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

Scientia Agricola

ISSN 1678-992X

Research article

AgroReg: main regression models in agricultural sciences implemented as an R Package

Gabriel Danilo Shimizu*, Leandro Simões Azeredo Gonçalves

Universidade Estadual de Londrina – Depto. de Agronomia, Rod. Celso Garcia Cid, Pr 445 km 380 – 86057-970 – Londrina, PR – Brasil.
*Corresponding author <shimizu@uel.br>

Edited by: Luiz Alexandre Peternelli

Received November 07, 2022
Accepted January 27, 2023

ABSTRACT: Regression analysis is highly relevant to agricultural sciences since many of the factors studied are quantitative. Researchers have generally used polynomial models to explain their experimental results, mainly because much of the existing software perform this analysis and a lack of knowledge of other models. On the other hand, many of the natural phenomena do not present such behavior; nevertheless, the use of non-linear models is costly and requires advanced knowledge of language programming such as R. Thus, this work presents several regression models found in scientific studies, implementing them in the form of an R package called AgroReg. The package comprises 44 analysis functions with 66 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 the respective *p*-values from the *t*-test, root mean square error (RMSE), Akaike's information criterion (AIC), Bayesian information criterion (BIC), maximum and minimum predicted values, and the regression plot. Furthermore, other measures of model quality and graphical analysis of residuals are also included. The package can be downloaded from the CRAN repository using the command: *install.packages*("*AgroReg*"). AgroReg is a promising analysis tool in agricultural research on account of its user-friendly and straightforward functions that allow for fast and efficient data processing with greater reliability and relevant information.

Keywords: agronomic experiments, regression analysis, model selection

Introduction

Agronomic experiments are generally laborious, expensive, and often take years to be performed. Moreover, they are often tricky and complex in planning and execution. They depend on many factors that affect both the efficiency and reliability of results due to the natural variability of biological and agricultural systems (Piepho and Edmondson, 2018).

Experimental design and analysis depend on the measurement structures of treatment factors, and their understanding is essential to a correct analysis (Piepho and Edmondson, 2018). In the case of qualitative factors, the levels do not have a specific level order on a numerical scale (Montgomery, 2017). In these cases, if the experiment has adequate repetitions, they can be compared by the standard error of the differences or tests of means (Hsu, 1996; Bretz et al., 2011; Piepho and Edmondson, 2018).

Treatment factors are quantitative when the levels can be ordered numerically (Montgomery, 2017). In this case, regression analysis is recommended, which uses distance information on the scale of quantitative predictors, allowing for the estimation of values even if they were not observed in the study (Cochran and Cox, 1986; Pimentel-Gomes, 2009; Banzatto and Kronka, 2013; Storck et al., 2016). However, an experiment for such a purpose requires at least three levels of the factor, although at least five are desirable.

Experiments that study quantitative factors in agricultural sciences have been reported in several articles such as studies of plant density or population (Van Roekel and Coulter, 2011; Williams et al., 2021), fruit post-harvest quality (Marodin et al., 2016), seed germination (Motsa et al., 2015), weed control (Noel et al., 2018), and growth curves (Lúcio and Sari, 2017). In these studies, polynomial models were predominantly used and, although this is not incorrect, many natural phenomena do not present such behavior but rather specific models (Archontoulis and Miguez, 2015).

Programming languages, such as SAS or R, usually perform non-linear models. For instance, regression analysis can be performed in the base R language using functions such as lm, nls or glm. Nevertheless, non-linear models performed by the nls function require the prior specification of values to obtain model coefficient estimates, which is time-consuming. Implementing the R package may help carry out these analyses, as it is more straightforward and accessible for users. Thus, this work presents regression models found in scientific studies and implements them as an R package called AgroReg.

Materials and Methods

Creation of the *AgroReg* package

The package was built using the R (version 4.1.0) language (R Core Team, 2021), and documentation and checks were generated by the devtools packages (Wickham et al., 2021) and roxygen2 (Wickham et al., 2021) to facilitate the construction and adequacy of the

CRAN policy. The drc (Ritz et al., 2015) and ggplot2 (Wickham et al., 2016) packages were imported and added as a dependency, the first for logistic regression analysis and the second for graphical representation. Other packages (boot, minpack.lm, dplyr, rcompanion, and broom) have also been added as a dependency to make it easier to build functions.

Installation

All developed functions were written in the R programming language; therefore, they can be executed in the R environment or any GUI (Graphical User Interface) that uses this language, such as RStudio (https://www.rstudio.com/). R can be installed on Windows, Linux, or Mac systems. Thus, the scientific community can freely use this package regardless of the operating system. The AgroReg package can be installed from the CRAN repository using the following command:

install.packages("AgroReg", dependencies = TRUE)

The following command must be run to load the package:

library(AgroReg)

The package documentation can be accessed via the link: https://cran.r-project.org/web/packages/AgroReg/AgroReg.pdf

Data set

The collection of functions available in the AgroReg package implements several methods to describe many of the phenomena observed in quantitative studies in agricultural sciences, as mentioned in Table 1. Thus, the data obtained in real experiments were implemented to better exemplify the applications of the package. The use of the functions available in the package and the interpretation of their results are best presented in the form of an applied example using real data.

"aristolochia"

The data for exemplifying the functions "LM", "LM_i", "LM13", "LM13i", "LM23", "LM23i", "LM2i3", "logistic", "LL", "CD", "BC", "GP", "SH", "gaussianreg", "loessreg", "newton", "valcam", and "VG" come from an experiment conducted at the Seed Analysis Laboratory of the Center for Agricultural Sciences at the State University of Londrina (UEL) (23°19′42.8″ S, 51°12′11.9″ W, altitude 580 m), in which five temperatures (15, 20, 25, 30, and 35 °C) were assessed as to their effect on the germination of *Aristolochia elegans*. The experiment was conducted in a completely randomized design with twelve replicates of 25 seeds each. Data can be accessed by the command *data("aristolochia")*.

"granada"

The "granada" dataset represents partial data from an experiment conducted at UEL to evaluate the drying kinetics of pomegranate peel over time. Mass loss was assessed at 60, 210, 390, 720, 930, 1410, 1890, 2370 min after the beginning of the experiment. This dataset was used to exemplify the following functions: "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", "lorentiz", and "mitscherlich" (Table 1). Data can be accessed by the command *data("granada")*.

Regression models

All regression models implemented in the package are shown in Table 1, in addition to functions and descriptions, as well as applications in articles in the field of agricultural sciences. The models were grouped into non-parametric (loess), polynomial, logistic or S-shaped, logarithmic, bell-shaped, segmented, and exponential models. They were primarily extracted from scientific journals with original works such as the one from Sadeghi et al. (2019) or review articles, including the one written by Archontoulis and Miguez (2015), aiming to cover as many regression models as possible. Furthermore, modifications of specific models were also implemented.

Polynomial models, also called linear models, were implemented from the lm base R function. The same procedure was used to obtain the logarithmic curves, the Valcam model, and specific exponential models. On the other hand, the non-parametric loess regression, also known as local regression, was performed using the loess base R function.

Logistic equation models, also called sigmoid curves, are S-shaped and mainly used to describe plant growth curves, seed germination over time, or herbicide dose-response studies (Archontoulis and Miguez, 2015). They are implemented in the drc (Ritz et al., 2015) and aomisc (https://github.com/OnofriAndreaPG/aomisc) packages. Thus, in AgroReg, these functions were imported and summarized in a more straightforward function with more information.

Finally, nls from the stats package was used for the other functions, relying on the methodology of ordinary least squares. In these cases, pre-established algorithms were used to automate the initial information. Thereby, most functions do not require initial information to generate models, although the problematic convergence of coefficients may occur, owing to not estimating good initial values. In such a situation, the user can specify a priori information by the "initial" argument, according to each regression model.

Table 1 – Functions, descriptions, mathematical model, and applications of the models implemented in the AgroReg package.

Function	Description	Model	Applications
Descriptive			
Nreg	Descriptive graphic	-	-
Non-parametric			
loess_model	Local loess non-parametrical of 0, 1, or 2 degree	-	-
Polynomial or linear models			
LM	Simple linear, quadratic, inverse quadratic, cubic or quartic	$y = \beta_o + \beta_1 x$ $y = \beta_o + \beta_1 x + \beta_2 x^2$ $y = \beta_o + \beta_1 x + \beta_2 x^{0.5}$ $y = \beta_o + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ $y = \beta_o + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$	Silicon doses and their influence on tomato post-harvest durability and quality (Marodin et al., 2016), *Chrysanthemum leucanthemum* seed imbibition curve (Pêgo et al., 2012), potato yield according to K_2O doses (Maingi and Mbuvi, 2020)
LM_i	Simple linear, quadratic, inverse quadratic, cubic, or quartic without intercept	$y = \beta_1 x$ $y = \beta_1 x + \beta_2 x^2$ $y = \beta_1 x + \beta_2 x^{0.5}$ $y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ $y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$	Drying kinetics of *Cydonia oblonga* (Tzempelikos et al., 2015)
LM13	Cubic without β_2	$y = \beta_o + \beta_1 x + \beta_3 x^3$	*Merremia aegyptia* straw dose in vegetable cowpea (Silva et al., 2020)
LM13i	Inverse cubic without β_2	$y = \beta_0 + \beta_1 x + \beta_3 x^{\frac{1}{3}}$	-
LM23	Cubic without β_1	$y = \beta_o + \beta_2 x^2 + \beta_3 x^3$	*M. aegyptia* straw dose in vegetable cowpea (Silva et al., 2020)
LM23i	Inverse cubic without β_1	$y = \beta_0 + \beta_2 x^{\frac{1}{2}} + \beta_3 x^{\frac{1}{3}}$	-
LM2i3	Cubic without β_1 and with inverse β_3	$y = \beta_0 + \beta_2 x^2 + \beta_3 x^{\frac{1}{3}}$	-
valcam	Valcam	$y = \beta_o + \beta_1 x + \beta_2 x^{1.5} + \beta_3 x^2$	Drying kinetics of *Bauhinia forficata* (Silva et al., 2017)
Logistic, sigmoid, or S-shaped models			
logistic	Logistic with three (npar="L.3"), four (npar="L.4"), or five (npar="L.5") parameters	$y = \frac{d}{1+e^{b(x-e)}}$ $y = c + \frac{d-c}{1+e^{b(x-e)}}$ $y = c + \frac{d-c}{1+e^{b(x-e)^f}}$	*Eleusine indica* germination curve (Kerr et al., 2018) and growth curve of satsuma mandarin (*Citrus unshiu* Marc.) (Yano et al., 2018) and strawberry production curve (Diel et al., 2019)
LL	Log-logistic with three (npar="LL.3"), four (npar="LL.4"), or five (npar="LL.5") parameters	$y = \frac{d}{1+e^{b(\log(x)-\log(e))}}$ $y = c + \frac{d-c}{1+e^{b(\log(x)-\log(e))}}$ $y = c + \frac{d-c}{1+e^{b(\log(x)-\log(e))^f}}$	Fungicide dose-response (Noel et al., 2018), Tomato brown rugose fruit virus (ToBRFV) progress in tomato cultivars (González-Concha et al., 2021)
BC	Brain-Cousens with four (npar="BC.4") or five parameters (npar="BC.5")	$y = \frac{d+fx}{1+e^{b(\log(x)-\log(e))}}$ $y = c + \frac{d-c+fx}{1+e^{b(\log(x)-\log(e))}}$	Herbicide dose-response (Schabenberger et al., 1999), *Trichoderma asperellum* doses in wheat seeds (Couto et al., 2021)
CD	Cedergreen-Ritz-Streibig with four (npar="CRS.4") or five (npar="CRS.5") parameters	$y = \frac{d+fe^{-\frac{1}{x}}}{1+e^{b(\log(x)-\log(e))}}$ $y = c + \frac{d-c+fe^{-\frac{1}{x}}}{1+e^{b(\log(x)-\log(e))}}$	Kiwi fruit drying kinetics (Sadeghi, et al., 2019) and dose-response of flufenacet and pendimethalin to control *Alopecurus myosuroides* (Metcalfe et al., 2017)

Continue.

Table 1 – Continuation.

weibull	Weibull with three (npar="w3") or four (npar="w4") parameters	$y = de^{-e^{b(\log(x)-\log(e))}}$ $y = c + (d-c)(e^{-e^{b(go(x)-\log(e))}})$	Substrate water retention curve (Bateman et al., 2019)
GP	Gompertz with two (npar="g2"), three (npar="g3"), or four quatro (npar="g4") parameters	$y = e^{-e^{-e^{b(x-e)}}}$ $y = d\ e^{-e^{-e^{b(x-e)}}}$ $y = c + (d-c)(e^{-e^{-e^{b(x-e)}}})$	Growth curve (Fang et al., 2022) and response-dose (Mendes et al., 2019)
VB	Von Bertalanffy	$y = L(1 - e^{-k(x-t0)})$	Strawberry production curve (Diel et al., 2019)
lorentz	Lorentz with three ("lo3") or four ("lo4") parameters	$y = \frac{d}{1 + b(x-e)^2}$ $y = c \frac{d-c}{1 + b(x-e)^2}$	-
Bell-shaped models			
beta_reg	Beta regression, developed by Yin et al. (1995)	$y = d\left\{\left(\frac{X - X_b}{X_o X_b}\right)\left(\frac{X_c - X}{X_c X_o}\right)^{\frac{X_c - X_o}{X_o - Xb}}\right\}^b$	Cardinal temperature estimate for cultivars of Quinoa (Mamedi et al., 2017) and *Alyssum homolocarpum* (Zaferanieh et al., 2020)
gaussianreg	Function analogous to Gaussian distribution or Bragg model	$y = d\ e^{-b(x-e)^2}$ $y = c + (d-c)\ e^{-b(x-e)^2}$	-
Segmented models			
linear.linear	Linear-linear segmented	$y = \beta_0 + \beta_1 x\ (x < bp)$ $y = \beta_0 + \beta_1 bp + wx\ (x > bp)$	Cardinal and optimal temperature estimate for seed germination (Motsa et al., 2015). Relationship of phosphorus content with wheat and maize yield (Xi et al., 2016)
linear.plateau	Linear-plateau segmented	$y = \beta_0 + \beta_1 x\ (x < bp)$ $y = \beta_0 + \beta_1 bp\ (x > bp)$	Estimate of optimal minimum plant density (Ferreira et al., 2020), onion bulb yield according to nitrogen doses (Gonçalves et al., 2019)
plateau.linear	Plateau-linear segmented	$y = \beta_0 + \beta_1 bp\ (x < bp)$ $y = \beta_0 + \beta_1 x\ (x > bp)$	Relationship between average daily temperature and weight and sunflower oil content (Angeloni et al., 2021)
quadratic. plateau	Quadratic-plateau segmented	$y = \beta_0 + \beta_1 x + \beta_2\ x^2\ (x < bp)$ $y = \beta_0 + \beta_1 bp + \beta_2 bp^2\ (x > bp)$	Density (Van Roekel and Coulter, 2011) or population of maize plants (Williams et al., 2021)
plateau. quadratic	Plateau-quadratic segmented	$y = \beta_0 + \beta_1 x + \beta_2 x^2\ (x > bp)$ $y = \beta_0 + \beta_1 bp + \beta_2 bp^2\ (x < bp)$	-
Logarithmic models			
LOG	Logarithmic	$y = \beta_0 + \beta_1 ln(x)$	Kiwi fruit drying kinetics (Sadeghi et al., 2019)
LOG2	Quadratic logarithmic	$y = \beta_0 + \beta_1 ln(x) + \beta_2 ln(x)^2$	-
thompson	Thompson or quadratic logarithmic without intercepto	$y = \beta_0\ \beta_1 ln(x) + \beta_2 ln(x)^2$	Kiwi fruit drying kinetics (Sadeghi et al., 2019)
Exponential models or equations that show exponential growth characteristics			
asymptotic	Logarithmic or Asymptotic	$y = \alpha\ e^{-\beta x} + c$	Kiwi fruit drying kinetics (Sadeghi et al., 2019)
asymptotic_i	Henderson and Pabis or exponential bi-parametric model	$y = \alpha\ e^{-\beta x}$	Orange seed drying (Rosa et al., 2015)
asymptotic_ineg	Negative asymptotic without intercepto	$y = -\alpha\ e^{-\beta x}$	-
biexponential	Biexponential	$y = A1\ e^{-e^{lrc1x}} + A2\ e^{-e^{lrc2x}}$	Atrazine degradation in soil (Swarcewicz and Gregorczyk, 2013)
mitscherlich	Mitscherlich's Law	$y = A\ (1 - 10^{-eb-ex})$	Evaluation of the effect of pH on phosphate availability (Barrow et al., 2020)
yieldloss	Yield loss	$y = \frac{ix}{1 + \frac{i}{A}x}$	Relationship between weed density and yield loss (Cousens, 1985)

Continue.

Table 1 – Continuation.

hill	Hill	$y = \frac{a \times x^c}{b \times x^c}$	-
MM	Michaelis-Menten or rectangular hyperbola	$y = \frac{Vm \times x}{k + x}$ $y = c + \frac{Vm \times x}{k + x}$	Pomegranate peel drying kinetics (Shimizu et al., 2020)
SH	Steinhart-Hart	$y = \frac{1}{A + B\ln(x) + C[\ln(x)]^3}$	Density ratio of maize plants to grain dry mass (Zhai et al., 2021)
PAGE	Page	$y = e^{-kx^n}$	Drying kinetics of *Cydonia oblonga* (Tzempelikos et al., 2015), and Kiwi (Sadegui et al., 2019)
newton	Newton ou Lewis	$y = e^{-kx}$	Kiwi (Sadegui et al., 2019) and onion drying kinetics (Sharma et al., 2005)
potential	Potential	$y = \beta_1 x^n$	Kiwi fruit drying kinetics (Sadegui et al., 2019)
midilli	Midilli	$y = \alpha e^{kx^n} + bx$	Kiwi fruit drying kinetics (Sadeghi et al., 2019) Effects of drying conditions on the functional and physical quality of dry autumn olives (Zannou et al., 2021)
midillim	Midilli modified	$y = \alpha e^{kx} + bx$	Kiwi fruit drying kinetics (Sadeghi et al., 2019)
AM	Avhad and Marchetti	$y = \alpha e^{kx^n}$	Drying kinetics of Hass avocado seeds (Avhad and Marchetti, 2016)
peleg	Peleg	$y = \frac{1 - x}{a + bx}$	Effects of drying conditions on the functional and physical quality of dry autumn olives (Zannou et al., 2021)

Statistical information and parameters

The functions were developed to provide estimates of the maximum and minimum predicted values obtained in the curve within the studied range. In addition, statistical parameters such as AIC (Akaike's information criterion), BIC (Bayesian information criterion), R^2 (coefficient of determination) or Pseudo-R^2 (correlation between observed and predicted outcome), RMSE (root mean square error), and *p*-value from the *t*-test of coefficients were also returned. In the case of polynomial models, the variance inflation factor (VIF) is also given.

The root mean square error is calculated by the following formula:

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

where $\hat{Y}_i$ is the response predicted by the model, Y_i the observed response, and n the sample size.

The Akaike's Information (AIC) and Bayesian Information (BIC) Criteria are calculated by the formula:

$$AIC_i = -2\log L_i + 2p_i \quad (2)$$

$$BIC_i = -2\log L_i + p_i \log n \quad (3)$$

where: L_i and p_i are the likelihood function and number of parameters for each model, and n the number of observations.

The VIF is calculated using the formula:

$$VIF_j = \frac{1}{1 - R_j^2} \; j = 1, 2, \ldots, p \quad (4)$$

where: p is the number of predictor variables; R_j^2 the multiple correlation coefficient, resulting from the X_j regression on the other p-1 regressors.

Other goodness-of-fit statistical parameters such as MBE – mean bias error, MBER – relative mean bias error, MAE – mean absolute error, RMAE – relative mean absolute error, SE – standard error, MSE – mean squared error, rMSE – relative mean square error, EF – modeling efficiency, SD – standard deviation of differences, and CRM – coefficient of residual mass are provided separately from the analyses, through the "stat_param" function. Graphical analysis of residuals can be performed by the command *"extract.model"* as follows:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right) \quad (5)$$

$$RMBE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)}{\hat{Y}_o} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{Y}_i - Y_i\right| \quad (7)$$

$$RMAE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|\hat{Y}_i - Y_i\right|}{\hat{Y}_i} \quad (8)$$

$$SE = \sum_{i=1}^{n}\left(\hat{Y}_i - Y\right)^2 \quad (9)$$

$$MSE = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{n} \quad (10)$$

$$rMSE = \frac{\frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{n}}{\hat{Y}_o} \tag{11}$$

$$EF = 1 - \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}_i\right)^2} \tag{12}$$

$$SD = \frac{\sum^{n}(\in - \bar{\in})^2}{n-1} \tag{13}$$

$$CRM = \frac{\bar{\in}}{\bar{Y}_i} \tag{14}$$

where: $\hat{Y}_i$ is the response predicted by the model, Y_i the observed response, $\bar{Y}_i$ the mean of the observed response, $\bar{\in}$ the mean of the difference between the predicted and the observed response, and n the sample size.

Results and Discussion

General information

The package has 44 regression analysis functions, which can also be accessed using the "regression" function and defining the "model" argument according to the requested regression model (Table 1). This function has the simple linear model (*model = "LM1"*) by default, as follows:

> *data("aristolochia")*

> *with(aristolochia, regression(trat, resp, model = "LM1"))*

For more information, access the documentation for the function (*"?regression"*).

In all analysis functions, the first two arguments are mandatory, representing the independent variable and dependent variable, respectively. In the case of polynomial functions (LM and LM_i), the argument "degree" defines the polynomial degree, while for logistic functions and some exponentials, such as logistic, LL, BC, CD, GP, weibull, lorentz, and MM, the argument "npar" sets the number of parameters.

Figures 1A and B show the plot of the simple linear regression analysis and Brain-Cousens four-parameter logistic model, respectively. Curves joining can be accessed by the "plot_arrange" function (Figure 1C), requiring a list with the outputs of each analysis as the only mandatory argument, as follows:

> *data("aristolochia")*

> *reg1 = with (aristolochia, LM(trat, resp))*

> *reg2 = with(aristolochia, BC(trat, resp))*

> *plot_arrange(list(reg1, reg2))*

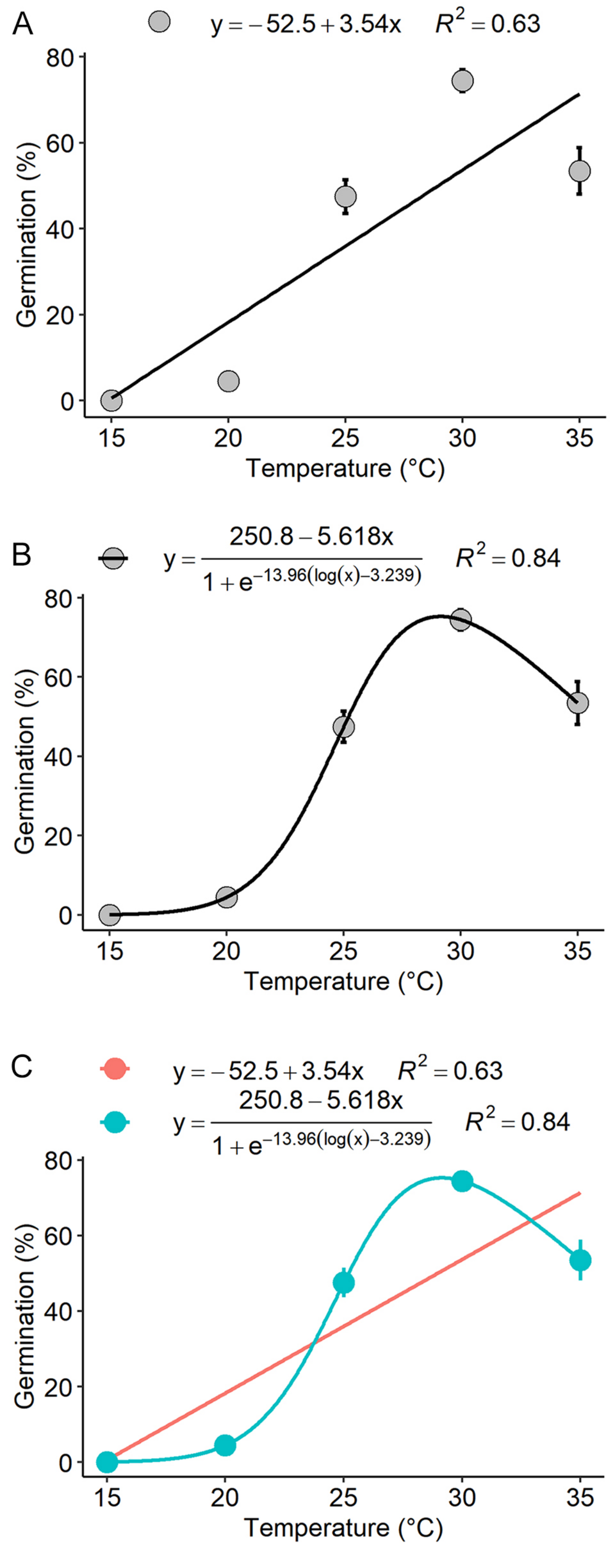

Figure 1 – Exemplification of the output of a linear (A) and Brain-Cousens logistic (B) function and union of the curves in a plot (C) using the functions from the AgroReg package for the "aristolochia" dataset of the germination of seeds of *Aristolochia elegans* depending on the temperature.

The graphical representation synthesis is a laborious and error-prone step, especially when the user needs more experience with the R language. Thus automatically providing the graphics such as equations and the coefficient of determination (R^2 or Pseudo R^2) avoids errors made by the researcher and optimizes time in this process within the regression analysis. In addition, the user can also change several graphic parameters such as shape, size, and markup color; titles and text size of axes; plotting standard error bars, standard deviation, or with no bars; and equations position among other arguments (for help, access *"?AgroReg"*).

Finally, the package provides essential information on selecting regression models such as AIC and BIC. For both statistical criteria, a lower value indicates a preferable model. BIC differs from AIC only in the second term of the equation which depends on n. Thus, as n increases, BIC favors the simpler models (fewer parameters), which is why, sometimes, AIC and BIC indices disagree (Archontoulis and Miguez, 2015). In addition, the coefficient of determination (R^2) is also returned, in which values close to 1 are desirable, although, in the case of linear models ("LM" function), attention should be paid to the problem of multicollinearity, which is evaluated as VIF in the function and should be less than 5 or 10 according to Myers and Montgomery (2002) and Petrini et al. (2012), respectively. The information can be summarized in a table using the *"comparative_model"* function and inserting a list with the variables returned in each analysis function.

Applied example

To exemplify and guide the use of the AgroReg package and interpret the results generated, an applied example with the dataset *"granada"* was inserted. The first step of any statistical analysis is to study descriptive exploratory information, obtaining, for example, position measures such as mean, median, maximum, minimum, and measures of dispersion such as variance and standard deviation. On the other hand, in the case of regression analyses, a procedure that must be carried out in advance is the graphical visualization of the results (Archontoulis and Miguez, 2015) because, with such information, it is possible to identify patterns and thus target specific models, avoiding unnecessary processes and clearing the path to reach a biologically acceptable explanation. This critical stage was further explored from the dataset known as the Anscombe quartet, proposed by the statistician Francis Anscombe in 1973, who observed that in four datasets, identical fitted and regression coefficients were produced; however, when viewed graphically, they revealed surprisingly different patterns of covariation between x and y.

In the case of the "granada" dataset, exploratory plots using the "Nreg" function were generated (Figure 2A and B). The dataset exhibited a visually noticeable low variability and a sharp rise in growth up to 1000 min with subsequent stability. This behavior is like exponential models or models that behave similarly, such as Michaelis-Menten, logistic models, or Mitscherlich. Segmented models, such as linear-linear, linear-plateau, and quadratic-plateau models, are also used to explain this behavior. Next, the routine of these functions and the appearance of the curve (Figure 3) is presented.

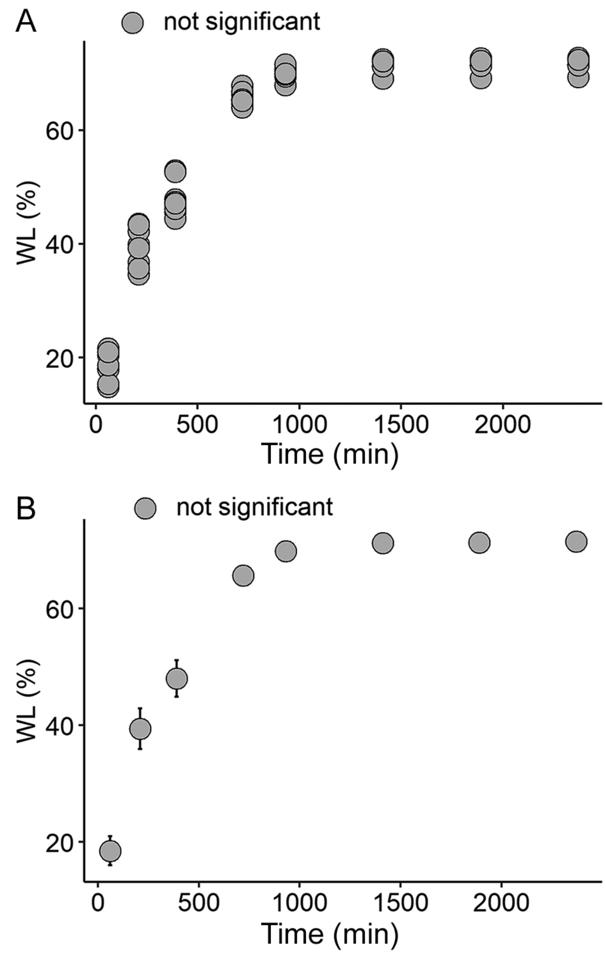

Figure 2 – Visualization of all observations by scatter plot (A) and mean and standard deviation (B) for the "granada" dataset. WL = weight loss, Time (min) = Time after pomegranate peel begins to dry. The function returns a "not significant" label as this function can be used to represent the absence of a trend when you want to join the plots.

```
> 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,")"))
```

After obtaining the model, the next step is the analysis of the residues. In AgroReg, this analysis can be performed graphically, as follows:

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

Based on the theoretical quantile graph (Figure 4A-J), all models presented points close to the normal distribution curve, even though there are better-fitted models, such as three-parameter log-logistic and four-parameter Brain-Cousens. Table 2 presents the statistical parameters of each model used in Figure 4A-J. In addition, the package also implemented bar graphs that facilitate the visualization of the model choice parameters (Figure 5). Thus, in this example, the biexponential model had the lowest AIC (307.46) and BIC (318.44) values and was among the models with the lowest RMSE and higher R^2, in addition to presenting all significant coefficients by the *t*-test ($p < 0.05$). However, although there are models statistically more adequate, almost all the models used could be applied to explain the behavior of this study.

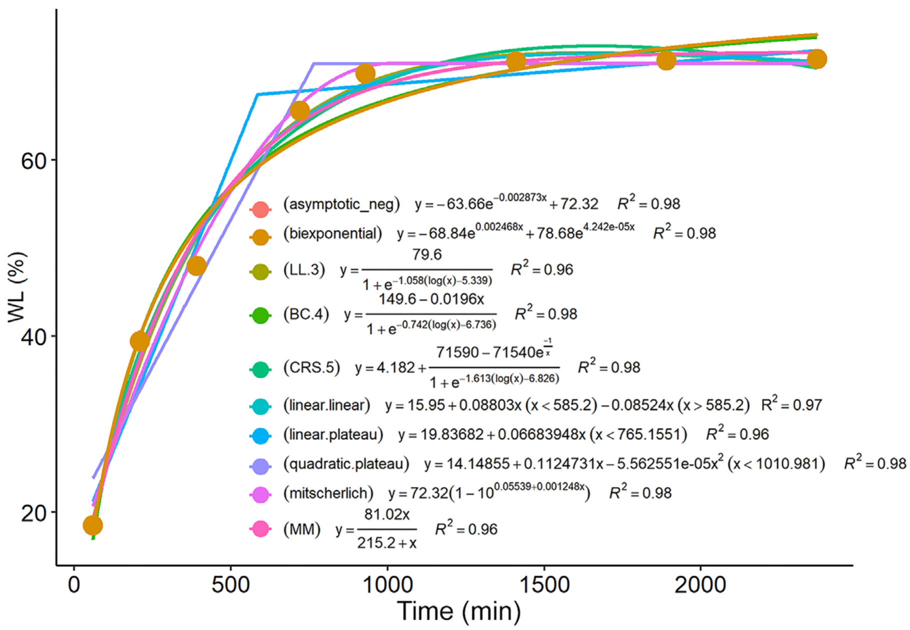

Figure 3 – Regression plot with the ten regression models used to exemplify the commands in the AgroReg package for the "granada" dataset. WL = weight loss, Time (min) = Time after pomegranate peel begins to dry.

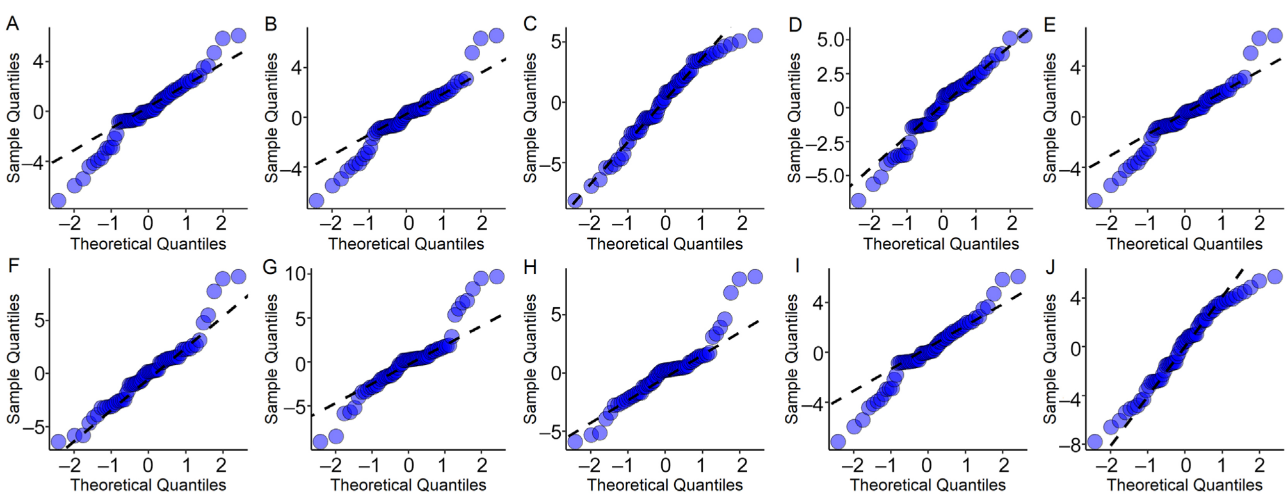

Figure 4 – QQ-plot plot of model residuals generated by the AgroReg package: negative asymptotic (A), biexponential (B), three-parameter log-logistic (C), four-parameter Brain-Cousens (D), five-parameter Cedergreen-Ritz (E), segmented linear-linear (F), segmented linear-plateau (G), segmented quadratic-plateau (H), Mistcherlich (I), and two-parameter Michaelis-Menten (J). Residuals generated from the models used in the "granada" dataset for the variable WL (weight loss) as a function of drying time.

Table 2 – Regression model, AgroReg package function, coefficient of determination (R^2), Akaike's information criterion (AIC), Bayesian information criterion (BIC), root mean square error (RMSE), and coefficients of the model for the example "granada" dataset for the variable WL (weight loss) as a function of drying time.

Model	Function	R^2	AIC	BIC	RMSE	Parameters
Negative asymptotic	asymptotic_neg	0.98	310.34	318.97	2.568	Alpha = –63.66** Beta = –0.00287** Theta = 72.32**
Biexponential	biexponential	0.98	307.46	318.44	2.475	A1 = –68.845** lrc1 = –6.004** A2 = 78.6817** lrc2 = –10.06778**
Log-logistic with 3 parameters	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* f = –0.0196017**
Cedergreen-Ritz	CD	0.98	308.01	320.97	2.444	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-plateau	linear.plateau	0.96	351.47	360.11	3.54	a = 19.84** b = 0.0668** c = 765**
Quadrático-plateau	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 significant at 5 %, 1 %, and not significant by the *t*-test, respectively.

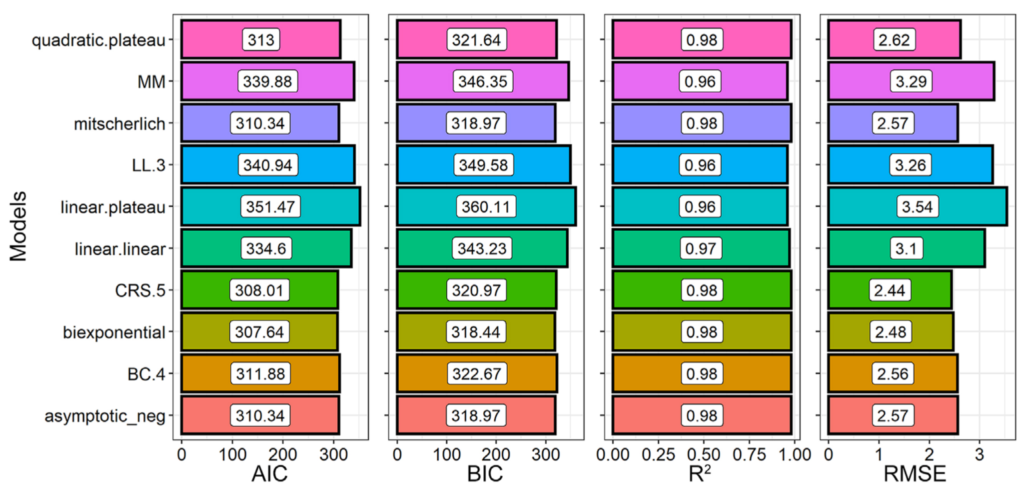

Figure 5 – Graphic representation of comparisons between regression models according to the criteria of AIC (Akaike information criterion), BIC (Bayesian Information Criterion), R^2 (coefficient of determination), and RMSE (root mean square error). The information is generated through the models used in the "granada" dataset for the variable WL (weight loss) as a function of drying time.

AgroReg is a promising analysis tool in agricultural research because it has simple functions that allow for fast and efficient data processing, aiming to offer greater reliability and relevant information. In addition, new functions and updates will be carried out to improve the package to meet the demands of the scientific community.

Authors' Contributions

Conceptualization: Shimizu GD, Gonçalves LSA. **Data acquisition**: Shimizu GD. **Investigation**: Shimizu GD. **Software**: Shimizu GD. **Supervision**: Shimizu GD, Gonçalves LSA. **Validation**: Shimizu GD. **Writing–original draft**: Shimizu GD. **Project administration**: Gonçalves LSA. **Writing–review & editing**: Gonçalves LSA.

References

Angeloni P, Aguirrezábal L, Echarte MM. 2021. Assessing the mechanisms underlying sunflower grain weight and oil content responses to temperature during grain filling. Field Crops Research 262: 108040. https://doi.org/10.1016/j.fcr.2020.108040

Anscombe FJ. 1973. Graphs in statistical analysis. The American Statistician 27: 17-21.

Archontoulis SV, Miguez FE. 2015. Nonlinear regression models and applications in agricultural research. Agronomy Journal 107: 786-798. https://doi.org/10.2134/agronj2012.0506

Avhad MR, Marchetti JM. 2016. Mathematical modelling of the drying kinetics of Hass avocado seeds. Industrial Crops and Products 91: 76-87. https://doi.org/10.1016/j.indcrop.2016.06.035

Banzatto DA, Kronka SN. 2013. Agricultural Experimentation = Experimentação Agrícola. 4ed. FUNEP, Jaboticabal, SP, Brazil (in Portuguese).

Barrow NJ, Debnath A, Sen A. 2020. Measurement of the effects of pH on phosphate availability. Plant and Soil 454: 217-224. https://doi.org/10.1007/s11104-020-04647-5

Bateman AM, Erickson TE, Merritt DJ, Veneklaas EJ, Muñoz-Rojas M. 2019. Water availability drives the effectiveness of inorganic amendments to increase plant growth and substrate quality. Catena 182: 104116. https://doi.org/10.1016/j.catena.2019.104116

Bretz F, Hothorn T, Westfall P. 2011. Multiple Comparisons Using R. CRC Press, Boca Raton, FL, USA.

Cochran WG, Cox GM. 1986. Experimental Design. 2ed. John Wiley, New York, NY, USA.

Cousens R. 1985. A simple model relating yield loss to weed density. Annals of Applied Biology 107: 239-252. https://doi.org/10.1111/j.1744-7348.1985.tb01567.x

Couto APS, Pereira AE, Abati J, Camargo Fontanela ML, Dias-Arieira CR, Krohn NG. 2021. Seed treatment with *Trichoderma* and chemicals to improve physiological and sanitary quality of wheat cultivars. Revista Caatinga 34: 813-823. http://dx.doi.org/10.1590/1983-21252021v34n408rc

Diel, MI, Sari, BG, Krysczun, DK, Olivoto T, Pinheiro, MVM, Meira, D, et al. 2019. Nonlinear regression for description of strawberry (*Fragaria* × *ananassa*) production. The Journal of Horticultural Science and Biotechnology 94: 259-273. https://doi.org/10.1080/14620316.2018.1472045

Fang SL, Kuo YH, Kang L, Chen CC, Hsieh CY, Yao MH, et al. 2022. Using Sigmoid Growth Models to Simulate Greenhouse Tomato Growth and Development. Horticulturae 8: 1021. https://doi.org/10.3390/horticulturae8111021

Ferreira AS, Zucareli C, Werner F, Fonseca ICB, Balbinot Junior AA. 2020. Minimum optimal seeding rate for indeterminate soybean cultivars grown in the tropics. Agronomy Journal 112: 2092-2102. https://doi.org/10.1002/agj2.20188

Gonçalves FDC, Grangeiro LC, Sousa VFL, Santos JP, Souza FI, Silva LRR. 2019. Yield and quality of densely cultivated onion cultivars as function of nitrogen fertilization. Revista Brasileira de Engenharia Agrícola e Ambiental 23: 847-851. http://dx.doi.org/10.1590/1807-1929/agriambi.v23n11p847-851

González-Concha LF, Ramírez-Gil JG, García-Estrada RS, Rebollar-Alviter Á, Tovar-Pedraza JM. 2021. Spatiotemporal analyses of tomato brown rugose fruit virus in commercial tomato greenhouses. Agronomy 11: 1268. https://doi.org/10.3390/agronomy11071268

Hsu J. 1996. Multiple Comparisons: Theory and Methods. Chapman & Hall, London, UK. https://doi.org/10.1201/b15074

Kerr RA, McCarty LB, Bridges WC, Cutulle M. 2018. Key morphological events following late-season goosegrass (*Eleusine indica*) germination. Weed Technology 33: 196-201. https://doi.org/10.1017/wet.2018.93

Lúcio ADC, Sari BG. 2017. Planning and implementing experiments and analyzing experimental data in vegetable crops: problems and solutions. Horticultura Brasileira 35: 316-327. https://doi.org/10.1590/S0102-053620170302

Maingi FM, Mbuvi HM. 2020. The effect of potassium dosage on selected growth parameters and yield response modeling on potatoes grown in Molo, Kenya. Journal of Scientific Agriculture 4: 101-107. https://doi.org/10.25081/jsa.2020.v4.6388

Mamedi A, Afshari RT, Oveisi M. 2017. Cardinal temperatures for seed germination of three quinoa (*Chenopodium quinoa* Willd.) cultivars. Iranian Journal of Field Crop Science 48: 89-100. https://doi.org/10.22059/ijfcs.2017.206204.654106

Marodin JC, Resende JTV, Morales RGF, Faria MV, Trevisam AR, Figueiredo AST, et al. 2016. Tomato post-harvest durability and physicochemical quality depending on silicon sources and doses. Horticultura Brasileira 34: 361-366. http://dx.doi.org/10.1590/S0102-05362016003009

Mendes RR, Oliveira Jr. RS, Constantin J, Silva VFV, Hencks JR. 2019. Identification and mapping of cross-resistance patterns to ALS-inhibitors in greater beggarticks (*Bidens* spp.). Planta Daninha 37: 1-10. https://doi.org/10.1590/S0100-83582019370100117

Metcalfe H, Milne AE, Hull R, Murdoch AJ, Storkey J. 2017. The implications of spatially variable pre-emergence herbicide efficacy for weed management. Pest Management Science 74: 755-765. https://doi.org/10.1002/ps.4784

Montgomery DC. 2017. Design and Analysis of Experiments. John Wiley, New York, NY, USA.

Motsa MM, Slabbert MM, Van Averbeke W, Morey L. 2015. Effect of light and temperature on seed germination of selected African leafy vegetables. South African Journal of Botany 99: 29-35. https://doi.org/10.1016/j.sajb.2015.03.185

Myers RH, Montgomery DC. 2002. Response Surface Methodology: Process and Product Optimization using Designed Experiments. 2ed. John Wiley, New York, NY, USA.

Noel ZA, Wang J, Chilvers MI. 2018. Significant influence of EC_{50} estimation by model choice and EC_{50} type. Plant Disease 102: 708-714. https://doi.org/10.1094/PDIS-06-17-0873-SR

Pêgo RG, Grossi JAS, Barbosa JG. 2012. Soaking curve and effect of temperature on the germination of daisy seeds. Horticultura Brasileira 30: 312-316. https://doi.org/10.1590/S0102-05362012000200021

Petrini J, Dias RAP, Pertile SFN, Eler JP, Ferraz JBS, Mourão GB. 2012. Degree of multicollinearity and variables involved in linear dependence in additive-dominant models. Pesquisa Agropecuária Brasileira 47: 1743-1750. https://doi.org/10.1590/S0100-204X2012001200010

Piepho HP, Edmondson RN. 2018. A tutorial on the statistical analysis of factorial experiments with qualitative and quantitative treatment factor levels. Journal of Agronomy and Crop Science 204: 429-455. https://doi.org/10.1111/jac.12267

Pimentel-Gomes F. 2009. Experimental Statistics Course = Curso de Estatística Experimental. 15ed. Nobel, Piracicaba, SP, Brazil (in Portuguese).

Ritz C, Baty F, Streibig JC, Gerhard D. 2015. Dose-response analysis using R. PloS One 10: e0146021. https://doi.org/10.1371/journal.pone.0146021

Rosa DP, Cantú-Lozano D, Luna-Solano G, Polachini TC, Telis-Romero J. 2015. Mathematical modeling of orange seed drying kinetics. Ciência e Agrotecnologia 39: 291-300. https://doi.org/10.1590/S1413-70542015000300011

Sadeghi E, Haghighi Asl A, Movagharnejad K. 2019. Mathematical modelling of infrared-dried kiwifruit slices under natural and forced convection. Food Science & Nutrition 7: 3589-3606. https://doi.org/10.1002/fsn3.1212

Schabenberger O, Tharp BE, Kells JJ, Penner D. 1999. Statistical tests for hormesis and effective dosages in herbicide dose response. Agronomy Journal 91: 713-721. https://doi.org/10.2134/agronj1999.914713x

Sharma G, Verma RC, Pathare P. 2005. Mathematical modeling of infrared radiation thin layer drying of onion slices. Journal of Food Engineering 71: 282-286. https://doi.org/10.1016/j.jfoodeng.2005.02.010

Shimizu GD, Paula JCB, Neves CSVJ, Pacheco CA. 2020. Quality of pomegranate peel cultivar Valenciana as determined by different drying methods. Revista Brasileira de Fruticultura 42: 1-12. https://doi.org/10.1590/0100-29452020431

Silva FP, Siqueira VC, Martins EAS, Miranda FMN, Melo RM. 2017. Thermodynamic properties and drying kinetics of *Bauhinia forficata* Link leaves. Revista Brasileira de Engenharia Agrícola e Ambiental 21: 61-67. https://doi.org/10.1590/1807-1929/agriambi.v21n1p61-67

Silva JN, Bezerra Neto F, Lima JSS, Barros Júnior AP, Nunes RLC, Chaves AP. 2020. Agronomic and economic feasibility indicators for cowpea-vegetable under green manure in a semiarid environment. Revista Ciência Agronômica 51: 1-11. https://doi.org/10.5935/1806-6690.20200040

Storck L, Lopes SJ, Estefanel V, Garcia DC. 2016. Plant Experimentation = Experimentação Vegetal. 3ed. Editora UFSM, Santa Maria, RS, Brazil (in Portuguese).

Swarcewicz MK, Gregorczyk A. 2013. Atrazine degradation in soil: effects of adjuvants and a comparison of three mathematical models. Pest Management Science 69: 1346-1350. https://doi.org/10.1002/ps.3510

Tzempelikos DA, Vouros AP, Bardakas AV, Filios AE, Margaris DP. 2015. Experimental study on convective drying of quince slices and evaluation of thin-layer drying models. Engineering in Agriculture, Environment and Food 8: 169-177. https://doi.org/10.1016/j.eaef.2014.12.002

Van Roekel RJ, Coulter JA. 2011. Agronomic responses of corn to planting date and plant density. Agronomy Journal 103: 1414-1422. https://doi.org/10.2134/agronj2011.0071

Wickham H, Danenberg P, Csárdi G, Eugster M. 2016. roxygen2: In-Line Documentation for R. R package version 7.1.1. https://CRAN.R-project.org/package=roxygen2

Wickham H, Hester J, Chang W, Bryan J. 2021. devtools: Tools to Make Developing R Packages Easier. R package version 2.4.2. https://CRAN.R-project.org/package=devtools

Williams JJ, Dodds DM, Buehring NW, Dhillon JS, Henry WB. 2021. Corn response to row spacing and plant population in the mid-south United States. Agronomy Journal 113: 4132-4141. https://doi.org/10.1002/agj2.20817

Xi B, Zhai LM, Liu J, Liu S, Wang HY, Luo CY, et al. 2016. Long-term phosphorus accumulation and agronomic and environmental critical phosphorus levels in Haplic Luvisol soil in northern China. Journal of Integrative Agriculture 15: 200-208. https://doi.org/10.1016/S2095-3119(14)60947-3

Yano T, Morisaki A, Matsubara K, Ito SI, Kitano M. 2018. Growth analysis of potted seedlings of Satsuma mandarin (*Citrus unshiu* Marc.) under different light conditions and air temperatures. Horticulture Journal 87: 34-42 https://doi.org/10.2503/hortj.OKD-051

Yin X, Kropff MJ, McLaren G, Visperas RM. 1995. A nonlinear model for crop development as a function of temperature. Agricultural and Forest Meteorology 77: 1-16. https://doi.org/10.1016/0168-1923(95)02236-Q

Zaferanieh M, Mahdavi B, Torabi B. 2020. 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 131: 259-266. https://doi.org/10.1016/j.sajb.2020.02.006

Zannou O, Pashazadeh H, Ghellam M, Hassan AM, Koca I. 2021. Optimization of drying temperature for the assessment of functional and physical characteristics of autumn olive berries. Journal of Food Processing and Preservation 45: 1-13. https://doi.org/10.1111/jfpp.15658

Zhai L, Li H, Song S, Zhai L, Ming B, Li S, et al. 2021. Intra-specific competition affects the density tolerance and grain yield of maize hybrids. Agronomy Journal 113: 224-235. https://doi.org/10.1002/agj2.20438