Rapporto n. 2 2005

dmsia @ unibg.it

S. Biffignandi, D. Toninelli

Sampling and Inference in Web Surveys

Serie Didattica

Dipartimento
di Matematica, Statistica,
Informatica e Applicazioni
"Lorenzo Mascheroni"

UNIVERSITÀ DEGLI STUDI DI BERGAMO

Index

Part I – Frames and Sampling

Part II – Inference

Abstract

Web-based and Internet surveys offer a much interesting potential. It is a wide spread belief that this data collecting tool makes it easy and cost effective to contact respondents. These guidelines do not cover web-based survey costs analysis. However, even if the costs of producing questionnaires, lists, etc. are an important component, marginal costs to reach a high number of respondents, even with a high level of dispersion in a territory, are very low, and could be even zero cost.

The target of these guidelines is to highlight the main problems and the points which should be focused on by the researcher when he decides to conduct a web-based survey.

In the first part of this paper some important themes linked to web sampling are discussed. The second part deals with the theme of inference using the same type of survey.

The first and second chapters briefly illustrate the main concepts related to sample surveys and sampling methodologies. Sampling techniques applicable to web researches are similar to those used in traditional surveys. Even in this case, therefore, it is important to divide the sampling methods into two different categories: probability and non-probability sampling.

Even if the classification is similar to the traditional one, there are, however, some important considerations to be made regarding specifically the use of sampling for web-based surveys.

The first important aspect concerns the availability of sample frames (Chap. 4): it is not always possible to obtain appropriate lists and, even if they are available, they are not always suitably drawn up, up-to-date or applicable for the research. In Par. 4.1 the main sources used for drawing up sampling frames are analyzed, pinpointing the problems that each of these involves. In Par. 4.2 a possible classification of sources is produced taking into consideration the probability or non-probability connotations of the samples which can be selected.

Even in the case when the sampling frame is suitably made up, some fundamental themes typical of the sampling for web-based surveys must not be ignored. Firstly, people who have access the Internet are characterized by particular and specific social-demographical and economic descriptions. Therefore, the application of a population to a web-based survey and the choice of an appropriate sampling technique cannot be carried out before one has adequate information on people who have or don't have access the Internet. Secondly, even if you restrict the analysis to certain arguments or to a particular population, potential problems connected with the presence of professional respondents together with the honesty of the answers must also be taken into careful consideration.

The second part of this paper deals with problems regarding inference in web based surveys.

Estimation theory provides both the statistical criteria for the transfer of sample data to population parameters (the so-called inference problem) as well as for the measurement of the level of estimate reliability. However, this is only valid if the sample selection is carried out on a random basis. The theory has been developed with reference to traditional data collection tools.

Regarding web based surveys, studies on inference problems are in progress and no final results are identified.

Nevertheless some observations can be made and some basic comments given. Above all it should be pointed out that estimation theory to make inference on web collected data is applicable in principle. The problem is that it is not always possible or easy to provide probabilistic samples, above all due to the difficulty in identifying a population frame suitable and similar to population (Pineau & Slotwiner, 2003).

From the point of view of inference, solutions to resolve situations where the sample is not representative of the target population are still being studied. An operative solution could be using a weighting system which takes into account different probabilities in the extraction of persons from a population which could exclude certain categories from being selected by using a specific sampling techniques (post stratified Horvitz and Thompson estimator).

Another possible solution would be to use the weighted system based on the propensity score, this is introduced with the aim of making it possible to apply inference principles to observational data (i.e. taken from samples not from random selections) (Biffignandi & Pratesi, 2005; Forsman, 2004).

Keywords

Sample, sampling techniques, probability sampling, non-probability sampling, sample frame, inference, confidence intervals, statistical hypothesis testing, errors, observational errors, non-observational errors, probabilistic selection, balancing score methods, propensity score methods.

Part I - Frames and Sampling

1 Introduction to Sampling

Data collection by surveys can use two different kind of approaches: one draws information from all the universe of the statistical units (this is called census), the other one considers only a part of all the statistical units (you do a partial observation, you select a sample of the statistical units).

In the first case, the harvested information enables you to gain a complete view of the target population, errors and inaccuracies depend on non correct coverage of the target population, from missing data, and so on.

In the second situation, the observed information is partial and the real population (the target population) value is unknown. However, depending on the statistical technique used for selecting statistical units of the sample the value of the target population can be estimated. The statistical inference technique makes it possible to obtain estimates of the value of the target population.

In the planning of a survey, the choice of a global or sampling collection of data is extremely critical and is connected to different factors, e.g. costs of data collection, timeliness, generalization of results, ...

Sampling means it is possible to reduce costs of carrying out a survey (Chisnall, 1990); in this way one can channel more funds into activities such as collecting data process quality controlling. Sampling surveys also require shorter completion times (giving more timely results. Moreover, In some cases it is necessary to use partial surveys: e.g., when it is not possible to contact all the statistical units, for instance when the observation of the future object of study, implies the destruction of the observed units. In Internet/Web survey the above-mentioned advantages hold too. In addition, comparative lower costs are often hypothesized, too.

In the first part of these guidelines, we again recall the definitions of "sampling survey", "sample", "sampling".

In the second part we will examine the problem of the choice between the different way of surveying (partial or global) with specific reference to web-surveys.

The fast and capillary diffusion of Internet worldwide offers many new opportunities and new tools of surveying; furthermore it concentrates the researchers' attention on solutions to new problems and on new important topics to face. Despite all the problems that exist now, web-based sampling surveys are considered a very versatile and useful tool that in the future will become even more widely used. Internet will mean searches can be optimized and, above all, the times of completion will be shortened considerably, leading to lower costs of survey production.

For a more exhaustive study of the web-based sampling survey, we propose a systematic classification of the different kinds of survey, as well as the different types of respondents that it is possible to involve in these and of the main tools which can be used to contact elements of the target population in building up a sample.

With this purpose we have concentred our attention on the most important problems, as apart from the choice between a probabilistic or a non-probabilistic approach, the non-coverage of the lists, the presence of professional respondents and how representative the sample is.

2 Some definitions

The statistical unit target of study composes the reference population (also called "universe" or "target population"), i.e. all the subjects from which one wants to collect information and which survey's conclusions must been referred or extended to.

To evaluate how a phenomenon is presented in a target population, it must be observed how it manifests itself on individual units.

Observing or collecting data can be extended to the entire population (census): usually this happens when one wants to analyze very small groups of units or groups which are difficult to find or contact, or when one wants to conduct a very detailed analysis (as happens in censuses of the population).

In many cases, however, this is not necessary, appropriate or possible to collect data from the statistical units that compose the target population. In these cases we cannot carry out a census, but only collect information from a part of the statistical units subject to the study only, i.e. a small group of those, called "sample".

If the data collection is based on a sample, even the survey being managed is called a "sampling survey". In sampling surveys, therefore, one can observe information in a standardized way, observing how the phenomenon that we want to study appears on the target units that belong to a representative sample. The major task of a sampling survey is to allow one to go back with sufficient approximation to the complex characteristics of the phenomenon, included those that are relevant to non observed cases.

By "representative" we mean that the sample units are chosen to make that sample represent the universe on a smaller scale, relative to some features of the phenomenon being studied or to other characteristics. The more the features of the sample are equal to that of the universe, the more the results of sampling survey can be considered reliable estimations of the state of the phenomenon in the entire population.

With "sampling" we mean the way in which, from a set of target units (the target population) we extract a reduced number of cases. These must be selected in such a way that makes it possible to generalise the results obtained from the sample to the entire population.

The first step for sampling is the sampling frame, i.e. the list with units that are eligible for inclusion in the survey and that are approachable and contactable. Firstly it is necessary to check that the list is suitable for the target population, and this can be done by submitting the list itself only after careful consideration. This allows us to limit, as much as possible, the influence of systematic error or error produced by eventual sample bias. It is helpful, in this situation, to establish inclusion and exclusion criteria that allow us to decide whether a statistical unit can be inserted in the sampling frame.

There are different sampling techniques used to obtain samples of the universe. Later we will indicate the main one classified in two different categories: probabilistic and non-probabilistic.

With "sample size" we intend the number of statistical units that constitute the sample, i.e. the number of selected units. For small populations, a sample may consist of 10/30% of the target population. In very large populations (150,000 people or more) the sample can consist of about 1% of the total units. There are simple formulas that allow the computation of sample size in connection with the sampling methodology chosen.

3 Sampling Methods

The choice of the methodology of sample selection depends, firstly, on the aims of the survey, and the most appropriate way to carry it out depends on the survey context. The method of selection should be chosen which best fits the situation.

The first classification of the sampling techniques can be made by distinguishing between probability and non-probability sampling.

3.1 Probability Sampling

Probability sampling aims at obtaining a representative sample of the target population. You can eliminate subjectivity in choosing the sample: there is a random selection of units that will be in the sample.

Each unit of the target population has a known and different from zero probability to be included in the sample. In the non probability sampling method, however, this probability of inclusion is unknown.

Knowing the inclusion probability means estimates can be provided of the features of the population from which sample has been selected respecting the maximum efficiency compatible with the limit of costs prefixed and to evaluate precision degree of the sample estimations. Probability sampling is used when you want to achieve conclusions or forecasts valid for the entire population.

To select a sample, it is necessary to have a well compiled, complete list of all statistical units of the universe. "Either the definition of population, or the choice of units (geographical units, social groups, foundations, households, housing, persons, facts,...) constitute foundations of any extracting methods.

The main probability sampling techniques are:

- Simple Radom Sampling
- Stratified Sampling
- Systematical Sampling
- Cluster (or Area) Sampling

- Multistage Sampling
- Double Sampling
- Replicated Sampling

The choice of one rather than the other methodology depends on the context of research in which one is involved, from the aims of the survey and from the target population involved in the project.

3.2 Non-probability Sampling

In non-probability sampling there is no random selection of the statistical units that should be involved in survey; the sample is made considering the target population characteristics and the purposes of the survey. Thus, differently from probability sampling, some members of the population can be chosen (or extracted), whereas others cannot be selected. This means that single units do not have equal probability to be selected: the probability that some of these will be chosen may be equal to zero.

In some cases, using a random selection of elements, results, in conclusion, are not applicable to the target population. In these situations non-probability sampling methods can be used to select representative units of the entire population.

Non-probability sampling is better than the probability one in these three cases:

1. When the survey regards groups which are difficult to identify (i.e. you do not have a complete list of the elements of the universe). This happens, e.g., when you want to research an illegal phenomenon (black economy), when it is difficult to obtain the cooperation of respondents, in the case you're dealing with difficult or tough themes or when population units are untrustworthy (as in survey regarding illegal immigrants, who fear expulsion). Unavailable units can be substituted with other considered equivalents, because it is not necessary to respect random principles of probability methods. This can contribute, in many situations, to reduce research costs.
2. When the target is carrying out a survey regarding specific groups of people: ethical reasons or reluctance in tackling all potential respondents favour these methods instead of the probabilistic ones.
3. In pilot surveys: when the purpose is to collect experimental data on target groups and when the data collected are not destined to be released, but to the planning of a survey or of a research. Non-probability sampling considerably reduces the costs of collecting data and research time.

The main non-probabilistic sampling methods are:

- Convenience sampling
- Snowball sampling
- Quota sampling
- Focus Group
- Judgement or purposive sampling

- Systematical non-probability sampling
- Self-selection sampling
- Representative elements sampling
- Area barometer sampling
- Plausability sampling

Non-probability sampling methods do not allow (as probability ones) the creation of an interval near the sample mean, in which it is possible to compute, with reasonable confidence, the unknown value of the population. Furthermore there is a relevant risk of systematic errors.

The various probability and non-probability sampling methods illustrated here can be used in combined ways. From the combination of probability sampling methods it is possible to obtain about 32 different other methods. (Malhorta 1999).

4 Web-based Surveys Sampling

In principle the traditional methods of probability and non-probability sample selection are applicable to Internet and web surveys.

The target and the sampling frame, as well as the respondents selection, are in many cases unique to the Internet and web. Mainly the difference between target and sampling population and the difficult in identifying the sampling frame causes coverage problems. In addition self-selection is specific to Internet and important from the sampling technique viewpoint. In fact, in most cases it is difficult to draw a probabilistic sample and in practice non probabilistic samples are used.

For these reasons it is important to classify the different approaches for the respondents selection which are currently used in Internet and web surveys, before entering sampling techniques.

4.1 The sampling frame

The "sampling frame" (also defined as "sampling list") is the list of the units who satisfies the criteria adopted in a survey for identifying the target population. The sampling frame is, in the empirical analysis, considered a proxy of the target population. Therefore a good target frame covers most part of the target (Malhorta 1999). From sampling frame you can select the units that will be part of the sample target of research.

One of the main problem in web-based surveys, moreover common with a lot of kind of surveys, is the availability of a suitable sampling frame. Indeed only when the sampling frame is a good representation of the target population a probabilistic approach can be used and statistical inference's principles can be applyed.

In web-based surveys there are, essentially, two kinds of approaches often used who bring to the building of partial, self-selected frames, that are not a good approximation of the target population and can be considered non-probabilistic samples (this means that a probabilistic sample can be extracted from them).

These approaches are:

- **Internal approach**: the respondents are detected on the web.

  They are visitors of web sites or people registered (for a different reasons) in mailing-lists.

  In the Alvarez classification (2003), the internal approach is divided in two different ways of respondents selecting that consider the kind of approach used.

  The first one is based on a way of contact called "Web Advertisements": the recruitment occurs through announces published on web sites or applied to newsgroups. The announces want to encourage the participation to the survey. This way of selection brings to non-probabilistic samples.

  The second one is based on subscription or on co-registration procedures: to people who are registered for a certain service is asked to provide a valid e-mail address. Afterwards these units are involved in different kinds of web-based surveys.

- **External approach**: the respondents are contacted through panel lists, paper lists, or other kind of list. The approach is different from the internal one: the respondents may be contacted through mail, bill posting, phone contact, and so on. The potential respondents are asked to go in Internet to participate to the survey. Internet, in this second approach, is a medium to collecting data and not a way to involve the potential respondents.

There are different kind of sources of information that can be used to build the sampling frame. Every kind of frame has advantages and disadvantages from viewpoint of sampling and estimating. The main sources are the following.

Announcements: they are published on web-site pages (invitation tags), sent via e-mail or through other ways of communications (like distribution of leaflets, newspaper printing, bill posting, etc.); announcements present the survey, ask for the participation and re-direct respondent directly on the web-page where the questionnaire is; otherwise they provide a point of contact. In this way a list of e-mail addresses can be built.

Invitations (banner, mail): you can use advertising banners published in web pages. These banners can be viewed in every moment and by every visitors of the page or only in prefixed times or by a fixed number of the visitors of the web-page. The efficiency of banner is higher than the invitations tag's one. Banners can be usually used in specific surveys: indeed they can be published in web-sites (or in their pages) that have a well defined contents; in this way a well selected and defined groups of people can detected.

Ipertextual links: these links are published on the web-pages, in search engines, etc. Links send visitors of the page to the questionnaire and can be underlined with graphic representations.

Pop-up surveys: questionnaire appears directly on the screen to the respondent. The web surfing is detected: the surfer request a specific page and, if he's selected from a specific mechanism, the system provides to open, in front of the asked page, a new window with the invitation to participate to the survey. A random sample of the web-site visitors can be selected.

Hijack or Browsing Intercept Software: it is similar to the Pop-up approach, but it is a much intrusive way to contact respondents. The reading of web-documents and the web-surfing is detected from a specific software. A sample of visitors, chosen with pre-defined criteria (like the Pop-up ones), is actively intercepted and re-directed, instead of to the asked page, to a page that contains the questionnaire or purposes survey (Pfleiderer & Gente, 1998). In this page there would be a "decline button", that allows to switch off the page with questionnaire and to go to the asked page.

Opinion Centre: some of the most relevant opinion centre are the AOL (America On Line) and the DMS (Digital Marketing Services), but there are a lot of other opinion centre (www.opinionplace.com). Surfers (through, in example, the use of banners) are invited to visit the "opinion place", a web-site where each person can give his availability to taking part in surveys. The visitor compiles a short questionnaire to provide some important information. Once given the consent to be involved, respondent, without knowing what is the survey topic, decides when ask to the questionnaire. Usually to a respondent is proposed not more than one survey a week.

Panel: respondents are recruited to participate to one ore more surveys or to a longitudinal surveys. The contact can be made trough phone, e-mail, mail or in other ways. Consensus is asked to be inserted in a group which, in the future, will be involved in participating in a certain number of surveys. Thanks to some relevant characteristics provided from the subject and to the stratification of the panel units it is possible to build a profile of the different kind of people. These profiles are useful to identify units that will be involved in the research. A particular way of recruitment is called "phone invitation": the units are contacted by phone and it is necessary to verify that these would be respondent of the target population of panel. After this how register on line must be explained to the subject and the URL that he can use for the registration in panel list must be given him.

E-mail lists: these lists are in large databases made for different purposes. Here there is the mean types of addresses lists:

- Lists of persons or enterprises that given consensus to be contacted again, usually requesting for some services. An example is the SSI (Survey Sampling Inc.). The difference from panel list is that a subject don't know to be included in a e-mail list.
- List of units registered in administrative archives (in example, business registers or social security databases). The administrative databases are build up to collecting some basic information for administrative purposes,

nevertheless other kinds of data, like e-mail address, are requested. The information available in these lists can be often missing or not updated. In some country the regulation about privacy don't allow institutes that manage an archive the diffusion of the information.

- List built with the e-mail addresses collected on the web-site pages (Harvested addresses). The addresses may be registered automatically by a specific software, or can be manually obtained by webmasters. In some cases national public lists of e-mail addresses are available to be used in the future (Batagelj, 1999). In other cases there are, on web-site pages, registration forms that allow to collect addresses and agree the use of certain services.
- List of customers registered by fidelity cards, subscriptions, and so on; usually at the time of the registration some information like e-mail addresses and others are asked. These information allow to process skills of the people who provided their e-mail address.
- Private lists (directories of customers, of furnishers, of personnel's member, ...). These lists are restricted, however, to the observation of definite and clear groups.
- List of members of Internet groups: discussion groups (Witmer et al., 1999), newsgroups, Usenet, discussion lists, Interest Groups,.... In these cases one can use software like Listserv o ListBot.

Snowballing sample criterion. At times a method to grow up e-mail addresses lists or lists of contacts is based on the snowballing sample criterion: people who are already inserted in the list signalise other persons who are interested to be inserted in. This approach is principally used in processing small population, when its members are difficult to contact, but often they are in contact with themselves.

4.2 Probability and Non-probability Sampling in Web surveys

As a consequence of the way of choosing respondents of web-based survey it is possible (or suitable) use the probability rather than the non-probability approach.

To select a probability sample is necessary to have a list of addresses or a system of contact with potential respondents (the access to a web-site, the visit to a specific web-page, ...) that allows the random selection of respondents.

Using a classification by Couper (2000) it is possible to consider situations where probabilistic rather than non-probabilistic methods are achievable used for recruiting sample.

The **probabilistic methods** can be classified in four different categories.

1. **Intercept surveys**: surveys based on the interception of visitors of particular web-sites. This category comprises Pop-up surveys or Hijack-based surveys. The sample frame include all visitors of a web-site or all the readers of a web-page. Among the visitors is selected a random sample. To this sample is asked to

participate to the survey. The mechanism of selection of sample can vary from case to case.

In Pop-up surveys, for example, visitors can be selected at fixed times or each *n* visitors that entered in web-site (or in web-page), an so on. This method of sample selection is frequently used and generally allows to achieve better results than the ones obtained with the help of a fixed banner (Malhotra 1999).

The presence of cookies in personal computers used by respondents, besides, provides for non contacting the same people twice. There are, anyway, a lot of problems: some people erase often cookies and so the risk to contact them twice is high (this is the case when one can use two or more PCs to access in Internet); but there are, also, people who share the same machine; and so on. The Pop-up approach is often used in survey that want to explain which kind of audience visits certain web-sites or for controlled experiment finalized to test on-line experiences.

In some situation you can use incentives. The debate on the use of an incentive is open; for instance Comley (2000) noticed that with this kind of collaboration encouragements there is the risk to attract fraudulent respondents that, to be contacted again, erase frequently cookies from their personal computers. The percentage of collaboration to the Pop-up survey proposed can vary from 2 to 15% (Fletcher Research); Virtual Surveys talk, in regard of its survey, of a agreement rate of about 25%.

The Hijack criterions to select units are similar to the Pop-up ones and are probabilistic. This kind of approach is not recommended: it is, in the opinion of researchers, excessively intrusive and risk to grow up the respondents' future resistance in the context of other surveys. The results that this approach consent to obtain are encouraging: in the UK, a BBC research about its audience reach a response rate of about 80%. The success of the survey was explained with the excellent relationship between the public broadcaster and its audience.

2. **E-mail directories**. There are a lot of different lists (as explained before). The quality of surveys depends by the coverage of these lists. The better situation for a probability survey is having a population characterized by an universal diffusion of the Internet connections, a compete addresses list without duplication and with one e-mail address for each unit. In the demographic populations and in the enterprises ones the degree of coverage of the list is a remarkable problem. It is very difficult to identify a list from which you can select a sampling frame that is a good approximation of the population. The problem is very critical and must be seriously tackled time to time. Less relevant is, instead, the problem of the coverage in the case of small and closed populations (for example, the students of one faculty or the employees of a great enterprise). In these cases you can obtain very sight samples because usually, in the same archives, a lot of other information about students/employees are registered.

3. **Pre-recruited panels**. These panels are based on contacting individuals to insert them, then, in one list of contacts. The quality of statistical survey depends from the respondents selection criteria and from the used frame. For example, if the probability sample (like the one selected with the Random Digital Dialing

methods, RDD) is adequate to the target population, target units can be directly selected. A relevant problem comes from the availability of the participants. The e-mail addresses, indeed, are frequently changed. Moreover technical issues limit the availability of potential respondents (Alvarez and VanBeselaere). It is, hence, improbable that the list of panel maintains a good quality from times to times. The response rate, besides, tends grow down in relation with the ramble of belongings to the panel. For this reason, a panel must be updated and modified from times to times.

The response rate, when recruiting is done by phone, is quite high: with a sample from a panel database recruited through telephone, the response rate is about 30-40%. However the response rate is not justified by the high costs of phone contact.

4. **Mixed-mode designs**. These surveys expect the integrated use of the different techniques seen before: the web-based survey is only one way to participate to a survey, a alternative to phone o other kind of surveys.

We discussed about the opportunity, in most situations, of using non-probability sampling methods. These techniques don't use probabilistic methods of selection. Therefore the sampling results can't be extended, through the statistic techniques, to a larger population: the inference step can be very difficult to be applied.

A classification **of non-probability methods** is built by Couper (2000) and provide three different typologies of techniques.

1. **Web surveys as entertainment**: this kind of surveys don't has scientific finality, but, in general, is proposed to visitors of a web-site. Visitors decide autonomously if participate to surveys or not. Finalities of this kind of surveys (very diffused in web) is usually to make web-sites contents more attractive. There are a lot of web-sites that ask their visitors to evaluate a movie, to signal the favourite song or singer, to chose the most loved public person, an so on.
2. **Unrestricted self-selected surveys**: through invitations (like announces, invitation tags, banners, iper-textual links, ...) published on the pages of a web-site it is possible to involve respondents in surveys. Only visitors of a specific web-site that give actively the consensus to participate to the survey can be part of the sample.
3. **Volunteer opt-in panels**: as in Opinion Centre, some people offer themselves voluntarily to be involved in various kinds of surveys. [1] Panelists are invited to go to a defined web-page where some fundamental information are required (between these, the e-mail address). All information are registered in a database that make possible, in a second time, the extraction of an opportune sample on which the research can be made. Even if this system of selection is non-

[1] Two examples of well known volunteer panels are Harris Interactive (http://vr.harrispollonline.com/register/main.asp) and Greenfield Online (http://greenfieldonline.com).

probabilistic, it is probable to obtain a selection of representative sample more than using others non-probability selection methods.

5 Concluding comments on Sampling

For web-based surveys it is possible to use the sampling techniques used in traditional ones. This kind of survey offers many advantages (e.g. low costs, timeliness, large territorial coverage at relatively low costs, etc.), but there are some important themes connected with the sampling to be looked into.

Given the special ways of obtaining contacts in web-surveys, one of the main problems in choosing the correct sampling method is connected with the availability of sample frames. Often no sample frame is available. When these are available, however, they are often unsuitable for the context, incomplete, inaccurate or no representative sample can be obtained from them. The consequence is that results obtained by a survey based on inaccurate frames cannot be extended to the entire population.

Another important problem concerning sample frames (when available) is the coverage error, or the differences between sample frames used and target population. Nua Ltd.[2] estimated (in February 2002) that only 8.96% of the population worldwide (544.2 million people) has access to the Internet. In North America the percentage grows up to 58.8 % (164.4 million people). In recent years the increase in web connections has grown up considerably, but these data demonstrate that it is still impossible to contact the entire generic population of a country.

Another critical point for web-based surveys is the difficulty in using probability sampling methods, considering the difficulty in obtaining lists and their unreliable representativeness. There are two ways to partially solve these difficulties: reduce the target population to people who have access to the Internet or use an alternative method of contacting respondents (e.g. the phone) in order to reach a larger group of respondents involved.

Currently, in web-based surveys a non-probabilistic sampling method of selection is usually used. It is often not possible to identify members of target population nor use a probability method of selection. If a random system to generate e-mail addresses is used, most of the addresses generated might be inexistent, invalid or unused. Moreover, spamming is forbidden and many e-mail servers provide for the automatic deletion of spam. All these factors contribute to an exponential rise in the risk of obtaining a non-representative sample of the population, a very different one from that selected by the sampling plan. To resolve this problem, a fundamental mixed mode approach is often used (Couper, 2000). Respondents can, for example, be contacted by phone with the objective of persuading them to participate in Internet surveys.

Another serious problem in drawing up sampling strategies is the scarcity and extreme approximation of information available about non-respondents.

Even if an appropriate sample can be obtained using a suitable list, it is still important to keep another aspect in mind, that is the identity of the respondent: this also being a

---

[2] http://www.nua.ie/surveys/how_many_online/index.html

problem with other kinds of surveys. web-based surveys can guarantee neither that respondents to a questionnaire are the same people selected for the sample, nor that the questionnaire has been filled in without consulting other people.

Therefore in web-based surveys sampling another important problem must not be ignored, i.e. the presence of professional respondents, or rather those who participate in a large number of surveys. It was estimated that in the United Kingdom the average housewife participates in about 11 phone interviews a year. A study conducted in the United States showed that 50% of all answers to surveys comes from only 4% of the population. The importance of professional respondents becomes clearer given that there are some web-sites (e.g. www.surveys4money.com, www.doughstrett.com, www.rewardsites.com) showing surfers where to find surveys where they can earn money or win a prize.

There are two risks regarding professional respondents: firstly they may become influenced by some topics, as they participate in numerous surveys, so they risk becoming non-representative of the real target population; secondly one is the honesty of answers, if the only reason for participating is to earn the incentive.

To contain this phenomenon, some research institutes (e.g. AOL/Opinion Place and the Millward Brown IntelliQuest Tech Panel) have strict rules about the frequency people are asked to participate on a panel. In spite of this, it is impossible for these panel companies to check if their respondents are registered with other panels. A study of SSI observed that about 90% of panel lists wanted to participate in at least one survey per week, and 60% in more than only one. Another study conducted by Wilke et al. (2000) demonstrated that 46% of people inserted in ACNielsen's BASES panel (people who participated in more than 20 tests per year) wanted to be interviewed more frequently.

It is clear, therefore from what has already been mentioned, how important it is to select an appropriate sampling technique, to firstly investigate thoroughly the study of the characteristics of people with access to the web, to go beyond the essential and basic information available e.g. the number of visitors to web-sites. The presence of mirror-sites, the possibility of hiding your identity, special frames or mechanical instruments can falsify statistical data on the number of visits (Smith, 1998 and Foan & Read, 1999).

When compiling a sample it must be remembered that web-surfers have specific social, demographical and economical characteristics. This is why McDonald (1999) pointed out that a sample of people who use the Internet is usually representative of Internet users only. In Coomber (1997) it is also stated the bias present in web-based surveys is caused by particular web-surfer characteristics, a very varied set of people compared with a set from a generic population. People who use the web usually come from a higher social and demographical background, they are white males who are relatively wealthy and well educated, with a higher level of education and can access high-tech resources unlike a generic population (Nielsen & CommercNet, 1995; Kehoe & Pitkow, 1996).

It is almost impossible to contact groups of a well defined population, e.g. old people, with low incomes or with a low level of education. For a more thorough and in-depth on the study of people who don't use the web frequently, see Coffey & Johnson (1998). It is also true (Coomber), thanks to the ever more widespread use of new technologies and ever faster Internet connections, the bias typical of the web-based survey sampling is destined to diminish in the near future.

Part II – Inference

1 Introduction to Inference

In web sample surveys, as in any sample survey, when studying a phenomenon not all the elements of the population involved (target population) are taken into consideration, but only a subgroup of respondents who represent the sample.

The data therefore regards the sample exclusively. Statistical inference has the aim of retracing the relevant data on the sample to the whole population being studied. By means of inference techniques one of more unknown parameters (or parameter functions) can be estimated relative to the universal population. In these cases inference brings to "parametric estimate".

It is possible to work in this way only if the subgroup in the survey is a statistically significant sample. Only in this situation is it possible to estimate the size (e. g. the average) relative to a defined population knowing the average sample and applying the statistical theory of the estimate. To formulate hypotheses on the parameters being studied (i.e. the average) sampling data is therefore used.

The average sample is a value (also called "estimate") of a specific casual variable (defined as "estimator") who represent the series of all the values the estimate can assume. Corresponding to the selection of a certain sample (i.e. the n-upla sampling is fixed), the estimator takes on a determined value, i.e. the estimate.

It can be seen by statistical inference that an exact estimate of the population parameters is not achieved, but a range of values is defined.

Statistical inference is applicable only in the case where it is possible to create a casual sample or when the distribution form of the casual variable who characterizes the phenomenon being studied is known. The theory that puts statistical criteria in order to transfer from sample data to population parameters is known as *estimation theory*.

The fact that sample selection in a casual way leaves the research to define the level of reliability of the estimate. In particular it becomes obvious that the real value of the unknown parameter belongs to a suitable set of sub series of values. These series generally coincide with specific intervals.

Give that, in general, the unknown distribution of the phenomenon being studied depends on the parameter that is to be estimated (as occurs with the average for normal distribution) this means that the real distribution belongs to a subset of the family including all the possible distributions of the casual variable.

Based on the principles of probability calculations and Inference Statistics, it is possible to construct the confidence intervals (Cap. 3) around the value to be estimated. This indicates with what probability the real value referring to the population is to be found near the estimated value. The problems of estimation are therefore linked to the definition of a possible subset of parametric values, from those probable (and therefore of a particular distribution family from those possible) and with the quantification of probability that could contain the real value referring to the universe being studied.

If, especially, the estimate is referred to one or more parameters it is called "parametric", otherwise if it refers to the distribution phenomenon it is called "functional".

Furthermore the techniques of inference statistics make it possible to use special statistical tests to verify the hypothesis (Cap. 4): it is therefore possible to evaluate when the estimate value can be, a more or less reliable approximation of the unknown parameter under study.

Furthermore, the result obtained is subject to the possibility of errors. What is interesting is that it is possible to qualify the importance of the probability error.

The fact is that in web surveys it is often not possible to apply probability sampling techniques. However it represents a severe limit on the use of statistical inference and leaves a series of problems still unsolved.

These guidelines propose ways of solving these problems that are mostly connected to sampling design.

After looking at some fundamental definitions regarding inference, a systematic analysis (using the most common classifications) will be carried out on the main types of error that can cause bias in sample results and their weight and some strategies are proposed to limit the effect on data quality. With this aim, two common approaches are introduced. These are still subject to extensive research and are based on *balancing score* and *propensity score*.

2 Concepts and definitions

To understand the characteristics of statistical inference the following definitions must be kept in mind and following concepts must be clear.

For example, the case is taken into consideration where it is necessary to estimate, for the reference population, the arithmetic average relative to a phenomenon (X) that presents normal distribution. If a casual sample of n individuals is extracted, the possible results can be represented by the multiple casual variable $(X_1, X_2, \ldots, X_n)$[3] which, once the n subjects of the sample has been extracted, take on precise values $(x_1, x_2, \ldots, x_n)$. It is then possible, from the results of the casual sample, to estimate the value of the average referred to the entire population.

Whit this aim, it is necessary to introduce the concept of "estimator".

- Estimator: this is a special function, for example $g(x_1, x_2, \ldots, x_n)$, of the sample's observed values used to estimate the unknown value of a certain parameter θ or of its functions $\tau(\theta)$. The estimator of $\tau(\theta)$, that can be called T, is a special casual variable, a function of the casual variables $(X_1, X_2, \ldots, X_n)$ which represent all the possible signs of a certain phenomenon being studied

[3] For further reading on casual variables, see any manual on statistical inference.

on the units making up the sample: $T = g(X_1, X_2, \ldots, X_n)$. The fact that the estimator is a function of the unit extracted means, in the case when a second sample is extracted, the value assumed bu the estimator could be different. For example, if when a sample is extracted, a certain variable is seen and using the average sample data are synthesized the estimate of the average value of the entire populations obtained. Extracting, using the same procedure, another sample of the same target population, the same estimator (the average sample) will probably give a different estimate. There is a series of methods that can be used to define estimators of a parameter: method of moments, maximum likelihood, minimum chi-square, and so on. Each estimator will be characterised by specific properties (correctness, efficiency, consistence, sufficiency) making it possible to choose the best one for the case being studied. For a more detailed study of punctual estimators and their properties see traditional manuals on Inference Statistics (see Further Reading).

- Estimate: Once the sample has been extracted the estimator, a function of sample results, takes on a well defined value which is defined as "estimate". E.g. the average of a sample is in fact an estimate. Once the *n* unit has been extracted and the x_i ($i = 1, 2, \ldots, n$) value which characterized them have been examined, it is possible to go from the concept of estimator to the concept of "estimate" of $\tau(\theta)$, which can be indicated with "t" and is such that: $t = g(x_1, x_2, \ldots, x_n)$.
- Parameters: When the unknown size of the population relative to the phenomenon being surveyed has to be estimated. E.g. the average unknown population from which the sample has been extracted is a parameter. Estimation operations can be carried out when dealing with a casual sample or when the distribution form of the casual variable which characterizes the phenomenon under consideration is known.

3 Confidence Intervals

Statistical inference theory provides the estimation of the unknown parameter by pinpointing a set of values ("confidence intervals") within which a given probability exists α (*alpha*) which includes the real value of the population parameter.

The definition of such a set of plausible values is called an "interval estimate" as this is often an interval. The result obtained is a "confidence region" (Azzalini, 2001).

Their identification is based on the hypothesis that, once observation has been carried out, the parametrical set (i.e. the set containing all the values which the unknown parameter could assume) contains, at a confidence level equal to $(1 - \alpha)*100$, the parameter being estimated. Therefore, the probability that given *n* as sample values, the estimator is contained within the confidence interval and that the real value of the parameter is contained in itself is equal to $1 - \alpha$.

Formally, to define a confidence interval it is first necessary to define two function of the type:

$T = g_1(X_1, X_2, \ldots, X_n)$ and

$V = g_2 (X_1, X_2, \ldots, X_n)$;

These are two statistics (and therefore casual variables as they function as casual variables) such that, if $T \leq V$:

$P_\theta [T < \tau(\theta) < V] = 1 - \alpha$, $(0 \leq \alpha \leq 1)$.

Otherwise the probability that, given *n* sample values, the estimator is included between T and V (the two statistics) is equal to $1 - \alpha$.

T is defined as lower extreme of the confidential interval for $\tau(\theta)$; V is analogously the upper extreme for $\tau(\theta)$.

The interval (T, V) is defined "confidence interval" at level $(1 - \alpha)*100$ for the function of the parameter θ, $\tau(\theta)$. $(1 - \alpha)$ represents, especially, the probability that the real value of the parameter (referred to the universe) is contained in the confidence interval identified with T and V.

Once the sample has been extracted and the observations on it made, x_i $(i = 1, 2, \ldots, n)$, T and V will be functions of sample values and therefore can be indicated respectively with the letters t and v:

$t = g_1 (x_1, x_2, \ldots, x_n)$ and

$v = g_2 (x_1, x_2, \ldots, x_n)$;

the interval (t, v) will again be a confidence interval for $\tau(\theta)$ characterized at the same level of confidence of the preceding (T, V) as $(1 - \alpha)$.

There can be cases where the distribution of the phenomenon being studied is one of the parameters to be estimated.

For example, in the case of a phenomenon which presents Normal type distribution it may be required to estimate the "average" parameter (μ). It is known that Normal distribution depends on two parameters: the average (μ) and the mean square error (σ).

In the case where distribution is a function of the type:

$Q = h (X_1, X_2, \ldots, X_n; \theta)$

that does not depend on θ, Q is defined as "pivotal quantity".

The pivotal quantity is used to construct the confidence interval.

Given a fixed level of confidence $(1 - \alpha)$, taking into consideration the functions Q (described above), $Q_1 = h_1 (X_1, X_2, \ldots, X_n; \theta)$ and $Q_2 = h_2 (X_1, X_2, \ldots, X_n; \theta)$ and observing sample values, an infinite couple of values will exist $(q_1 < q_2)$ such that:

$P (q_1 < Q < q_2) = (1 - \alpha)$.

From the infinite couples (and therefore from infinite intervals identified by the two extremes, the one that minimizes the difference $(q_1 - q_2)$, i.e. the width of the confidence interval, will be chosen. This is because for this type of interval the probability of including false values (or not probable) in the parameter is minimal.

Having chosen the couple (q_1, q_2) to obtain the confidence interval for $\tau(\theta)$ it is necessary to rewrite (inverting them) the inequality:

$q_1 < Q (X_1, X_2, \ldots, X_n; \theta) < q_2$

in a way to isolate, the centre, the parameter to estimate, obtaining an inequality which it includes of the type:

$T < \tau(\theta) < V$;

The inversion can be made on each different sample that can be extracted.

However, the functions T and V must be identified so that the inequality

$q_1 < Q(X_1, X_2, \ldots, X_n; \theta) < q_2$ and

$T(X_1, X_2, \ldots, X_n) < \tau(\theta) < V(X_1, X_2, \ldots, X_n)$

are equivalent, i.e. they have the probability of containing the term $\tau(\theta)$ equal to $(1 - \alpha)$.

For example, in the case where phenomenon X is distributed normally, in order to identify the confidence interval relative to the average it is possible to use the pivotal quantity:

$$Q = \frac{(\overline{X} - \mu)}{\sqrt{\frac{1}{n}}}$$

which is pivotal because it has normal standardized distribution and therefore is independent of the average. It is known that infinite couples exist (q_1, q_2) for which:

$P(q_1 < Q < q_2) = (1 - \alpha)$

is valid. Substituting the pivotal quantity Q in inequality and proceeding with inversion to isolate the average μ, an interval is obtained (for the average) indicated as such:

$$\overline{X} - q_2\sqrt{\frac{1}{n}} < \mu < \overline{X} - q_1\sqrt{\frac{1}{n}}$$

which is equivalent to:

$T < \mu < V$;

Both inequalities therefore are created with the probability $(1 - \alpha)$.

It must be kept in mind that in the case of Normal distribution, an interval which minimizes (to an equal level of confidence) its own size, is symmetrical in respect of the average, i.e. type $q_1 = - q_2$.

Therefore q_2 is defined in the following way:

$q_2 = z_{\left(1-\frac{\alpha}{2}\right)}$ [4];

in the case of Normal distribution, a confidence interval for the sample average can be identified as follows:

$$\overline{X} - z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{1}{n}} < \mu < \overline{X} + z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{1}{n}} \;;$$

---

[4] The identification of q_2 is based on equal division of probability α at the two extremis (left and right) of the confidence interval.

In the more general case of a confidence interval for the average (not referred to Normal standardized casual variable) the formula used is as follows:

$$\overline{X} - z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\sigma^2}{n}} < \mu < \overline{X} + z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\sigma^2}{n}}$$ (in the case when the value σ^2 is known), or

$$\overline{X} - t_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{S^2}{n}} < \mu < \overline{X} + t_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{S^2}{n}}$$ (when the value of σ^2 is not unknown).

It must be remembered that the pivotal quantity

$$T = \frac{\left(\overline{X} - \mu\right)}{\sqrt{\frac{S^2}{n}}},$$

with the increase of n, tends to take a Normal standardized casual variable. Therefore it is possible to substitute the value $t_{\left(1-\frac{\alpha}{2}\right)}$ with $z_{\left(1-\frac{\alpha}{2}\right)}$.[5]

In the same way it is possible to construct confidence intervals even for other parameters of interest.

For example, in the case of variance the pivotal quantity is considered $U = \sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^2}$, the distribution of which (like chi-square, with n degree of freedom) does not depend on σ^2.

Likewise for the average, from the tables relative to chi-square distribution it is possible to obtain two values ($q_1 < q_2$) so that:

P ($q_1 < U < q_2$) = $(1 - \alpha)$.

Normally to determinate the values q_1 and q_2 the following is used:

P ($U \leq q_1$) = ($\alpha/2$) = P ($U \geq q_2$).

Applying inequality inversion (to isolate the parameter involved):

$$q_1 < \sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^2} < q_2 ,$$

a confidence interval is obtained for σ^2:

$$\frac{1}{q_2}\sum_{i=1}^{n}(X_i - \mu)^2 < \sigma^2 < \frac{1}{q_1}\sum_{i=1}^{n}(X_i - \mu)^2$$

which is equal to. Both inequalities are created with a probability $(1 - \alpha)$ of inclusion unknown value of σ^2.

---

[5] For quite large n it is usually understood the value of the size of the sample is greater or equal to thirty.

This, in the same way as the first, has a probability equal to (1 – α) of including the unknown value of σ^2.

4 Statistical hypothesis testing

In the area of statistical hypothesis testing, as mentioned before, the researcher generally has a theory which translated into statistical language becomes the affirmation of the values a parameter can assume from those possible. Likewise, the affirmation can be referred to the distribution of the phenomenon: it is necessary to define the subset of distribution contained within the possible distributions; this subset corresponds to the values that can be assumed by an unknown parameter. Statistical inference techniques are usually used to confirm or otherwise, on the basis of sample results, the affirmation made.

Generally, the affirmation to be verified is defined as "null hypothesis" and is conventionally indicated by H_0. The null hypothesis can refer to the value of a parameter (average, variance, etc.), the set of values a parameter can assume, the equality between two unknown parameters and so on.

However, generally in the case of the value of an unknown parameter of the universe population two distinct hypotheses are formulated: the null hypothesis H_0, and the alternative hypothesis H_1. The aim is to establish which of the two is more probable (or, however, preferable), on the base of sample result.

Apart from simple hypothesis which for example fixes two hypothetical values of a parameter, it is possible to also consider composite hypotheses. This happen, for example, when the alternative hypothesis does not regard a single value of the parameter, but is the negation of null hypothesis: in these cases, therefore, the alternative hypothesis can be both bi and mono-directional (to be examined later). In both situation the alternative hypothesis does not specify the precise value of the parameter but a range of several values. There are also cases in which even the null hypothesis takes in consideration a set of values and not one specific value of the parameter.

When the hypothesis H_0 takes in consideration a certain value of the parameter, the shape of the distribution of the phenomenon is unequivocally identified, in the case in which the null hypothesis is composite, a family of possible distributions is identified.

Depending on the type of hypothesis to verify there are a series of statistical tests available; e.g.: the tests for the value of an average, for the difference between two averages, for the difference between two proportions, for the equality of two variances, and so on.

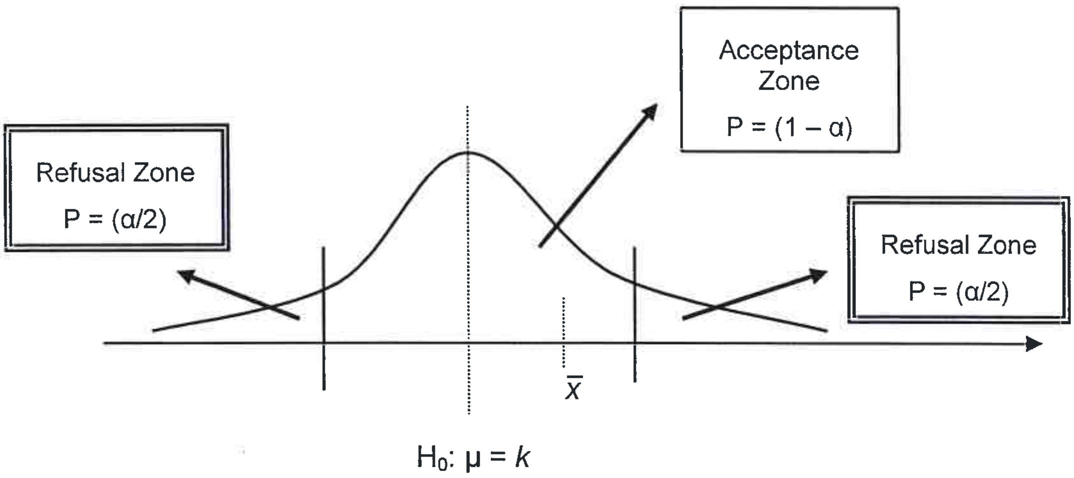
See further reading for more details on statistical inference manuals.

It is possible to apply one of the tests regarding the average value of a variable observed (e.g. X) on a defined population. The hypothesis H_0 to be verified refers to one of the possible numerical values the average can assume; one example of hypothesis H_0 could be the following:

$\mu_X = k$.

The equality of two average values can be established by a test, a certain level of reliability is fixed (α), if the null hypothesis ($\mu_X = k$) can be accepted or must be refused. The conclusion is reached using the sample data collected: if the sample average ($\bar{x}$) is sufficiently near the value to be tested (hypothesis H_0, therefore k), it can be concluded that the hypothesis can be proved by sample data and can be accepted at the pre-established level of reliability α. If the distance between the two values ($\bar{x}$ e k) are excessively high, it is necessary to accept the alternative hypothesis H_1: this means the observed difference is significant and the null hypothesis cannot be accepted.

In this as in other test it is possible to draw up a graphic representation of the distribution of the phenomenon in which refusal zone and acceptance zone of the same test is highlighted. In the case of the average, it is presumed that the phenomenon is normally distributed and that there is a zone of refusal both to the right and to the left of the value fixed by the hypothesis (bi-directional test). This means the hypothesis will be refused both for excessively high values for the sample average as well as for excessively low values.



If the sample average $\bar{x}$ falls in the refusal zone (therefore excessively far from the value indicated by the hypothesis H_0), the null hypothesis is refused, otherwise (as shown in the graph above) it must be accepted.

It is already been mentioned that test exists to verify both mono-directional e bi-directional hypothesis. In the former (for example the equality test on two proportion) the refusal zone corresponds to only one of the extremes of the distribution of the phenomenon being studied (mono-directional test). In the latter (for example, as already seen, the average value test) the refusal zone of the hypothesis will have values which are excessively far from that indicated in the hypothesis H_0 in both direction (bi-directional tests).

In carrying up a test to verify the hypotheses two types of error can be encountered:

- The error of the first type is verify when the sample value falls in the refusal zone, even though the null hypothesis is true. The possibility of making a first type error is equal to alpha (α) and is usually fixed at a level equal to 0.05. In the case of the hypothesis of a sample average, the test for which is fixed alpha equal to 0.05 indicate that, whenever hypothesis H_0 is true, the difference between the sample value and the value fixed by the hypothesis is considered significant in 5% of the test. On the basis of sampling data, therefore, there is a probability of refusing the true hypothesis, i.e. a 5% risk of making an error. This is why α also indicate the significance of the test, which in this case con be considered equal to 5%.

- The error of the second type is when there is a null hypothesis, although false, it is not refused (i.e. when the sample value is in the acceptance zone). The probability of making an error of the second type is indicated by *beta* (β). Fixing, for example, $\beta = 0,3$ indicates that in the 30% of the cases using a test the null hypothesis is accepted even if it is not true.

The power of the test is indicated by 1-β and indicates the probability that the null hypothesis is refused when it is false.

The following table summarizes what has been explained.

		Null Hypothesis (H_0)	
		True	False
Decision	Accept H_0	$1-\alpha$ (correct decision)	β (2nd type error)
	Refuse H_0	α (1st type error; test significance)	$1-\beta$ (correct decision; power of test)

It should be especially noted that when it is not possible to refuse the null hypothesis it is neither possible to conclude that it should be accepted. For Example, in the case of a test on difference between two proportions, if the test result show a non-refusal of the null hypothesis, it is possible to (NCES, 2002)...

- ...confirm that none conclusion has been reached;
- ...indicate that no significance difference were met;
- ...underline the fact that the conclusion could be due to reduced sample size or to a high value of standard error, if this is evident or other research shows a significance difference;
- ...indicate that, if the estimate is not reliable, the standard error is so high that the large observed differences are not considered significant;
- ...highlight the fact that, if a significant statistical difference for the target population is observed, but similar differences of subgroup relative to the same size are associated with low sample numbers or high standard errors, these differences do not result as statistically significant.

There is a large quantity of statistical software (SPSS, SAS, ...) for analyzing data (also very complex data) obtained from sample surveys to verify the hypothesis by means of statistical tests.

However, as it can be noted, the statistical tests used in inference techniques are based on probabilistic principles and require, therefore, a probabilistic sampling procedure. In fact, if any sample of subjects interviewed on the web were probabilistic there will no problems in applying the principle of inference statistics to the results following in data collection. However in the applicative field this rarely happens as almost all the samples for web surveys are created on a self-selection bases and risk (depending on the target population) of not being representative and, therefore, only rarely are (or can be) built with probabilistic selection method.

The consequence is that the conclusions of surveys carried out on Internet in general cannot be extended to a population bigger than that making up the sample itself. In any case it is necessary refer in detail, together with the results, all the aspect regarding the way in which the sample was selected.

5 Errors

If one of the principal objectives of the sample surveys (based on web use or other types) is the estimate of parameter regarding parameters population, it is not possible to ignore that fact that, for however accurate the choice of the sample can be (i.e. for as representative as it can be), sample data generally means reaching only an approximate estimate of the characteristics of the parameters being studied and can be a long way from the real value in a greater or lesser degree.

It has already been pointed out that estimation theory requires, as a prerequisite for its statistical variability, that the chosen sample is probabilistic. Furthermore the estimate theory provides results which allow the researcher to quantify the identity of the error. The researcher can then apply methodological criteria of estimation theory to define the size of error which is deemed acceptable. The type of error mentioned here, which is found in estimation theory, is called sampling error.

The researcher must, however, pay attention as this is not the only type of error which can affect the results of a survey. The other types of error do not originate in sampling and can be hidden at different stages of the survey, they are difficult to deal with and often depend on the correctness and precision of the survey methodology.

It is possible to define an error called *Total Error* which includes each error which causes the distance of the estimated value from the datum relative to the entire population. Total error is the result of the contribution of two different types of error: *Sampling* and *Non-Sampling error*.

5.1 The components of Total Errors in traditional surveys

One possible classification of different types of error was introduced by Kish (1967) and was later developed by Bethlehem (1999).

Total Errors are considered to be the result of the presence of two different types of error, *sampling error* and *non-sampling error*.

Sampling error. This is due to the fact that the data collected does not cover the whole target population but only a sub total (i.e. the sample).

The relevance of sampling errors falls with the increase of sample size and disappears in the case where the sample covers the entire target population.

Sampling Error can, in turn, be subdivided into:

1. *Estimation Error*: this originates in random sample selection and is closely linked to statistical inference theory. The resulting problem can be solved or reduced by modifying the sampling plan; increasing the number of samples, using proportional probability selection with a suitably chosen auxiliary and so on.
2. *Selection Error*: this is not closely linked to the problems estimate inference but originates from the risk of choosing a wrong sample unit, i.e. one different from that expected. The case can be verify, for example, when the same individual is present more than once on the sampling list: on an e-mail list of addresses it frequently happens that an individual has more than one address. The same mistake can happen when the data declared by a panel member are not true. Selection error is quite difficult to avoid if a careful study of the sampling lists used is not carried out; the cleaning procedure is even more difficult in the field of web surveys, due to the fact it is almost impossible to connect the e-mail address with the person who owns it (to eliminate possible duplications), as well as to the tendency of users to frequently change e-mail addresses.

Non-sampling Error. This can only happen in cases when dealing with census surveys, i.e. when dealing with a survey regarding the whole population and not a limited sample unit.

Non-Sampling Error can in turn be subdivided into:

1. *Observation Error*: this regards the phases of data collection which can also be subdivided :
   - *Over-Coverage Error*: this happens when elements not belonging to the target population are included in the sample. If the aim of a survey proposed by a web page is to interview people interested in a specific *topic*, it is possible that the person is not really interested, but has opened the page by chance and has decided anyway to answer the questions. Likewise, if those members of a panel give false information when they register, then the choice of including these people in the sample on the basis of their profile could cause over-coverage errors, given that in reality these individuals are not part of the target population. However in web surveys it is more frequent to find under-coverage errors (see below).
   - *Measurement Error*: this causes discrepancies between the actual value and that which will be processed in the data base. This type of error can be met when a respondent makes a mistake in replying (by mistake or cannot reply, etc.), or when an untrue response is given or when the question is not clearly understood. In web surveys measurement error

risks having a quite serious importance: a series of studies has in fact faced delicate subject of the respondents' which on the web does not seems to be particularly guaranteed. Furthermore, in comparison with more traditional surveys (e.g. face-to-face), there is not the presence of an interviewer to clarify the meaning of not readily understood questions.

- *Processing Error*: this can happen during the data processing phase and is caused by transcription errors of the collected data. Generally processing errors are considered errors which originate at data-entry, however in the case of web surveys this type of error has virtually no importance seeing the data are inserted by the respondent and are automatically recorded in the data-base.

2. *Non-observation Error*: this occurs when it is not possible to observe a statistical unit or a series of statistical units. This can be subdivided into two categories:

   - *Under-Coverage*: this occurs when some elements of the target population do not appear in the *frame population*. Certain units (or representatives of a certain group of units) of the target population have no probability of being extracted. In these cases the results of sample survey can only be extended to the frame population and not to the whole target population. If the web survey exclusively regards individuals who have Internet access, this means potentially there is no risk of coverage error. Usually, however, the target population is much more generic and not only limited to Internet users. Under-coverage errors therefore take on an importance which is not easy to ignore. On the other hand, as the research carried out by Couper (2000) and Dillman & Bowker (2001) confirms, when the Internet is used as a selection mechanism, certain groups of individuals will certainly be under-represented in the sample. Coverage error, given that Internet users are very different from a generic population, especially regards well defined categories that use the Internet less frequently: the over-55s, people with a low level of education, predominantly female, ... (Bethlehem e Hoogendoorn, 2003).

   - *Non-Response Error*: this happens when the respondents taking part in the sample do not give the replies asked for. This is called *item non-response* if the respondent has not replied to some of the questions, and *unit non-response* if the person selected refuses to take part in the survey. The latter case can cause a serious reduction in the sample number with the risk of making the estimate less accurate. The problem can be solved by increasing the size of the sample. The presence of unit non-response is not a problem of great importance, given that it can be resulted by increasing the size of the sample. It has, however, a negative impact on the quality of the results, principally for two reason. On one hand the percentage of replying: the error is much more important when it is higher than the percentage of non-response. On the other hand the degree of difference between respondents and non-respondents, can lead to biased estimates of population parameters. It is evident that not only in this situation is it possible to find high non-response percentages, but also the distribution of unit non-response causes an over or under-representation of some groups of the target population when these groups behave in a different way from the characteristics being analyzed. There are different factors which can influence the non-response rate: the topic of the survey, the target population, the timing of the survey, the length of the questionnaire and so on. An accurate study of all these factors (perhaps with a pilot survey) is indispensable to limit high rates of non-response. Furthermore in web surveys the lack of an interviewer has an important relevance: it is not possible to clarify not clearly understood questions nor insist in trying to convince the individual to participate in the survey (as happens in face-to-face interviews). There are also technological reasons which influence the non-response rate. The respondents can suffer a series of technical problem such as slow connection, unreliable connection, high costs connection and a series of other problems which discourage the respondents from participating in the survey (Couper, 2001; Dillman & Bower, 2001; Fricker & Schonlau, 2002; Heerwegh & Loosveldt, 2002). The researcher has to try to minimise the non-response rate to reduce to the minimum any bias of the results. It is equally important to study the characteristics of those who refused to participate in the survey. (Pineau & Slotwiner, 2003).

The presence of over and under-coverage has been mentioned. Generally speaking these are called *Coverage Errors* when the *frame population* is different from the *target population*. This type of error often carries a considerable weight in web surveys, as not all the target population has web access and therefore cannot be part of the frame. Furthermore, it is possible that people who do have web access can be very different from the other units. This can cause serious bias of the results. Coverage errors cannot be estimated using statistical inference theory, but an in-depth study of them is extremely important given that they are directly linked to the quality of the survey.

5.2 The Error problem in Web surveys

In the specific area of web surveys, one special type of error is a relevant one: *Self-selection Error*: This term indicates the distortion which happens when sample of respondents is made up of volunteers instead of a random sample or one chosen by the researchers. To make sure the sample is representative, research must be carried out on a sample chosen according to pre-established criteria. However, in the great majority of web surveys this does not happen.

Even if a sample panel made up of volunteers recruited on web sites is used as a base the problem of self-selection is still there. Nor is it possible in these cases to talk about probabilistic selection methods, given that the volunteers in the sample panel are subject to self-selection.

It should already be observed that in web surveys non sampling errors are a quite important problems. Especially under-coverage problems affect web surveys results.

Non response errors can have a high impact on the results. The question is open and further studies are required.

5.3 How can Errors be reduced

The biggest challenge facing researchers is how to measure every potential bias and apply adjustment measures to correct or minimise them. There are various approaches which can be adopted to limit the effect of different types of error. Pineau & Slotwiner (2003) described potential impact of different type of error on survey results.

The first and very important step is to clean up the data and where possible, correct them, before final processing. However, with this operation it is only possible to prevent or limit errors of measurement, data development and over-coverage.

It has already been mentioned that the effect of sampling-error decreases with an increase in the number of samples and disappears when the sample covers the whole population of reference.

Regarding selection or under-coverage and errors caused by non response it is not worth correcting each single error found, but it is necessary to different types of approach, those used in estimate procedures, adjustment methods based on suitable weighting systems. With regard to the low-representativeness or to the under-coverage problems of sample, in web surveys the growing up of the sample size can reveal having the opposite effect and can also bring also to an higher bias of sampling results.

Problems caused by the presence of self-selection samples can be limited by using *cookies* [6]; in this way it is possible to stop a person from participating several times in the same survey. However, *cookies* cannot be relied on if the user cancels the *cookies* regularly; there could be further problems when two people share the same computer and you want to interview both.

To limit the effect of self-selection error it is also necessary to have a broad knowledge of the population being studied and the coverage that can be obtained from the sampling list used in the research. At the moment, however, it is not possible to statistically define the web using public. Connections to the Internet and e-mail addresses don't provide a standardized representation method that could represent every people who have a web access. Therefore it is impossible to obtain a probabilistic sample which represents the web using public in an statistically acceptable way. (Pineau & Slotwiner, 2003).

6 Probabilistic selection

The principle of Horvitz and Thompson (1952), universally known, maintains that unbiased estimates of the characteristics of the population can be obtained from sample surveys only when every unit of the population has the probability of being extracted

---

[6] *Cookie*: "this is a text file generated by the user's browser following a message sent from the web server in reply to the request for connection received. The cookie is memorized on the user's PC and contains information which positively identifies the user who contacted the Internet site generating the cookie. At each successive connection the web server ask the browser for the cookie previously memorized. If this is found the server can use the information contained for various aims: to offer personalized contents, trace behaviour profiles, update statistics on visits, etc.". (Diodati).

known and different from zero. Furthermore only in this case is it possible to calculate the accuracy of the estimate.

The principle of Horvitz and Thompson is similar to a probabilistic sampling plan as an essential pre-requisite for the possibility of estimating using inference statistics techniques, the population parameters and to evaluate the reliability of the estimates. Therefore only one casual selection makes it possible to apply the inference principles to sample data.

A significant problem linked to web surveys it is represented by the difficulty in carrying out probabilistic selection of the sample.

For Example, for entertainment surveys it is not possible to have any control over who answers the questionnaire or how many times a person participates in the survey. Likewise, there is no control over respondents when selecting samples of individuals using invitations such as pop-up or banner adverts on web-sites. Neither is it possible to know what the probability is of a statistical unit being intercepted by one or more invites or who accepts answering the questions in the survey. Both cases are non-probabilistic sample selection types: the sample is defined "self-selected web surveys" and the units are not chosen (or extracted) by the researcher. Nor is it possible to calculate the probability of those who will take part in the sample.

Even when using a panel made up of volunteers recruited through very popular web sites as a sampling basis the problem of self-selection still arises. In fact panel surveys are based on vast data-basis of individual who have voluntarily given permission to be involved in future surveys. From these data-bases suitable samples of the target population are selected or extracted. The procedure of choosing individuals who take part in the sample is made easier by the fact that the volunteers, when they register, provide a series of personal data which can prove extremely useful in the selection of units chosen for the survey. It is not possible, in these cases, either to consider these probabilistic selection methods, given that at the beginning the volunteers who form the panel were self-selected.

It should be pointed out that statistical inference methodology can be applied only when a probabilistic sampling technique is used (Kish, 1965; Cochran, 1977). Only in these conditions, is it possible to calculate the accuracy of the estimate.

In web surveys, however, it is often not possible (or it is not easy) to resort to the use of probabilistic sampling techniques: in the majority of cases, the sample in fact, results as being self-selected (i.e. *self-selected web surveys*). This type of unit selection does not allow the calculation of the probability a subject has of entering to take part in the sample.

Bethlehem e Hoogendoorn (2003), using the Dutch election results, compared the results of a survey carried out on a sample of self-selected electors with the results of another sample using a casually selected sample. Evaluating the distribution of seats in the Dutch parliament it was noted that notwithstanding all the attempts made to obtain a representative sample from a set of self-selected units, the results closest to reality were those based on units selected by probabilistic sampling.

Even when using convenience sampling a serious problem is the application of inference principles for the estimation of frequencies or percentages regarding the whole population. Unfortunately this problem link to the application to samples of these type is often ignored. (Schonlau et al., 2002, 2004).

The real problem regarding inference in web-surveys is the fact that inference is reliable only in the case when the requisite of probabilistic sampling is verified. The starting point, therefore, must be the selection of a casual sample within a group of units that have the probability of being extracted different of zero and known. For example, simple casual sampling would be an ideal situation to apply the techniques of inference statistics: in fact, all the units have the same probability (different from zero and known) of being extracted). However these prerequisite in web-based surveys, as seen in most cases, cannot be verified.

Danielsson (2004) as "non casual surveys" also included where the sample is extracted from a sub-population and not from the entire target population. This means that, referring to the entire target-population, the probability of inclusion (shown by π_i) is sometimes positive (sometimes known) and others equal to zero. Therefore a casual sample with non-response unit (or cannot be contacted as they have no Internet connection) can be interpreted as non-casual survey. Two groups of individuals can be identified (respondents and non respondents) and information is obtained only from the respondents. Web surveys are therefore an example of non-response surveys and consequently also non casual surveys, given that this is aimed only Internet users the rest of the population is missed. The only difference between non response surveys and web surveys is that the size of the sample is casual for the former and fixed for the latter. (Danielsson, 2004).

7 Non-representative and representative samples

Another problem must be emphasised, which is linked to the sampling technique used, i.e. a representative sample is obtained. In web surveys certain strata of the population (e.g. people with a low level education, older individuals, etc.) risk being under represented or not represented at all.

This means that, relevant to the target population, the statistical units being studied do not have the same probability of being extracted. For example, in a web survey on the movement of citizens within a certain area, it is possible to note that some special types of citizens (e.g. elderly people) are not intercepted in proportion to their percentage of the universe population (very few elderly people have web access); these particular categories in the sample either under-represented or not-represented at all. As a consequence it can be maintained the elderly people in a web survey have a lower probability of being extracted in comparison of either individual.

Casual extraction system is used for those who have web access. Furthermore, studies carried out in this area have shown the differences between populations who have or do not have Internet access involve numerous variables which could be the subject of further study.

7.1 How to obtain a representative sample

There are two ways to obtain a representative sample for web surveys, selected on the basis of probabilistic principles, as in Pineau & Slotwiner (2003):

- Use a multi-modal (telephone, post, personal, etc...) system to contact respondents: the individuals contacted are asked for their e-mail addresses, they are later contacted by e-mail to propose participation in the survey. Those who do not have a connection system (not e-mail) are interviewed using an alternative method.
- A representative selection by phone using the RDD (Random Digit Dialling) selection method. An Internet connection is provided for the individuals selected who do not have one (Intern connection).

This involves, in both cases, a rather expensive and work intensive procedure.

One operating approach to solve situations where the sample is not representative of the target population is to use a weighting system to calculate the general results (once the data has been collected). This weighting system takes into consideration the probability of extracting an individual from the population which could exclude some categories from being selected when using certain sampling techniques.

This is true in the case of non casual extraction of the sample to reach a conclusion which can bee extended to the whole target population it is first necessary to study of the relation between the sub-population sample and the universe of reference. The second step is to apply a suitable weighting system which makes it possible to also apply statistical inference to survey results obtained from non casual samples (for in-depth studies see Little & Rubin, 1987 and Danielsson, 2004).

One quite common estimation is that of Horvitz and Thompson which provides for a special weighting system of the average sample (post stratification). To apply a weighting system it is necessary to have quite detailed information on the target population, at least regarding some significant variables in order to make important comparisons with the distribution of the sample obtained. The next step is to carry out a proportional adjustment on the data regarding demographical, geographical, and aptitude variables from other sources, or deriving from independent points of reference i.e. Census data. The aim of using Horvitz and Thompson's estimator is to create a group of respondents more similar to the population who do not have Internet access, adjusting the results obtained in the web survey. This technique helps to minimize the bias effects (Pineau & Slotwiner, 2003).

Given the various approaches available, it is often not possible to improve the quality of the results. By the balancing of the results it is only possible to force the specific estimates to make them similar to those of the target population. However it is not guaranteed that the estimates being studied gained from the survey are really representative of how the phenomenon presents itself on the entire population (Pineau & Slotwiner, 2003). On the other hand, if a probabilistic sample is balanced against a reliable benchmark, this is a statistically valid technique (and often used), using instead balancing for data obtained from non-probabilistic sampling there is a risk of obtaining heavily biased results. (Deming, 1943)

The bias problem is further motivated by the fact that samples are mostly made up from volunteers recruited exclusively on the web as well as the fact that there is not sufficient auxiliary information to operate with a suitable weighting system.

8 The Balancing Score method

In the case when the casual requirement is missing (as happens in most Internet surveys) Rosenbaum and Rubin (1983) introduced a *balancing score*, a system of weights which can be used to form a group which shows a shape similar to the target population being studied.

Firstly it is necessary to carry out a sub-classification of the sample based on the *balancing score* and secondly carry out a rebalancing of the data with appropriate system of weight (i.e. *balancing score*) to calculate the general results of the survey.

The balancing score are defined (Danielsson, 2004) as a function $b(X)$ so that the distribution conditioned by the considering phenomenon X, fixed $b(X)$, is the same for the two groups of individuals in consideration (it is possible to define the two groups as group 1 and group 2).

Formalizing the above, the phenomenon Z shows the belonging to group 1 rather than group 2:

distr. $(X \mid b(X), Z = 1)$ = distr. $(X \mid b(X), Z = 2)$.

This means that, conditionally to $b(X)$, X and Z are independent phenomena.

To calculate the general results of a web survey it is possible to use the weighting system based on balancing score $b(X)$. For more in-depth information on balancing score see Rosenbaum & Rubin (1984) and D'Agostino & Rubin (2000).

The most common balancing score used is represented by the *propensity score* which is special function of the *balancing score*.

8.1 The Propensity Score method

The weighting system based on the *propensity score* was introduced with the aim of making it possible to apply inference principles to observational data (i.e. taken from non casually selected samples). (Rosenbaum & Rubin, 1983; Rosenbaum, 2002)

To reach significant results regarding a certain variable involved, a vector of auxiliary variables is used and each one refers to each individual of the population. These variables can be of different types: demographical, webographics[7], and so on. Before the survey is carried out, regarding each individual there is therefore a vector of variables. The probability that each individual has been included in the web sample is defined as propensity score. The propensity scores are therefore a special type of balancing score which meets specific requirements. If the size of a synthesis is to be calculated, a very reliable and realistic estimate is obtained using the weighting system of the propensity score.

The procedure for the functioning of propensity score methods will be now schematically illustrated.

---

[7] Webographics variables are a special type of variable obtained from webgeographic questions. Some researchers define them as lifestyle or attitudinal variables. They have the aim of picking out the differences between web survey respondents and the general population. (Schonlau et al., 2004)

The evaluation of the results of a survey applied to the same group of people, but with two different application methods, this is the same as having two groups (1 and 2) where two different methods are applied.

It is possible indicate with Y the variable of interest observed (e.g. the results of a surveys relative to a specific question). Each individual can participate only once in the survey: so it is, therefore, necessary to use a dichotomous variable, e.g. Z ($Z = 1, 2$), which indicates if the i-th individual has participated in a web survey ($Z_i = 1$) [8] or in an alternative survey ($Z_i = 2$), e.g. by telephone (i indicate the i-th individual).

The variable Y, that is the answer given by each individual to the question under consideration, can be indicated as Y_Z, i.e. Y_1 or Y_2 (given that $Z = 1,2$), depending on whether the individual has participate in a web survey or not. Therefore Y_{1i} and Y_{2j} will be the notation used to indicate the answer of the i-th element of a web survey and j-th to indicate the individual interviewed by phone.

Obviously it is unnecessary to make a direct comparison to evaluate the outcome of the same question from two different type of surveys with the same individual. This means that, in the case being considered, the difference $Y_1 - Y_2$, i.e. $Y_{1i} - Y_{2i}$, has no significance. It should be possible, however, to carry up more generally the comparison between the expected value from the answers obtained from a group of web respondents on one hand and with those who were interviewed by phone on the other hand.

What is interesting is the difference:

$\tau = E[Y_2] - E[Y_1]$.

Using the notation indicated, now X will indicate not one, but a vector of variables x, each referring to every individual of the population. These variables can be of different type: demographics, webographics, and so on.

Before the survey can be carried out, for each individual there will be a vector X_i of variables x_i (where i indicates the i-th individual).

The conditioned probability that the i-th individual has been included in a web sample (given the characteristics presented in the vector variables X_i) is indicated with $e(X_i)$:

$e(X_i) = P(Z_i = 1 \mid X_i)$.

This quantity is called "propensity score".

Propensity scores are basically special *balancing scores* which meet the following requirement for certain functions *f*:

$e(X) = f(b(X))$.

As previously mentioned the interesting point is the difference τ.

In the case of a probabilistic sampling technique for the selection of units, an estimate is achieved that is not biased by τ. In fact, as a consequence of the casual choice of the unit, the distribution of Y_1 e Y_2 are identical; therefore Y_1 e Y_2 are independent of Z.

---

[8] From here the sample who accepted to participate in a web survey will be called "web sample".

In the case, however, of a study where sampling is not casual (or one of the two groups being studies is not casually selected), the distributions of Y_1 e Y_2 will be equal only conditionally to X[9]; only with this condition will Y_1 e Y_2 be independent of Z and an unbiased estimate of τ can be obtained. The second case covers most of web surveys.

What has just been mentioned can be formulized in the following way:

$$E_X [E(Y_1 \mid X, Z = 2) - E(Y_0 \mid X, Z = 1)] = E[Y_2 - Y_1] = \tau.$$

Apart of being able to use a variable of reference X, it is also possible to use one of its function, e.g. a balancing score $b(X)$. That is to say it is possible to use every type of balancing score, even the propensity score that, as seen, is nothing more than a function of the balancing score $f(b(X))$, that is $e(X)$.

In the case of calculating synthesis (e.g. the average) of the two groups in consideration, it is possible to obtain more reliable and realistic estimates using propensity score weighting method.

For further in depth information on propensity scores and their application in web surveys see Danielsson (2004).

9 Conclusions: problems with Propensity Score and the researcher's task

The approach to web surveys using *balancing* and *propensity scores* is quite interesting. In practice, however, the *balancing score* or other useable points for the adjustment of results are often missing: weighting system, for example, can only be estimated through available data.

Furthermore, another problem linked to the implementation of these points must not be ignored: the propensity scores can be estimated through the corresponding proportions taken from the sample only if the auxiliary variable can assume only low values and the number of samples is to high. The method of re-weighting of the results based on auxiliary variables is also useful if the values that it assumes are few, otherwise problems linked to its size could arise (for further reading see Clements, 1997).

Other important consideration include in general the method of re-weighting results on the base of an auxiliary variable is useful if the possible values of X are few. However, even in the case when this is verified, there can still be problems due to the size of X. For example, some combinations of X might not be present in the sample extracted. The use of balancing and propensity score greatly simplifies things, given that the multi-dimensional structure of X is substituted by a function (b) mono-dimensional ($b(X)$ or $e(b(X))$. (Danielsson, 2004)

The function $b(X)$ (or $e(X)$) is normally a continuous function therefore necessary to recode it into classes: Cochran (1965) and Rosenbaum & Rubin (1984) claim that five classes are sufficient to obtain unbiased of estimate τ. To confirm the successful outcome of the re-classification made with $b(X)$ or $e(X)$ it is necessary to verify that the distribution of X is approximately identical in the five classes.

---

[9] This assumption is defined by Danielsson (2004) as *strong ignorability*.

Regarding the final estimate it is possible to compare each value of X and add up the weighted results of each subgroup. Therefore once the weights has been estimated using the relative fraction for each classified value of $e(X)$, e.g. using sample fraction, it is possible to obtain an estimate of the totally effect τ using the correctly weighted differences. (Danielsson, 2004)

One last consideration. It has been noted that, with known *propensity scores*, the results of the study can be weighted to estimate parameters of the population involved. However in the situations in which non-respondents are present, there is often no information on the section of the population which cannot be included in the sample. This can present a serious problem, given that to estimate the *propensity score* with e.g. the logistic regression it is necessary to have a complementary sample of the population not in the sample or of the entire population referred to. The estimated *propensity score* therefore must be corrected with known fractions relative to the subpopulation not taken into consideration (Danielsson, 2004). The estimate of the *propensity score* can be calculated with parallel surveys carried out on limited sub surveys taken of the population.

The weighting method based on propensity score seems very promising (Danielsson, 2004) and are being applied in numerous fields and context.

An applicative example for propensity score is represented by Harris Interactive, a company which organizes web surveys. Their approach is based on dividing propensity scores into a categorical variable which, then, with other variable is used for post-stratification. For further reading: Schonlau et al. (2004).

Research carried out by Harris Interactive has demonstrated that applying propensity score to web surveys produce surprisingly good results (Duncan & Stasny, 2001).

Schonlau et al. (2004) discuss when and under what hypotheses estimates obtained by adjustment made using propensity score are comparable to those based on casual samples. In particular the idea is discussed of whether it is possible or not to apply inference to financial goods and the main domicile owned based on data exclusively obtained from web surveys carried out by the Health and Retirement Study (HRS). The study was based on the comparison between the estimates obtained from a casual sample and that obtained from a sub-sample of people who having an Internet connection and agreed to participate in web surveys. This method made it possible to isolate and study any bias due to the way in which the survey was carried out.

In conclusion, all the methods of research contain, potentially, the risk of bias which can bring about a difference, also in a significant way from the correct estimate and lead to wrong measurements and conclusions. The task of the researcher is to minimize the potential sources of error, and at the same time, keeping in mind signs of these sources highlighting their characteristics and importance clearly and exhaustively.

10 Further reading

Concerning traditional sampling methods:

Babbie, E. (1990). *Survey Research Methods.* Belmont, California, USA: Wadsworth Publishing Company.

Barnett, V. (1991). *Sample Survey – Principles & Methods.* London, England: Edward Arnold.

Czaja, R., Blair, J. (?). *Designing Surveys – A Guide to Decisions and Procedures.* Thousand Oaks, California, USA: Pine Forge Press.

Deming, W.E. (1990). *Sample Design in Business Research.* New York, USA: John Wiley & Sons Inc.

ISTAT (1989). *Manuale di Tecniche di Indagine – Vol. 6: Il sistema di Controllo della Qualità dei Dati.* Roma, Italia: ISTAT – Istituto Nazionale di Statistica.

Kish, L. (1987). *Statistical Design for Research.* New York, USA: John Wiley & Sons, Inc.

Maisel, R. (1996). *How Sampling Works.* Thousand Oaks, California, USA: Pine Forge Press.

Moser, C.A., Kalton, G. (1979). *Survey Methods in Social Investigation.* Aldershot, England: Gower Publishing Company Limited.

Sarndal, C.E., Swensson, B., Wretman, J. (1992). *Model Assisted Survey Sampling.* New York, USA: Springer-Verlag.

Utts, J.M. (1996). *Seeing through Statistics.* Belmont, California, USA: Wadsworth Publishing Company.

WebSM (Web Survey Methodology). http://www.websm.org/.

Zikmund, W.G. (1997). *Exploratoring Marketing Research – Sixth Edition.* Orlando, Florida, USA: The Dryden Press.

Concerning statistical inference:

Azzalini, A., (2001). *Inferenza Statistica – Una presentazione basata sul concetto di verosimiglianza*. Springer: Milano, Italy.

Casella, G., Berger L.R. (2002). *Statistical Inference – Second Edition.* Duxbury Thomson Learning: California, U.S. .

De Cristofaro, R. (1992). *Un'introduzione critica all'inferenza statistica – II. Statistica.* G. Giappichelli Editore: Torino, Italy.

Hays, W. L., (1994). *Statistics*. Fifth Editino. Harcourt College Publishers: Fort Worth, TX.

Naddeo, A. (1987). *Inferenza Statistica*. La Nuova Italia Scientifica: Roma, Italy.

Pieraccini, L. (1991). *Fondamenti di Inferenza Statistica*. G. Giappichelli Editore: Torino, Italy.

Silvey, S. D. (1970). *Statistical Inference*. Chapman & Hall: Cambridge, England.

Zenga, M. (1996). *Inferenza Statistica*. G. Giappichelli Ed.: Torino, Italy.

Concerning web surveys:

Biffignandi, S., Pratesi, P. (2002). *Modelling the Respondents' Profile in a web Survey on Firms in Italy*. In A. Ferligoj, A. Mrvar (ed.), *Developments in Social Science Methodology*, *Metodoloski zveski*, 18, Ljubljana, 2002.

Biffignandi S., Pratesi M., Toninelli D. (2003), *Potentiality of propensity score methods in weighting for web surveys: a simulation study based on a statistical register*, ISI Conference, Berlino, august 2003.

Danes, E. E. *Model-based Inference for Complete-Data Statistics from Voluntary Responses to Internet Surveys*. From http://www.w3journal.com/3/s3.danes.html.

Lenhart, A., Horrigan, J. et al. (2003). *The Ever-Shifting Internet Population: A New Look at Internet Access and the Digital Divide*. Pew Internet & American Life Project: Washington, D.C. .

Nicolini, G., Lo Presti, A. *Composite Estimators for Complex Sampling*. From http://www.economia.unimi.it/pubb/wp152.pdf.

WebSM (Web Survey Methodology). http://www.websm.org.

Case studies regarding bias effect and the application of propensity score:

Biffignandi S., Pratesi M., *Indagini web: propensity scores matching e inferenza. un'analisi empirica ed uno studio di simulazione*, in Falorsi, Pallara, Russo (a cura di): *L'Integrazione di dati di fonti diverse. Tecniche e applicazioni del Record Linkage e metodi di stima basati sull'uso congiunto di fonti statistiche e amministrative.*, F. Angeli, 2005, Milano, pag. 131 a 158.

Biffignandi S. , Pratesi M., (2002) *Modeling The Respondents' Profile In A web Survey On Firms In Italy*, In A. Ferligoj, A. Mrvar (ed.), *Developments in Social Science Methodology*, *Metodoloski zveski*, 18, Ljubljana, 2002.

Biffignandi S., Pratesi M., *Potentiality of propensity scores methods in weighting for web surveys: a simulation study*, Quaderni del Dipartimento di Matematica, Statististica, Informatica e Applicazioni, n. 1, 2003

Duncan, K.B., Stasny, E.A. (2001). *Using Propensity Scores to Control Coverage Bias in Telephone Survey.* From *Survey Methodology* 27 (121-130).

Hirano, K., Imbens, G. V. (2004). *The Propensity Score with Continuous Treatments*, from http://elsa.berkeley.edu/~imbens/hir_07feb04.pdf

Pineau, V., Slotwiner, D. (2003). *Probability Samples vs. Volunteer Respondents in Internet Research: Defining Potential Effects on Data and Decision-Making in Marketing Applications*, Knowledge Networks, Inc., from http://www.knowledgenetworks.com/info/press/papers/Volunteer%20white%20paper%2011-19-03.pdf

Schonlau, M., van Soest, A., Kapteyn, A., Couper, M., Winter, J. (2004). *Adjusting for selection bias in web surveys using propensity scores: the case of the Healt and Retirement Study.* Paper presented in Joint Statistical Meetings, Toronto (August 2004), from http://www.schonlau.net/publication/04jsm_hrs_web_survey_propensity.pdf

Zanutto, E., Lu, B., Hornik, R. (2005). *Using Propensity Score Subclassification for Multiple Treatment Doses to Evaluate a National Anti-Drug Media Campaign.* From http://www-stat.wharton.upenn.edu/~zanutto/bio/papers/propensity.adcampaign.jebs.pdf

11 References

Frames and Sampling:

Alvarez, R. M., Sherman, R. P., VanBeselaere, C. E. (2003). *Subject Acquisition for Web-Based Surveys*.

Bourque, L. B., Fielder, E. P. (1995). *How to Conduct Self-Administered and Mail Surveys - The Survey Kit, Vol. 3*. Thousand Oaks, California, U.S.: SAGE Publications, Inc. .

Bradley, N. (1997). *Sampling for Internet Surveys. An examination of respondent selection for Internet research*. University of Westminster. Retrieved 10.07.2005, from http://users.wmin.ac.uk/~bradlen/papers/sam06.html

Cochran, W. (1977). *Sampling Techniques (3rd edition)*. John Wiley and Sons: New York.

Comley, P. (2000). *Pop-up Surveys. What works doesn't work and what will work in the future*. Proceedings of the ESOMAR Worldwide Internet Conference Net Effects 3, Dublin, 2000.

COMMERCENET and NIELSEN MEDIA RESEARCH (1996). *Internet Hyper-Growth Continues, Fuelled by New, Mainstream Users*. Press Release, http://www.commerce.net/work/pilot/nielsen_96/press.html .

Coomber, R. (1997). *Using the Internet for Survey Research*. Sociological Reserach Online, vol. 2 no. 2, from http://www.socresonline.org.uk/socresonline/2/2/2.html.

Couper, M. P. (2000). *Web Surveys: A Review of Issues and Approaches*. Public Opinion Quarterly, 64, pp. 464-494.

Couper, M. P., Traugott, M. W., Lamias, M. J. (2001). *Web Surveys Design and Administration*. Public Opinion Quarterly, 65, pp. 230-253.

Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method (2nd edition)*. John Wiley and Sons: New York.

ESOMAR (1998). *Conducting Marketing and Opinion Research Using the Internet*. NL: Esomar Guideline.

ESOMAR (1998). *Internet Conference*. Paris (January 1998): papers 117-133, 165,182, 168-182, 219-232, .

ESOMAR (1999). *Internet Conference*. London (February 1999): papers 119-129.

Fink, A. (1995). *How to Sample in Surveys – The Survey Kit, Vol. 6*. SAGE Pubblication, Inc.: Thousand Oaks, California, U.S. .

Foan, R., Read M. (1999). *Website Auditing*. London: MRG Meeting (10-2-1999).

Godin, S. (1999). *Permission Marketing.* Simon & Schuster.

Hansen, M. H., Hurwitz, W. N., Madow, W. G. (1953). *Sample Survey Methods and Theory - vol. 1: Methods and applications; vol. 2: Theory.* Whiley: New York.

Hansen, M. H., Hurwitz, W. N., (1942). *Relative Efficiencies of Various Sampling Units in Population Enquiries.* Journal of the American Statistical Association, 37, pp. 89-94.

Jeavons, A. (1999). *Research on the Internet.* ESOMAR Handbook of Market and Opinion Research (pp. 1085-1101).

Kehoe, C.M., Pitkow, J.E. (1996). *Suveying the Territory: GVU's Fire WWW User Survey.* Worldwide Web Journal, vol. 1, issue 3, from www.w3.org/pub/WWW/Journal/3/s3.kehoe.html.

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

Knowledge Networks (2002). *Decoding the Consumer Genome.* http://www.knowledgenetworks.com/press/presskit.html.

Malhotra, N. K. (1999). *Marketing Research. An Applied Orientation.* (International Edition, 3 ed.) London: Prenctice Hall.

McDonald, M., Wilson H. (1999). *E-Marketing: Improving Marketing Effectiveness in a Digital World.* London: Financial Times Prentice Hall.

Mugo, F.W., "*Sampling in Research*". From http://www.socialresearchmethods.net/tutorial/Mugo/tutorial.htm

NIELSEN MEDIA RESEARCH and COMMERCENET (1995). *The CommerceNet Nielsen Internet Demographics Survey.* www.nielsenmedia.com .

Pfleiderer, R., Gente J. (1998). *A New Dimension of Internet Research.*

Poynter, R. (2001). *A Guide to Best Practice in Online Quantitative Research.* ASC International Conference. May, 2001. from http://www.asc.org.uk/Events/May01/Resource/Poynter2.pps

Smith, P.R. (1998). *Marketing Communications.* London: Kogan Page.

Talmage, P. A. (1988). *Dictionary of Market Research.* London: MRS/ISBA.

Tse, A., altri (1994). *A comparison of the effectiveness of mail and facsimile as survey media on response rate, speed and quality.* Journal of the Market Research Society, n. 36, pp. 349-355.

Turner, W.J. (1989). *Small Business data collection by area censusing: a filed test of "Saturation Surveying" methodology.* Journal of Market Reserach Society (Vol. 3, No.2, April 1989).

U.S. Department of Commerce (2000). *Falling Through the Net: Toward Digital Inclusion.* http://www.ntia.doc/ntiahome/fttn00/falling.htm.

WebSM (Web Survey Methodology). http://www.websm.org/.

Wilke, J., Lundy, S., Mustard D. (2000). *Burning Out Internet Respondents. Avoiding the Mistakes of the Past.* Proceedings of the ESOMAR Worldwide Internet Conference Net Effects 3, Dublin, 2000.

Witmer, D. F., Colman, R. W., Katzman S. L. (1999). *From Paper-and-Pencil to Screen-and-Keyboard: Toward a Methodology for Survey Research on the Internet.* Reproduced on pp. 145-162 in Jones, S. (1999). *Doing Internet Research. Critical Issues and Methods for Examining the Net.* California, U.S.: SAGE Publications, Inc. .

Inference:

Bethlehem, J., Hoogendoorn, A. (2003). *Methodological guidelines for Blaise web surveys*. Statistics Netherlands / Free University Amsterdam, from http://www.blaiseusers.org/IBUCPDFS/2003/Methodological_guidelines_for.pdf .

Bethlehem, J.G. (1999). *Cross-sectional Research.* In Adèr, H.J., Mellenbergh, G.J., *Research Methodology in the Social, Behavioural & Life Science (pp. 110-142),* Sage Pubblications: London.

Bowker, D., Dillman, D.A. (2000). *An experimental evaluation of left and right oriented screens for Web questionnaires*. Presentation to Annual Meeting of the American Association for Public Opinion Research.

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

Couper, M.P. (2000). *Web surveys: A Review of Issues and Approaches*. The Public Opinion Quarterly 64(4) (pp. 469-494).

Couper, M.P., Kapteyn, A., Schonlau, M., Winter, J. (2004). *Noncoverage and nonresponse in an Internet Survey*. In *Proceedings of the International Conference on Social Science Methodology* (RC33): Amsterdam (August 2004).

Couper, M.P., Traugott, M.W., Lamias, M.J. (2001). *Web survey design and administration*. Public Opinion Quarterly 65(2) (pp. 230-253).

D'Agostino, R.B., Rubin, D.B. (2000). *Estimating and Using Propensity Score with Partially Missing Data.* From *Journal of the American Statistical Association* 95 (pp. 749-759).

Danielsson, S. (2004). *The propensity score and estimation in non-random surveys: an overview*. .Department of Statistics, University of Linkoeping. Report no. 18 from the project *Modern Statistical Survey Methods*: August 2004. From http://www.statistics.su.se/modernsurveys/publ/11.pdf

Deming, W.E. (1943). *Statistical Adjustment of Data*. New York Wiley & Sons: London.

Fricker, R.D., Schonlau (2002). *Advantages and disadvantages of Internet research surveys: Evidence form the literature*. Filed Methods 14 (pp. 347-367).

Heerwegh, D., Losveldt, G. (2002). *An evaluation of the effect of response formats on data quality in web surveys.* Paper presented at the International Conference on Improving Surveys, Copenhagen.

Kish, L. (1967). *Survey Sampling.* Wiley & Sons: New York, U.S. .

Little, R., Rubin, D (1987). *Statistical Analysis with Missing Data.* Wiley & Sons: New York, U.S. .

NCES (2002). *NCES Statistical Analysis Manual 2002.* NCES: Washington, D.C. .

Pineau, V., Slotwiner, D. (2003). *Probability Samples vs. Volunteer Respondents in Internet Research: Defining Potential Effects on Data and Decision-Making in Marketing Applications*, Knowledge Networks, Inc., from http://www.knowledgenetworks.com/info/press/papers/Volunteer%20white%20paper%2011-19-03.pdf

Rosenbaum, P.R. (2002). *Oservational Studies* (2nd edition). Springer-Verlag: New York.

Rosenbaum, P.R., Rubin, D.B. (1983). *The central role of the propensity score in observational studies for causal effects.* From *Biometrika* 70 (pp. 41-55).

Rosenbaum, P.R., Rubin, D.B. (1984). *Reducing Bias in Observational Studies Using Subclassification on the Propensity Score.* From *Journal of the American Statistical Association* 79 (pp. 516-524).

Schonlau, M., Fricker, R., Elliott, M. (2002). *Conducting Research Surveys via Email and the Web.* RAND: Santa Monica, CA

Schonlau, M., Zapert, K., Payne, S.L., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kann H-J., Turner, R., Berry, S. (2004). *A comparison between a propensity weighted web survey and an identical RDD survey.* From *Social Science Computer Review* 22 (1)

Schonlau, M., van Soest, A., Kapteyn, A., Couper, M., Winter, J. (2004). *Adjusting for selection bias in Web surveys using propensity scores: the case of the Healt and Retirement Study.* Paper presented in Joint Statistical Meetings, Toronto (August 2004), from http://www.schonlau.net/publication/04jsm_hrs_web_survey_propensity.pdf

WebSM (Web Survey Methodology). http://www.websm.org/.

This paper was made within research project

WebSM *(Web Survey Methodology – 6th EU Framework Program)*

www.websm.org

- Paper published in December 2005 -