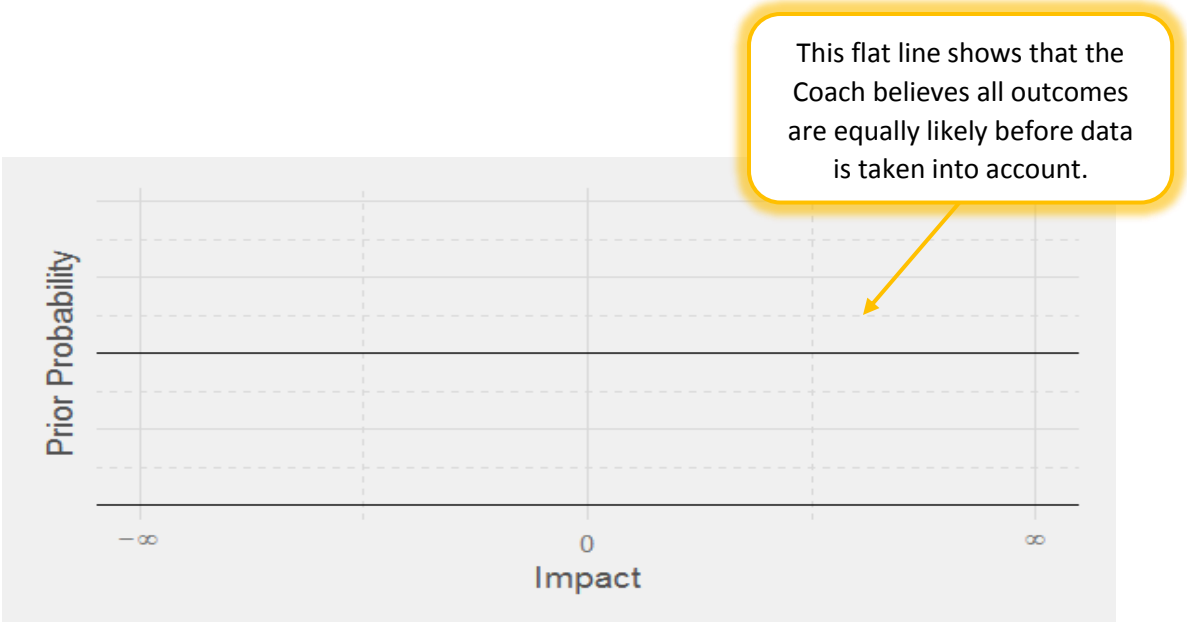
Impact Estimation Technical Appendix

OVERVIEW

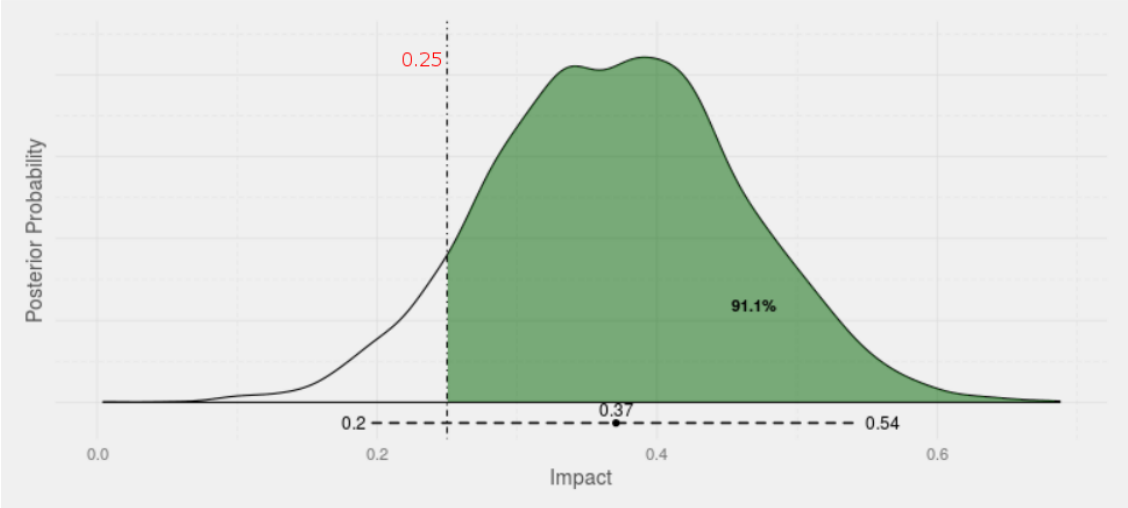
The RCE Coach will help you determine if an educational technology is moving the needle. Before you load the data, the Coach is unable to determine whether or not the technology is effective. That is, the Coach believes it is equally likely that the educational technology is good, bad, or has no effect at all. This is what statisticians call the prior distribution, Figure 1. A prior distribution describes the analyst's beliefs about a measurable outcome before any data are taken into account. Without data, we assume all outcomes are equally likely. That is, the probability the impact is 0 is equal to the probability the impact is 5 is equal to the probability the impact is -5, etc.

Figure 1. *Prior distribution of the effect of the educational technology on the outcome*



Once the data is loaded, and the other inputs are specified, the RCE Coach calculates the probability that the effect of an educational technology is above or below a threshold selected by the user (in Figure 2, the threshold, as indicated by the vertical line, is 0.25). This is the probability that the educational technology is moving the needle. Figure 2 illustrates this calculation.

Figure 2. *Posterior distribution of the effect of the educational technology on the outcome*



This figure tells us that:

- There is a 91.1% probability that the intervention increases the outcome by 0.25 units or more.
- There is a 95% probability that the true impact of the intervention is between 0.20 and 0.54 units.
- The point estimate (e.g., our best guess) for the effect of the educational technology on the outcome is 0.37. This number is calculated as the mean of the posterior distribution.

This information is included in the Findings Brief for each evaluation.

In this example, the user told the RCE Coach that any increase on the outcome larger than 0.25 units would be considered a success. The Coach calculated that there is a 91.1% probability that the educational technology moved the needle by at least that amount.

Additionally, the Coach estimates a credible interval. In this case, the user told the Coach to calculate the 95% credible interval. This tells the user that there is a 95% probability that the true impact of the educational technology is between 0.20 and 0.54 units, as we see in the second bullet above.

Note that the answer to whether or not the educational technology is moving the needle will depend on the threshold chosen by the user (0.25 units in this example) and how much uncertainty he or she is willing to tolerate. For example, some users may conclude that a technology is moving the needle if there is at least a 75% probability that the technology is meeting the threshold they chose, while others may want to be more confident, leading them to select a higher probability. In the example above, if the user sets a certainty level of 95%, then this technology would not be found to be moving the needle, since 95 is larger than 91.1.

THE RCE COACH

The RCE Coach uses a program called Stan (see Carpenter et. al) to estimate a linear model using Bayesian statistics. In particular, the Coach uses the R package RStan. The Engine acts as a simple graphic interface to help the user construct a formula. The user has to select the outcome, treatment indicator, and any additional covariates that the user may want to include. Then, the Coach estimates the following model:

$$Y = \alpha + \eta T + \beta X + \varepsilon,$$

where Y is the outcome of interest, T is the treatment indicator, and X is a matrix with the other covariates.[1] The term η tells you the effect of the educational technology, i.e., how much higher the outcome was for the average student using the educational technology, $T = 1$, compared to the average student not using it, $T = 0$. The Coach computes the posterior distribution of η, which describes our understanding of the educational technology's effect after observing the data. Using this distribution, the Coach can calculate the probability that the treatment effect is above or below a threshold selected by the user, as shown in Figure 1.

WHICH VARIABLES SHOULD I INCLUDE AS CONTROL VARIABLES?

You should include all the variables that you think can affect the outcome of interest and for which you have good data. In the case of student achievement, it's common to include previous achievement and other characteristics like indicators for English as a second language, free or reduced price lunch eligibility, etc.

Additionally, sometimes you may need to include an element to the model to account for "clusters" in the data. This applies when technology users and non-users are determined based on existing groups they belong to, rather than determined on an individual basis. In order to correctly estimate the effect of the intervention, you will need to tell the Coach that you used clusters during step 5. By giving the Coach this information you allow it to account for the possibility that students belonging to a cluster can do better or worst just because they belong to that cluster and not because they are using the technology. For example, if users and non-users were determined based on classroom, you would want to indicate which variable identifies classroom clusters so that the Engine can account for classroom-based differences in outcomes, such as performance differences associated with teacher quality.

All the code for the RCE Coach is open-source under the GPL-V3 license, and available on our github repository: https://github.com/mathematica-mpr/MPRDashboards

[1] In the clustered case, we add cluster-specific random effects to the regression model that's given.

CITATIONS:

- Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2016. Stan: A probabilistic programming language. Journal of Statistical Software (in press).

- Stan Development Team. 2016. RStan: the R interface to Stan, Version 2.10.1. http://mc-stan.org