

Likelihood ratio test for the analysis of germination percentages

Yongha Rhie^{1,†}, Soyeon Lee^{2,†} and Hohsuk Noh² ¹Department of Horticulture, Kangwon National University, Kangwondaehakgil 1, Chuncheon 24341, South Korea and ²Department of Statistics, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul 04310, South Korea

Research Paper

†These authors contributed equally to this work.

Cite this article: Rhie Y, Lee S, Noh H (2024). Likelihood ratio test for the analysis of germination percentages. *Seed Science Research* 1–7. <https://doi.org/10.1017/S0960258524000023>

Received: 12 July 2023

Revised: 8 January 2024

Accepted: 23 January 2024

Keywords:

ANOVA; complete separation; germination percentage; likelihood ratio test; logistic regression

Corresponding author:

Hohsuk Noh;

Email: hsnoh@sookmyung.ac.kr**Abstract**

The germination percentage (GP) is commonly employed to estimate the viability of a seed population. Statistical methods such as analysis of variance (ANOVA) and logistic regression are frequently used to analyse GP data. While ANOVA has a long history of usage, logistic regression is considered more suitable for GP data due to its binomial nature. However, both methods have inherent issues that require attention. In this study, we address previously unexplored challenges associated with these methods and propose the utilization of a likelihood ratio test as a solution. We demonstrate the advantages of employing the likelihood ratio test for GP data analysis through simulations and real data analysis.

Introduction

Germination percentage (GP) is utilized as an estimation of the viability of a seed population. The analysis of GP involves examining experiments that measure the proportion of germinated seeds under different treatments or conditions. Various statistical tests, such as ANOVA, logistic regression or Kruskal–Wallis, are commonly employed to assess the impact of different treatments on GP. However, it is often overlooked whether these specific tests and their underlying assumptions are suitable for the analysis. Several studies have explored this issue to address the appropriateness of statistical analyses in GP studies. Both Jaeger (2008) and Warton and Hui (2011) argued that ANOVA's assumptions are violated by binomial data, thus advocating for the use of logistic regression instead. Sileshi (2012) conducted a study on considerations when utilizing ANOVA and its nonparametric counterpart, the Kruskal–Wallis test, and proposed logistic models or their extensions with random effects as alternative approaches. Additionally, Onofri et al. (2010) and Ritz et al. (2013) recommended employing time-to-event survival analysis for the analysis of germination data. Despite the availability of these studies, we believe that there are still significant statistical issues to be explored concerning the analysis of germination data.

Our main focus is the statistical issue that arises when analysing data where GPs are consistently either 0 or 100% across all replicates. This situation is commonly observed in experiments that study methods of breaking dormancy. In such cases, the absence of variability around the treatment mean significantly violates two crucial assumptions of ANOVA: normality and homogeneity of error variances. Consequently, it is not feasible to employ ANOVA methods or conduct subsequent mean comparison tests. Sileshi (2012) asserted that such data should conform strictly to the binomial distribution. However, he did not outline how to compare GP using the binomial distribution. It is important to emphasize that logistic regression or mixed logistic regression cannot be applied in this scenario either (Heinze and Schemper, 2002; Gianinetti, 2020), as it corresponds to the situation of perfect separation in logistic regression. As Albert and Anderson (1984) investigated, when there is a combination of predictor variables that can perfectly predict the outcome variable, logistic regression models may yield infinite parameter estimates or extremely large values. These infinite estimates lead to very inaccurate test results, as will be demonstrated in our numerical simulations. Some researchers, cognizant of this issue, have eliminated data where GPs are all zeros or ones within a treatment before performing statistical analysis (see Yang et al., 2011, Table 1). Considering the experimental significance of conditions where GPs are consistently zero or one, we believe that such an approach constitutes an incomplete solution.

The second issue pertains to the utilization of GP as raw data for statistical analysis. While many studies have examined whether the GP or its transformation satisfies ANOVA assumptions (Ahrens et al., 1990; Sileshi, 2012; Malik and Piepho, 2016), our perspective differs. We contend that when only GPs are employed as data, an issue arises since crucial information, such as the number of tested seeds utilized to produce each GP, is overlooked in the analysis. Figure 1 shows the ANOVA result from analysis of GP data according to gibberellic acid (GA) concentration in Rhie et al. (2016), when the number of tested seeds is reduced to one-

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

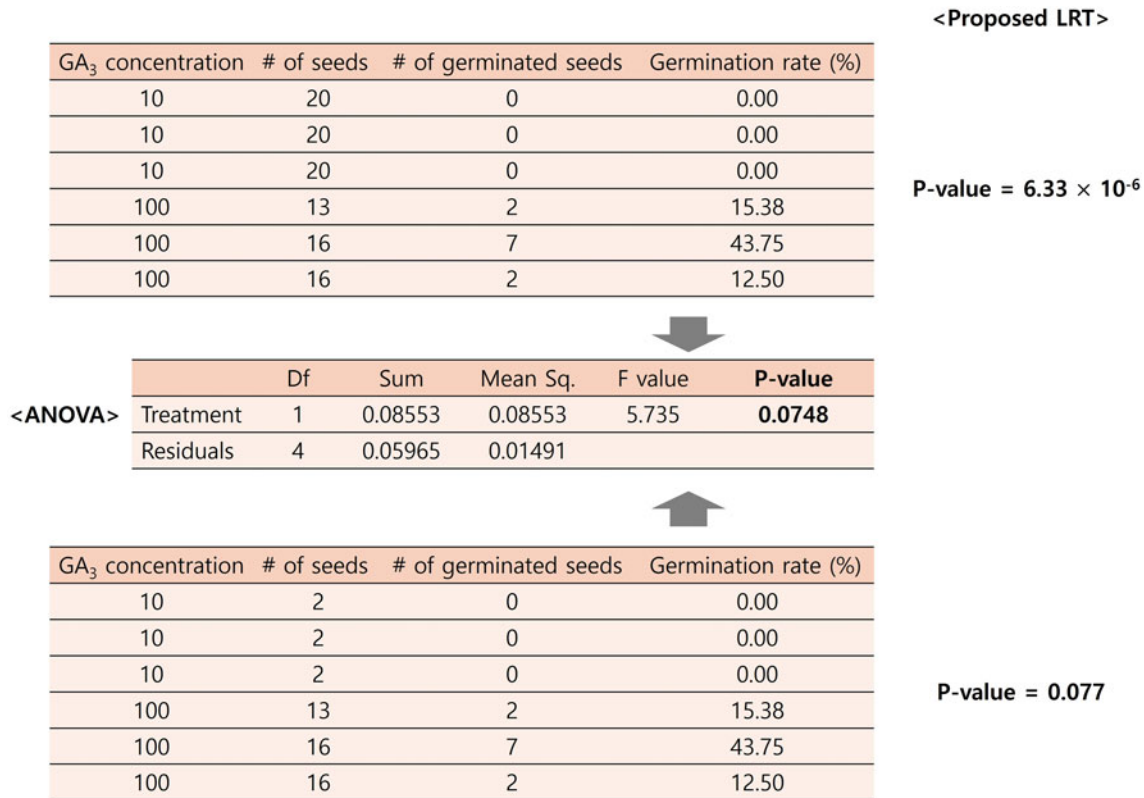


Figure 1. The challenges associated with analysing GP data using the ANOVA method are compared to the results obtained from the proposed LRT.

tenth for GA concentration 10. The distinction in confidence or reliability of a GP that is zero between the two analyses is not considered by the ANOVA, resulting in the same P -values. Given this consideration, we believe that analysing the data solely based on GP or its transformation may not be appropriate. Our objective is to propose a statistical method for comparing GP using binomial data (i.e., the number of tested and germinated seeds) but without the first problem. To overcome this issue, we use a likelihood ratio test (LRT), which is a well-established statistical method. Although the LRT has been utilized before to test the significance of regression coefficients in logistic regression and assess variations in GP among treatments, no previous efforts have been made to utilize the LRT directly for testing the significance of treatment differences in GP. The LRT we employed does not have the first and second problems and exhibits superior power in both the overall mean comparison and post-hoc analysis in numerical simulations. The rest of our paper is organized as follows. A detailed description of the proposed LRT method for analysing GP is presented in the second section. The third section compares the performance of the proposed method with existing germination rate comparison methods in various GP data scenarios through numerical simulation. In the fourth section, the effectiveness of the proposed LRT method is validated by reanalysing GP data that was previously analysed using existing methods. Finally, key discussion points for further consideration are provided in the last section.

LRT for analysing GP

We assume the situation where one treatment factor with I levels exists and N_i repetitions (e.g. Petri dishes, trays, plates, etc.) are done for each treatment level i . We use i for denoting the treatment level and j for the replication. Let n_{ij} and g_{ij} be the number

of tested seeds and the number of germinated seeds for the j th replication of the i th treatment level, respectively. Finally, p_i is the GP of the i th treatment level. In this situation, we are interested in testing whether there is a significant difference in GP between the treatment levels ($H_0: p_1 = \dots = p_I$ vs. $H_1: \text{not } H_0$) and which levels differ significantly from one another in case there is a difference. Our GP data are (g_{ij}, n_{ij}) , $i = 1, \dots, I$, $j = 1, \dots, N_i$.

The LRT is a statistical test used to compare the fit of two nested statistical models. A nested model can be obtained by imposing restrictions on the parameters of a more general model. In our case, the model specified under H_1 is $g_{ij} \sim \text{binomial}(n_{ij}, p_i)$ for $i = 1, \dots, I$, while the model under H_0 is $g_{ij} \sim \text{binomial}(n_{ij}, p)$ for $i = 1, \dots, I$, subject to the constraint $p_1 = \dots = p_I$. Consequently, under H_0 , there is only one parameter. The likelihood function measures the probability of the observed data, given the model and its parameters. Therefore, the likelihood ratio compares the support provided by the data for one model relative to the other.

In our case of analysing GP, the estimated parameters under H_0 and H_1 through the maximum likelihood method are

$$\hat{p}_{i,H_0} = \frac{\sum_{i=1}^I \sum_{j=1}^{N_i} g_{ij}}{\sum_{i=1}^I \sum_{j=1}^{N_i} n_{ij}} \quad \text{and} \quad \hat{p}_{i,H_1} = \frac{\sum_{j=1}^{N_i} g_{ij}}{\sum_{j=1}^{N_i} n_{ij}}, \quad (1)$$

respectively. Note that \hat{p}_{i,H_0} does not depend on the treatment level i , but \hat{p}_{i,H_1} does depend. \hat{p}_{i,H_0} is the pooled estimate across all treatment groups. Furthermore, the log-likelihoods of the two models are given as follows:

$$l(H_0) = l(\hat{p}_{i,H_0}; \text{data}) = \log \left(\prod_{i=1}^I \prod_{j=1}^{N_i} \hat{p}_{i,H_0}^{g_{ij}} (1 - \hat{p}_{i,H_0})^{n_{ij} - g_{ij}} \right) \quad (2)$$

Table 1. Detection rate of whether there is a difference in germination rates when $(p_1, p_2, p_3, p_4) = (0, 0, 0.15, 0.5)$ (Scenario A)

	ANOVA	ANOVA-AS	KW	Logistic	MCPAN	LRT
$n_{ij} = 15$	1.000	1.000	1.000	1.000	1.000	1.000
$n_{ij} = 30$	1.000	1.000	1.000	1.000	1.000	1.000

$$l(H_1) = l(\hat{p}_{i,H_1}; \text{data}) = \log \left(\prod_{i=1}^I \prod_{j=1}^{N_i} \hat{p}_{i,H_1}^{g_{ij}} (1 - \hat{p}_{i,H_1})^{n_{ij} - g_{ij}} \right) \quad (3)$$

The LRT typically calculates two times the difference of the two log-likelihoods $\Lambda = 2(l(H_1) - l(H_0))$ and decides which model to support depending on the size of Λ . When Λ is large, we support the model under H_1 and when Λ is close to zero, we support the model H_0 . Note that $\Lambda \geq 0$ since the model under H_1 nests the model under H_0 . Thanks to Wilks' theorem (Wilks, 1938), we know that under H_0 the test statistic Λ is asymptotically chi-squared distributed with degrees of freedom $I - 1$, which is the difference in the number of parameters of the two models. Simply speaking, when Λ is greater than the $(1 - \alpha)$ -quantile of the chi-square distribution with degrees of freedom $I - 1$, then we conclude that there is a significant difference in germination rates at significance level α .

One of the significant benefits of the LRT framework we are exploring is that it permits testing even when the GPs are all zeros or ones at particular treatment levels. Moreover, unlike ANOVA, it considers the difference in confidence in the germination rate that arises from variations in the number of seeds used in each repetition when testing the difference in GP. Figure 1 shows that the LRT test does consider the number of tested seeds utilized to compare GP across levels. Once it is confirmed that the GP is not equal as a whole, a post-hoc test can be conducted to pinpoint the specific treatment levels contributing to the GP difference. This can be achieved by conducting our LRT test pairwise and adjusting the P -values for multiple comparisons. Alternatively, other tests designed for comparing two proportions can be utilized for pairwise comparisons. Based on our simulation results (not presented in the paper), it appears that the

pairwise comparison with the LRT yields slightly better outcomes compared to the two-sample proportion test. Our LRT method, which we have implemented in our developed R package called *gerstat*, can be executed by utilizing the function *ger_test*. For more information about our R package, please refer to the Appendix.

Numerical simulation

This section assesses the efficacy of our proposed LRT method in analysing GP data relative to conventional methods under various scenarios. The methods we compare include ANOVA, ANOVA with arcsine transformed response, logistic regression, and Kruskal–Wallis test. Furthermore, we consider the test suggested by Sill (2007) and Schaarschmidt et al. (2008) and will henceforth refer to this method as the MCPAN method. For our simulations, we employ a fully randomized design with three replications at each level and four treatment levels.

Three primary simulation scenarios were taken into consideration: (A) a situation where specific levels exhibit seed dormancy, (B) a scenario where there is no seed dormancy, but a difference in GP exists among the levels and (C) a situation where there is no seed dormancy, and there is no difference in GP. For Scenario A, the first two treatment levels exhibit seed dormancy ($p_1 = p_2 = 0$), but the third and fourth treatment levels have GP of $p_3 = 0.15$ and $p_4 = 0.5$, respectively. For Scenario B, we have $(p_1, p_2, p_3, p_4) = (0.1, 0.2, 0.3, 0.4)$. Finally, we have the same GP ($p_1 = p_2 = p_3 = p_4 = 0.3$) all across the levels. We generate all the data from the model $g_{ij} \sim \text{binomial}(n_{ij}, p_i)$, $i = 1, \dots, 4$, with n_{ij} being 15 or 30 (the number of seeds per replication). The number of simulations is 1000 and the significance level α is 0.05.

Once it is established that the GP values are not uniformly equal, a post-hoc test is performed to identify the specific

Table 2 Percentage that identified a difference in germination rate for a given treatment pair when $(p_1, p_2, p_3, p_4) = (0, 0, 0.15, 0.5)$ (Scenario A)

		p_1 vs. p_2	p_1 vs. p_3	p_1 vs. p_4	p_2 vs. p_3	p_2 vs. p_4	p_3 vs. p_4
$n_{ij} = 15$	ANOVA	0.000	0.344	0.992	0.344	0.992	0.909
	ANOVA-AS	0.000	0.742	1.000	0.742	1.000	0.915
	KW	0.000	0.000	0.891	0.000	0.891	0.001
	Logistic	0.000	0.000	0.000	0.000	0.000	0.840
	MCPAN	0.000	0.500	1.000	0.500	1.000	0.844
	LRT	0.000	0.914	1.000	0.914	1.000	0.924
	$n_{ij} = 30$	ANOVA	0.000	0.637	1.000	0.637	1.000
ANOVA-AS		0.000	0.970	1.000	0.970	1.000	0.998
KW		0.000	0.000	0.994	0.000	0.994	0.000
Logistic		0.000	0.000	0.000	0.000	0.000	0.992
MCPAN		0.000	0.989	1.000	0.989	1.000	0.996
LRT		0.000	0.999	1.000	0.999	1.000	1.000

Note. The treatment pair (p_1 vs. p_2) has no difference in germination rate.

Table 3 Detection rate of whether there is a difference in germination rates when $(p_1, p_2, p_3, p_4) = (0.1, 0.2, 0.3, 0.4)$ (Scenario B)

	ANOVA	ANOVA-AS	KW	Logistic	MCPAN	LRT
$n_{ij} = 15$	0.641	0.621	0.427	0.870	0.869	0.870
$n_{ij} = 30$	0.915	0.916	0.745	0.997	0.996	0.997

treatment levels contributing to the GP difference. This post-hoc test entails conducting multiple pairwise comparison tests. To control the overall rate of false positives, multiple testing methods are implemented during the post-hoc test to control either the false positive rate (FPR) or the family-wise error rate (FWER). As multi-step methods are more efficient in controlling the FPR or the FWER than single-step methods, we utilized the Holm procedure, a multi-step method, for the entire simulation. To conduct pairwise comparison tests in both ANOVA and logistic regression, we utilized the `glht` function from the R package `multcomp`. This package streamlines pairwise comparisons across diverse models by furnishing suitable inference procedures for testing general linear hypothesis tests. In the context of Kruskal–Wallis, we conducted pairwise comparisons similar to Dunn’s test (Dunn, 1964), a frequently employed post-hoc test in nonparametric scenarios. As for the LRT method, we executed the pairwise comparison as outlined in the section ‘LRT for analysing GP’. To apply the MCPAN method, we conducted pairwise comparisons and adjusted *P*-values following the guidelines provided by Sill (2007). The MCPAN method was implemented using the function `binomRDtest` in the R package MCPAN.

In our comparison of the methodologies under consideration, we evaluate their performance on two key dimensions. Firstly, we calculate the proportion of overall tests that have detected differences between groups, if any. Secondly, we assess the accuracy of the follow-up tests in detecting treatment pairs with distinct GPs. Although various GP vectors could be examined for a given scenario, such simulations (which are not included in the paper) did not yield significantly different conclusions compared to those presented in the paper.

Scenario A ($(p_1, p_2, p_3, p_4) = (0, 0, 0.15, 0.5)$)

In this subsection, we present the results obtained when certain treatment levels exhibit seed dormancy. As shown in

Table 1, the performance differences between the methodologies are insignificant when it comes to testing for differences in GP. However, as highlighted in Table 2, the methods exhibit noticeable performance differences in identifying the specific treatment levels that contribute to the difference in GP. Logistic regression struggles with detecting GP differences due to the issue of complete separation (Mansournia et al., 2017). On the other hand, the Kruskal–Wallis test has a lower power, which is a characteristic of nonparametric methods, and can only detect pairs with relatively large differences in germination rates ($p_1 = 0$ vs. $p_4 = 0.5$ or $p_2 = 0$ vs. $p_4 = 0.5$). In most cases, the proposed LRT demonstrates superior performance compared to ANOVA, arcsine-transformed response-based ANOVA and the MCPAN method in detecting treatment pairs with distinct germination rates.

Scenario B ($(p_1, p_2, p_3, p_4) = (0.1, 0.2, 0.3, 0.4)$)

In this subsection, we investigate the scenario where there is no seed dormancy, but a difference in GP exists among the treatment levels. Table 3 shows that the proposed LRT, logistic regression and the MCPAN method outperform the other methods in detecting whether there are differences in germination rates. Concerning the identification of specific treatment levels contributing to the difference in GP (Table 4), two primary trends are evident. Firstly, the Kruskal–Wallis test continues to exhibit low detection power. Secondly, the proposed LRT, logistic regression and the MCPAN method consistently outperform ANOVA and arcsine-transformed response-based ANOVA in the majority of cases. Notably, the proposed LRT demonstrates the best performance.

Scenario C ($(p_1, p_2, p_3, p_4) = (0.3, 0.3, 0.3, 0.3)$)

In Scenario C, there are no differences in germination rates among the treatment levels. Consequently, it is crucial to control the power of the overall test close to the significance level. Similarly, when employing multiple testing methods, it is

Table 4 Percentage that identified a difference in germination rate for a given treatment pair when $(p_1, p_2, p_3, p_4) = (0.1, 0.2, 0.3, 0.4)$ (Scenario B)

		p_1 vs. p_2	p_1 vs. p_3	p_1 vs. p_4	p_2 vs. p_3	p_2 vs. p_4	p_3 vs. p_4
$n_{ij} = 15$	ANOVA	0.033	0.219	0.507	0.066	0.233	0.070
	ANOVA-AS	0.066	0.260	0.495	0.044	0.147	0.027
	KW	0.000	0.014	0.243	0.004	0.008	0.000
	Logistic	0.049	0.350	0.731	0.061	0.269	0.054
	MCPAN	0.122	0.468	0.824	0.071	0.311	0.057
	LRT	0.150	0.471	0.802	0.084	0.311	0.065
$n_{ij} = 30$	ANOVA	0.104	0.503	0.859	0.147	0.507	0.139
	ANOVA-AS	0.182	0.571	0.862	0.126	0.396	0.084
	KW	0.000	0.016	0.524	0.000	0.006	0.000
	Logistic	0.283	0.805	0.988	0.172	0.674	0.139
	MCPAN	0.266	0.810	0.992	0.146	0.666	0.122
	LRT	0.329	0.837	0.990	0.191	0.690	0.145

Table 5 Detection rate of whether there is a difference in germination rates when $(p_1, p_2, p_3, p_4) = (0.1, 0.2, 0.3, 0.4)$ (Scenario C)

	ANOVA	ANOVA-AS	KW	Logistic	MCPAN	LRT
$n_{ij} = 15$	0.030	0.034	0.016	0.044	0.043	0.044
$n_{ij} = 30$	0.034	0.036	0.016	0.042	0.041	0.042

important to ensure that the percentage of detected differences in each treatment pair is controlled at a value near the significance level. Upon analysing the findings presented in Tables 5 and 6, we observe that all the methods exhibit these characteristics. However, it is notable that the Kruskal–Wallis test proves to be conservative, with a detection rate lower than the significance level.

An important observation across all scenarios is that both logistic regression and the proposed LRT yield identical results concerning the test for overall differences. This alignment arises from the equality between the deviation in deviance calculated between the full model and the null model (the model with only the intercept) used by logistic regression and our proposed LRT statistic. Despite the potential instability in coefficient estimates when there are consistent zeros or ones across all replicates in logistic regression, it does not adversely affect the likelihood values of the estimated model. Consequently, the detection rate of overall differences in logistic regression remains robust in situations with perfect separation. However, concerning pairwise comparisons, reliance on coefficient estimates significantly hampers testing performance in situations of perfect separation.

Reanalysis of the data from Rhie et al. (2016)

To demonstrate the advantages of our method, we conducted a reanalysis of the data from Rhie et al. (2016)'s study that examines the effects of GA₃ on germination under low temperatures. Figure 2 presents the raw data and the corresponding testing results. In Rhie et al. (2016)'s original analysis, ANOVA was employed to test the overall difference, followed by the utilization of Tukey's honestly significant difference (HSD) test as a post-hoc test with a specific significance level $\alpha = 0.05$. In our reanalysis, we applied the LRT described in the section 'Numerical simulation'

with the same significance level $\alpha = 0.05$. It is important to note that both the Holm–Bonferroni correction, which we used for our LRT analysis, and Tukey's HSD method for ANOVA are designed to control the family-wise error rate. Upon comparing the two analyses, it was determined that both reached the same conclusion that there is a difference in GP. However, there were variations in the results for the treatment pair that exhibited a difference in GP. In Rhie et al. (2016)'s analysis, it was observed that there was no significant difference in GP between the 100 and 10 mg L⁻¹ doses of GA₃. However, it is challenging to accept the conclusion that there was no significant difference in germination between the two treatments. This is due to the fact that no germination occurred at the 10 mg L⁻¹ dose of GA₃, while certain replications at the 100 mg L⁻¹ dose showed high germination rates of up to 43%. The substantial disparity in germination outcomes between the two treatments raises doubts about the conclusion of no significant difference. In contrast, the LRT-based analysis classifies the 0 and 10 mg L⁻¹ treatments, which displayed no germination, into the same germination rate group (*a*). Furthermore, it distinguishes the 100 and 1000 mg L⁻¹ treatments, which exhibited a significant difference in average germination, into distinct germination rate groups (*b* and *c*). Considering our reanalysis of Rhie et al. (2016)'s data and the findings from our simulations, where the ANOVA method occasionally failed to detect differences between treatments with zero germination and those without, we can conclude that the proposed LRT method can provide more reliable conclusions than traditional methods in various GP data analysis situations.

Discussion

This paper introduces a novel germination rate comparison method based on LRT, which performs effectively even in

Table 6 Percentage that identified a difference in germination rate for a given treatment pair when $(p_1, p_2, p_3, p_4) = (0.3, 0.3, 0.3, 0.3)$ (Scenario C)

		p_1 vs. p_2	p_1 vs. p_3	p_1 vs. p_4	p_2 vs. p_3	p_2 vs. p_4	p_3 vs. p_4
$n_{ij} = 15$	ANOVA	0.006	0.008	0.008	0.007	0.003	0.007
	ANOVA-AS	0.008	0.007	0.007	0.006	0.005	0.005
	KW	0.000	0.004	0.001	0.000	0.001	0.001
	Logistic	0.005	0.005	0.005	0.007	0.004	0.004
	MCPAN	0.011	0.006	0.013	0.009	0.008	0.007
	LRT	0.008	0.006	0.007	0.007	0.005	0.006
$n_{ij} = 30$	ANOVA	0.006	0.008	0.006	0.008	0.002	0.003
	ANOVA-AS	0.008	0.007	0.006	0.009	0.002	0.002
	KW	0.000	0.002	0.002	0.002	0.001	0.002
	Logistic	0.007	0.006	0.004	0.008	0.004	0.005
	MCPAN	0.013	0.007	0.010	0.009	0.007	0.009
	LRT	0.008	0.007	0.006	0.008	0.005	0.005

GA ₃ concentration	# of seeds	# of germinated seeds	Germination rate (%)
0	20	0	0.00
0	20	0	0.00
0	20	0	0.00
10	20	0	0.00
10	20	0	0.00
10	20	0	0.00
100	13	2	15.38
100	16	7	43.75
100	16	2	12.50
1000	15	12	80.00
1000	12	11	91.67
1000	13	10	76.92

<ANOVA>		<Proposed LRT>	
GA ₃ concentration	Final germination (%)	Final germination (%)	
0 mg·L ⁻¹	0.0 a	0.0 a	
10 mg·L ⁻¹	0.0 a	0.0 a	
100 mg·L ⁻¹	23.9 a	23.9 b	
1000 mg·L ⁻¹	82.9 b	82.9 c	
Significance	***	***	

Figure 2. Reanalysis of the data from Rhie et al. (2016)'s study. (***) means that the germination rate difference is significant at $P < 0.001$. The use of different letters to mark the averages signifies that the corresponding treatments have been identified as having significantly different germination rates through multiple comparisons.)

scenarios where all germination rates for a particular treatment can be either 0 or 1. However, it is crucial to emphasize that our paper exclusively addresses the analysis of a single treatment factor. In real germination rate experiments, the consideration of multiple treatment factors is plausible. We are currently working on extending our proposed methodology to encompass cases involving two treatment factors. Furthermore, it is worth mentioning that our methodology did not address the issue of overdispersion, which is a significant concern in the analysis of germination rate data. In practice, when different Petri dishes are employed within the same treatment, the Petri dish effect can influence the germination rate. It is customary to model this effect as a random effect in order to account for its impact on the observed data. Considering and incorporating this aspect of overdispersion is crucial for a more comprehensive analysis of germination rate data. The development of a methodology that simultaneously considers the random effect and addresses the scenario where germination rates in a specific treatment can be 0 or 1 is a topic that deserves thorough research. In line with this direction, a promising avenue would involve extending Firth's method (Heinze and Schemper, 2002), an alternative approach for resolving complete separation in logistic regression, to accommodate a random effect setting. Conducting a study in this direction would contribute significantly to the field of germination rate data analysis.

Acknowledgements. The authors are indebted to one reviewer and the Editor for their excellent comments and suggestions. Yongha Rhie would like to thank the support of R&D Program for Forest Science Technology provided by the Korea Forest Service (Korea Forestry Promotion Institute). Partial financial support is acknowledged from the Korea Forest Service (Korea Forestry Promotion Institute) (Project No. 2021380A00-2123-BD02).

Competing interest. The authors declare no conflict of interest.

References

- Ahrens WH, Cox DJ and Budhwar G (1990) Use of the arcsine and square root transformations for subjectively determined percentage data. *Weed Science* 38, 452–458.
- Albert A and Anderson JA (1984) On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 71, 1–10.
- Dunn OJ (1964) Multiple comparisons using rank sums. *Technometrics* 6, 241–252.
- Gianinetti A (2020) Basic features of the analysis of germination data with generalized linear mixed models. *Data* 5, 6.
- Heinze G and Schemper M (2002) A solution to the problem of separation in logistic regression. *Statistics in Medicine* 21, 2409–2419.
- Jaeger TF (2008) Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59, 434–446.
- Malik WA and Piepho H-P (2016) On generalized exponential transformations for proportions. *Communications in Statistics – Theory and Methods* 45, 5857–5870.

- Mansournia MA, Geroldinger A, Greenland S and Heinze G** (2017) Separation in logistic regression: causes, consequences, and control. *American Journal of Epidemiology* **187**, 864–870.
- Onofri A, Gresta F and Tei F** (2010) A new method for the analysis of germination and emergence data of weed species. *Weed Research* **50**, 187–198.
- Rhie YH, Kim J, Lee SY and Kim KS** (2016) Non-deep simple morphophysiological dormancy in seeds of heavenly bamboo (*Nandina domestica* Thunb.). *Scientia Horticulturae* **210**, 180–187.
- Ritz C, Phipper CB and Streibig JC** (2013) Analysis of germination data from agricultural experiments. *European Journal of Agronomy* **45**, 1–6.
- Schaarschmidt F, Sill M and Hothorn LA** (2008) Approximate simultaneous confidence intervals for multiple contrasts of binomial proportions. *Biometrical Journal* **50**, 782–792.
- Sileshi GW** (2012) A critique of current trends in the statistical analysis of seed germination and viability data. *Seed Science Research* **22**, 145–159.
- Sill M** (2007) *Approximate Simultaneous Confidence Intervals for Multiple Comparisons of Binomial Proportions*. Technical Report, Institute of Biostatistics, Leibniz University Hannover.
- Warton DI and Hui FKC** (2011) Arcsine is asinine: the analysis of proportions in ecology. *Ecology* **92**, 3–10.
- Wilks SS** (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* **9**, 60–62.
- Yang CJ, Chien CT, Liao YK, Chen SY, Baskin JM, Baskin CC and Kuo-Huang LL** (2011) Deep simple morphophysiological dormancy in seeds of the basal taxid *Cephalotaxus*. *Seed Science Research* **21**, 215–226.

APPENDIX

Using gerstat for method implementation

We have created the R package *gerstat* to empower researchers in analysing germination rate data with a wide range of germination rate comparison methods, including the proposed LRT approach. The R package *gerstat* can be installed from our github repository. To install the package, please verify that the R base (version $\geq 4.3.1$) is installed on your system and follow these R code instructions:

```
if (!requireNamespace("devtools", quietly = TRUE)) {
  install.packages("devtools")
}
library(devtools)
devtools::install_github("Noh-Hohsuk/gerstat")
```

As *gerstat* is an optional package, it needs to be loaded in each new R session before using its functions. To load the *gerstat* package, you can use the `library()` command:

```
library(gerstat)
```

For germination rate testing, the function *ger_test* can be used.

Germination rate test using gerstat

The *ger_test* function executes a germination rate comparison test based on the specified method. If the difference in germination rate comparison is statistically significant at the designated significance level, it generates a compact letter display that groups treatments with similar germination rates, utilizing the specified multiple comparison method. The basic usage is the following:

```
ger_test(gdata = data, method = "LRT", p_adjust_mtd = "holm", ctr_lv = 0.05)
```

The *gdata* is expected to be an R data frame with three variables: *treatment*, *num_seed* and *ger_seed*. The *treatment* variable should be a factor representing the treatment level, *num_seed* should indicate the number of seeds used in each treatment iteration and *ger_seed* should indicate the number of seeds that germinated. The method parameter specifies the statistical method for germination rate comparison. The available options for the method are 'LRT' (likelihood ratio test), 'ANOVA' (analysis of variance), 'ASANOVA' (ANOVA with arcsine square root transformed response), 'logistic' (logistic regression) or 'KW' (Kruskal–Wallis test). The *p_adjust_mtd* parameter specifies the *P*-value correction method for multiple comparisons. The *ctr_lv* parameter represents both the significance level for the overall test and the control level for multiple testing. For more information, you can use the command `?ger_test`. To reproduce the analysis of the data from Rhie et al. (2016) using the *ger_test* function, you can execute the following R code:

```
data(nandina)
ger_test(gdata = nandina, method = "LRT", p_adjust_mtd = "holm", ctr_lv = 0.05)
The output is
$p_value [1]
0$clld
GA0 GA10 GA100 GA1000
"a" "a" "b" "c"
```

The *P*-value represents the result from the overall test, while the *clld* (compact letter display) is generated based on multiple testing. When the *P*-value is lower than the specified significance level, the *clld* value will be NULL.