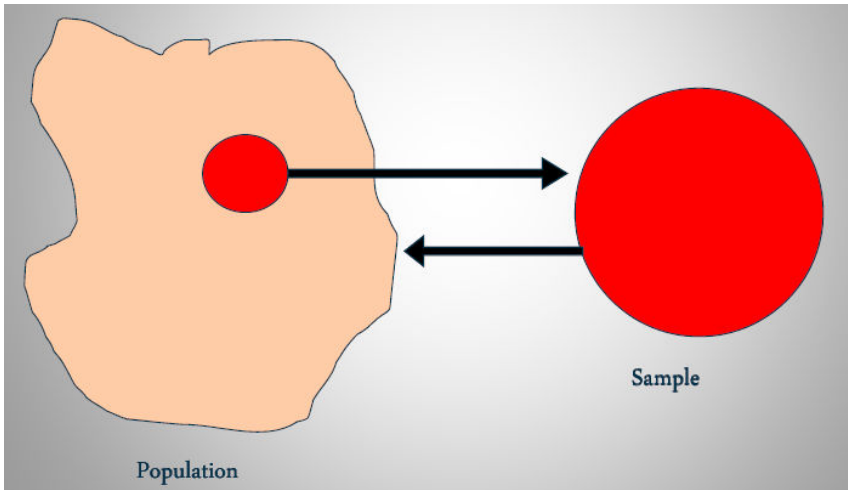
UNIVERSITY OF LONDON

Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

4.1 Introduction to sampling

Sampling is a key component of any research design. The key to the use of statistics in research is being able to take data from a sample and make inferences about a large population. This idea is depicted below.



Sampling design involves several basic questions.

- *Should* a sample be taken?
- If so, what *process* should be followed?
- What *kind* of sample should be taken?
- How *large* should it be?

We now consider how to answer these questions.

Sample or census?

We introduce some important terminology.

- **Population** – The aggregate of all the elements, sharing some common set of characteristics, which comprise the universe for the purpose of the problem being investigated.
- **Census** – A complete enumeration of the elements of a population or study objects.
- **Sample** – A subgroup of the elements of the population selected for participation in the study.

To determine whether a sample or a census should be conducted, various factors need to be considered.

- A census is very costly, so a large budget would be required, whereas a small budget favours a sample because fewer population elements are observed.
- The length of time available for the study is important – a sample is far quicker to collect.
- How large is the population? If it is 'small', then it is feasible to conduct a census (it would not be too costly nor too time-consuming). However, it might not be practical to enumerate a 'large' population.
- We will be interested in some particular characteristic, such as the heights of a group of adults. If there is a small variance of the characteristic of interest, then population elements are 'similar', so we only need to observe a few elements to have a clear idea about the characteristic. If the variance is large, then a sample may fail to capture the large dispersion in the population, hence a census would be more appropriate.
- Sampling errors occur when the sample fails to adequately represent the population. If the consequences of making sampling errors are extreme (i.e. the 'cost' is high), then a census would appeal more since it eliminates sampling errors completely.
- If non-sampling errors are costly (for example, an interviewer incorrectly questioning respondents) then a sample is better because fewer resources would have been spent on collecting the data.
- Measuring sampled elements may result in the destruction of the object, such as testing the road-life of a tyre. Clearly, in such cases a census is not feasible as there would be no tyres left to sell!
- Sometimes we may wish to perform an in-depth interview to study elements in great detail. If we want to focus on detail, then time and budget constraints would favour a sample.

The conditions which favour the use of a sample or census are summarised in the table below. Of course, in practice, some of our factors may favour a sample while others favour a census, in which case a balanced judgement is required.

Factors	Conditions favouring the use of a:	
	Sample	Census
Budget	Small	Large
Time available	Short	Long
Population size	Large	Small
Variance of the characteristic	Small	Large
Cost of sampling errors	Low	High
Cost of non-sampling errors	High	Low
Nature of measurement	Destructive	Non-destructive
Attention to individual cases	Yes	No

Classification of sampling techniques

We draw a sample from the **target population**, which is the collection of elements or objects which possess the information sought by the researcher and about which inferences are to be made. We now consider the different types of sampling techniques which can be used in practice, which can be decomposed into **non-probability sampling techniques** (covered here in Section 4.1) and **probability sampling techniques** (covered in Section 4.2).

Non-probability sampling techniques are characterised by the fact that some units in the population do not have a chance of selection in the sample. Other individual units in the population have an *unknown probability of being selected*. There is also an inability to measure sampling error. Examples of such techniques are:

- **convenience sampling**
- **judgemental sampling**
- **quota sampling**
- **snowball sampling**.

We now consider each of the listed techniques, explaining their strengths and weaknesses. To illustrate each, we will use the example of 25 students (labelled '1' to '25') who happen to be in a particular class (labelled 'A' to 'E') as follows:

A	B	C	D	E
1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Convenience sampling

Convenience sampling attempts to obtain a sample of convenient elements (hence the name!). Often, respondents are selected because they happen to be in the right place at the right time. Examples include using students and members of social organisations; also 'people-in-the-street' interviews.

Suppose class D happens to assemble at a convenient time and place, so all elements (students) in this class are selected. The resulting sample consists of students 16, 17, 18, 19 and 20. Note in this case there are no students selected from classes A, B, C and E.

A	B	C	D	E
1	6	11	**16**	21
2	7	12	**17**	22
3	8	13	**18**	23
4	9	14	**19**	24
5	10	15	**20**	25

Strengths of convenience sampling include being the cheapest, quickest and most convenient form of sampling. Weaknesses include selection bias and lack of a representative sample.

Judgemental sampling

Judgemental sampling is a form of convenience sampling in which the population elements are selected based on the judgement of the researcher. Examples include purchase engineers being selected in industrial market research; also expert witnesses used in court.

Suppose a researcher believes classes B, C and E to be 'typical' and 'convenient'. Within each of these classes one or two students are selected based on typicality and convenience. The resulting sample here consists of students 8, 10, 11, 13 and 24. Note in this case there are no students selected from classes A and D.

A	B	C	D	E
1	6	**11**	16	21
2	7	12	17	22
3	**8**	**13**	18	23
4	9	14	19	**24**
5	**10**	15	20	25

Judgemental sampling is achieved at low cost, is convenient, not particularly time-consuming and good for 'exploratory' research designs. However, it does not allow generalisations and is subjective due to the judgement of the researcher.

Quota sampling

Quota sampling may be viewed as two-stage restricted judgemental sampling. The first stage consists of developing control categories, or **quota controls**, of population elements. In the second stage, sample elements are selected based on convenience or judgement. Suppose a quota of one student from each class is imposed. Within each class, one student is selected based on judgement or convenience. The resulting sample consists of students 3, 6, 13, 20 and 22.

A	B	C	D	E
1	**6**	11	16	21
2	7	12	17	**22**
3	8	**13**	18	23
4	9	14	19	24
5	10	15	**20**	25

Quota sampling is advantageous in that a sample can be controlled for certain characteristics. However, it suffers from selection bias and there is no guarantee of representativeness of the sample.

Snowball sampling

In snowball sampling an initial group of respondents is selected, usually at random. After being interviewed, these respondents are asked to identify others who belong to the target population of interest. Subsequent respondents are selected based on these referrals.

Suppose students 2 and 9 are selected randomly from classes A and B. Student 2 refers students 12 and 13, while student 9 refers student 18. The resulting sample consists of students 2, 9, 12, 13 and 18. Note in this case there are no students from class E included in the sample.

A	B	C	D	E
1	6	11	16	21
2	7	**12**	17	22
3	8	**13**	**18**	23
4	**9**	14	19	24
5	10	15	20	25

Snowball sampling has the major advantage of being able to increase the chance of locating the desired characteristic in the population and is also fairly cheap. However, it can be time-consuming.