Leibniz-Institut
für Sozialwissenschaften

GESIS Survey Guidelines

Weighting

Siegfried Gabler, Jan-Philipp Kolb, Matthias Sand, and Stefan Zins

December 2016, Version 2.0

Abstract

This contribution deals with the fundamentals of weighting and with the different types of weights. Terms such as *design weighting* and *adjustment weighting* are explained, and the Horvitz-Thompson estimator and the GREG estimator are presented.

Citation

Gabler, S., Kolb, J.-P., Sand, M., & Zins, S. (2016). Weighting. *GESIS Survey Guidelines*. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_007

1. What is it about?

In most surveys, the underlying samples are not drawn by means of simple random sampling. Rather, *complex sampling procedures* are used. One example is disproportionate stratified random sampling. In such samples, the sample mean is no longer an adequate estimator for the population mean of the variable of interest. For this reason, suitable weights are applied to the individual elements in the sample, and the individual data sets are enhanced with a weighting variable. When these weights are the inverse of the inclusion probabilities, this is referred to as design weighting, which is a statistically sound procedure.

Another case is present when the realised sample deviates from the target sample due to nonresponse, for example. Here, weighting is applied by adjusting the sample data to known marginal distributions of key variables, thereby endeavouring to correct skewness in the sample. However, adjustment weighting should never be applied independently of design weighting. Fundamental articles on this subject can be found in Bethlehem (2002), Dorofeev and Grant (2006), Kish (1965), Lohr (1999), and Särndal et al. (1992).

2. Design weighting

Let *U* denote a population, and let u_i denote the elements of the population, where $i = 1,\ldots,N$. *N* is the population size. Frequently, only the index set $U = \{1,\ldots,N\}$ is used for the target population. A sample *S* comprising *n* units is a sequence of *n* elements $(i_1,\ldots,i_n)$ drawn from *U*. Index i_k represents the unit selected in the *k*th draw; *n* is the *sample size*.

Random samples are characterised by the fact that each possible sample *S* is assigned a known probability *P*(*S*). The set of all samples where *P*(*S*) > 0 is referred to as the *sample space*. The probability of selection, or inclusion,

$$\pi_{ij} = \sum_{S:i,j\in S} P(S)$$

gives the probability that the units *i* and *j* will be selected into the sample. Instead of π_{ii}, we use the shorter notation π_i. If *i* is not equal to *j*, this is referred to as *second-order inclusion probability*; if *i* is equal to *j*, it is termed *first-order inclusion probability*. In the case of simple random sampling (without replacement) of *n* units from a population with *N* units,

$$\pi_i = \frac{n}{N}$$

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)} \quad \text{for } i \neq j.$$

A sample drawn in such a way that all possible samples of *n* elements from the population are equally likely is called a *simple random sample*. All other samples are called *complex samples*. Complex samples also include cluster samples, where the first-order selection probabilities for every element of the population are identical but the second-order selection probabilities are not.

Consider, as a further example, stratified random samples. Assume that the population U is divided into H strata. The stratification of Germany according to federal states would yield 16 strata, for example. If a simple random sample of exactly n_h units is drawn from the hth stratum comprising N_h units, then

$$\pi_i = \frac{n_h}{N_h} \quad \text{for } i \text{ from stratum } h$$

$$\pi_{ij} = \frac{n_h(n_h-1)}{N_h(N_h-1)} \quad \text{for } i \neq j \text{ both from stratum } h$$

$$\pi_{ij} = \frac{n_h n_k}{N_h N_k} \quad \text{for } i \text{ from stratum } h \text{ and } j \text{ from stratum } k \text{ with } h \neq k.$$

When the (unequal) probabilities of selection of the sample units yielded by the sampling design are taken into account at the estimation stage, this is referred to as *design weighting*. Design weights are computed as the inverses of the probabilities of selection of the sampled units and applied to the data. The weights are often normalised, so that the sum of the weights equals the number of observations. In practice, second-order selection probabilities are usually unknown for complex sampling designs. Thus, for variance and confidence interval estimation, most statistical software packages will either ignore the complex design, try to approximate the sampling variance, or use a model-based approach.

The selection probabilities for units (e.g., municipalities) in the first stage of sampling are often unequal. Cities are given a higher probability of selection than small municipalities. However, if an element that had a very low probability a priori of being selected into the sample is selected after all, it has a *larger weight* than an element that had a very high probability a priori of being selected. Hence, the element with a *low probability of selection* has a *large weight*, whereas the element with a *high probability of selection* has a *small weight*. In order to avoid an extreme spread in the weight distribution, a transformation can be used that introduces an upper, and sometimes also a lower, bound.

3. Which estimators are typically used?

The Horvitz-Thompson estimator is used as an unbiased estimator for the total $Y = \sum_{i=1}^{N} Y_i$

$$\hat{Y}_{HT} = \sum_{i=1}^{N} L_i \frac{Y_i}{\pi_i}$$

where $L_i = \begin{cases} 1 & \text{if } i\text{th unit is selected} \\ 0 & \text{otherwise} \end{cases}$ for $i = 1,\ldots,N$.

It is assumed that all π_i are positive. For the variance of the Horvitz-Thompson estimator, we obtain

$$\operatorname{var}\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j} \left(\pi_{ij} - \pi_i \pi_j\right).$$

In the case of a sampling procedure with a fixed sample size n

$$\sum_{j=1}^{N} \pi_{ij} = n\pi_i \quad \text{and} \quad \sum_{i=1}^{N}\sum_{j=1}^{N} \pi_{ij} = n^2$$

applies, and the Yates-Grundy variance estimator

$$\mathrm{v}_{YG} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{L_iL_j}{\pi_{ij}}\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2\left(\pi_i\pi_j-\pi_{ij}\right)$$

yields an unbiased estimate of the $\mathrm{var}\left(\hat{Y}_{HT}\right)$ when all π_{ij} are positive. It is clearly non-negative when $\pi_i\pi_j \geq \pi_{ij}$ applies to all *i* and *j*.

In the case of simple random sampling (without replacement) or stratified sampling, the Horvitz-Thompson estimator and the variance estimator correspond to the usual estimates.

Although the Horvitz-Thompson estimator is always unbiased, its variance may be very large. If this is the case, another estimator should be used. An amusing example of this is given in Basu (1971).

4. GREG estimator

Incorporation of auxiliary information into the estimator

Even in the case of simple random sampling, it may be expedient to incorporate auxiliary information into the estimator, for example from official statistics sources. Examples of this are the ratio estimator and the regression estimator.

Besides the *y* values of interest, *x* values are also assigned to the units of the population. The estimator used then takes the form

$$\overline{y}_S - B\left(\overline{x}_S - \overline{x}_U\right)$$

where $\overline{y}_S$ is the sample mean of the *y* values, $\overline{x}_S$ is the sample mean of the *x* values, and *B* is a known parameter, which is often estimated from the sample.

The advantage of estimators of the aforementioned type is that they then have a smaller variance than the sample mean if the relationship between the *y* values and the *x* values is nearly linear.

GREG estimator

In complex samples, the Horvitz-Thompson estimator can be modified in a similar way, and the resulting general regression (GREG) estimator is

$$\hat{Y}_{GREG} = \sum_{i=1}^{N} L_i w_i Y_i$$

with

$$L_i = \begin{cases} 1 & \text{if } i\text{th unit is selected} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1,\ldots,N\,.$$

and

$$w_i = \frac{1}{\pi_i q_i}\left(1 + c_i\left(\sum_{k=1}^{N}\mathrm{x}_k - \sum_{k\in S}\frac{1}{\pi_k q_k}\mathrm{x}_k\right)'\left(\sum_{k\in S}\frac{c_k}{\pi_k q_k}\mathrm{x}_k\mathrm{x}_k{}'\right)^{-1}\mathrm{x}_i\right)$$

where c_i are positive numbers specified by the statistician and q_i are the response probabilities of the ith element. c_i is often set equal to 1 and then corresponds to the two-phase GREG estimator proposed by Särndal and Lundström (2005: 64). The values of the auxiliary variable K for the kth person are summarised in the K-dimensional vector $\mathbf{x}_k$. As a rule, the response probability q_i for the unit i is unknown. Homogeneous response groups are frequently assumed, or q_i is estimated using a logistic regression model. In the majority of applications, q_i is set equal to 1, which yields the calibration weights

$$w_i = \frac{1}{\pi_i}\left(1 + c_i\left(\sum_{k=1}^{N}\mathrm{x}_k - \sum_{k\in r}\frac{1}{\pi_k}\mathrm{x}_k\right)'\left(\sum_{k\in r}\frac{c_k}{\pi_k}\mathrm{x}_k\mathrm{x}_k{}'\right)^{-1}\mathrm{x}_i\right).$$

For the GREG estimator

$\hat{Y}_{GREG} = \sum_{i=1}^{N} L_i w_i X_i = \sum_{i=1}^{N} X_i$ clearly applies for all auxiliary variables x. In this sense, the GREG estimator is “representative”. The concept of calibration is dealt with in Deville and Särndal (1992).

Example

Consider a simple example where the response probabilities are known. In a company that has 300 male and 1000 female employees, a random sample comprising 100 men and 100 women is drawn. Thirty of the men and 50 of the women answer questions about their job satisfaction. Of the 30 men, 20 are satisfied; of the 50 women, 10 are satisfied. If one wishes to estimate the number of employees who are satisfied with their jobs, one computes

$$20\cdot\frac{1}{\frac{100}{300}\cdot 0.3} + 10\cdot\frac{1}{\frac{100}{1000}\cdot 0.5} = 400$$

as an estimated value. The proportion of employees who are satisfied with their jobs is therefore estimated to be 31%. Here, it is assumed that the response probability is 0.3 for all the men and 0.5 for all the women. But because one knows from experience that the expected response rate of men is 24% and that of women is 50%, the estimate would be

$$20\cdot\frac{1}{\frac{100}{300}\cdot 0.24} + 10\cdot\frac{1}{\frac{100}{1000}\cdot 0.5} = 250 + 200 = 450\,.$$

However, this estimate has one disadvantage: If one had asked about the number of employees who were dissatisfied with their jobs, one would have obtained

$$10 \cdot \frac{1}{\frac{100}{300} \cdot 0.24} + 40 \cdot \frac{1}{\frac{100}{1000} \cdot 0.5} = 125 + 800 = 925$$

and would have therefore estimated that the total number of employees was 1,375. The estimated percentage of employees who are satisfied with their jobs plus the estimated percentage of dissatisfied employees does not add up to one. This can be taken into account by calibrating an all-ones vector. If one sets for the calibration weights w_i, the values $x_i = 1$, $c_i = 1$ for all *i* and $\pi_i = 100/300 = 1/3$ for the men and $\pi_i = 100/1000 = 0.1$ for the women, and $q_i = 0.24$ for the men and $q_i = 0.5$ for the women, and therefore $w_i = \frac{1}{\pi_i q_i} \cdot \frac{N}{\sum_{\text{Respondents}} \frac{1}{\pi_j q_j}} = \frac{130}{11}$ for the men and $\frac{208}{11}$ for the women, the

estimator for the number of employees who are satisfied with their jobs would be

$$20 \cdot \frac{130}{11} + 10 \cdot \frac{208}{11} = \frac{4680}{11} = 425.45$$

that is, 32.72%, and the estimate for the number of employees who are dissatisfied with their jobs would be

$$10 \cdot \frac{130}{11} + 40 \cdot \frac{208}{11} = \frac{9620}{11} = 874.55$$

that is, 67.28%. The estimated total of all employees would then be 1,300, which is the actual number of employees.

Nonresponse weighting

In sample-based survey practice, nonresponse is an unavoidable problem. The level of nonresponse varies across countries. However, within countries, too, response behaviour may depend on factors such as the length of the questionnaire, the choice of survey topic, the survey mode, or the interviewer. If a sampled person does not take part in the survey (e.g., because of refusal or non-contact), this is referred to as *unit nonresponse*. If a respondent fails to answer only some of the survey questions (e.g., the income question), this is known as *item nonresponse*.

What can be done when, as a result of nonresponse, the sample distributions of known variables such as age, sex, and region differ significantly from those in the population?

Figure 1 below illustrates the nonresponse process in the case of a two-phase sampling procedure in which first the sample S is selected and then the subset $r \subset S$.

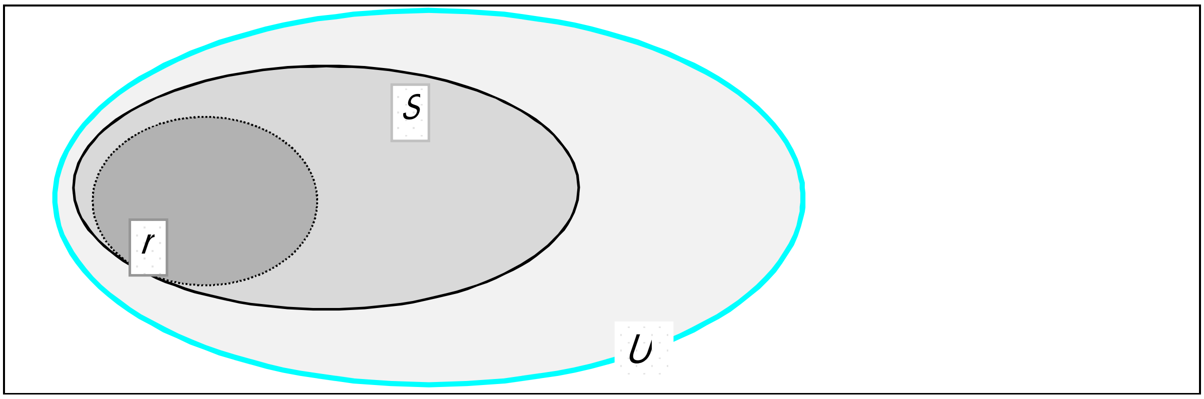

Figure 1. Response set r, sample S, population U with $r \subset S \subset U$

A simple way of taking nonresponse into account in the estimator is to modify the GREG estimator by replacing the sample S in the formula with the response set r:

$$w_i = \frac{1}{\pi_i q_i}\left(1 + c_i\left(\sum_{k=1}^{N} \mathrm{x}_k - \sum_{k\in r}\frac{1}{\pi_k q_k}\mathrm{x}_k\right)'\left(\sum_{k\in r}\frac{c_k}{\pi_k q_k}\mathrm{x}_k \mathrm{x}_k{}'\right)^{-1}\mathrm{x}_i\right).$$

Further deliberations on nonresponse weighting can be found in Elliot (1991).

Of fundamental importance when dealing with nonresponse is the question of the nonresponse model. Typically, a distinction is made between (a) MCAR (missing completely at random), when nonresponse is completely random, (b) MAR (missing at random), when nonresponse in subgroups is completely random, and (c) MNAR (missing not at random), when neither MCAR nor MAR is present. The most frequent application is MCAR, whereby the subgroups must be defined by constellations of variables.

5. Adjustment weighting

Cell weighting, or adjustment weighting, is frequently applied in addition to design weighting. It can be considered when the distributions of external variables (e.g., age, sex, federal state) in the population are known, but they differ from those in the sample. By calibrating the marginal sample distributions to those in the population, it is hoped that the estimator will be improved. In many cases, design weighting is dispensed with, and only adjustment weighting is applied. For statistical reasons, however, design weighting should always precede adjustment weighting. The probabilities of selection are then included in the adjustment weighting.

Depending on whether the joint distribution of the adjustment variables or only their marginal distributions are known, two cases are distinguished:

1. The joint distribution of K variables is known. By means of simple weighting (target/actual), the distribution in the sample is adjusted to match the distribution in the population. This type of weighting is also referred to as *poststratification weighting*. A general formula in the case of K=2 categorical variables is:

   $$w_{ijk} = \frac{1}{\pi_k}\frac{N_{ij}}{\hat{N}_{ij}};\hat{N}_{ij} = \sum_{k\in S_{ij}}\frac{1}{\pi_k}$$

where π_k denotes the first-order inclusion probabilities and s_{ij} contains all units of the sample from the *ij*th cell. When the nonresponse process is random in every cell, a good estimate is usually obtained. Cases where many cells in the sample are empty are problematic. Adjoining cells must then be aggregated.

2. When only the marginal distributions of the *K* variables are known, so-called raking methods are used. The best known of these methods is based on the iterative proportional fitting (IPF) algorithm developed by Deming and Stephan (1941), which is also applied in loglinear data analysis.

6. Other weighting methods

In addition to design weighting and adjustment weighting, the following weighting methods play a role at the analysis stage in surveys across several countries and rounds:

- Analyses on the basis of one country in one round
- Analyses on the basis of several countries in one round
- Analyses on the basis of the combined data sets of one country across several rounds
- Analyses on the basis of the combined data sets of several countries across several rounds

An example of this within the framework of the European Social Survey can be found in Gabler and Ganninger (2004).

Own weighting procedures are needed in the case of longitudinal weighting. The German Socio-Economic Panel (SOEP) is a well-known example. The weighting applied can be found in Schupp (2004).

References

Basu, D. (1971). An essay on the logical foundations of survey sampling, Part 1 (with discussion), In: Godambe & Sprott (Eds.), *Foundations of statistical inference,* 203–242, Holt, Reinhart and Winston, Toronto.

Bethlehem, J. (2002): Weighting nonresponse adjustments based on auxiliary information. S. 275-288 in Robert Groves, Don Dillman, John Eltinge, & Roderick Little (Eds.), *Survey nonresponse*. New York: Wiley.

Deming, E., & Stephan, F. (1941). On a least squares adjustment of a sampled frequency table when the expected marginal tables are known. *The Annals of Mathematical Statistics 11*: 427–444.

Deville, J-C., & Särndal, C-E. (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association, *87*, 376-382.

Dorofeev, S., & Grant P. (2006). *Statistics for real-life sample surveys: Non-simple-random samples and weighted data*. Cambridge University Press.

Elliot, D. (1991). *Weighting for non-response: A survey researcher's guide.* Office of Population Census and Surveys, Social Survey Division.

Gabler, S. (2004). Gewichtungsprobleme in der Datenanalyse, In: A. Diekmann (Ed.), Methoden der Sozialforschung, Sonderheft 44, *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, pp. 128–147.

Gabler, S., & Ganninger, M. (2010). Gewichtung, In: C. Wolf, & H. Best (Eds). *Handbuch der sozialwissenschaftlichen Datenanalyse,* 143–164. Wiesbaden: VS Verlag für Sozialwissenschaften.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Lohr, S. L., (1999): *Sampling: Design and Analysis*. Belmont, CA: Duxbury Press.

Särndal, C-E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. Wiley, New York.

Särndal, C-E., Swensson, B., & Wretman, J, (1992). *Model assisted survey sampling*. New York: Springer Verlag.

Schupp, J. (2004). *Gewichtung in der Umfragepraxis – Das Beispiel SOEP*. http://eswf.uni-koeln.de/lehre/04/04_05/schupp.pdf