

Sampling & Questionnaire Design

EE489 Seminar in Industrial Economics

Dr. Wanwiphang Manachotphong

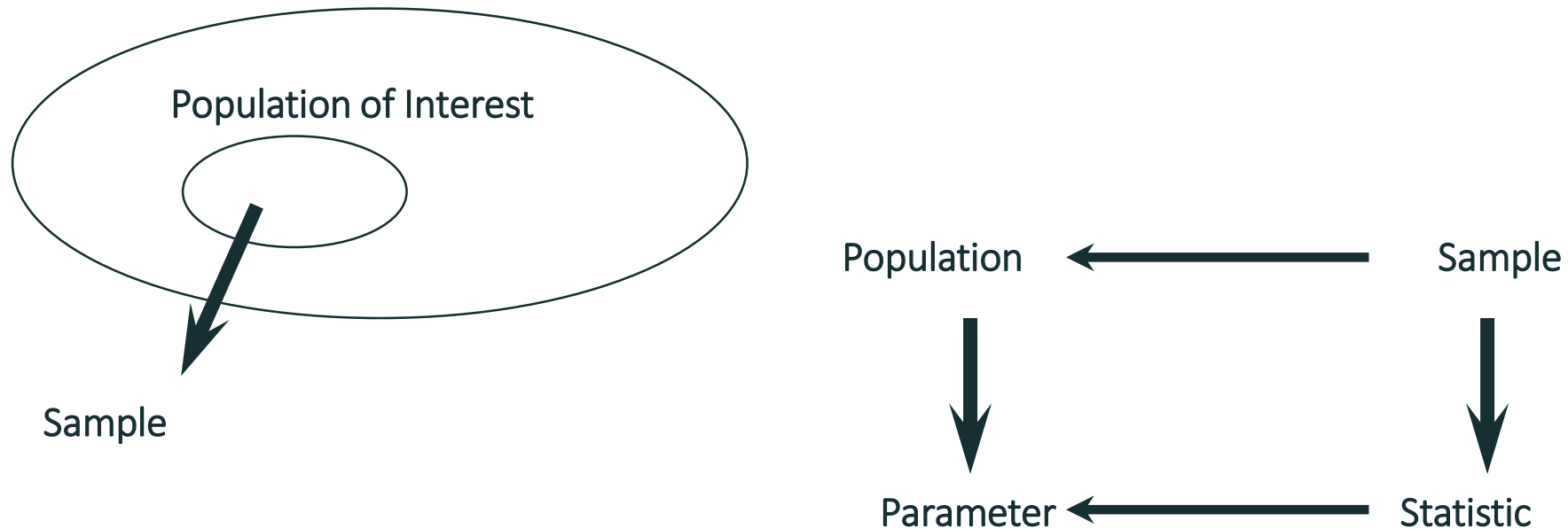
Sampling Design

- Obtain information from a subset of a larger group
- The results used to make estimates of the larger group
- Faster and cheaper than asking the entire population

Two keys

1. Selecting the right people
2. Selecting the right number of the right people

Population vs. Sample



We measure the sample using statistics in order to draw inferences about the population and its parameters.

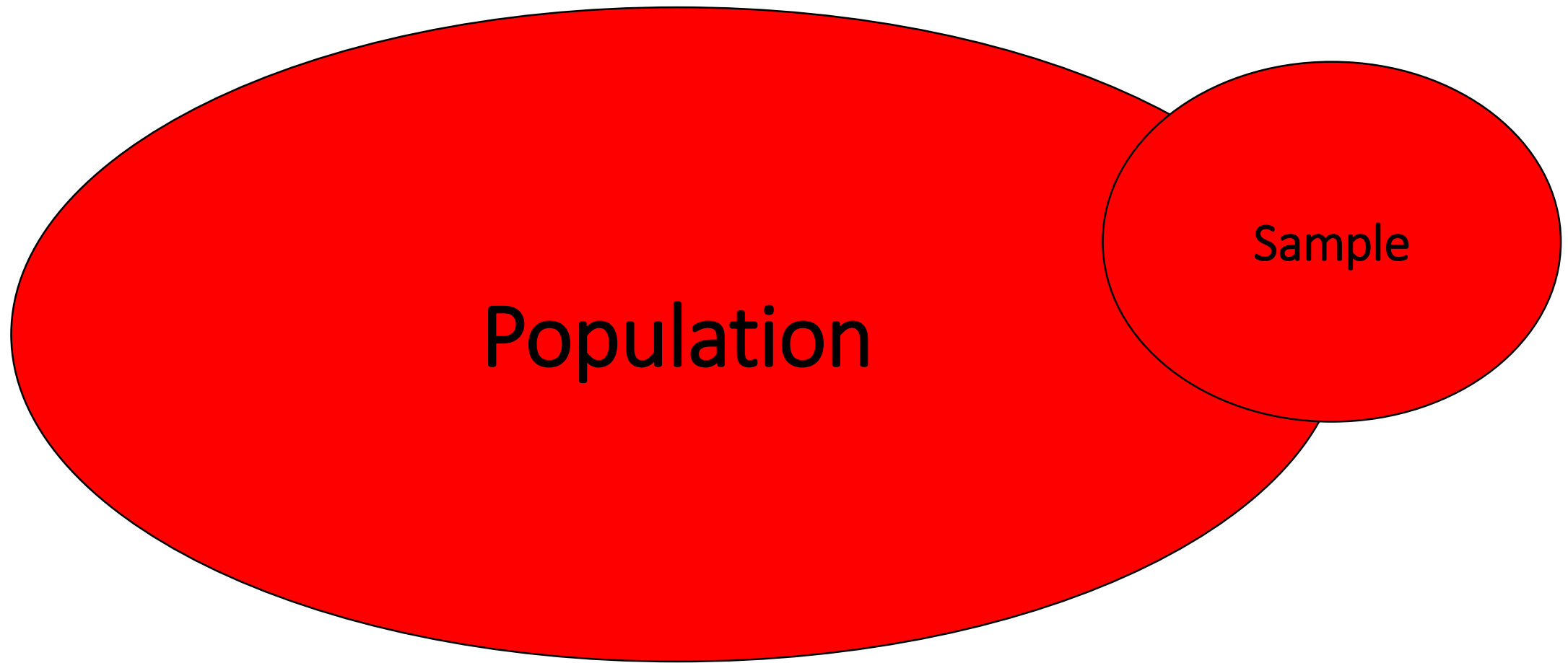
Characteristics of Good Samples

Representative

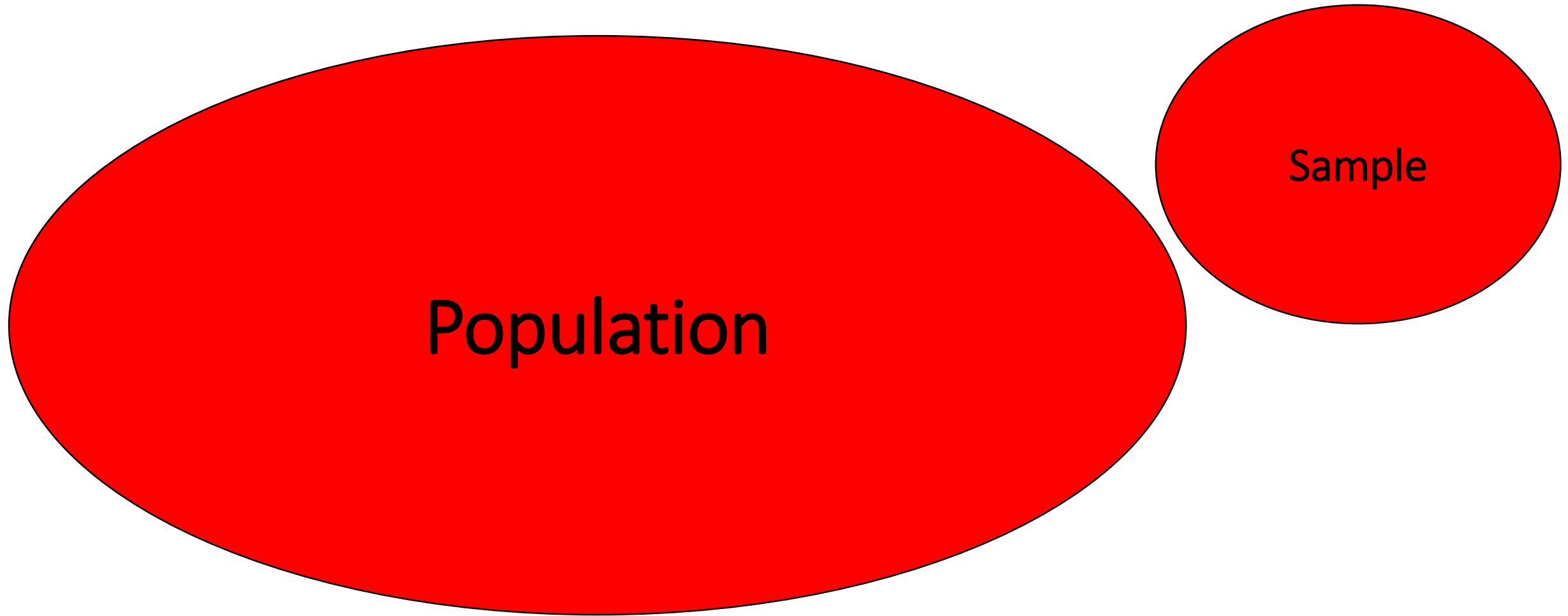
Accessible

Low cost

...This (bad)...



...or this (VERY bad)...



Terminology

❖ *Population*

The entire group of people of interest from whom the researcher needs to obtain information.

❖ *Element (sampling unit)*

one unit from a population

❖ *Sampling*

The selection of a subset of the population

❖ *Sampling Frame*

Listing of population from which a sample is chosen

❖ *Census*

A polling of the entire population

❖ *Survey*

A polling of the sample

Terminology

❖ *Parameter*

The variable of interest

❖ *Statistic*

The information obtained from the sample about the parameter

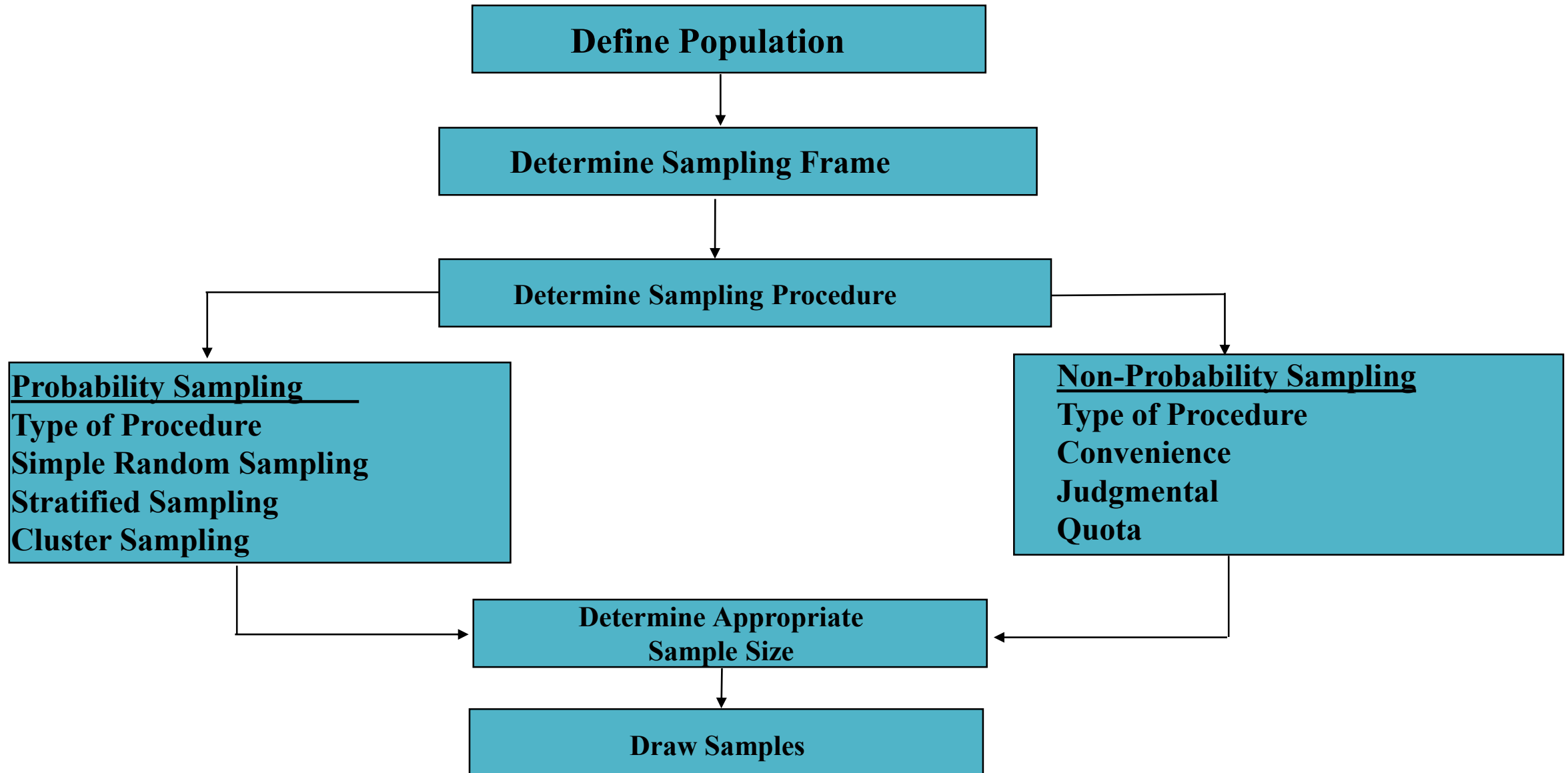
❖ *Goal*

To be able to make inferences about the population parameter from knowledge of the relevant statistic - to draw general conclusions about the entire body of units

❖ *Critical Assumption*

The sample chosen is representative of the population

Sampling Design Process



1. Define the target population

❑ “Ideally, who do you want to survey?”

i.e. those who have the information sought What are their characteristics. Who should be excluded?

❑ It involves

- defining population units
- setting population boundaries
- Screening (e.g. security questions, product use)

2. Determine the Sampling Frame

Select “*sample units*”

- Individuals
- Household
- Streets
- Telephone numbers
- Companies

3. Selecting a Sampling Design

PROBABILITY SAMPLING

Equal chance of being included in the sample (random)

- simple random sampling
- systematic sampling
- stratified sampling
- cluster sampling

NON-PROBABILITY SAMPLING

Unequal chance of being included in the sample (non-random)

- convenience sampling
- judgment sampling
- snowball sampling
- quota sampling

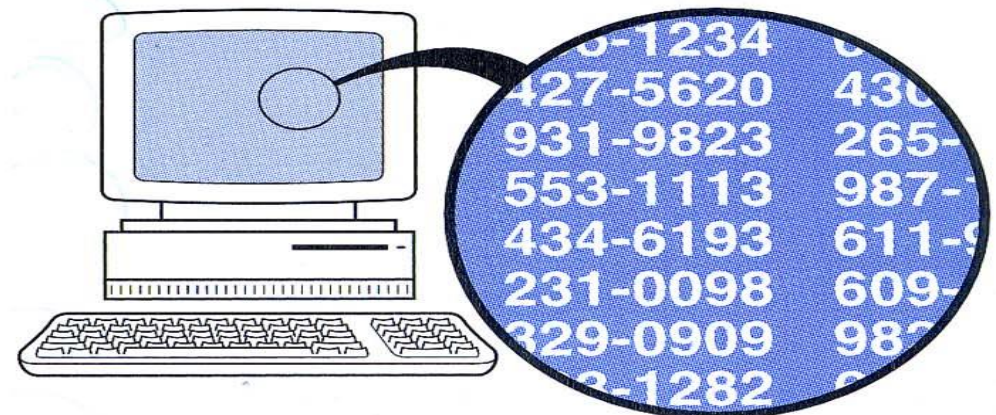
- Simple Random Sampling (SRS)

Population members are selected directly from the sampling frame

Equal probability of selection for every member (sample size/population size)

$$400/10,000 = .04$$

Use random number table or random number generator



- Systematic Sampling

Order all units in the sampling frame based on some variable and number them from 1 to N

Choose a random starting place from 1 to N and then sample every k units after that



- Systematic Sampling

$N = 100$

want $n = 20$

$N/n = 5$

**select a random number from 1-5:
chose 4**

start with #4 and take every 5th unit

1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

• Stratified Sampling (I)

The chosen sample is forced to contain units from each of the segments, or strata, of the population

equalizing "important" variables

years in school, geographic area, product use, etc.

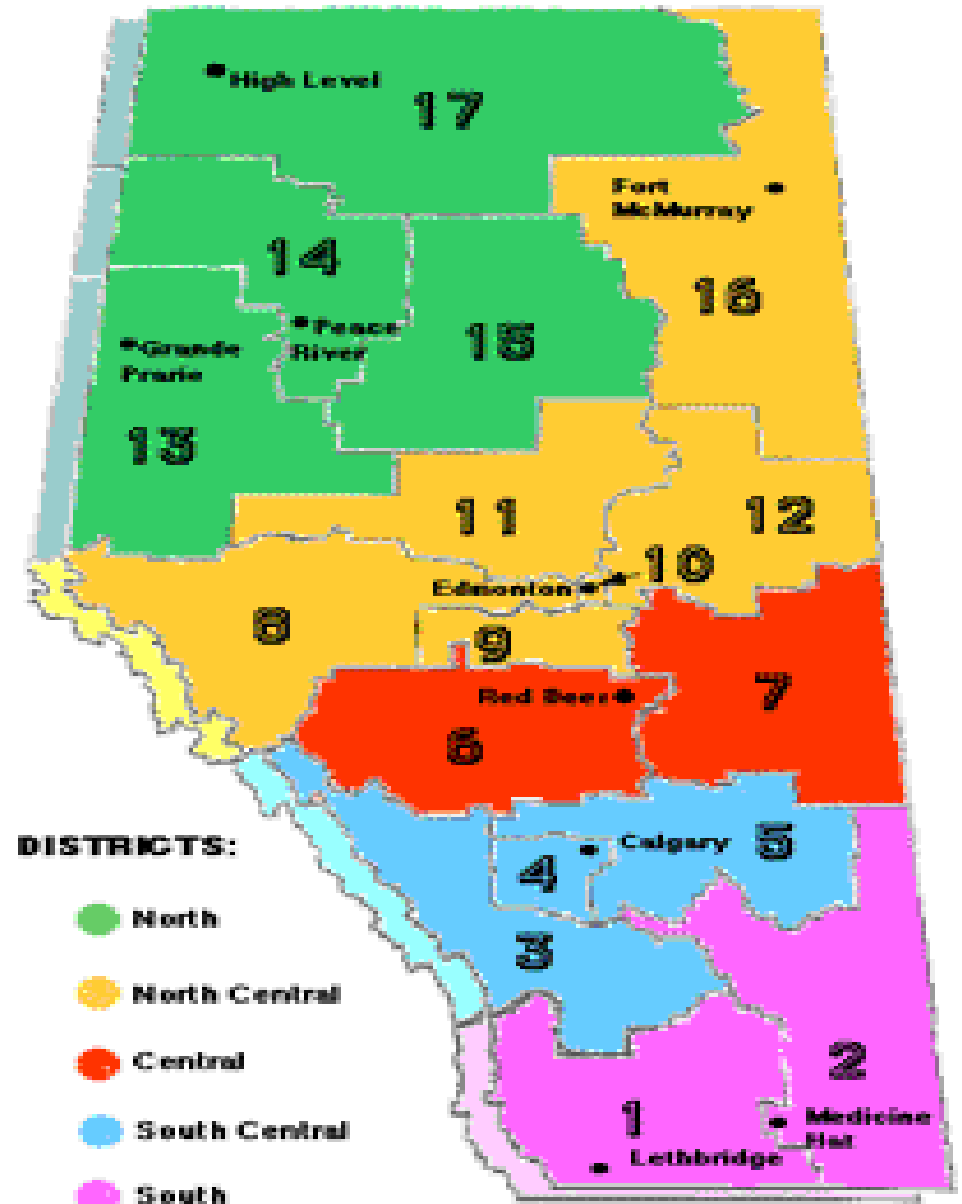
Steps:

Population is divided into mutually exclusive and exhaustive strata based on an appropriate population characteristic. (e.g. age, gender etc.)

Simple random samples are then drawn from each stratum.

• Cluster Sampling

- Divide population into clusters (usually along geographic boundaries)
- Randomly sample clusters
- Measure units within sampled clusters



3. Selecting a Sampling Design

When to use stratified sampling

- To compare groups of different types of samples.

When to use cluster sampling

- Costly to move between each data collection location.
- There is a list of clusters but not of individual population members

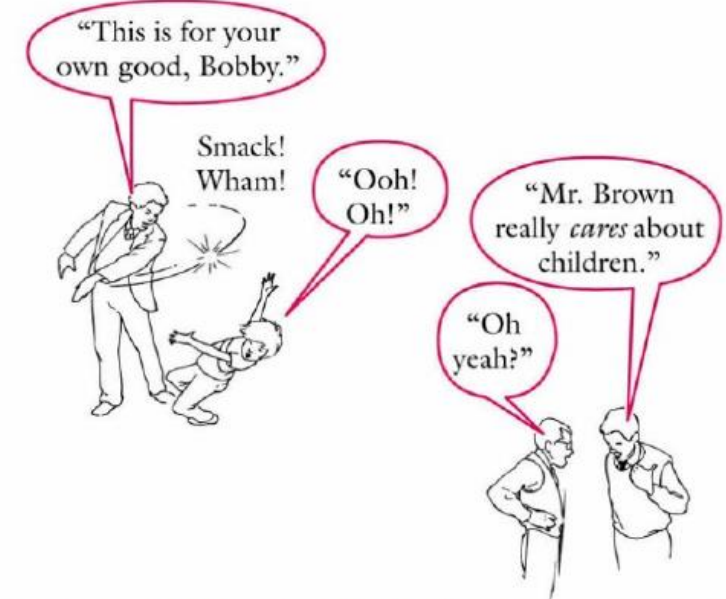
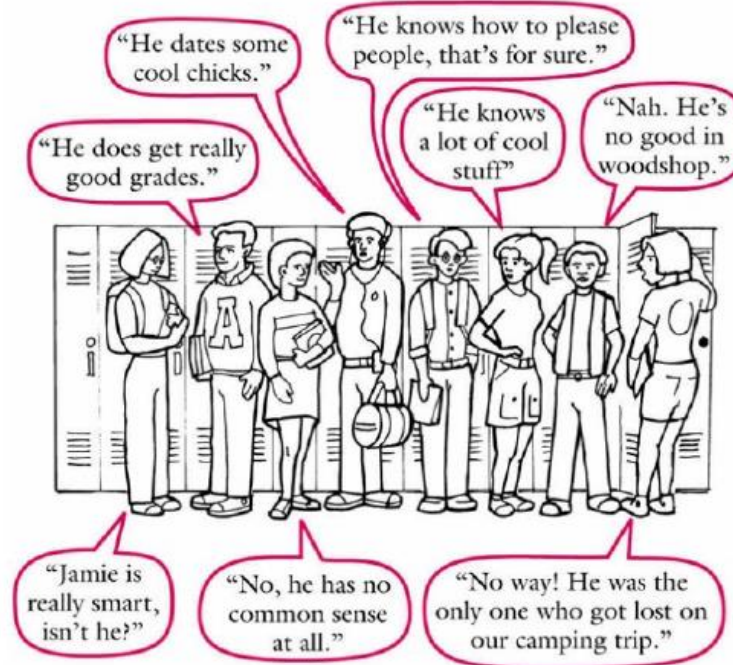
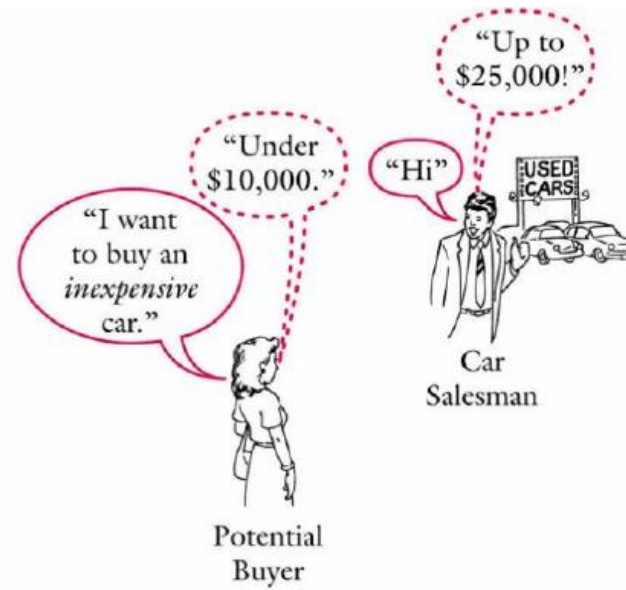
Questionnaire Design

adapted from Taylor-Powell (1998) and Burgess (2001)

1. define your research aims
2. identify the population and sample
3. decide how to collect replies
4. design your questionnaire
5. run a pilot survey
6. carry out the main survey
7. analyze the data

Questions ...

and how they are
perceived



Gay, L. R., & Airasian, P. (2003). *Educational Research: Competencies for Analysis and Applications*. New jersey: Merrill Prentice Hall.

Decide how to collect replies

- Interview
 - need to train all the interviewers.
- Self-administered questionnaires
 - questions have to be clear.

Questionnaire design

Collect information that answers your question.

Issue:	Question focus:
What reasons might candidates give for undertaking an MBA?	Is the candidate looking for: career change career advancement higher remuneration etc.
Could past experience affect the reasons?	How many years work experience does candidate have?
Could gender differences affect the reasons?	Is the candidate male or female?
Could educational background and attainment affect the reasons?	What is highest educational qualification obtained? What subject area(s) is this qualification in?
Etc.	Etc.

Layout and sequence

- Concise
- Has some introductory statement
- Logical
- Be consistent on your wordings
- Give clear instruction

e.g. tick the relevant box. Consider using bold for the questions themselves or for headings. Symbol fonts can be useful for characters such as ☐✓

Open vs. closed questions

Do you think football hooliganism is caused by: *(tick if appropriate)*

Lack of discipline at home	☐
Players' behaviour on pitch	☐
Family breakdown	☐
Youth unemployment	☐
Poor schooling	☐
Violence on T.V.	☐
Other (please specify)	☐

e.g. What was the main problem you encountered with your wheelchair?

☐

Single vs. Multiple responses

Which of the following means do you use to travel to college?

Bus	☐
Car	☐
Bike	☐

What is your most usual means of travelling to college?

(Tick one box only)

Bus	☐
Car	☐
Bike	☐

Complicated question

Never be tempted to use the following structure:

Select up to three of the options below and enter in the boxes opposite

Option A

Option B

Option C

Option D

Option E

Option F

--	--	--

Whilst this may be logically sound you will have problems when summarising these results. A very tedious and finally unsatisfactory search is required to determine how any given option performs.

Ranked responses

Place in order of importance to you the following features of a camping holiday

(Indicate by numbering from 1-4 in order where 1 is the most important)

Open air	
Mobility	
Cost	
People	
Atmosphere	

Rated responses

(Circle the number under the initials that applies.

VI=Very important; I=Important; N=Neutral; U=Unimportant;

VU=Very Unimportant).

Indicate your view of the following aspects of a camping holiday

	VI	I	N	U	VU
Community life	1	2	3	4	5
Low cost	1	2	3	4	5
Outdoor life	1	2	3	4	5
Ability to move around	1	2	3	4	5

Avoid double questions

Do you think the British should eat less and exercise more?

instead ask:

(Please circle relevant number)

	Yes	No
Do you think the British should eat less	1	2
Do you think the British should exercise more	1	2

Notice the consistent use of circled number responses for Yes/No questions.

Avoid questions that involve negatives

Don't confuse the respondent by language like this:

(Please circle relevant number)

Yes No

Are you against a ban on smoking	1	2
---	---	---

Ask for precise answers

Give your age on 1st September 2001

--	--

years

is preferable to:

Are you...

Under 18

18-65

Over 65

Avoid leading questions

Leading questions such as “Do you agree with the majority of people that the health service is failing?” should be avoided for obvious reasons that any right-minded individual can see. Don’t you agree?

Reference

Burgess (2001), "A general Introduction to the Design of Questionnaires for Survey Research", *Information Systems Services*, University of Leeds