

Quantitative methods

Lesson 5

Daróczi Gergely

Corvinus University of Budapest, Hungary

2011 March 8



- 1 Sample-bias
- 2 Sampling theory

Time magazine reported in the late 1950s that

"the average Yaleman,
class of 1924,
makes \$ 25,111 a year"

which would be equivalent to well over \$ 150,000 today!

Sample-bias

Cause of errors

Time's estimate turns out to have been based on replies received to a sample survey questionnaire mailed to those members of the Yale class of 1924 whose addresses were known in the late 1950s by the Yale administration.

- 1 selection bias,
- 2 nonresponse bias,
- 3 response bias.

Sample-bias

Other historical examples

1936: the *American Literary Digest* magazine collected over two million postal surveys and predicted that the Republican candidate in the U.S. presidential election, Alf Landon, would beat the incumbent president, Franklin Roosevelt by a large margin.

- records of registered automobile owners and telephone users,
- George Gallup: quota sampling with 50.000 respondents.

1948: *Chicago Tribune* printed the headline „DEWEY DEFEATS TRUMAN” based on a Gallup poll.

- telephone interviews,
- quota matrix had changed a lot!

Sampling theory

Elements

- 1 population,
- 2 respondents, units of analysis,
- 3 sampling frame,
- 4 sampling methods.

Kish posited four basic problems of sampling frames:

- 1 Missing elements: Some members of the population are not included in the frame.
- 2 Foreign elements: The non-members of the population are included in the frame.
- 3 Duplicate entries: A member of the population is surveyed more than once.
- 4 Groups or clusters: The frame lists clusters instead of individuals.

Sampling theory

A not so well chosen sampling frame

We started a small research company and someone proposed to use the public phonebook to build samples:

- 1 based on public phonebook: only those are on the list who holds a phone,
- 2 only those with *public* phone number,
- 3 mobile numbers are not called for surveying (expensive),
- 4 repeated calls to the same number are forbidden,
- 5 only those are reached, who are willing to answer to our questions on the line.

Sampling theory

Sampling frame

Propose a well chosen sampling frame for the following research subjects:

- 1 Missing elements: Some members of the population are not included in the frame.
- 2 Foreign elements: The non-members of the population are included in the frame.
- 3 Duplicate entries: A member of the population is surveyed more than once.
- 4 Groups or clusters: The frame lists clusters instead of individuals.

Sampling theory

Basic types of probability sampling

Next week!

- 1 Simple Random Sampling (SRS)
- 2 Systematic Sampling
- 3 Stratified Sampling



A subset of the population.

Sampling theory

Basic sample size

For SRS:

$$SE = \frac{S^*}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}$$

$$S^* = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Sampling theory

An example of a stratified sample

We asked 4 student about the number of cats at home:

	Rockers	Rappers
Girls	9	7
Boys	3	1

Imagine, what would be the results if the sample was chosen randomly and if it was stratified?

Choosing samples of $n=2$:

- 1 SRS: 6 possible samples: (1,7) (1,9) (3,7) (3,9) (1,3) (7,9)
 $\bar{x} = \frac{4+5+5+6+2+8}{6} = 5, S^* = \frac{1+0+0+1+9+9}{6} = 3.33$
- 2 Strat. Sampling: 4 possible samples: (1,7) (1,9) (3,7) (3,9)
 $\bar{x} = \frac{4+5+5+6}{4} = 5, S^* = \frac{1+0+0+1}{4} = 0.5$
- 3 Strat. Sampling: 4 possible samples: (1,3) (1,9) (3,1) (3,7)
 $\bar{x} = \frac{2+5+2+5}{4} = 3.5, S^* = \frac{1.5+1.5+1.5+1.5}{4} = 1.5$

It was a pleasure!

Daróczy Gergely
daroczy.gergely@btk.ppke.hu