

ARTNeT Greater Mekong Sub-region (GMS) initiative

Session 4

Methods sampling and when they are appropriate

ARTNeT Consultant

Witada Anukoonwattaka, PhD

Thammasat University, Thailand

witada@econ.tu.ac.th

Asia-Pacific Research and Training Network on Trade

www.artnetontrade.org

Outline

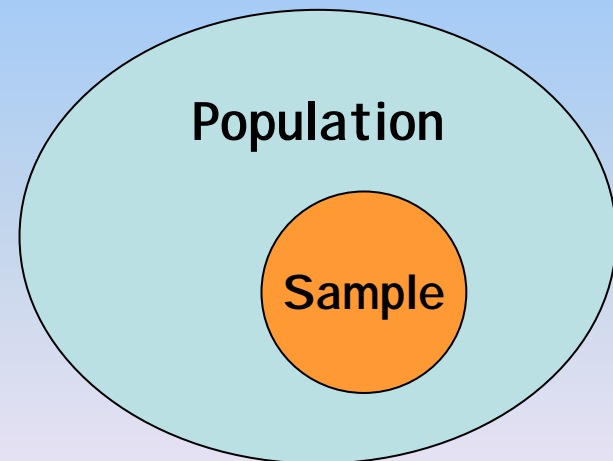
- Important technical terms
 - Population vs. Sample
 - Statistic vs Parameter
- Accuracy of sampling
- Types and methods of sampling

Samples

- *Samples* are taken from populations.
- Estimates are made about the total population based on information derived from the sample.

E.g. You ask a group of people a list of questions and based on their results, you draw conclusions about the population as a whole.

- A sample *must be large enough* to give a good *representation* of the population, but *small enough to be manageable*.



Variables

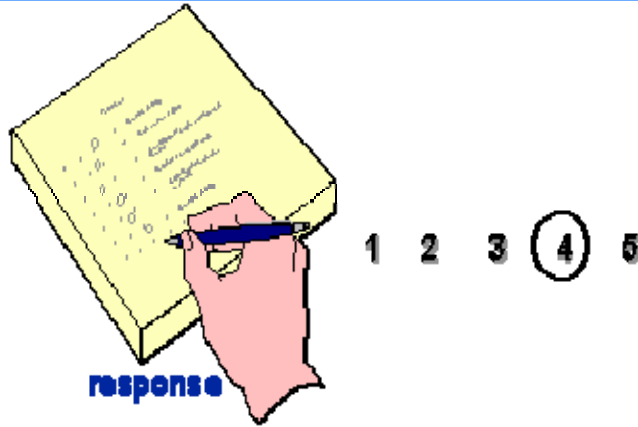
A characteristic of interest drawn from the data. It is usually classified into qualitative and quantitative types.

Example:

<u>Data</u>	<u>Variable</u>	<u>Qualitative or Quantitative</u>
Gender of the worker	Gender	?
Age of the worker	Age	?
Speed of assembly	Speed	?
Country of production	Country	?

Statistic and Parameter

Variable



When we sample, the person is responding to a survey instrument and gives a response of '4'.

Statistic



When we look across the responses that we get for our entire sample, we use a **statistic**.

Parameter



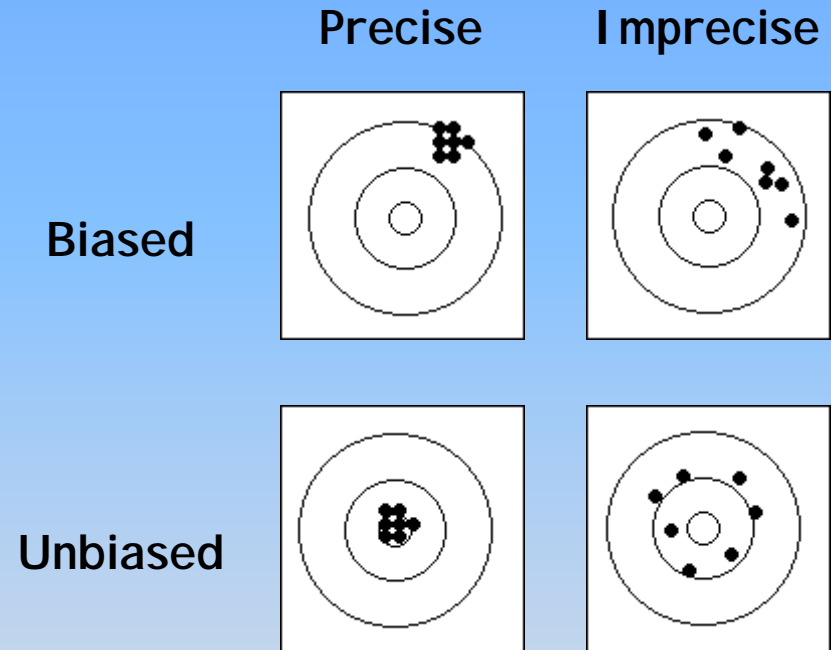
If you measure the entire population and calculate a value like a mean or average, we call it a **parameter** of the population

Accuracy of Sampling (1)

- You will need to know how many people are in the entire group and how "accurate" you want your results to be.
- Accurate sampling provides “unbiased” and “precise” estimates of the population parameters.

Accuracy of Sampling (2)

Bias is a term which refers to how far the average statistic lies from the parameter it is estimating



Precision means **low measurement error**. If a particular sampling methodology produces very similar results every time it is used, then that methodology is *precise*.

Factors that determine sampling accuracy

- **Sampling size**
 - Are there enough samples to make the results a good representation of the entire target population ?
- **Sampling methods**
 - Example: if you are surveying customers and your survey is internet-based, you are not likely to get much feedback from people who don't use computers.

Sample Size

- As large as possible.
- Do you feel confident with the following results?
 - a) 10 out of 10 said “strongly agree”.
 - b) 7 out of 10 said “strongly agree”.
 - c) 700 out of 1000 said “strongly agree”.
- If you have a smaller number of respondents, you need much stronger results in order to draw conclusions from the numbers.

Types of sampling methods

- **Random sampling**
 - All items have some chance of selection.
- **Non-random sampling**
 - Some items are certainly excluded. Bias?

Types of Sampling Methods

Which one is random sampling?

1. Drawing a token from a box filled with 10 tokens.
2. A TV program asked people to phone in to give their opinion on the government's performance.

Main methods of random sampling

- Simple random sampling (SRS)
 - Each item in a population has an equal chance of being selected. Example: lotto
- Systematic sampling (interval sampling)
 - There is a gap, or interval, between each selection. Example: select every 10th person who enters a particular store, after *randomly selecting* the first person.
- Stratified sampling
 - You form the population into groups (called strata), and sample from each group (strata).

Example of Simple Random Sampling

Which one is simple random sampling?

- a) **A bingo game**: all the numbers (total population) are put into a barrel and some numbers (sample) are drawn.
- b) **An election**: each member of the population (18 years or older) can participate if they desire.
- c) **A census**: every member of the target population must be included.

Tips: Excel can help you run a random sampling

1. Copy and paste a list of every person in the group into a single column. You can use names, email addresses, employee numbers, or whatever.
2. In a second column, fill the entire column with Excel's "Randomize" function. The exact value of each cell should be "=rand()" (do not include the quotation marks). Only fill the cells next to where you pasted the group info in step #1.
3. Sort both columns by the "Randomize" column. (It does not matter whether you sort them in ascending or descending order.)
4. Scroll down to the row number of the group size. Everybody from this row up is a part of your sample.

joe@custominsight.com	0.914294686
heather@custominsight.com	0.240840596
jason@custominsight.com	0.695627907
julian@custominsight.com	0.267099684
todd@custominsight.com	0.943071458
erin@custominsight.com	0.618678821
robert@custominsight.com	0.702098512
betty@custominsight.com	0.599022136
harold@custominsight.com	0.568510442
monique@custominsight.com	0.310769667
stacy@custominsight.com	0.483876425

Column of random numbers "=rand()"

Select both columns and sort by this column to randomize the list of names.

Then select the number of names that you need.

Column of names

Every person in the population or subgroup gets listed here.

Disadvantage of simple random sampling

For instance:

- Population: 90 people, of which 45 are men and 45 are women.
- Sample: 25 out of 30 people that were randomly selected are men.

Is the selection procedure biased?

Is the sample biased?

Stratified Sampling

- You want to understand the differences between productivity of men and women.

If the population consists of 90 people that are a mixture of 45 men and 45 women. Then,...?...% of sample should be men.

- You select a separate random sample from each of the subgroups rather than just taking a single random sample from the entire group.

Example: Stratified Sampling

- Number of garment firms by firm sizes

<u>Firm sizes</u>	<u>Number of firms</u>
Small (less than 50 workers)	900
Medium (51-100 workers)	600
Large (more than 100 workers)	300
<u>Total</u>	<u>1,800</u>

- a) What is the total population of garment firms ?
- b) If we want to sample 50% of the population. How many firms would this be?
- c) To keep the correct proportion of firm sizes in the sample, how many small, medium, and large firms should be included in the sample ?

Disadvantage of Stratified Sampling

- Criterion of selection is subjective, and may be biased.

Example: Systematic Sampling

A computer manufacturer is running an assembly line that assembles 2,700 computers per day. The manager wants to systematically randomly select 300 computers for quality checking.

- What should the sampling interval be?
- If the number 8 was your first randomly drawn number, what would be the first 5 numbers of your sample?

Reading

- Bartlett, J. E., II, Kotrlik, J. W., & Higgins, C. (2001). Organizational research: Determining appropriate sample size for survey research. *Information Technology, Learning, and Performance Journal*, 19(1) 43-50.
- Cochran, William G. (1977). *Sampling Techniques* (Third ed.). Wiley.
- Lohr, Sharon L. (1999). *Sampling: Design and Analysis*. Duxbury.

Exercise 3

Using random number tables.

The following is a list of company employees (the sampling frame). The list has been numbered and put in alphabetical order for ease of use. The gender (Male, Female) of each person is also given:

01	James Allen	M	21	Alan Masters	M
02	James Amos	M	22	Wai Fan Ng	F
03	Tom Baines	M	23	Ramil Obispo	M
04	Fred Baker	M	24	Andy Pinder	M
05	Seema Bhatti	F	25	Sanjay Rajendera	M
06	Maxine Bramley	F	26	Dave Sargent	M
07	Jack Chatman	M	27	Babu Sharma	M
08	Alex Cluskie	M	28	Dave Thompson	M
09	Chintu Desai	M	29	Pat Truscott	F
10	Jim Fairman	M	30	Mal Tynedale	M
11	Bert Graham	M	31	Jim Watson	M
12	Denise Green	F	32	Maggie Weakley	F
13	Maggie Greenway	F	33	Paul Woodhouse	M
14	Nick Harrison	M	34	Anne Woods	F
15	Jean Hope	F	35	Helen Yates	F
16	Joan Hutton	F	36	Steve Yendell	M
17	Pat James	M	37	Mew Ling Yeo	F
18	Chris Light	M	38	Chris Yeung	M
19	Genie Lin	F	39	Julia Young	F
20	Frances Mann	F	40	Nicola Zainu	F

Note: each number allocated to the sampling frame must contain the same number of digits otherwise you cannot read the tables consistently. Any number outside the range, or repeated, is not used.

1. Simple random sampling *The following is a string of random numbers taken from tables: 06 73 29 12 98 06 33 80 05*

By starting at 06 and reading to the right, select a Simple Random Sampling of size 5 from the list of employees given above. List the five names selected and their gender. Is this sample representative of the population?

2. Stratified sampling *For the following string of random numbers, and by starting at 84 and reading to the right, select a stratified sample of size 5, based on the sex of employees. List the five names and their gender: 84 37 10 05 85 12 11 28*

What other criteria could be used to stratify the population?

3. Systematic sampling

Using a random start number of 3, select a systematic sample size of 5. List the five names and their gender.