Statistical Modeling, Causal Inference, and Social Science

"The most important aspect of a statistical analysis is not what you do with the data, it's what data you use" (survey adjustment edition)

Posted on August 7, 2018 9:24 AM by Andrew

Dean Eckles pointed me to this recent report by Andrew Mercer, Arnold Lau, and Courtney Kennedy of the Pew Research Center, titled, "For Weighting Online Opt-In Samples, What Matters Most? The right variables make a big difference for accuracy. Complex statistical methods, not so much."

I like most of what they write, but I think some clarification is needed to explain why it is that complex statistical methods (notably Mister P) *can* make a big difference for survey accuracy. Briefly: complex statistical methods can allow you to adjust for more variables. It's not that the complex methods alone solve the problem, it's that, with the complex methods, you can include more data in your analysis (specifically, population data to be used in poststratification). It's the sample and population data that do the job for you; the complex model is just there to gently hold the different data sources in position so you can line them up.

In more detail:

I agree with the general message: "The right variables make a big difference for accuracy. Complex statistical methods, not so much." This is similar to something Hal Stern told me once: the most important aspect of a statistical analysis is not what you do with the data, it's what data you use. (I can't remember when I first heard this; it was decades ago, but here's a version from 2013.) I add, though, that better statistical methods can do better to the extent that they allow us to incorporate more information. For example, multilevel modeling, with its partial pooling, allows us to control for more variables in survey adjustment.

So I was surprised they didn't consider multilevel regression and poststratification (MRP) in their comparisons in that report. The methods they chose seem limited in that

they don't do regularization, hence they're limited in how much poststratification information they can include without getting too noisy. Mercer et al. write, "choosing the right variables for weighting is more important than choosing the right statistical method." Ideally, though, one would not have to "choose the right variables" but would instead include *all* relevant variables (or, at least, whatever can be conveniently managed), using multilevel modeling to stabilize the inference.

They talk about raking performing well, but raking involves its own choices and tradeoffs: in particular, simple raking involves tough choices about what interactions to rake on. Again, MRP can do better here because of partial pooling. In simple raking, you're left with the uncomfortable choice of raking only on margins and realizing you're missing key interactions, or raking on lots of interactions and getting hopelessly noisy weights, as discussed in this 2007 paper on struggles with survey weighting.

Also, I'm glad that Mercer et al. pointed out that increasing the sample size reduces variance but does nothing for bias. That's a point that's so obvious from a statistical perspective but is misunderstood by so many people. I've seen this in discussions of psychology research, where outsiders recommend increasing N as if it is some sort of panacea. Increasing N can get you statistical significance, but who cares if all you're measuring is a big fat bias? So thanks for pointing that out.

But my main point is that I think it makes sense to be moving to methods such as MRP that allow adjustment for more variables. This is, I believe, completely consistent with their main point as indicated in the title of their report.

One more thing: The title begins, "For Weighting Online Opt-In Samples…". I think these issues are increasingly important in *all* surveys, including surveys conducted face to face, by telephone, or by mail. Nonresponse rates are huge, and differential nonresponse is a huge issue (see here). The point is not just that there's nonresponse bias, but that this bias varies over time, and it can fool people. I fear that focusing the report on "online opt-in samples" may give people a false sense of security about other modes of data collection.

I sent the above comments to Andrew Mercer, who replied:

> *I'd like to do more with MRP, but the problem that we run into is the need to have a model for every outcome variable. For Pew and similar organizations, a typical survey has somewhere in the neighborhood of 90 questions that all need to be analyzed together, and that makes the usual approach to MRP impractical. That said, I have recently come across this paper by Yajuan Si, yourself, and others about creating calibration*

weights using MRP. This seems very promising and not especially complicated to implement, and I plan to test it out.

I don't think this is such a problem among statisticians, but among many practicing pollsters and quite a few political scientists I have noticed two strains of thought when it comes to opt-in surveys in particular. One is that you can get an enormous sample and that makes up for whatever other problems may exist in the data and permit you to look at tiny subgroups. It's the survey equivalent of saying the food is terrible but at least they give you large portions. As you say, this is obviously wrong but not well understood, so we wanted to address that.

The second is the idea that some statistical method will solve all your problems, particularly matching and MRP, and, to a lesser extent, various machine learning algorithms. Personally, I have spoken to quite a few researchers who read the X-box survey paper and took away the wrong lesson, which is that MRP can fix even the most unrepresentative data. But it's not just, MRP. It was MRP plus powerful covariates like party, plus a population frame that had the population distribution for those covariates. The success of surveys like the CCES that use sample matching leads to similar perceptions about sample matching. Again, it's not the matching, it's matching and all the other things. The issue of interactions is important, and we tried to get at that using random forests as the basis for matching and propensity weighting, and we did find that there was some extra utility there, but only marginally more than raking on the two-way interactions of the covariates, and no improvement at all when the covariates were just demographics. And the inclusion of additional covariates really didn't add much in the way of additional variance.

You're absolutely right that these same underlying principles apply to probability-based surveys, although I would not necessarily expect the empirical findings to match. They have very different data generating processes, especially by mode, different confounds, and different problems. In this case, our focus was on the specific issues that we've seen with online opt-in surveys. The fact that there's no metaphysical difference between probability-based and nonprobability surveys is something that I've written about elsewhere (e.g. https://academic.oup.com/poq/article/81/S1/250/3749176), and we've

> *got lots more research in the pipeline focused on both probability and*
> *nonprobability samples.*

I think we're in agreement. Fancy methods can work because they make use of more data. But then you need to include that "more data"; the methods don't work on their own.

To put it another way: the existence of high-quality random-digit-dialing telephone surveys does not mean that a sloppy telephone survey will do a good job, even if it happens to use random digit dialing. Conversely, the existence of high-quality MRP adjustment in some surveys does not mean that a sloppy MRP adjustment will do a good job.

A good statistical method gives you the conditions to get a good inference—if you put in the work of data collection, modeling, inference, and checking.

This entry was posted in Bayesian Statistics, Multilevel Modeling, Political Science, Sociology by Andrew. Bookmark the permalink [https://statmodeling.stat.columbia.edu/2018/08/07/important-aspect-statistical-analysis-not-data-data-use-survey-adjustment-edition/] .

---

22 THOUGHTS ON ""THE MOST IMPORTANT ASPECT OF A STATISTICAL ANALYSIS IS NOT WHAT YOU DO WITH THE DATA, IT'S WHAT DATA YOU USE" (SURVEY ADJUSTMENT EDITION)"

Anonymous
on **August 7, 2018 9:37 AM at 9:37 am** said:

*Too noisy

Andrew on **August 7, 2018 9:43 AM at 9:43 am** said:

Typo fixed; thanks for pointing it out!

ghenly
on **August 7, 2018 12:15 PM at 12:15 pm** said:

Is there a "not" missing in your penultimate sentence?