Birnbaum-Saunders Linear and Quantile Regression: A comparative application to Castor seeds' growth data

Tailine J. S. Nonato

Embrapa

Context

Castor seed is the source of castor oil (**óleo de rícino**), which is very famous for its **cosmetic use**.

This oil is widely used in the industry due to a **unique chemical characteristic** that no other commercially produced vegetable oil has.

Tailine J. S. Nonato

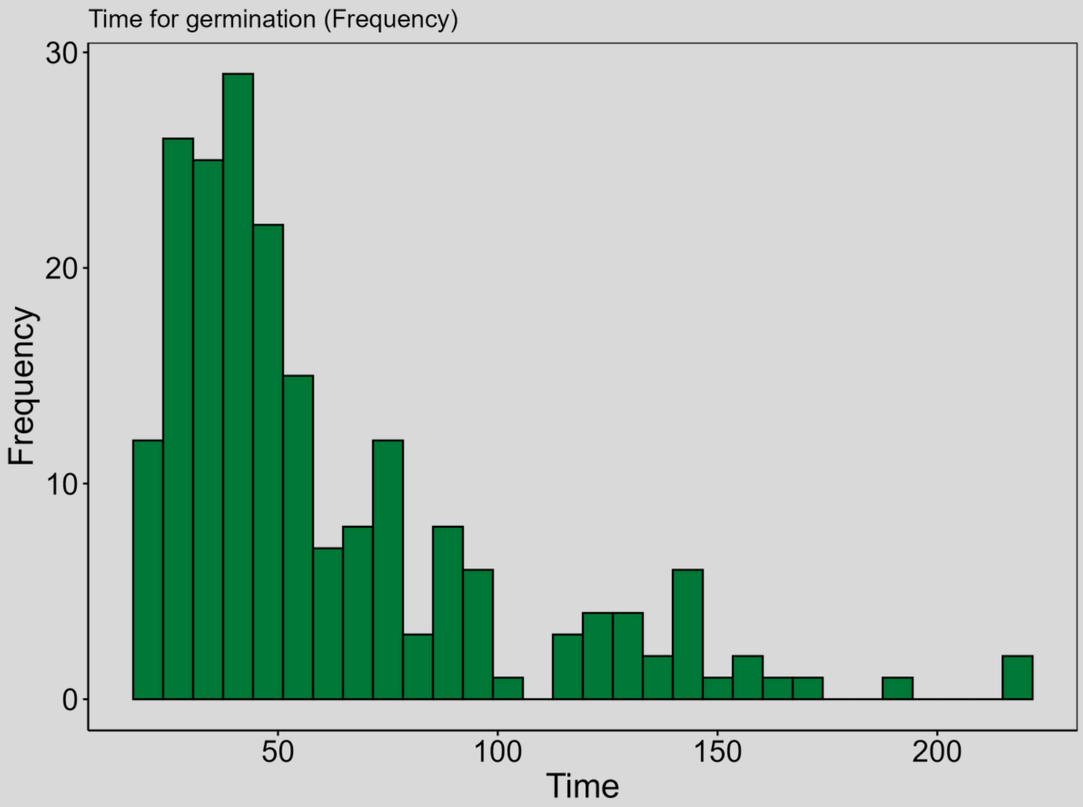
Dataset

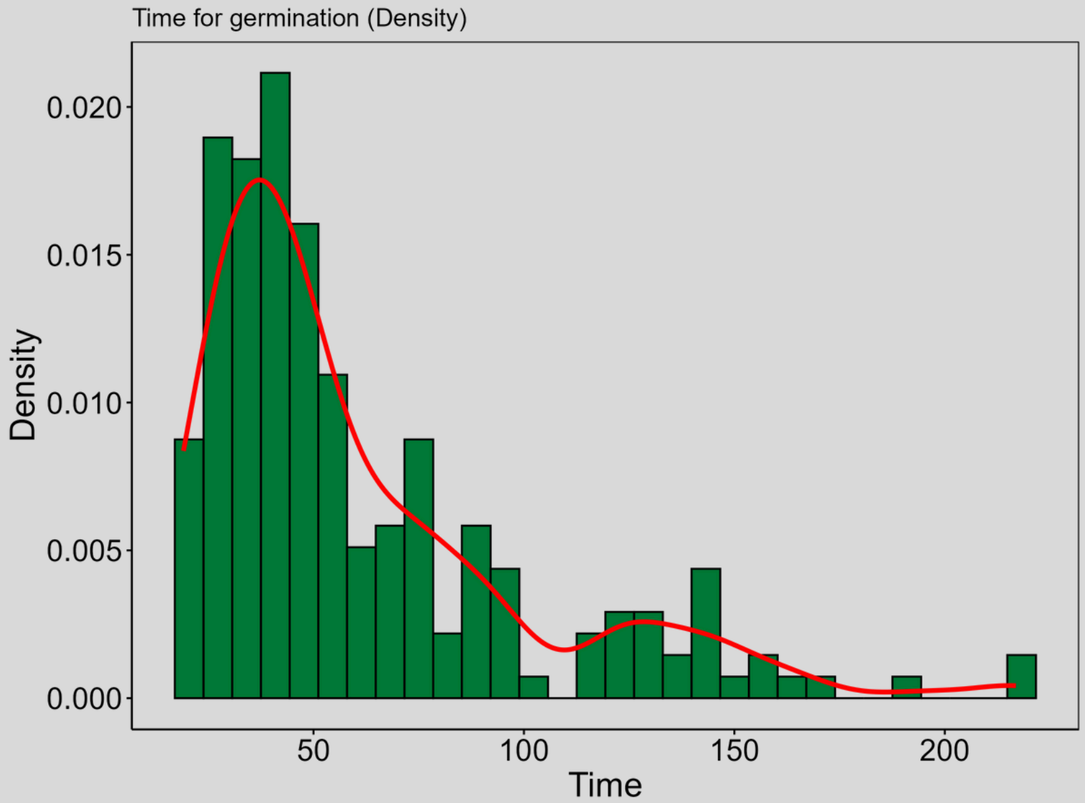
The data was collected in an experiment with objective to **measure the progress of hydration of castor seed under temperatures between 16 and 34°C**, and if making a puncture in the seed coat would make the hydration to occur faster. The **time for germination** was also measured (SEVERINO, 2024).

Tailine J. S. Nonato

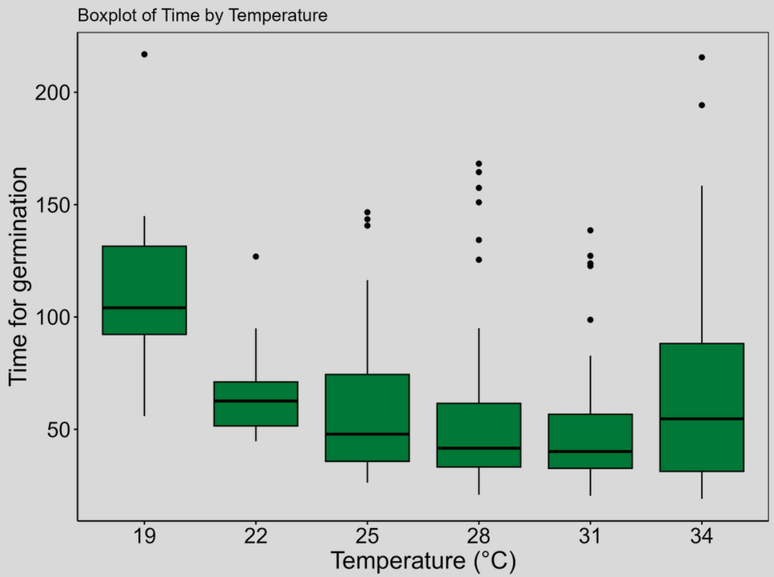
Embrapa

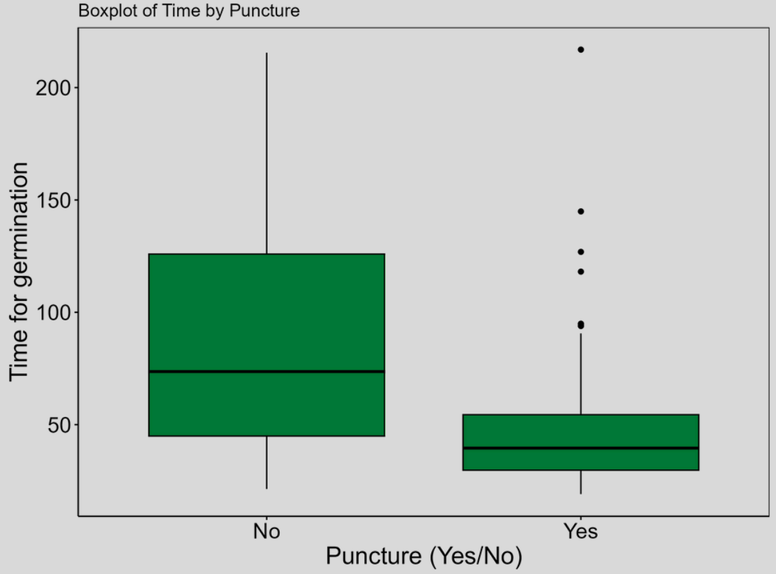
Exploratory Data Analysis

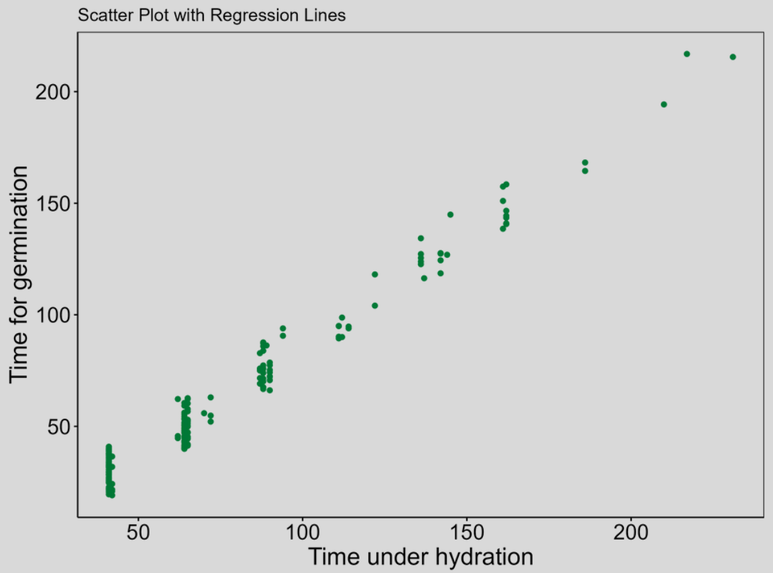




Exploratory Data Analysis







Tailine J. S. Nonato

Proposal

Hypothesis

The time of germination of Castor seeds be explained by the variables related to hydration and puncture of the seed.

Objectives

- Fit a linear regression model based on the Birnbaum-Saunders distribution;
- Fit a quantile regression model based on the Birnbaum-Saunders distribution;
- Compare the two models to see which one fits better to the study.

Tailine J. S. Nonato

Methods

Variables

Interest variable: time of germination.
Explanatory variables: Puncture (yes/no), Temperature and Time under hydration.

Software

R, Version 4.5.0
Packages: MASS, maxLik, VGAM, gamlss

Tailine J. S. Nonato

Methods

GAMLSS Model (used for the Linear Regression based on the BS Distribution)

response variable $Y \sim \mathcal{D}(\boldsymbol{\theta}_k)$ where $\mathcal{D}$ is a probability distribution and $\boldsymbol{\theta}_k = (\theta_1, \ldots, \theta_p)^\top$ is its vector of parameters.

$$g_k(\theta_k) = \eta_k = X_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} s_{jk}(x_{jk}),$$

where $g_k(\cdot), k = 1, \ldots, p$ is a link function and $\boldsymbol{\beta}_k^\top = (\beta_{1k}, \ldots, \beta_{J_k^0 k})$ is the parameter vector.

(ROQUIM et al., 2021)

Tailine J. S. Nonato

Methods

Quantile Regression Model [BS Distribution]

let $T_i \sim \mathrm{BS}(\alpha, Q_i)$ then, define the model as

$$h(Q_i) = \eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad i = 1, \ldots, n$$

such that $Q_i = h^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta})$ where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$

for p<n, , is a vector of unknown regression parameters to be estimated,

and $\mathbf{x}_i^\top = (1, x_{i1}, \ldots, x_{i(p-1)})$ represents the values of the p covariates.

(SÁNCHEZ et al., 2021)

Tailine J. S. Nonato

Literature Review

The application for both GAMLSS for BS distribution and the Quantile Regression based on the BS distribution on agriculture (seeds growth) would be news to the literature, especially if the comparison is possible.

The data was published recently (<1 year). The published studies related to it are authored by the same researcher (Liv Soares Severino) and none of them take the statistical modeling approach.

References

SÁNCHEZ, Luis; LEIVA, Víctor; GALEA, Manuel; SAULO, Helton. Birnbaum-Saunders quantile regression and its diagnostics with application to economic data. Appl Stochastic Models Bus Ind. 2021;37:53–73. Disponível em: https://doi.org/10.1002/asmb.2556.

ROQUIM, Fernanda V.; RAMIRES, Thiago G.; NAKAMURA, Luiz R.; RIGHETTO, Ana J.; LIMA, Renato R.; GOMES, Rayne A. Building flexible regression models: including the Birnbaum-Saunders distribution in the gamlss package. Semina: Ciências Exatas e Tecnológicas, [S. l.], v. 42, n. 2, p. 163–168, 2021. Disponível em: https://doi.org/10.5433/1679-0375.2021v42n2p163.

SEVERINO, Liv Soares. 2024, Castor seed hydration and germination influenced by temperature and puncture. Disponível em: https://doi.org/10.48432/6N3WQA.