Two Measures for Assisting Sample Size Discussions

Beat Hulliger* Philippe Eichenberger† Jann Potterat‡

Abstract

Social surveys often must estimate the sizes or the proportions of many small groups and small differences between group sizes. The discussion of the needed precision of the estimators and the corresponding sample size is difficult since many different objectives must be considered and often lay persons are involved. The size resolution and the difference resolution of a sample are two measures which are derived from approximations to the probability of not observing a group or a difference in the sample. The size resolution is an operationalization of the smallest group which can be estimated from a sample. The difference resolution describes a minimal difference between two group sizes which can be estimated from a sample. Since these resolution measures embody elements of statistical hypothesis tests without the need of a complete test specification they orient the users to a reasonable sample size while remaining simple enough to assist the discussion with various stakeholders. The European Social Survey serves as an example of the application of the resolution measures.

Key Words: Sampling, precision, size resolution, difference resolution

1. Introduction

The discussion of the necessary sample size for a survey is difficult because of the many research questions a survey has to answer. A further complication arises when discussing sample size with lay persons. They often have problems to specify the needed precision for the survey. To answer this question the client would have to determine a relevant effect size and a power for a test he or she wants to obtain. This is difficult even for trained persons and more so for lay persons. In addition usually a survey has to answer multiple questions and often has multiple clients with different needs. The discussion of sample size then may become painful.

The type of survey we are looking at is a multi-purpose survey among persons or households. Actually the development of the proposed resolution measures was initiated during the discussion of the precision of the Swiss Population Survey. This survey should recover some of the informations that were no longer available after the Swiss population census was abolished. The discussion of the needed precision with politicians (e.g. city mayors) and many other users of the Swiss Population Survey was demanding and a simpler description of the needed precision or needed sample size had to be developed.

The critical point for sample size determination for the Swiss Population Survey were various small groups within (possibly small) domains, usually municipalities. For example a mayor of a village may want to estimate the number of inhabitants who speak Portuguese or who commute by train to the nearby city.

The problem is similar in other social surveys like the European Social Survey where rather the size of subgroups within socio-demographic domains should be estimated. For example a user might want to estimate the number of elderly persons with a particular type of education.

*University of Northwestern Switzerland FHNW, Riggenbachstrasse 16, 4600 Olten, Switzerland
†Federal Statistical Office, Espace de l'Europe 10, 2010 Neuchtel, Switzerland
‡Federal Statistical Office, Espace de l'Europe 10, 2010 Neuchtel, Switzerland

Two measures were developed to support the discussion of sample size in a transparent and simple way with stakeholders that lack statistical training: The size resolution and the difference resolution. The size resolution describes how small a group in the population can be such that its size can still be estimated from the survey. The difference resolution describes how small a difference between the sizes of two groups either in different domains or at different time points can be such that the differenct can be still be estimated from the survey.

Eichenberger et al. (2011) describe the derivation of the resolution measures fully. Here we treat the main properties and the use of the resolution measures. We concentrate on the approximate versions, which are most useful for the discussion with lay persons. We introduce the size resolution in Section 2, the difference resolution in Section 3 and consider the derivation of sample sizes based on the resolution measures in Section 4. Some examples of the use are presented in Section 5 and Section 6 concludes.

2. Size Resolution

Consider the case of a simple random sample without replacement and the goal of estimating the size N_A of a group A within a domain D of size N_D. The population proportion is $p_A = N_A/N_D$ and the sample proportion is $\hat{p}_A = n_A/n_D$. The sample proportion $\hat{p}_A$ is an unbiased estimator of the population proportion p_A. Its standard error, conditional on n_D, the sample size in domain D, is

$$\sigma(\hat{p}_A) = \sqrt{(1-f_D)\frac{p_A(1-p_A)}{f_D N_D}\frac{N_D}{N_D-1}} \approx \sqrt{\frac{p_A(1-p_A)}{n_D}}, \qquad (1)$$

where $f_D = n_D/N_D$ is the sampling fraction in D. In principle this settles the relation between sample size n_D and proportion n_A. The usual discussion about sample size then uses the margin of error of a confidence interval for p_A based on $\hat{p}_A$ and an estimate of the standard error. Since many users do not understand these concepts and therefore can not state a desired margin of error they are locked out from the discussion about the relevance of the survey. Cochran (1977) and Lenth (2001) discuss the problems of communication with users. It is of no help then to show the users the typical table of approximate standard errors (Table 1) because the strong dependence on the population proportion makes the discussion even more difficult.

Table 1: Approximate standard error of $\hat{p}_A$ (in percentage points)

	$p_A \cdot 100\%$						
n	0.1	1	2	5	10	25	50
5	1.4	4.4	6.3	9.7	13.4	19.4	22.4
10	1.0	3.1	4.4	6.9	9.5	13.7	15.8
20	0.7	2.2	3.1	4.9	6.7	9.7	11.2
50	0.4	1.4	2.0	3.1	4.2	6.1	7.1
100	0.3	1.0	1.4	2.2	3.0	4.3	5.0
250	0.2	0.6	0.9	1.4	1.9	2.7	3.2
1000	0.1	0.3	0.4	0.7	0.9	1.4	1.6

To come up with a statement on a relevant margin of error a user must imagine a proportion p_A to estimate and an alternative proportion whose difference to p_A he or she considers relevant. What difference actually is relevant depends on the problem the user has in mind and might be difficult to define. Then, in principle, the user would have to specify the desired power of a test at the alternative (see, e.g. Lenth (2001)). However the usual discussion restricts to defining a margin of error which should be smaller than the relevant difference. An alternative at the margin of error will then usually have an approximate power of 0.5.

For the development of the size resolution we first avoid the problem of the estimation of $\sigma(\hat{p}_A)$ by using the tolerance interval instead of a confidence interval. The tolerance interval describes an interval which covers an area of specified size (in probability) of a random variable. In our case we use, in addition, the normal approximation to the (scaled) hypergeometric distribution of $\hat{p}_A$ and arrive at the approximate $100(1-\alpha)$% tolerance interval for $\hat{p}_A$:

$$[p_A - z\sigma(\hat{p}_A),\ p_A + z\sigma(\hat{p}_A)], \tag{2}$$

where $z = \Phi^{-1}(1-\alpha/2)$. In addition we limit the tolerance interval to the allowed range $[0,1]$. The coverage of a tolerance interval is closer to its nominal value than for a normality based confidence interval but it is, of course, still only approximately $100(1-\alpha)\%$.

If a small group should be estimable the probability that no member of the group turns up in the sample should be sufficiently small. In other words we want to control $P[\hat{p}_A = 0]$. Approximately, if we use a $100(1-\alpha)$% tolerance interval, the probability of $P[\hat{p}_A = 0] \leq \alpha/2$. In other words, to control $P[\hat{p}_A = 0]$ the lower limit of the tolerance interval must be larger than 0:

$$p_A - z\,\sigma(\hat{p}_A) > 0. \tag{3}$$

This inequality not only implies that $P[\hat{p}_A = 0] \leq \alpha/2$ (approximately) but in addition that $\sigma(\hat{p}_A) \leq p_A/z$ and therefore imposes a bound on the variability of the estimator $\hat{p}_A$. In other words condition (3) implies that we reach a certain precision for the estimation of $\hat{p}_A$. Thus not only the probability not to observe any member of the group is controlled but the estimation of p_A has a minimal precision. These two ingredients justifies to say that the size of the group A is estimable when (3) is fulfilled. In addition the tolerance interval also implies $P[\hat{p}_A > z\sigma(\hat{p}_A)] \leq \alpha/2$ (approximately) and therefore bounds the probability to observe an unusually large $\hat{p}_A$, too.

Solving the inequality (3) for N_A leads to the expression

$$N_A > N_D \frac{z^2(1-f_D)\frac{N_D}{N_D-1}}{f_D N_D + z^2(1-f_D)\frac{N_D}{N_D-1}}. \tag{4}$$

Since the size N_A itself is simpler to understand and communicate to clients than the proportion p_A we write the above bound for N_A, but dividing by N_D the inequality for p_A is seen readily. Further approximations (assuming $N_D/(N_D-1 \approx 1$ and $1-f_D \approx 1$) and majorisation (dropping the second summand in the denominator) lead to the approximate $100(1-\alpha)$% size resolution

$$\tilde{R}_s = z^2/f_D. \tag{5}$$

Setting $N_D/(N_D-1)$ to 1, a minor approximation, the bound (4) is called the size resolution R_s in Eichenberger et al. (2011). The approximate size resolution $\tilde{R}_s$ is free of the domain size N_D and the particular population proportion p_A to estimate. It is therefore sort of a universal measure for the survey. The approximations and majorisations used are much smaller than the usual replacement of $p_A(1-p_A)$ by the majorising 0.25. However, for small samples the difference between the tight bound (4) and the approximate size resolution is noticeable. Table 2 shows the size resolution R_s and the approximate size resolution $\tilde{R}_s$ for some sample and domain sizes. Note that the dependence of the approximate size resolution is only through the different sampling rates n_D/N_D and not directly through the domain sizes N_D.

Table 2: 95% size resolution and approximate size resolution

n_D	N_D	R_s	$\tilde{R}_s$
5	500	217	385
5	10000	4344	7683
10	500	137	193
10	10000	2774	3842
20	500	78	97
20	10000	1609	1921
50	500	33	39
50	10000	711	769

3. Difference Resolution

The difference resolution is designed to help in the situation where independent subsamples of a simple random sample without replacement are considered in two separate domains D_1 and D_2. Alternatively the same domain may be considered at two time points for which independent samples are drawn. Since the domains may have different sizes a difference of the group sizes N_{A1} and N_{A2} often is less interesting than a difference of the proportions p_{A1} and p_{A2} of the groups in the two domains. For example we may want to compare the proportion of Portuguese speaking inhabitants in a village with a neighbour village of different size. The interesting question is whether the difference between the two proportions is significant. In other words we look at the classical problem of comparing two proportions from independent samples.

Assuming $p_{A2} \geq p_{A1}$ we establish a similar condition as (3) for the difference of proportions $\delta_A = p_{A2} - p_{A1}$

$$\delta_A - z\,\sigma(\hat{\delta}_A) > 0, \tag{6}$$

where $\hat{\delta}_A = \hat{p}_{A2} - \hat{p}_{A1}$ and $\sigma(\hat{\delta}_A) = (\sigma^2(\hat{p}_{A2}) + \sigma^2(\hat{p}_{A1}))^{1/2}$. A solution for δ_A of (6) can be derived (see Eichenberger et al. (2011)) with the parameterization $p = (p_{A1} + p_{A2})/2$. However, the resulting formula depends on p and the two sampling rates and the two domain sizes. Thus it is too complex to be of direct use in communicating with lay persons. Assuming equal domain sizes and sampling fractions and an approximate difference resolution (for equal N_D and f_D) can be derived:

$$\tilde{R}_d = \sqrt{\frac{z^2 N_D}{2 f_D}}. \tag{7}$$

In order to get rid of the dependence on the mean proportion p the majorisation $p(1-p) \leq 0.25$ was used, which is a loose bound for small p. Remember, however, that here the objective is not the bound p away from 0 but to bound δ_A away from 0. In any case the approximate difference resolution still depends on the domain size N_D. Instead of the difference in absolute numbers the proportional difference may be more interesting. Dividing $\tilde{R}_d$ by the domain size N_D we arrive at the relative approximate difference resolution $\tilde{r}_d = z/\sqrt{2 f_D N_D} = z/\sqrt{2 n_D}$. Thus the relative approximate difference resolution $\tilde{r}_d$ depends only on the sample size of the domain.

Table 3 shows that the majorisation for p has a considerable effect when the true $p = 0.2$. At $p = 0.5$ the approximate difference resolution is not much larger than the difference resolution. The main point is that for a sample of a specified size and for a specified domain size the approximate difference resolution indicates the lower limit of the difference of group sizes which is estimable, i.e. the estimated difference will be larger than 0 with high probability and its standard error will be bounded. For example we can be confident that for a domain of size $N_D = 10000$ any difference which is larger than 1960 will be estimable even with a relatively small sample of $n_D = 50$.

Table 3: 95% diff. resolution at $p = 0.2$ and $p = 0.5$ and approx. diff. resolution

n_D	N_D	$R_d(0.2)$	$R_d(0.5)$	$\tilde{R}_d(0.5)$
5	500	*210	263	310
5	10000	*4214	5268	6198
10	500	160	200	220
10	10000	3210	4013	4383
20	500	117	146	155
20	10000	2366	2958	3099
50	500	74	92	98
50	10000	1535	1919	1960

* impossible values since $0.2 \cdot N_D - R_d(0.2)/2 < 0$.

4. Sample Size

The size resolution leads immediately to a necessary overall sampling fraction or the a necessary overall sample size for a simple random sample or a stratified random sample with proportional allocation, which is the case of the Swiss Population Survey. It suffices to set $f = z^2/\tilde{R}_s$ for a given desired approixmate size resolution and the overall sample size needes is

$$n = Nz^2/\tilde{R}_s, \tag{8}$$

where N is the population size. The latter result holds because the approximate resolution does not depend on N_D but on f_D, such that the sample size n_D depends linearly on the domain size. The resulting net sample size n would have to be augmented to take into account non-response. For complex samples a correction for the design effect would have to be applied in addition.

Using the approximate difference resolution to determine the sample size we obtain $n_D = (N_D z/\tilde{R}_d)^2/2$. Since the sample size is not linear in N_D this expression does not scale to the overall population size. The relative approximate difference resolution $\tilde{r}_d$ may be more useful. In other words we may be able to determine a relevant difference of proportions δ_A we would like to be estimable and set $\tilde{r}_d = \delta_A$. Then the needed sample size per domain is $n_D = (z/\tilde{r}_d)^2/2$. However, this expression is not linear in the domain size either and thus does not scale to the whole population. Only for domains that can be defined as the strata of a stratified sample design we can derive an overall sample size from the difference resolutions in the domains. The resulting sample sizes will usually be different from those derived from the size resolution and compromises will be necessary. Therefore, the difference resolution may be rather used to check whether the sample sizes derived from the size resolution yield useful possibilities to compare proportions.

5. Examples

The European Social Survey (European Social Survey, 2010) is also carried out in Switzerland. The Swiss survey of 2008 had a sample of size $n = 1819$ which corresponds to a sampling rate of $f = 0.0003$ approximately. A two stage stratified sample of households and of one randomly selected person per household is used. For the illustration here we assume a design effect of 1.24. The 95% approximate size resolution is $1.96^2/0.0003 = 12806$. Using the design effect 1.24 we arrive at a corrected approximate size resolution of $\tilde{R}'_s = 1.24 \cdot 12806 = 15879$. Therefore we cannot hope to estimate sizes of groups smaller than 15879 but any(!) group larger than this is estimable.

If we would like to estimate the size of groups which are smaller, e.g. $N_A = 5000$, we may use (8) to arrive at an effective sample size of $n = 4610$ and a net sample size adjusted for the design effect of $n = 5717$.

The approximate difference resolution corrected for the design effect is $\tilde{R}'_d = \sqrt{1.24} \cdot \sqrt{N_D} z/\sqrt{2 f_D} = \sqrt{1.24 \cdot N_D} 80 \approx \sqrt{N_D} \cdot 90$. Therefore, it will only be reasonable to estimate differences for domains of size larger than $90^2 = 8100$. This is not astonishing since with the overall sample size of the European Social Survey of 1819 the sample size in a domain of size $N_D = 8100$ is about $n_D = 3$! For two domains of size $N_D = 90000$ the approximate difference resolution will be 27000. Any(!) difference above this threshold will be estimable within these domains. To estimate a difference of proportions of 10% an effective net sample of $n_D = 1.24 \cdot (z/r_d)^2/2 = 1.24 \cdot (1.96/0.1)^2/2 \approx 238$ per domain is needed independently of the size of the domains. For a domain of size $N_D = 90000$ which received an expected sample size of $n_D = 27$ in the European Social Survey of 2008 this would mean a nearly 9 times larger sample.

6. Final remarks

The resolution measures indicate the smallest group size or difference of group sizes which is estimable. The measures are more intuitive than standard errors or confidence intervals. Mainly the size resolution was helpful in the discussion of the precision of the Swiss Population Survey with lay persons. When discussing the precision of samples and the needed sample size the usual discussion turns around the precision for a given sample size. Rarely the sample size for a given precision or resolution is sought. This is natural because for a given sample size there are many different types of questions to be answered and many different resolutions to be considered. Therefore also the difference resolution, which does not lead as easily to an overall sample size as the size resolution, is a valuable measure when the discussion needs to address comparisons of proportions.

References

Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). Wiley.

Eichenberger, P., B. Hulliger, and J. Potterat (2011). Two measures for sample size determination. *Survey Research Methods 5/1*, 27–37.

Lenth, R. (2001). Some practical guidelines for effective sample size determination. *The American Statistician 55/3*, 187–193.

European Social Survey (2010). ESS4 - 2008 documentation report, edition 3.01. Technical report, The ESS Data Archive.