Dissertation Module
Research Skills Program

LECTURE 6

SAMPLING AND SAMPLE SIZE

LEARNING OBJECTIVES

- Understand the concept of a probability sample and give examples.
- Understand the importance of random sampling for the representativeness of a sample.
- Be able to calculate required sample size for a descriptive cross-sectional study.

Sampling

- **THE SAMPLE SHOULD REPRESENT THE TARGET POPULATION**

AVOID SELECTION BIAS

- **THE SAMPLE SHOULD HAVE SUFFICIENT SIZE TO CONFIRM OR REJECT RESEARCH HYPOTHESIS WITH STATISTICAL CONFIDENCE**

SMALL RANDOM ERROR

Sampling Techniques

THE SAMPLE SHOULD REPRESENT THE TARGET POPULATION

AVOID SELECTION BIAS

ISSUE OF CONFORMITY

PROBABILITY SAMPLE

Sampling Techniques

PROBABILITY SAMPLE

A sample is a **probability sample** if each individual from the target population has a known chance to be part of the sample.

Non-probability sampling implies that either some groups of the target population have no chance to be sampled, or the chance to be sampled cannot be accurately determined.

Probability Sampling Techniques

Sampling Techniques

- **Simple Random Sampling**
- **Systematic Sampling**
- **Stratified Random Sampling**
- **Cluster Sampling**

“Simple” Random Sampling

Simple random sampling is a method of selecting n individuals out of a target population of size N in the following way:

- **A complete list of the target population exists.**
- **Individuals are independently selected one at a time until desired sample size is achieved (“sampling without replacement”).**
- **Each person in the target population has an equal chance of being included in the sample.**

Example: Simple Random Sampling

A survey wants to estimate the prevalence of smoking in Queensland's adolescents (15 to 21 years old).

Target population: All 15 to 21 year olds living in Queensland.

Sampling Procedure:

- **Compile a complete list of 15 to 21 year olds living in QLD.**
- **Number the list of persons in the target population from 1 to N.**
- **Use list of available random numbers to identify individuals of sample.**

Random Numbers

39591	66082	48628	93780	48931	…….
68198	08752	26829	35817	97112	…….
66451	18823	34849	42409	11459	…….
…….					

“Simple” Random Sampling

- Statistical procedures assume random samples (or samples that can be treated like random samples).
- Simple random sampling requires a complete list of the target population.
- Random sampling is not equal to randomization!

Systematic Sampling

Systematic sampling is a method of selecting n individuals out of a target population of size N in the following way:

- **Calculate** $k = \mathrm{rnd}\left(\frac{N}{n}\right)$

  that is, divide N by n and round this ratio to a natural number.

- **An initial random number r between 1 and k is chosen and the sample of size n is every k'th consecutive person:**

  r, r+k, r+2k, r+3k, r+4k,……..

Example: Systematic Sampling

Assume a sample of size 200 is to be drawn from a target population of size 100,000.

Sampling Procedure:

- **Compile a complete list of the 100,000 persons in the target population (1 to N).**
- **Calculate**

$$k = \text{rnd}\left(\frac{100{,}000}{200}\right) = \text{rnd}(500) = 500$$

- **Choose random number r between 1 and 500: r = 77.**
- **Sample consists of individuals: 77, 577, 1077, 1577,…….**

Systematic Sampling

- Systematic sampling requires only **one** random number.
- Systematic sampling samples **evenly** over the entire target population (every k'th individual).
- Periodic sequence in list of target population might introduce **bias**.
- If initial list of target population is a random list, sample derived from systematic sampling can be treated **like a simple random sample**.

Stratified Random Sampling

Stratified random sampling involves simple random sampling in strata of the target population.

Example: Stratified Random Sampling

A survey wants to describe sun protection behavior in outdoor workers in Queensland by using workers employed by one large company. Most (95%) outdoor workers employed by this company are male, however, the researchers also want to refer to women.

Sampling Procedure:

- **Compile a complete list of outdoor workers of this company.**
- **Stratify the outdoor workers by gender.**
- **Create TWO random samples:**
  one male and one female sample.

Stratified Random Sampling

- Stratified random sampling takes additional information about target population into account.
- Assures that each strata is represented in sample.
- Stratification might create more homogenous sub-samples. That is, estimations might be more precise compared to simple random sampling.

Cluster Sampling

Cluster sampling involves "natural" groups (clusters).

One-stage cluster sampling

A random sample of clusters is selected and ALL individuals in those clusters are used.

Two-stage cluster sampling

A random sample of clusters is selected and within each cluster a random sample of individuals are used.

Example: Cluster Sampling

A survey wants to estimate the prevalence of smoking in Queensland's adolescents (12 to 17 years old).

Target population: All 12 to 17 year olds living in Queensland.

Sampling Procedure (for example):

- Simple random sample of schools
- Simple random sample of adequate classes within each school
- All students within each of the selected classes

Cluster Sampling

- Cluster sampling might be considered if a complete list of target population is unavailable and if targeted persons form "natural" groups (e.g. schools, suburbs, day-care centers).
- Cluster sampling is often more convenient and economical than simple random sampling.
- Cluster sampling requires sample size adjustment ("design effect").

Non-Probability Sampling Techniques

Sampling Techniques

- **Convenience sampling**
- **Snowball sampling**
- **Purposive sampling**

Sampling

- **THE SAMPLE SHOULD REPRESENT THE TARGET POPULATION**
  AVOID SELECTION BIAS

- **THE SAMPLE SHOULD HAVE SUFFICIENT SIZE TO CONFIRM OR REJECT RESEARCH HYPOTHESIS WITH STATISTICAL CONFIDENCE**
  SMALL RANDOM ERROR

Sample Size

Assume: Experimental study to identify better treatment, measured as proportion cured.

TRIAL 1		Cured	Proportion cured
Group A:	n = 10	6	60%
Group B:	n = 10	8	80%

TRIAL 2		Cured	Proportion cured
Group A:	n = 100	60	60%
Group B:	n = 100	80	80%

We are more confident in results of trial 2,

as it was based on a larger sample size!

Sample Size

The operational research hypothesis quantifies expected outcome:

Difference in treatment outcome
Prevalence of disease
Estimation of risk, etc.

Calculation of sample size allows to estimate expected outcome with pre-specified statistical confidence.

Sample Size for Descriptive Cross-Sectional Studies

$$\text{Sample Size } n = \frac{z^2 \times p\,(1 - p)}{d^2}$$

z Describes statistical confidence;
z = 1.96 translates to 95% confidence
z = 1.68 translates to 90% confidence
(Choice depends on cost and actions arising)

p Expected prevalence

d Describes intended precision;
d = 0.1 means that estimate falls +/-10 percentage points of true p with the considered confidence

Formula for Sample Size in Descriptive Cross-Sectional Studies

$$\text{Sample Size } n = \frac{z^2 \times p\,(1 - p)}{d^2}$$

d Describes intended precision;
d = 0.1 means that estimate falls +/-10 percentage points of true p with the considered confidence (e.g. p = 0.4; 0.3 to 0.5); d = half the confidence interval

If you decrease "d", you increase:
- Precision
- Sample size
- Cost & Time

Example: Sample Size – Nutrition Survey

$$\text{Sample Size } n = \frac{z^2 \times p\,(1 - p)}{d^2}$$

Simple random sample

p = 0.3

95% confidence (z = 1.96)

d = 0.05 (5% precision)

$$n = \frac{1.96^2 \times 0.3\,(1 - 0.3)}{0.05^2} = 322.7 \Rightarrow 323 \text{ people}$$

Formula for Sample Size in Descriptive Cross-Sectional Studies

$$\text{Sample Size } n = \frac{z^2 \times p\,(1 - p)}{d^2}$$

p	p (1-p)
0.5	0.25
0.4	0.24
0.3	0.21
0.2	0.16
0.1	0.09

If p unknown, choose p = 0.5

Formula for Sample Size in Descriptive Cross-Sectional Studies

$$\text{Sample Size } n = \frac{z^2 \times p\,(1 - p)}{d^2}$$

d	Sample size (for p = 0.5; 95% confidence)
0.2	24
0.1	96
0.05	384
0.025	1537

SUMMARY

- A sample is a probability sample if each individual in the target population has a known chance of being part of the sample.
- Non-probability sampling means that either some groups of the target population have no chance of being sampled, or that the chance of being sampled cannot be accurately determined.
- Random sampling, systematic sampling, and cluster sampling are probability sampling approaches.
- A sample size calculation allows the researcher to compute the optimal size for the planned study based on the operational research hypothesis.
- There are numerous different formulae available for sample size calculations, which depend on the study design, the operational research hypothesis, and the outcome measure.
- Special software programmes are available for dealing with sample size calculations.