

Article

The Use of the Generalized Linear Model to Assess the Speed and Uniformity of Germination of Corn and Soybean Seeds

Deoclecio Jardim Amorim ¹, Amanda Rithieli Pereira dos Santos ², Gabriela Nunes da Piedade ², Rute Quelvia de Faria ³, Edvaldo Aparecido Amaral da Silva ² and Maria Márcia Pereira Sartori ^{2,*}

¹ Department of Exact Sciences, ESALQ—University of São Paulo, Av. Pádua Dias, 11, P.O. Box 9, Piracicaba, SP 13418-900, Brazil; deocleciojardim@hotmail.com

² Department of Plant Production, School of Agriculture, UNESP—São Paulo State University, Av. Universitária, 3780, P.O. Box, Botucatu, SP 18610-034, Brazil; amandarithieli@hotmail.com (A.R.P.d.S.); gabriela.piedade@unesp.br (G.N.d.P.); amaral.silva@unesp.br (E.A.A.d.S.)

³ Department of Agricultural Engineering, Goiano Federal Institute, Rod. Geraldo Silva Nascimento, Km-2.5, P.O. Box, Urutaí, GO 75790-00, Brazil; rute.faria@ifgoiano.edu.br

* Correspondence: maria.mp.sartori@unesp.br

Abstract: The use of seeds with high physiological quality allows rapid growth and establishment of seedlings in the field to be obtained. Therefore, the accuracy of the information obtained during the determination of the physiological quality of seeds is of great importance. The objective was to use generalized linear models, investigating which link function (Probit, Logit and Complementary log-log) is suitable to predict T50 and uniformity during germination of soybean and corn seeds. To perform the experiments, we used seeds from five commercial hybrids and/or cultivars of corn and soybean. The germination speed was calculated by counting the germinated seeds and the results were expressed in the form of proportions. Germination uniformity was calculated by the difference in the times required for germination. The best model was selected according to the criteria of the test of Deviance, AIC and BIC. The Logit model showed accurate results for most cultivars. The evaluation of germination in the form of proportions considering the assumption of binomial response is satisfactory, and the choice of the link function is dependent on the characteristics of each lot and/or species evaluated. The use of this methodology makes it possible to estimate any germination time and uniformity.

Keywords: physiological quality; vigor; *Zea mays* L.; *Glycine max* (L.) Merrill; link functions



Citation: Jardim Amorim, D.; Pereira dos Santos, A.R.; Nunes da Piedade, G.; Quelvia de Faria, R.; Amaral da Silva, E.A.; Pereira Sartori, M.M. The Use of the Generalized Linear Model to Assess the Speed and Uniformity of Germination of Corn and Soybean Seeds. *Agronomy* **2021**, *11*, 588. <https://doi.org/10.3390/agronomy11030588>

Academic Editors: Petr Smýkal and Juan Pablo Renzi

Received: 19 February 2021

Accepted: 16 March 2021

Published: 19 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Physiological seed quality is the ability of the seed to perform vital functions, characterized by germination, vigor and longevity, which directly affects the implantation of a culture under field conditions. High-potential seeds guarantee the growth and development of plants and the eventual yield of crops [1–3].

The correct identification of lots and/or cultivars with seeds of high physiological potential is one of the main tasks of researchers and professionals working in seed physiology and technology. For this, two basic physiological components are taken into account—germination and vigor—these being the factors that theoretically govern the ability of seeds to express their vital functions under biotic and abiotic conditions [4,5].

There are several ways to assess the physiological quality of a seed lot; however, the most common method is the germination test [2,6]. As a result, the germination test is performed to determine the final germination percentage of the seed lots. However, the time for 50% of the seeds to germinate (T50) and the germination uniformity, which is the time difference between two percentage germinations and other defined parameters [7–10], complement the germination data and can indicate the performance of seeds in the field.

To assess which seed lot has superior physiological quality, the response variable (percentage of germinated seeds) is commonly assessed by non-linear regressions, such as

the Hill function [7,10,11]. Another way used to analyze the germination data is through linearization, in which a link function is used, highlighting the Probit function as the most used [12–14].

However, such approaches are imprecise, as they consider the response variable to be continuous. For instance, in the use of non-linear regressions, the proportions of germinated seeds are cumulative and residual autocorrelation may occur [15], whereas in the use of linearization, the germination percentages are transformed into Probit units, considering that the data follow normal distribution, which ends up generating inaccurate models or simply a lack of linearization [14,16].

As the germination process is qualitative, with a binary result—that is, the seed does or not germinate—errors are not normally distributed. Thus, the classic regression analysis approach is not indicated [15]. An alternative for analyzing this type of data is the theory of generalized linear models, with the binomial distribution being a particular case and indicated for proportion data [13,17,18]. Therefore, our hypothesis is that the generalized linear models are more indicative, since it may provide the most accurate information about the problem exposed [19].

In this context, the objective of this work was to use the generalized linear models, investigating which link function (Probit, Logit and Complementary log-log) is suitable to predict T50 and uniformity during germination of soybean and corn seeds.

2. Materials and Methods

2.1. Plant material

The commercial corn seeds used were: AS 1633 PRO3, 2B587 RR, 2A401PW, AL Bandeirante and BRS 4103. The first three refer to hybrid corn seeds, while the last two are cultivars. The commercial soybean cultivars used were: DS59716 IPRO, CD2737 RR, CD251 RR, CD2820 IPRO and CD2857 RR. Corn seeds were produced in the center of southern Brazil and soybean seeds were produced in the central region of Brazil (Goiás Province) grown in the crop season of 2016/17.

2.2. Germination, T50 and Uniformity Study

The corn and soybean seeds were distributed on sheets of paper towels, moistened with an amount of water equivalent to 2.5 times the dry paper mass in Petri dishes. For each hybrid and/or cultivar, 20 seeds were used with four replications at 15 independent points in time. Then, the Petri dishes were transferred to a BOD (Biochemical Oxygen Demand) germinator type, set at a constant temperature of 25 °C. In Table 1, the average water content of seeds for each hybrid and/or cultivar used in this research is demonstrated.

Table 1. Average water content for corn and soybean hybrids and/or cultivars.

Corn		Soybean	
Hybrids/Cultivars	* (%U)	Cultivars	* (%U)
AS 1633 PRO3	10.72	DS59716 IPRO	11.73
2B587 RR	11.36	CD2737 RR	11.82
2A401PW	12.00	CD251 RR	12.00
AL Bandeirante	11.15	CD2820 IPRO	12.26
BRS 4103	11.42	CD2857 RR	12.60

* (%U) = average water content calculated based on wet weight [20].

The counting of the germinated seeds was carried out at regular intervals of 6, 12 and 24 h and continued until 204 h (15 independent points in time), with the protrusion of the primary root ≥ 2 mm being adopted as the germination criterion [20]. The results were expressed by the percentage of the number of seeds germinated in each time interval and the number of viable seeds obtained at the end of the experiment, given by a sequence of Bernoulli tests.

2.3. Binomial Regression Models and Selection Criteria

Since the response variable, that is, the proportion of seeds germinated over time, has an exponential distribution, the use of the generalized linear models proposed by [21] was considered. These models are made up of three parts: a random component of the model in which the response variable has a distribution belonging to the exponential family; a systematic component formed by a set of independent variables; and the link function, which makes the link between the random component and the systematic component.

The default distribution assumed for the random variable Y_i , $i = 1, \dots, n$, which is the number of seeds germinated, is the binomial with parameter m_i and probability of success π_i —that is, $Y_i \sim \text{Binomial}(m_i, \pi_i)$. Thus, the link functions used for binomial regression were Probit, Logit and Complementary log-log. The canonical forms of these models are presented below and, according to [17], the use of these functions in canonical form has as advantages an adequate scale for modeling with practical interpretation of the model parameters and the simplification of the estimation algorithms. Thus, these models in canonical form are:

$$g(\mu_i) = \Phi^{-1}(\mu_i/m_i) = \Phi^{-1}(\pi_i) \quad (1)$$

$$g(\mu_i) = \log\left(\frac{\mu_i}{m_i - \mu_i}\right) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) \quad (2)$$

$$g(\mu_i) = \log\left\{-\log\left(1 - \frac{\mu_i}{m_i}\right)\right\} = \log\{-\log(1 - \pi_i)\} \quad (3)$$

in which μ_i is the mean and Φ^{-1} is the standard normal cumulative distribution function.

The model parameters were obtained using the maximum likelihood method. It is recommended to use the maximum likelihood method due to its excellent, consistent and asymptotic properties [17,21]. When the parameters β_0 and β_1 were obtained, the adjustment was verified by the Deviance criteria and Akaike (AIC) and Bayesian (BIC) information criteria.

The Deviance criterion is used as a measure of discrepancy, measuring the quality of fit of the models; for this reason, the log likelihood rate statistic is used. By this criterion, a model is considered ideal when it results in a significantly small value of the deviation, that is, generating evidence that, for a small number of parameters, an adjustment as good as the saturated model is obtained [21,22]. With regard to binomial distribution, Deviance was defined by:

$$D_B = 2 \sum_{i=1}^n \left[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) + (m_i - y_i) \log\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right) \right] \quad (4)$$

in which y_i represents the realization of the random variable; $m_i = n$: sample size; $\hat{\mu}_i$, $i = 1, 2, \dots, n$, are the adjusted values for the model of interest.

Thus, when Deviance, divided by the number of degrees of freedom, is not close to 1, there is an indication that the model may be poorly adjusted, since the binomial models have a dispersion parameter $\psi = 1$, which is fixed. Values with parameter dispersion $\psi < 1$ and $\psi > 1$ indicate underdispersion and overdispersion, respectively.

The checked sub-effect or overdispersion is incorporated into a constant parameter dispersion (ψ) using the maximum estimation method of “quasi-likelihood”. The dispersion parameter ψ was estimated from the square root of the Deviance quotient by the degrees of freedom. Thus, with this new dispersion parameter, the standard errors of the parameters (β_0 and β_1) were corrected, and the inferences in the significance tests were reformulated. Such correction was performed as described by Equation (5):

$$SE_{corrected} = \hat{\psi} \times SE \quad (5)$$

in which SE and $SE_{corrected}$ is the standard error and standard error corrected, respectively.

Among the more specific criteria and those used in the literature to assess quality setting are the Akaike (AIC) [23] and Bayesian (BIC) information criteria [24]. Both criteria penalize the lack of adjustment to the data and the complexity of the model, so that the lowest values are preferred [19]. The expressions are shown in (6) and (7), respectively.

$$AIC = -2LL + 2p \quad (6)$$

$$BIC = -2LL + p \log(n) \quad (7)$$

in which p is the number of parameters estimated in the model, LL is the log likelihood evaluated at the value of the estimated parameters; and n is the total number of observations used.

2.4. Germination Times and Germination Uniformity

After selecting the most parsimonious model, the times needed to germinate were calculated at 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90 and 99% for each corn and soybean hybrid and/or cultivar, by repetition. Germination uniformity was calculated by the difference in germination times of 75 and 25% (U7525).

The following formulas and Table 2 were used to calculate each time already mentioned, depending on the model selected:

$$Probit (\hat{\theta}_p) = \frac{1}{\hat{\beta}_1} [\Phi^{-1}(p) - \hat{\beta}_0] \quad (8)$$

$$Logit (\hat{\theta}_p) = \frac{1}{\hat{\beta}_1} \left[\log\left(\frac{p}{1-p}\right) - \hat{\beta}_0 \right] \quad (9)$$

$$CLL (\hat{\theta}_p) = \frac{1}{\hat{\beta}_1} \{ \log[-\log(1-p)] - \hat{\beta}_0 \} \quad (10)$$

in which $\hat{\theta}_p$ is specific time for a determined p -value and $(\hat{\beta}_0, \hat{\beta}_1)$ is the pair of estimates for a simple linear regression model.

Table 2. Standard table for calculating the germination probabilities obtained by the Probit, Logit and Complementary log-log (CLL) functions.

% Germination	Probability	Probit *	Logit **	CLL ***
	$p = (\%Germination/100)$	$\Phi^{-1}(p)$	$\log(p/(1-p))$	$\log[-\log(1-p)]$
1	0.0100	-2.3263	-4.5951	-4.6001
2	0.0200	-2.0537	-3.8918	-3.9019
3	0.0300	-1.8808	-3.4761	-3.4914
4	0.0400	-1.7507	-3.1781	-3.1985
5	0.0500	-1.6449	-2.9444	-2.9702
6	0.0600	-1.55478	-2.7515	-2.7826
7	0.0700	-1.4758	-2.5867	-2.6232
8	0.0800	-1.4051	-2.4423	-2.4843
9	0.0900	-1.3408	-2.3136	-2.3612
10	0.1000	-1.2816	-2.1972	-2.2504
16	0.1600	-0.9945	-1.6582	-1.7467
20	0.2000	-0.8416	-1.3863	-1.4999
25	0.2500	-0.6745	-1.0986	-1.2459
30	0.3000	-0.5244	-0.8473	-1.0309
40	0.4000	-0.2533	-0.4055	-0.6717
50	0.5000	0.0000	0.000	-0.3665
60	0.6000	0.2533	0.4055	-0.0874
70	0.7000	0.5244	0.8473	0.1856
75	0.7500	0.6745	1.0986	0.3266

Table 2. Cont.

% Germination	Probability	Probit *	Logit **	CLL ***
	$p = (\% \text{Germination}/100)$	$\Phi^{-1}(p)$	$\log(p/(1-p))$	$\log[-\log(1-p)]$
80	0.8000	0.8416	1.3863	0.4759
84	0.8400	0.9945	1.6582	0.6057
90	0.9000	1.2816	2.1972	0.8340
91	0.9100	1.3408	2.3136	0.8788
92	0.9200	1.4051	2.4423	0.9265
93	0.9300	1.4758	2.5867	0.9780
94	0.9400	1.5548	2.7515	1.0344
95	0.9500	1.6449	2.9444	1.0972
96	0.9600	1.7507	3.1781	1.1690
97	0.9700	1.8808	3.4761	1.2546
98	0.9800	2.0537	3.8918	1.3641
99	0.9900	2.3263	4.5951	1.5272

* Values in Probit units; ** Values in Logit units; *** Values in CLL units.

2.5. Statistical Analysis of Germination Times

The germination times T10, T50, T99 and U7525 (uniformity) were submitted to the Shapiro–Wilk (modified) normality test [25], followed by univariate analysis, considering the completely randomized design with the application of Fisher’s LSD test, at 5% probability.

Subsequently, aiming to know the germination behavior of corn and soybean hybrids and/or cultivars globally, the germination times (T10, T20, T30, T40, T50, T60, T70, T80, T90 and T99) and uniformity (U7525) were submitted to the multivariate classificatory technique of grouping by the Ward method, using the Euclidean distance, by the standardized data matrix for each variable by the formula:

$$Z = \frac{X - \bar{X}}{S} \quad (11)$$

in which Z is the standardized value observed; X is the observed value; \bar{X} is the average of the observed values; and S is the standard deviation of the observed values.

2.6. Computational Aspects

The work was carried out using SAS Software (Version 9.4, Cary, NC, USA) through GENMOD procedures for model adjustments and selection, and GLM for univariate analyses. For multivariate analysis and preparation of graphics, we used MINITAB Statistical Software (Version 17, State College, PA, USA).

3. Results and Discussion

In Figure 1 (experimental data), we present the representations of the germination process for the different hybrids and/or cultivars of corn and soybeans. Initially, germination is slow, with a low proportion, then there is a period of acceleration and finally, stabilization with all viable seeds germinating.

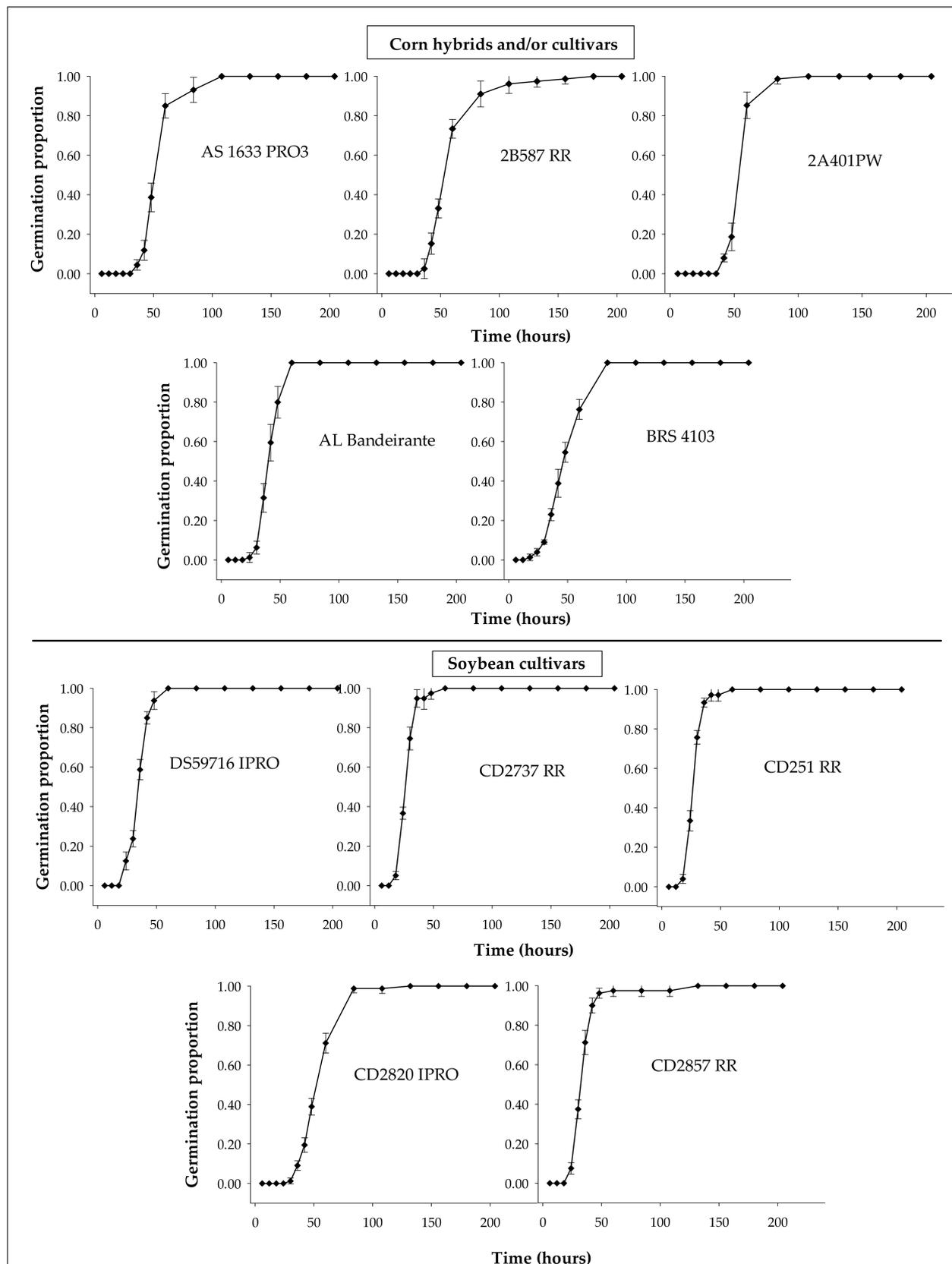


Figure 1. Germination process of corn and soybean hybrids and/or cultivars (mean \pm standard deviation of germination proportions at 15 independent points in time).

When adjusting the Probit, Logit and Probit, and Logit and Complementary log-log link functions for the 10 hybrids and/or cultivars tested, it was observed that no cultivar presented adequate adjustment simultaneously by the link functions when considering the Deviance criterion. The corn hybrid and soybean cultivars with dispersion parameter closest to 1 were AS 1633 PRO3 and CD251 RR, both adjusted by the Logit link function. The overdispersion phenomenon was observed in six cultivars simultaneously, with functions by link values to Deviance from 1.3, whereas only the 4103 corn cultivar BRS presented underdispersion, indicated by three functions being used (Table 3).

Table 3. Criteria for adjustment and selection of models for corn and soybean hybrids and/or cultivars.

Corn				
Cultivars	Functions	Deviance	AIC	BIC
AS 1633 PRO3	Probit	1.2323	94.97	99.16
	Logit	1.0959	89.24	92.81
	* CLL	2.3024	139.92	143.49
2B587 RR	Probit	2.1019	175.06	179.11
	Logit	1.7032	153.53	157.58
	* CLL	4.2072	288.75	292.8
2A401PW	Probit	0.6858	63.7	67.27
	Logit	0.5467	57.86	61.43
	* CLL	1.7637	108.97	112.54
AL Bandeirante	Probit	1.8101	102.83	105.99
	Logit	1.8553	104.36	107.53
	* CLL	2.2246	116.92	120.09
BRS 4103	Probit	0.3884	77.05	80.43
	Logit	0.4170	78.14	81.52
	* CLL	0.3350	75.02	78.4
Soybean				
Hybrids/Cultivars	Functions	Deviance	AIC	BIC
DS59716 IPRO	Probit	1.5779	104.76	107.93
	Logit	1.6267	106.42	109.59
	* CLL	2.0302	120.14	123.31
CD2737 RR	Probit	1.4961	100.38	103.55
	Logit	1.3030	93.81	96.98
	* CLL	2.8422	525.07	529.18
CD251 RR	Probit	1.1934	89.73	92.9
	Logit	1.0035	83.27	86.44
	* CLL	2.4862	133.69	136.85
CD2820 IPRO	Probit	1.8319	133.17	136.83
	Logit	1.7514	129.47	133.21
	* CLL	3.1479	193.71	197.45
CD2857 RR	Probit	4.3249	259.93	263.67
	Logit	2.8754	193.25	196.99
	* CLL	7.6990	415.14	418.88

* Complementary log-log; Akaike information criterion (AIC) and Bayesian information criterion (BIC).

The phenomena of under- and overdispersion were defined by [17], as a variance of the response variable above or below the variance expected by the model adopted. The main consequences of these phenomena are the estimation of standard errors, which consequently can induce an inappropriate choice of models, potentially compromising the conclusions [26]. Even in the face of researchers' efforts to control experimental conditions, the occurrence of phenomena such as overdispersion is common for agricultural systems, as there is great variability [27,28].

Among the tested link functions, the complementary log-log was the one that gave the highest values for Deviance—in other words, most models formulated with that link function are overdispersed, except BRS 4103. While functions Probit and Logit showed the smallest deviations, the latter being closer, the Logit function was the one that provided the best adjustments, presenting the appropriate adjustment in 7 of the 10 hybrids and/or cultivars used in this research (Table 3).

These results demonstrate the importance of considering the choice of the correct link function, since the use of inaccurate models is potentially generating misleading conclusions [26,28]. Although the logit model is currently preferred in some areas—for example, in biometrics [22,29,30]—in this study, smaller deviances were obtained for most hybrid and/or cultivar studied; it is necessary to study this by comparing the link seeking functions that best describe the probability of interest [31].

For all hybrids and/or cultivars tested, there was agreement between the Deviance criterion and the AIC and BIC information criteria—that is, those functions with the least deviations were also those with the lowest AIC and BIC values (Table 3). This agreement facilitated the selection of the most parsimonious models to evaluate the germination of corn and soybean seeds. Thus, the Probit model was chosen to evaluate the germination of two cultivars, the Logit model of seven cultivars and the Complementary log-log of one cultivar (Table 4).

Table 4. Estimation of the parameters of the binomial regression models adjusted to the proportion data in the germination of corn and soybeans.

Corn							
Hybrids/ Cultivars	Functions	Parameter	Estimated	SE	SE _{corrected}	p Value	p Value _{corrected}
AS 1633 PRO3	Logit	$\hat{\beta}_0$	−8.6418	0.7731	0.8093	<0.0001	<0.0001
		$\hat{\beta}_1$	0.1653	0.0158	0.0165	<0.0001	<0.0001
		ψ	1.0468	-	-	-	-
2B587 RR	Logit	$\hat{\beta}_0$	−5.9099	0.3861	0.5039	<0.0001	<0.0001
		$\hat{\beta}_1$	0.1009	0.0074	0.0096	<0.0001	<0.0001
		ψ	1.3051	-	-	-	-
2A401PW	Logit	$\hat{\beta}_0$	−12.9499	1.2264	0.9068	<0.0001	<0.0001
		$\hat{\beta}_1$	0.2417	0.0237	0.0175	<0.0001	<0.0001
		ψ	0.7394	-	-	-	-
AL Bandeirante	Probit	$\hat{\beta}_0$	−5.8063	0.4596	0.6183	<0.0001	<0.0001
		$\hat{\beta}_1$	0.1616	0.0126	0.0170	<0.0001	<0.0001
		ψ	1.3454	-	-	-	-
BRS 4103	CLL	$\hat{\beta}_0$	−6.0826	0.4315	0.2497	<0.0001	<0.0001
		$\hat{\beta}_1$	0.1521	0.0108	0.0063	<0.0001	<0.0001
		ψ	0.5788	-	-	-	-
Soybean							
Cultivars	Functions	Parameter	Estimated	SE	SE _{corrected}	p value	p value _{corrected}
DS59716 IPRO	Probit	$\hat{\beta}_0$	−4.4746	0.3207	0.4028	<0.0001	<0.0001
		$\hat{\beta}_1$	0.1291	0.0091	0.0114	<0.0001	<0.0001
		ψ	1.2562	-	-	-	-
CD2737 RR	Logit	$\hat{\beta}_0$	−7.1362	0.5913	0.6749	<0.0001	<0.0001
		$\hat{\beta}_1$	0.2666	0.0214	0.0245	<0.0001	<0.0001
		ψ	1.1415	-	-	-	-
CD251 RR	Logit	$\hat{\beta}_0$	−7.5645	0.6589	0.6600	<0.0001	<0.0001
		$\hat{\beta}_1$	0.2810	0.0238	0.0239	<0.0001	<0.0001
		ψ	1.0017	-	-	-	-

Table 4. Cont.

Corn							
Hybrids/ Cultivars	Functions	Parameter	Estimated	SE	SE _{corrected}	p Value	p Value _{corrected}
CD2820 IPRO	Logit	$\hat{\beta}_0$	-7.4894	0.5670	0.7504	<0.0001	<0.0001
		$\hat{\beta}_1$	0.1410	0.0116	0.1110	<0.0001	<0.0001
		ψ	1.3234	-	-	-	-
CD2857 RR	Logit	$\hat{\beta}_0$	-6.0158	0.4445	0.7537	<0.0001	<0.0001
		$\hat{\beta}_1$	0.1799	0.0130	0.0221	<0.0001	<0.0001
		ψ	1.6957	-	-	-	-

Estimated parameters at the 5% probability level; $\hat{\beta}_0$ = intercept; $\hat{\beta}_1$ = slope; ψ = scale parameter calculated by the deviation divided by the degrees of freedom; SE = standard error; and SE_{corrected} = standard error corrected.

The information criteria presented penalize the lack of adjustment to the data and the complexity of the model; therefore, models with lower values were chosen [19,32]. According to these criteria, the Logit function stands out as a good alternative to evaluate the germination of corn and soybean seeds, generating theoretical bases for other areas, such as, for example, in seed science and technology, countering the idea that the model Probit should always be used to assess physiological quality, from germination to longevity in thermal models [12,14].

Keeping the focus on the estimation of the dispersion parameter, the selected models had their standard errors of the estimates corrected by quasi-likelihood using Deviance to estimate the constant dispersion parameter, where it was used in the procedure in (5). With the application of this correction, as expected, standard errors showed an increase for models with overdispersion and a decrease for models with subdispersion; however, the significance of the parameters was not affected (Table 4).

The use of quasi-likelihood to correct estimated standard errors is recommended by [33] and has been adopted in seed germination studies [18,29], in entomology data [34,35], in the assessment of ecological data [26], and in the modeling of the number and dry matter mass of Rhizobium nodules in bean culture [28]. Therefore, there is a solid body of literature on the use of this methodology.

With the selected models (Table 4), it is possible to estimate the germination times of interest to the researcher, using the formulas presented in (8), (9) and (10). The average germination time or time required for 50% of the seeds used in the germination experiments (T50) is considered to be the preferred one to describe the germination and physiological quality of the seeds submitted to different treatments or to compare different batches of seeds [7–9,11,14].

In the case of the Probit and Logit models, the T50 can be obtained easily, since the parameters β_0 and β_1 form a linear equation of the type $y = a \pm bx$, where, upon equaling the terms of the equation to zero, the T50 is obtained, because both models have symmetry around zero. Thus, to calculate the T50, we can use the following formula:

$$T50 = -\frac{\hat{\beta}_0}{\hat{\beta}_1} \quad (12)$$

in which $\hat{\beta}_0$ is the intercept and $\hat{\beta}_1$ the slope angle.

Table 5 shows the values for T50 obtained in the 10 hybrids and/or cultivars, adopting the models provided in Table 4 and the expected interval for this parameter obtained in an experimental way. Thus, we can consider the methodology of generalized linear models, adopting the efficient binomial distribution to evaluate the germination of corn and soybeans, since all the results estimated by the chosen functions are contained in the experimental intervals.

Table 5. Estimates of the T50 obtained from the equations in Table 4 and the experimental interval in which the T50 is contained.

	Hybrids/Cultivars	T50 (95% CI)	T50 (Experimental Interval)
Corn	AS 1633 PRO3	52.23 (48.90; 56.76) h	48–60 h
	2B587 RR	58.57 (54.42; 62.27) h	48–60 h
	2A401PW	53.57 (49.67; 57.52) h	48–60 h
	AL Bandeirante	35.93 (32.02; 39.87) h	30–36 h
	BRS 4103	37.59 (33.65; 41.50) h	36–42 h
	Cultivars	T50 (95% CI)	T50 (Experimental Interval)
Soybean	DS59716 IPRO	34.66 (29.56; 39.70) h	30–36 h
	CD2737 RR	26.77 (21.73; 31.87) h	24–30 h
	CD251 RR	26.92 (21.86; 32.01) h	24–30 h
	CD2820 IPRO	53.12 (48.31; 58.45) h	48–60 h
	CD2857 RR	33.44 (28.38; 38.52) h	30–36 h

95% CI = 95% confidence interval.

As much as authors defend the use of linear models to estimate germination times, considering the assumption of normality of the data [12,14,36], often a simple transformation of the percentages of germination using a certain link function, such as the Probit model (inv. Norm function in Microsoft Excel) does not allow the dataset to be linearized [14,16]. Thus, an approach considering germination as a binary variable, in which seeds may or may not germinate, has been more indicated [13,18].

It is worth mentioning that another differential of the work is that the calculated germination times are obtained based on the number of viable seeds—that is, the correct definition for the T50 in this research is the time required for 50% of the viable seeds to germinate, not requiring additional formulas to calculate the actual amount of germinated seeds.

In addition to the traditional T50, other parameters can be used to evaluate the germination of a seed batch; for example, it is possible to calculate the time for 10% of viable seeds to germinate, or it is also necessary to identify whether two batches with final germinations have the same germination uniformity or even understand germination as a global process, not being restricted to just a few parameters that can lead to false conclusions about the physiological quality of a seed lot.

As an example of using other parameters to assess germination, we have the work of [7,11], using Hill's nonlinear function of four parameters to estimate beyond T50, the maximum germination time and germination uniformity, which is the time interval between two predefined germinations. The germination times of 10 and 90% of the seeds have also been calculated [14,37] to evaluate the physiological quality of seeds. However, the two most widespread statistics for evaluating the germination of a seed lot are the T50 and germination uniformities [7,10,37].

For uniformity of germination in contrast to the T50, there is no standardization, being adopted in several ways: U9010 (time for 90% germination—time for 10% germination) [37], U8416 (time for 84% germination—time for 16% germination) [10], U7525 (time for 75% germination—time for 25% germination) [7], the latter being the most traditional.

Thus, seed lots or cultivars that exhibit the lowest values of T50 or any other germination time can be considered of higher physiological quality; the same reasoning is valid for germination uniformity [7,8,37].

Currently, research involving the evaluation of seed germination to determine previously mentioned parameters largely uses non-linear models or some link function directly on the percentage data, without paying attention to the type of variable studied and its probability distribution, which often causes convergence problems or even severe errors in parameter estimation [7,10,14]. However, when we use generalized linear models, these

difficulties are overcome, as we are working with a simple linear equation. As a demonstration, the germination of the corn cultivar BRS 4103 was modeled by the Complementary log-log function, showing the T50 and the germination uniformity (see Figure 2). The substitution of T50 = 37.59 in the equation shown in Figure 2 will return the value of ~ -0.3665 , corresponding to 50% germination as indicated in Table 2.

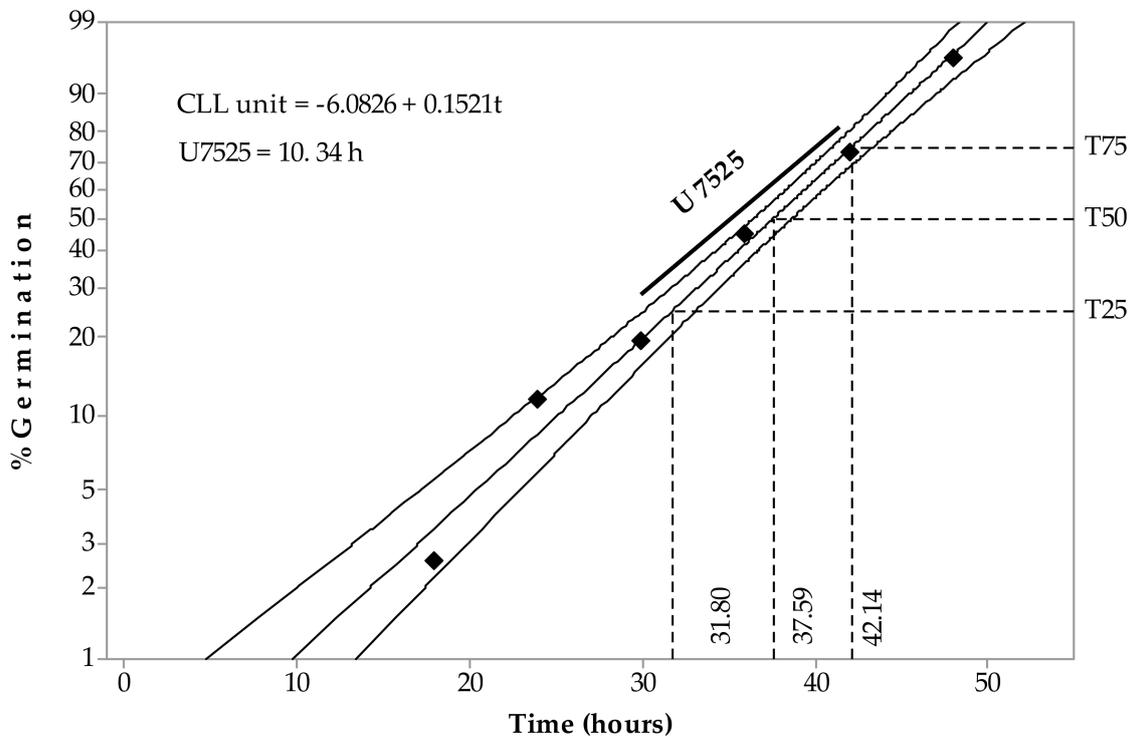


Figure 2. Germination times of BRS 4103 corn cultivar estimated by the Complementary log-log link function. T25 = 31.80 h; T50 = 37.59 h; T75 = 42.14 h and uniformity of germination U7525. At 5% probability.

Following selection of link functions better suited to evaluate the germination of each plot, germination times T10, T50, T99 and U7525 were determined for all sampling data, following which, analysis of variance was performed complemented with the means test (LSD) in order to compare the physiological potential of corn and soybean cultivars. According to the (modified) Shapiro–Wilk test [25], the four parameters evaluated have a normal distribution. The results of the analysis of variance revealed that the F test was significant for all corn parameters, whereas, for soybean cultivars, only germination uniformity was not significant at 5% probability.

When evaluating T10, it was possible to observe that the corn cultivar 2A401PW was the one that showed the slowest germination, while the cultivars AL Bandeirante and BRS 4103 showed faster germinations. For evaluated soybean varieties, cultivating CD2820 IPRO showed slower germination, different to other cultivars using twice the time to reach 10% germination, compared with cultivars CD251 RR and CD2737 RR (Table 7).

Table 6. Germination times and uniformity for corn and soybean hybrids and/or cultivars.

	Hybrids/Cultivars	T10 _(95% IC) *	T50 _(95% IC) *	T99 _(95% IC) *	U7525 _(95% IC) #
Corn	2A401PW	44.83 (41.43; 48.22) a	53.57 (49.67; 57.52) a	71.94 (59.73; 84.14) b	8.77 (4.33; 13.21) b
	AS 1633 PRO3	40.15 (36.76; 43.55) ab	52.23 (48.90; 56.76) a	79.34 (67.14; 91.55) b	12.68 (8.24; 17.12) b
	2B587 RR	37.44 (34.04; 40.83) b	58.57 (54.42; 62.27) a	102.07 (89.9; 114.3) a	20.91 (16.47; 25.35) a
	AL Bandeirante	29.45 (26.06; 32.85) c	35.93 (32.02; 39.87) b	47.73 (35.53; 59.94) c	6.83 (2.40; 11.27) b
	BRS 4103	25.26 (21.86; 28.65) c	37.59 (33.65; 41.50) b	49.96 (37.75; 62.16) c	10.28 (5.84; 14.72) b

Table 7. Germination times and uniformity for corn and soybean hybrids and/or cultivars.

	Cultivars	T10 _(95% IC) *	T50 _(95% IC) *	T99 _(95% IC) *	U7525 _(95% IC) #
Soybean	CD2820 IPRO	40.30 (34.69; 45.90) a	53.12 (48.31; 58.45) a	80.74 (71.17; 90.30) a	13.08 (9.59; 16.58) ns
	DS59716 IPRO	25.73 (20.12; 31.33) b	34.66 (29.56; 39.70) b	50.79 (41.22; 60.36) bc	9.37 (5.88; 12.86) ns
	CD2857 RR	22.05 (16.44; 27.66) b	33.44 (28.38; 38.52) bc	57.29 (47.73; 66.86) b	11.40 (7.91; 14.90) ns
	CD251 RR	19.44 (13.84; 25.05) b	26.92 (21.86; 32.01) c	42.62 (33.05; 52.18) c	7.50 (4.00; 10.99) ns
	CD2737 RR	19.01 (13.40; 24.61) b	26.77 (21.73; 31.87) c	43.11 (33.55; 52.68) c	7.80 (4.30; 11.29) ns

* T10_(95% CI), T50_(95% CI) and T99_(95% CI) = time required for germination of 10, 50 and 99% of viable seeds and 95% confidence interval; # U7525_(95% CI) = germination uniformity given by the difference between the 75th and 25% percentiles and 95% confidence interval; ns = not significant at the 5% probability level; averages followed by equal lowercase letters in the columns do not differ between themselves by Fisher's LSD test at 5%.

The behavior of cultivars at T50 was altered in relation to T10, only for cultivar 2B587 RR, the corn cultivars AL Bandeirante and BRS 4103 were also faster in reaching 50% germination in approximately 16 h. For soybeans, the cultivar CD2820 IPRO continued to be less vigorous, whereas cultivars CD251 RR and CD2737 RR continued to exhibit greater physiological quality (Table 7).

The time for 99% of germinated seeds was calculated in order to determine the behavior of hybrids and/or cultivars when they are near to complete 100% germination. Thus, it is observed that the corn hybrid 2B587 RR, which did not present statistical difference in the previous times with the cultivar AS 1633 PRO3, showed difference in more than 22 h. This differentiation may be due to uniformity since the corn hybrid 2B587 RR proved to be less uniform. The best performing corn cultivar in time T99 was AL Bandeirante, also being the most uniform. For soybean cultivars at time T99, it was proved that CD2820 IPRO is the least vigorous and the CD251 RR and CD2737 RR cultivars have the highest physiological quality.

The lower or higher speed of germination of one cultivar in relation to the other is due to the time spent in the restoration of the damaged organelles and tissues before beginning the development of the embryonic axis, during the germination process [8,38]. According to [38], cultivars or seed lots with higher germination speed and uniformity are considered the most vigorous.

The effect of a seed lot can be defined as the sum of the properties that determine the activity and performance of seed lots as acceptable in germination in a wide range of environments [4,5]. Thus, the identification of high-performance seed lots or cultivars is an important initiative for the success of agricultural production [2].

Under these assumptions, we list the corn cultivars AL Bandeirante and BRS 4103 and the soybean cultivars CD251 RR and CD2737 RR as those of greater vigor based on germination and uniformity times. Several authors who, using a simple radicle count, managed to predict the vigor of several species [8,39,40], support this statement.

The higher vigor of hybrids and/or cultivars is evidenced when analyzing germination in a broader context, considering various germination times (T10, T20, T30, T40, T50, T60, T70, T80, T90 and T99) and uniformity standard (U7525) through multivariate classificatory analysis. In general, the cophenetic coefficient was above 0.90, indicating little distortion with the original data matrix [41]. It was possible to verify the existence of three groups, in the dendrogram, for both corn and soybeans (Figure 3).

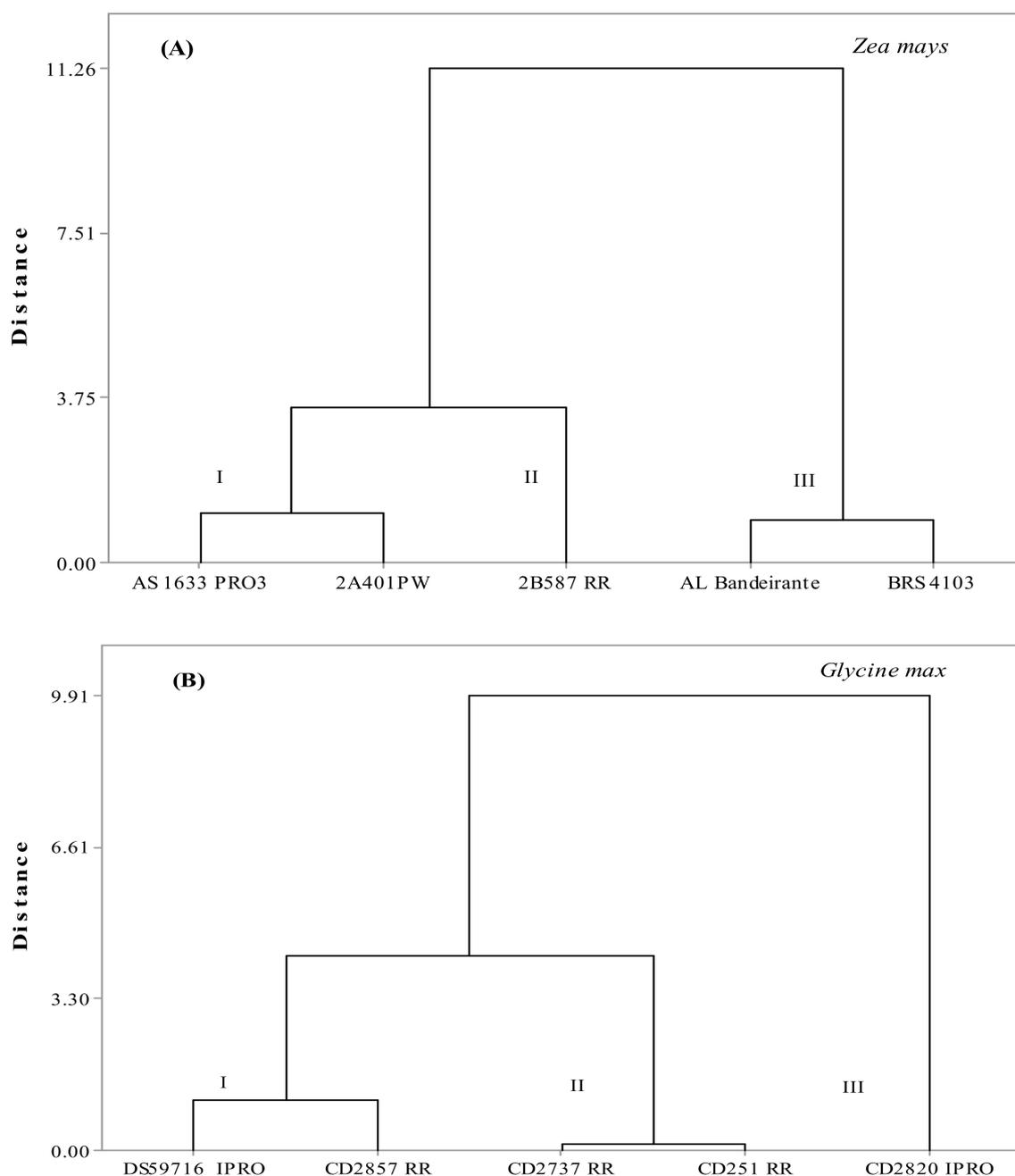


Figure 3. Dendrogram including germination and uniformity times for corn (A) and soybean (B) seeds. I, II and III = groups formed are after multivariate analysis by the Ward method using the Euclidean distance.

The corn plants with smaller distance were AL Bandeirante and BRS 4103, with Euclidean distance of 0.96 (Table 8), or had similar physiological quality, confirming the results for T10, T50 and T99 (Table 7). The greatest difference in physiological quality was observed between the cultivars BRS 4103 and 2B587 RR with Euclidean distance greater than 7 (Table 8). Regarding soybean cultivars, the closest or less distant were cultivars CD2737 and CD251 RR, with Euclidean distance equal to 0.15 (Table 8), confirming the results introduced previously (Table 5). Additionally, the most distant physiological potential was observed for the cultivar CD2820 IPRO with cultivars CD251 RR and CD2737 RR, with Euclidean distance equal to 8.02 (Table 8).

Table 8. Euclidean distances between corn and soybean hybrids and/or cultivars obtained using Ward's method.

Corn					
Hybrids/Cultivars	2A401PW	2B587 RR	AL Bandeirante	AS 1633 PRO3	BRS 4103
2A401PW	0.00	0.00	0.00	0.00	0.00
2B587 RR	3.34	0.00	0.00	0.00	0.00
AL Bandeirante	5.26	7.15 #	0.00	0.00	0.00
AS 1633 PRO3	1.13	2.53	5.07	0.00	0.00
BRS 4103	5.22	6.71	0.96 **	4.85	0.00
Soybean					
Cultivars	CD251 RR	CD2737 RR	CD2820 IPRO	CD2857 RR	DS59716 IPRO
CD251 RR	0.00	0.00	0.00	0.00	0.00
CD2737 RR	0.15 **	0.00	0.00	0.00	0.00
CD2820 IPRO	8.02 #	8.02 #	0.00	0.00	0.00
CD2857 RR	2.61	2.54	5.79	0.00	0.00
DS59716 IPRO	2.32	2.31	5.72	1.11	0.00

** Smallest and # largest distance obtained.

4. Conclusions

Germination evaluation in the form of proportions, considering the assumption of binomial response, is satisfactory, and the choice of the link function will depend on the characteristics of each lot and/or species evaluated. The presented methodology allows calculation of any germination time and uniformity in a more robust way.

Author Contributions: Conceptualization, D.J.A. and M.M.P.S.; methodology, D.J.A., A.R.P.d.S., G.N.d.P. and R.Q.d.F.; software, D.J.A.; validation, M.M.P.S., A.R.P.d.S. and R.Q.d.F.; formal analysis, D.J.A., A.R.P.d.S., R.Q.d.F. and M.M.P.S.; investigation, D.J.A., A.R.P.d.S., R.Q.d.F. and M.M.P.S.; writing—original draft preparation, D.J.A., M.M.P.S. and E.A.A.d.S.; writing—review and editing, A.R.P.d.S., G.N.d.P., M.M.P.S., E.A.A.d.S.; supervision, M.M.P.S. and E.A.A.d.S.; project administration, D.J.A., A.R.P.d.S., M.M.P.S. and E.A.A.d.S.; funding acquisition, M.M.P.S. and E.A.A.d.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Council for Scientific and Technological Development (CNPq-Brazil), finance code 130773/2017-4 and the scholarship of the first author; the fellowship granted to E.A. Amaral da Silva under the grant number 309718-2018-0 and Maria Marcia Pereira Sartori under the grant number 305524/2015-1. Was also funded by the FAPESP-São Paulo Research Foundation (Proc. n° 2018/25698-4) as financial support.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. França-Neto, J.B.; Krzyzanowski, F.C.; Henning, A.A.; Pádua, G.P. Tecnologia de produção de soja de alta qualidade. *Inf. ABRATES* **2010**, *20*, 26–32.
2. Marcos Filho, J. Seed vigor testing: An overview of the past, present and future perspective. *Sci. Agric.* **2015**, *72*, 363–374. [[CrossRef](#)]
3. Pizá, M.C.P.; Prevosto, L.; Zili, C.; Cejas, E.; Kelly, H.; Baleastrasse, K. Effects of non-thermal plasmas on seed-borne *Diaporthe/Phomopsis* complex and germination parameters of soybean seeds. *Innov. Food Sci. Emerg. Technol.* **2018**, *49*, 82–91. [[CrossRef](#)]
4. Association of Official Seed Analysts—AOSA. *Seed Vigor Testing Handbook*; AOSA: Ithaca, NY, USA, 1983; p. 93.
5. Baalbaki, R.; Elias, S.; Marcos-Filho, J.; McDonald, M.B. *Seed Vigor Testing Handbook*; Contribution to the Handbook on Seed Testing; AOSA: Ithaca, NY, USA, 2009; Volume 32.
6. Diniz, F.O.; Reis, M.S.; Dias, L.A.S.; Araújo, E.F.; Sediyaama, T.; Sediyaama, C.A. Physiological quality of soybean seeds of cultivars submitted to harvesting delay and its association with seedling emergence in the field. *J. Seed Sci.* **2013**, *35*, 147–152. [[CrossRef](#)]
7. 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. *Plant J.* **2010**, *62*, 148–159. [[CrossRef](#)]

8. Matthews, S.; Noli, E.; Demir, I.; Khajeh-Hosseini, M.M.-H. Wagner Evaluation of seed quality: From physiology to international standardization. *Seed Sci. Res.* **2012**, *22*, S69–S73. [[CrossRef](#)]
9. Hatzig, S.V.; Matthias, F.; Breuer, F.; Nesi, N.; Ducournau, S.; Wagner, M.-H.; Leckband, G.; Abbadi, A.; Snowdon, R.J. Genome-wide association mapping unravels the genetic control of seed germination and vigor in *Brassica napus*. *Front. Plant Sci.* **2015**, *6*, 221. [[CrossRef](#)]
10. Neto, V.G.; Ribeiro, P.R.; Del-Bema, L.E.; Bernal, D.T.; Lima, C.S.T.; Ligterinke, W.; Fernandez, L.G.; Castro, R.D. Characterization of the superoxide dismutase gene family in seeds of two *Ricinus communis* L. genotypes submitted to germination under water restriction conditions. *Environ. Exp. Bot.* **2018**, *155*, 453–463. [[CrossRef](#)]
11. El-Kassaby, Y.A.; Moss, I.; Kolotelo, D.; Stoehr, M. Seed germination: Mathematical representation and parameters extraction. *For. Sci.* **2008**, *54*, 220–227.
12. Ellis, R.H.; Roberts, E.H. Improved equations for the prediction of seed longevity. *Ann. Bot.* **1980**, *45*, 13–30. [[CrossRef](#)]
13. Hay, F.R.; Mead, A.; Bloomberg, M. Modelling seed germination in response to continuous variables: Use and limitations of Probit analysis and alternative approaches. *Seed Sci. Res.* **2014**, *24*, 165–186. [[CrossRef](#)]
14. Daibes, L.F.; Cardoso, V.J.M. Seed germination of a South American forest tree described by linear thermal time models. *J. Therm. Biol.* **2018**, *76*, 156–164. [[CrossRef](#)] [[PubMed](#)]
15. Onofri, A.; Gresta, F.; Tei, F. A new method for the analysis of germination and emergence data of weed species. *Weed Res.* **2010**, *50*, 187–198. [[CrossRef](#)]
16. Gazola, S.; Scapim, C.A.; Guedes, T.A.; Braccini, A.L. Proposta de modelagem não-linear do desempenho germinativo de sementes de milho híbrido. *Ciência Rural* **2011**, *41*, 551–556. [[CrossRef](#)]
17. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*; Chapman and Hall: London, UK, 1989; p. 511.
18. Murphey, M.; Kovach, K.; Elnacash, T.; He, H.; Bentsink, L.; Donohue, K. DOG1-imposed dormancy mediates germination responses to temperature cues. *Environ. Exp. Bot.* **2015**, *112*, 33–43. [[CrossRef](#)]
19. Emiliano, P.C.; Vivanco, M.J.F.; Menezes, F.S. Information criteria: How do they behave in different models? *Comput. Stat. Data Anal.* **2014**, *69*, 141–153. [[CrossRef](#)]
20. Brasil Ministério da Agricultura. *Pesca e Abastecimento. Regras Para Análise de Sementes*; Secretaria Nacional de Defesa Agropecuária: Brasília, Brazil, 2009; p. 399.
21. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear models. *J. R. Stat. Soc. A* **1972**, *135*, 370–384. [[CrossRef](#)]
22. Demétrio, C.G.B.; Hinde, J.; Moral, R.A. Models for overdispersed data in entomology. In *Ecological Modeling Applied to Entomology*; Springer: Cham, Switzerland, 2014; pp. 219–259.
23. Akaike, H. A new look at the statistical model identification. *Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
24. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
25. Rahman, M.M.; Govindarajulu, Z.A. Modification of the test of Shapiro and Wilk for normality. *J. Appl. Stat.* **1997**, *24*, 219–236. [[CrossRef](#)]
26. Harrison, X.A.A. Comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ* **2015**, *3*, e1114. [[PubMed](#)]
27. Suriyagoda, L.D.B.; Ryan, M.H.; Lambers, H.; Renton, M. Comparison of novel and standard methods for analysing patterns of plant death in designed field experiments. *J. Agric. Sci.* **2012**, *150*, 319–334. [[CrossRef](#)]
28. Rizzardi, D.A.; Contreras-Soto, R.I.; Figueredo, A.S.T.; Andrade, C.A.B.; Santana, R.G.; Sacapim, C.A. Generalized mixed linear modeling approach to analyze nodulation in common bean inbred lines. *Pesquisa Agropecuária Brasileira* **2017**, *52*, 1178–1184. [[CrossRef](#)]
29. Nunes, J.A.R.; Morais, A.R.; Bueno Filho, J.S.S. Modelagem da superdispersão em dados por um modelo linear generalizado misto. *Rev. Mat. Estat.* **2004**, *22*, 55–70.
30. Nattino, G.; Lu, B. Model assisted sensitivity analyses for hidden bias with binary outcomes. *Biometrics* **2018**. [[CrossRef](#)]
31. Freitas, L.; Filho, S.; Júnior, J.; Silva, F. Comparação das funções de ligação Probit e Logit em regressão binária considerando diferentes tamanhos amostrais. *Encicl. Biosf.* **2013**, *9*, 16.
32. Resende, M.D.V. *Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético*, 1st ed.; Embrapa Florestas: Colombo, Brazil, 2007; p. 362.
33. Hinde, J.; Demétrio, C.G.B. Overdispersion: Models and estimation. *Comput. Stat. Data Anal.* **1998**, *27*, 151–170. [[CrossRef](#)]
34. Ribeiro, L.P.; Vendramim, J.D.; Bicalho, K.U.; Andrade, M.S.; Fernandes, J.B.; Moral, R.A.; Demétrio, C.G.B. *Annona mucosa* Jacq. (Annonaceae): A promising source of bioactive compounds against *Sitophilus zeamais* Mots. (Coleoptera). *J. Stored Prod. Res.* **2013**, *55*, 6–14. [[CrossRef](#)]
35. Liska, G.R.; Silveira, E.C.; Reis, P.R.; Cirillo, M.Â.; Gonzalez, G.G.H. Seleção de um modelo de regressão binomial para taxa de predação de *Euseius concordis* (Chant, 1959). *Coffee Sci.* **2015**, *10*, 113–121.
36. Bradford, K.J. Applications of hydrothermal time to quantifying and modeling seed germination and dormancy. *Weed Sci.* **2002**, *50*, 248–260. [[CrossRef](#)]
37. Coolbear, P.; McGill, C.R. Effects of a low-temperature pre-sowing treatment on the germination of tomato seed under temperature and osmotic stress. *Sci. Hortic.* **1990**, *44*, 43–54. [[CrossRef](#)]
38. Villiers, T.A. Ageing and longevity of seeds in field conditions. In *Seed Ecology*; Heydecker, W., Ed.; The Pennsylvania State University: London, PA, USA, 1973; pp. 265–288.

39. Khajeh-Hosseini, M.; Lomholt, A.; Matthews, S. Mean germination time in the laboratory estimates the relative vigour and field performance of commercial seed lots of maize (*Zea mays* L.). *Seed Sci. Technol.* **2009**, *37*, 446–456. [[CrossRef](#)]
40. Khajeh-Hosseini, M.; Nasehzadeh, M.; Matthews, S. Rate of physiological germination relates to the percentage of normal seedlings in standard germination tests of naturally aged seed lots of oilseed rape. *Seed Sci. Technol.* **2010**, *38*, 602–611. [[CrossRef](#)]
41. Silva, A.R.; Dias, C.T.S.A. Cophenetic correlation coefficient for Tocher's method. *Pesquisa Agropecuária Brasileira* **2013**, *48*, 589–596. [[CrossRef](#)]