Computational Statistics (2023) 38:1791–1810
https://doi.org/10.1007/s00180-022-01287-4

ORIGINAL PAPER

Bayesian multilevel logistic regression models: a case study applied to the results of two questionnaires administered to university students

Cristian David Correa-Álvarez[1] · Juan Carlos Salazar-Uribe[2] · Luis Raúl Pericchi-Guerra[3]

Received: 19 October 2021 / Accepted: 14 September 2022 / Published online: 25 October 2022

Abstract

Bayesian multilevel models—also known as hierarchical or mixed models—are used in situations in which the aim is to model the random effect of groups or levels. In this paper, we conduct a simulation study to compare the predictive ability of 1-level Bayesian multilevel logistic regression models with that of 2-level Bayesian multilevel logistic regression models by using the prior Scaled Beta2 and inverse-gamma distributions to model the standard deviation in the 2-level. Then, these models are employed to estimate the correct answers in two questionnaires administered to university students throughout the first academic semester of 2018. The results show that 2-level models have a better predictive ability and provide more precise probability intervals than 1-level models, particularly when the prior Scaled Beta2 distribution is used to model the standard deviation in the second level. Moreover, the probability intervals of 1-level Bayesian multilevel logistic regression models proved to be more precise when Scaled Beta2 distributions, rather than an inverse-gamma distribution, are employed to model the standard deviation or when 1-level Bayesian multilevel logistic regression models, are used.

Keywords Bayesian inference · Prior distributions · Multilevel models · Bayesian logistic models

Juan Carlos Salazar-Uribe and Luis Raúl Pericchi-Guerra have contributed equally to this work.

✉ Cristian David Correa-Álvarez
cristiancorrea@itm.edu.co

Extended author information available on the last page of the article

1 Introduction

Multilevel models—also known as hierarchical or mixed models—are used to model data with levels or hierarchies. They naturally occur in various everyday situations given the hierarchical or nested structure of the sampling units and are widely employed in fields such as social sciences, medicine, education, and reliability (Gaviria Morera 2005). According to Gaviria Morera (2005) and McElreath (2015), some of their most notable advantages are that they (1) provide better estimates for sampling when observations arise from the same individual, (2) explicitly model variation when the research questions include variation among individuals, and (3) make it possible to avoid data transformations.

Multilevel models are mainly characterized by modeling the random effect within groups at different levels. In Bayesian statistics, these models are said to occur naturally because there exists a multilevel structure in which the prior distribution represents one level of the model and the likelihood function constitutes the final stage, which results in the posterior distribution (Pinheiro and Bates 2006; Mason 1985; Bornmann et al. 2016; Gelman et al. 2013; Ntzoufras 2011).

Given the hierarchical nature of everyday situations, Bayesian multilevel logistic regression models have been increasingly used in recent years (Ayalew 2020; Aychiluhm et al. 2020; Gañan-Cardenas et al. 2021; Młynarczyk et al. 2021; Jabessa and Jabessa 2021; Fagbamigbe et al. 2021; Grogan-Kaylor et al. 2021; Sherwood et al. 2021). For example, Fagbamigbe et al. (2021) employed a Bayesian multilevel model to examine the main risk factors and regional variations in maternal mortality in Ethiopia. Similarly, Grogan-Kaylor et al. (2021) analyzed the association of physical punishment and nonphysical discipline with child socio-emotional functioning using Bayesian multilevel logistic regression models.

Other authors who have also employed multilevel models in their research studies include: Jara et al. (2008), Birlutiu et al. (2010), Tang and Duan (2014), De la Cruz et al. (2016), Lu et al. (2017) and Wang et al. (2019). For instance, Birlutiu et al. (2010) modeled multitask learning in hearing-impaired subjects using a hierarchical approach. Wang et al. (2019) presented a 10-minute trivia game-based activity for the intuitive understanding of confidence intervals. They fitted a mixed-effects Bayesian logistic model using noninformative priors for the coefficients and variance parameters. In this paper, we will develop an activity similar to that proposed by Wang et al. (2019), based on the results of two questionnaires that were administered to university students throughout the first academic semester of 2018.

For their part, Pérez et al. (2017) proposed the Scaled Beta2 (SBeta2) distribution as an alternative to the inverse-gamma distribution for modeling variances and standard deviations (or, more generally, for scales). They presented it as a flexible and tractable family that can be used to model scales in both hierarchical and nonhierarchical settings. In addition, they demonstrated that, for certain values of the hyperparameters, the mixture of a normal and an SBeta2 distribution gives a closed form marginal. Hence, we will here employ the SBeta2 and inverse-gamma distributions to model the variance parameters in the 2-level Bayesian multilevel logistic model. In addition, the resulting estimates of the models fitted with these two distributions will be compared by means of a simulation study to determine which distribution has the best predictive ability.

The rest of this paper is structured as follows. In Sect. 2, we provide an overview of Bayesian multilevel logistic regression and present the general models employed here. In Sect. 3, we introduce the SBeta2 distribution used to model the variance parameters. In Sect. 4, we present a simulation study to compare the performance of the proposed models. In Sect. 5, we apply the proposed models to real-life data. Finally, we outline the conclusions of this study.

2 Bayesian multilevel logistic regression

Multilevel models are extensions of regression models, with data structured in groups and coefficients that can vary by group. From the Bayesian perspective, Bayesian models are said to have an inherently multilevel or hierarchical structure. The prior distribution, $\xi(\theta \mid \phi)$, of a model with prior parameters ϕ can be considered one level or hierarchy, with the likelihood as the final stage of a Bayesian model, which results in the posterior distribution, $\xi(\theta \mid y) \propto f(y \mid \theta)\xi(\theta \mid \phi)$, obtained using Bayes' theorem. A key aspect in these models is that ϕ is unknown; therefore, it will have its own prior distribution, $\xi(\phi)$. The appropriate Bayesian posterior distribution is that of vector (ϕ, θ) (Mason 1985; Bornmann et al. 2016; Gelman et al. 2013; Ntzoufras 2011). The joint prior distribution is given by

$$\xi(\phi, \theta) = \xi(\theta \mid \phi)\xi(\phi). \tag{1}$$

Then, the joint posterior distribution is given by

$$\xi(\phi, \theta \mid y) \propto \xi(\phi, \theta)f(y \mid \phi, \theta) = \xi(\theta \mid \phi)\xi(\phi)f(y \mid \theta). \tag{2}$$

This latter simplification holds because y and ϕ are conditionally independent given θ. That is, the data distribution $f(y \mid \phi, \theta)$, depends only on θ; the hyperparameters ϕ affect y only through θ (Gelman et al. 2013).

In Bayesian multilevel logistic regression models, the assumption of interchangeability is employed when there is no information on the structure of the parameters of interest. It is assumed that all random parameters come from a common distribution and that their ordering does not affect the model or the results. In other words, the prior distributions of the hyperparameters are invariant to the random permutations of these latter (Ntzoufras 2011). For further elaboration on this concept, see the studies by Bernardo and Smith (2000) and Gelman et al. (2013).

Bayesian multilevel (or hierarchical) logistic regression models can be used to model clustered data having a binary response variable. Such is the case of logistic regression, which is considered a standard way of modeling this type of variables (i.e., data that take the values of 1 or 0). Results of these variables can be found in studies into, for instance, social, medical, and natural matters, in which the response is usually the presence or absence of a characteristic of interest (Peng et al. 2002; Gelman and Hill 2006; Pregibon 1981; Ntzoufras 2011). According to this, the dependent or response variable, y, is assigned a value of 1 if it has the characteristic of interest or a value of 0 if it does not.

In logistic regression, a single outcome variable, y_i $(i = 1, ..., n)$, follows a Bernoulli probability function that takes the value of 1 with probability π_i or the value of 0 with probability $1 - \pi_i$ and is denoted by $y_i \sim$ Bernoulli (π_i). This type of regression adopts the logit link, which, besides being the most suitable choice because it is the canonical link, has a smooth and pleasant interpretation based on the probability ratio, $\pi_i/(1 - \pi_i)$ (King and Zeng 2001; Ntzoufras 2011).

Let us suppose that, in a study seeking to model the correct answers in a questionnaire administered to groups of students, $y_{ij} = 1$ $(i = 1, \ldots, N_j;\ j = 1, \ldots, J)$ if the i^{th} question is correctly answered by the individual in the j^{th} group and $y_{ij} = 0$ if it is not. In this case, the Bayesian hierarchical logistic model is given by

$$y_j \sim \text{ Binomial } (\pi_j, N_j), \tag{3}$$

where N_j is the number of questions in group j and $y_j = \sum_{i=1}^{N_j} y_{ij}$. The link function is given by

$$\text{logit } (\pi_j) = \log\left(\frac{\pi_j}{1 - \pi_j}\right) = \theta_j, \tag{4}$$

$$\theta_j \sim N(\mu_\theta, \sigma_\theta^2), \tag{5}$$

for $j = 1, \ldots, J$, with J number of groups, μ_θ follows a noninformative normal distribution and σ_θ^2 has a defined distribution on positive values.

If the questionnaire is assumed to be administered twice within each group, the previous model is then given by

$$y_{kj} \sim \text{ Binomial } (\pi_{kj}, N_{kj}), k = 1, 2 \text{ and } j = 1, \ldots, J. \tag{6}$$

The link function is given by

$$\text{logit } (\pi_{kj}) = \log\left(\frac{\pi_{kj}}{1 - \pi_{kj}}\right) = a_j + \theta_j I(k = 1), \tag{7}$$

$$\theta_j \sim N(\mu_\theta, \sigma_\theta^2), \tag{8}$$

where y_{kj} are the correct answers in the j^{th} group in the k^{th} questionnaire. Parameter θ_j denotes the log odds of their correct answers, while a_j is the log odds of their incorrect answers.

Noninformative prior distributions are commonly employed for parameters μ_θ, a_j, and σ_θ^2. In the specialized literature, $\mu_\theta \sim N(0, 1000), a_j \sim N(0, 1000)$, and σ_θ^2 are often modeled using inverse-gamma, half-Cauchy, uniform, and SBeta2 distributions (Ntzoufras 2011; Pérez et al. 2017). We will here use the SBeta2 and inverse-gamma distributions to model the variance parameters.

In this paper, data from Application 1 in Sect. 5 are modeled using Bayesian logistic models and following a two-step approach. First, the models are fitted per questionnaire both 1-level and 2-level models. Then, a similar procedure is carried out, but this time an independent variable that takes into account the results of the two questionnaires is included. In both scenarios, the results are compared by means of the Deviance Information Criterion (DIC) in order to identify which model exhibits the best predictive ability. In addition to being employed for assessing model adequacy (Spiegelhalter et al. 2002), this criterion is considered a generalization of the Akaike Information Criterion (AIC) and used to compare hierarchical or multilevel models (Ntzoufras 2011).

3 Scaled beta2 distribution

The SBeta2 distribution is a scaled version of the inverse beta distribution and is employed to model precisions and variances in Bayesian hierarchical models. This distribution extends previous proposals, such as those of Gelman (2006) and Berger (2006), for modeling the variance parameter. It is defined as

$$\text{SBeta2}\,(\psi \mid p, q, b) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)b} \frac{\left(\frac{\psi}{p}\right)^{(p-1)}}{\left(\frac{\psi}{p}+1\right)^{(p+q)}} \quad \text{for} \quad \psi > 0, b > 0, p > 0, q > 0. \tag{9}$$

Variable ψ can be represented as the odds ratio, $\psi = \frac{\pi}{1-\pi}$, where $\pi \sim$ Beta $(\pi \mid p, q)$. The expected value and variance of ψ are given by

$$E(\psi) = \frac{p}{q-1}p \quad \text{for} \quad q > 1. \tag{10}$$

$$Var(\psi) = \frac{p(p+q-1)}{(q-1)^2(q-1)} \quad \text{for} \quad q > 2. \tag{11}$$

In the SBeta2 (p, q, b) distribution, parameter p controls the distribution behavior at the origin; parameter q, that in the right-hand tail of the distribution; and parameter b, the scale. This distribution is characterized by its reciprocity, i.e., if $\psi \sim$ SBeta2 (p, q, b), then $1/\psi \sim$ SBeta2 $(p, q, 1/b)$. Other properties of this distribution are described in the study by Pérez et al. (2017). Expression $\psi = \frac{\pi}{1-\pi}b$, with $\pi \sim$ Beta (p, q), can be used as a way to generate random variables, $\psi \sim$ SBeta2 (p, q, b). The main limitations of SBeta2 are that, in practice, it is difficult to choose the values of the scale parameter, and that SBeta2 is not conjugate distribution and hinders the computational process. Pérez et al. (2017) suggest that, in order to obtain a robust inference considering an SBeta2 (p, q, b) distribution, the value of parameter q must be between 0 and 1 because the smaller the q value, the heavier-tailed the distribution. They also recommend selecting a value between 0.5 and 1 for p when the variance distribution is centered around zero to avoid a very large shrinkage towards the mean.

In light of the above, we here use two forms of the SBeta2 distribution—SBeta2(1,1,4) and SBeta2 $(1/2, 1/2, 625)$ plotted in the Fig. 1 and the inverse-gamma distribution—IG(α, β), where parameter α denotes the shape; and parameter β, the scale. In particular, we consider $\alpha = \beta = 0.001$ to model the variance parameters in the 2-level Bayesian multilevel logistic regression model (Llera and Beckmann 2016; Gelman 2006; Pérez et al. 2017; Rojas and Ramírez 2019). It is important to be careful when using the inverse-gamma distribution because studies such as those of Berger (2006), Gelman (2006), Pérez et al. (2017), and Rojas and Ramírez (2019) have outlined the drawbacks of using it as a prior distribution. It, for instance, can lead to posterior distributions with incorrect values or improper posterior distributions.

4 Simulation study: comparing the 1-level and 2-level Bayesian multilevel logistic regression model

This section presents a simulation study in which the scenarios were created to best show the behavior of the Bayesian multilevel logistic regression model. Data were generated from a binomial distribution for different groups and proportions of defects. The focus of this simulation study is to predict the proportions of successes, as well as to compare the predictive ability of the 1-level and 2-level Bayesian multilevel logistic model. The procedure we used is as follows:

1. Four groups (N_j) of different sizes (j) are considered. The sample sizes are 3, 6, 10, and 15, as shown in Table 1.

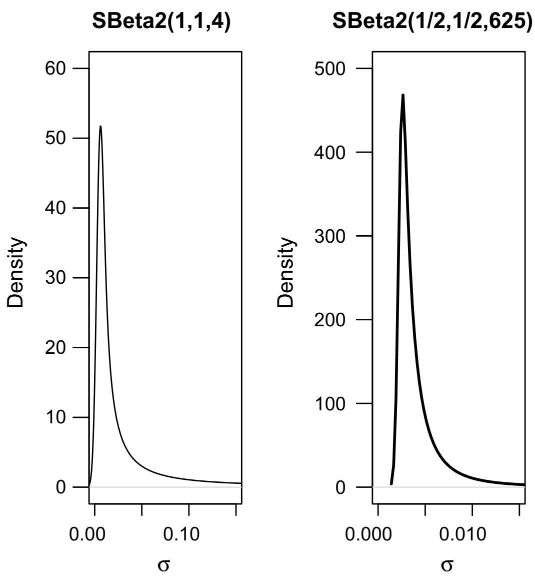

Fig. 1 SBeta2 distribution for the variance parameter with different values of p, q *and* b

2. For each group, the number of correct answers (successes) is estimated using a binomial distribution with a probability of success in three percentiles ($p = 25\%$, 50%, and 75%).
3. With each group, the 1-level and 2-level Bayesian multilevel logistic regression models are fitted, considering the *bugs* function in the R library (Sturtz et al. 2010). This function takes into account, among other values, the group sizes, the number of successes obtained in Item 2, the iterations (40,000), and a burn-in rate of 10% (4,000) to eliminate unstable autocorrelations.
4. The convergence of the posterior distributions of the parameters is tested using the chains obtained in Item 3. Then, the KPSS statistic (Kwiatkowski et al. 1992) (it is implemented in the *tseries* package (Trapletti et al. 2019) in R (R Core Team 2019)) and the p-value of the truncation parameter are calculated. If the p-value is below $\alpha = 0.05$, H_0 (the Markov chain has reached the stationary distribution) is rejected. Finally, graphical analysis are performed based on the lags, ergodic averages, and densities. Other convergence proofs can be found in the studies by Cowles and Carlin (1996) and Brooks and Roberts (1998).
5. The probability intervals, along with the DIC, are calculated in each scenario to assess the predictive ability of the models.

The prior distributions considered to model the variance parameters in 2-level Bayesian multilevel logistic regression were selected based on the recommendations provided by Pérez et al. (2017). In particular, we employed the SBeta2(1,1,4), SBeta2(1/2,1/2,625), and the inverse-gamma distribution traditionally used to model said effect, i.e., IG(0.001, 0.001). For the 1-level Bayesian multilevel logistic regression model, we used noninformative normal distributions, $N(0, 1000)$ (Ntzoufras 2011). The following are the models proposed here:

- Model 1 (M1). With SBeta2(1,1,4)

$$
\begin{aligned}
y_j &\sim \text{ Binomial } (\pi_j, N_j) \\
\text{logit } (\pi_j) &= \theta_j \\
\theta_j &\sim N(\mu_\theta, \sigma^2_\theta) \\
\mu_\theta &\sim N(0, 1000) \\
\sigma^2_\theta &\sim \text{ SBeta2 } (1, 1, 4)
\end{aligned}
$$

Table 1 Sample sizes per group

N_j	j														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3	30	45	60												
6	60	75	90	105	120	135									
10	135	150	165	180	195	210	225	240	255	270					
15	270	285	300	315	330	345	360	375	390	405	420	435	450	465	480

- Model 2 (M2). SBeta2(1,1,4) in model 1 is replaced with SBeta2(1/2,1/2,625).

$$\sigma_\theta^2 \sim \text{ SBeta2 } (1/2, 1/2, 625)$$

.

- Model 3 (M3). SBeta2(1,1,4) in model 1 is replaced with IG(0.001, 0.001).

$$\sigma_\theta^2 \sim \text{ IG } (0.001, 0.001)$$

.

- Model 4 (M4). 1-level model

$$y_j \sim \text{ Binomial } (\pi_j, N_j)$$
$$\text{logit } (\pi_j) = \theta_j$$
$$\theta_j \sim N(0, 1000)$$

Convergence was tested by analyzing the autocorrelation, trace plot, density, and ergodic average. The autocorrelation between the various values of the parameters generated at different lags for π_1 are shown in Table 2, with three groups and a 25% of correct answers (or successes). According to the information in this table, there is a low association between the values of the parameter generated with the different lags.

Figures 2 and 3 present the trace plots, density, and ergodic average of the chain of values generated with the different lags of the posterior distribution of π_1. As can be seen in these figures, after burning the first 4000 values, the unstable autocorrelations are eliminated, and the posterior distribution of the parameter of interest is found to exhibit a stationary behavior. In addition to the aforementioned graphs, the KPSS test was applied to parameter π_1, yielding a value of 0.29075, with a truncation parameter of 17 and a p-value of 0.1. This indicates that there is not enough sample evidence to reject H_0, which states that the Markov chain has reached the stationary distribution. Similar results were obtained with M2 and M3, considering different parameters, groups, and percentages of successes.

Figures 4 and 5 display the probability intervals of M1, M2, M3, and M4 for the different groups and percentages of correct answers (p = 25%, 50%, and 75%). From these figures, we observe that M4 provides the widest intervals; and M2, the smallest ones. This is explained by the fact that the variability within each group is not

Table 2 Lag autocorrelations for π_1 of the model with $\sigma_\theta^2 \sim$ SBeta2 $(1, 1, 4)$

Lags	Autocorrelation for π_θ with $\sigma_\theta^2 \sim$ SBeta2 $(1, 1, 4)$
0	1.0000000
1	– 0.0095379
5	0.0132617
10	– 0.0003188
50	– 0.0006554

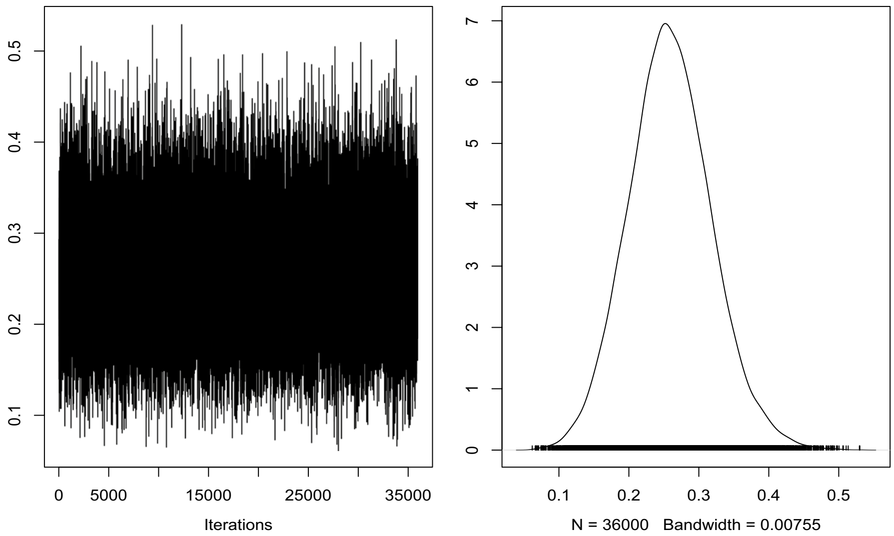

Fig. 2 Trace plot and density function of one of the simulated chains of π_1 *in model 1*

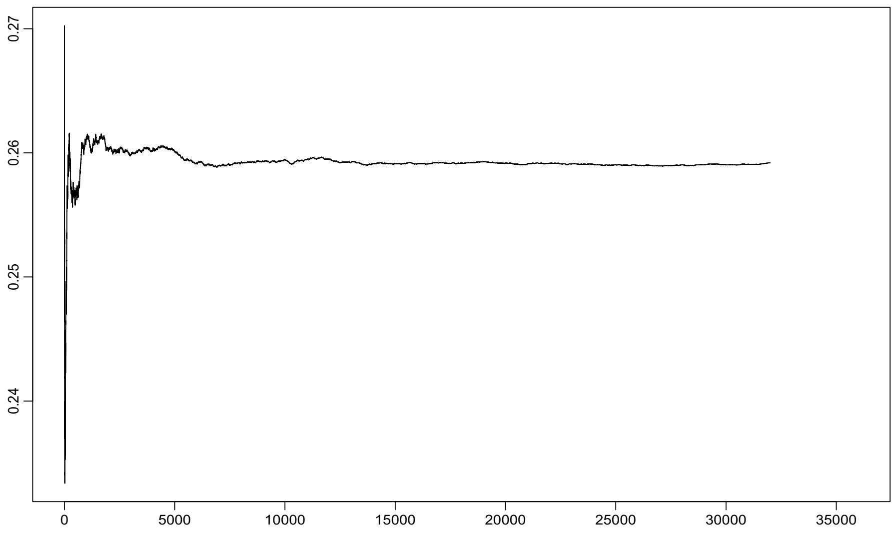

Fig. 3 Ergodic averages of one of the simulated chains of π_1 *in model 1*

considered in the 1-level model. Also, the intervals are found to be more precise as the sample size increases. Similar results were obtained with 10 and 15 groups (N_{10} and N_{15}).

Table 3 provides the results of the DIC, considering the four models (M1, M2, M3, and M4), the different groups, and the percentages of correct answers (p). As this table shows, M2 has a better predictive capacity because its DIC is lower in

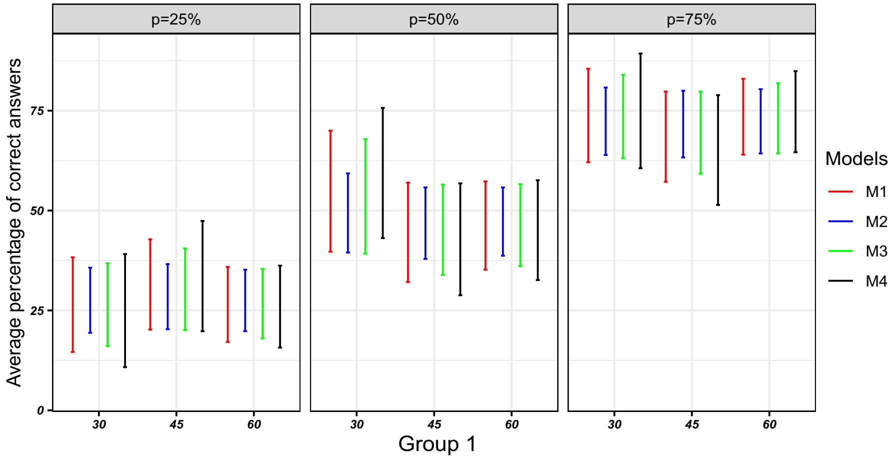

Fig. 4 Average percentage of correct answers in the subsequent results of group 1 in the different models for $p = 25\%$, 50%, and 75%

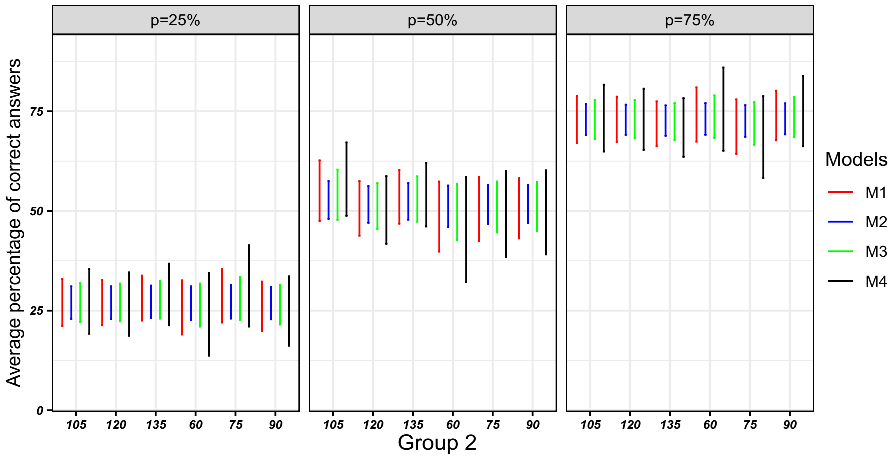

Fig. 5 Average percentage of correct answers in the subsequent results of group 2 in the different models for $p = 25\%$, 50%, and 75%

all scenarios, whereas the 1-level model (M4) exhibits the lowest predictive ability because of its higher DIC. Based on this, the prior distributions used to model the variance parameters in 2-level Bayesian multilevel logistic regression models are found to produce better results in terms of predictive ability because they provide more precise intervals and better DIC values.

Table 3 Deviance Information Criterion (DIC) of the 1-level and 2-level models by group (N_j), considering the percentages of successes (p)

N_j	$p = 25\%$				$p = 50\%$				$p = 75\%$			
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
3	16.34	15.29	15.89	17.98	17.79	17.15	17.63	18.65	16.35	15.47	15.91	17.99
6	34.73	32.58	33.32	40.28	37.92	35.93	36.70	41.66	34.78	32.32	33.30	40.28
10	69.19	68.52	68.62	74.84	67.93	65.37	66.65	77.68	69.21	68.42	68.61	74.81
15	102.0	99.48	101.4	121.2	111.8	109.5	110.9	125.4	102.3	99.42	101.4	121.1

5 Application: modeling the correct answers of university students

Table 4 shows the correct answers of the university students by group. These data were collected from two questionnaires administered to students from two Higher Education Institutions (HEIs) in Medellín (Colombia)—Universidad Nacional de Colombia (UN) (Medellín campus) and Instituto Tecnológico Metropolitano (ITM)—and using a procedure similar to that proposed by Wang et al. (2019). In their study, these authors present a 10-minute trivia game-based activity for the intuitive understanding of confidence intervals. In addition, they explain how the activity can be implemented one or more times in an inferential statistics course.

For data collection, we followed a three-step procedure. First, we formulated 10 general interest questions, the answers to which were provided in the form of an interval (a minimum and maximum value within which respondents considered the answer to each question would fall). Then, students were given 10 minutes to respond. Finally, the students' responses were tabulated.

The questionnaires were administered to statistical inference students from the aforementioned universities (nine groups from UN and three from ITM) at the beginning and end of the first academic semester of 2018. It should be noted that permission was requested from the governing body of each HEI to conduct the experiment, and students were explained that the information provided would be employed for academic

Table 4 Students' correct answers per questionnaire and university

University	Groups	Questionnaire 1	Questionnaire 2	Total
ITM	ITM1	30	29	210
ITM	ITM2	22	37	130
ITM	ITM3	17	8	80
UN	UN1	24	28	150
UN	UN2	102	68	430
UN	UN3	104	129	440
UN	UN4	30	40	160
UN	UN5	35	26	150
UN	UN6	28	20	130
UN	UN7	13	10	70
UN	UN8	45	40	210
UN	UN9	18	24	90

purposes. This experiment could be useful in examining how students understand basic statistical concepts such as confidence intervals and confidence levels in statistical tests.

To analyze the results, we first considered students' correct answers in the two questionnaires administered throughout the academic semester. Then, a 2-level Bayesian multilevel logistic regression model was fitted to model the proportion of correct answers from questionnaires 1 and 2 separately, considering M1, M2, and M3. The purpose of this was to predict students' percentage of correct answers in the two questionnaires, as well as to see whether there were any discrepancies in the forecasts when taking into account other values of the parameters in the prior distributions. A 1-level model (M4) was also considered and compared to the other models in order to identify which has the best predictive ability (Tables 5 and 6).

Subsequently, a single model with an independent variable that considered the two questionnaires and the prior distributions used in the aforementioned models was fitted to model the 2-level. The 1-level model was also fitted. The results of the models' fitting are presented in Table 7.

To model students' performance in terms of the probabilities of correct answers within the groups, M1, M2, M3, and M4 were fitted. If students in group j correctly answers question i, then $y_{ij} = 1$, and if they do not, then $y_{ij} = 0$ for $j = 1, 2, \dots 12$ and $i = 1, 2, \dots N_j$, with N_j number of questions per group.

The model in Eq. (5), with link function (7), was used to simultaneously model both questionnaires. If students in group j correctly answers question i in questionnaire k, then $y_{ikj} = 1$, and if they do not, then $y_{ikj} = 0$, for $j = 1, 2, \dots 12$, and $i = 1, 2, \dots N_{kj}$ and $k = 1, 2$, with N_{kj} number of questions per group, considering questionnaire k. The following are the models proposed here:

- Model 5 (M5). With SBeta2(1,1,4)

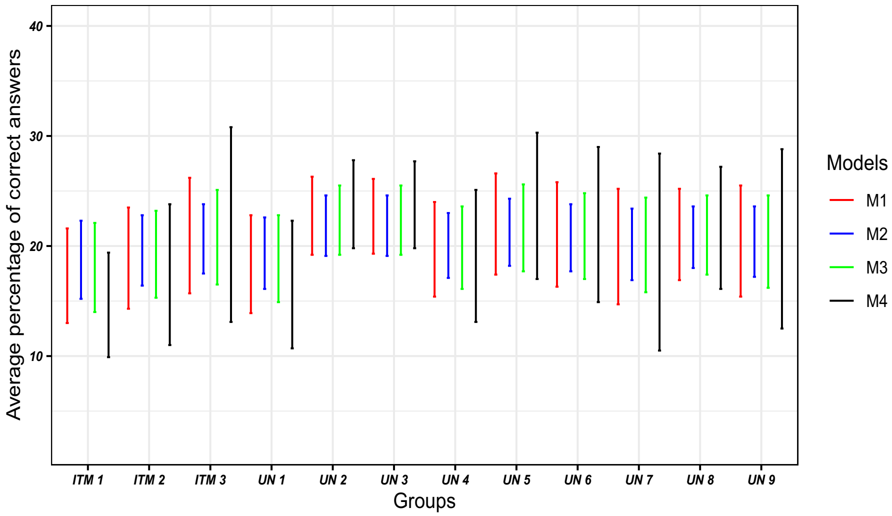

Fig. 6 Average percentage of correct answers per group in the subsequent results of questionnaire 1 for the different models

Table 5 Subsequent summary of the results of questionnaire 1 using a 1-level and 2-level Bayesian multilevel logistic regression models

	M1				M2			
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
μ_θ	– 1.375	0.082	– 1.546	– 1.222	– 1.355	0.061	– 1.480	– 1.240
σ_θ	0.179	0.079	0.049	0.360	0.069	0.061	0.002	0.220
π_θ	0.202	0.013	0.176	0.228	0.205	0.010	0.185	0.224
π_1	0.175	0.022	0.130	0.216	0.196	0.018	0.152	0.223
π_2	0.190	0.023	0.143	0.235	0.201	0.016	0.164	0.228
π_3	0.206	0.026	0.157	0.262	0.206	0.015	0.175	0.238
π_4	0.186	0.023	0.139	0.228	0.200	0.016	0.161	0.226
π_5	0.225	0.018	0.192	0.263	0.214	0.014	0.191	0.246
π_6	0.224	0.018	0.193	0.261	0.213	0.014	0.191	0.246
π_7	0.197	0.022	0.154	0.240	0.203	0.015	0.171	0.230
π_8	0.216	0.023	0.174	0.266	0.209	0.015	0.182	0.243
π_9	0.208	0.023	0.163	0.258	0.207	0.015	0.177	0.238
π_{10}	0.199	0.026	0.147	0.252	0.204	0.016	0.169	0.234
π_{11}	0.209	0.021	0.169	0.252	0.207	0.014	0.180	0.236
π_{12}	0.203	0.025	0.154	0.255	0.205	0.015	0.172	0.236
	M3				M4			
μ_θ	– 1.364	0.071	– 1.514	– 1.234	– 1.414	0.063	– 1.538	– 1.292
σ_θ	0.125	0.071	0.029	0.295	0.294	0.055	0.192	0.408
π_θ	0.204	0.011	0.180	0.226	0.196	0.010	0.177	0.216
π_1	0.185	0.021	0.140	0.221	0.143	0.024	0.099	0.194
π_2	0.196	0.020	0.153	0.232	0.170	0.033	0.110	0.238
π_3	0.206	0.021	0.165	0.251	0.212	0.045	0.131	0.308
π_4	0.193	0.020	0.149	0.228	0.160	0.030	0.107	0.223
π_5	0.220	0.016	0.192	0.255	0.237	0.020	0.198	0.278
π_6	0.220	0.016	0.192	0.255	0.236	0.020	0.198	0.277
π_7	0.200	0.019	0.161	0.236	0.187	0.031	0.131	0.251
π_8	0.213	0.020	0.177	0.256	0.233	0.034	0.170	0.303
π_9	0.207	0.019	0.170	0.248	0.216	0.036	0.149	0.290
π_{10}	0.202	0.021	0.158	0.244	0.186	0.046	0.105	0.284
π_{11}	0.208	0.018	0.174	0.246	0.214	0.028	0.161	0.272
π_{12}	0.204	0.020	0.162	0.246	0.200	0.042	0.125	0.288

$$
\begin{aligned}
&Y_{kj} \sim \text{ Binomial } (\pi_{kj}, N_{kj})\\
&\text{logit } (\pi_{kj}) = a_j + \theta_j I(k=1)\\
&\theta_j \sim N(\mu_\theta, \sigma^2_\theta)\\
&\mu_\theta \sim N(0, 1000)\\
&a_j \sim N(0, 1000)\\
&\sigma^2_\theta \sim \text{ SBeta2 } (1, 1, 4)
\end{aligned}
$$

Table 6 Subsequent summary of the results of questionnaire 2 using a 1-level and 2-level Bayesian multilevel logistic regression

	M1				M2			
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
μ_θ	– 1.440	0.130	– 1.706	– 1.186	– 1.432	0.117	– 1.674	– 1.207
σ_θ	0.375	0.111	0.206	0.635	0.325	0.102	0.170	0.568
π_θ	0.192	0.020	0.154	0.234	0.193	0.018	0.158	0.230
π_1	0.150	0.022	0.109	0.196	0.153	0.022	0.112	0.199
π_2	0.261	0.035	0.196	0.333	0.255	0.034	0.194	0.326
π_3	0.140	0.032	0.081	0.206	0.146	0.031	0.087	0.209
π_4	0.189	0.028	0.137	0.246	0.189	0.027	0.140	0.245
π_5	0.162	0.017	0.130	0.196	0.164	0.017	0.133	0.198
π_6	0.283	0.021	0.242	0.325	0.280	0.021	0.239	0.322
π_7	0.237	0.030	0.181	0.299	0.233	0.029	0.180	0.295
π_8	0.179	0.027	0.129	0.235	0.181	0.026	0.131	0.235
π_9	0.166	0.028	0.114	0.224	0.169	0.027	0.118	0.224
π_{10}	0.166	0.034	0.103	0.237	0.170	0.033	0.108	0.238
π_{11}	0.191	0.025	0.145	0.242	0.191	0.024	0.147	0.241
π_{12}	0.241	0.038	0.173	0.322	0.237	0.037	0.172	0.316
	M3				M4			
μ_θ	– 1.436	0.125	– 1.693	– 1.196	– 1.483	0.067	– 1.618	– 1.354
σ_θ	0.359	0.114	0.188	0.629	0.477	0.078	0.338	0.646
π_θ	0.193	0.019	0.155	0.232	0.185	0.010	0.166	0.205
π_1	0.151	0.022	0.110	0.197	0.138	0.024	0.095	0.188
π_2	0.258	0.035	0.195	0.331	0.285	0.039	0.212	0.365
π_3	0.142	0.031	0.083	0.206	0.100	0.033	0.045	0.174
π_4	0.189	0.027	0.138	0.246	0.186	0.032	0.129	0.252
π_5	0.163	0.017	0.131	0.197	0.158	0.018	0.126	0.194
π_6	0.282	0.021	0.242	0.325	0.293	0.022	0.252	0.336
π_7	0.236	0.030	0.181	0.298	0.250	0.034	0.186	0.320
π_8	0.179	0.027	0.129	0.234	0.173	0.030	0.118	0.236
π_9	0.167	0.028	0.116	0.223	0.154	0.032	0.097	0.221
π_{10}	0.167	0.034	0.104	0.237	0.143	0.042	0.073	0.234
π_{11}	0.191	0.024	0.146	0.241	0.191	0.027	0.140	0.246
π_{12}	0.240	0.038	0.173	0.320	0.267	0.047	0.180	0.363

- Model 6 (M6). SBeta2(1,1,4) in model 5 is replaced with SBeta2(1/2,1/2,625)

$$\sigma_\theta^2 \sim \text{ SBeta2 } (1/2, 1/2, 625).$$

- Model 7 (M7). SBeta2(1,1,4) in model 5 is replaced with IG(0.001,0.001)

$$\sigma_\theta^2 \sim \text{ IG } (0.001, 0.001)$$

.

- Model 8 (M8). 1-level model

$$Y_{kj} \sim \text{ Binomial } (\pi_{kj}, N_{kj})$$
$$\text{logit } (\pi_{kj}) = a_j + \theta_j I(k = 1)$$
$$\theta_j \sim N(0, 1000)$$
$$a_j \sim N(0, 1000)$$

Tables 5 and 6 show the subsequent results of the models' fitting for questionnaires 1 and 2, respectively; and Table 7 provides the results for both questionnaires in the same model. Tables 5 and 6 present the average (mean), the Standard Deviation (SD), and the probability interval (2.5%, 97.5%) per group, while Table 7 reports similar results but without the standard deviation. As can be seen from the tables above, σ_θ is lower in M2 and M6 and higher in M4 and M7. This is explained by the fact that, although this variability within the groups exists, it is not modeled in the latter models.

As observed in Table 5, UN2 and UN3 (i.e., π_5 and π_6, respectively) are the groups with the highest probability of correct answers in the different models,

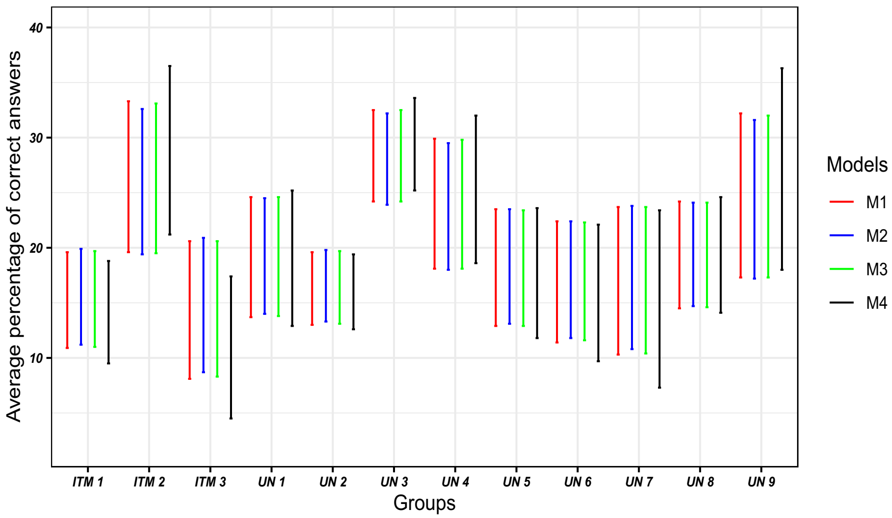

Fig. 7 Average percentage of correct answers per group in the subsequent results of questionnaire 2 for the different models

Table 7 Subsequent summary of the results of questionnaires 1 and 2 using 1-level and 2-level Bayesian multilevel logistic regression models with an explanatory variable

	M1			M2			M3			M4		
	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%	Mean	2.5%	97.5%
μ_θ	0.038	– 0.23	0.32	0.032	– 0.19	0.27	0.036	– 0.22	0.30	0.068	– 0.11	0.25
σ_θ	0.356	0.16	0.65	0.243	0.01	0.53	0.322	0.10	0.64	0.552	0.37	0.77
OR_t	1.049	0.79	1.37	1.040	0.83	1.30	1.046	0.80	1.36	0.517	0.47	0.56
$\pi_{1.1}$	0.143	0.10	0.19	0.142	0.11	0.19	0.143	0.10	0.19	0.143	0.10	0.19
$\pi_{1.2}$	0.197	0.14	0.26	0.207	0.15	0.27	0.200	0.14	0.26	0.169	0.11	0.24
$\pi_{1.3}$	0.179	0.11	0.26	0.171	0.11	0.25	0.177	0.11	0.26	0.213	0.13	0.31
$\pi_{1.4}$	0.167	0.12	0.22	0.170	0.12	0.22	0.168	0.12	0.22	0.160	0.11	0.22
$\pi_{1.5}$	0.229	0.19	0.27	0.221	0.18	0.26	0.227	0.19	0.27	0.237	0.20	0.28
$\pi_{1.6}$	0.243	0.21	0.28	0.248	0.21	0.29	0.244	0.21	0.28	0.236	0.20	0.28
$\pi_{1.7}$	0.202	0.15	0.26	0.207	0.15	0.26	0.204	0.15	0.26	0.187	0.13	0.25
$\pi_{1.8}$	0.222	0.17	0.29	0.216	0.16	0.28	0.220	0.16	0.28	0.233	0.17	0.30
$\pi_{1.9}$	0.202	0.14	0.27	0.197	0.14	0.26	0.200	0.14	0.27	0.216	0.15	0.29
$\pi_{1.10}$	0.174	0.11	0.26	0.171	0.11	0.25	0.173	0.11	0.25	0.185	0.11	0.28
$\pi_{1.11}$	0.211	0.16	0.26	0.209	0.16	0.26	0.210	0.16	0.26	0.214	0.16	0.27
$\pi_{1.12}$	0.219	0.15	0.30	0.225	0.16	0.30	0.221	0.15	0.30	0.200	0.12	0.29
$\pi_{2.1}$	0.138	0.10	0.18	0.139	0.10	0.18	0.138	0.10	0.18	0.138	0.09	0.19
$\pi_{2.2}$	0.257	0.19	0.33	0.247	0.19	0.32	0.254	0.19	0.33	0.284	0.21	0.36
$\pi_{2.3}$	0.133	0.08	0.20	0.141	0.08	0.21	0.136	0.08	0.20	0.100	0.05	0.18
$\pi_{2.4}$	0.179	0.13	0.24	0.177	0.13	0.23	0.179	0.13	0.24	0.186	0.13	0.26
$\pi_{2.5}$	0.167	0.13	0.20	0.174	0.14	0.21	0.169	0.14	0.21	0.158	0.13	0.19
$\pi_{2.6}$	0.287	0.25	0.33	0.281	0.24	0.32	0.286	0.25	0.33	0.293	0.25	0.34
$\pi_{2.7}$	0.236	0.18	0.30	0.230	0.18	0.29	0.234	0.18	0.30	0.250	0.17	0.32
$\pi_{2.8}$	0.185	0.13	0.25	0.190	0.14	0.25	0.187	0.13	0.25	0.173	0.12	0.24
$\pi_{2.9}$	0.167	0.11	0.23	0.173	0.12	0.23	0.169	0.12	0.23	0.154	0.10	0.22
$\pi_{2.10}$	0.155	0.09	0.23	0.158	0.10	0.23	0.156	0.09	0.23	0.143	0.07	0.23
$\pi_{2.11}$	0.194	0.15	0.25	0.196	0.15	0.24	0.194	0.15	0.24	0.191	0.14	0.25
$\pi_{2.12}$	0.248	0.17	0.33	0.242	0.17	0.32	0.245	0.17	0.33	0.267	0.18	0.36
OR_1	1.067	0.68	1.62	1.052	0.71	1.52	1.062	0.68	1.59	1.084	0.59	1.81
OR_2	0.725	0.41	1.12	0.818	0.47	1.16	0.753	0.42	1.13	0.533	0.28	0.92
OR_3	1.501	0.82	2.80	1.312	0.83	2.37	1.445	0.82	2.70	2.812	1.01	6.65
OR_4	0.944	0.57	1.44	0.970	0.62	1.39	0.951	0.58	1.43	0.868	0.45	1.51
OR_5	1.506	1.08	2.07	1.380	0.96	1.96	1.470	1.04	2.04	1.686	1.18	2.34
OR_6	0.803	0.59	1.05	0.851	0.62	1.10	0.816	0.60	1.07	0.753	0.55	1.01
OR_7	0.837	0.51	1.24	0.893	0.56	1.24	0.855	0.53	1.25	0.716	0.40	1.18
OR_8	1.290	0.81	2.03	1.201	0.82	1.85	1.261	0.81	1.97	1.519	0.83	2.57
OR_9	1.306	0.79	2.12	1.207	0.81	1.92	1.271	0.80	2.05	1.601	0.81	2.90
OR_{10}	1.198	0.66	2.08	1.132	0.70	1.86	1.177	0.67	2.04	1.530	0.56	3.41
OR_{11}	1.136	0.75	1.66	1.104	0.78	1.58	1.127	0.76	1.64	1.194	0.71	1.89
OR_{12}	0.881	0.49	1.40	0.930	0.55	1.34	0.898	0.51	1.38	0.729	0.33	1.37

while ITM1 (i.e., π_1) is the one with the lowest probability. Regarding the intervals, M4 is found to have the lowest precision, whereas M2 showed the highest precision. Also, from Fig. 6, we can see that M4 produces the widest intervals, while M2 provides the smallest intervals. This is may be due to the fact that variability within each group is not considered in M4.

After the models were compared based on the resulting DIC values, M3 was found to be the model with the best ability to predict the percentage of correct answers because it had the lowest DIC (76.81), while M4 exhibited the lowest predictive ability, with a DIC of 84.81. It should be noted that M1 and M2 also yielded good results because their DIC was 77.01 and 77.21, respectively.

Table 6 presents the posterior results for questionnaire 2. As can be seen, UN3 (i.e., π_6) is the group with the highest probability of correct answers in the different models, while ITM3 (i.e., π_1) is the one with the lowest probability. As for the intervals shown in this table, we observe a pattern (illustrated in Fig. 7) similar to that in Table 5: M4 produces the widest intervals, whereas M2 provides the smallest intervals. The resulting DIC values of M1, M2, M3, and M4 are similar to those calculated with the results of questionnaire 1 (81.78, 82.12, 81.93, and 84.14, respectively). However, in this case, M1 exhibits the best ability to predict the percentages of correct answers.

Finally, Table 7 summarizes the results of M5, M6, M7, and M8 in terms of the probabilities of correct answers, which are similar to those presented in Tables 5 and 6. From this table, it can be seen that UN3, ITM1, and ITM3 (i.e., π_6, π_1, and π_3, respectively) are the groups with the highest and lowest probability of correct answers. Also, when both questionnaires were compared using the odds ratio, we observed that students improved their probability of success in the ITM2, UN1, UN3, UN4, and UN9 groups (i.e., OR_2, OR_4, OR_6, OR_7, and OR_{12}, respectively), with ITM2 showing the greatest improvement probability. In the rest of the groups, students were not found to improve such probability.

The behavior of the probability intervals is similar to that observed in Figs. 6 and 7: The model with the least precise probability intervals is M8 and that with the most precise intervals is M6. The DIC values of M5, M6, M7, and M8 were 165.1, 167.5, 165.7, and 169.2, respectively. According to this, M5 is the model with the best predictive ability.

6 Concluding remarks

Bayesian multilevel models are commonly used to represent data with levels or hierarchies, particularly in cases involving students within universities. Since there is often no information about the variance parameter, it is usually modeled using noninformative distributions. In this paper, we employed two versions of the SBeta2 distribution and the inverse-gamma distribution to model such parameter.

According to the results, the probability intervals are wider in the 1-level model than in the 2-level models. This is explained by the fact that, although variability within the groups exists, it is not considered in the former. In addition, the 1-level model can provide erroneous results when comparing the groups. For instance, as observed in Fig. 6, when considering variability within each group, the percentages of correct answers in UN2 and UN3 differ from those in ITM1 when there is, in fact, no difference. Likewise, the 2-level models were found to have a greater predictive ability than the 1-level model because they had a lower DIC value.

When the probability of success remains the same, 2-level Bayesian multilevel logistic regression models exhibit a better predictive ability when SBeta2 distributions, rather than a 1-level Bayesian multilevel logistic regression model or an inverse-gamma distribution, are used to model the variance parameter. Based on the estimated DIC values, similar results in terms of the improved predictive ability of the 2-level Bayesian multilevel logistic regression model with respect to the 1-level Bayesian multilevel logistic regression model are obtained when the probabilities of success and sample sizes vary.

Finally, the probability intervals of 2-level Bayesian multilevel logistic regression models proved to be more precise when SBeta2 distributions, rather than 1-level Bayesian multilevel logistic regression model or an inverse-gamma distribution, are employed to model the random effect.

Funding Open Access funding provided by Colombia Consortium.

References

Ayalew MM (2020) Bayesian hierarchical analyses for entrepreneurial intention of students. J Big Data 7:711–23

Aychiluhm SB, Gelaye KA, Angaw DA, Dagne GA, Tadesse AW, Abera A, Dillu D (2020) Determinants of malaria among under-five children in Ethiopia: Bayesian multilevel analysis. BMC Public Health 20:10–2011

Berger J (2006) The case for objective Bayesian analysis. Bayesian Anal 1(3):385–402

Bernardo J, Smith A (2000) Bayesian theory. Wiley, New York

Birlutiu A, Groot P, Heskes T (2010) Multi-task preference learning with an application to hearing aid personalization. Neurocomputing 73(7–9):1177–1185

Bornmann L, Stefaner M, de Moya Anegón F, Mutz R (2016) Excellence networks in science: A web-based application based on Bayesian multilevel logistic regression (bmlr) for the identification of institutions collaborating successfully. J Informet 10(1):312–327

Brooks S, Roberts G (1998) Assessing convergence of Markov chain Monte Carlo algorithms. Stat Comput 8(4):319–335

Cowles M, Carlin B (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. J Am Stat Assoc 91(434):883–904

De la Cruz R, Meza C, Arribas-Gil A, Carroll R (2016) Bayesian regression analysis of data with random effects covariates from nonlinear longitudinal measurements. J Multivar Anal 143:94–106
Fagbamigbe AF, Uthman AO, Ibisomi L (2021) Hierarchical disentanglement of contextual from compositional risk factors of diarrhoea among under-five children in low-and middle-income countries. Sci Rep 11(1):1–17
Gañan-Cardenas E, Jiménez JC, Pemberthy-R JI (2021) Bayesian hierarchical modeling of operating room times for surgeries with few or no historic data. J Clin Monit Comput 36:1–16
Gaviria J, Morera M (2005) Modelos jerárquicos lineales. Editorial La Muralla
Gelman A (2006) Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). Bayesian Anal 1:13515–534
Gelman A, Hill J (2006) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge
Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D (2013) Bayesian data analysis, 3rd edn. Chapman and Hall/CRC, Reading
Grogan-Kaylor A, Castillo B, Pace GT, Ward KP, Ma J, Lee SJ, Knauer H (2021) Global perspectives on physical and nonphysical discipline: a Bayesian multilevel analysis. Int J Behav Dev 45(3):216–225
Jabessa S, Jabessa D (2021) Bayesian multilevel model on maternal mortality in Ethiopia. J Big Data 8(1):1–17
Jara A, Quintana F, San Martín E (2008) Linear mixed models with skew-elliptical distributions: a Bayesian approach. Comput Stat Data Anal 52(11):5033–5045
King G, Zeng L (2001) Logistic regression in rare events data. Polit Anal 9(2):137–163
Kwiatkowski D, Phillips P, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? J Econom 54(1–3):159–178
Llera A, Beckmann C (2016) Estimating an inverse gamma distribution. arXiv:1605.01019
Lu Z-H, Khondker Z, Ibrahim JG, Wang Y, Zhu H, Initiative ADN (2017) Bayesian longitudinal low-rank regression models for imaging genetic data from longitudinal studies. Neuroimage 149:305–322
McElreath R (2015) Statistical rethinking: a Bayesian course with examples in r and stan. Chapman and Hall/CRC, New York
Młynarczyk D, Armero C, Gómez-Rubio V, Puig P (2021) Bayesian analysis of population health data. Mathematics 9(5):577
Ntzoufras I (2011) Bayesian modeling using winbugs, vol 698. Wiley, New York
Peng C-Y, Lee K, Ingersoll G (2002) An introduction to logistic regression analysis and reporting. J Educ Res 96(1):3–14
Pérez M-E, Pericchi L, Ramírez I (2017) The scaled beta2 distribution as a robust prior for scales. Bayesian Anal 12(3):615–637
Pinheiro J, Bates D (2006) Mixed-effects models in S and S-PLUS mixed-effects models in s and s-plus. Springer, Berlin
Pregibon D (1981) Logistic regression diagnostics logistic regression diagnostics. Ann Stat 9(4):705–724
R Core Team (2019) R: a language and environment for statistical computing [Computer software manual]. Vienna, Austria. https://www.R-project.org/
Rojas J, Ramírez I (2019) Ajuste de un modelo jerárquico desde un enfoque bayesiano (Unpublished master's thesis). Universidad Nacional de Colombia-Sede Medellín
Sherwood RJ, Oh HS, Valiathan M, McNulty KP, Duren DL, Knigge RP, Middleton KM (2021) Bayesian approach to longitudinal craniofacial growth: the craniofacial growth consortium study. Anat Rec 304(5):991–1019
Spiegelhalter D, Best N, Carlin B, Van Der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc Ser B (Stat Methodol) 64(4):583–639
Sturtz S, Ligges U, Gelman A (2010) R2openbugs: a package for running openbugs from r. http://cran.rproject.org/web/packages/R2OpenBUGS/vignettes/R2OpenBUGS. pdf)
Tang N-S, Duan X-D (2014) Bayesian influence analysis of generalized partial linear mixed models for longitudinal data. J Multivar Anal 126:86–99
Trapletti A, Hornik K, LeBaron B, Hornik M (2019) Package 'tseries' Package 'tseries'
Wang X, Reich N, Horton N (2019) Enriching students' conceptual understanding of confidence intervals: an interactive trivia-based classroom activity. Am Stat 73(1):50–55
Wong GY, Mason WM (1985) The hierarchical logistic regression model for multilevel analysis. J Am Stat Assoc 80(391):513–524

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Cristian David Correa-Álvarez[1] · Juan Carlos Salazar-Uribe[2] · Luis Raúl Pericchi-Guerra[3]

Juan Carlos Salazar-Uribe
jcsalaza@unal.edu.co

Luis Raúl Pericchi-Guerra
luis.pericchi@upr.edu

1 Department of Quality and Production, Instituto Tecnológico Metropolitano (ITM), Calle 73 No. 76A - 354, Medellín, Antioquia 050034, Colombia

2 School of Statistics, Universidad Nacional de Colombia (Medellín campus), Carrera 65 Nro. 59A - 110, Medellín, Antioquia 050034, Colombia

3 Department of Mathematics and Center for Biostatistics and Bioinformatics, University of Puerto Rico, 14, 2534 Av. Universidad Ste, Rio Piedras 00925, San Juan, Puerto Rico