

Sampling Techniques

Introduction

In *Women and Love: A Cultural Revolution in Progress* (1987) **Shere Hite** obtained several impacting results:

- 84% of women are not satisfied emotionally with their relationships.
- 70% of all women married 5 or more years have sex outside of their marriages.
- 95% of women report forms of emotional and harassment from men in love relations with them.
- 84% of women show forms of condescension from the men in their love relationships.

Wide criticism of the book, e.g. *Time magazine*: ...conclusions *dubious* and of *limited value*...

But... although S. Hite gave a voice for many women to share their experiences and points of view, it is doubtful that Hite's conclusions can apply to the whole feminine population. Because,

- The sample were self selected as the recipients of questionnaires decided to enter into the sample. Hite mailed 100000 questionnaires and only a 4.5% of the questionnaires were returned.
- Questionnaires were sent mainly to women's associations with different points of view, but **joined** in a *all women* group.
- The survey had 127 questions with several parts: very time-consuming task.
- Many terms of the questionnaire were vague and subjective like the concept of *love*.

Criticism

- Hite writes *...Does research that is not based on a probability or random sample give one the right to generalize from the results of the study to the population at large? If a study is large enough and the sample broad enough and if one, generalizes carefully yes...* (p. 778)
- But most survey statisticians would answer Hite's question with a resounding **NO**.
- The women who sent questionnaires were *purpose fully* chosen and an **extremely** small percentage of the women returned the questionnaires.

- The final sample is not representative of women in USA and the statistics can *only* be used to describe women who would have responded to the survey.
- Age, educational and occupational profiles of women in the sample matched those for the population of women in the United States. But the women in the sample are a **minority** who had the time and interest to fill out a long questionnaire offering very personal information to a researcher.

Requirements of a valid sample

- In the old movie entitled *Magic Town*, a public opinion researcher (played by James Steward) discovers a town with exactly the same characteristics of the whole *USA*: same proportion of voting people, same proportion of poverty and unemployment, etc.
- If he interviewed in that town he could know the situation of the whole country: this is the *perfect* sample... A perfect sample is a **re-scaled** version of the whole population.
- It will be representative if each sample unit will represent the characteristics of a known number of units in the population.

Design of a survey

Main issues are:

- The mode of data collection
 - face-to-face interview,
 - telephone interview,
 - self-completion form,
- The framing of the questions to be asked.
- The method of processing the data, as well as the sample design.
- The economics involved in the data collection process.

Definitions

- One of the first steps in survey design is to define the population to be studied.
- The term *population* is the totality of the elements under study, where the *elements* are the units of analysis.
- The elements may be *persons*, or be *households*, *farms*, *schools*, or any other unit.
- The population definition needs to be precisely specified according to the survey objectives, because the results will depend on the definition adopted.

Example

- A survey to be carried out in a city to discover the degree of support for the introduction of a new bus system.
- **Questions:**
 - Should the survey be confined to persons living within the city boundaries?
 - Which is the minimum age for the population to be surveyed?
 - Should residents ineligible to vote in city elections be included?
 - Should visitors living temporarily in the city be excluded, and if so, how are they to be defined?
- Many of questions like these, arise in defining most populations, making the definitional task *not straightforward*.

Methodology

- First start by defining the population as the ideal one required to meet the survey objectives: the *target population*.
- **Example:** many national surveys would ideally include servicemen based abroad and people living in hospitals, hotels, prisons, army barracks, and other institutions.
- There are many problems involved in collecting responses from such persons: frequently excluded from the survey population.
- Advantage of starting with the ideal target population: exclusions are explicitly identified, enabling to control the magnitude and consequences of the restrictions.

Taking a sample

- A *naif* approach is to take a complete enumeration of all the elements in the population, but it is better and more economic to collect data from a *part* of the population.
- By concentrating resources on only a part of the population, the quality of the data collection may be superior to that of a complete enumeration.
- A sample survey may in fact produce more accurate results: unless the population is small, sampling is almost always used.

Sampling selection

A basic distinction to be made is whether the sample is selected by a probability mechanism or not.

- With a **probability sample**, each element has a known, nonzero chance of being included in the sample. Thence, there are not selection *biases*, and Statistical Theory can be used to derive properties of the survey estimators.
- **Non-probability sampling**: use of volunteers or choice of elements for the sample supposing they are *representative* of the population.
- There is such subjectivity that precludes the development of a theoretical framework for it.

Probabilistic Sampling

- It is essential for any form of probability sample the existence of a sampling frame from which the sampled elements can be selected.
- When a list of all the population elements is available, the frame may be the list.
- When there is no list, the frame is some *equivalent* procedure for identifying the population elements.
- **Example:** *area sampling* each element of the population is associated with a particular geographical area. For instance, people or households are associated with the area of their residence, or main residence if they have more than one.
- A sample of areas is drawn, and either all elements in the selected areas are included in the survey or a sample of these elements is included.

Considerations about Probabilistic Sampling

- Organization of the sampling frame and the information: it is necessary to know about the population elements which have a strong influence on the choice of sample design.
- Defects in the frame, such as a failure to cover all the elements in the survey population, can have harmful effects on the sample.
- A variety of probability **sampling techniques** have been developed to provide efficient practical sample designs. Among the most widely used are
 - Systematic Sampling.
 - Stratification.
 - Multistage (Cluster) Sampling.
 - Probability proportional to size sampling.

Other Questions to Consider

In any study, either from a sampling design or from a complete enumeration survey there are some questions to be answered:

- What is your main research question? (**study purpose**).
- What is your population of interest? (**target population**).
- What do you know about this population? (**previous study**).
- Do you have a sampling frame? (**access to the population**).
- How good is the sampling frame? (**appropriateness**).
- Do you have an existing questionnaire? (**data gathering instrument**).
- When do you need your data and analysis? (**time frame**).
- How much money do you have? (**cost of the study**).

Software

- Software for survey analysis was specialized and their origin came from USA national agencies: SUDAAN, VPLX, WesVar.
- Other general purpose packages have some support like SAS and Stata.
- R has become nowadays a free *lingua franca*, open-source, software for **research statisticians**.
(<http://cran.us.r-project.org>).
- There are several libraries specialized in survey sampling: `survey`, `surveyNG`, `pps`, `sampling`, `epiR` and `spsurvey` among others.