



SC101 Researching Social Life I
Week 7: Sampling and Data
Collection I

Dr. Neli Demireva
Department of Sociology

Types of Samples

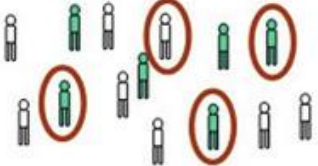

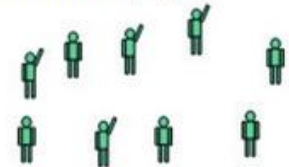
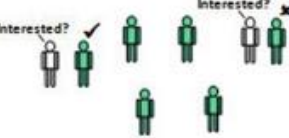
<p>Random sampling</p> 	<p>Every member of a population has an equal chance of being selected</p> <p>E.g. Pulling names out of a hat</p>	<p>For very large samples it provides the best chance of an unbiased representative sample</p>	<p>For large populations it is time-consuming to create a list of every individual.</p>
<p>Stratified sampling</p> 	<p>Dividing the target population into important subcategories</p> <p>Selecting members in proportion that they occur in the population</p> <p>E.g. 2.5% of British are of Indian origin, so 2.5% of your sample should be of Indian origin... and so on</p>	<p>A deliberate effort is made to make the sample representative of the target population</p>	<p>It can be time consuming as the subcategories have to be identified and proportions calculated</p>
<p>Volunteer sampling</p> 	<p>Individuals who have chosen to be involved in a study. Also called self-selecting</p> <p>E.g. people who responded to an advert for participants</p>	<p>Relatively convenient and ethical if it leads to informed consent</p>	<p>Unrepresentative as it leads to bias on the part of the participant. E.g. a daytime TV advert would not attract full-time workers.</p>
<p>Opportunity sampling</p> 	<p>Simply selecting those people that are available at the time.</p> <p>E.g. going up to people in cafés and asking them to be interviewed</p>	<p>Quick, convenient and economical.</p> <p>A most common type of sampling in practice</p>	<p>Very unrepresentative samples and often biased by the researcher who will likely choose people who are 'helpful'</p>

Photo credit: McLeod, S. A. (2014). Sampling Methods. Retrieved from www.simplypsychology.org/sampling.html

Basic Terms and Concepts (1)

- **Population:** the universe of units from which the sample is to be selected
- **Sample:** the segment of population that is selected for investigation
- **Sampling frame:** list of all units
- **Representative sample:** a sample that reflects the population accurately
- **Sample bias:** distortion in the representativeness of the sample

Basic Terms and Concepts (2)

- **Probability sample:** sample selected using random selection
- **Non-probability sample:** sample selected not using random selection method
- **Sampling error:** difference between sample and population
- **Non-sampling error:** findings of research into difference between sample and population
- **Non-response:** when members of sample are unable or refuse to take part
- **Census:** data collected from entire population

Sampling Error

Definition: Difference between sample and population

- Biased samples do not represent the population .
 - some groups are over-represented; others are under-represented

Sources of bias

- non-probability sampling, inadequate sample frame, non-response

Probability sampling reduces sampling error and allows for inferential statistics.

Types of Probability Samples

Simple random sample

Systematic sample

Stratified random sample

Multi-stage cluster sample

Nature of Probability Sampling

Probability Sampling

- All elements are considered and each has equal chance of being selected

Simple
Random

- Every n th element chosen started at random & picking every n th element in succession

Systematic

Stratified

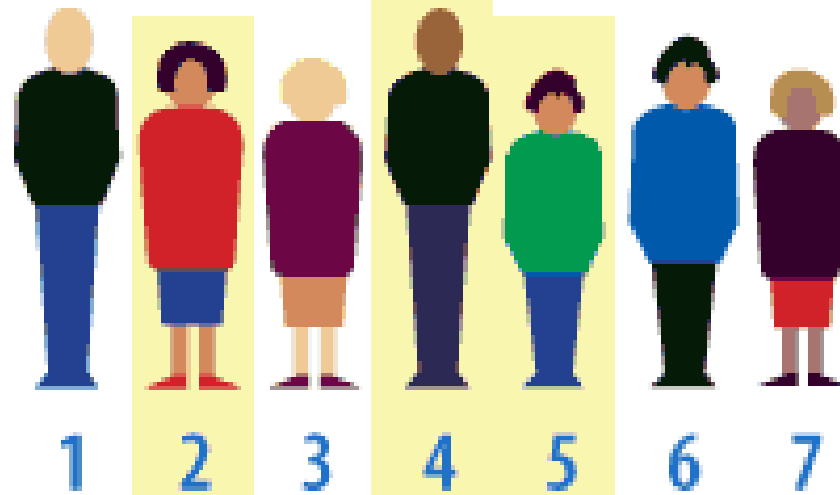
- Population is divided into sub-population/stratum & subjects selected randomly

Cluster

- Population divided into clusters, random sample of clusters is selected from as simple random design



Simple Random Sampling



**Assign Numbers,
Auto-Generate Random
Selections**

Systematic Sampling

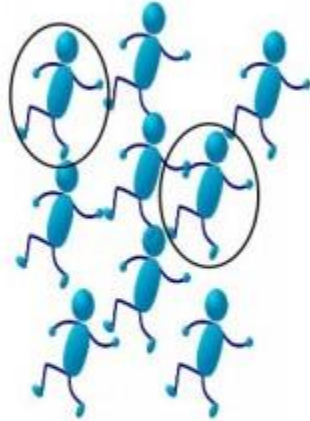
SYSTEMATIC SAMPLING



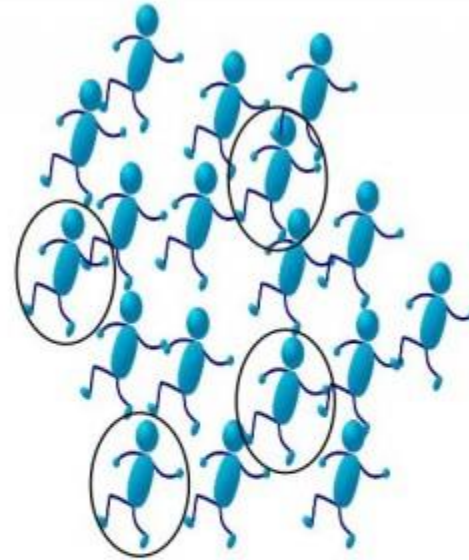
Stratified Random Sampling

- Starting point is to categorise population into 'strata' (*relevant divisions, or departments of companies for example*).
- So the sample can be proportionately representative of each *stratum*.
- Then, randomly select within each category as for a simple random sample.

Consider This Example: Two Strata – Boys and Girls



Girls



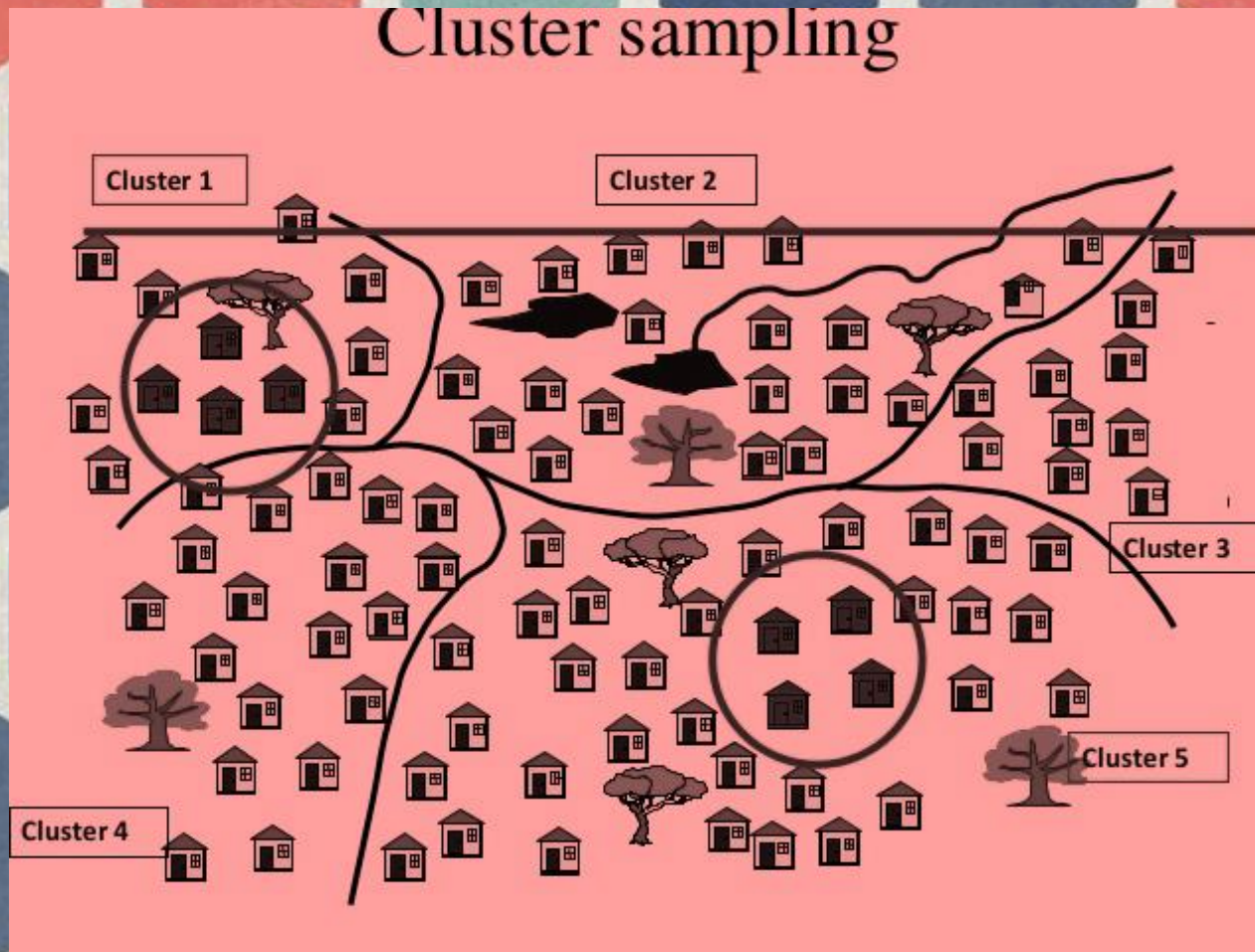
Boys

There are twice as many boys as girls in the population...
...so you need twice as many boys as girls in a stratified sample.

Cluster Sampling

- Useful for widely dispersed populations
- First, divide population into groups (clusters) of units, *like geographic areas, or industries, for example*
- Sub-clusters (sub-groups) can then be sampled from these clusters, if appropriate
- Now randomly select units from each (sub)cluster
- Collect data from each cluster of units, consecutively

Cluster Sampling



Advantages of Probability Sampling

- Representative - allows for generalization from sample to population
- Inferential statistical tests
- Sample means can be used to estimate population means
- Standard error (SE): estimate of discrepancy between sample mean and population mean
- *95% of sample means fall between +/- 1.96 SE from population mean*

Sample Size

1. **Absolute size** matters more than **relative size**.
2. The **larger** the sample, the **more precise** and **representative** it is likely to be.
3. As **sample size** increases, ***sampling error*** decreases.
4. Important to be honest about the **limitations** of your sample.

Bryman 2012:
p.198

Factors Affecting Sample Size (1)

- **Time and cost**

- after a certain point ($n=1000$), increasing sample size produces less noticeable gains in precision
- very large samples are decreasingly cost-efficient (Hazelrigg, 2004)

- **Non-response**

- response rate = % of sample who agree to participate (or % who provide usable data)
- responders and non-responders may differ on a crucial variable

Factors Affecting Sample Size (2)

- **Heterogeneity of the population:**
 - the more varied the population is, the larger the sample will have to be
- **Kind of analysis to be carried out:**
 - some techniques require large sample (e.g. contingency table; inferential statistics)

Types of Non-probability Sampling (1)

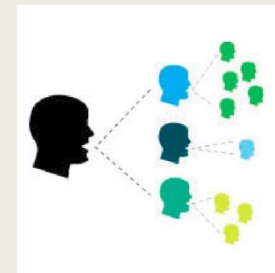
1. Convenience sampling

- the most easily accessible individuals
- useful when piloting a research instrument
- may be a chance to collect data that is too good to miss



2. Snowball sampling

- researcher makes initial contact with a small group
- these respondents introduce others in their network



Types of Non-probability Sampling (2)

3. Quota sampling

- often used in market research and opinion polls
- relatively cheap, quick and easy to manage
- proportionately representative of a population's social categories (*strata*)
- but non-random sampling of each stratum's units
- interviewers select people to fit their quota for each category, so the sample may be biased towards those who appear friendly and accessible (e.g. in the street), leading to under-representation of less accessible groups



4. What about **Purposive** Sampling?

1. The form of sampling typically used in **qualitative** research.
2. **Strategic** in nature: individuals/cases are selected on the basis of their relevance to research questions.
3. **Not possible** to **extrapolate** results to the **general** population.
4. Important to have clear and specific **criteria** to determine **inclusion/exclusion** of units of analysis.

Sampling in Ethnography

- Often a combination of convenience and snowball sampling.
- May involve purposive sampling (a set of criteria).
- The researcher has to get the information from whoever is prepared to divulge it.
- Stratified sampling might be possible.

Limits to Generalization

- **findings can only be generalized to the population from which the sample was selected**
 - be wary of over-generalizing in terms of locality
- time, historical events and cohort effects

Panel and **cohort** studies are both types of longitudinal research design, in which a sample is surveyed on more than one occasion to monitor changes over time. In a panel study, such as the British Household Panel Survey, this is usually a randomly selected sample of people who are consulted on their views of particular topics. In a cohort study, such as the National Child Development Study, the sample is selected on the basis of a shared characteristic (such as date of birth), and studied periodically over a relatively long time.

- results may no longer be relevant and so require updating (replication)

Important to remember!

Sampling is
not just
about
people!!!!!!!
!!!!

- Documents can be selected for their relevance to research questions
- Time periods need to be sampled if observations are made at different times of the day or on different days of the week
- Contexts need to be sampled if observations are made in different locations (Hammersley & Atkinson, 1995)

Summary: Keywords

- sample, population, sampling error, probability sampling, simple random, stratified, systematic, cluster, non-probability sampling, purposive sampling, convenience, snowball, quota, generalizability of results, non-response, survey costs

SC111

- Carlos Gigoux
and Poverty

Lecture 5 & 6. Inequality