search ReStore here   [Search]

PEAS Home   ReStore site ▽   This web resource ▽   Help ▽   Contact us

FAQs   Contact us   Give us feedback

web version

5. Adjusting for non-response by weighting

>> 5.1 Why?
>> 5.2 Post stratification
    >> 5.2.1 How to post-startify >> 5.2.2 Why it is rarely elaborate
    >> 5.2.3 Raking and calibration
    >> 5.2.4 Are non-response weights the same as design weights?
    >>5.2.5 Software for poststratification

5.3 From sampling frame 5.4 Assumptions 5.5 Effect on standard errors: is it worth it?

▲ top

5.1 Why adjust for non-response?

For voluntary surveys the greatest threat to the accuracy of the survey estimates is non-response. Different surveys achieve different response rates, the surveys with the highest response rates tending to be those that ask questions that seem relevant and interesting to respondents. But, even with 'popular' surveys, response rates have been declining in recent years, and, as a direct consequence, worries about survey bias have been increasing.

Non-response is only a problem if the non-respondents are a non-random sample of the total sample. Unfortunately, this seems almost always to be the case. In household surveys, for instance, there is lots of evidence that non-respondents are younger than respondents, and that men are harder to persuade to take part than women. Response rates also tend to be lower than average in cities and in deprived areas.

The result of these patterns is that the achieved samples for surveys often don't reflect the population they are meant to represent very well. Surveys typically over-represent women, and those over the age of 30. And they often under-represent those living in cities and deprived areas.

Rather than accept a poor match between the sample and the population, it is now common for survey data-sets to use weights to bring the two more closely into line. This is known as 'non-response weighting'.

Another approach to survey non-response is imputation. This is covered in section 6 of the background theory in the P|E|A|S site, where there is a brief discussion of the choice between the two approaches.

▲ top

5.2 Adjusting for non response with post-stratification

5.2.1 How the weights are calculated using population data

Most non-response weighting schemes involve 'post-stratification '. This is essentially a two-step procedure:

(i) identify a set of 'control totals' for the population that the survey ought to match;

(ii) calculate weights to adjust the sample totals to the control totals.

As a very simple illustration, suppose the adult population distribution by age and sex is:

population	%
men 16-39	22.1
men 40-59	14.9
men 60+	11.0

women 16-39	22.0
women 40-59	15.0
women 60+	15.1

but the sample distribution is:

population	%
men 16-39	20.2
men 40-59	13.8
men 60+	12.5
women 16-39	21.5
women 40-59	15.8
women 60+	16.2

To poststratify the sample, weights would be calculated that bring the sample distribution into line with the population. The weight applied to men aged 16-39 would be 22.1/20.2; the weight applied to men aged 40-59 would be 14.9/13.8; and so on.

▲ top

5.2.2 Why post-stratification is rarely very elaborate

The main constraint on what can be done using post-stratification is that, for the calculations, we need to know the population distributions. This automatically limits what can be done: the only control totals that can be used are the ones available and known to be accurate.

So, for most national surveys, control totals tend to be age and sex within geographical areas. Other control totals that might be useful, such as social class, are not available in GB except in census years, and, even then, concerns about the reliability of the match between social class as captured by the census and social class as captured by surveys, tends to rule it out.

▲ top

5.2.3 Different approaches to post-stratification and : raking and calibration

In its simplest form post-stratification compares an n-way table for the population with the equivalent n-way table for the sample. A weight is then calculated per cell of the table to adjust the sample to the population.

There are alternative methods to use in the setting where the full n-way table for the population is not available, but the marginal distributions are. (If, for example, you know the sex distribution and the age distribution, but not the sex by age distribution.) The usual approach is to use a method called raking or iterative proportional fitting. Exemplar 1 uses the Family Resources Survey which uses this method.

Another post-stratification approach that is rapidly gaining ground is 'calibration weighting'. Calibration is now becoming the standard for surveys of individuals selected via households, especially in the particular case where more than one individual is selected per household. The reasoning is that standard post-stratification will tend to give a different non-response weight to each member of a household and this can create difficulties for household-level analyses. Calibration uses an iterative procedure to create weights that bring individual level survey data into line with the population, but with the constraint that all individuals within a household must have the same weight. The underlying assumption is that non-response is primarily a household decision rather than an individual decision (in the sense that non-response is usually determined on the doorstep before any individuals are selected) and it is household level non-response that creates most of the discrepancy between the sample and population distributions.

Regression models can also be used in simpler situations (See exemplar 5) as an alternative to simple post-stratification . They are most useful when a lot of information is available from the population so that the cells for post-stratification would be too small and might result in very variable weights that have a lot of sampling error. The regression models effectively smooth the weights so as to get more stable estimates.

▲ top

5.2.4 Are non-response weights the same as design weights?

Nearly but not quite. We are not quite sure if it makes any difference in theory, but in practice it doesn't seem to. A possible difference is that design weights are known exactly but non-response weights are only estimated.

For design weights we know how many units were selected and how many were in the sampling frame. Assuming that either we had no non-response (we should be so lucky!) or that we had no non-responding PSUs and, within PSUs, non-response was not related to any survey features (MCAR - see imputation theory section 6.2), then we can calculate the exact probability that any unit is present in the sample.

Non-response weights are estimated by comparing responding units to totals from the population or from the sampling frame. If we repeated the sampling procedure many times we would get different numbers of non-responding units in each post-strata, even if the MCAR assumption was true. This would give different non-response weights in each possible sample. This uncertainty in the exact value of non-response weights should be reflected in the standard errors of the non-response adjusted analyses. Only replication methods (see design effects section 4.4) such as jackknives and bootstrap methods include this adjustment. If the ⓘpost-stratification is based on a simple model (as recommended for other reasons in section 5.2.2) the contribution to the standard error from the uncertainty in the weights will be very small.

5.2.5 Software considerations for postratification and raking surveys

Post-stratification involves two steps:

1. the reweighting of the sample to match the totals
2. the subsequent analyses

The first of these is easy for just a single post-stratification factor, but it is more tricky when raking or calibration are used. Subsequent analyses strictly requires replication methods if the full benefits of post-stratification are to be seen in the standard errors of the estimates.

Both the R survey package (R section A.5)and Stata's replication package (Stata section D.6) can carry out both of steps 1, and 2. in an integrated fashion. The R package allows you to analyse post-stratified surveys without using replication methods, based on a generalised calibration method.

Raking can be carried out by the SAS macro CALMAR that can be downloaded from the web site of INSEE (the French National Statistics Service) link to INSEE web site for CALMAR . Documentation is in French. Some limited translation is available in the SAS code for exemplar 1 which shows how to use it. A new version (CALMAR 2) with documentation in English is promised in the next few months.

Step 2. could be programmed in SAS, and we give a simple example of how to program a Jacknife method in exemplar 1. . A paper with some code is available.

▲ top

5.3 How the weights are calculated using survey information from sampling frame

Sometimes non-response re-weighting can be carried out by comparing the characteristics of those who responded to a survey with the whole group who the survey attempted to reach. This will not be very helpful when (as in most household surveys) we don't know much about the people who do not respond.

This approach is most helpful when we are selecting a sample form an informative sampling frame. Examples might be surveys of a workforce, where we know the grade, age, length of service of all employees. Another circumstance when this is used is in the context of longitudinal surveys when a survey is re-contacting people who responded at a previous wave of the survey. This is done in three steps.

(i) Carry out an investigation of which factors predict that a response has been received.
(ii) Apply a weight to the responding cases that is proportional to 1/(probability of responding).
(iii) Finally, the weights, at this stage generally above 1.0, are usually rescaled so as to add to the responding sample numbers.

At the first step a response variable is attached to the sampling frame that is coded as 1 for responders and 0 for non-responders. Where the sampling frame only tells us a few things about the units we can divide the sample up into groups (called response classes) and the probability of response is simply the proportion who respond in each response class.

For example, in a survey of community groups one might know only a limited number of things about them such as their location and their funding source. This gives four response classes, as shown below, and the probability of response can be calculated for each.

Sampling frame group	Total	responses	probability response	weight	rescaled weight
urban- public funding	68	34	0.50	2.00	1.18
urban - not public funded	39	12	0.31	3.25	1.92
rural public funded	32	20	0.63	1.60	0.94
rural not public funded	56	49	0.88	1.14	0.67
TOTAL sample sum of weights	195	115	0.59	195	115

Table 5.1 Sampling Frame Group

The most underrepresented group (urban, not publicly funded) gets the highest weight to bring the sample into line with the sampling frame.

When the sampling frame contains more detailed information about the non-responders the factors that influence non-response are often investigated via logistic regression. The resulting model is then used to calculate the probability of response at step (i) above, and the subsequent steps are carried out in the same way as above.

When this type of regression model is used we need to strike a balance between having a powerful model to predict non-reponse (and so reduce bias) and the introduction of extreme weights that will affect precision see weighting section 3.7. Models are often simplified at the final stage to avoid extreme weights (either large or small). Another practice that some surveys employ is to cap the weights, for example by replacing all weights above 2.5 with the value of 2.5.

▲ top

5.4 The assumptions made

Survey statisticians claim that non-response weighting should reduce bias,
but they acknowledge it will not eliminate bias. The reasoning is:

- The reasons individuals choose to take part in surveys are complex, and depend upon lots of factors specific to individuals. As a result, the non-response biases in surveys are likely to be complex;
- Post-stratification works on the assumption that, by aligning the survey to the population along a small number of dimensions (such as age and sex), many of these complex biases will reduce. But it would be misleading to suggest they will be eliminated.

To illustrate the point: In the example above, where a survey is post-stratified to the population for age and sex, for non-response bias to be removed we have to make the assumption that the men aged 16-39 who responded to the survey are representative of men aged 16-39 in the population. The same assumption has to be made for the other age and sex groups too. But it is easy to imagine scenarios where this is not the case: for instance, amongst men aged 16-39 it could be that those who live alone were less likely to take part in the survey that men who lived with others, or it could be that men aged 16-39 in professional occupations were less likely to take part than others. Or more intangibly, but realistically, it could be that men aged 16-39 who are uncomfortable about talking to strangers about themselves are less likely to take part than others. Post-stratification by age and sex will not eliminate survey bias such as these.

This is an illustration of the idea of the missing respondents not fulfilling the requirement of being missing at random (MAR - see imputation section 6.3)

▲ top

5.5 The effect of non-response weighting on standard errors . Is it worth it?

the main motivation for post-startification is to remove bias. But poststratification should, if

done well, reduce the standard errors of most survey estimates (relative to only applying inverse selection weights). The exceptions are:

(i) if the variables used as control totals are unrelated to the survey variables;
(ii) if there are small numbers of extremely large weights.

In most cases you should be able to assume that the survey organisation responsible for the calculation of the weights will have checked for both of these. (The standard practice is, for instance, to trim very large weights.) However, it is good practice to check the distribution of the weights, and if there are some very large weights, to make sure you understand how and why they have arisen. Chances are they are there for a reason but errors do happen.

But post-stratification also brings problems.

- It relies on the totals being correct.
- If post-stratification cuts accross strata correct standard errors require more complicated methods of analysis. Replication methods are the usual way this is dealt with and they are complicated to use. The alternative of using a calibration method, is only available in one of the packages we include (R, Survey package). It seems to give results very close to those from the replication methods.

Additionally, the benefits of ⓘpost-stratification (like stratification) are largely confined to whole sample analyses. WIth subgroups that cut accross strata they both of these procedures may have much less benefit (see exemplar 1 section 1,6)