Lesson 9

Introduction to post-stratification

Often, despite your best efforts, you will still have bias in your sample. One or several characteristics (such as age, education, sex, etc.) that are highly correlated with your variables of interest, will be distributed quite differently in your sample compared to their actual distribution among a population (University of Illinois, 2017). For example, in mVAM surveys it is very common for the sample to consist of 80% men, despite the fact that we know women make up at least 50% of the population. This could be because you accidentally oversampled some parts of the population and undersampled others. But most often it is because there are structural biases in your survey that you just cannot feasibly remove. Phone ownership bias is the most frequent type of bias with these types of surveys. In many of the countries and contexts where WFP works, phone-owners are typically wealthier and better educated, and men are much more likely to respond and use a phone than women. This structural bias simply cannot be removed. Another example are internally displaced persons who, for many reasons, are much more difficult to reach than non-displaced households.

Whatever the case may be, this disparity in the joint distribution of the subpopulations in your sample from their true distribution introduces bias into your estimate, as any estimation procedure will give greater weight to those people you oversampled. You can partially correct for these biases mathematically by constructing post-stratification survey weights. Post-stratification adjusts the weights of undersampled and oversampled subpopulations so the overall sample is more representative of the true subpopulation distributions of the actual target population.

Although it sounds simple, post-stratification is generally an advanced topic in survey statistics because it requires quite a bit of experienced judgement on your part. First and foremost, post-stratification creates a trade-off between precision and accuracy. Adjusting the weights increases the stratum variance and thereby the design-effect. In some cases this can be considerable. Too much post-stratification and you can decrease your effective samples by as much as 80%! Your estimates also become much more sensitive to outliers; counter-intuitively, you can further bias your results with post-stratification.

Post-stratification requires finding a sweet-spot between bias and precision while avoiding creating even more bias. However, it's almost always necessary to produce good estimates.

This document will provide a brief overview but putting it into practice does require a skilled statistical analyst. **Please do not try to do so yourself if you do not have the proper statistical training.**

Overview

First, to know if your survey requires construction of post-stratification weights, we need to compare our survey results to an existing dataset, either a census data or current population survey data, that can serve as the 'truth benchmark' for demographic characteristics (Debell & Krosnick, 2009). You would then identify demographic variables in your survey likely to be measured with little error and low non-response (such as age, sex, education, language, etc.) and compare them stratum-by-stratum to the benchmark values. If there are more than a handful of instances where the difference exceeds a few percentage points, then you should consider constructing post-stratification weights (Debell & Krosnick, 2009).

For the purpose of this document, the term post-stratification will imply 'any post-survey reweighting technique used to correct for bias.' There are two families of said techniques: post-stratification and calibration. The key difference between the two families is in the kind of *auxiliary information* -- information about the population from outside the survey -- that they require:

Post-stratification re-weights observations based solely on the **joint-distribution** of the stratification variables and post-stratification variables. That is, you will require as auxiliary information, the population counts of each and every subgroup belonging to your post-stratification variables within each and every stratum of your survey. You can choose one or several sociodemographic characteristics to construct post-stratification weights, but again these variables must have been measured by your survey with little error and low non-response for said weights to be reliable (University of Illinois, 2017). **Please note that all subgroups in your sample that are to be post-stratified must first have a reliable number of observations** (30 being a good rule of thumb for minimum required samples)**.**

Example (from University of Illinois, 2017) - Suppose we have the following survey, stratified by geographies A,B,C for which we did optimal sample allocation (by stratum variance) and we desire to post-stratify on Age and Gender (table 1).

Table 1

Geography	Population	Sample	Inverse Probability Weight
A	20,000	200	20000/200=100
B	30,000	200	30000/200=150
C	50,000	400	50000/400=125

In this case you need population breakdowns by age and gender for each stratum (i.e. the joint distribution) as auxiliary information. The post-stratification weight is then computed as:

Wpoststr =Population %/Sample %

And the final weights are computer by multiplying the inverse probability weight by the post-stratification weight:

*Wfinal =Wipw * Wpoststr*

We illustrate below for stratum A (table 2).

Table 2

Stratum	Age and gender	Population %	Sample Count	Sample %	Post-stratification Weight	Final Weight (IPW=100)
A	18-34, Male	25%	30	30/200=15%	0.25/0.15 =1.67	100*1.67=167
A	18-34, Female	22%	40	40/200=20%	0.22/0.2 =1.10	100*1.10=110
A	35-64, Male	18%	50	50/200=25%	0.18/0.25 =0.72	100*0.72=72
A	35-64, Female	20%	60	60/200=30%	0.2/0.3 =0.67	100*0.67=67
A	65+, Male	5%	10	10/200=5%	0.05/0.05 =1.00	100*1.00=100
A	65+, Female	10%	10	10/200=5%	0.1/0.05 =2.00	100*2.00=200

Suppose you do not have the precise joint-distributions you need to perform post-stratification. This situation can easily arise in many of the countries where remote food security surveys are conducted. Census data is often outdated and there may not exist a similar current population survey because we are often the first ones on the scene after some sort of adverse shock. Calibration allows us to re-weight observations based on both whatever joint-distributions and marginal-distributions of various socio-demographic variables you happen to have in your survey. That is, you can use as auxiliary information the population counts along just a single socio-demographic variable such as age group, not intersected by stratum; joint-distributions of strata and one or more socio-demographic variables; and even joint-distributions of two or more socio-demographic variables not intersected by strata. You can almost think of calibration as using the entire 'mix' of auxiliary information that is available. See Kott, P.S., (2012), for more on calibration. However this is a highly advanced topic in survey statistics not typically introduced without proper graduate-level coursework. **It is extremely difficult to do calibration correctly** and the mathematics involved are also quite advanced. **We do not recommend anyone perform calibration without first properly understanding the mathematics, risks, and intuitions involved.**

Weight Trimming

After you have post-stratified or calibrated your survey we may find that some subgroups have either extremely small or extremely large post-stratification weights. As mentioned earlier this can actually make your estimate worse by increasing your variance and sensitivity to outliers. Statisticians often 'trim' these weights, truncating the extreme high or low weight values in order to reduce their impact on the variance of the estimates, especially for subgroup estimates. By truncating extreme weight values one generally lowers sampling variability but may incur some bias (Battaglia et al. 2009). However, the mean squared error of key outcome estimates will only be lower if the reduction in variance is larger than the relative increase in bias arising from weight trimming. There are no established rules for weight trimming; rather most people use a general set of guidelines. Some common truncation points are (Battaglia et al. 2009):

1. the median weight plus and minus 5-6x the interquartile range (IQR) of the weights
2. 5x and 0.2x of the mean weight
3. the 5th and 95th percentile of the weights
4. 0.2 and 5

Weight Normalization

An undesired consequence of weight trimming is now the weights of the entire sample will not add up to the known population size. If the discrepancy is minor between the sum of the weights of the population, we have no need to normalize weights. This is actually common practice in many large population surveys (Battaglia et al. 2009). However a significant difference between the two requires we normalize weights to be consistent with our original population. Fortunately this is very simple to do, as the term 'normalize' implies, we simply divide the new weights by their average.

References

Battaglia M, Hoaglin DC, Frankel, MR. (2009). Practical considerations in raking survey data. *Survey Practice* (Online Journal). http://www.surveypractice.org/index.php/SurveyPractice/article/view/176/html

Debell, M. & Krosnick, J. A. (2009). Computing weights for American National Election Study Survey Data, (Online Journal). http://www.electionstudies.org/Library/papers/nes012427.pdf

University of Illinois. 'Post-stratification Weights'. Retrieved March 06, 2017, from atlas.illinois.edu/support/stats/resources/spss/create-post-stratification-weights-for-survery-analysis.pdf

Kott, P.S., (2012). An Introduction to Calibration Weighting for Establishment Surveys. *American Statistical Association*. https://ww2.amstat.org/meetings/ices/2012/papers/302286.pdf

WFP vam food security analysis

For more information contact:
wfp.mvam@wfp.org