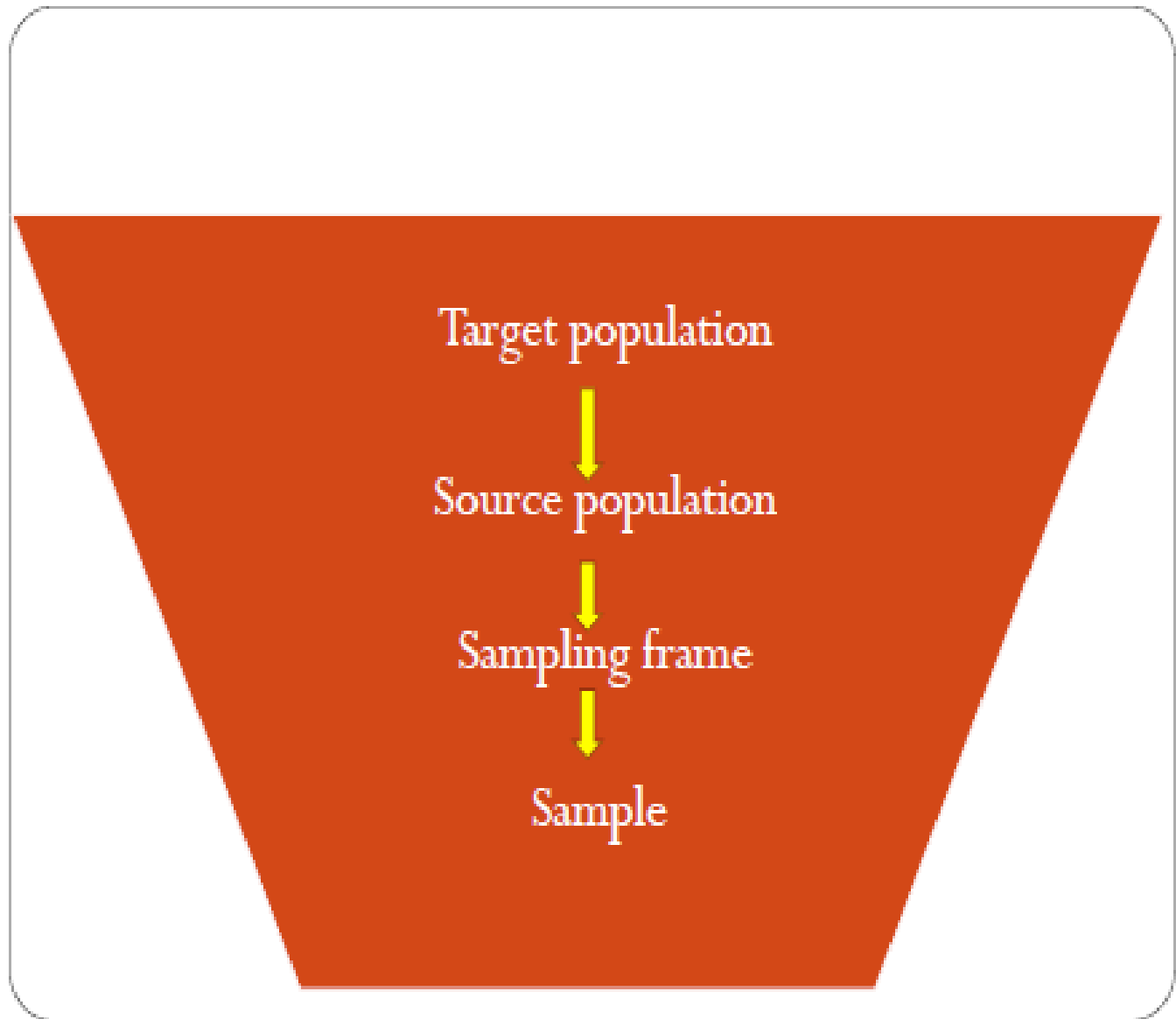


Sampling



Sampling

- **Sampling:** is the process of selecting a small number of elements from a larger defined target group of elements such that the information gathered from the small group will allow judgments to be made about the larger groups.
- The population is too large for us to consider collecting information from all its members.
- Instead we select a sample of individuals hoping that the sample is representative of the population.

Idea Behind Sampling

- The term sampling refers to strategies that enable us to **pick a subgroup from a larger group** and then use this subgroup as a basis for making inferences about the larger group.
- The researcher's goal is always **to generalize** about the population based on observations of the sample.
- Sampling strategies not only make it possible to collect data from a smaller number of respondents, but these strategies also make it possible to go into **greater depth** with this smaller number by asking more and deeper questions or by following up the structured questions with more open-ended or qualitative questions than would be possible with a larger group of respondents.

Sampling Strategy

- The manner in which a sample is drawn is an important factor in determining how useful the sample will be for making inferences about the population from which it is drawn.
- It is quite possible to have a very large sample upon which no sound decision can be based.
- This occurs because the respondents in the sample are not really similar to the population about which we want to make generalizations.
- To be useful, the **sample must be representative** of the **population** about which we wish to make generalizations.

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

- There is always a sampling error.
- Create a feeling of discrimination with the population.
- Inadvisable where every unit in the population is legally required to have a record.

Definition of sampling terms

→ **Target/Reference/source population**

- Is that population about which an investigator wishes to draw a conclusion.
- The population of interest, to which the investigators would like to generalize the results of the study, and from which a representative sample is to be drawn.
- **E.g.** for the study of prevalence of trachoma in children, all children in A.A

➔ **Study population (population sampled):**

- Population from which the sample actually was drawn and about which a conclusion can be made.
- The population included in the sample.
- In some instances, the target population and the population sampled are identical.
- e.g. number of students selected from all children attending schools in A.A

➔ Sampling unit

- **Sampling unit:** The *unit of selection* in the sampling process.
- For example, in a sample of districts, the sampling unit is a district; in a sample of persons, a person, etc.

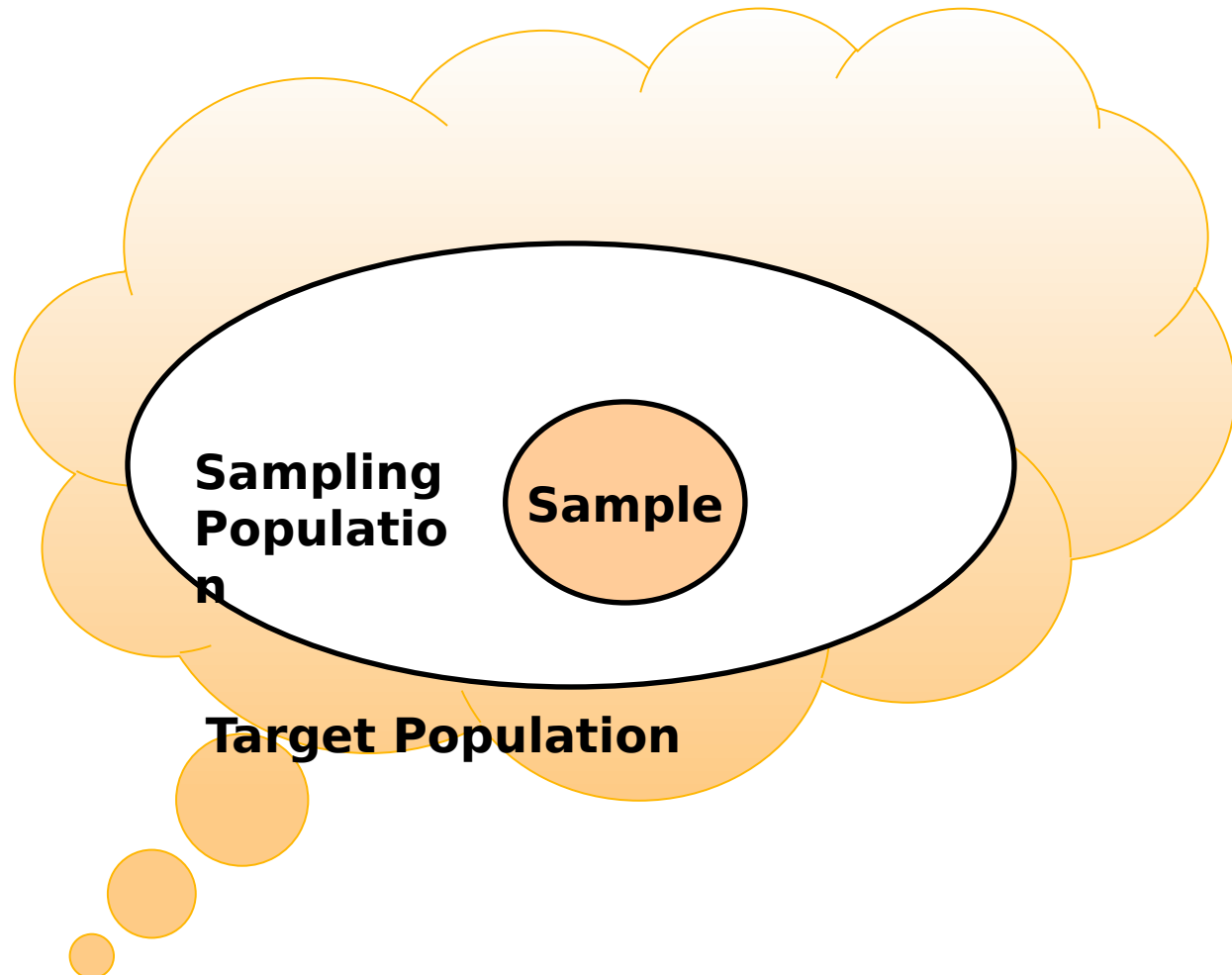
→ Study unit

- **Study unit:** The unit on which the ***observations will be collected***/ the unit on which information is collected.
- For example, persons in a study of disease prevalence, or households, in a study of family size.

Sampling unit and study unit ...

- If the objective is to determine the availability of latrine, then the study unit would be the **household**;
- If the objective is to determine the prevalence of trachoma, then the study unit would be the **individual**.
- **Sampling frame** - the **lists** of **all the units** in the reference population, from which a sample is to be picked.

Sampling and representativeness



Target Population □ **Sampling Population** □
Sample

Type of sampling technique

- **Depends on the**
 - ▢ **Type of research,**
 - ▢ **Population size and composition**
 - ▢ **Availability** of sampling frame

Types of Sampling methods

- I. **Non probability** sampling methods
- II. **Probability** sampling methods

Specifying the sampling method

▮ Probability Sampling

- ➔ Every element in the target population or universe [sampling frame] has **equal probability** of being chosen in the sample for the survey being conducted.
- ➔ Scientific, operationally convenient and simple in theory.
- ➔ Results may be **generalized**.

• Non-Probability Sampling

- ➔ Every element in the universe [sampling frame] **does not have equal probability** of being chosen in the sample.
- ➔ Operationally convenient and simple in theory.
- ➔ Results may **not be generalized**.

I. Non-probability sampling methods

- **When to use Non probability Sampling**

- ✦ **Group that represents the target population already exists.**
- ✦ **Difficult or impossible to obtain the list of names for sampling**
- ✦ **All of the cases of interest may not be identified ahead of time.**
- ✦ **For rare population.**

Cont'd...

Advantages

- ☑ Used ***when a sampling frame does not exist.***
- ☑ Quick, inexpensive and convenient.
- ☑ Good for ***pretests, pilot studies, In-depth interviews.***
- ☑ Used when ***precise representativeness is not necessary.***

Disadvantage

- No random selection(selection bias)
- Reliability cannot be measured.
- **No way to measure the precision of the resulting sample.**
- Inappropriate for **generalizing** findings

Types of Non probability Sampling

- 1. Convenience / haphazard**
- 2. Judgmental /Purposive**
- 3. Quota**
- 4. Snow ball**

1. Convenience/Haphazard sample

- ❖ Drawn at the ***convenience of the researcher.***
- ❖ Common in **exploratory research.**
- ❖ **Based on availability & accessibility**
- ❖ Often used in **face to face interviews**
- ❖ ***Does not lead to any conclusion.***

2. Purposive/Judgmental

- Sampling based on ***some judgment(characteristic), gut-feelings or experience of the researcher.***
- If **inference drawing is not necessary**, these samples are quite useful.
- Units from pre-specified groups are purposively sought out & sampled.
- Appropriate when the ***study subjects are difficult to locate.***
- **More efficient and economic**

Purposive sampling ...

In purposive sampling, we sample with a *purpose* in mind

1. Modal Instance Sampling

- ✓ In statistics, the *mode* is the most frequently occurring value in a distribution.
- ✓ we are sampling the most frequent case, or the "typical" case
- ✓ We could say that the modal voter is a person who is of average age, educational level, and income in the population

Purposive sampling...

II. Expert Sampling

- ✓ It involves the assembling of a sample of persons with known or demonstrable experience and expertise in some area.

3. Quota Sampling

- ✓ In quota sampling, ***you select people non randomly according to some fixed quota.***
- ✓ The participants of the study will be selected non randomly until the quota sample is fulfilled.

4. Snowball sampling

- ✓ In SBS, you begin by **identifying someone** who meets the criteria for inclusion in your study
- ✓ Then you ***ask them to recommend others*** who they may know who also meet the criteria
- ✓ SBS is especially useful when you are trying to reach populations that are inaccessible or hard to find
- ✓ For instance, if you are studying the homeless, you are not likely to be able to find good lists of homeless people within a specific geographical area.
- ✓ However, if you go to that area and identify one or two, you may find that they know very well who the other homeless people in their vicinity are and how you can find them

Snow ball...

- ❑ *Used in studies involving **respondents who are rare to find.***
- ❑ *To start with, the researcher compiles a short list of sample units from various sources.*
- ❑ *Each of these respondents are contacted to provide names of other probable respondents.*
- ❑ Uses a process of **“chain referrals”**
- ❑ **Used when trying to interview hard to reach groups**
- ❑ Also called **network, chain, or**

II. Probability sampling methods

- *Is more complex, more time-consuming and usually more costly*
- **Reduces possibility of selection bias**
- The *available sampling frame* should be used,
- ***Inferences can be made*** about the population.

The method selection

- the available sampling frame,
- how spread out the population is,
- how costly it is to survey members of the population and
- How users will analyze the data.

Methods used in probability samples

- A. Simple random sampling**
- B. Systematic sampling**
- C. Stratified random sampling**
- D. Cluster sampling**
- E. Multistage sampling**

A. Simple random sampling

- **SRS** is a method of probability sampling in which every unit has an **equal, non-zero and known chance** of being selected.
- **SRS** approach is conducted ***without replacement***.

➔ Principle

- ***Equal, known and non-zero chance/probability of drawing each unit***

Simple random sampling ...

→ Procedure

- Take study/sampling population
- Need listing of all sampling units (“**sampling frame**”)
- Number all units
- **Randomly draw units by:**
 - Lottery method
 - Table of random number
 - Computer generated random number

→ Assumption of the population

- **Homogeneity** with respect to the variable of interest
- **Availability of Sampling frame**

Simple random sampling ...

- **Advantages**

- Simple
- Sampling error easily measured

- **Disadvantages**

- Need complete list of units
- Does not always achieve best representativeness
- Units may be scattered and poorly accessible

Simple random sampling ...

Example: evaluate the prevalence of tooth decay among 1200 children attending a school

- List of children attending the school
 - Children numerated from 1 to 1200
 - Sample size = 100 children
 - Random sampling of 100 numbers between 1 and 1200
- **How to randomly select?**

How to Use Random Number Tables

1. Assign a unique number to each population element in the sampling frame. Start with serial number 1, or 01, or 001, etc. upwards depending on the number of digits required.
2. Choose a random starting position.
3. Select serial numbers across rows or down columns.
4. Discard numbers that are not assigned to any population element and ignore numbers that have already been selected.
5. Repeat the selection process until the required number of sample elements is selected.

How to Use a TRN to Select a Sample

- ***Your Biostatistics lecturer wants to randomly select 10 students from your class of 50 students. Here is how he can do it using a random number table.***

Step 1: Assign all the 50 members of the population a unique number. You may identify each element by assigning a two-digit number. Assign 01 to the first name on the list, and 50 to the last name.

- If this is done, then the task of selecting the sample will be easier as you would be able to use a 2-digit random number table.

Cont

NAME	NUMBER	NAME	NUMBER
A	01	P	32
.....			...
E	05	R	41
.....	...		...
G	10	T	47
.....	...		...
K	20		...
.....	...	W	49
N	30	Z	50

Generating random number using computer: OpenEpi

Expand All | Collapse

OpenEpi

Start Enter Results Examples Help

Home

Info and Help

Language/Options/Settings

Calculator

Counts

Person Time

Continuous Variables

Sample Size

Power

Random numbers

Download OpenEpi

Searches

Net Links

Development

dlm control ©2002-2003
by [Saskia Luchin](#)

Calculate

Clear

Random Number Generator		
Range	Lowest value	
	Highest value	
How many would you like?		
Format into how many columns?		5
Omit text from output?		No

Feed the information

Expand All | Collapse

OpenEpi

Start Enter Results Examples Help

Home

- Info and Help
- Language/Options/Settings
- Calculator
- Counts
- Person Time
- Continuous Variables
- Sample Size
- Power
- Random numbers
- Download OpenEpi
- Searches
- Net Links
- Development

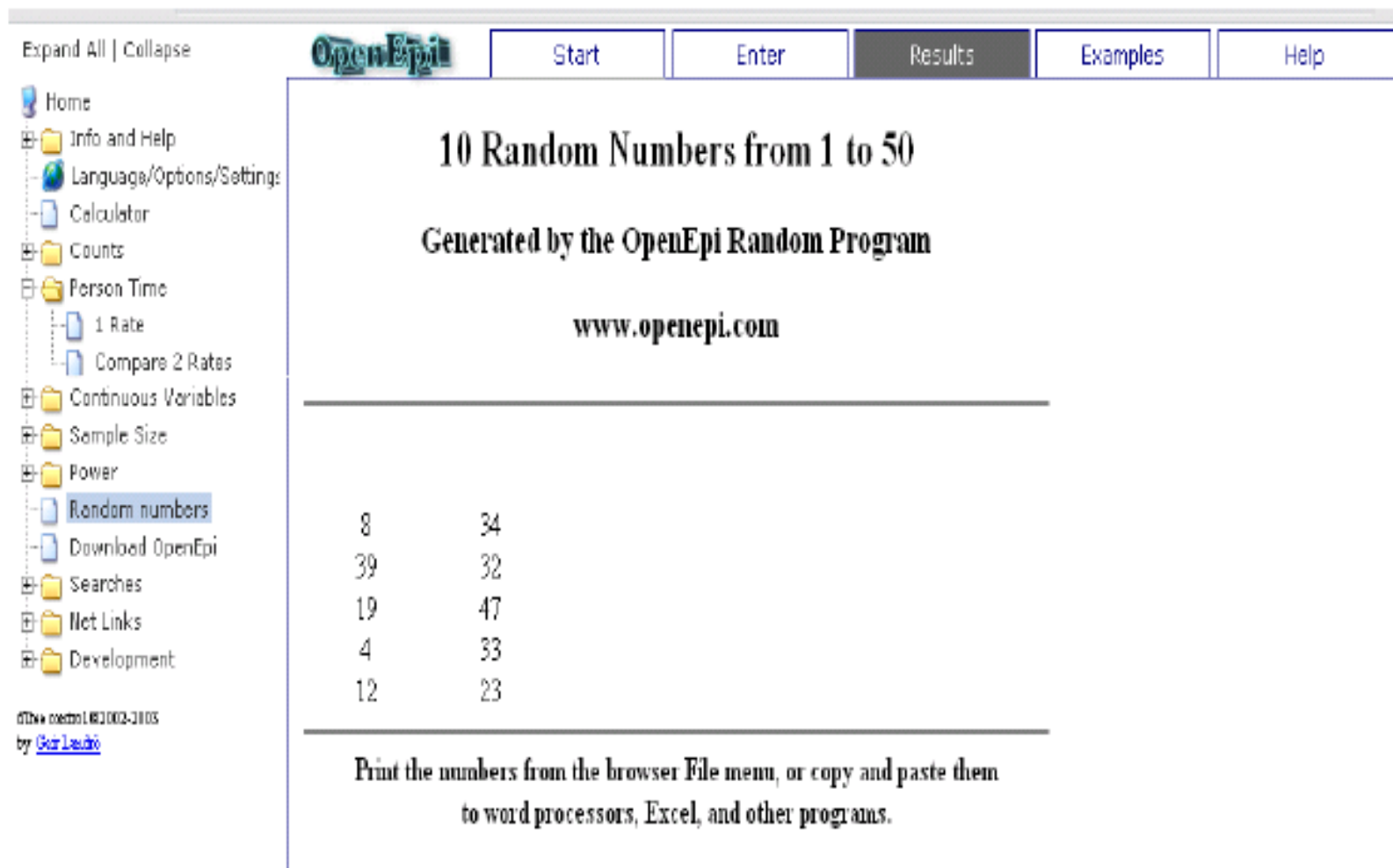
date: 0000-00-00 00:00
by: [OpenEpi](#)

Calculate

Clear

Random Number Generator		
Range	Lowest value	1
	Highest value	50
How many would you like?		10
Format into how many columns?		2
Omit text from output?		NO

The numbers generated are:



The screenshot shows the OpenEpi web application interface. The top navigation bar includes links for 'Expand All | Collapse', 'OpenEpi', 'Start', 'Enter', 'Results' (highlighted), 'Examples', and 'Help'. The left sidebar contains a tree view of the application's features, with 'Random numbers' selected under the 'Power' category. The main content area displays the title '10 Random Numbers from 1 to 50', the text 'Generated by the OpenEpi Random Program', and the website 'www.openepi.com'. Below this, a table lists 10 random numbers in two columns. At the bottom, a message instructs the user to print the numbers from the browser File menu or copy and paste them into word processors, Excel, and other programs.

Expand All | Collapse

OpenEpi

Start Enter Results Examples Help

Home

Info and Help

Language/Options/Settings

Calculator

Counts

Person Time

1 Rate

Compare 2 Rates

Continuous Variables

Sample Size

Power

Random numbers

Download OpenEpi

Searches

Net Links

Development

©The contents of 2002-2005
by [Gerr Leach](#)

10 Random Numbers from 1 to 50

Generated by the OpenEpi Random Program

www.openepi.com

8	34
39	32
19	47
4	33
12	23

Print the numbers from the browser File menu, or copy and paste them to word processors, Excel, and other programs.

Table of random numbers

57172	42088	70098	11333	26902	29959	43909	49607
33883	87680	28923	15659	09839	45817	89405	70743
77950	67344	10609	87119	15859	74577	42791	75889
11607	11596	01796	24498	17009	67119	00614	49529
56149	55678	38169	47228	49931	94303	67448	31286
80719	65101	77729	83949	83358	75230	56624	27549
93809	19505	82000	79068	45552	86776	48980	56684
40950	86216	48161	17646	24164	35513	94057	51834
12182	59744	65695	83710	41125	14291	74773	66391
13382	48076	73151	48724	35670	38453	63154	58116
38629	94576	48859	75654	17152	66516	78796	73099
60728	32063	12431	23898	23683	10853	04038	75246
01881	99056	46747	08846	01331	88163	74462	14551
23094	29831	95387	23917	07421	97869	88092	72201
15243	21100	48125	05243	16181	39641	36970	99522
53501	58431	68149	25405	23463	49168	02048	31522
07698	24181	01161	01527	17046	31460	91507	16050
22921	25930	79579	43488	13211	71120	91715	49881
68127	00501	37484	99278	28751	80855	02035	10910
55309	10713	36439	65660	72554	77021	46279	22705
92034	90892	09853	06175	61221	76825	18239	47687
50612	84077	41387	54107	09190	74305	68196	75634
81415	98504	32168	17822	49946	37545	47201	85224
38461	44528	30953	08633	08049	68698	08759	45611
07556	24587	88753	71626	64864	54986	38964	83534
60557	50031	75829	05622	30237	77795	41870	26300

B. Systematic Random Sampling

→ Systematic random sampling is a method of probability sampling in which the defined target population is **ordered** and the sample is selected according to **position** *using a skip interval*.

→ Interval Sampling used

→ *Individuals are chosen at regular intervals*

Advantages

- Easier to perform
- Require less time than SRS
- It is important when the *population is homogeneous*

Disadvantages

- Patterns/periodicity or **cyclic repetition**

Systematic sampling should not be used when a cyclic repetition is inherent in the sampling frame

Steps in Drawing a SRS

1. Obtain a list of units that contains an acceptable frame of the target population
2. Determine the number of units in the list and the desired sample size
3. Compute the **skip interval**
 - The population is listed in a particular order, then **every k^{th} unit** is **selected**, $k = N/n$ (interval size)
 - Randomly select an integer between 1 to k
4. Determine a random start point
5. Beginning at the start point, select the units by choosing each unit that corresponds to the skip interval.

More on Systematic Sampling

- Performance depends strongly on response patterns
 - Linear trend yields an implicit stratification, and works well
 - However cyclic variation of period k (or some multiple of k) can result in huge variability.

Systematic random sampling

- **Example:** the researcher wants to know the prevalence of malnutrition among under 5 children in woreda X.

children in woreda x

N = 100

want n = 20

N/n = 5

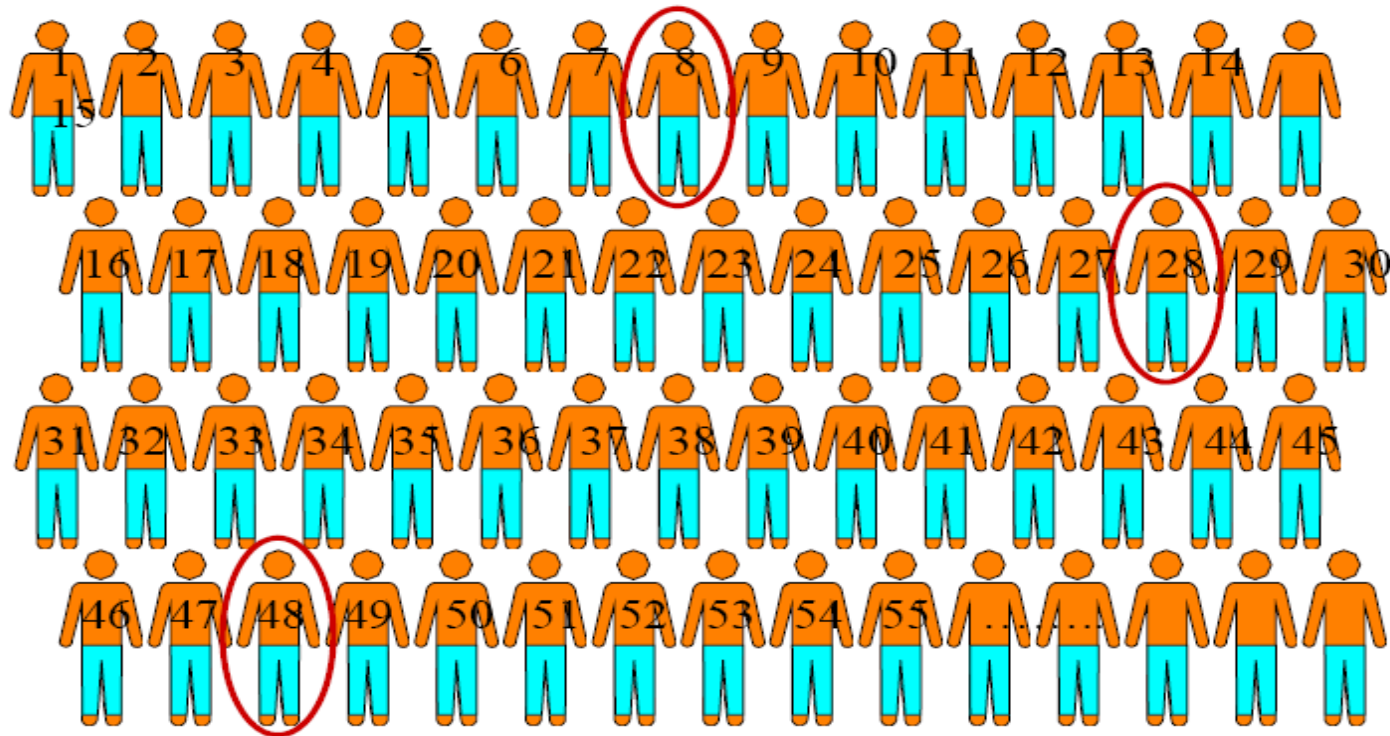
**select a random number from 1-5:
chose 4**

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

E.g. Systematic sampling

- $N = 1200$, and $n = 60$
 $\Rightarrow$ **sampling fraction** $= 1200/60 = 20$
- List persons from 1 to 1200
- Randomly select a number between 1 and 20 (ex : 8)
 $\Rightarrow$ 1st person selected = the 8th on the list
 $\Rightarrow$ 2nd person = $8 + 20 =$ the 28th
etc



Systematic sampling should not be used when a cyclic repetition is inherent in the sampling frame

Cont ...

Even though there is
frame, the
Population may be
heterogeneous.
So What?

C. Stratified sampling

- Appropriate when the population is **heterogeneous**.
- The population is divided into groups called **strata**
- A separate sample is then taken from each stratum by SRS

Stratified random sampling ...

- **Stratified Random Sampling** involves dividing your population into homogeneous subgroups and then taking a simple random sample in each subgroup.
- **Objective:** Divide the population into non-overlapping groups (i.e., *strata*) $N_1, N_2, N_3, \dots, N_i$, such that $N_1 + N_2 + N_3 + \dots + N_i = N$. Then do a simple random sample depending on the type of allocation
 - Proportional allocation: $n_i = \frac{n}{N} * N_i$

Example:

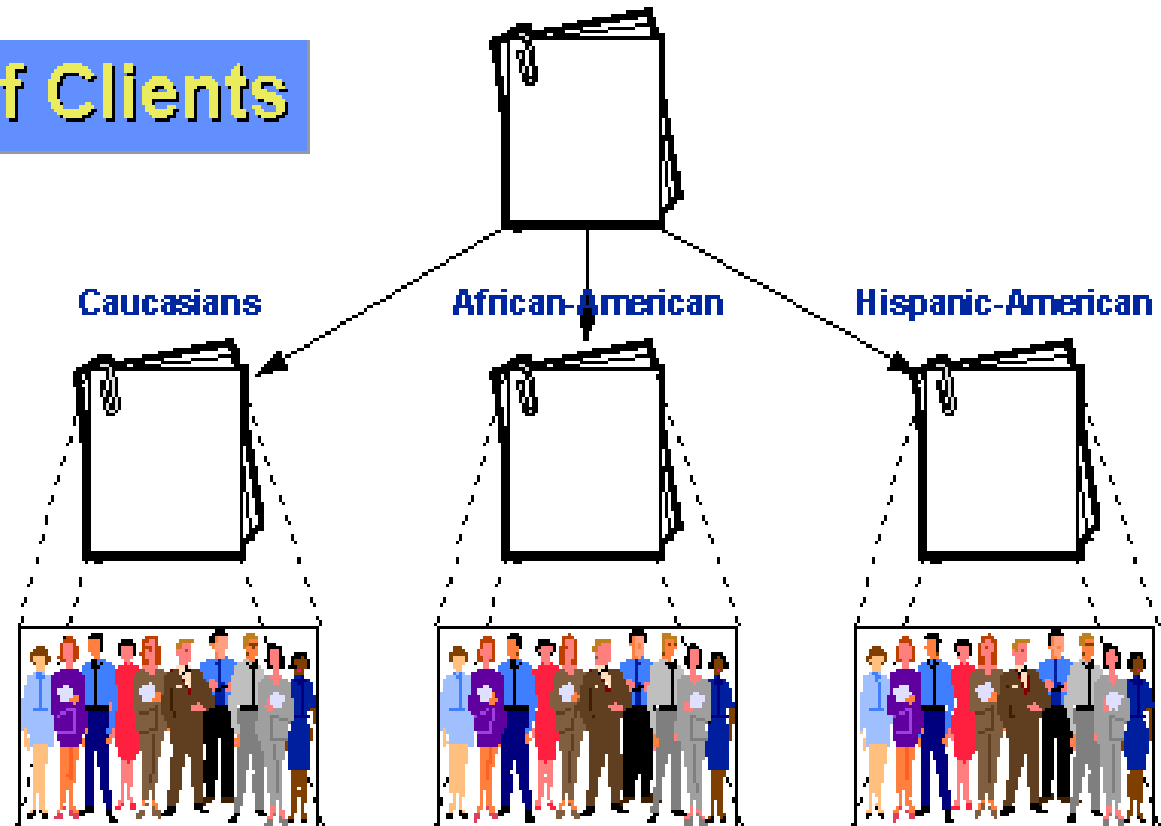
- An agency has clients from three ethnic groups and the agency wants to assess clients' view of quality of service for the last year.

Stratified random sampling

...

List of Clients

Strata



Random Subsamples of n/N

Stratified random sampling

...

Advantages

- More precise
- All subgroups represented

Disadvantages

- Difficult to select relevant stratification variables
- Not useful when no homogeneous subgroups
- Expensive
- Requires accurate information

D. Cluster sampling

- The problem with random sampling methods when we have to sample a population that's **disbursed across a wide geographic region** is that you will have to cover a lot of ground geographically in order to get to each of the units you sampled
- **Principle**
 - Sample units not identified independently but in a group (or “**cluster**”)
 - **Provides logistical advantage.**
- Involves selection of groups called **clusters**.
- Observation is made on all individual units within the selected cluster.
- **Used** in **homogenous** and **geographically scattered** population.

Cluster sampling ...

In cluster sampling, we follow these steps:

- ✓ divide population into clusters (usually along geographic boundaries) **e.g. *woreda, kebele or Got***
- ✓ **randomly sample clusters**
- ✓ **measure all units within sampled clusters**

Cluster Sampling ...

Example

- Consider the problem of sampling students in grade schools.
- We might begin with a national sample of **school districts** stratified by educational level.
- Within selected districts, we might do a SRS of schools.
- Within schools, we might do a SRS of classes.
- And, within **classes**, we might even do a SRS of **students**.
- In this case, we have 3 or 4 stages in the

Advantage of cluster sampling

- Sampling frame of the reference population is **not required**
- Less travel/resources required- **cost effective**

Disadvantage of cluster sampling

- ▮ Potential problem is that cluster members are more likely to be **alike**, than those in another cluster (homogenous)....
 - This “dependence” needs to be taken into account in the sample size....and the analysis (**“design effect”**)
- **Imprecise if clusters are homogeneous and therefore sample variation greater than population variation.**
- ▮ It is based on the assumption that the study units are uniformly distributed throughout the reference population.
- ▮ **You do not have total control over the final sample size**

E. Multi-stage sampling

- Many surveys use complex sample designs that combine several of the above elements in a multistage sampling framework
- Sometimes the population is too large and scattered for it to be practical to make a list of the entire population from which to draw a SRS.
- Suppose that each unit in the population can be divided into a number of smaller units, or subunits
- picking a sample from each chosen cluster
- Selection is done in stages
- Finally study subjects will be selected

Multistage ...

- The prime stimulus for multi-stage sampling is administrative convenience.
- As with cluster sampling, we select c of C clusters, but now instead of sampling **all units** in each cluster, we take a **random sample**.
- **Most large surveys carried out this way**

Multistage ...

Advantage

- Cuts the cost of preparing the sampling frame

Disadvantage

- Sampling error is high compared with simple random sampling

Conclusions

→ **Probability samples are the best**

❖ Ensure

➤ Representativeness

➤ Precision

→ within available constraints

Errors in sampling

- **Sampling (Random) error**: is any type of bias that is attributable to mistakes in either drawing a sample or determining the sample size.
 - **Errors introduced due to errors in selection of a sample.**
- **Non-sampling error (*measurement error*)**: occurs during the course of all stages or activities of research.
 - **Is any error which will be committed during data collection, coding, entry, and so on**

N.B: No sample is the exact mirror image of the population

Error in sampling ...

❑ **Sampling error (chance)**

✓ **Can not be avoided or totally eliminated**

➤ **Causes of sampling error**

- ▮ **One is chance:** is just because of bad luck
- ▮ **Design error**
- ▮ **Unrepresentativeness** of the sample

Cont'd

- **Sampling error can be reduced by**
 - **Large sample size** produces smaller errors than do small samples
 - **Homogeneous population** produce smaller errors than heterogamous population
 - **Well identified target and study population**
 - **Standardized study design and instrument use**

Error in Sampling ...

❑ **Non-sampling error**

- ✓ Observational error
- ✓ Respondent error
- ✓ Lack of preciseness of definition
- ✓ Error in editing, analysis and tabulation of the data