KERMAN UNIVERSITY OF MEDICAL SCIENCES

HIV/STI Surveillance Research Center, and
WHO Collaborating Center for HIV Surveillance

Sampling

Sirjan School of Medical Sciences

Dr. Hamid Sharifi
Professor in Epidemiology
Kerman University of Medical Sciences

چند نمونه بگیرم تا هم هزینه تحقیق کم شود و هم به اهداف مطالعه برسم؟
این نمونه‌ها را چگونه از بین افراد جامعه مورد نظر انتخاب کنم؟

Sampling vs. Census

When might you sample the entire population?

- When your population is very small
- When you have extensive resources
- When you don't expect a very high response

Advantage of sampling

Feasibility	• it may be the only feasible method of collecting data.
Reduced cost	• sampling reduces demands on resource such as finance, personal and material.
Greater accuracy	• sampling may lead to better accuracy of collecting data.
Greater speed	• data can be collected and summarized more quickly.

Disadvantage of Sampling

Validity and Reliability	• When the sampling is biased, or not representative or too small
Complicated	• When the population is very large and there are many sections and subsections
Needs technical knowledge	• The researcher does not possess the necessary skill and technical knowledge in sampling procedure, then the outcome will be devastated.

Sampling Theory Concepts

- Acceptance rate or response rate
- Representativeness
- Generalizability

Response Rate

- Want to go for a high response rate
  - A higher response rate increases the representativeness of sample and generalizability of the study results
- The characteristics of the responders can be different than the ones of the non-responders

Factors Affecting Response Rates

- Length of survey
- How intense is the intervention?
- Is there any incentive for the participants?

- Is it an RCT?
  - People do not usually like to be randomized

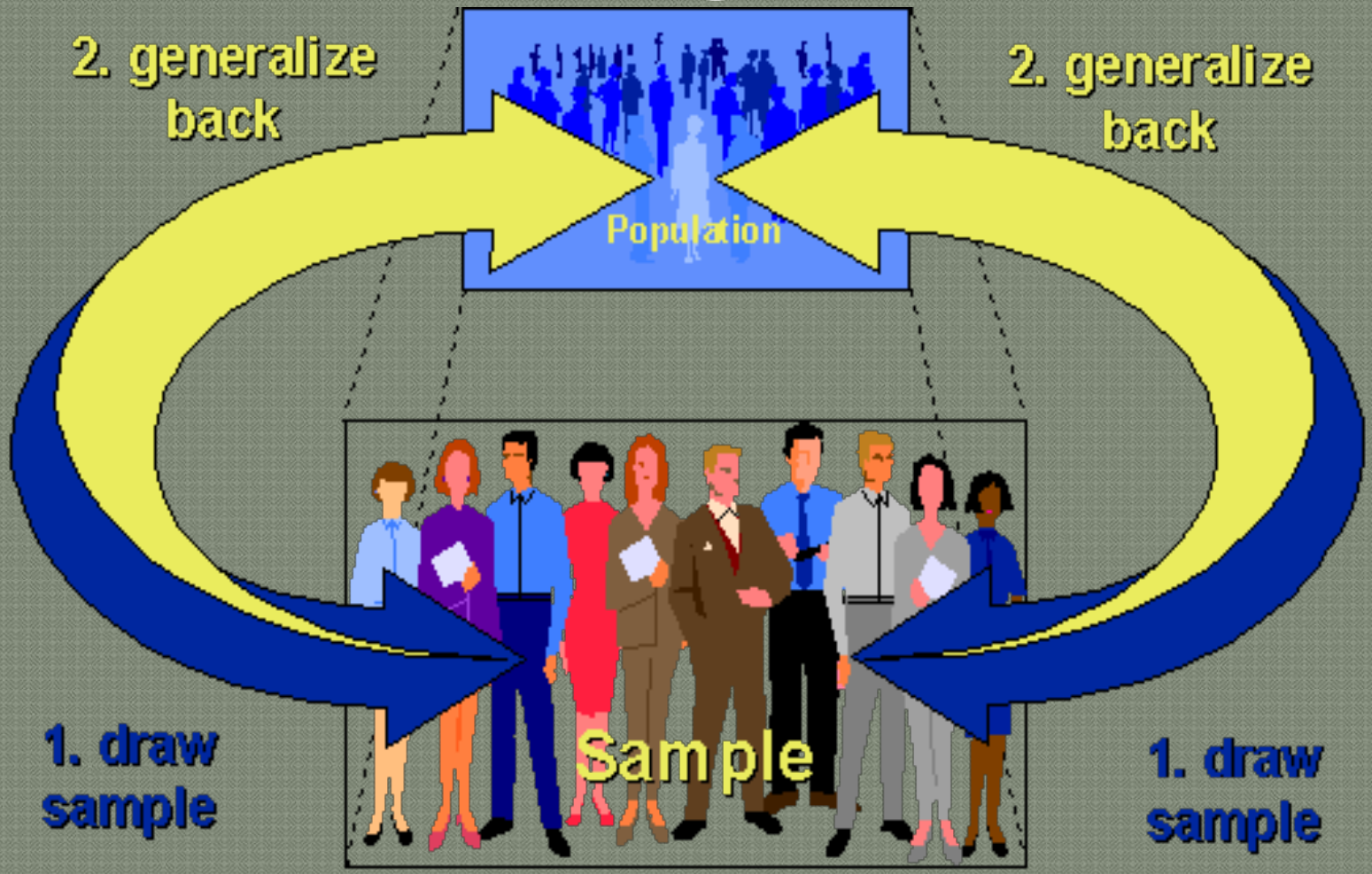
Representativeness

- Does the sample represent the general population of the persons with the specified problem?
  - Example: Sample of 1,200 students from Kerman city compare to the total population of 60,000 students on age, gender, smoking, and behaviors, etc.?

Generalizability

- Who is the sample generalizable to?
- The results are generalizable to the sampling frame
  - Example: The research results from the random sample of 1,200 students would be generalizable to the population of 60,000 students in Kerman city

Sampling and Generalization



Characteristics of Good Samples

- Three factors that influence sample representativeness
  - Sampling procedure
  - Sample size
  - Participation (response)

Symbols for Population and Sample Variables

Variable	Population	Sample
Mean	μ	$\bar{X}$
Proportion	π	p
Variance	σ^2	s^2
Standard deviation	σ	s
Size	N	n
Standard error of the mean	$\sigma_{\bar{X}}$	$s_{\bar{X}}$

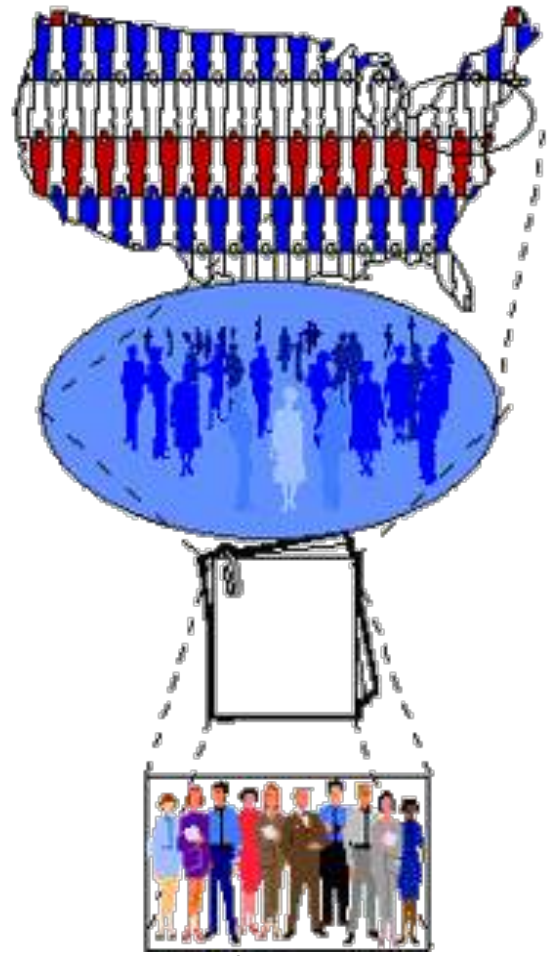
Sampling (Eligibility) Criteria

- Inclusion criterion
  - Who is in?
  - Need to specify demographic and clinical characteristics
- Exclusion criterion
  - Who do you want to keep out to avoid bias because they would provide poor data, be likely lost, or have ethical concerns?

Sampling Frame

- List of units from which sample is drawn
  - Defines your population
  - E.g., List of members of organization or community
- Ideally you'd like to list all members of your population as your sampling frame
  - Randomly select your sample from that list
- Often impractical to list entire population

Basics of sampling



Who do you want to generalize to?

The Theoretical Population

What population can you get access to?

The Study Population

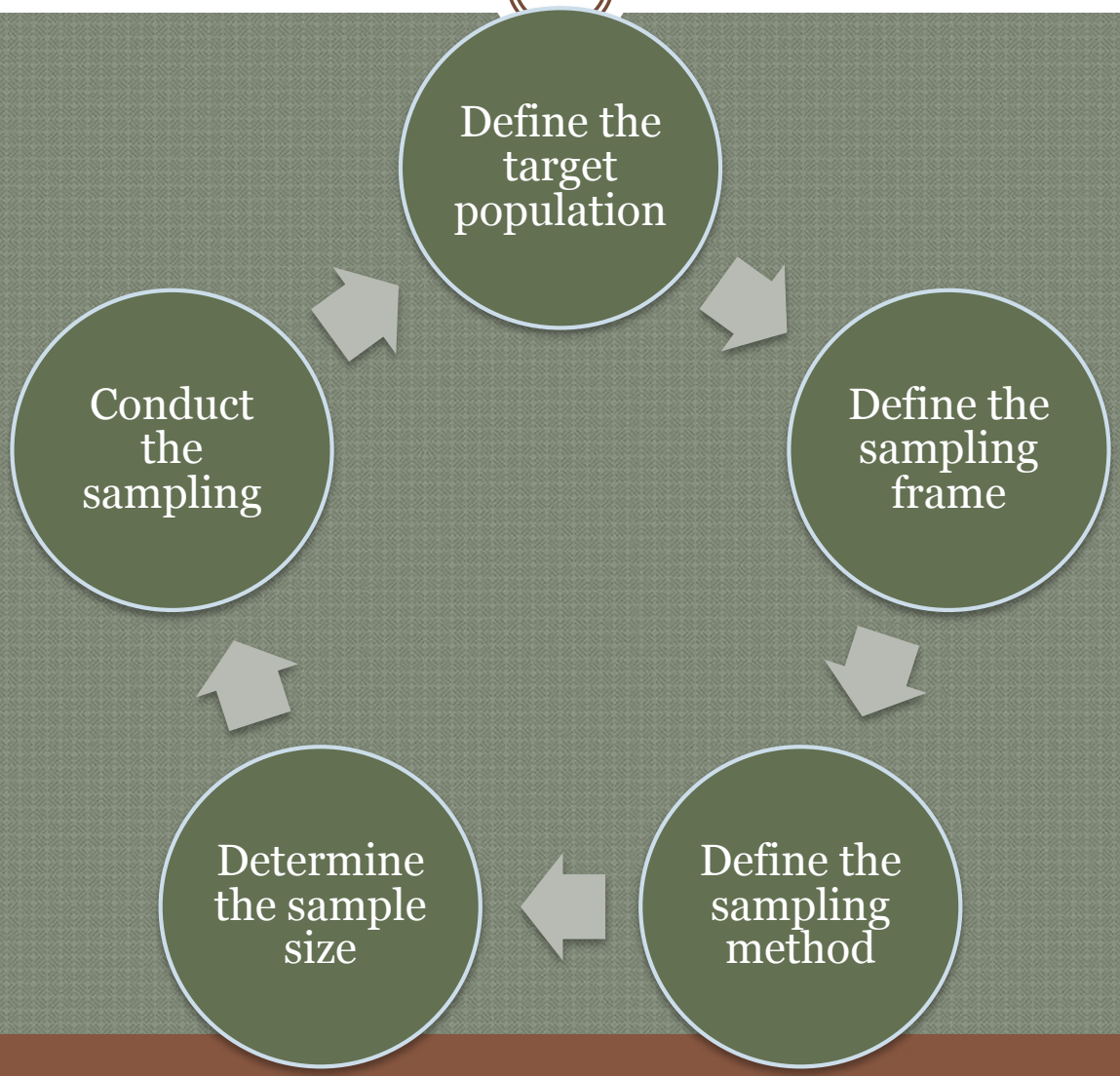
How can you get access to them?

The Sampling Frame

Who is in your study?

The Sample

Sampling approach



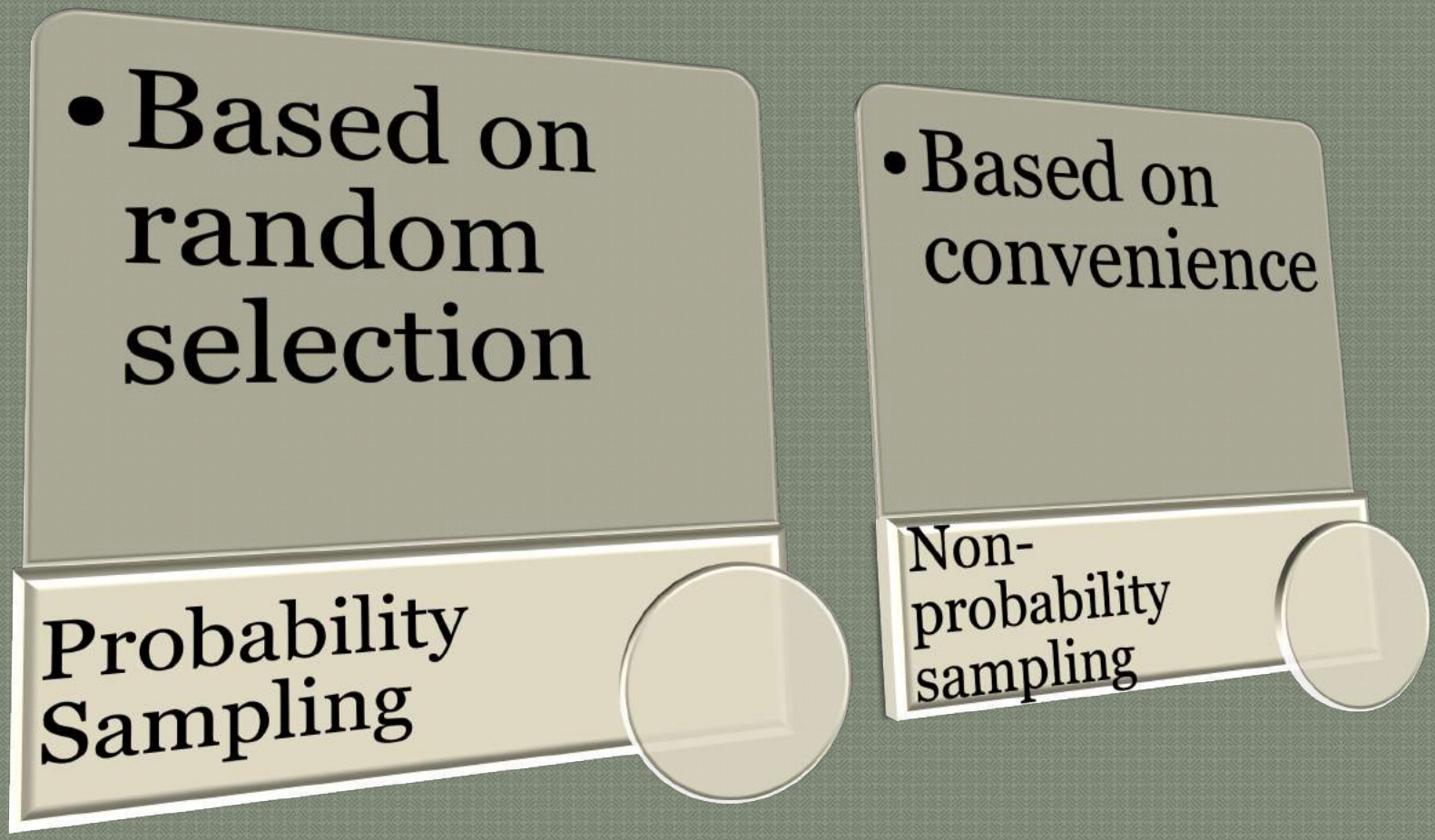
Part I

SAMPLING PROCEDURE

Characteristics of a Good Sampling Design

- Sample design must result in a truly representative sample.
- Sample design must be such which results in a small sampling error.
- Sample design must be viable in the context of funds available for the research study.
- Sample design must be such so that systematic bias can be controlled in a better way.
- Sample should be such that the results of the sample study can be applied with a reasonable level of confidence.

Sampling methods



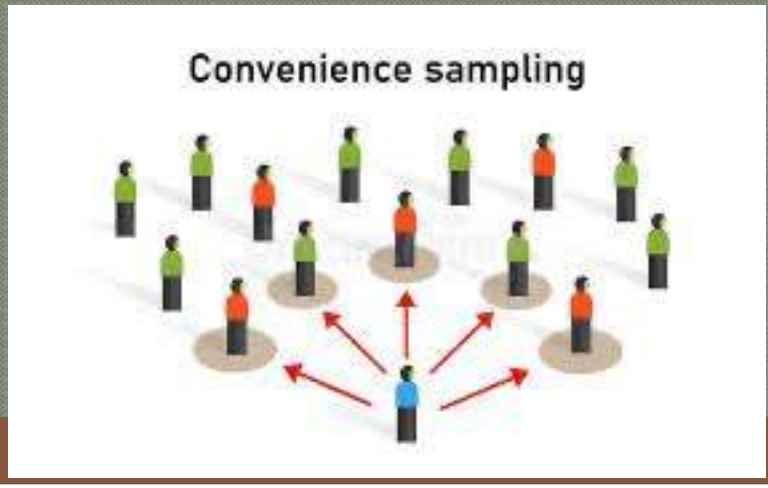
Non-probability Sampling

- In situations where sampling frame for random sampling doesn't exist

- Types of non-probability samples:
  1. convenience sampling (Reliance on available subjects)
  2. Purposive or judgmental sampling
  3. Quota sampling
  4. Network sampling methods

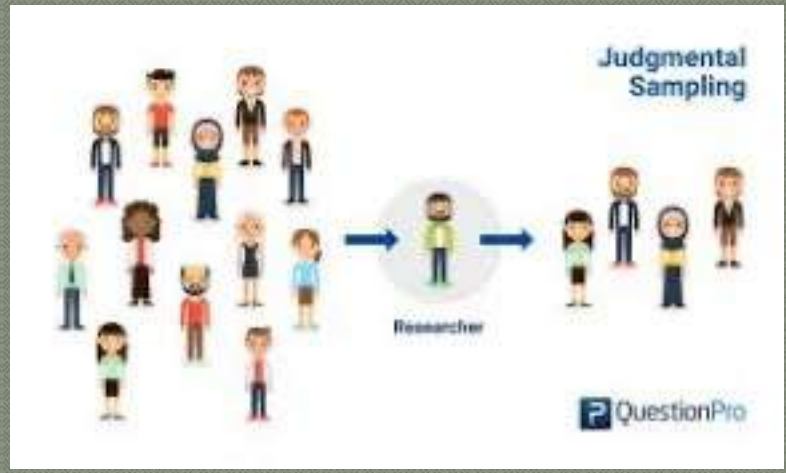
Convenience sampling

- easily accessible sample
- Examples:
  - Mall intercepts, college students, persons on the street
- Frequently used, but usually biased
- Notoriously inaccurate
  - Especially in making inferences about larger population



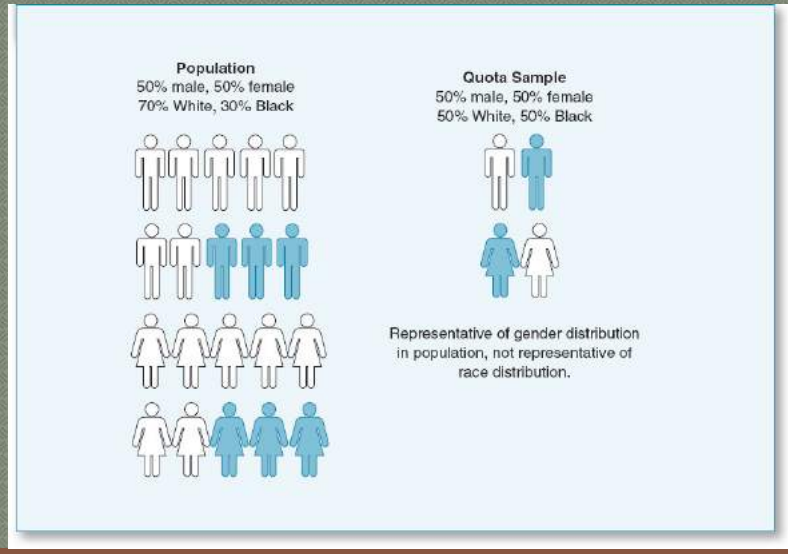
Purposive or Judgmental Sampling

- Dictated by the purpose of the study
  - Situational judgments about what individuals should be surveyed to make for a useful or representative sample
  - e.g., Hospitalized populations



Quota Sampling

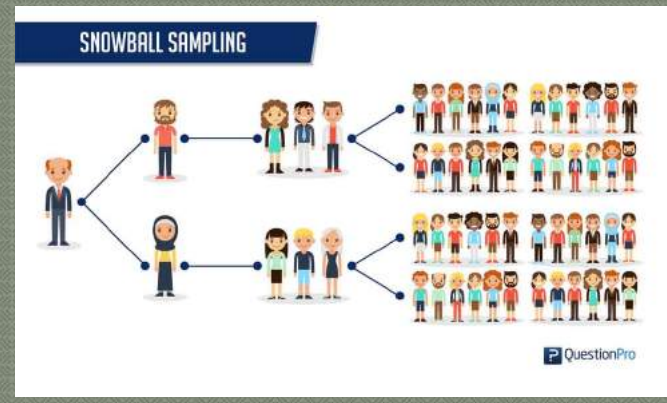
- Begins with a table of relevant characteristics of the population
  - Proportions of Gender, Age, Education, Ethnicity from census data
  - Selecting a sample to match those proportions
- Problems:
  - 1. Quota frame must be accurate
  - 2. Sample is not random



Network Sampling

- Snowball Sampling, Respondent Driven Sampling, Time-Location Sampling

- Used when population of interest is difficult to locate
  - e.g., homeless people
- Research collects data from of few people in the targeted group
  - Initially surveyed individuals asked to name other people to contact
    - Good for exploration
    - Bad for generalizability

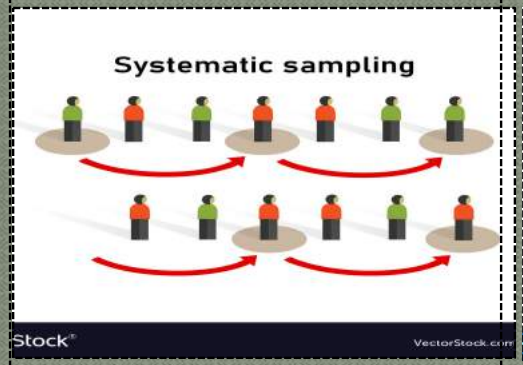


Probability Sampling

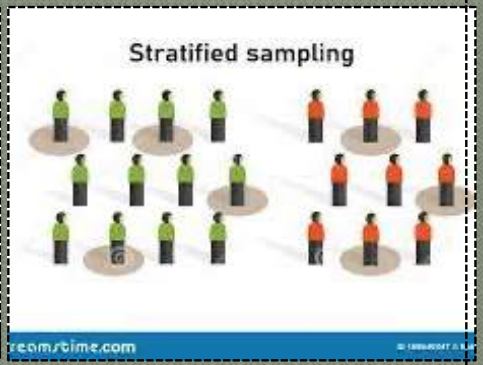
- Goal: Representativeness
  - Sample resembles larger population
- Random selection
  - Enhancing likelihood of representative sample
  - Each unit of the population has a chance of being selected into the sample

Types of Random Sampling Designs

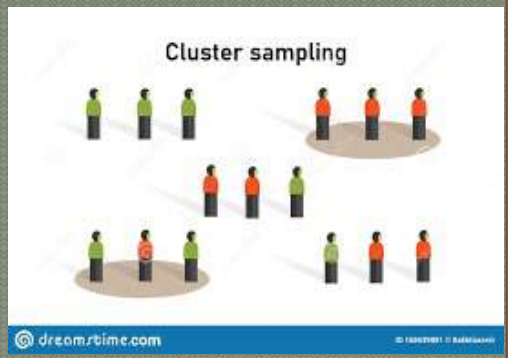
Simple Random Sampling



Systematic Sampling



Stratified Sampling



Cluster Sampling

Simple Random Sampling

- Establish a sampling frame
  - A number is assigned to each element
  - Numbers are randomly selected into the sample

انتخاب نمونه تصادفی ساده

- با کمک جدول **اعداد تصادفی** می توان **شانس یکسان** به افراد جامعه جهت انتخاب در نمونه داد.

- روش استفاده از این جدول بدین ترتیب است که :

بطور تصادفی از نقطه ای از جدول شروع به خواندن و یادداشت کردن اعداد می کنیم تا نمونه مورد نظر تامین گردد.

تعداد ارقام اعداد تصادفی که در جدول خوانده می شود با تعداد ارقام شماره آخرین فرد از جامعه مساوی باشد.

جدول ارقام تصادفی

01703	49894	57579	98505	85008	98681	56862	41860
87556	95669	39885	31669	31460	96413	84398	31562
84254	60541	73290	54685	80208	77044	14771	33378
12429	43566	32578	38935	75460	98133	18386	12417
63055	26768	63609	92424	50808	95416	12795	50787
18348	79628	05778	72095	90754	90430	00791	38023
19827	95727	02372	23485	54372	89732	67768	72151
30236	52309	99971	44890	28522	92140	40703	16888
32160	42795	04959	73840	99110	07527	73725	19291
14832	30334	18047	38712	32931	85481	15378	25011
21151	02668	44154	95153	63213	70014	67531	52581
89677	82090	42211	75118	36233	25131	13314	33063
67129	12388	41678	51286	80948	91599	52652	02519
27808	23807	25424	35877	96308	45847	88287	88419
24646	88222	66395	24060	98186	81741	08675	36931
10030	79086	89464	28282	89252	14777	02033	42852
26512	51935	86185	75646	51698	89313	57145	85070
43334	27009	27879	73339	74387	14314	42078	

مثال:

می خواهیم نمونه ای به حجم n = ۱۰ از هفتاد نفر دانشجویان یک کلاس به صورت تصادفی انتخاب نمائید.

ابتدا به هر یک از دانشجویان شماره ای از ۰۱ تا ۷۰ اختصاص می دهیم.

۰۱ ۰۲ ۰۳ ..... ۶۹ ۷۰

چون شماره آخرین فرد دو رقمی است با کمک **جدول اعداد تصادفی** ده عدد دو رقمی انتخاب می نمائیم. بدیهی است که ارقام دو رقمی بزرگتر از هفتاد در نظر گرفته نمی شود. ارقام تکراری نیز در نظر گرفته نمی شود. به عبارت دیگر:

از یک نقطه تصادفی شروع کرده و ده عدد دورقمی غیرتکراری بین ۰۱ تا ۷۰ پیدا می کنیم.

جدول ارقام تصادفی

01703	49894	57579	98505	85008	98681	56862	41860
87556	95669	39885	31669	31460	96413	84398	31562
84254	60541	73290	54685	80208	77044	14771	33378
12429	43566	32578	38935	75460	98133	18386	12417
63055	26768	63609	92424	50808	95416	12795	50787
18348	79628	05778	72095	90754	90430	00791	38023
19827	95727	82372	23485	54372	89732	67768	72151
30236	52309	99971	44890	28522	92140	40703	16888
32160	42795	04959	73840	99110	07527	73725	19291
14832	30334	18047	38712	32931	85481	15378	25011
21151	02668	44154	95153	63213	70014	67531	52581
89677	82090	42211	75118	36233	25131	13314	33063
67129	12388	41678	51286	80948	91599	52652	02519
27808	23807	25424	35877	96308	45847	88287	88419
24646	88222	66395	24060	98186	81741	08675	36931
10030	79086	89464	28282	89252	14777	02033	42852
26512	51935	86185	75646	51698	89313	57145	85070
43334	27009	27879	73339	74387	14314	42078	

نقطه شروع

شماره افراد انتخاب شده: 48, 28, 52, 29, 21, 40, 70, 31, 68, 32

بنابراين دانشجويان با شماره هاي زير نمونه تصادفي را تشكيل مي دهند:

48, 28, 52, 29, 21, 40, 70, 31, 68, 32

محقق مي تواند به افراد مذكور مراجعه نموده و بررسي خود (تكميل پرسشنامه، معاينه باليني، نمونه برداري براي آزمايش هاي پاراكلينيكي، ....) را به انجام رساند.

Systematic Sampling

- Establish sampling frame
  - Select every k^{th} element with random start
  - e.g., 1000 on the list, choosing every 10^{th} name yields a sample size of 100
- Sampling interval: standard distance between units on the sampling frame
  - Sampling interval = population size / sample size
- Sampling ratio: proportion of population that are selected
  - Sampling ratio = sample size / population size
- Note: Order of the sampling frame could lead bias

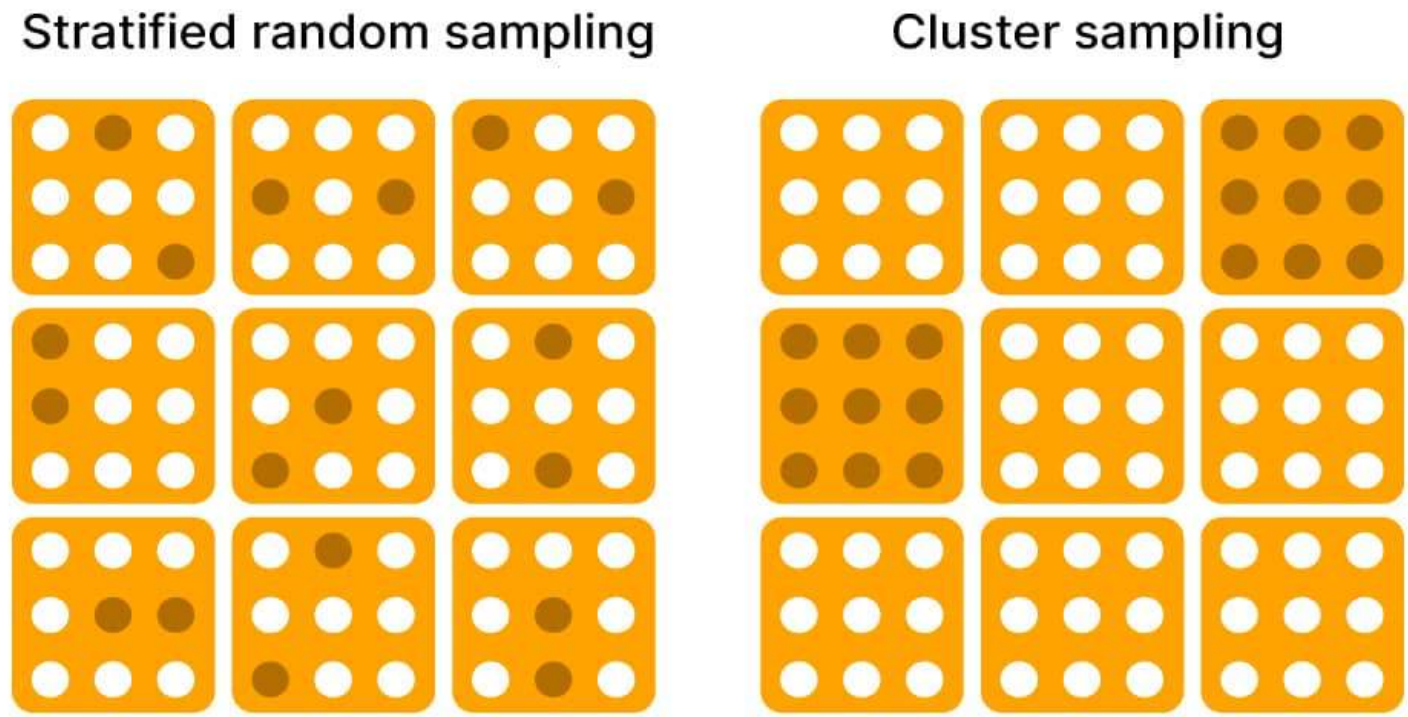
Stratified Sampling

- Modification used to reduce potential for sampling error
  - Research ensures that certain groups are represented proportionately in the sample
    - e.g., Stratifying by socioeconomic status of the people to make sure that each socioeconomic is proportionately represented

Cluster Sampling

- Frequently, there is no convenient way of listing the population for sampling purposes
  - Hard to get a list of the population members
- Cluster sample
  - Sample of schools
    - List of people for selected schools
    - Select sub-sample of students are in each school

Stratification vs. clustering



Sample in each cluster or stratum

- Proportion to size of each cluster or stratum
- Fixed sample in each cluster or stratum
  - Some stratification groups can be over-sampled for sub-group analysis
  - Samples are then weighted to restore population proportions

Multi-stage Sampling

- Sampling done in a series of stages:
  - List, then sample within each level
- Example:
  - Stage 1: Listing provinces
    - Randomly selecting provinces
  - Stage 2: List cities in each selected province
    - Randomly select a few cities
  - Stage 3: List schools (boys and girls) in each city
    - Try to include both genders schools AND select schools
  - Stage 4: List classes on selected schools
    - Randomly select schools
  - Stage 4: List students of selected classes
    - Randomly select students to interview

Multi-stage Sampling and Sampling Error

- Error is introduced at each stage
- One solution is to use stratification at each stage to try to reduce sampling error

Part II

SAMPLE SIZE

Factors to determine sample size

- Size of population
- Resources – subjects, financial, manpower
- Method of Sampling- random, stratified
- Degree of difference to be detected
- Variability (S.D.) – pilot study, historical
- Degree of Accuracy (or errors)
  - Type I error (alpha) $p<0.05$
  - Type II error (beta) less than 0.2 OR Power of the test: more than 0.8 (80%)
- Statistical Formulae
- Dropout rate, non-compliance

Incorrect sample size

- Wrong conclusions
- Poor quality research (Errors)
  - Type II error can be minimized by increasing the sample size
- Waste of resources
- Loss of money
- Ethical problems
- Delay in completion

Estimation of one proportion

$$n = \frac{Z^2_{(1-\alpha/2)}\, P(1-P)}{d^2}$$

Estimation one mean

$$n = \frac{Z^2_{(1-\alpha/2)} S^2}{d^2}$$

two-sample comparison of proportions

$$n = \frac{[Z_{1-\alpha/2}\sqrt{2\bar{P}(1-\bar{P})} + Z_{1-\beta}\sqrt{P_1(1-P_1) + (P_2(1-P_2)}]^2}{(P_1 - P_2)^2}$$

two-sample comparison of means

$$n = \frac{(Z_{(1-\alpha/2)} + Z_{(1-\beta)})^2 (sd_1^2 + sd_2^2)}{d^2}$$

$d = m1 - m2$

Design effect

- **Design effect = 1 + (n-1)ρ** where ρ is the intra-class correlation.
- So if we know the simple random sample size required for a given power we need to multiply this by the design effect.
- For example our data has ρ=16.205/(16.205+139.367)=0.104
- So for schools of size 10 pupils we would need 1+9*0.104=1.94 times as many students (in total) to get the same power.

Finite population correction

SAMPLE SIZE DETERMINATION USING THE FINITE POPULATION CORRECTION FACTOR

$$n = \frac{n_0 N}{n_0 + (N - 1)} \qquad \textbf{(8.14)}$$

Critique The Sample

- What was the sampling frame?
- Were the inclusion and exclusion criteria identified?
- What sampling methods were used?
- Was there rationale for the sampling method?
- What was the response rate?
- Was there a power analysis?
- Was the sample large enough?
- Were the characteristics of the sample described?
- Was the sample representative of the population they were studying?
- Who is the sample generalizable to?

How Do I Deal with Sample Size In the Real World?

- Know that to detect a small effect, you need a larger sample
- Know that for every extraneous variable, you need a bigger sample
- Know that if you have a small sample, you may be underpowered
- Look to see if your results are in the expected direction

Sample Section for Your Research Proposal

- What is your sampling frame?
- What are the inclusion and exclusion criteria?
- What sampling methods will be used?
- What is the rationale for the sampling method?
- About how big will the sample be?
- Explain how subjects will be assigned to groups
- Who is the sample generalizable to?
- Discuss strengths and weaknesses of sampling method

Power Analysis

- Standard power of 0.8
- Level of significance
  - The alpha value can be set at .05, .01, .001
- Effect size
- Sample size

Effect Size

- Small, medium, or large effect of dependent (outcome) variable
  - Example: Change on the blood pressure. Do we want to get a change of 10mg., 20mg., or 30mg. mercury?
- Look at other studies to see what kinds of effect sizes they get and what kind of sample sizes they had to get those

Useful links

Sampling Sirjan School/sample_size.xls

https://sample-size.net/

https://www.gigacalculator.com/calculators/power-sample-size-calculator.php