Jaboticabal: FUNEP. 237p.
- BARROW, N. J.; DEBNATH, A.; SEN, A. Measurement of the effects of pH on phosphate availability. **Plant and Soil**, v. 454, p. 217-224, 2020. doi: 10.1007/s11104-020-04647-5
- BATEMAN, A. M.; ERICKSON, T. E.; MERRITT, D. J.; VENEKLAAS, E. J.; MUÑOZ-ROJAS, M. Water availability drives the effectiveness of inorganic amendments to increase plant growth and substrate quality. **Catena**, v. 182, p. 1-10, 2019. doi: 10.1016/j.catena.2019.104116
- BRETZ, F.; HOTHORN, T.; WESTFALL, P. 2011. **Multiple comparisons using R**. Boca Raton, FL: CRC Press.
- COCHRAN, W. G.; COX, G. M. 1986. **Experimental design**. 2ed. New York: John Wiley. 611p.
- COUSENS, R. A simple model relating yield loss to weed density, **Annals of applied biology**, v. 107, p. 239-252, 1985. doi: 10.1111/j.1744-7348.1985.tb01567.x
- COUTO, A. P. S.; PEREIRA, A. E.; ABATI, J.; CAMARGO FONTANELA, M. L.; DIAS-ARIEIRA, C. R.; KROHN, N. G. Seed treatment with Trichoderma and chemicals to improve physiological and sanitary quality of wheat cultivars. **Revista Caatinga**, v. 34, p. 813-823, 2021. doi: 10.1590/1983-21252021v34n408rc

- FERREIRA, A. S.; ZUCARELI, C.; WERNER, F.; FONSECA, I. C. B.; BALBINOT JUNIOR, A. A. Minimum optimal seeding rate for indeterminate soybean cultivars grown in the tropics. **Agronomy Journal**, v. 112, p. 2092-2102, 2020. doi: 10.1002/agj2.20188
- GONÇALVES, F. D. C.; GRANGEIRO, L. C.; DE SOUSA, V. D. F.; SANTOS, J. P. D.; SOUZA, F. I. D.; DA SILVA, L. R. Yield and quality of densely cultivated onion cultivars as function of nitrogen fertilization. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 23, p. 847-851, 2019. doi: 10.1590/1807-1929/agriambi.v23n11p847-851
- GONZÁLEZ-CONCHA, L. F.; RAMÍREZ-GIL, J. G.; GARCÍA-ESTRADA, R. S.; REBOLLAR-ALVITER, Á.; TOVAR-PEDRAZA, J. M. Spatiotemporal Analyses of Tomato Brown Rugose Fruit Virus in Commercial Tomato Greenhouses. **Agronomy**, v. 11, p. 1268, 2021. doi: 10.3390/agronomy11071268
- HSU, J. C. 1996. **Multiple comparisons**. Theory and methods. London, UK: Chapman & Hall. doi: 10.1007/978-1-4899-7180-7
- KERR, R. A.; MCCARTY, L. B.; BRIDGES, W. C.; CUTULLE, M. Key Morphological Events following Late-Season Goosegrass (*Eleusine indica*) Germination. **Weed Technology**, v. 33, p. 196-201, 2019. doi: 10.1017/wet.2018.93
- LÚCIO, A. D. C.; SARI, B. G. Planning and implementing experiments and analyzing experimental data in vegetable crops: problems and solutions. **Horticultura Brasileira**, v. 35, p. 316-327, 2017. doi: 10.1590/S0102-053620170302
- MAINGI, F. M.; MBUVI, H. M. Effect of potassium dosage on selected growth parameters and yield response modeling on potatoes grown in Molo, Kenya. **Journal of Scientific Agriculture**, v. 4, p. 0101-0107, 2020. doi: 10.25081/jsa.2020.v4.6388
- MAMEDI, A.; TAVAKKOL AFSHARI, R.; OVEISI, M. Cardinal temperatures for seed germination of three quinoa (*Chenopodium quinoa* Willd.) cultivars. **Iranian Journal of Field Crop Science**, v. 48, p. 89-100, 2017. doi: 10.22059/ijfcs.2017.206204.654106
- MARODIN, J. C.; RESENDE, J. T. V.; MORALES, R. G. F.; FARIA, M. V.; TREVIZAM, A. R.; FIGUEIREDO, A. S. T.; DIAS, D. M. Tomato post-harvest durability and physicochemical quality depending on silicon sources and doses. **Horticultura Brasileira**, v. 34, p. 361-366, 2016. doi: 10.1590/S0102-05362016003009

- MENDES, R. R.; OLIVEIRA, R.; CONSTANTIN, J.; SILVA, V. F. V.; HENCKS, J. R. Identification and mapping of cross-resistance patterns to ALS-inhibitors in greater beggarticks (*Bidens* spp.). **Planta Daninha**, v. 37, p. 1-10, 2019. doi: 10.1590/S0100-83582019370100117
- METCALFE, H.; MILNE, A. E.; HULL, R.; MURDOCH, A. J.; STORKEY, J. The implications of spatially variable pre-emergence herbicide efficacy for weed management. **Pest management science**, v. 74 p. 755-765, 2017. doi: 10.1002/ps.4784
- MONTGOMERY, D. C. **Design and analysis of experiments**. Hoboken, NJ, John Wiley & Sons. 757p, 2017.
- MOTSA, M. M.; SLABBERT, M. M.; VAN AVERBEKE, W.; MOREY, L. Effect of light and temperature on seed germination of selected African leafy vegetables. **South African Journal of Botany**, v. 99, p. 29-35, 2015. doi: 10.1016/j.sajb.2015.03.185
- MYERS, R. H.; MONTGOMERY, D. C. **Response surface methodology: process and product optimization using designed experiments**. 2^oed. Nova York, John Wiley. 798p, 2002.
- NOEL, Z. A.; WANG, J.; CHILVERS, M. I. Significant influence of EC50 estimation by model choice and EC50 type. **Plant disease**, v. 102, p. 708-714, 2018. doi: 10.1094/PDIS-06-17-0873-SR
- PÊGO, R. G.; GROSSI, J. A. S.; BARBOSA, J. G. Soaking curve and effect of temperature on the germination of daisy seeds. **Horticultura Brasileira**, v. 30, p. 312-316, 2012. doi: 10.1590/S0102-05362012000200021
- PETRINI, J.; DIAS, R. A. P.; PERTILE, S. F. N.; ELER, J. P.; FERRAZ, J. B. S.; MOURÃO, G. B. Degree of multicollinearity and variables involved in linear dependence. **Pesquisa Agropecuária Brasileira**, v. 47, p. 1743-1750, 2012. doi: 10.1590/S0100-204X2012001200010
- PIEPHO, H. P.; EDMONDSON, R. N. A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels. **Journal of Agronomy and Crop Science**, v. 204, p. 429-455, 2018. doi: 10.1111/jac.12267
- PIMENTEL GOMES, F. **Curso de estatística experimental**. 15ed. Piracicaba: Nobel. 451p, 2009.

- RIBEIRO, T. D.; MATTOS, R. W. P. D.; MORAIS, A. R. D.; MUNIZ, J. A. Description of the growth of pequi fruits by nonlinear models. **Revista brasileira de fruticultura**, v. 40, p. 1-11, 2018. doi: 10.1590/0100-29452018949
- RITZ, C.; BATY, F.; STREIBIG, J. C.; GERHARD, D. Dose-response analysis using R. **PloS one**, v. 10, p. e0146021, 2015. doi: 10.1371/journal.pone.0146021
- ROSA, D. P.; CANTÚ-LOZANO, D.; LUNA-SOLANO, G.; POLACHINI, T. C.; TELIS-ROMERO, J. Mathematical modeling of orange seed drying kinetics. **Ciência e Agrotecnologia**, v. 39, p. 291-300, 2015. doi: 10.1590/S1413-70542015000300011
- SADEGHI, E.; HAGHIGHI ASL, A.; MOVAGHARNEJAD, K. Mathematical modelling of infrared-dried kiwifruit slices under natural and forced convection. **Food science & nutrition**, v. 7, p. 3589-3606, 2019. doi: 10.1002/fsn3.1212
- SCHABENBERGER, O.; THARP, B. E.; KELLS, J. J.; PENNER, D. Statistical tests for hormesis and effective dosages in herbicide dose response. **Agronomy Journal**, v. 91, p. 713-721, 1999. doi: 10.2134/agronj1999.914713x
- SHARMA, G.; VERMA, R.; PATHARE, P. Mathematical modeling of infrared radiation thin layer drying of onion slices. **Journal of Food Engineering**, v. 71, p. 282-286, 2005. doi: 10.1016/j.jfoodeng.2005.02.010
- SHIMIZU, G. D.; PAULA, J. C. B.; NEVES, C. S. V. J.; PACHECO, C. A. Quality of pomegranate peel cultivar Valenciana as determined by different drying methods. **Revista Brasileira de Fruticultura**, v. 42, p. 1-12, 2020. doi: 10.1590/0100-29452020431
- SILVA, F. P. D.; SIQUEIRA, V. C.; MARTINS, E. A.; MIRANDA, F.; MELO, R. M. Thermodynamic properties and drying kinetics of *Bauhinia forficata* Link leaves. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 21, p. 61-67, 2017. doi: 10.1590/1807-1929/agriambi.v21n1p61-67
- SILVA, J. N. D.; BEZERRA, F.; LIMA, J. S. S. D.; BARROS, A. P.; NUNES, R. L. C.; CHAVES, A. P. Agronomic and economic feasibility indicators for cowpea-vegetable under green manure in a semiarid environment. **Revista Ciência Agronômica**, v. 51, p. 1-11, 2020. doi: 10.5935/1806-6690.20200040
- STORCK, L.; LOPES, S. J.; ESTEFANEL, V.; GARCIA, D. C. **Experimentação Vegetal**. 3Ed. Santa Maria, Editora UFSM. 198p, 2016.

SWARCEWICZ, M. K.; GREGORCZYK, A. Atrazine degradation in soil: effects of adjuvants and a comparison of three mathematical models. **Pest management science**, v. 69, p. 1346-1350, 2013. doi: 10.1002/ps.3510

TZEMPELIKOS, D. A.; VOUIROS, A. P.; BARDAKAS, A. V.; FILIOS, A. E.; MARGARIS, D. P. Experimental study on convective drying of quince slices and evaluation of thin-layer drying models. **Engineering in Agriculture, Environment and Food**, v. 8, p. 169-177, 2015. doi: 10.1016/j.eaef.2014.12.002

VAN ROEKEL, R. J.; COULTER, J. A. Agronomic responses of corn to planting date and plant density. **Agronomy Journal**, v. 103, p. 1414-1422, 2011. doi: 10.2134/agronj2011.0071

WICKHAM, H.; DANENBERG, P.; CSÁRDI, G.; EUGSTER, M. 2020. **roxygen2: In-Line Documentation for R**. R package version 7.1.1. Disponível em: <https://CRAN.R-project.org/package=roxygen2>. Acesso em: 14 mar. 2024.

WICKHAM, H.; HESTER, J.; CHANG, W. 2021. **devtools: Tools to Make Developing R Packages Easier**. R package version 2.4.2. Disponível em: <https://CRAN.R-project.org/package=devtools>. Acesso em: 14 mar. 2024.

WILLIAMS, J. J.; DODDS, D. M.; BUEHRING, N. W.; DHILLON, J. S.; HENRY, W. B. Corn response to row spacing and plant population in the Mid-South United States. **Agronomy Journal**, v. 113, p. 4132-4141, 2021. doi: 10.1002/agj2.20817

XI, B.; ZHAI, L. M.; LIU, J.; LIU, S.; WANG, H. Y.; LUO, C. Y.; REN, T. Z.; LIU, H. B. Long-term phosphorus accumulation and agronomic and environmental critical phosphorus levels in Haplic Luvisol soil, northern China. **Journal of Integrative Agriculture**, v. 15, p. 200-208, 2016. doi: 10.1016/S2095-3119(14)60947-3

YANO, T.; MORISAKI, A.; MATSUBARA, K.; ITO, S. I.; KITANO, M. Growth analysis of potted seedlings of Satsuma mandarin (*Citrus unshiu* Marc.) under different light conditions and air temperatures. **Horticulture Journal**, v. 87, p. 34-42, 2018. doi: 10.2503/hortj.OKD-051

YIN, X.; KROP, M. J.; MCLAREN, G.; VISPERAS, R. M. A nonlinear model for crop development as a function of temperature. **Agricultural and Forest Meteorology**, v. 77, p. 1-16, 1995. doi: 10.1016/0168-1923(95)02236-Q

ZAFERANIEH, M.; MAHDAVI, B.; TORABI, B. Effect of temperature and water potential on *Alyssum homolocarpum* seed germination: Quantification of the cardinal temperatures and

using hydro thermal time. **South African Journal of Botany**, v. 131, p. 259-266, 2020. doi: 10.1016/j.sajb.2020.02.006

ZANNOU, O.; PASHAZADEH, H.; GHELLAM, M.; HASSAN, A. M.; KOCA, I. Optimization of drying temperature for the assessment of functional and physical characteristics of autumn olive berries. **Journal of Food Processing and Preservation**, v. 45, p. 1-13, 2021. doi: 10.1111/jfpp.15658

ZHAI, L.; LI, H.; SONG, S.; ZHAI, L.; MING, B.; LI, S.; XIE, R.; JIA, X.; ZHANG, L. Intra-specific competition affects the density tolerance and grain yield of maize hybrids. **Agronomy Journal**, v. 113, p. 224-235, 2021. doi: 10.1002/agj2.20438

ARTIGO C

***seedreg*: UM PACOTE R PARA ÁREA DE TECNOLOGIA DE SEMENTES.**

RESUMO

Estudos de germinação e emergência de sementes apresentam, comumente, problemas de pressuposições, o que invalida a utilização da maioria dos testes paramétricos, como à análise de variância. Assim, o objetivo desse trabalho é construir um pacote em linguagem R denominado *seedreg* que aborda cálculos tradicionalmente usados na área de tecnologia de sementes e outras abordagens potenciais, bem como testes estatísticos com aplicações na área. O pacote apresenta 18 funções, realizando o cálculo de índice de velocidade de germinação (IVG) ou emergência (IVE) tradicionais ou baseados em regressão logística, tempo médio de germinação ou emergência tradicionais e baseados em T50 (TME ou TMG) e velocidade de emergência ou germinação (v). O pacote apresenta funções para Anova/Teste de médias, modelos de regressão linear e não-linear, análise por modelo linear generalizado para fatores qualitativos e quantitativos com aplicações na área de sementes. Para exemplificar os resultados, foi usado um conjunto de dados oriundo de um experimento que avaliou cinco temperaturas (15, 20, 25, 30 e 35°C) na germinação de *Aristolochia elegans*. Os modelos foram comparados quanto aos critérios estatísticos, como o AIC, BIC, R^2 , RMSE, significância dos coeficientes e o sentido biológico. O pacote *seedreg* pode ser instalado pelo repositório CRAN, utilizando o comando `install_packages("seedreg")`. Para o exemplo utilizado, foi constatado que a estimativa e os critérios estatísticos, bem como o sentido biológico do modelo proposto por Brain-Cousens e Cedergreen-Ritz-Streibig com quatro parâmetros apresentaram as melhores explicações dentre os modelos testados. O pacote *seedreg* será atualizado frequentemente e espera-se que facilite às análises de experimentos na área de sementes.

Palavras-chave: gráficos; regressão não-linear; *Software R*

INTRODUÇÃO

Os dados oriundos de experimentos voltados para área de tecnologia de sementes possuem várias características que os diferenciam de outros dados frequentemente coletados em experimentos agrícolas (SCOTT *et al.*, 1984). Primeiro, que a germinação é considerada como uma variável qualitativa, pois são variáveis binárias, ou seja, germinou ou não germinou (RANAL; SANTANA, 2006). Segundo que, para uma dada população de sementes, a germinação completa raramente ocorre. Dentre as causas, destaca-se a inviabilidade inicial, a perda de viabilidade pelo efeito do próprio tratamento, a censura aleatória, por exemplo, perda ou contaminação acidental, e a censura à direita devido ao término prematuro do teste, por exemplo, devido à dormência.

Testes para avaliação da qualidade das sementes são rotineiramente utilizados em experimentação. Um dos principais é o teste de germinação, quase sempre associado a algum teste indicativo de vigor, como o teste de crescimento de plântulas (NAKAGAWA, 1999). No teste de germinação, é comum a contagem diária do número de sementes germinadas, e esses dados são frequentemente utilizados para cálculo do índice de velocidade de germinação (MAGUIRE, 1962), tempo médio de germinação e velocidade de germinação (EDMOND; DRAPALA, 1958). Este mesmo procedimento é aplicado aos dados de emergência de plântulas, além de ser usado analogamente a outros campos da pesquisa, como na brotação de estacas de videira (RIBEIRO *et al.*, 2021).

As planilhas eletrônicas são utilizadas de forma generalizada para tabular dados, realizar os mais diversos cálculos experimentais e processar informações. No entanto, erros nas fórmulas ou no manuseio das planilhas podem comprometer significativamente as informações geradas (POWELL *et al.*, 2009). Assim, o desenvolvimento de ferramentas que minimizem esses riscos e possam aumentar a confiabilidade dos resultados é importante.

Embora muitos *softwares* de baixo custo já estejam disponíveis para as mais diversas áreas, há poucas que permitem cálculos voltados para área de tecnologia de sementes e que também sejam de fácil de uso. No entanto, há algumas planilhas gratuitas que não são realmente um *software*, mas um conjunto de planilhas desenvolvidas no pacote Microsoft® Excel (JOOSEN *et al.*, 2010; MIRANDA *et al.*, 2014; RANAL *et al.*, 2009). Implementações desses cálculos podem ser encontradas também em pacotes estatísticos, como no caso do *SeedCalc* (SILVA *et al.*, 2019), que possibilita calcular diversos parâmetros de germinação/emergência, ou em *software*, como no caso GerminaR/GerminaQuant (LOZANO-ISLA *et al.*, 2019; MARQUES *et al.*, 2015), que foi desenhado para facilitar o entendimento dos processos de germinação e a análise dos dados.

Vale ressaltar que, dentre esses citados, a maioria não necessariamente apresenta todas as aplicações no cotidiano dos pesquisadores ligados à área de sementes, parte pela baixa divulgação em ambiente acadêmico, parte por não apresentar a realidade da maioria dos experimentos dessa área. Isso é notoriamente verdadeiro, quando analisamos a infinidade de possíveis experimentos, a exemplo no caso do estudo da germinação de sementes sob diferentes temperaturas.

A estimativa da faixa térmica quanto a germinação de sementes, embora seja um tópico de grande interesse, sobretudo visando obter as temperaturas ótimas, mínimas e máximas para que ocorra a germinação, são poucos explorados por meio de modelagem matemática. Por exemplo, estudos como o de Belmehdi *et al.* (2018) utilizaram teste de Tukey para comparar as diferentes temperaturas na germinação de *Origanum elongatum*, o mesmo foi realizado para *Origanum compactum* por Laghmouchi *et al.* (2017) e *Psidium guajava* por Alves *et al.* (2015). Embora o teste de médias não seja necessariamente errado e bastante popular em manuscritos, podem gerar resultados equivocados, além de muitas vezes não apresentar sentido biológico. Todavia, a utilização de regressões, por ser uma análise mais complexa, tem sido subutilizada nesses ensaios e poderia ser mais bem explorada, sobretudo utilizando modelos não-lineares. Esse ponto é particularmente verdadeiro quando as análises requerem um conhecimento mais aprofundado de estatística e a necessidade de usar linguagem de programação, como o R (R CORE TEAM, 2022).

Essa modelagem pela linguagem R pode ser acessada ainda na própria base da linguagem, com funções como o *lm*, *nls* ou *glm*. Em contrapartida, a implementação dos chamados R *packages* podem ser úteis na modelagem, pois facilitam a utilização por usuários inexperientes, assim como ocorre no pacote *drc* (RITZ *et al.*, 2016). Assim, o trabalho tem a finalidade de apresentar alguns modelos de regressão comuns nos trabalhos científicos e outros com potenciais para uso, além de implementar gráficos, análises e cálculos usados na área da tecnologia de sementes, na forma de um R *package* denominado *seedreg*.

MATERIAIS E MÉTODOS

CRIAÇÃO DO PACOTE SEEDREG

O pacote foi construído utilizando a linguagem R 4.2.1 (R CORE TEAM, 2022), sendo a documentação gerada pelos pacotes *devtools* (WICKHAM *et al.*, 2021) e *roxygen2* (WICKHAM *et al.*, 2021). Os pacotes *drc* (RITZ *et al.*, 2016) e *ggplot2* (WICKHAM,

2016) foram importados e adicionados como dependência, sendo o primeiro para análise de regressão não-linear e o segundo para representação gráfica. A análise de *deviance* e o teste de comparação múltipla com ajuste de Tukey foi realizado com base nos pacotes *emmeans* (LENTH; LENTH, 2018), *multcomp* (HOTHORN *et al.*, 2016) e *car* (FOX *et al.*, 2012). Foi implementado oito funções de análises:

- *LM_model()*: função para regressões polinomiais de primeiro (*grau = 1*), segundo (*grau = 2*) ou terceiro (*grau = 3*) grau.
- *LL_model()*: função para regressão logística com três (*npar = "LL.3"*) ou quatro (*npar = "LL.4"*) parâmetros.
- *BC_model()*: função para regressão logística de Brain-Cousens de quatro (*npar = "BC.4"*) ou cinco parâmetros (*npar = "BC.5"*).
- *CD_model()*: função para regressão logística de Cedergreen-Ritz-Streibig de quatro (*npar = "CRS.4"*) ou cinco (*npar = "CRS.5"*) parâmetros.
- *normal_model*: função para regressão análogo ao modelo da densidade de probabilidade de uma distribuição gaussiana.
- *piecewise_model*: função para regressão segmentada linear-linear.
- *loess_model()*: análise para regressão não-paramétrica loess.
- *quali_model()*: análise considerando o fator qualitativo. É retornado a análise de deviance e o teste de comparação múltipla. Nessa função é possível escolher entre a distribuição binomial ou quasibinomial.

O pacote apresenta funções para calcular de forma automatizada algumas variáveis de germinação, como o índice de velocidade de germinação ou emergência (IVG ou IVE), usando a metodologia de Maguire (1962), tempo médio de germinação ou emergência (TME ou TMG) e a velocidade de emergência ou germinação (*v*) (LABOURIAU, 1983), conforme a seguir:

$$IVG = E_1/N_1 + E_2/N_2 + \dots + E_n/N_n$$

em que: E_1, E_2 e E_n é o número de plântulas normais, computadas na primeira, segunda e última contagem; N_1, N_2, N_n é o número de dias de semeadura à primeira, segunda e última contagem.

$$TMG = \frac{G_1T_1 + G_2T_2 + \dots + G_nT_n}{G_1 + G_2 + \dots + G_n}$$

em que: TMG é o tempo médio, em dias, necessário para atingir a germinação máxima; G_1, G_2 e G_n é o número de sementes germinadas nos tempos T_1, T_2 e T_n , respectivamente.

Além disso, foi implementada a possibilidade de calcular esses mesmos parâmetros, mas usando uma abordagem alternativa, baseada na análise de regressão logística. Para isso, para cada repetição, poderá ajustar um modelo de regressão logístico de três ou quatro parâmetros, conforme especificado pelo usuário, entre o tempo e o número de sementes germinadas/emergidas acumuladas. Após as análises, é calculada a área abaixo da curva, o que representa uma medida similar ao índice de velocidade de emergência ou germinação. O tempo médio de germinação ou emergência poderá ser calculada com base no E50, o que representaria o tempo até a germinação de 50% das sementes e por fim, a velocidade de germinação/emergência seria o inverso do tempo médio ($1/v$). Tais medidas poderão ser comparadas com as medidas clássicas.

O pacote *seedreg* pode ser instalado através do repositório CRAN, conforme a seguir:

```
install.packages("seedreg")
```

A documentação do pacote pode ser acessada pelo link: <https://cran.r-project.org/web/packages/seedreg/seedreg.pdf>

EXEMPLO COM CONJUNTO DE DADOS

Os dados para exemplificação são oriundos de um experimento conduzido no Laboratório de Análises de Sementes do Centro de Ciências Agrárias da Universidade Estadual de Londrina, no qual foram avaliadas cinco temperaturas (15, 20, 25, 30 e 35°C) na germinação de *Aristolochia elegans*. O experimento foi conduzido em delineamento inteiramente casualizado com doze repetições de 25 sementes cada. O conjunto de dados também apresenta dados de velocidade de germinação (dias^{-1}). Os dados estão disponíveis no pacote, podendo ser acessado pelo comando `data("aristolochia")`.

Foi utilizado onze modelos de regressão: linear ($Y = \beta_0 + \beta_1 X$); quadrático ($Y = \beta_0 + X + \beta_2 X^2$); cúbico ($Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$); logístico de três ($Y = \frac{d}{1+e^{(b(\log(x)-\log(e)))}}$) ou quatro ($Y = c + \frac{d-c}{1+e^{(b(\log(x)-\log(e)))}}$) parâmetros; logístico de Brain-Cousens com quatro ($Y = \frac{d+fx}{1+e^{(b(\log(x)-\log(e)))}}$) ou cinco parâmetros ($Y = c + \frac{d-c+fx}{1+e^{(b(\log(x)-\log(e)))}}$); logístico de Cedergreen-Ritz-Streibig com quatro ($Y = \frac{d+fx \exp(-\frac{1}{x})}{1+e^{(b(\log(x)-\log(e)))}}$) ou cinco parâmetros ($Y = c + \frac{d-c+fx \exp(-\frac{1}{x})}{1+e^{(b(\log(x)-\log(e)))}}$); modelo análogo a

função densidade de probabilidade de uma distribuição normal ($Y = ae^{-\frac{(x-b)^2}{2c^2}}$), modelo segmentado ($Y = \beta_0 + \beta_1x(x < bp); Y = \beta_1bp + wx(x > bp)$) e o modelo não-paramétrico loess.

Os modelos foram comparados quanto aos critérios estatísticos, como o AIC (critério de inferência de Akaike), BIC (critério de inferência Bayesiana), R^2 (coeficiente de determinação), RMSE (raiz quadrada do erro-quadrático), significância dos coeficientes e, o sentido biológico pela visualização de dados.

RESULTADOS E DISCUSSÃO

A função *loess_model* retorna a regressão não-paramétrica de loess, que é usada normalmente para análise preliminar. A função *quali_model* realiza a análise de *deviance* considerando a distribuição binomial ou quasibinomial e posterior teste de médias, metodologia essa comum em manuscritos em dados dicotomizados. As funções *LL_model*, *LM_model*, *BC_model*, *CD_model* e *normal_model* do pacote *seedreg* retornam além dos gráficos e estimativas dos coeficientes, importantes critérios estatísticos para escolha do modelo, como o AIC, BIC, RMSE e R^2 . Todas essas análises podem ser facilmente obtidas inserindo apenas o vetor de tratamento, vetor resposta e o argumento do modelo, conforme a seguir:

```
library(seedreg)
data("aristolochia")
attach(aristolochia)
LM_model(trat, germ) # modelo linear
LM_model(trat, germ, grau = 2) # modelo quadrático
LM_model(trat, germ, grau = 3) # modelo cúbico
LL_model(trat, germ, npar = "LL.3()") # modelo logístico de três parâmetros
LL_model(trat, germ, npar = "LL.4()") # modelo logístico de quatro parâmetros
BC_model(trat, germ, npar = "BC.4()") # modelo Brain-Cousens de quatro parâmetros
BC_model(trat, germ, npar = "BC.5()") # modelo Brain-Cousens de cinco parâmetros
CD_model(trat, germ, npar = "CRS.4") # modelo Cedergreen-Ritz-Streibig de quatro parâmetros
CD_model(trat, germ, npar = "CRS.5") # modelo Cedergreen-Ritz-Streibig de cinco parâmetros
normal_model(trat, germ) # modelo análogo a função densidade de probabilidade de uma normal
piecewise_model(trat, germ) # modelo segmentado linear-linear
loess_model(trat, germ) # modelo loess
quali_model(trat, germ, n=25, family="quasibinomial") # Teste de médias
```

No conjunto de dados “aristolochia”, foi testado todos os parâmetros estatísticos, sendo representado na Tabela 5. Pelos valores de AIC, BIC e RMSE, o modelo de Brain-Cousens e o de Cedergreen-Ritz-Streibig de quatro parâmetros (BC.4 e CRS.4) foram os melhores dentre os implementados no pacote. Além disso, foi um dos poucos modelos que permitiram estimar a temperatura ótima e a temperatura mínima para que ocorra a germinação. Nesse ensaio, não foi possível estimar a temperatura máxima limitante. Em relação a estimativa dos pontos, nota-se que há similaridade em relação ao TOE entre o modelo cúbico, loess, CRS.4 e o BC.4, enquanto para TME, os modelos BC.5, BC.4, CRS.5, CRS.4 e o quadrático apresentaram estimativas próximas (Tabela 5). Vale ressaltar que embora o coeficiente de determinação (R^2) do modelo cúbico seja superior ao BC.4 e CRS.4, este primeiro apresenta multicolinearidade alta, o que indica que o modelo não está adequado.

Tabela 5 – Critério de inferência de Akaike (AIC); critério de inferência bayesiana; coeficiente de determinação (R^2); raiz quadrado do erro quadrático (RMSE); temperatura ótima estimada (TOE); temperatura mínima estimada (TME) dos modelos linear, quadrático, cúbico, logístico de três (LL.3) ou quatro parâmetros (LL.4), modelo de Brain-Cousens de quatro (BC.4) ou cinco parâmetros (BC.5), Cedergreen-Ritz-Streibig de quatro (CRS.4) ou cinco parâmetros (CRS.5), segmentado linear-linear (Piecewise) loess e teste de comparação múltipla (TCM).

Modelo	AIC	BIC	R^2	RMSE	TOE	TME
Linear	706,93	714,08	0,74	19,33	-	-
Quadrático	693,88	703,40	0,82	17,60	34,25	16,33
Cúbico	643,07	654,98	0,99	12,65	30,97	19,00
LL.3	661,30	670,82	0,79	14,36	-	18,37
LL.4	662,77	674,68	0,80	14,31	-	-
BC.4	642,34	654,25	0,84	12,59	29,14	16,91
BC.5	663,97	678,26	0,80	12,59	-	16,68
CRS.4	642,34	654,25	0,84	12,59	28,96	16,84
CRS.5	663,97	678,26	0,80	14,24	-	16,68
Normal	645,58	645,05	0,83	12,67	30,48	-
Piecewise	664,16	673,69	0,79	39,75	30,28	16,64
loess	-	-	-	-	30,31	18,99
TCM	-	-	-	-	30,00	15,00

A escolha de modelos de regressão nas ciências agrárias tem se resumido basicamente a modelos polinomiais de primeiro ou segundo grau, visto a facilidade de análise e a representação gráfica. Além disso, normalmente esses modelos apresentam ajuste significativo ($p < 0,05$) e de coeficiente de determinação alto. Por outro lado, assim como descrito por Altman e Krzywinski (2017), interpretar e adotar o modelo unicamente pelo valor de probabilidade (p -valor), ou seja, se ele é significativo, pode gerar desvios de interpretação. Além disso, muitas vezes os modelos lineares não representam os processos biológicos de uma cultura (ARCHONTOULIS; MIGUEZ, 2015). Nesse sentido, o pesquisador deve estar ciente de que o ajuste de curva deve levar em conta não só parâmetros de qualidade da análise, como a explicação biológica do fenômeno.

Outro ponto importante é em relação a utilização de teste de médias para comparação entre as temperaturas. Embora na literatura é descrito que é recomendado a realização de ensaios variando a temperatura em 2°C e a realização de teste de médias, como descrito por Brasil (2009), esses resultados podem desviar das reais temperaturas limitantes. Por exemplo, no conjunto de dados “*aristolochia*”, o teste de comparação múltipla indica que a menor temperatura (15°C) apresenta-se limitante para germinação (Figura 29 H). Todavia, baseado nas demais curvas, estimamos que a temperatura limite é inferior a 17°C. Nesse sentido, a abordagem por modelagem estatística pode ser tão eficiente quanto analisar diversas faixas de temperatura, além disso, reduz custos e otimizam tempo.

O pacote *seedreg* fornece também a exibição de gráficos com as equações de forma automática, o que facilita a visualização dos dados e evita erros de editoração. Por exemplo, na Figura 29 B, nota-se que embora o ajuste quadrático possa ser usado para explicar os dados, baseado em parâmetros estatísticos, ele estima valores negativos em certo ponto, o que biologicamente é inexplicável. Por outro lado, o modelo de Brain-Cousens de quatro parâmetros (Figura 29 E) apresentou além dos melhores parâmetros estatísticos, melhor sentido biológico e visualmente melhor ajuste de curva, não estimando valores negativos em nenhum momento da curva no intervalo estudado.

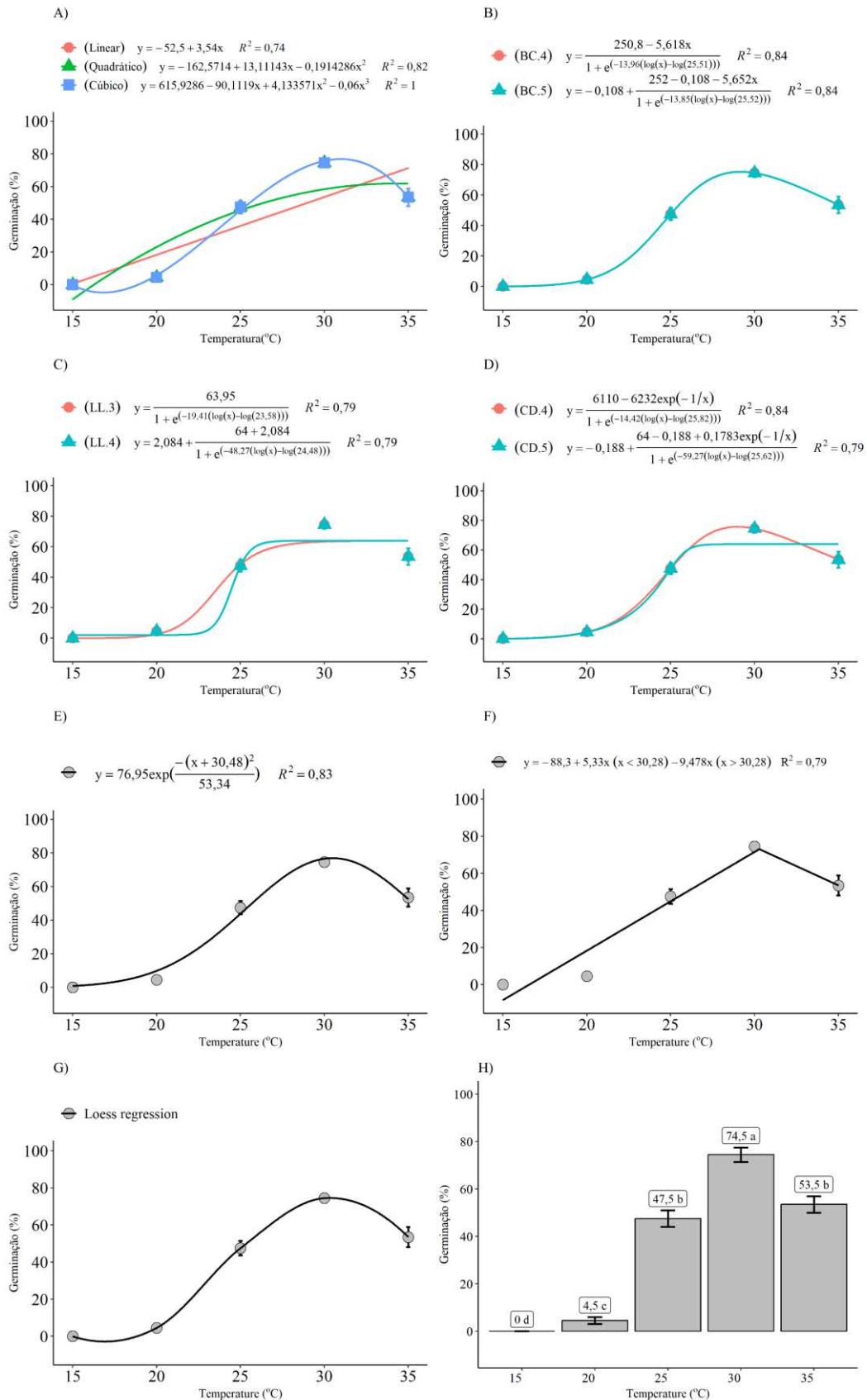


Figura 29 – Modelo de gráfico do pacote seedreg com os modelos: linear, quadrático, cúbico (A); LL.3 e LL.4 (B); BC.4 e BC.5 (C), CRS.4 e CRS.5 (D), normal (E), piecewise_model (F); loess (G) e teste de comparação múltipla (G).

A partir do pacote *seedreg*, e a adoção de modelos pré-estabelecidos na literatura, porém adotados em outras situações, podemos otimizar a estimativa das faixas térmicas para germinação de sementes, dando maior sentido a real comportamento dos dados. Além disso, com futuras atualizações, será possível ampliar as possibilidades e gerar ajustes que atendam a um leque maior de espécies.

CONCLUSÃO

- A estimativa e os parâmetros estatísticos, bem como o sentido biológico do modelo proposto por Brain-Cousens e Cedergreen-Ritz-Streibig com quatro parâmetros apresentou a melhor explicação dentre os modelos testados, sugerindo esses modelos para esses ensaios.
- O pacote *seedreg* pode ser uma importante ferramenta para análise de dados de germinação em função da temperatura, apresentando funções simples e saídas com informações pertinentes e vitais para escolha do modelo.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALTMAN, N.; KRZYWINSKI, M. Points of significance: interpreting p values. **Nature methods**, v. 14, n. 3, 2017. doi: 10.1038/nmeth.4210
- ALVES, C. Z.; SILVA, J. B. D.; CÂNDIDO, A. C. D. S. Metodologia para a condução do teste de germinação em sementes de goiaba. **Revista Ciência Agronômica**, v. 46, n. 3, p. 615-621, 2015. doi: 10.5935/1806-6690.20150045
- ARAÚJO, A. M.; ASSIS, L. C. S. L. C.; NOGUEIRA, W. N.; FREITAS, R. M. O.; TORRES, S. B. Substrates and temperatures for the germination of seeds of *Senegalia tenuifolia* (L.) Britton & Rose. **Revista Caatinga**, v. 29, n. 1, p. 113-118, 2016. doi: 10.1590/1983-21252016v29n113rc
- ARCHONTOULIS, S. V.; MIGUEZ, F. E. Nonlinear regression models and applications in agricultural research. **Agronomy Journal**, v. 107, n. 2, p. 786-798, 2015. doi: 10.2134/agronj2012.0506

BELMEHDI, O.; EL HARSAL, A.; BENMOUSSI, M.; LAGHMOUCHI, Y.; SENHAJI, N. S.; ABRINI, J. Effect of light, temperature, salt stress and pH on seed germination of medicinal plant *Origanum elongatum* (Bonnet) Emb. & Maire. **Biocatalysis and agricultural biotechnology**, v. 16, p. 126-131, 2018. doi: 10.1016/j.bcab.2018.07.032

BRASIL - Ministério da Agricultura, Pecuária e Abastecimento. **Regras para análise de sementes**. Brasília: Mapa/ACS, 2009. 399p.

DÜRR, C.; DICKIE, J. B.; YANG, X. Y.; PRITCHARD, H. W. Ranges of critical temperature and water potential values for the germination of species worldwide: contribution to a seed trait database. **Agricultural and forest meteorology**, v. 200, p. 222-232, 2015. doi: 10.1016/j.agrformet.2014.09.024

EDMOND, J. B.; DRAPALA, W. J. The effects of temperature, sand and soil, and acetone on germination of okra seed. **Proceedings of the American Society for Horticultural Science**, v. 71, p. 428-434, 1958.

FOX, J.; WEISBERG, S.; ADLER, D.; BATES, D.; BAUD-BOVY, G.; ELLISON, S.; ...; MURDOCH, D. (2012). **Package ‘car’**. Vienna: R Foundation for Statistical Computing.

HOTHORN, T.; BRETZ, F.; WESTFALL, P.; HEIBERGER, R. M.; SCHUETZENMEISTER, A.; SCHEIBE, S.; HOTHORN, M. T. (2016). **Package ‘multcomp’**. **Simultaneous inference in general parametric models**. Project for Statistical Computing, Vienna, Austria.

JOOSEN, R. V. L.; KODDE, J.; WILLEMS, L. A. J.; LIGTERINK, W.; VAN DER PLAS, L. H. W.; HILHORST, H. W. M. Germinator: A software package for high-throughput scoring and curve fitting of *Arabidopsis* seed germination. **The Plant Journal**, v. 62, p. 148-159, 2010. doi: 10.1111/j.1365-313x.2009.04116.x

LAGHMOUCHI, Y.; BELMEHDI, O.; BOUYAHYA, A.; SENHAJI, N. S.; ABRINI, J. Effect of temperature, salt stress and pH on seed germination of medicinal plant *Origanum compactum*. **Biocatalysis and agricultural biotechnology**, v. 10, p. 156-160, 2017. doi: 10.1016/j.bcab.2017.03.002

LAMARCA, E. V.; SILVA, C. V.; BARBEDO, C. J. Limites térmicos para a germinação em função da origem de sementes de espécies de *Eugenia* (Myrtaceae) nativas do Brasil. **Acta Botanica Brasilica**, v. 25, n. 2, p. 293-300, 2011. doi: 10.1590/S0102-33062011000200005

LENTH, R.; LENTH, M. R. Package ‘lsmeans’. **The American Statistician**, v. 34, n. 4, p. 216-221, 2018.

LOPES, R. R.; FRANK, L. B. Aspectos térmico-biológicos da germinação de sementes de cornichão anual sob diferentes temperaturas. **Revista Brasileira de Zootecnia**, v. 40, n. 10, p. 2091-2096, 2011. doi: 10.1590/S1516-35982011001000004

MARCOS FILHO, J. **Fisiologia de sementes de plantas cultivadas**. 2ª ed. Londrina: ABRATES, 2015.

MELLO, J. I. O.; BARBEDO, C. J. Temperatura, luz e substrato para germinação de sementes de Pau-Brasil (*Caesalpinia echinata* Lam., Leguminosae – Caesalpinioideae). **Revista Árvore**, v. 31, n. 4, p. 645-655, 2007. doi: 10.1590/S0100-67622007000400009

MIRANDA, R. Q.; CORREIA, R. M.; ALMEIDA-CORTEZ, J. S.; POMPELLI, M. F. Germinação de *Prosopis juliflora* (Sw.) D.C sementes em diferentes potenciais e temperaturas osmóticas. **Biologia de Espécies Vegetais**, v. 29, p. e9-e20, 2014. doi: 10.1111/1442-1984.12025

NAKAGAWA, J. Testes de vigor baseados no desempenho das plântulas. In: KRZYZANOSKI, F. C.; VIEIRA, R. D.; FRANÇA NETO, J. B. (Ed.). **Vigor de sementes: conceitos e testes**. Londrina: ABRATES, 1999. p.2.1-2.24.

PEREIRA, K. T. O.; PAIVA, E. P. D.; SOUZA NETA, M. L. D.; BENEDITO, C. P.; TORRES, S. B. Physiological quality evaluation of *Piptadenia stipulacea* (Benth.) Ducke seeds by tetrazolium test. **Revista Ciência Agronômica**, v. 51, n. 1, 2020. doi: 10.5935/1806-6690.20200019

POWELL, S. G.; BAKER, K. R.; LAWSON, B. Errors in operational spreadsheets. **Journal of Organizational and End User Computing**, v. 21, n. 3, p. 24-36, 2009. doi: 10.4018/joeuc.2009070102

R CORE TEAM (2023). **R: A language and environment for statistical computing**. **R Foundation for Statistical Computing, Vienna, Austria**. Disponível em: <https://www.R-project.org/>. Acesso em: 14 mar. 2024.

RANAL, M. A.; SANTANA, D. G. D. How and why to measure the germination process? **Brazilian Journal of Botany**, v. 29, n. 1, p. 1-11, 2006. doi: 10.1590/S0100-84042006000100002

RANAL, M. A.; SANTANA, D. G.; FERREIRA, W. R.; MENDES-RODRIGUES, C. Calculating germination measurements and organizing spreadsheets. **Brazilian Journal of Botany**, v. 32, n. 4, p. 849-855, 2009. doi: 10.1590/s0100-84042009000400022

RIBEIRO, L. T. M.; BINOTTO, C.; SHIMIZU, G. D.; HIGUCHI, M. T.; ROBERTO, S. R.; SATO, A. J. Bud burst induction in BRS-Carmem vine minicuttings under controlled environment. **Comunicata Scientiae**, v. 12, p. e3678-e3678, 2021. doi: 10.14295/cs.v12.3678

RITZ, C.; STREBIG, J. C.; RITZ, M. C. (2016). **Package ‘drc’**. Creative Commons: Mountain View, CA, USA.

SCOTT, S. J.; JONES, R. A.; WILLIAMS, W. Review of data analysis methods for seed germination I. **Crop science**, v. 24, n. 6, p. 1192-1199, 1984. doi: 10.2135/cropsci1984.0011183X002400060043x

SILVA, L. J. D.; MEDEIROS, A. D. D.; OLIVEIRA, A. M. S. SeedCalc, a new automated R software tool for germination and seedling length data processing. **Journal of Seed Science**, v. 41, n. 2, p. 250-257, 2019. doi: 10.1590/2317-1545v42n2217267

TESFAY, S. Z.; MODI, A. T.; MOHAMMED, F. The effect of temperature in moringa seed phytochemical compounds and carbohydrate mobilization. **South African Journal of Botany**, v. 102, n. 1, p. 190-196, 2016. doi: 10.1016/j.sajb.2015.07.003

WICKHAM, H. **ggplot2**. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 3, n. 2, p. 180-185, 2016.

LOZANO-ISLA, F.; BENITES-ALFARO, O. E.; POMPELLI, M. F. GerminaR: An R package for germination analysis with the interactive web application “GerminaQuant for R.” **Ecological Research**, v. 34, n. 2, p. 339-346, 2019. doi: 10.1111/1440-1703.1275

MARQUES, F. R. F.; MEIADO, M. V.; CASTRO, N. M. C. R. D.; CAMPOS, M. L. D. O.; MENDES, K. R.; SANTOS, O. D. O. D.; POMPELLI, M. F. GerminaQuant: a new tool for germination measurements. **Journal of Seed Science**, v. 37, n. 3, p. 248-255, 2015. doi: 10.1590/2317-1545v37n3145605

CONCLUSÕES GERAIS

Os pacotes e aplicativos desenvolvidos em linguagem R realizam a análise estatística de experimentos agrícolas e atendem a demanda pela comunidade científica, sobretudo nas ciências agrárias, por serem ferramentas simples e com potencial para a pesquisa.

O pacote *AgroR* e seu aplicativo shiny atende à demanda por um pacote/aplicativo de fácil uso na experimentação agrícola e que apresenta as informações necessárias para o usuário, bem como permite a visualização gráfica dos resultados.

O pacote *AgroReg* atende a demanda por um pacote destinado a análise de regressão, abordando diversos modelos comumente utilizados no âmbito das ciências agrárias.

O pacote *seedreg* permite estimar temperaturas cardinais por diferentes métodos estatísticos, além de apresentar funções para calcula automatizado de importantes parâmetros da área de tecnologia de sementes.