

# Bayesian and Frequentist Multilevel Modeling

Andy Grogan-Kaylor

2025-09-25

## Table of contents

1	Introduction	2
2	Conceptual Appropriateness	2
3	Accepting the Null Hypothesis ( $H_0$ )	3
4	Model Comparison	3
5	Prior Information	3
6	Multiple Comparisons	3
7	Smaller Samples	4
8	Full Distribution of Parameters	4
9	Distributional Models	5
10	Non-Linear Terms	6
11	Maximal Models	6
	References	7

# 1 Introduction

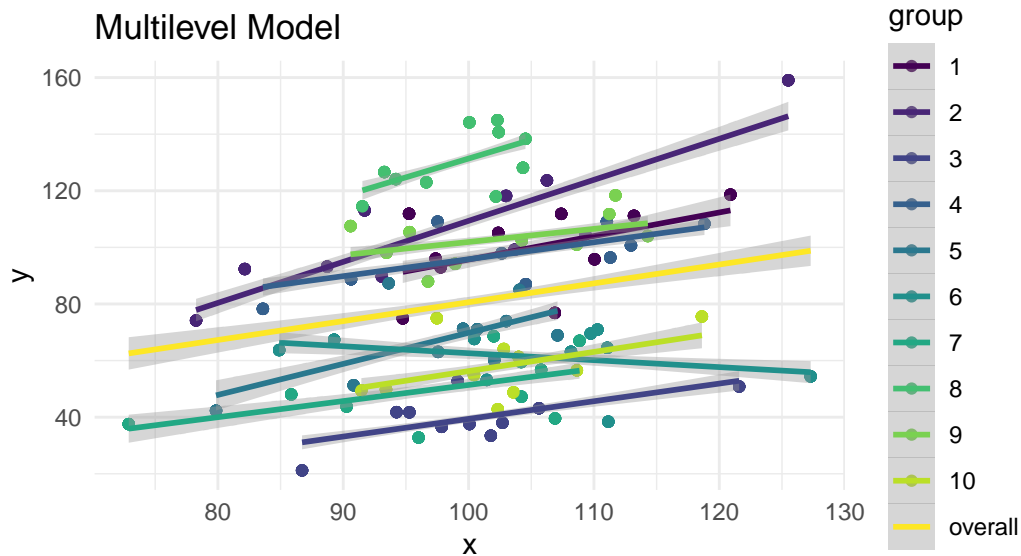


Figure 1: Simulated Multilevel Model

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + u_{0j} + u_{1j}x + e_{ij}$$

All multilevel models account for group structure, in estimating the association of  $x$  and  $y$ , by including a random intercept ( $u_0$ ), and possibly one or more random slope terms ( $u_1, u_2, etc....$ ).

Bayesian models may offer some advantages over frequentist models, but may be *substantially* slower to converge.

## 2 Conceptual Appropriateness

Following Kruschke (2014) all Bayesian models have a *conceptual appropriateness*.

In frequentist reasoning we are estimating the probability of observing data at least as extreme as our data, while assuming a null hypothesis ( $H_0$ ). As has been noted,  $H_0$ , e.g.  $\beta = 0$ , or  $\bar{x}_A - \bar{x}_B = 0$ , is very often not a substantively interesting or substantively meaningful hypothesis.

In Bayesian analysis, we are not rejecting a null hypothesis. Instead, we are *directly estimating the value of a parameter* such as  $\beta$  and are indeed estimating a *full probability distribution* for this parameter.

### 3 Accepting the Null Hypothesis ( $H_0$ )

The Bayesian approach means that we have the ability to accept the null hypothesis  $H_0$  (Kruschke & Liddell, 2018). This ability to accept  $H_0$  might possibly lead to theory simplification (Gallistel, 2009; Morey et al., 2018), as well as to a lower likelihood of the publication bias that results from frequentist methods predicated upon the rejection of  $H_0$  (Kruschke & Liddell, 2018).

### 4 Model Comparison

Relatedly, Bayesian approaches allow one to compare an alternative model  $H_A$  with a null model  $H_0$ , or to simply compare two alternative statistical models ( $H_1$  vs.  $H_2$ ). Bayesian models may have a better perspective on these kinds of statistical comparisons than do frequentist approaches. As Jarosz et al. (2014) note:

“All Bayesian approaches are comparisons of models. This means that a Bayes factor considers the likelihood of both the null and the alternative hypothesis. From the researcher’s standpoint, this is likely closer to their overall goal than simply rejecting the null hypothesis.”

### 5 Prior Information

Bayesian models allow one to incorporate prior information about a parameter of interest.

Prior information may come from the prior research literature, e.g. from systematic reviews or meta-analyses, or expert opinion or clinical wisdom.

Kruschke (2014) points out that “No analysis is immune to false alarms, because randomly sampled data will occasionally contain accidental coincidences of outlying values.” However, according to Kruschke (2014) careful use of priors may reduce the probability of false alarms.

### 6 Multiple Comparisons

As Kruschke (2014) observes, multiple comparisons (especially when they are *post hoc*) are less of a concern for Bayesian analysis than they are for a frequentist analysis:

“In a Bayesian analysis, however, there is just one posterior distribution over the parameters that describe the conditions. That posterior distribution is unaffected by the intentions of the experimenter, and the posterior distribution can be examined from multiple perspectives however is suggested by insight and curiosity.”

## 7 Smaller Samples

Bayesian multilevel models may be better with small samples (Rognli et al., 2022; van de Schoot et al., 2015), especially samples with small numbers of Level 2 units (Hox et al., 2012). Some of this advantage may occur when parameters are not normally distributed (Muthen & Asparouhov, 2012). It appears that much of this improvement in performance is dependent upon the use of informative priors (Smid et al., 2020), and that Bayesian models with small samples may provide worse estimates than maximum likelihood estimates when less informative, or inaccurate, priors are used (McNeish, 2016).

## 8 Full Distribution of Parameters

Bayesian models of all kinds provide full distributions of the parameters (e.g.  $\beta$ 's and random effects ( $u$ 's))—both singly and jointly—rather than only point estimates.

Information about the full distribution of a parameter, such as the estimate of the probability distribution of values of a risk factor, a protective factor, or the effect of an intervention, may be substantively meaningful (Rindskopf, 2020). Such information may be especially important when the distribution of a regression parameter is non-normal (Finch & Bolin, 2017; van de Schoot et al., 2014) e.g. in smaller samples.

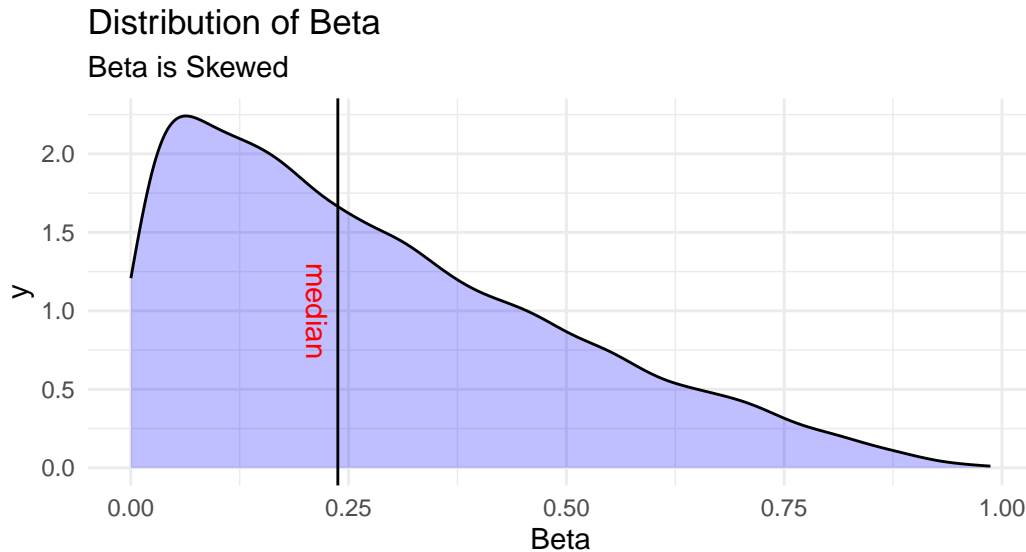


Figure 2: Distribution of a Single Parameter

## Joint Distribution of Random Effects

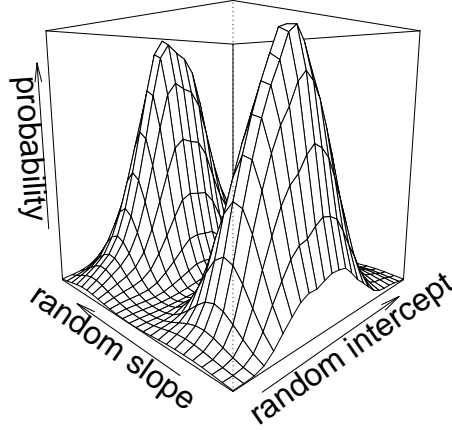


Figure 3: Joint Distribution of Parameters

As Stata Corporation notes, “In a Bayesian multilevel model, *random effects* are model parameters just like regression coefficients and variance components” (StataCorp, 2020). This ability to estimate the *distributions* of these random effects means that the *distribution* of the random effect for one group can be compared to another. For example, the *distribution* of a parameter in one country could be directly compared to the *distribution* of that same parameter in another country. One could even estimate the probability that a particular  $\beta$  had a higher value in one group (e.g. country), than in another. As Balov (2016) suggests, this Bayesian approach allows us to “quantify the credibility” of these comparisons, which would not be possible with a frequentist approach. Rognli et al. (2022) make a similar point about estimating the probability of a practically or substantively meaningful effect size.

As an example, Stunnenberg et al. (2018) conducted a Bayesian analysis where the results of the multilevel analysis were used to inform treatment decisions. Here the data were repeated measures on patients, and thus the patients were the groups: “On completion of each treatment set, a Bayesian analysis was conducted to calculate the posterior probability of mexiletine [treatment] producing a clinically meaningful difference in the individual patient.”

## 9 Distributional Models

Bayesian estimators allow one to directly model  $\sigma_{u_0}$ , the variance of the Level 2 units, as a function of covariates (Burkner, 2018). This *potentially* allows for the opportunity for this variation to become an outcome parameter of substantive interest.

## 10 Non-Linear Terms

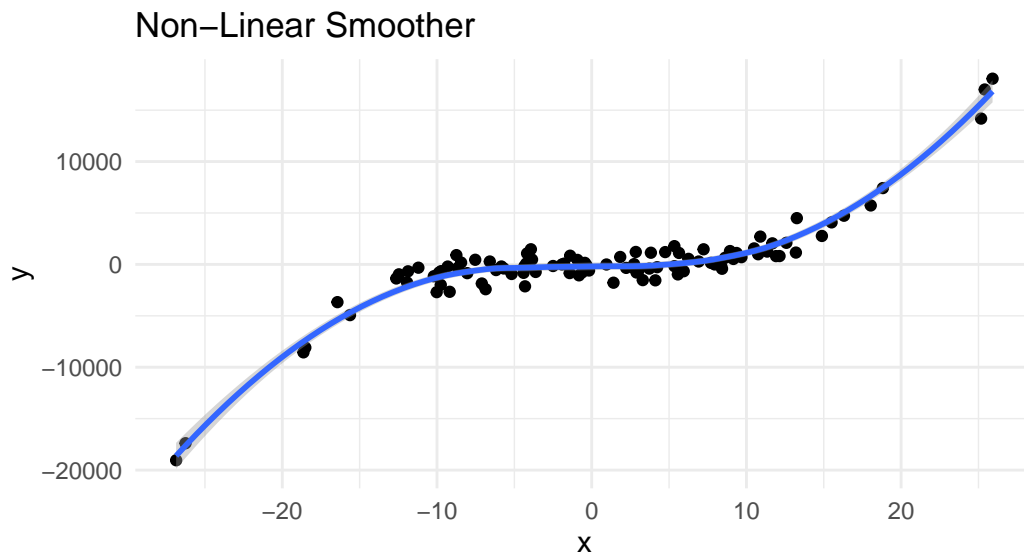


Figure 4: Non-Linear Terms

Bayesian estimators allow for the incorporation of non-linear terms (Burkner, 2018). Such non-linear terms offer ways of non-parametrically fitting curvature. An open question is whether such methods represent a kind of over-fitting. A related question is the degree to which non-linear terms provide substantively interpretable results.

## 11 Maximal Models

Hypothetically, one might imagine that there could be group level unobserved factors which affect many regression slopes—i.e. the relationship between multiple predictors (e.g.  $x_1$ ,  $x_2$ ,  $x_3$ , etc.) and outcome variable  $y$ . Arguably, were one to ignore these unobserved factors in statistical estimation, they would show up either in an error term ( $e_i$  or  $u_0$ ), or in the regression coefficients ( $\beta$ ). Were they to show up in the regression coefficients this would represent statistical bias and a substantive mis-estimation of important effects. Thus, there is a conceptual argument for including as many random effects—i.e. random slopes—in a statistical model as possible.

Bayesian estimators more readily allow for the estimation of so called *maximal models* (Barr et al., 2013; Frank, 2018), which allow for the inclusion of a large number of random slopes, e.g.  $u_1, u_2, u_3, \dots$ , etc. even when some of those estimated slopes are close to 0.

It should be noted that Matuschek et al. (2017) argue that such a *maximal* approach may lead to a loss of statistical power and further argue that one should adhere to “a random effect structure that is supported by the data.”

In contrast, Nalborczyk et al. (2019) argue that maximal models are supported under the Bayesian approach. Oberauer (2022) also argues for including multiple random slopes. Schielzeth & Forstmeier (2009) make a similar argument from a frequentist perspective.

## References

- Balov, N. (2016). Bayesian hierarchical models in stata. *2016 Stata Conference*. <https://ideas.repec.org/p/boc/scon16/30.html>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Burkner, P.-C. (2018). Advanced bayesian multilevel modeling with the r package brms. *The R Journal*, 10(1), 395–411.
- Finch, W. H., & Bolin, J. E. (2017). *Multilevel modeling using mplus*. <https://doi.org/10.1201/9781315165882>
- Frank, M. (2018). *Mixed effects models: Is it time to go bayesian by default?* <http://babieslearninglanguage.blogspot.com/2018/02/mixed-effects-models-is-it-time-to-go.html>
- Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review*, 116(2), 439–453. <https://doi.org/10.1037/a0015251>
- Hox, J., Hox, J. J. C. M., van de Schoot, R., & Matthijsse, S. (2012). How few countries will do? Comparative survey analysis from a bayesian perspective. *Survey Research Methods*. <https://doi.org/10.18148/srm/2012.v6i2.5033>
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *Journal of Problem Solving*. <https://doi.org/10.7771/1932-6246.1167>
- Kruschke, J. K. (2014). *Doing bayesian data analysis: A tutorial with r, JAGS, and stan, second edition* (pp. 1–759). Elsevier Science. <https://doi.org/10.1016/B978-0-12-405888-0.09999-2>
- Kruschke, J. K., & Liddell, T. M. (2018). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of Memory and Language*. <https://doi.org/10.1016/j.jml.2017.01.001>
- McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5), 750–773. <https://doi.org/10.1080/10705511.2016.1186549>

- Morey, R. D., Homer, S., & Proulx, T. (2018). Beyond statistics: Accepting the null hypothesis in mature sciences. *Advances in Methods and Practices in Psychological Science*. <https://doi.org/10.1177/2515245918776023>
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. In *Psychological Methods* (No. 3; Vol. 17, pp. 313–335). American Psychological Association. <https://doi.org/10.1037/a0026802>
- Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & Burkner, P.-C. (2019). *An introduction to bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard indonesian*. PsyArXiv. [https://doi.org/10.1044/2018\\_JSLHR-S-18-0006](https://doi.org/10.1044/2018_JSLHR-S-18-0006)
- Oberauer, K. (2022). The Importance of Random Slopes in Mixed Models for Bayesian Hypothesis Testing. *Psychological Science*. <https://doi.org/doi:10.1177/09567976211046884>
- Rindskopf, D. (2020). Reporting Bayesian Results. *Evaluation Review*. <https://doi.org/10.1177/0193841X20977619>
- Rognli, E. W., Zahl-Olsen, R., Rekdal, S. S., Hoffart, A., & Bertelsen, T. B. (2022). Editorial perspective: Bayesian statistical methods are useful for researchers in child and adolescent mental health. *Journal of Child Psychology and Psychiatry*, n/a(n/a). <https://doi.org/https://doi.org/10.1111/jcpp.13662>
- Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, 20, 416–420. <https://doi.org/10.1093/beheco/arn145>
- Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian Versus Frequentist Estimation for Structural Equation Models in Small Sample Contexts: A Systematic Review. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 131–161. <https://doi.org/10.1080/10705511.2019.1577140>
- StataCorp. (2020). *Bayesian multilevel models*. <https://www.stata.com/features/overview/bayesian-multilevel-models/>
- Stunnenberg, B. C., Raaphorst, J., Groenewoud, H. M., Statland, J. M., Griggs, R. C., Wouterman, W., Stegeman, D. F., Timmermans, J., Trivedi, J., Matthews, E., Saris, C. G. J., Schouwenberg, B. J., Drost, G., Van Engelen, B. G. M., & Van Der Wilt, G. J. (2018). Effect of mexiletine on muscle stiffness in patients with nondystrophic myotonia evaluated using aggregated n-of-1 trials. *JAMA - Journal of the American Medical Association*. <https://doi.org/10.1001/jama.2018.18020>
- van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology*, 6(1), 25216. <https://doi.org/10.3402/ejpt.v6.25216>
- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to bayesian analysis: Applications to developmental research. *Child Development*, 85(3), 842–860. <https://doi.org/10.1111/cdev.12169>