ISSN 1977-0375

eurostat
Methodologies and
Working papers

Survey sampling reference guidelines

Introduction to sample design and estimation techniques

2008 edition

eurostat
EUROPEAN COMMISSION

Luxembourg: Office for Official Publications of the European Communities, 2008

ISBN 978-92-79-07840-8
ISSN 1977-0375
Cat. No. KS-RA-08-003-EN-N

Theme: General and regional statistics
Collection: Methodologies and working papers

The publication was prepared by Professor Risto Lehtonen and Senior Researcher Kari Djerf.
Manuscript finalised in August 2007.

EUROSTAT
L-2920 Luxembourg — Tel. (352) 43 01-1 — website http://ec.europa.eu/eurostat

Eurostat is the Statistical Office of the European Communities. Its mission is to provide the European Union with high-quality statistical information. For that purpose, it gathers and analyses figures from the national statistical offices across Europe and provides comparable and harmonised data for the European Union to use in the definition, implementation and analysis of Community policies. Its statistical products and services are also of great value to Europe's business community, professional organisations, academics, librarians, NGOs, the media and citizens.

Eurostat's publications programme consists of several collections:

- **News releases** provide recent information on the Euro-Indicators and on social, economic, regional, agricultural or environmental topics.
- **Statistical books** are larger A4 publications with statistical data and analysis.
- **Pocketbooks** are free of charge publications aiming to give users a set of basic figures on a specific topic.
- **Statistics in focus** provides updated summaries of the main results of surveys, studies and statistical analysis.
- **Data in focus** present the most recent statistics with methodological notes.
- **Methodologies and working papers** are technical publications for statistical experts working in a particular field.

Eurostat publications can be ordered via the EU Bookshop at http://bookshop.europa.eu.

All publications are also downloadable free of charge in PDF format from the Eurostat website http://ec.europa.eu/eurostat. Furthermore, Eurostat's databases are freely available there, as are tables with the most frequently used and demanded short- and long-term indicators.

Eurostat has set up with the members of the 'European statistical system' (ESS) a network of user support centres which exist in nearly all Member States as well as in some EFTA countries. Their mission is to provide help and guidance to Internet users of European statistical data. Contact details for this support network can be found on Eurostat Internet site.

Contents

1. Introduction

The guidelines concern basic principles and methods of survey sampling. This includes survey planning, survey quality, sampling and estimation, and nonresponse. The approach is non-technical; only necessary technical materials are included. The methods are illustrated with practical examples, and references to statistical software are given when relevant.

Because a comprehensive treatment of the various aspects of survey sampling is not possible in some brief guidelines, we have concentrated on selected topics we believe are of importance for readers. We have aimed at a practical guide intended for experts whose practical experience in survey sampling is limited but who have some background knowledge in basic statistics. For further information on topics covered and extensions, we refer to selected literature.

The guidelines are organized as follows. Chapter 2 discusses survey planning and reporting. A number of basic concepts and definitions are given, also including survey quality. Basic sampling techniques are introduced in Chapter 3. We discuss methods such as simple random sampling, systematic sampling and cluster sampling. The use of auxiliary information plays a key role in modern survey sampling, and methods are discussed such as PPS sampling, stratified sampling and model-assisted methods including ratio and regression estimation. Sample size determination is treated and illustrated. Chapter 4 covers nonresponse and discusses reweighting and imputation methods. A brief summary of software available for survey sampling and analysis is included in Chapter 5. We have included a comprehensive list of references on current survey sampling literature in Chapter 6. Chapter 7 includes a list of selected links to web materials relevant to the area.

2. Survey planning and reporting

2.1. Basic concepts and definitions

Definition of a survey

A *survey* refers to any form of data collection. A *sample survey* is more restricted in scope: the data collection is based on a sample, a subset of total population - i.e. not total count of target population which is called a *census*. However, in sample surveys some sub-populations may be investigated completely while the most sub-populations are subject to selected samples. In the subsequent chapters the term *survey* is devoted to sample surveys.

Descriptive surveys versus analytical surveys

Descriptive surveys, including censuses, are typical in statistical offices. They tend present information on parameters like totals, averages or proportions at the total population level or some well-defined sub-populations. In surveys where the emphasis is on *analysis*, the interest is focused on connections and interdependences between phenomena. The parameters of interest are connected with statistical models, such as linear models, and are represented by correlation or regression coefficients. However, it is important for both types of surveys to estimate the unknown parameters as reliably as possible.

Social surveys vs. Business surveys

In social surveys the focus is related with persons and households: e.g. population statistics, labour force participation, wages and salaries, household consumption, poverty and income distribution, education, cultural activities, health and other interested topics.

In business surveys the focus is related with enterprises, establishments and/or other business units like the local kind of activity units, including farms. The interest may vary from production composition and amount to investment plans, employment, use of energy, output waste etc.

Social surveys and business surveys differ from each other also in other aspects. In official statistics business surveys are often mandatory while social surveys tend to be voluntary; the data collection modes are more versatile in social surveys; even the sampling designs can be different.

2.2. Overall survey design

In recent years many textbooks have been published on survey methodology. Groves et al. (2004) provide a good overview on the whole process from the design to the analysis and interpretation. In addition there is a number of specific literature on various data collection modes, testing questionnaires and questions, interviewing strategies etc.

Operational phases of a survey are described e.g. by Sundgren (1999). It includes various tasks from the definition of the main objectives, data collection strategy, processing of data, production of results, evaluation of quality till archiving. All tasks are important to guarantee the various uses of data and their quality. The readers are recommended to obtain more information from appropriate literature like Lyberg et al. (1997), or Biemer & Lyberg (2003).

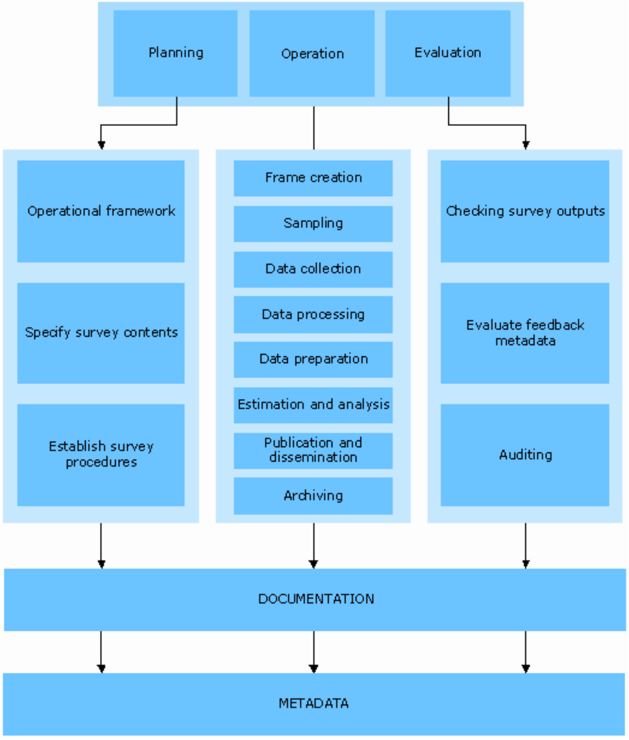


Figure 1. Flow chart of survey process (see e.g. Statistics Finland)

2.3. Reporting of survey quality

The users should be reported with appropriate information on survey quality, preferably from all stages of the survey process. It has been a tradition to report of survey quality by distinguishing various sources of error which may occur during the many stages of survey operations. For example, Biemer & Lyberg (2003) describe following types of errors: Specification, Frame, Nonresponse, Measurement, Processing, and Sampling error. Some may be born randomly but unfortunately various sources tend to introduce systematic errors.

Sampling errors

Standard errors for the estimable parameters, often point estimates are the oldest quality measures. They (and other estimates derived from those like coefficients of variation or confidence intervals) were introduced during the rise of survey methodology in 1940s.

Measurement errors

Besides the sampling errors the other types of errors were introduced quite early. The first UN recommendations on reporting survey quality were given already in 1950s and the measurement errors were already included. However, the implementation of systematic reporting took much longer.

Total survey error

The total survey error of a parameter θ is measured by the mean square error (MSE), i.e. sum of the variance and squared bias: $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + Bias^2(\hat{\theta})$.

Sampling variance is derived from the sampling design, the other components affecting its estimate are sample size, the variability of the parameter of interest and sampling weights. Sampling error, i.e. square root of sampling variance, is a random error by definition. Bias is the difference between the true value and the expectation of the estimator, and when nonzero it represents systematic error. Unfortunately the MSE estimation requires repeated sampling and thus cannot easily be carried out with large-scale surveys. Some subtle methods have, however, been suggested to evaluate the total error (see e.g. Lessler & Kalsbeek 1992).

The quality dimensions and standards of the European Statistical System provide a good frame to report on quality. The quality dimensions are Relevance, Accuracy, Timeliness and Punctuality, Comparability, Coherence, and Accessibility and Clarity. Relevance describes how the statistical survey meets the user needs and requirements. Accuracy contains the traditional measures on survey quality (like standards errors, confidence intervals and coefficients of variation etc.). Timeliness and punctuality measure the freshness of data and the results. Comparability and coherence are related with various forms of comparisons: different sources describing the same phenomenon, comparability of the same survey over various domains, like geographical areas, comparability over time etc. Finally, Accessibility and clarity describe the various form data are available and results disseminated, metadata and other user support etc.

Furthermore, a list of quality indicators have been constructed to make the follow-up easier for those surveys which are repeated more or less regularly.

The Eurostat Quality website presents all relevant documents on quality reporting and also some current practices and guidelines on the issue: http://ec.europa.eu/eurostat/quality.

The International Monetary Fund (IMF) and Organization for Economic Co-operation and Development (OECD) have created own standards which are also widely used especially in the field of economic statistics (see International Monetary Fund 2003).

2.4. Sampling frame issues

Population and frame

Target population is the population we theoretically are interested in. It is assumed to be fixed (and finite).

Frame population is the population we can obtain.

Survey population is the intersection of those above.

Those three populations do not quite coincide because the frame population tends to contain some erroneous elements called coverage errors. Below we present some typical reasons for coverage errors:

- time lags between the moment the sample frame was created and it was actually used
- failure to include new births in the frame
- failure to include or exclude elements which have moved (physical removals, enterprises which have changed their industry etc.)
- failure to remove deaths and similar out-of-scope elements

Overcoverage means that our sampling frame contains elements which do not belong to our target population. Overcoverage can normally be detected during the field-work.

Undercoverage is a much more problematic phenomenon since often it cannot be detected and assessed in a reliable manner.

There may be no realistic way to include all possible differences between the target population and the ultimate sampling frame but those known should be included. Kish (1965) advocated a stratum of surprises to include those cases.

Sometimes no good frame exists for the target population and one has to find other solutions described below.

Multiple frames

Multiple frames may occur if the target population can be compiled from several independent sources. Use of many frames is not uncommon in developing countries but can also used in developed societies when new phenomena are investigated.

Clustered frames

It may well happen that there is not a good population frame for the ultimate sampling units, or that the creation of such would be much too expensive. Then the next solution is to seek for an alternative from the combinations of the elements, i.e. seek for clustered frames.

Consider, for example, a study of school children: even if a population frame would be available covering all children attending the schools, the field work will become much cheaper if the schools and/or classes are selected instead of the pupils randomly over the whole population.

In large population and household surveys we most often deal with clusters which comprise to some natural combination of elements, e.g. people living in enumeration districts or administrative regions.

Other issues

Double listings of the same elements should always be removed from the frame if found.

Small sub-populations may sometimes be quite impossible to reach although they are known, e.g. people living in remote mountainous villages. For cost and other reasons they may be removed from the sampling frame. Then a difficult question arises: do the estimates from the reduced population reflect the properties of those from removed sub-populations?

Cut-off samples are another example related with the same problem. Normally cut-off samples are applied in business surveys where the smallest units do not contribute too much to the parameter of interest. However, since one part of the target population is deliberately excluded there is a chance to obtain bias in estimation.

Auxiliary information

Information obtained from “background” variables to be used either at the sampling stage (e.g. to create strata or clusters, calculate measure of size etc.) or after data collection to calculate weights etc. Sometimes auxiliary data cannot be obtained from the sampling frame but can be available after the survey from other sources, such as official statistics.

3. Techniques for sample selection and estimation

3.1. Preliminaries

In a sample survey, a probability sample is drawn from the frame population by using a specified sampling design. Typically, the sampling design consists of a combination of various sample selection techniques. A complex sampling design can involve clustering and stratification and several stages of sampling. In simple cases, sampling of elements is carried out directly from the sampling frame. In all cases, some of the well-documented sample selection techniques are used in the sampling procedure. Good examples of relevant literature on sampling techniques are Kish (1965), Cochran (1977); Lohr (1999), and Lehtonen & Pahkinen (2004), which is the primary source for this section. Helpful supplemental materials on survey sampling and estimation, including computational examples using real survey data, can be found in VLISS-virtual laboratory in survey sampling, representing a web extension of the Lehtonen and Pahkinen textbook. The application can be accessed freely at http://www.math.helsinki.fi/VLISS/. Many of the common sample selection techniques can be readily implemented by statistical software products, such as the SAS procedure SURVEYSELECT.

The properties of sampling techniques vary with respect to statistical efficiency and certain practical aspects, such as suitability to a given sampling task, requirements for application and user friendliness. Often the study design and time and budget constraints affect the choice of the sampling design in a given survey setting. An important additional aspect is the role of auxiliary information in a given sampling procedure. Let us first discuss the standard sample selection schemes from this point of view.

Use of auxiliary information in sampling and estimation

It is often useful to incorporate auxiliary information on the population in a sampling procedure. In practice, there are different ways to obtain auxiliary information. For example, in the so-called register countries (e.g. Scandinavian countries), sampling frames used in official statistics production often include auxiliary information on the population elements, or these data are extracted from administrative registers and are merged with the sampling frame elements at the micro level. In other cases, aggregate-level auxiliary information can be obtained from different sources such as published official statistics. Use of auxiliary information in sampling and estimation is an expanding feature in official statistics production. Auxiliary information can be useful in the construction of an efficient sampling design and further, at the estimation stage for improved efficiency for the actual sample. To be useful, auxiliary information should be related to the variation of the study variables.

In *simple random sampling* (SRS), the sample is drawn without using auxiliary information on the population. Therefore, SRS provides a reference scheme when assessing the gain from the use of auxiliary information in more complex designs or in improving the efficiency of estimation for a given sample.

Auxiliary information does not play a role in standard application of *systematic sampling* (SYS). Thus, the efficiency of SYS tends to be similar than that of SRS. This also holds if population elements in the sampling frame are in random sort order with respect to the study variable. In a method called *implicit stratification*, auxiliary information can be used in the form of the list order of elements in the frame. Now, SYS can be more efficient than SRS if there is a certain relationship between the ordering of elements in the sampling frame and the values of the study variable.

Sampling with probability proportional to size (PPS) is a method where auxiliary information has a key role. An auxiliary variable is assumed to be available as a measure of the size of a population element. Varying inclusion probabilities for population elements can be assigned using the size variable. Efficiency improves relative to SRS if the relationship between the study variable and the size variable is strong. PPS is often used in business surveys and in general, for situations where the sampling units vary with a size measure.

Stratified sampling (STR) relies strongly on the use of auxiliary information. In STR, the frame population is first divided into non-overlapping subpopulations called *strata*, and sampling is executed independently within each stratum. If the strata are internally homogeneous with respect to the study variable, i.e. if the within-stratum variation of the study variable is small and a large share of the total variation is captured by the variation between the strata, then STR can be more efficient than SRS.

In *cluster sampling* (CLU), the population is assumed to be readily divided into naturally formed subgroups called *clusters*. A sample of clusters is first drawn from the population of clusters. In the next stage, all elements of the sampled clusters are taken in the element sample (*one-stage cluster sampling*), or a sample of elements is drawn from each sample cluster (*two-stage cluster sampling*). If the clusters are internally homogeneous, which is usually the case, then CLU is less efficient than SRS. This *clustering effect* can be reduced by stratifying the population of clusters, tending to improve efficiency.

The sampling techniques introduced above can be used to construct a manageable sampling design for a sample survey, either using a particular method or more usually a combination of methods. In all methods excluding SRS, auxiliary information in the form of auxiliary variables can be incorporated in the sampling procedure. Note that the use of auxiliary information in SRS, SYS and stratified sampling requires that the values of auxiliary variables must be available for every population element. Auxiliary information in cluster sampling concerns at least the grouping of the population elements into clusters. If additional auxiliary data are available on the population of clusters, these data can be used for example for stratification or PPS sampling purposes.

Use of auxiliary information in the sampling phase is typical in *descriptive surveys* where the number of study variables is small. Efficiency gains can be obtained if the association between the study variable(s) and the auxiliary variables is strong.

Auxiliary information can be used for the selected sample in the *estimation phase*. Use of auxiliary information in the estimation phase involves flexibility: the sample design can be kept simple and in the estimation phase, the use of auxiliary information can be tailored for diverse study variables. In addition, requirements for auxiliary data in standard methods are weaker than in the previous case, because unit-level auxiliary data only are needed for the sampled elements, and the auxiliary data can be incorporated at an aggregate level in the estimation procedure. Some of the standard methods are ratio estimation, regression estimation and post-stratification. All these methods use statistical models as assisting or working models when incorporating the auxiliary data in the estimation procedure. The methods thus are called *model-assisted*.

In *ratio* and *regression estimation,* the population total of a continuous auxiliary variable is assumed known. The assisting model is of regression-type linear model. In ratio estimation, the model is without an intercept term, i.e. the intercept is assumed zero. Efficiency can improve if the study variable and the auxiliary variable are correlated. But the method can be ineffective if there is a nonzero intercept term in the true model. In regression estimation, the assisting model is again of regression-type, but now with an intercept term. Efficiency can improve if the study variable and the auxiliary variable are correlated.

Post-stratification resembles stratified sampling, but the stratification is carried out after the sample selection. The selected sample is divided into non-overlapping subgroups called *post-strata* according to a categorical or classified auxiliary variable (or several such variables), and the estimation follows that of stratified sampling. Similarly as in stratified sampling, efficiency can improve if the post-strata are internally homogeneous with respect to the study variable. Post-stratification is often used for adjusting for unit nonresponse (see Section 4.1).

Thus, auxiliary information on the population can be used in the construction of the sampling design and, for a given sample, to improve the efficiency in the estimation phase. As a rule, efficiency of estimation can improve by the proper use of auxiliary information.

Parameters, estimators and quality measures

Let our *parameter of interest* be a fundamental parameter in survey sampling, the *population total* $T = \sum_{k=1}^{N} y_k$ of study variable *y*. In the formula for the total, y_k are the (unknown) values of the study variable and *N* is the number of elements in the population. Many parameters routinely used in survey sampling, such as means, proportions, ratios and regression coefficients, can be expressed as functions of totals. To have an *estimate* for the unknown population total *T*, a sample is drawn from the population and the sample values of the study variable are measured. An *estimator* of the population total *T* is denoted by $\hat{t}$. The concept estimator refers to a calculation formula or algorithm that is used for the sample to obtain a numerical value for the estimate. A simple example is the sample mean $\bar{y} = \sum_{k=1}^{n} y_k / n$, which is calculated using the *n* sample measurements. Using the sample mean, an estimate for the population total is calculated as $\hat{t} = N \times \bar{y}$. These derivations hold for simple sampling designs; more complex derivations are needed for complex sampling designs.

In survey sampling, estimators are preferred that fulfil certain theoretical properties. These are *unbiasedness*, meaning that the expectation of an estimator coincides with the target parameter, i.e. $E(\hat{t}) = T$, and the bias is defined as $Bias(\hat{t}) = E(\hat{t}) - T$. *Consistency* is a somewhat weaker property, referring to the behaviour of an estimator to better match with the value of the target parameter when sample size *n* increases, and to reproduce the target parameter when the sample size coincides *N*, the population size. *Precision* of an estimator refers to its variability and is measured by the *design variance* $Var(\hat{t})$. The smaller is the design variance, the better is the precision. A precise estimator is called *efficient*. And *accuracy* of an estimator refers to combined bias and precision properties of an estimator and is measured by the mean square error: $MSE(\hat{t}) = Var(\hat{t}) + Bias^2(\hat{t})$.

In survey sampling practice, estimators are used that are unbiased or at least consistent. A challenge for survey statistician is for a given sampling task to obtain efficient estimators whose design variances are as small as possible. This is for high reliability of the results calculated by using the collected sample survey data.

The *standard error* (s.e), *coefficient of variation* (c.v) and *design effect* (deff) of an estimator are commonly used quality measures of estimators. The quality measures are derived from the theoretical properties introduced above. For an estimator $\hat{t}$ of population total, the measures are defined as follows.

Estimated standard error: $\text{s.e}(\hat{t}) = \sqrt{\hat{v}(\hat{t})}$, where $\hat{v}(\hat{t})$ is the *estimated design variance* or *sampling variance* of the total estimate $\hat{t}$.

Estimated coefficient of variation or *relative standard error*: $\text{c.v}(\hat{t}) = s.e(\hat{t}) / \hat{t}$, i.e. the estimated standard error divided by the estimate itself. Coefficient of variation is often expressed in percentages, $100 \times \text{c.v}\%$. Coefficient of variation is routinely reported in official statistics. C.v is often used as a quality standard in the context of the ESS (see Section 3.3).

Design effect (deff) (Kish 1965) measures the *statistical efficiency* of the sampling design with respect to simple random sampling (SRS) and is given by

$$\text{deff}\,(\hat{t}) = \frac{\hat{v}(\hat{t})}{\hat{v}_{SRS}(\hat{t})},$$

where the numerator is the sampling variance of the total estimator under the actual (possibly complex) sampling design and the denominator represents the sampling variance under an assumption of simple random sampling of a sample of similar size. Using the design effect, *effective sample size* is determined as $n_{\text{eff}} = n / \text{deff}(\hat{t})$, that is, the actual sample size *n* divided by the design effect of the total estimate.

The formula for deff gives rise to the following remarks:

(a) deff < 1 The actual sampling design is *more effective* than SRS. Correspondingly, effective sample size is larger than the actual sample size.

(b) deff = 1 The efficiency of the actual sampling design is similar to that of SRS.

(c) deff > 1 The actual sampling design is *less effective* than SRS. In this case, effective sample size is smaller than the actual sample size.

In survey sampling practice, a natural goal is the case (a). In this effort, the use of the available auxiliary information in the sampling design is beneficial. Stratified sampling and PPS sampling are often used for this purpose. In addition, efficiency can be improved in the estimation phase by incorporating auxiliary data in the estimation procedure via model-assisted techniques. In cluster sampling, the case (c) is often encountered because of the internal homogeneity of the clusters with respect to the variables of interest.

3.2. Basic sampling techniques

Basic sampling techniques include *simple random sampling*, *systematic sampling* and *sampling with probabilities proportional to size* (PPS). These methods are used in sampling designs as the final methods for selecting the elementary or *primary sampling units* (PSU:s) and for working out randomization. A manageable sampling design for a survey often involves stratification, clustering and multiple stages of sampling. *Stratification* of the population into non-overlapping subpopulations is a popular technique where auxiliary information can be used to improve efficiency. In *cluster sampling*, the practical aspects of sampling and data collection are the main motivation for the use of auxiliary information in the sampling design.

3.2.1. Simple random sampling

Simple random sampling (SRS) is often regarded as the basic form of probability sampling. SRS is applicable to situations where there is no previous information available on the population structure. Simple random sampling directly from the frame population ensures that each population element has an equal probability of selection. Thus, SRS is an *equal-probability sampling design*.

As a basic sampling technique, simple random sampling can be included as an inherent part of a sampling design. In addition, simple random sampling sets a baseline for comparing the relative efficiency of a sampling design by using the design effect statistic introduced above.

In simple random sampling of *n* elements, every element *k* in the population frame of *N* elements has exactly the same inclusion probability, that is, $\pi_k = \pi = n / N$. Recall that inclusion probability is the probability of a population element to be included in a *n* element sample. An inclusion probability is assigned for every population element before carrying out the sampling procedures. Inclusion probabilities depend on the sampling design and are by definition greater than zero for all population elements.

In practice, SRS can be performed either without replacement (SRS-WOR) or with replacement (SRS-WR). WOR type sampling refers to the case where a sampled element is not replaced in the population; this also means that a population element can be sampled only once. In a WR scheme, a sampled element is replaced in the population. In both cases, the inclusion probability $\pi = n/N$ remains, and the only difference is in the variance formula of the statistic of interest. As a general rule, WOR-type SRS is more efficient that WR-type SRS, that is, the variance in SRS-WOR tends to be smaller than that in a SRS-WR counterpart. This property also holds for the other sampling designs and explains the frequent use of without replacement type designs in survey sampling practice.

Under SRS, an estimator of the target parameter *T* can be written simply as

$$\hat{t} = N\sum\nolimits_{k=1}^{n} y_k / n = N\overline{y} \,,$$

where $\overline{y} = \sum\nolimits_{k=1}^{n} y_k / n$ is the sample mean. Alternatively, by using the SRS inclusion probabilities π, the estimator can be expressed in the form

$$\hat{t} = \sum\nolimits_{k=1}^{n} y_k / \pi = \sum\nolimits_{k=1}^{n} y_k /(n/N) = \sum\nolimits_{k=1}^{n} w_k y_k \,,$$

where $w_k = N/n$ is the *sampling weight*, i.e. the inverse inclusion probability. Note that in SRS, the sampling weights are equal for all sample elements. In more complex designs to be addressed, the sampling weights can vary between elements (as in PPS sampling) or groups of elements (as in stratified sampling).

Using the estimated total, the population average or mean $\overline{Y} = \sum\nolimits_{k=1}^{N} y_k / N$ can be estimated by $\overline{y} = \hat{t} / N$. Note that we assumed here a known population size *N*, which is a realistic assumption in practice. But if *N* is unknown at the estimation stage, an estimator $\hat{N} = \sum\nolimits_{k=1}^{n} w_k$ can be used for the population size.

For an estimator $\hat{t}$ of population total under SRS-WOR, the sampling variance of $\hat{t}$ is given by

$$\hat{v}(\hat{t}) = N^2 (1 - n/N)(1/n)\hat{s}^2 \,,$$

where $\hat{s}^2 = \sum\nolimits_{k=1}^{n} (y_k - \overline{y})^2 /(n-1)$ is the sample variance of the study variable *y*. The quantity $(1 - n/N)$ in the sampling variance formula is called the *finite population correction* (fpc). Note that if the sampling fraction n/N is small, as is the case in typical sampling designs for persons or households, practical importance of the fpc is minor, because fpc is close to one. But this is not necessarily so in sampling designs for business surveys where sampling fractions can be much larger.

For SRS-WR, the only difference in the sampling variance $\hat{v}(\hat{t})$ is that the fpc is given by $(1 - 1/N)$. This difference also indicates better efficiency for the SRS-WOR design: the design effect of $\hat{t}$ under SRS-WR is $\text{deff}(\hat{t}) = (1 - 1/N)/(1 - n/N) > 1$, assuming that sample size *n* is larger than one and smaller than population size *N*. Note that we used SRS-WOR as the reference SRS design in the deff formula; this is a natural choice but sometimes, SRS-WR is put in this role in certain statistical software.

To summarize, if the sampling fraction (*n/N*) is small the fpc for SRS-WOR will be close to 1. And vice versa: if the sample size *n* approaches the population size *N*, the variance estimate $\hat{v}(\hat{t})$ will reduce. Thus, in a census the sampling variance is zero.

In practice SRS is executed with an appropriate piece of software. For example, the SAS procedure SURVEYSELECT can be used for both SRS-WR and SRS-WOR. In real life sampling with SRS we mostly deal with the without-replacement type SRS design.

Example. *Bernoulli sampling* provides an example of an SRS-WOR type sampling scheme. In this method, the sample size is not fixed in advance but is a random variate whose expectation is *n*, the desired sample size. This property leads to a variation in the sample size with the expected value $N\pi$ and variance $N(1 - \pi)\pi$, where π stands for the inclusion probability. The randomness in the sample size is relatively unimportant in large samples.

Let us briefly introduce the technique. To carry out Bernoulli sampling, we need to carry out the following steps:

Step 1. Fix the value of the inclusion probability π, where $0 < \pi < 1$, so that the expected sample size will be $N\pi$, the product of the population size and the inclusion probability. If the desired sample size is *n*, then $\pi = n/N$.

Step 2. Append three variables, let say PROB, IND and UNI, to the sampling frame data set. PROB is set equal to the chosen value of π, and IND is set to zero, for all *N* population elements. For UNI, a value from a uniform distribution over the range (0, 1) is drawn independently for each population element, starting from the first element. A pseudo random number generator can be used in generating the random numbers.

Step 3. The decision rule for inclusion of a population element in the sample is the following. The *k*th population element is included in the sample if UNI < π, and correspondingly, we set IND = 1 for the selected element (otherwise, the value of IND remains zero).

Step 4. Treat all population elements sequentially by using Step 3.

When Steps 1 to 4 are completed, the sum of IND over the sampling frame appears to be close (or, equal) to the desired sample size *n*. The elements having IND = 1 constitute the Bernoulli sample. The procedure can be easily programmed for example with Excel, SAS or SPSS. Appendix 1. contains a short example of Bernoulli sampling.

3.2.2. Systematic sampling

Systematic sampling (SYS) is a widely used sampling technique in situations where the sampling frame is an ordinary electronic (or manual) data base, such as a population register, a register of business firms or farms, or a list of schools. SYS also is an equal probability sampling design because the inclusion probability of a population element in an *n* element sample is $\pi = n / N$.

Steps in the selection of a systematic sample of *n* elements from a population of *N* elements are the following:

1. Define the sampling interval $q = N/n$, where an integer *q* is assumed.

2. Select a random integer a with an equal probability of $1/q$ between 1 and q (a pseudo random number generator for uniform distribution over the range (1, q) of e.g. Excel, SAS, SPSS can be used).

3. Select elements numbered a, $a + q$, $a + 2q$, $a + 3q$,..., $a + (n-1)q$ in the sample.

Thus, with an integer q, SYS results in an n element sample. If q is not an integer, all sampling intervals can be defined as of equal length except one.

In practice, there are several ways of selecting a systematic sample. The one we introduced above represents an example of SYS sampling with one random start. Alternatively, two, or more generally m, independent systematic samples can be taken using the procedure above. The size of each SYS sample is then n/m elements and the length of the sampling interval is $m \times q$. This technique is suitable if variance estimation is to be carried out using so-called *replication techniques* (see Wolter 2007).

Further, a systematic sample can be drawn by treating the elements in the sampling frame as a closed loop. Beginning from the randomly selected integer A from [1, N], the selection proceeds successively by drawing elements $A + q$, $A + 2q$, … , till the end of the frame, and then the selection continues from the beginning of the frame. The loop will be closed when n elements have been drawn. These random start methods lead to the selection of a SYS sample of n elements, and the techniques are equivalent with respect to the estimation.

In statistical software products, such as the SAS procedure SURVEYSELECT, there are advanced sampling algorithms for SYS that use fractional intervals to provide exactly the specified sample size n.

For SYS, there is no known analytical variance estimator for the design variance, even for such a simple estimator as the total. Therefore, approximate variance estimators are used in practice (see e.g. Wolter 2007; Lehtonen and Pahkinen 2004, Section 2.4).

Estimation under systematic sampling depends on the knowledge on the sorting order of the sampling frame:

1. If the sorting order of the sampling frame can be assumed random with respect to the study variables and all auxiliary variables, estimation with SYS will correspond to that of SRSWOR. Thus, formulas derived for SRS can be used.

2. If the sampling frame is sorted by an auxiliary variable (or, several such variables), SYS sampling will produce a sample which tends to mirror correctly the structure of population with respect to the variables used in sorting. Sorting the frame before SYS sampling is called *implicit stratification.* For example, in some cases it is a good idea to sort the frame according to the regional population structure. Then a systematic sample will retain the appropriate population distribution across regions. Additional cases are those where the population is already stratified or a trend exists that follows the population ordering, or there is a periodic trend (all these situations can also be reached by appropriate sorting procedures). Periodicity may be harmful in some cases, especially if harmonic variation coincides with the sampling interval. The estimation under implicit stratification corresponds to the estimation under stratified sampling.

Systematic sampling, including implicit stratification, can be carried out for example with the SAS procedure SURVEYSELECT.

Example. Let us consider SYS sampling of *n* = 200 elements from a population of *N* = 2000 elements. The sampling interval is *q* = *N*/*n* = 2000/200 = 10. We next draw a random integer *a* between 1 and 10, let *a* = 7. The SYS sample of *n* = 200 elements consists of population elements numbered 7, 17, 27,…,1997. The inclusion probability for every population element is $\pi_k = \pi = n/N = 200/2000 = 0.1$ and the constant sampling weight for the sampled elements is $w_k = w = 10$.

3.2.3. Sampling with probability proportional to size

In *sampling with probability proportional to size* (PPS), the inclusion probability depends on the size of the population element. Reduction in variance can then be expected if the size measure and the study variable are closely related. It is assumed that the value Z_k of the auxiliary size variable *z* is known for every population element *k*. Typical size measures are variables that physically measure the size of a population element. In business surveys, for example, the number of employees in a business firm can be used as a measure of size, and in a school survey the total number of pupils in a school is also a good size measure. PPS sampling can be very efficient, especially for the estimation of the total, if a good size measure is available.

In PPS sampling, the inclusion probability of an element in a *n* element sample is $\pi_k = np_k = nZ_k / T_z$, where $T_z = \sum_{k=1}^{N} Z_k$ is the sum of size measures over the *N* element population and p_k is called the *single-draw selection probability*. In PPS, the inclusion probabilities π_k vary between elements and thus, PPS is an unequal probability sampling design.

A PPS sample can be drawn either without or with replacement. Calculation of the inclusion probabilities is easier to manage under WR type sampling, because the population remains unchanged after each draw. In PPS-WOR, the population changes after each draw and the inclusion probabilities must be re-calculated for the remaining elements.

The basic principles of estimation under PPS sampling are introduced here only briefly. Under PPS-WOR, an unbiased estimator of the population total T is given by

$$\hat{t} = \sum_{k=1}^{n} w_k y_k = \sum_{k=1}^{n} y_k / \pi_k ,$$

where $w_k = 1/\pi_k$ is the sampling weight. The estimator is called the *Horvitz-Thompson (HT) estimator* or *expansion estimator*. The HT estimator is design unbiased and is very popular in practice. An estimator of the variance of the estimated total is

$$\hat{v}(\hat{t}) = \sum_{k=1}^{n} \sum_{l=1}^{n} (w_k w_l - w_{kl}) y_k y_l ,$$

where $w_{kl} = 1/\pi_{kl}$. The variance estimator of the HT estimator contains the second-order inclusion probabilities π_{kl} (i.e. probabilities to include both elements *k* and *l* in the sample), whose computation is often impractical, especially for large samples. Therefore, approximations are often used in practice. One alternative is

$$\hat{v}(\hat{t}) = N^2 (1/n) \sum_{k=1}^{n} (y_k / (Np_k) - \bar{y})^2 / (n-1) ,$$

which corresponds to a with-replacement PPS scheme where the second-order inclusion probabilities are zero, because the draws are mutually independent.

There are different versions of PPS sampling schemes available for practical purposes. Examples are the *cumulative total method* with replacement or without replacement, *systematic PPS sampling with unequal probabilities* and *Poisson sampling*. For example, Poisson sampling as a without-replacement type design resembles Bernoulli sampling where the sample size is a random quantity; the difference is in the calculation of the inclusion probabilities. Despite of the property of a random sample size, Poisson sampling is sometimes considered attractive because the second-order inclusion probabilities reduce to $\pi_{kl} = \pi_k \pi_l$ which simplifies the calculation of the sampling variance. The book by Brewer & Hanif (1983) provides a good source for the various PPS methods. The most commonly used PPS techniques are implemented in the SAS procedure SURVEYSELECT.

3.2.4. Stratified sampling and allocation techniques

In *stratified sampling* (STR) the target population is divided into non-overlapping subpopulations called *strata*. These are regarded as separate populations in which sampling of elements can be performed independently. Within the strata, some of the basic sampling techniques, SRS, SYS or PPS, are used for drawing the sample of elements. Stratification involves flexibility because it enables the application of different sampling techniques for each stratum.

In general, there are several reasons for the popularity of stratified sampling:

1. For administrative reasons, many frame populations are readily divided into natural subpopulations that can be used in stratification. For example, strata are identified if a country is divided into regional administrative areas that are non-overlapping.

2. Stratification allows for flexible stratum-wise use of auxiliary information for both sampling and estimation. For example, PPS technique can be used in sampling within the stratum, and ratio or regression estimation can be used for the selected sample, depending on the availability of additional auxiliary information in the stratum.

3. Stratification can involve improved efficiency if each stratum is homogeneous with respect to the variation of the study variables. Hence, the within-stratum variation will be small, which is beneficial for efficiency.

4. Stratification can guarantee representation of small subpopulations or domains in the sample if desired. This means that inclusion probabilities can vary between strata. The variation is controlled by the so-called *allocation techniques*.

In stratified sampling, the population is divided into *H* non-overlapping subpopulations of size $N_1, N_2, \ldots, N_h, \ldots, N_H$ elements such that their sum is equal to *N*. For stratification, auxiliary information is required in the sampling frame. Regional, demographic and socioeconomic variables are typical stratifying variables. A sample is selected independently from each stratum, where the stratum sample sizes are $n_1, n_2, \ldots, n_h, \ldots, n_H$ elements, and their sum is equal to *n*, the overall sample size.

There are alternative strategies to determine stratum sample sizes for a given survey. In some cases, the overall sample size *n* is first fixed and then allocated to the strata. This is typical in cases where the strata themselves are not of interest (i.e. producing statistics for the separate strata is not the primary aim). If the survey involves statistics production for each stratum (e.g. a regional area or industrial group), then it is important to ascertain large enough stratum sample sizes. In this case, the stratum sample sizes n_h are first determined (see Section 3.3).

The most common allocation techniques for defining the stratum sample sizes are *proportional allocation*, *equal allocation*, *optimal* or *Neyman allocation* and *power* or *Bankier allocation.* To give an idea of allocation, let us introduce briefly the three first mentioned methods (Bankier allocation requires more detailed additional information on the population distribution within strata, see for example Lehtonen & Pahkinen 2004, Section 3.1).

Proportional allocation is the simplest allocation scheme and is widely used in practice. It presupposes knowledge of the stratum sizes, since the sampling fraction n_h / N_h is constant for each stratum. The number of sample elements n_h in stratum *h* is given by $n_h = n \times W_h$, where $W_h = N_h / N$ is the stratum weight, and *n* is the specified overall sample size. Proportional allocation guarantees an equal share of the sample in all the strata and involves an equal probability sampling design where the inclusion probability $\pi_{hk} = \pi = n / N$ of population element *k* in stratum *h* is constant. Thus, the sampling weight also is a constant $w_{hk} = w = N / n$, and the design is called *self-weighted*.

Equal allocation provides an equal sample size $n_h = n / H$ for each stratum, where *H* is the number of strata. If the stratum sizes N_h vary, inclusion probabilities also vary and are given by $\pi_{hk} = n_h / N_h = n /(H \times N_h)$ for element *k* in stratum *h*. Thus, sampling weights are $w_{hk} = H \times N_h / n$. If all stratum sizes N_h are equal, then $\pi_{hk} = \pi = n / N$ and an equal-probability design is obtained.

Optimal or Neyman allocation is usable if the population standard deviations S_h for individual strata of the study variable *y* are known or a reliable figure is available. In practice, close approximations to the true standard deviations may be made from experience gained in past surveys. Thus, Neyman allocation is often used in continuous business surveys. The stratum sample sizes are first calculated. The number of sample units n_h in stratum *h* under optimal allocation is calculated as

$$n_h = n \frac{N_h S_h}{\sum_{h=1}^{H} N_h S_h} .$$

The overall sample size *n* is then the sum of stratum sample sizes. In optimal allocation, a stratum which is large or has a large within-stratum variance has more sampling units than a smaller or more internally homogeneous stratum.

The three allocation schemes are illustrated in an example below. Allocation under STR sampling is further illustrated, with additional computational examples, in the VLISS application, the web extension of Lehtonen & Pahkinen (2004).

In stratified sampling, an estimator $\hat{t}$ of population total T_y is the sum of stratum total estimators, given by $\hat{t} = \sum_{h=1}^{H} \hat{t}_h$, where $\hat{t}_h = \sum_{k=1}^{n_h} y_{hk} / \pi_{hk} = \sum_{k=1}^{n_h} w_{hk} y_{hk}$ is the Horvitz-Thompson estimator of the stratum total T_h. Because the samples are drawn independently from each stratum, the sampling variance of $\hat{t}$ is the sum of within-stratum variances $\hat{v}(\hat{t}_h)$, that is, $\hat{v}(\hat{t}) = \sum_{h=1}^{H} \hat{v}(\hat{t}_h)$. Because in STR sampling, the sampling variance only depends on the within-stratum variances, it is a good idea to try to construct internally homogeneous strata with respect to the study variable *y*.

For example, assuming SRS-WOR in each stratum, the total estimator is given by $\hat{t} = \sum_{h=1}^{H} N_h / n_h \sum_{k=1}^{n_h} y_{kh}$, where N_h / n_h is the stratum-specific sampling weight. With proportional allocation this simplifies as $\hat{t} = \sum_{h=1}^{H} N_h / n_h \sum_{k=1}^{n_h} y_{kh} = N / n \sum_{h=1}^{H} \sum_{k=1}^{n_h} y_{kh}$, because the weights N_h / n_h are equal to constant N / n. This reflects the self-weighting property of proportional allocation.

Estimation under stratified sampling is discussed in more detail in standard sampling textbooks; good sources are Kish (1965) and Lohr (1999). Stratified sampling can be carried out for example with the SAS procedure SURVEYSELECT, which allows for several discrete variables as stratification variables.

Example. As a simple example, consider STR sampling with proportional, equal and Neyman allocation schemes. A stratified SRS-WOR sample of *n* = 200 elements is drawn from a population of *N* = 2000 elements (Table 1).

There are *H* = 5 strata in the population. In proportional allocation, a 10% sample is drawn from each stratum, involving a constant sampling weight $w_{hk} = w = 2000/200 = 10$ for every sample element. In equal allocation, a sample of $n_h = 200/5 = 40$ elements is drawn from each stratum, involving varying sampling weights $w_{hk} = N_h / n_h$ for each stratum *h*. For Neyman allocation, we assume that reliable knowledge on S_h, the population standard deviation (Std. Dev.) of *y*, is available, and that figure is equal to all strata except Stratum 3, whose standard deviation is larger indicating larger variation for the study variable. Stratum-wise sample sizes are calculated as $n_h = N_h S_h / 64000$, *h* = 1,…,5. This allocation scheme provides larger relative sample size for Stratum 3 and correspondingly, smaller sampling weight, when compared to the other strata. In those strata, the weights are nearly equal resembling proportional allocation.

Table 1. Proportional, equal and Neyman allocation schemes for STR sampling of *n* = 200 elements from an *N* = 2000 element population.

		Proportional allocation			Equal allocation		Neyman allocation			
Stratum h	Stratum size N_h	N_h / N	Sample size n_h	Sampling weight w_{hk}	Sample size n_h	Sampling weight w_{hk}	Std. Dev. S_h	$N_h S_h$	Sample size n_h	Sampling weight w_{hk}
1	500	0.25	50	10	40	12.5	20	10000	31	16.1
2	100	0.05	10	10	40	2.5	20	2000	6	16.7
3	800	0.40	80	10	40	20.0	50	40000	125	6.4
4	200	0.10	20	10	40	5.0	20	4000	13	15.4
5	400	0.20	40	10	40	10.0	20	8000	25	16.0
All	2000	1.00	200		200			64000	200	

3.2.5. Cluster sampling

To carry out *cluster sampling*, a sample of clusters (naturally formed groups of population elements such as clusters of employees in establishments, clusters of pupils in schools and clusters of people in households) is first drawn from the population of clusters by using one of the basic sampling techniques (SRS, SYS or PPS). Moreover, the population of clusters can be stratified before sample selection. In *one-stage cluster sampling*, all elements of the sampled clusters are included in the element sample. In *two-stage cluster sampling*, an element-level sample is drawn from the sampled clusters by using again the chosen basic sampling techniques.

An important advantage in cluster sampling is that a sampling frame at the element level is not needed for the whole population. The only requirements are for cluster-level sampling frames and, in two-stage cluster sampling, frames for sampling of elements from the sampled clusters. Cluster-level frames are often accessible, for example, for establishments, schools, villages, farms, blocks or block-like units in a city, etc. Auxiliary information in cluster sampling therefore concerns not only the grouping of the population elements into clusters but also the properties of the clusters needed if stratification is used. Stratification is typical in multi-stage sampling designs employed for example in business surveys. For example, the frame population of business firms can be stratified by type of industry or by size group before sampling of the individual firms.

In two-stage cluster sampling designs, PPS sampling is sometimes used for the first-stage units, that is, the clusters (for example regional units from the population of regions, enterprises from a business register, etc.). An equal probability or self-weighting sampling design is obtained if the elements are sampled from the sampled clusters with an equal sample size.

Cluster sampling is often motivated by cost efficiency, that is, the low cost of data collection per sample element. This is especially true for populations that have a large regional spread. Using cluster sampling, the traveling costs of interviewers can be substantially reduced as the workload for an interviewer can be regionally planned. The *cost efficiency* of cluster sampling can therefore be high. But there are also certain drawbacks of cluster sampling that concern statistical efficiency. If each cluster closely mirrors the population structure, we would attain efficient sampling such that standard errors of estimates would not exceed those of simple random sampling. However, in practice, clusters tend to be internally homogeneous, and this *intra-cluster homogeneity* increases standard errors and thus decreases *statistical efficiency*.

Cluster sampling is discussed at practical and more technical level in standard sampling textbooks. A good example is Kish (1965). Textbook by Lehtonen & Pahkinen (2004, Chapters 5, 7−9) gives several real-world examples on this phenomenon, and further illustrations can be found in the web extension VLISS-virtual laboratory in survey sampling.

Example. PISA 2000 Survey. The efficiency in cluster sampling is measured with *design effect* estimates. The design effect statistic was introduced in Section 3.2. For a sample mean $\bar{y}$, deff is given by $\text{deff}(\bar{y}) = \hat{v}(\bar{y}) / \hat{v}_{SRS}(\bar{y})$, where $\hat{v}(\bar{y})$ is the variance estimate calculated under the actual cluster sampling design and $\hat{v}_{SRS}(\bar{y})$ is the counterpart from simple random sampling. For cluster samples, design effect estimates tend to be larger than one, indicating poorer efficiency relative to simple random sampling. Correspondingly, the *effective sample size* decreases: $n_{\text{eff}} = n / \text{deff}$ becomes smaller than the original sample size *n*, if deff is larger than one. Effective sample size gives the SRS sample size that produces equal precision than the actual cluster sample of size *n* elements. We illustrate these properties by an example taken from Lehtonen & Pahkinen (2004, Section 9.4).

The data are from the OECD's Programme for International Student Assessment (PISA).The first PISA Survey was conducted in 2000 in 28 OECD member countries and 4 non-OECD countries. We discuss here the area of reading literacy. We selected from the PISA database the following countries: Brazil, Finland, Germany, Hungary, Republic of Korea, United Kingdom and United States. The survey data set from these 7 countries comprised a total of 1388 schools and 32,101 pupils.

Stratified two-stage cluster sampling was used in most PISA countries. The first stage consisted of sampling of individual schools with systematic PPS sampling. The number of students in a school was used as the measure of size in PPS sampling. In most cases, the population of schools was stratified before sampling operations. In the second stage, samples of students were selected within the sampled schools with equal probability.

The study variable *y* is the student's combined reading literacy score, scaled so that the mean over the participating countries is 500 and the standard deviation is 100. In Table 2, selected descriptive statistics are given The design effect accounts for weighting, stratification and clustering. The deff figures indicate a strong clustering effect for most countries.

The effective sample sizes of students are calculated by dividing the number of sample students by the design effect estimate. The effective sample size is the equivalent sample size needed to achieve the same precision in estimation if simple random sampling from a student population without any clustering were used. If the observations are not independent from each other, as is the case here, the effective sample size decreases: the higher the design effect, the smaller the effective sample size. Though the nominal sample sizes of students are large (several thousands) in all countries, some of the effective sample sizes are quite small (only a few hundred). Design-effect estimates also indicate that standard errors calculated under an (erroneous) assumption of simple random sampling would be much smaller than the (correct) design-based standard error estimates for most countries, tending to lead to unreliable statistical conclusions.

Table 2. Descriptive statistics for combined reading literacy score in the PISA 2000 Survey by country (in alphabetical order).

				Effective sample size of students	Number of observations in data set	
Country	Mean	Standard error	Design effect		Students	Schools
Brazil	402.9	3.82	8.33	476	3961	290
Finland	550.7	2.15	2.79	1600	4465	147
Germany	497.4	5.68	13.47	305	4108	183
Hungary	485.7	6.02	20.00	231	4613	184
Republic of Korea	526.6	3.66	12.99	351	4564	144
United Kingdom	531.4	4.08	14.08	564	7935	328
United States	517.0	5.16	6.93	354	2455	112
All	500.0			3881	32101	1388
Data source: OECD PISA database, 2001.						

3.3. Sample size determination

Survey statisticians face very often questions on how to determine the appropriate sample size for a sample survey. There is no simple answer and, thus, the statistician must ask about the needs, like:

I. Which are the most important study variables, and the parameters to be estimated?
II. Is there any guess about the (statistical) distribution of the study variables?
III. What is the level of precision one would like to have for the parameter estimates?
IV. What are the most important domains where the estimates must be provided and how precisely?
V. Are there any specific questions which must be taken in to account, e.g. special populations to be covered, certain analysis to be carried out, methodology used etc.?
VI. What will be the anticipated nonresponse rate?
VII. What are the financial and time constraints?

All these questions (and actually many others) should be considered before a proper sampling design including the sample size can be made. The first misunderstanding among non-survey practitioners is that the population size matters. By and large the population size does influence the sample size – except that the sample size cannot exceed the population size. [1]

Example. Community Labour Force Survey. The Council Regulation No 577/98 lays down the basic principles to be obeyed in calculation of the sample size. For simplicity we take only the first paragraph of Article 3:

Article 3
Representativeness of the sample

1. For a group of unemployed people representing 5% of the working age population the relative standard error for the estimation of annual averages (or for the spring estimates in the case of an annual survey in the spring) at NUTS II level shall not exceed 8% of the sub-population in question.

Regions with less than 300 000 inhabitants shall be exempt from this requirement.

Thus one must consider various aspects before the actual calculation:

1. Working age population? - Often considered as aged 15 or more. Sometimes 15-74.
2. Proportion of unemployed of that population (even though young people, say less than 20 or retired most often do not belong to labour force).
3. Sampling variance to be estimated according to the applied sampling design.
4. NUTS II domains.

Sample allocation

The Regulation requires that the sample size calculation must begin at NUTS II regions, i.e. at geographical domains. Generally, NUTS II regions are very different in size between countries and even within countries. Thus it is not a bad idea to apply stratified sampling where the strata consist of NUTS II regions.

[1] However, in some multinational surveys the national sample sizes can be adjusted to reflect differences of the population size. That is the case in EU-SILC, for example.

For example, in Finland there are five NUTS II regions. The population size for 15-74 year old people was in 2006:

Region number	Major Region	Population (15-74)
1	Southern Finland	1,987,000
2	Western Finland	999,000
3	Eastern Finland	496,000
4	Northern Finland	470,000
5	Autonomous Territory of Åland Islands	20,000

Since the smallest NUTS II region (no. 5) has population which is smaller than 300,000 it is not necessary to begin calculation from there but from the second smallest region (no. 4).

The yearly average of the number of unemployed was 204,000. Since the total population is about 4 million, the proportion is about 5 per cent of population - by chance it is exactly the proportion mentioned in the regulation. Furthermore assume that proportion of unemployed is roughly equal in all regions.

Stratified simple random sampling

We start the calculation of sample size from the simple random sampling of elements, i.e. individual persons. The regulation states that the coefficient of variation may not exceed 8 per cent. Thus for each NUTS II region (except the last one) one has to obey the condition:

$$\text{c.v}(\hat{p}) = \frac{\sqrt{\hat{v}(\hat{p})}}{\hat{p}} = 0.08$$

Our parameter estimate, proportion of unemployed from the working age population is $\hat{p}$ = 0.05 and we assume it is roughly equal in all NUTS II regions. Since unemployment indicator is a dichotomous variable with values [0,1] we can apply *binomial distribution* to approximate the sampling variance:

$$\hat{v}(\hat{p}) = \hat{p}(1-\hat{p})/n$$

It must be plugged into the formula above:

$$\text{c.v}(\hat{p}) = \frac{\sqrt{\hat{v}(\hat{p})}}{\hat{p}} = \frac{\sqrt{\hat{p}(1-\hat{p})/n}}{\hat{p}} = 0.08$$

Next we must raise the two components to the second power and rearrange:

$$\hat{p}(1-\hat{p})/n = 0.08^2\,\hat{p}^2$$

Thus

$$n = \frac{\hat{p}(1-\hat{p})}{0.08^2\,\hat{p}^2} = \frac{0.05\times 0.95}{0.08^2\times 0.05^2} = \frac{0.95}{0.08^2\times 0.05} \approx 2{,}969$$

Thus the sample size would be roughly 3,000. In addition, one must consider other issues. Namely,

1. What is the anticipated nonresponse rate?
2. Can we apply the sampling design in this case, or should we consider some other design?

The above sample size was calculated with an assumption of 100% response. In real life it should be inflated by the anticipated nonresponse rate and possible undercoverage problems in the sampling frame. Consider, for example, that it is about 15% in comparable social surveys. The inflated sample size would be about 3,500.

Next question is whether to use equal or proportional allocation (see section 3.2.4). *Equal allocation* will yield approximately the same precision to all strata. Hence the sample size for the whole country would be more than 14,000 individuals (4 × 3,500 = 14,000 + specifically chosen sample size for the 5th NUTS II region).

For *proportional allocation* the sample size must be fixed to the NUTS II region no 4 according to calculation above. The sampling rate of that region is 3,500/470,000 i.e. about 0.75 per cent of the population. Using the same sampling rate the sample sizes for other NUTS II regions would become 14,800; 7,450; 3,700 and 150. Thus the total sample size will increase to more than double: 29,600 – depending on the last region sample size.

Other main types of allocation (e.g. Neyman or power allocation, see 3.2.4) are not easily applied in this example since their aim is to allocate the sample optimally over the whole population which might not fulfil the regulation requirements on the precision over NUTS II domains.

One-stage cluster design

In many countries the sampling frames are old or somehow not amenable for selecting individual persons or households directly for social surveys, there for the LFS. In those cases some cluster sampling design is applied. *One-stage cluster sampling design* requires that the units to be selected contain the ultimate sampling units, individual persons. The primary sampling units (PSU) can be addresses, houses or dwelling units, for example. When a *two-stage cluster design* is applied, the first-stage sampling units are often some administrative entities: municipalities, villages or census enumeration areas. The second-stage sampling units are again addresses, houses or dwelling units, and the ultimate sampling units consists of individual persons (see section 3.2.5).
In any case, the number of primary and secondary sampling units must be calculated differently. However, the basic calculation for the element sampling design above can serve as an input to that purpose, too.

For simplicity we deal here with a stratified one-stage cluster design, i.e. instead of individuals we collect information from all members of households, selection either from address frame or dwelling unit frame. Furthermore we assume that we apply *equal allocation* scheme in NUTS II strata. What we need for a cluster design is the average size of clusters and some knowledge on design effect estimate. From another survey we can anticipate that the *design effect* (see p. 12) of the number of unemployed from a one-stage cluster designs is about 1.5. If we assume that the average cluster size (i.e. average number of eligible members which fulfil the required conditions) is 3 and that the nonresponse rate calculated from the individuals remains at 15% (i.e. response rate $\tau = 0.85$) we can obtain the sample size by

$$m_{CLU} = \frac{n_{SRS} / \tau}{\bar{n}_{CLU}} \times \mathrm{deff}\,(\hat{p}) \approx \frac{3,500}{3} \times 1.5 = 1,750$$

Using the equal allocation scheme the overall sample size would be over 7,000 households (4 × 1,750 + specifically chosen sample size for the last NUTS II region). So the requirement could be fulfilled with a relatively large sample size: 7,000 households contain 21,000 individuals. The fieldwork cost can, of course, become smaller than that of the stratified SRS sampling design.

Business surveys

Sample size determination for business surveys can appear much more difficult than the example above. The distributions of main study variables are often very skewed which can lead to a need of balancing various requirements. The basic calculation follows, of course, the same path described above but the sampling designs are typically heavily stratified element sample designs. Stratification is often carried out by the industry (NACE classification) and size class.

The most influential units are often considered as certainty units, and each of them form technically a stratum of its own. Thus there is no sampling variance involved (unless some units fail to respond). For smaller units the sampling is carried out by stratified systematic random sampling, simple random sampling or PPS sampling. The sample size calculation must take all the necessary information into account, and in the case of detailed stratification, calculation must be carried out separately for each stratum.

3.4. Use of auxiliary information in the estimation phase

In modern survey sampling practice, auxiliary information is often used to improve the efficiency of estimation for a given sample, by using *model-assisted estimation techniques*. Thus, in addition to the sampling design, an *estimation design* enters on the scene. The concept of *estimation strategy* is sometimes used referring to a combination of a sampling design and an estimation design. Table 3 shows examples of strategies, including design-based strategies where auxiliary information is used in some strategies, and model-assisted strategies where auxiliary data are incorporated in the estimation phase, and for some strategies, also in the sampling design.

Table 3. Examples of estimation strategies.

	Auxiliary information	Assisting model	Example strategies (sampling design * estimation design)
Design-based strategies			
SRS-WOR	Not used	No explicit model	SRS-WOR strategy
SYS	Not used	No explicit model	SYS strategy
PPS	Size variable	No explicit model	PPS strategy
STR	Stratification variables	No explicit model	STR * SRS-WOR
Model-assisted strategies			
Ratio estimation	Continuous	Regression (no intercept)	SRS-WOR* Ratio estimation
Regression estimation	Continuous	Regression (with intercept)	STR * Regression estimation
Post-stratification	Discrete	linear ANOVA	SYS * Post-stratification

In model assisted estimation, the auxiliary data are incorporated in the estimation procedure of the population total by using statistical models. Linear models are most often used for this purpose, especially when the response variable is of continuous type. *Ratio estimation* and *regression estimation* use a linear regression model, where the explanatory variables are assumed continuous. In ratio estimation, the intercept term is excluded from the regression model, whereas the intercept is included in the regression model underlying regression estimation. In practice, several continuous auxiliary variables can be incorporated in a regression estimation procedure. For both methods, the auxiliary data consists of population totals of one or several continuous variables, which can come from a source such as official statistics.

In *post-stratification*, a linear analysis of variance or ANOVA model is used as the assisting model, and the explanatory variables are of discrete type. The auxiliary data consists of population cell and marginal frequencies of one or several categorical variables. A benefit in all these methods is that an access to unit-level auxiliary data on the population is not assumed.
Model parameters are estimated from the observed sample by using the weighted least squares (WLS) technique. Nonlinear models also can be used as an assisting model. This is the case especially if the study variable is binary or polytomous (see e.g. Lehtonen, Särndal &Veijanen 2005).

Ratio and regression estimation and post-stratification are special cases of *generalized regression (GREG) estimators*. The book by Särndal, Swensson & Wretman (1992) provides the main reference text for model-assisted survey sampling. Model-assisted methods are discussed, at a more practical level, in Lehtonen & Pahkinen (2004). The methods are illustrated (with computational examples) in the VLISS application, the web extension of the book.

4. Treatment of nonresponse

During data collection we will lose information, the main reason being unit nonresponse. Thus the number of responses is smaller than the originally selected sample. If the sample size was not inflated by the anticipated nonresponse the analyses and findings will contain greater risk to incorrect conclusions than originally thought.

Unit nonresponse is the most common reason for missing data. The classification below shows how they affect the data to be analyzed. The textbook by Groves et al. (2002) gives a through presentation of those effects and the ways to take them into account in analysis.

Type of missingness:	Effect on data set:
- unit nonresponse	The whole data vector remains empty (or all items were rejected)
- item nonresponse	One of more items empty/are rejected
- sub-unit/partial nonresponse	All data from one or more ultimate cluster elements missing, e.g. one or two family members refuse, data accepted from others

Unit nonresponse

All survey organisations keep track on the magnitude and distribution of response and nonresponse. However, despite efforts there is no standardised way of calculating those entities (see e.g. Lynn et al. 2003). Typically, nonresponse is classified into following categories in household surveys: Non-contacts, Refusals and Other reasons. In addition, the coverage errors should be distinguished whenever possible.

In business surveys – which are often compulsory – the coverage issues tend to be more dominant than in household surveys. In some cases enterprises refuse to respond because they may feel their competitors can somehow dig out strategic market information. Statistical agencies strive to ensure that the most dominant enterprises will respond in any case to preserve the credibility of the results.

4.1. Reweighting to adjust for unit nonresponse

If respondents and nonrespondents were distributed quite randomly over the whole sample, the nonresponse could be detected as *ignorable*, and one could apply a naive model $\tau = m /n$ which means just replacing the sample size by the number of respondents in weighting. Thus the only punishment would be increased sampling variance. Unfortunately such a naive model is normally not valid and data may be distorted by the nonresponse effects, i.e. *non-ignorable nonresponse.* The most often applied reweighting methods are post-stratification and ratio estimator. Below we present also some other methods.

Post-stratification

We can easily *post-stratify* our data set (see Section 3.4). For nonresponse adjustment the best way is to find subgroups where the response/nonresponse rates are as different as possible. The idea is to find homogenous sub-groups with the responding behaviour. The task requires substantial amount of tabulation but once done the weighting is relatively straightforward.

Ratio estimator

The *ratio estimator* (see 3.4) may as well be used after data collection either directly for estimating the study variable *y*, or by weighting. We need the totals of continuous variable(s) either at population or domain level, and the corresponding observation x_k for each respondent.

Ratio estimator yields good results for a study variable *y* only if there is a strong positive correlation between *x* and *y*. Note that ratio estimator is biased by definition if the true regression line does not go through origin.

Empirical response homogeneity groups (RHG)

Empirical *response homogeneity groups* (RHG), often called as weighting classes mean that the sample elements are divided into mutually disjoint sub-groups according to the response probabilities. It is technically similar to post-stratification but operates only with sample-level information – not population information.

Assume SRS-WOR design and that the responding part of the sample can be divided according RHG's into *L* categories:

$$\tau_l = \frac{m_l}{n_l} \in (0,1],\ l = 1,...,L\,.$$

Thus the final modified weight is $w_k = a_k / \tau_l$, $k \in G_l$, where a_k is the design weight and G_l refers to response homogeneity group *l*.

Derivation of good unbiased variance estimator for the weighting class estimator is very tedious and thus will be omitted here.

Explicit response probability modelling

Modelling is a continuation of the tabulation information. Thus models can either be fit at the individual level or at the frequency level. The idea is again to divide the sample into separate groups according to response probabilities. However, a model is used to help the task and furthermore to predict the model-based response probabilities. *Logistic regression models* (see 3..4) are most popular in this case due to the character of phenomenon.

A good model takes the main effects of many variables into account and, if necessary, also the interactions. As a result the predicted values are smoother than the original empirical response probabilities.

Calibration of weights

The *calibration estimator* is probably the mostly used reweighting method at present. It is versatile and combines the good properties of modelling the nonresponse with the need to have consistency of auxiliary information, i.e. the sample estimates of chosen auxiliary information distributions match with the population distributions. Also the sampling variance approximations are available (see e.g. Deville & Särndal 1992; Särndal & Lundström 2005).

4.2. Imputation to adjust for item nonresponse

Imputation is the main method for compensating the item non-response. It is also applied to correct for erroneous values found in data checking. Therefore the phase of work is often called “editing and imputation”. The main methods are logical imputation, real-donor imputation and model imputation. The European Research Framework Programmes 4 and 5 contained projects, called AUTIMP and EUREDIT which provided a lot of tools for statistical offices on editing and imputation.

Logical imputation is connected with editing of data. Currently it takes place as a part of data entry program: computer assisted interview programs, data extract programs from administrative sources and various business survey data entry programs are equipped with a set of logical checking routines which can detect errors and also give either right or “best guess” imputed values for the missing or incorrect ones.

Real-donor imputation relies on the responses from the same unit or similar unit. In business surveys it has been a tradition to impute missing values with the values from the previous measurement of the same unit. That method is called *cold-deck imputation.* Sometimes the previous values are multiplied by the average change of the variable estimated from the same industry.

If the value is borrowed from a neighbour unit – physically the next or previous observation – the method is *hot-deck imputation.* Traditional hot-decking leads often to implausible situations and therefore it is improved with some *nearest neighbour imputation* method where similar units are sought with some metrics or by choosing sub-populations. The value may be taken directly from the closest unit or taken as lottery from a pool of similar units.

Model-donor imputation refers to the situation where some statistical model is used to predict the missing (or incorrect) value. Typically that occurs again in business surveys where the acceptable responses can be used to fit the model and then just applied to the units where there are missing values. The simplest method is to take an overall or suitable domain mean. Mean imputation cannot be regarded as good because of its tendency of reducing the variation very much in the case there are a lot of missing values. Imputation based on a regression or some other statistical model yields much better results in that respect. One can also lessen the tendency of variance reduction by adding some extra random variation from appropriate distribution which will lead to *stochastic imputation* instead of *deterministic imputation*.

In principle all the methods above deal with a *single imputation* case, i.e. one missing value is imputed with one value. *Multiple imputation* is another approach which solves many problems by imputing one missing value with a number of new values (see Rubin 1987, or Schafer 1997). It is statistically a sound method but the statistical offices (and often their clients) have not yet taken the methods in their work programmes.

5. Software

Two large scale statistical programs, SAS and SPSS contain specific modules for sample selection and survey data analysis. In SAS the procedures are included in STAT module from version 8 on. In SPSS there is an add-on module Complex Samples which must be purchased separately, available from version 11 and above.

The programming language R is widely used in academic studies and some authors have provided codes for survey sampling and analysis. Those are found on different web-pages containing R program codes, e.g. www.r-project.org/.

Software for sample selection

SAS/Stat v. 8 and above	Proc SurveySelect (www.sas.com)
SPSS/Complex Samples v. 11 and above	CSPLAN together with CSSELECT (www.spss.com)

Software for weight derivation

SAS add-on programmes:	CALMAR2	(INSEE/France)
	CLAN97	(Statistics Sweden)
SPSS add-on programmes:	g-CALIB	(Statistics Belgium)
BLAISE component	BASCULA	(Statistics Netherlands)

Software for editing and imputation

SAS/Stat v. 8 and above	Proc MI and MIANALYZE (multi-imputation)	
SAS add-on programmes:	BANFF/GEIS (Statistics Canada)	
	CONCORD, DIESIS,	
	QUIS	(ISTAT/Italy)
	IVEWARE	
		(www.isr.umich.edu/src/smp/ive/)

Software for “standard” estimation

SAS/Stat v. 9 and above	Proc Surveymeans Proc Surveyfreq	
SAS add-on programmes:	CLAN97	(Statistics Sweden)
	POULPE	(INSEE/France)
	GES	(Statistics Canada)
SPSS v. 11 and above Complex Samples	CSDESCRIPTIVES CSTABULATE	
SPSS add-on programmes:	g-CALIB	(Statistics Belgium)
BLAISE component	BASCULA	(Statistics Netherlands)
SUDAAN	(www.rti.org)	

Software for analytical purposes

SAS/Stat v. 9 and above	Proc Surveyreg Proc Surveylogistic (GLM, MIXED, NLMIXED,…)
SPSS v. 11 and above Complex Samples	CSGLM, CSORDINAL, CSLOGISTIC
SUDAAN	
WesVAR	(www.westat.com)

6. References

Andersson, C. & Nordberg, L. (1994). A Method for variance estimation of non-linear functions of totals in surveys ! theory and a software implementation. *Journal of Official Statistics*, Vol. 10, No. 4, 395-406.

Andersson C. (2002). CLAN97 v. 3.1. Supplement to A User's Guide to CLAN97. (Unpublished).

Biemer P. & Lyberg L. (2003). *Introduction to Survey Quality*. New York: John Wiley & Sons.

Brewer K. & Hanif F. (1983). *Sampling with Unequal Probabilities*. New York: Springer-Verlag.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd ed. New York: John Wiley & Sons.

Deville J-C., & Särndal C-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, Vol. 87, No. 418, 376!382.

Deville J-C., Särndal C-E., & Sautory O. (1993). Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, Vol. 88, No. 423, 1013-1020.

European Union (1998). *Council regulation (EC) No 577/98 of 9 March 1998 on the organisation of a labour force sample survey in the Community*.

Groves R.M., Dillman D.A., Eltinge J.L., & Little R.J.A. (eds.). (2002). *Survey Nonresponse*. New York: John Wiley & Sons.

Groves R.M., Fowler F.J., Couper M.P., Lepkowski J.M., Singer E. & Tourangeau R. (2004). *Survey Methodology*. Hoboken: John Wiley & Sons.

Hansen M.H., Hurvitz W.N. & Madow W.G. (1953). *Sample survey methods and theory*. New York: John Wiley & Sons.

Kish L. (1965). *Survey sampling*. New York: John Wiley & Sons.

Lehtonen R. & Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys*. 2nd ed. Chichester: John Wiley & Sons.

Lehtonen R., Särndal C.-E. & Veijanen A. (2005). Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition*, 7, 649–673.

Lessler J. & Kalsbeek W. (eds.). 1992. *Nonsampling Error in Surveys*. New York: John Wiley & Sons.

Lohr S. (1999). *Sampling: Design and analysis*. Pacific Grove: Duxbury Press.

Lyberg L., Biemer P., Collins M., de Leeuw E. D., Dippo C., Schwarz N. & Trewin D. (eds.). (1997). *Survey Measurement and Process Quality*. New York: John Wiley & Sons.

Lynn P., Beerten R., Laiho J. & Martin J. (2003). Towards standardisation of survey outcome categories and response rate calculations. *Research in Official Statistics,* Vol. 5 1/2002, 61-84.

Oh J.L. & Scheuren F. (1983). Weighting Adjustment for Unit Nonresponse, in W.G. Madow, I. Olkin & D.B. Rubin (eds.) *Incomplete Data in Sample Surveys*, Vol 2., 143-184. New York: Academic Press.

Rubin D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Sautory O. (1993). La macro CALMAR. Redressement d'un échantillon par calage sur marges. I.N.S.E.E., Série des documents de travail nE F 9310. Paris: I.N.S.E.E.

Sautory, O. & Le Guennec, J. (2005). La macro CALMAR2. Redressement d'un échantillon par calage sur marges. I.N.S.E.E. Paris: I.N.S.E.E.

Schafer J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Smith T.M.F. (1990). Post-stratification. *The Statistician*, 40, 315!323.

Sundgren B. (1999). Information systems architecture for national and international statistical offices. Guidelines and recommendations. *Conference of European Statisticians. Statistical Standards and Studies.* No 51. Geneva: United Nations.

Särndal C-E. & Lundström S. (2005). *Estimation in Surveys with Nonresponse*. Chichester: John Wiley & Sons.

Särndal C-E., Swensson B. & Wretman J. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

United Nations (1950). The Preparation of Sampling Survey Reports. *Statistical Papers* Ser. C No. 1 (Revised). New York.

United Nations (1964). Recommendation for the Preparation of Sample Survey Reports (Provisional Issue). *Statistical Papers* Ser. C No. 1 Rev.2. New York.

Wolter K. (2007). *Introduction to Variance Estimation,* Second Edition. New York: Springer.

7. Web links

AUTIMP (AUTomatic IMPutation software for business surveys and population censuses)
see EUREDIT

EUREDIT (The Development and Evaluation of New Methods for Editing and Imputation):
http://www.cs.york.ac.uk/euredit/

Eurostat Quality site:
http://ec.europa.eu/eurostat/quality

The European Statistics Code of Practice
ESS Quality standards:
Standard Quality Report
Standard Quality Indicators
How to Make a Quality Report
Handbook on improving quality by analysis of process variables
DESAP – self-assessment for survey managers

International Monetary Fund. Data quality assessment framework. 2003,
http://dsbb.imf.org/vgn/images/pdfs/dqrs_Genframework.pdf

Organisation for Economic Cooperation and Development. OECD PISA database, 2001:
http://pisaweb.acer.edu.au/oecd/

Organisation for Economic Cooperation and Development. Quality framework and guidelines for OECD statistics. 2003. http://www.oecd.org/dataoecd/26/42/21688835.pdf

Statistics Finland: Quality Guidelines for Official Statistics:
http://stat.fi/tk/tt/laatuatilastoissa/alku_en.html

VLISS-virtual laboratory in survey sampling:
http://www.math.helsinki.fi/VLISS/.

Appendix 1. Example of sample selection using Bernoulli sampling

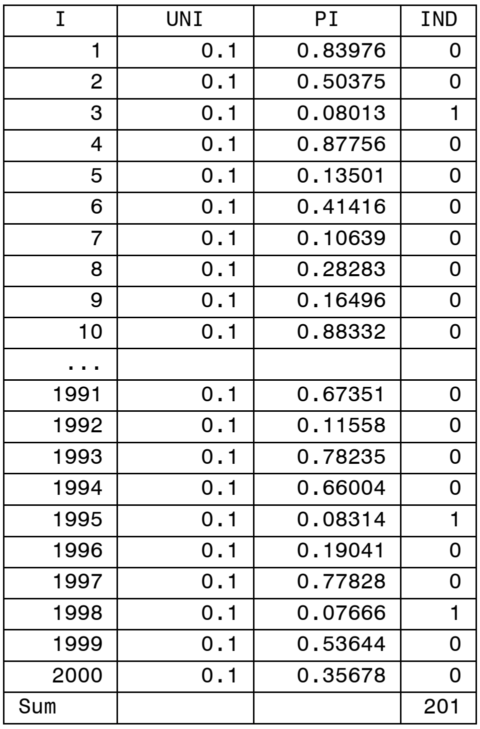
We create a sampling frame consisting 2000 elements and want to select about 200 units to the sample. The symbols are the same as used on p. 15.

Sampling fraction UNI = 200/2000 = 0.1. All elements in the frame are assigned a pseudo random number from Uniform distribution, PI. Those elements with PI ≤ 0.1 are selected and selection indicator IND is given value 1. If the unit was not selected, IND is set 0.

The program code for SAS is very simple:

```
Data Bernoulli;
        UNI=200/2000;             /* Set the limit for selection    */
Do I=1 to 2000;                   /* Create a frame of 2000 elements    */
        PI=Ranuni(0);             /* Attach every element with a random number*/
        If PI<=UNI then IND=1;    /* Check whether the unit is selected */
        Else IND=0;               /* or not */
        Output;
End;

Proc Print;
Sum IND;
Run;
```



I	UNI	PI	IND
1	0.1	0.83976	0
2	0.1	0.50375	0
3	0.1	0.08013	1
4	0.1	0.87756	0
5	0.1	0.13501	0
6	0.1	0.41416	0
7	0.1	0.10639	0
8	0.1	0.28283	0
9	0.1	0.16496	0
10	0.1	0.88332	0
...			
1991	0.1	0.67351	0
1992	0.1	0.11558	0
1993	0.1	0.78235	0
1994	0.1	0.66004	0
1995	0.1	0.08314	1
1996	0.1	0.19041	0
1997	0.1	0.77828	0
1998	0.1	0.07666	1
1999	0.1	0.53644	0
2000	0.1	0.35678	0
Sum			201

In Bernoulli sampling the sample size is a random quantity and this example shows that we received one unit too much. The simplest way to obtain a fixed sample size would be to sort the frame by the random number and select exactly 200 cases (from the beginning, end or just at any point as long as the random numbers are used for selection).

European Commission

Survey sampling reference guidelines – Introduction to sample design and estimation techniques

Luxembourg: Office for Official Publications of the European Communities

2008 — 36 pp. — 21 x 29.7 cm

ISBN 978-92-79-07840-8
ISSN 1977-0375