

# Sampling

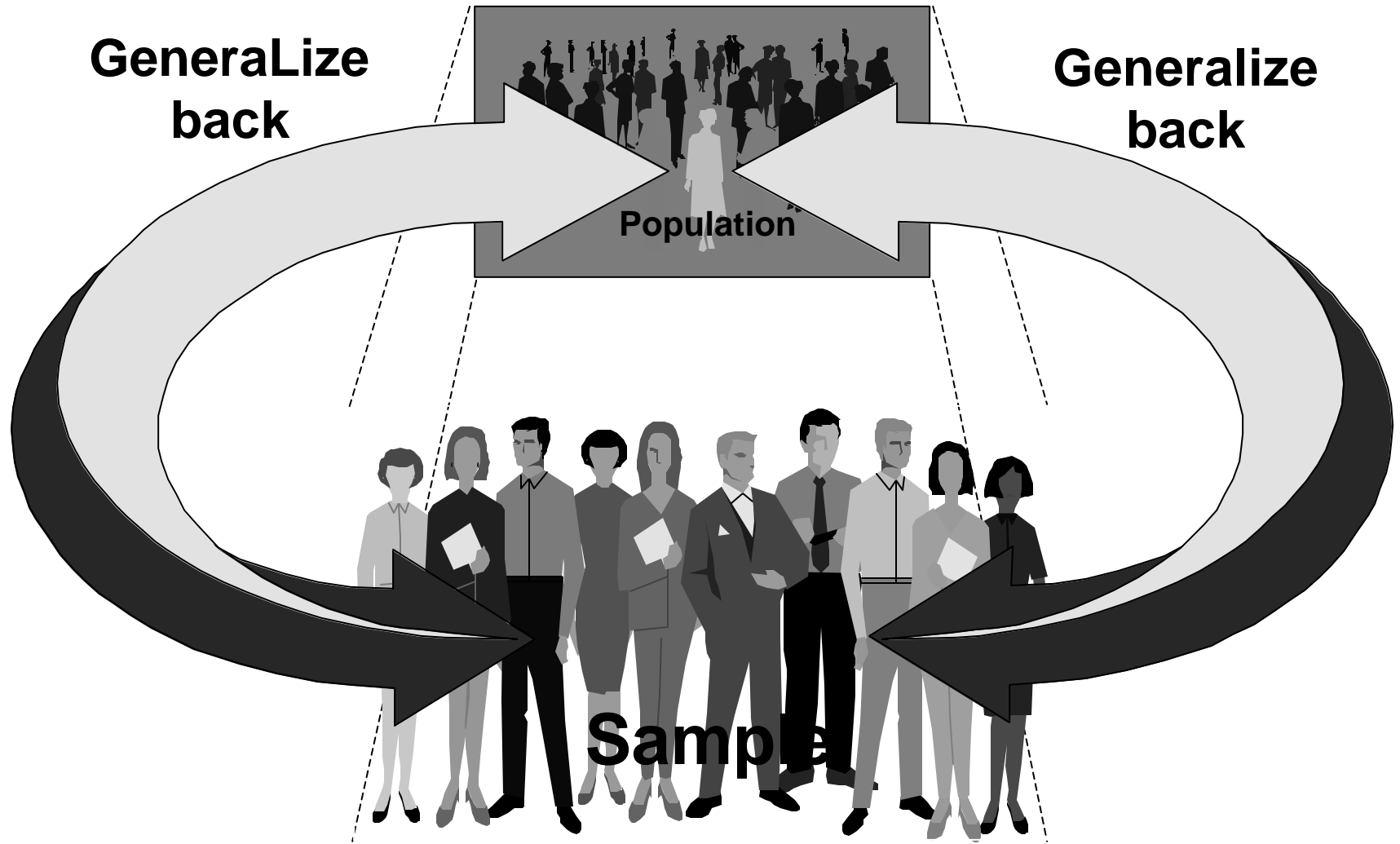
# Why Sample?

Because *heterogeneity* or *variability* in population requires sampling

--homogenous populations require only a single sample

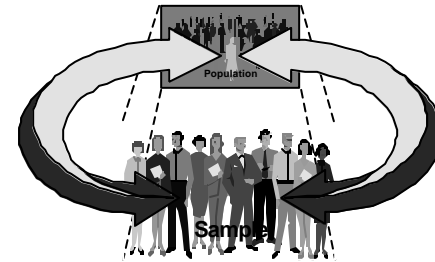
A sample is said to be “*representative*” if the aggregate characteristics of sample approximate the same characteristics of population

# Sampling affects External Validity

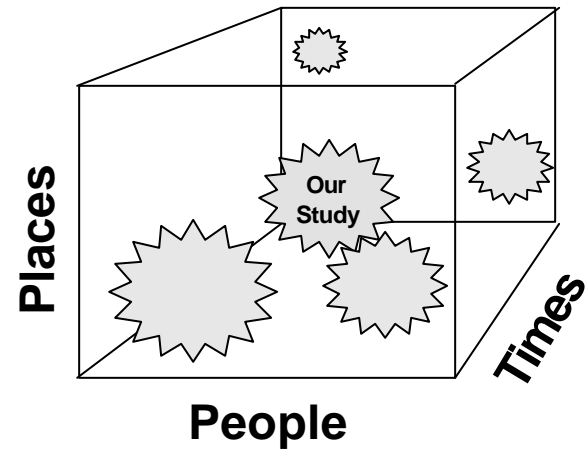


# Improving External Validity

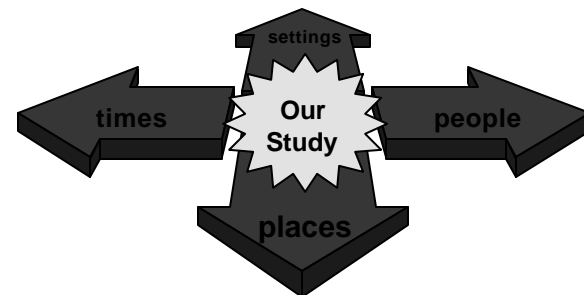
Random Sampling



Replicate,  
Replicate,  
Replicate



Use theory



# Sampling Terminology

sampling—process of selecting observations

- ***parameter***—summary description of a given variable in a population
- ***statistic***—summary description of a given variable in a sample
- ***population***—theoretically specified aggregation of sampling elements—aggregation of elements from which the sample is actually selected
- ***element***—unit about which information is collected
- ***sampling unit***—element or set of elements considered for selection in some stage of sampling
- ***sampling frame***—actual list of sampling units from which the sample is selected--sampling frames often define the survey population

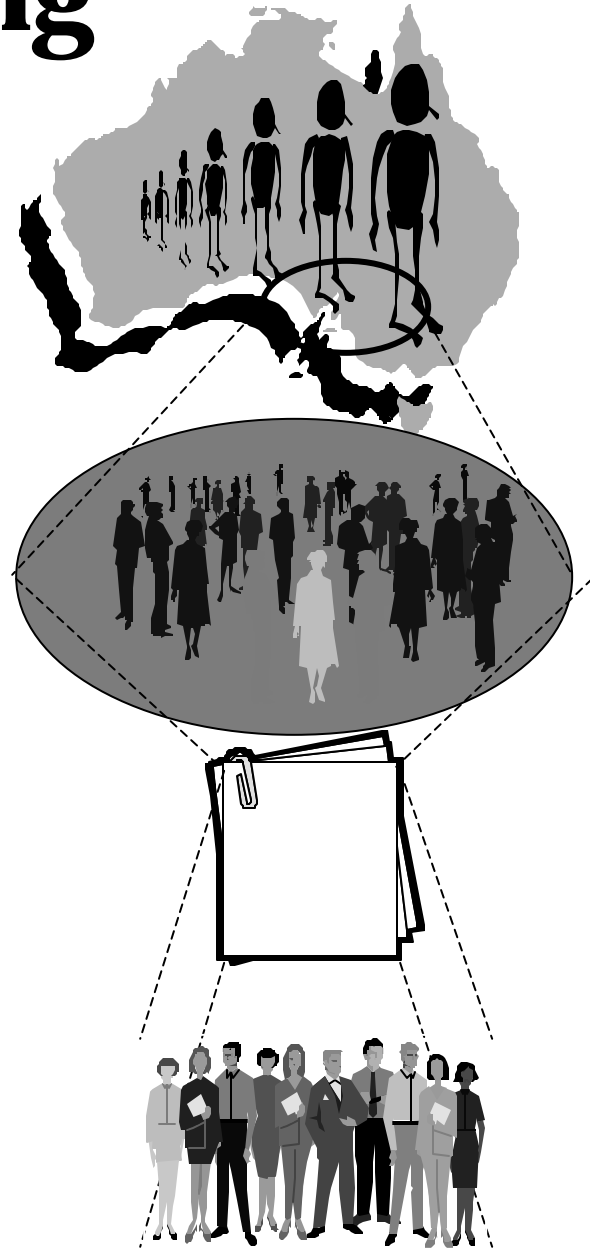
# Groups in Sampling

The theoretical population

The study population

The sampling frame

The sample



# What Can Go Wrong?

The theoretical population

?

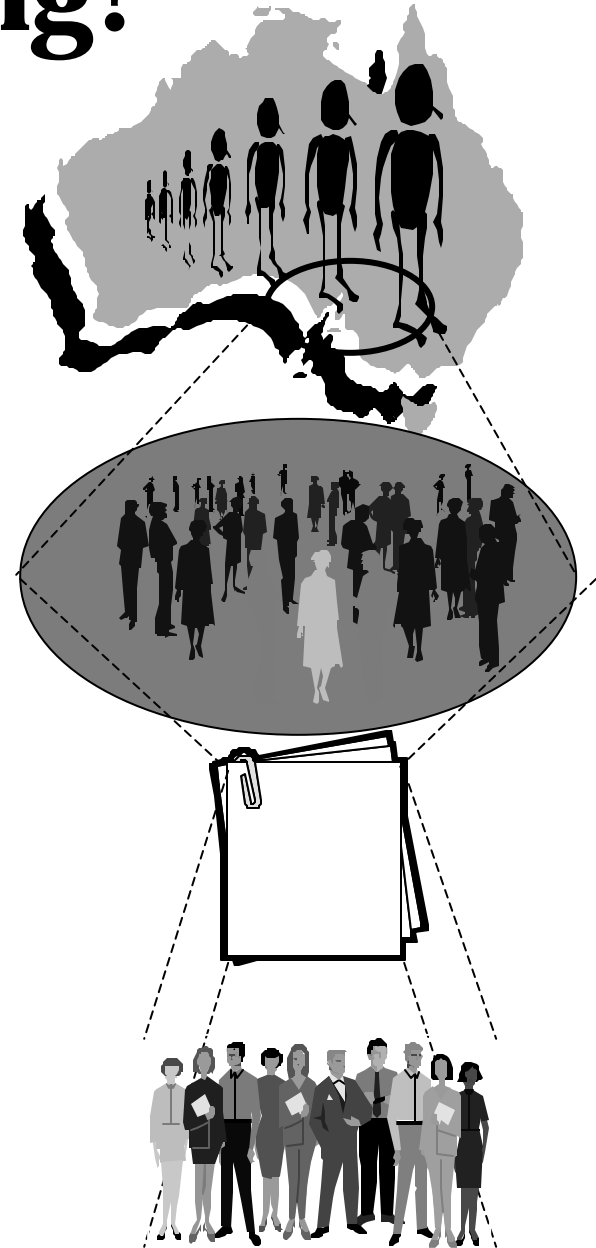
The study population

?

The sampling frame

?

The sample



# Two Major Types of Sampling Methods

## Probability sampling

**Uses some form of random selection.**

**Requires that each unit have a known (often equal) probability of being selected.**

## Nonprobability sampling

**Selection is systematic or haphazard, but not random.**

# **Types of Probability Sampling Designs**

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Cluster (area) sampling
- Multistage sampling

# Simple Random Sampling

- Objective: Select  $n$  (sample) units out of  $N$  (population) such that every element has an equal chance of being selected
- Procedure: Use table of random numbers, computer random number generator or random device
- Can sample with or without replacement.
- $f=n/N$  is the sampling fraction

# Simple Random Sampling

## Example

1. Select 30 random students from our Applied Research class
2. Assign each student a number from 1 to 140 (N)
3. Select students based on last 3 digits

**39634 62349 74088 65564 16379 19713 39153 69459 17986 24537  
14595 35050 40469 27478 44526 67331 93365 54526 22356 93208  
30734 71571 83722 79712 25775 65178 07763 82928 31131 30196  
64628 89126 91254 24090 25752 03091 39411 73146 06089 15630  
42831 95113 43511 42082 15140 34733 68076 18292 69486 80468  
80583 70361 41047 26792 78466 03395 17635 09697 82447 31405  
00209 90404 99457 72570 42194 49043 24330 14939 09865 45906  
05409 20830 01911 60767 55248 79253 12317 84120 77772 50103  
95836 22530 91785 80210 34361 52228 33869 94332 83868 61672  
65358 70469 87149 89509 72176 18103 55169 79954 72002 20582  
72249 04037 36192 40221 14918 53437 60571 40995 55006 10694  
41692 40581 93050 48734 34652 41577 04631 49184 39295 81776**

Result: Select students 50,126,113,37,88,47,90,82...etc.

# Systematic Random Sampling

- Number the units in population from 1 to  $N$ .
- Decide on the  $n$  that you want or need.
- $N/n=k$  the interval size.
- Randomly select a number from 1 to  $k$ .
- Take every  $k$ th unit.

# Example of Systematic Random 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

**Stratified sampling—divide population into homogenous subsets and select the appropriate number from each subset**

- To insure representation of each strata
- Oversample population subgroups
- Convenience
- Increase precision if strata are homogeneous within

# Stratified Random Sampling

List of students in Applied Research

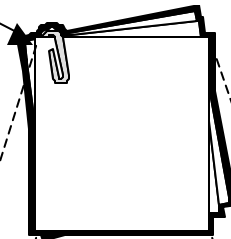


LBDV

LBRM

Others

Strata



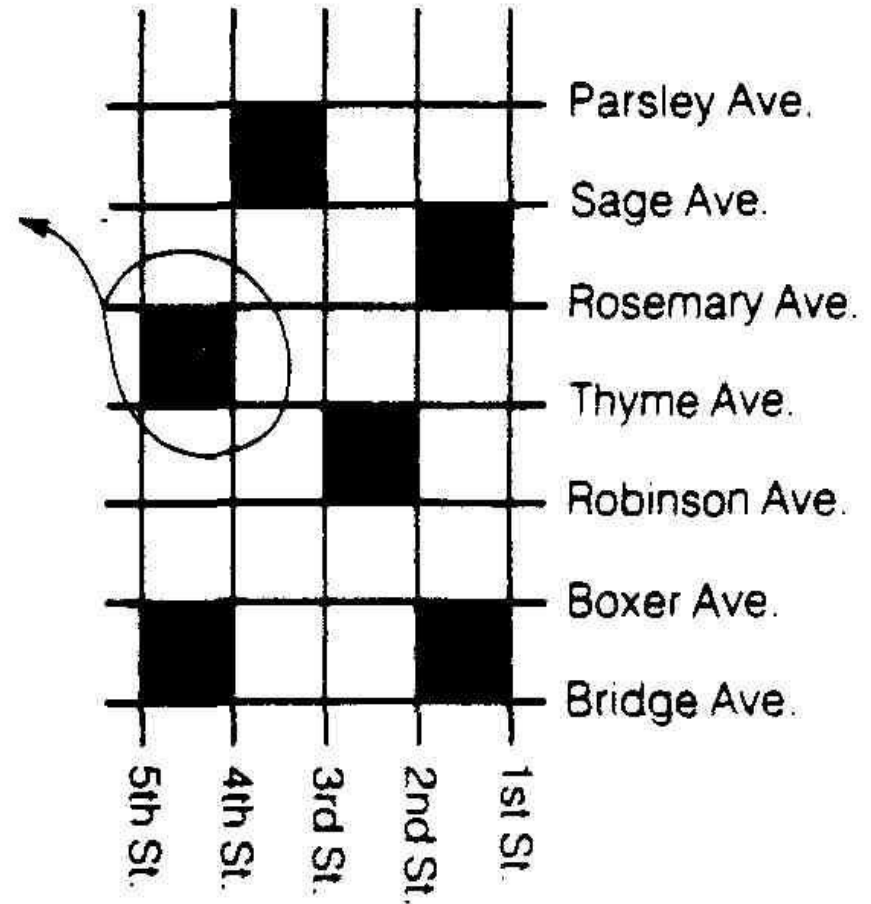
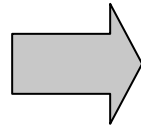
Random subsamples of  $n/N$

# Cluster (Area) Random Sampling

Cluster sampling—used when impossible or impractical to compile an exhaustive list of the elements composing the target population

- Divide population into clusters.
- Randomly sample clusters.
- Sample units within sampled clusters.

Example: Selecting city residents by sampling city blocks



# What is the difference between a “stratum” and a “cluster”?

- strata are relatively “homogeneous” subsets of a population based on some attribute (e.g., class, race, gender)
- clusters are relatively “heterogenous” subsets of population that are organized by space or geography (e.g., block, park entrance)

# Multi-Stage Sampling

- Cluster (area) random sampling can be multi-stage.
- Any combinations of single-stage methods.

# Multi-Stage Sampling

Example: Choosing students from schools

- 1<sup>st</sup> Stage: List all schools then select a *sample* of schools
- 2<sup>nd</sup> Stage: List all classes within the sampled schools then select a sample of classes
- 3<sup>rd</sup> Stage List all students within each class then select a sample of students from each class.

# Proportionate vs. Disproportionate Sampling

*Concept: sample subpopulations disproportionately to insure sufficient numbers of cases from each*

- Proportionate: If sampling fraction ( $n/N$ ) is equal for each stratum (subpopulation)
- Disproportionate: Unequal sampling fraction ( $n/N$ ) in each stratum (subpopulation)
- Needed to enable better representation of smaller (minority groups)

# Non-probability Sampling

- Likely to misrepresent (non-representative) the population
- May be difficult or impossible to detect degree of misrepresentation

# Types of Nonprobability Sampling

## ■ Convenience or accidental sampling

“Man on the street” ...available or accessible clients or volunteer samples

*Problem:* No evidence for representativeness

## ■ Modal instance

Sample for the *typical* case, e.g., typical voter, typical Australian...

*Problem:* May not represent the modal group proportionately

## ■ Purposive

Sample pre-defined groups or deliberately sampling an *extreme* group

*Problem:* Proportionality and often lack of linkage to theory

## ■ Quota

Select people non-randomly according to quotas, either proportionately or nonproportionately

*Problem:* How do you pick the population characteristics? How do you know their true proportion in population?

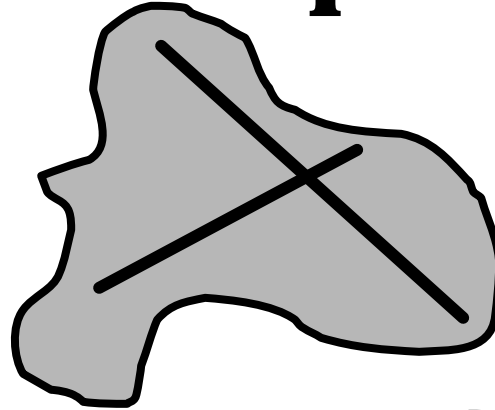
## ■ Snowball

One person recommends another, who recommends another, who recommends another, etc. Good way to identify hard-to-reach populations, for example, homeless persons

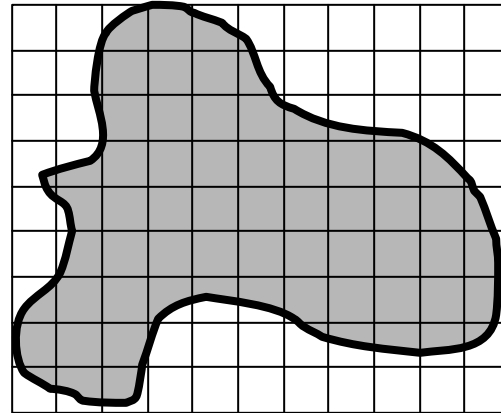
*Problem:* No evidence for sample representativeness

# Ecological Sampling Methods

- Transects



- Grids



- Points

