



## Data, Measures and Methods

# Analyzing Quota Sample Data and the Peer-review Process

Thomas Gschwend

Mannheimer Zentrum für Europäische Sozialforschung (MZES), University of Mannheim, 68131 Mannheim, Germany. E-mail: Thomas.Gschwend@mzes.uni-mannheim.de

There is no random sampling in France. Thus, when analyzing French survey data, every scholar has to deal with an important methodological challenge: How can one use the standard panoply of significance tests on quota sample data? This essay suggests some strategies for successfully dealing with such enquiries during the peer-review process. Scholars should gather as much external evidence as possible to argue that their achieved sample represents the population on as many dimensions as possible. The more evidence they are able to compile, the more confidence there is that their estimation results are robust even based on quota sample data.

*French Politics* (2005) 3, 88–91. doi:10.1057/palgrave.fp.8200068

**Keywords:** survey research; significance testing; simple random sample; quota sample; French public opinion

---

Scholars studying voting behavior or public opinion in France face a serious problem getting their work published in peer-reviewed journals. While there is no simple random sampling in France,<sup>1</sup> significance testing is appropriate only in probability samples. Hence scholars potentially have to go a long way to convince reviewers and journal editors that their modeling attempts are justified. Essentially all survey data available for secondary analysis in France are not probability samples. The use of quota samples is a common practice among French polling organizations. While the particular quota sampling methodology might vary slightly from institute to institute, at least the last selection stage, to decide at the end on who gets administered the questionnaire, is not random, as it is the case in probability sampling. Instead, it is left to the arbitrariness of the respective interviewer of how the given quotas are fulfilled that are representative of the whole population. Herein lies the crux of the quota *vs* probability sample controversy. Instead of repeating the main arguments of this debate, I would like to focus on the practical implications of this debate for scholars using French survey data and, finally, suggest possible ways to convince the involved actors in the peer-review process that some modeling attempts might be justified after all — even with quota sample data.



Undeniably, quota sampling has advantages over any type of probability sampling. It is a cheap and convenient method to generate a sample quickly because it is easy to administer and does not need any sampling frame. Proponents of quota samples are not tired of arguing that they can approximate a stratified random sampling scheme employing elaborate and very restrictive quotas, that is, many fine-grained so-called quota controls such as gender, age, occupation, social class, education, size of city and region. Quota sampling, though, has — also undeniably — its disadvantages. Validity and reliability of quota controls depend on precision and objectivity of administrative data sources. Moreover, even well-applied quota controls do not remove the possibility of a biased selection of respondents within each cell. This is grounded in the nature of the interview situation in such a setting since interviewers do not push hard for a particular case and rather go with the most cooperative respondents. Furthermore, it is also difficult to check whether interviewers ‘creatively’ redefine a respondent to fulfill a given quota instead of having to make many more calls. Most important, though, is its theoretical weakness. In general, it is neither clear according to statistical theory how to compute a standard deviation, nor how to estimate standard errors, or whether there is any other way to systematically assess the expected variability in quota sampling.<sup>2</sup> Significance testing is only appropriate in probability samples.

Thus when analyzing French survey data, every scholar has to deal with this important methodological challenge: How can one use the standard panoply of significance tests on quota sample data? One solution — used quite frequently — is to just analyze the achieved sample as if it were generated through a random process assuming that the sample is a probability sample. What if reviewers or journal editors rightfully insist, however, that every estimate is worthless unless accompanied by any measure of uncertainty and ask how to justify standard errors with quota sample data? Purist advice would be to abandon all modeling attempts using this kind of data since there is no statistical theory that justifies the use of any measure of uncertainty estimated from a non-probability sample. Since we are interested in modeling aspects of French politics we ought to make the best of this situation, given these data constraints. I would now like to provide some practical advice and suggest some strategies for dealing with such enquiries during the peer-review process.<sup>3</sup>

Using standard significance tests on data generated through any sampling methods assumes in addition to other things that the achieved sample behaves as if it were generated by a random selection process. Thus scholars could run their most preferred regression models even on quota sample data, but must understand that any deviation from the probability sampling standard has potentially important consequences for the robustness of their findings (Oakes, 1986, 156). Without the independence of observations, ensured through randomization, or an explicitly modeled dependence structure of observations



in the sample at hand, every derived variance estimate is biased. Hence, the robustness of significance tests is all the more questionable the more the achieved sample deviates from a 'hypothetical' probability sample. I will concentrate on two ways to bolster the robustness of estimation results based on quota sample data.

One way to pre-empt the robustness issue is to show that the achieved sample represents the population along as many controls as possible — not only the ones that were used to generate the sample in the first place. One can either try to show that marginal distributions of socio-demographic variables fit administrative records or show whether one can successfully predict political outcomes. Both strategies are not unproblematic, though. My experience is that socio-demographic variables that are not used in the quota selection stage are less reliable. Take, for instance, religion or trade-union membership, two frequently reported socio-demographic variables. They are typically not used as controls — for good reasons. The strict separation of church and state in France implies that the state does not have its own statistics. One has to rely on the disclosure of membership statistics of various religious affiliations with all the problems that come with this. Moreover, trade-union membership is a very political issue, since the trade unions in France are known to be more politically radical than in other countries. Depending on the source, one will find different estimates. Furthermore, the problems with predicting political outcomes, that is, turnout estimates and reported vote intentions, from survey data are well known. One cannot expect to get a perfect fit. Thus, on the one hand, such comparisons are easier to make for the data analyst since every misfit can be readily explained. On the other hand, though, such comparisons alone generate less convincing evidence that the achieved sample is in fact 'broadly' representative of the population, and not only by design along dimensions defined by the controls of the quota sampling process.

Another way is to compare marginal distributions or cross-tabs among various surveys that are conducted at about the same time. If the distributions of certain variables of the achieved sample is replicated with different survey data, then one has gathered strong evidence that the achieved sample represents the same population within appropriate confidence intervals due to sampling and measurement error.<sup>4</sup> If possible, show that the distribution of (potentially a combination of) core variables of your analysis can be replicated within other samples, too. This argument is, of course, particularly convincing, if you can successfully replicate distributions of some of your variables in probability samples. Even if you have only quota samples as in the case of France, this still makes the case that your sample is not systematically different from the other quota samples, at least along these variables. This does imply that the achieved sample is likely to be replicated, a characteristic only probability samples have according to statistical theory. Depending on the



nature of the core variables, one might nevertheless have to acknowledge that some items are particular to the survey one is analyzing and are not asked in any other survey.

To sum up, the panoply of significance testing can, in general, not be justified on quota sample data. Since in all likelihood quota samples are all they have if scholars want to analyze French voting behavior and public opinion, they have to go the extra mile in order to pass peer-review. Scholars should gather as much evidence as possible, at least to ensure themselves that their achieved sample represents the population on as many dimensions as possible. The more evidence they are able to compile the more confidence there is that their estimation results are robust, even based on quota sample data.

### Acknowledgements

I would like to thank Sarah Birch, Siegfried Gabler, Jean-Marie Jungblut, Mike Lewis-Beck, Markus Klein, Evi Scholz, Hermann Schmitt, and particularly Dirk Leuffen for challenging questions and helpful comments.

### Notes

- 1 For administrative reasons, it was in the past not possible to draw a valid random sample. This is no longer the case. For instance, the French module of the European Social Survey (the ESS), a biennial multi-country survey covering more than 20 nations, is a probability sample.
- 2 There might be instances where inferences from quota samples can in fact be justified, see Smith (1983) and King (1985).
- 3 For a successful application of this strategy, see Gschwend and Leuffen (2005).
- 4 A successful replication of certain variable distributions does not imply an existing representativeness along these dimensions, though, because the distributions of these surveys could still be incorrect. Fortunately, the more successful the replications, the less likely this caveat becomes.

### References

- Gschwend, T. and Leuffen, D. (2005) 'Divided we stand — unified we govern? The issue of cohabitation in the French elections of 2002', *British Journal of Political Science*, forthcoming.
- King, B.F. (1985) 'Surveys combining probability and quota methods of sampling', *Journal of the American Statistical Association* 80: 890–896.
- Oakes, M. (1986) *Statistical Inference: A Commentary for the Social and Behavioural Sciences*, Chichester: Wiley.
- Smith, T.M.F. (1983) 'On the validity of inferences from non-random samples', *Journal of the Royal Statistical Society, Series A (General)* 146(4): 394–403.