Bayesian Workflow for Model Building in Educational Research

M. Shane Tutwiler, EdD

Abstract

Educational data often contain a large degree of measurement error, missingness, and complexity due to the clustering of students. Traditional statistical inference via frequentist models and null hypothesis significance testing, while computationally efficient and relatively easy to interpret, fail to address these issues directly and often mask the degree of uncertainty in findings by focusing on point estimates and p-values. In this poster, we demonstrate a proposed workflow for modeling and communicating relationships in educational data via Bayesian multilevel modeling. Benefits include the ability to encode domain knowledge via informative priors, using said priors to regularize estimates to prevent the over-fitting of models to data, inherent mechanisms to impute missing data, and the capacity to handle measurement error in the outcome and predictors.

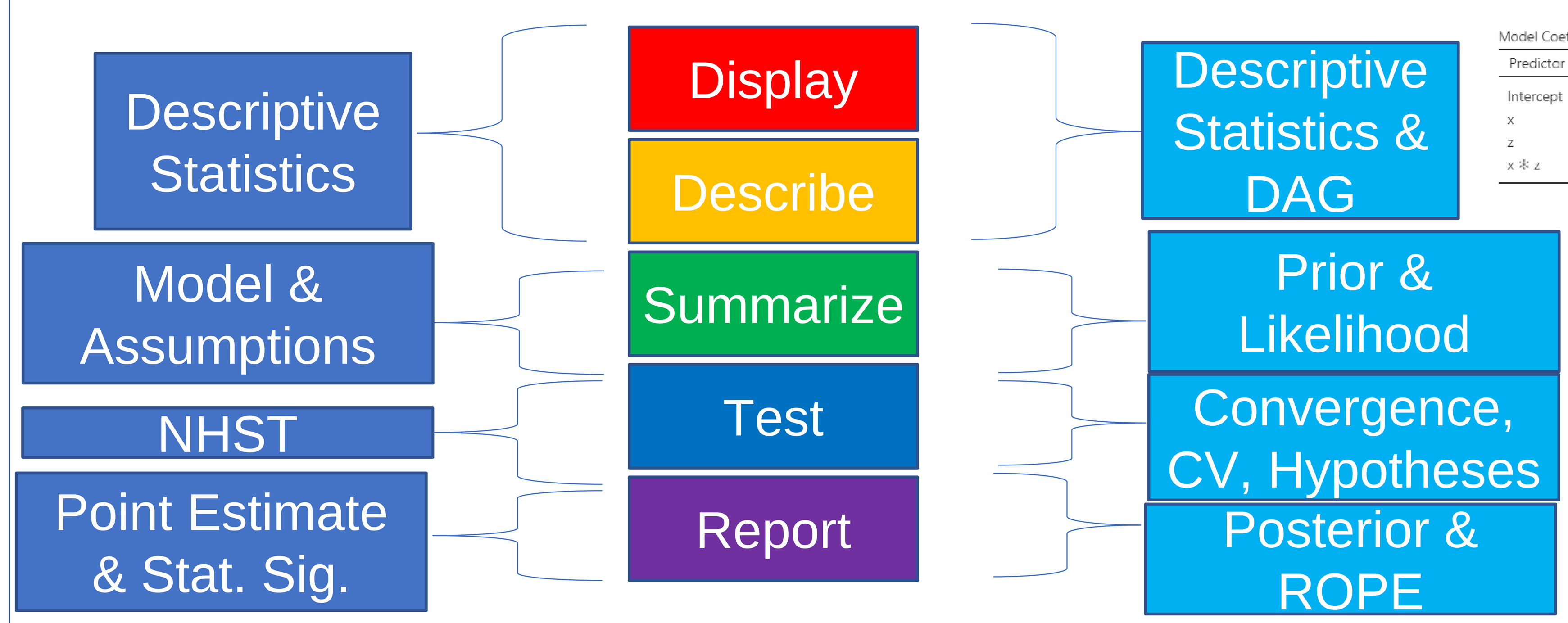
Statistical Modeling in Educational Research

Traditional statistical inference in educational research is based on frequentist statistical modeling and null hypothesis statistical testing (e.g., Murnane & Willett, 2010). This is problematic given the often small, nested, and idiosyncratic nature of data in many educational interventions (Tutwiler & Bressler, in press; Gelman & Carlin, 2014). Many tools designed to handle this data complexity in the traditional modeling toolkit require robust sample sizes and the acceptance of untenable assumptions (McElreath, 2016). Taking a Bayesian modeling perspective allows for the modeling of sparse, complex data at the cost of computational time (Gelman et al., 2020; McElreath, 2016).

Table 1. Potential Bayesian statistical modeling solutions to common issues in statistical modeling in educational research

Traditional Statistical Modeling Issue	Potential Bayesian Solutions
Assumes static trend and random sampling	1) Data held constant and estimates sampled via MCMC
Testing a null hypothesis isn't meaningful much of the time	1) Report full posterior & ROPE 2) Posterior contrast
Interpretation assumes no measurement error	1) Report full posterior & ROPE 2) Bayesian modeling of measurement error
Interpretation of NHST assumes adequate statistical power	1) No NHST 2) Report full posterior & ROPE
Missingness can lower power and/or ruin generalizability	1) No NHST 2) Bayesian imputation
Outliers can majorly bias estimates	1) Regularizing priors
Models can "over-fit" the data and do a bad job predicting trends in new data sets	1) Regularizing priors 2) Bayesian model averaging/stacking
Statistical interactions can be difficult to interpret, especially in non-linear models	1) Posterior contrast

Comparison of Workflows



Basic Bayesian Workflow

1) Specify hypotheses via Directed Acyclic Graphs
2) Specify priors & likelihood
3) Assess model convergence, predictive strength, and robustness across priors
4) Communicate full Posterior, confidence, and Region of Practical Equivalence
5) Evaluate evidence for given hypotheses

Directed Acyclic Graphs (DAGs)

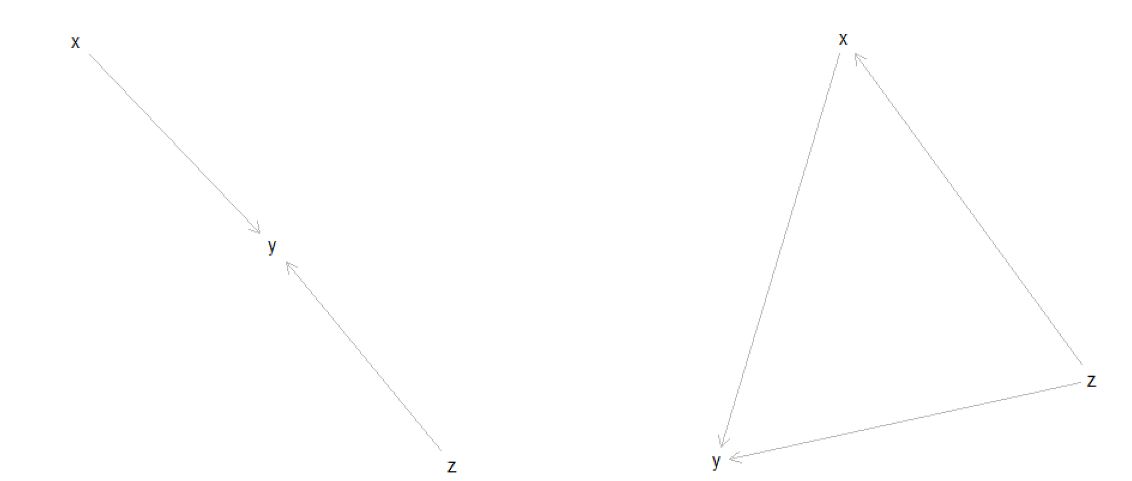


Figure 1. Two DAGs depicting the effects of x and z on an outcome of interest (y). The left DAG assumes independent effects of x and z, the right DAG assumes x is influenced by z.

Priors & Likelihoods

Eqn. 1. Likelihood & priors for linear model of Y given X and Z

$$Y_i \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 (X_i * Z_i)$$
$$\beta_0 = Normal(0, \sigma_{class})$$
$$\beta_{1,2,3} = Normal(0,1)$$
$$\sigma \sim Exponential(1)$$
$$\sigma_{class} \sim Exponential(1)$$

Eqn. 2. Likelihood & priors for linear model of Y given X and Z accounting for measurement error in Y

$$\boldsymbol{Y_{est,i} \sim Normal(\mu, \sigma)}$$
$$\mu = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 (X_i * Z_i)$$
$$\boldsymbol{Y_{obs,i} \sim Normal(Y_{est,i}, Y_{se,i})}$$
$$\beta_0 = Normal(0,10)$$
$$\beta_{1,2,3} = Normal(0,1)$$
$$\sigma \sim Exponential(1)$$
$$\sigma_{class} \sim Exponential(1)$$

Eqn. 3. Likelihood & priors for linear model of Y given X and Z accounting for missingness in X

$$Y_i \sim Normal(\mu, \sigma)$$
$$\mu = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 (X_i * Z_i)$$
$$\boldsymbol{X_i = Normal(\pi, \sigma_x)}$$
$$\beta_0 = Normal(0,10)$$
$$\beta_{1,2,3} = Normal(0,1)$$
$$\boldsymbol{\pi = Normal(0,1)}$$
$$\sigma \sim Exponential(1)$$
$$\sigma_{class} \sim Exponential(1)$$
$$\boldsymbol{\sigma_x \sim Exponential(1)}$$

Table 2. Comparison of traditional linear (LEFT) and Bayesian multilevel model (RIGHT) estimates of Y on X and Z (Eqn. 1).

Model Fit Measures

Model	R	R^2
1	0.525	0.276

Model Coefficients - y

Predictor	Estimate	SE	t	p
Intercept	1.036	0.163	6.356	< .001
x	0.440	0.166	2.659	0.013
z	0.291	0.195	1.491	0.148
x ✻ z	0.158	0.200	0.791	0.436

Parameter	Median	CI	CI_low	CI_high	pd	ps	ROPE_CI	ROPE_low	ROPE_high	ROPE_Percentage	Rhat	ESS
(Intercept)	1.02030524	0.95	0.46051376	1.6061352	0.9966667	0.9960000	0.95	-0.09527025	0.09527025	0.00000000	1.0010991	905.4142
x	0.51131106	0.95	0.21511953	0.8041783	0.9990000	0.9970000	0.95	-0.09527025	0.09527025	0.00000000	1.0018351	2,407.7884
z	0.32797552	0.95	-0.01588164	0.6795958	0.9726667	0.9130000	0.95	-0.09527025	0.09527025	0.06526316	1.0010148	2,517.3161
x:z	0.08976467	0.95	-0.26942791	0.4460506	0.6946667	0.4893333	0.95	-0.09527025	0.09527025	0.37017544	0.9999908	2,241.1051

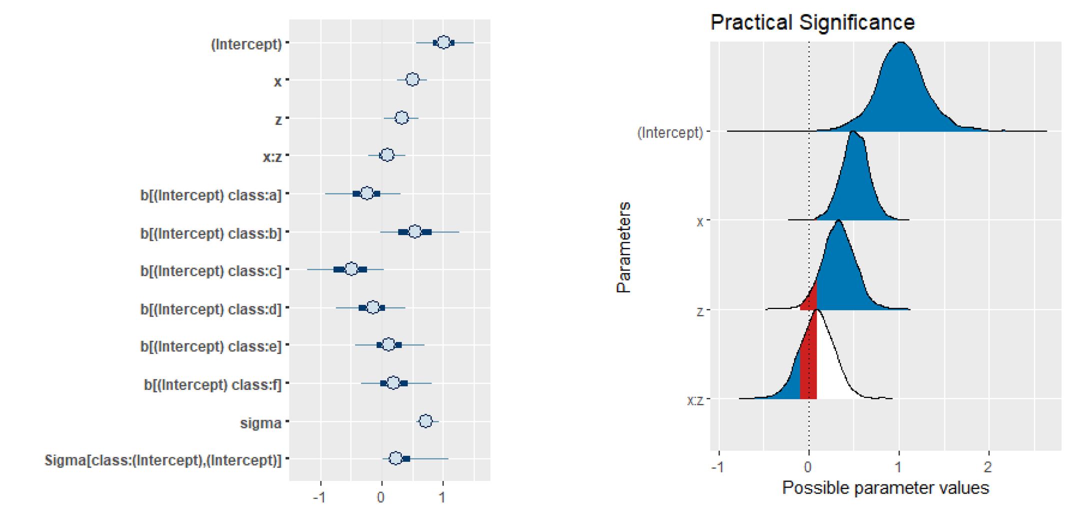


Figure 2. Posterior estimates of Eqn. 1, showing random effects (LEFT) and Region of Practical Equivalence (RIGHT)

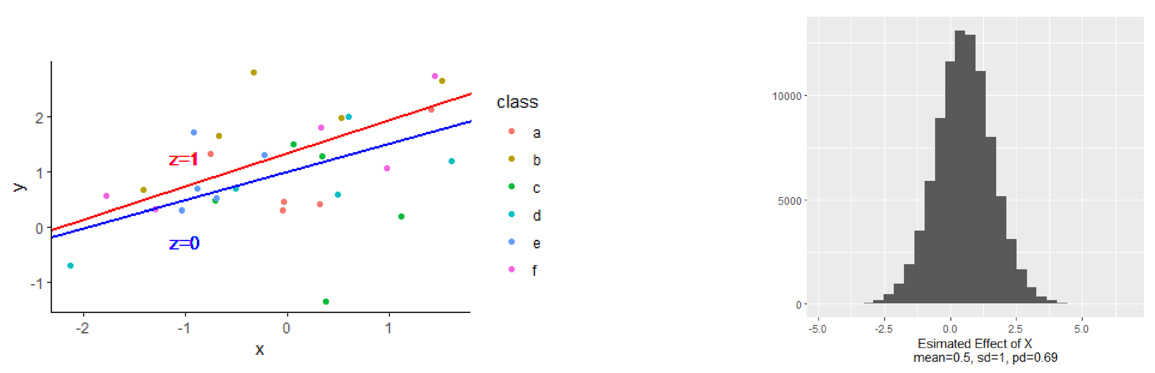


Figure 3. Relationship between X and Y for high (1) and average (0) values of Z (LEFT), and posterior contrast estimate X (RIGHT).

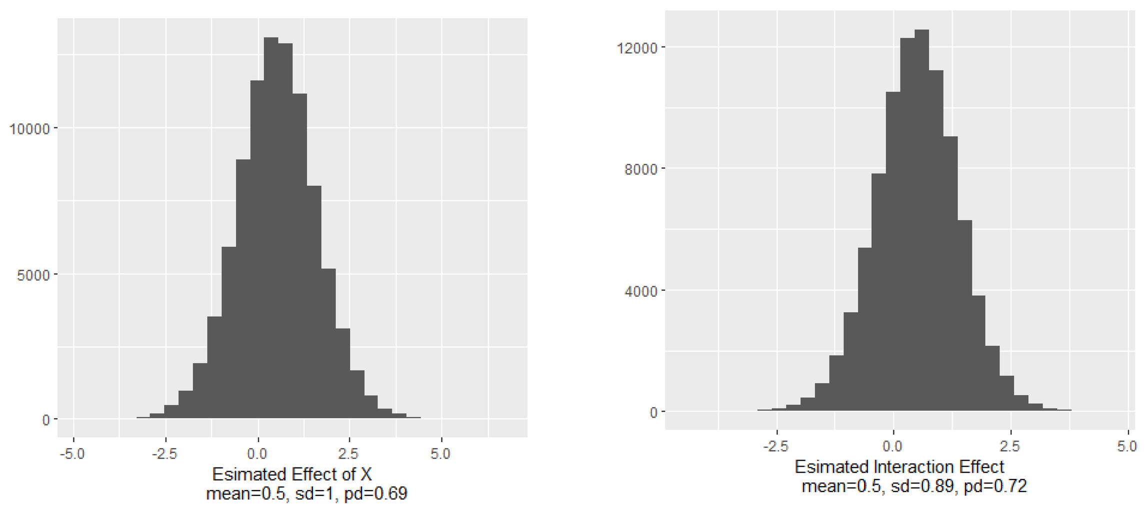


Figure 4. Relationship between X and Y for high (1) and average (0) values of Z (LEFT), and posterior contrast estimate X (RIGHT).

Discussion

In this poster, I demonstrated the basic workflow required to test hypotheses in educational data via the rstanarm package in R (Goodrich et al., 2023). Expansions to the given example include accounting for measurement error directly in the model and imputing missing variables directly in the model fitting process. Given the joint occurrence of noisy, nested data and often small sample sizes with frequent missing values, moving to a Bayesian analytic framework would allow educational researchers to more robustly recover data generating models and test scientific hypotheses of interest, vice the often uninteresting and unrealistic hypothesis that a given relationship is precisely zero in a population of interest.

Contact

M. Shane Tutwiler, EdD
University of Rhode Island
School of Education
shane_tutwiler@uri.edu

References

Goodrich B, Gabry J, Ali I & Brilleman S. (2023). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.4 https://mc-stan.org/rstanarm.
Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. Perspectives on Psychological Science, 9(6), 641-651.
Gelman, A., Hill, J., & Vehtari, A. (2020). Regression and other stories. Cambridge University Press.
McElreath, R. (2016). Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC.
Murnane, R. J., & Willett, J. B. (2010). Methods matter: Improving causal inference in educational and social science research. Oxford University Press.
Tutwiler, M.S., & Bressler, D.M. (in press). Toward a Framework for Robust Design Based Research. Educational Innovates & Emerging Technologies.