Population & Sample

Professor Dr Abubakir M. Saleh

Biostatistics NUR304

Fall semester

3rd week

21/10/2023

Outline

- **Study population.**
- **Sampling.**
- **Sample**
- **Reasons for sampling population.**
- **Methods of sampling.**
- **Errors in samples**
- **Sources of errors in Samples**
- **Summary**

Objectives

- Define sample and sampling;
- Describe different methods of sampling;
- Choose appropriate sampling methods for a specific research question;
- Identify different source of errors in samples.

Study population

Is the entire group of **individuals** having certain common **characteristics** about which statistical **inferences** can be made.

- The population must be clearly defined in terms of any characteristics relevant to the study: e.g.
  - Geographical location: e.g. Kurdistan, or Erbil city
  - Age: e.g. children under 5 years old, adults
  - Job: e.g. workers in tobacco factory
  - Gender: males or females
  - Others: e.g. students in medical university
  - Combination of the above: e.g. male adults in Erbil city

Reasons for sampling

- Generally, the population is too large for us to consider collecting information from all its members.

- Usually, it is not possible to study the entire population as this would take too much time and waste limited resources.
- Even when it is possible to survey a whole population, it is better to select a sample because with fixed study resources, more time and effort can then be spent on getting complete data from every individual sampled.

Sample

- Therefore, we select a sample of individuals, and just collect information from these, hoping that the sample is representative of the population.

- The sample selected must reflect the population and is dependent on the way the sample is selected.

Definitions

Target (or Reference) Population

- The population in which we are interested

  e.g. university students, children under 5 years old

Study Population

- The population from which we draw our sample

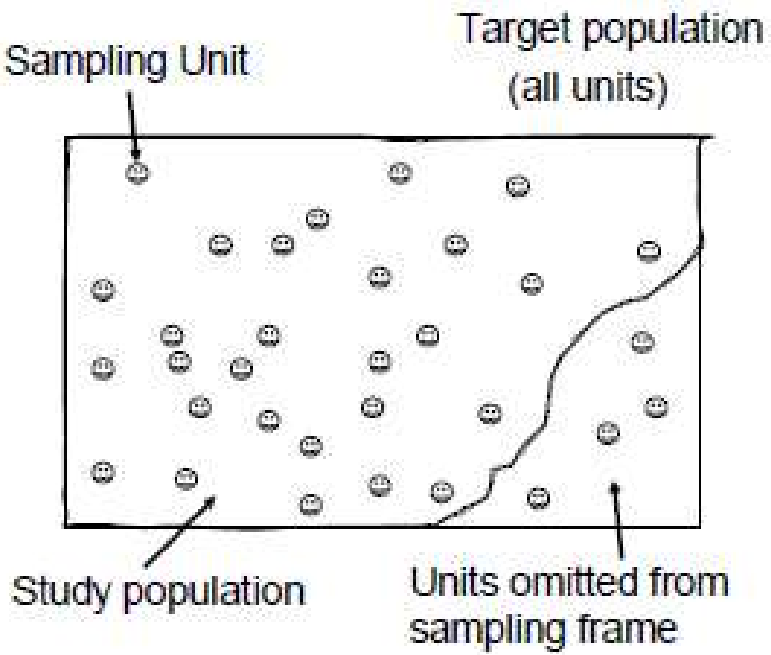
  e.g. student of faculty of nursing, children under 5 years in Erbil

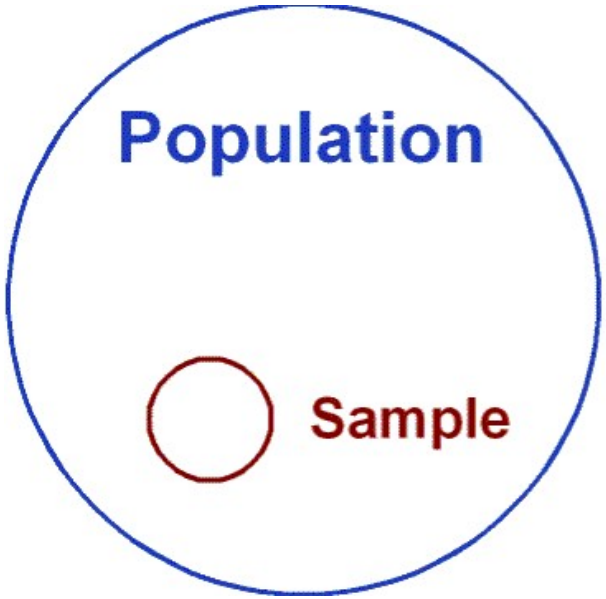
Sampling Unit

- A person (a child, a student), a household, a school, etc.

Sampling Design (or *Scheme)*

- The procedure for selecting the sampling units from the study population


Sampling Unit
Target population
(all units)
Study population
Units omitted from
sampling frame


Population
Sample

Definitions

Sampling Frame

- List of sampling units from which sample is selected

- List defines your study population

- *Examples:*
  - *University students – university registrar*
  - *Erbil city – list of quarters (Municipality, Governor)*

- A sampling frame may not give any chance of selection to some members of the study population (‘coverage error’)

Examples of definition

We conducted a study on the children in primary schools of Erbil city to measure their weights and heights

- **Target population:**

  Children in primary schools

- **Study population:**

  Children in primary schools in Erbil city

- **Sampling unit:**

  A primary school child

- **Sampling frame:**

  List of schools and list of children in the schools

Methods of sampling

- Random sampling methods (Probability sampling).
- Non-probability sampling methods (Non-random).

Definition of random sampling

Every person in the population has the same and equal chance (opportunity) to be included in the sample.

How to select a sample
"Random sampling"

- **Random** selection of subjects from the population
- **Chance alone** determines who will be included in your sample
- Sometimes called '**probability sampling**'
- Removes the possibility of bias
- Probability of selecting each unit **is known**

Random sampling methods

1. Simple random sampling (SRS).
2. Systematic sampling.
3. Cluster sampling.
4. Stratified random sampling.

1. Simple random sampling (SRS)

- The most basic sampling design

  Steps:

  1. Obtain a sampling frame that includes all units in the study population

  2. Number each unit from 1 to N (e.g. 1000)

  3. Generate a sequence of **n random** numbers between 1 and N (e.g. using random number tables, a calculator or computer)

Simple random sampling (SRS)(Cont.)

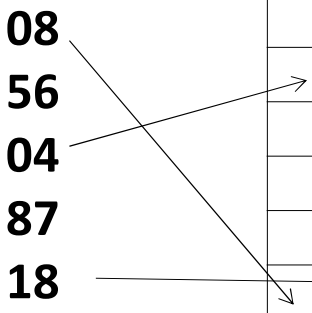
4.If any random number selected is greater than N, or has already been selected, ignore it and select another

5. Select each unit corresponding to the random numbers

TABLE 27. RANDOM SAMPLING NUMBERS

Each digit is an independent sample from a population in which the digits 0 to 9 are equally likely, that is, each has a probability of $\frac{1}{10}$.

84	42	56	53	87	75	18	91	76	66	64	83	97	11	69	41	80	92	38	75
28	87	77	03	57	09	85	86	46	86	40	15	31	81	78	91	30	22	88	58
64	12	39	65	37	93	76	46	11	09	56	28	94	54	10	14	30	73	80	30
49	41	73	76	49	64	06	70	99	37	72	60	39	16	02	26	91	90	16	54
06	46	69	31	24	33	52	67	85	07	01	33	16	33	43	98	17	62	52	52
75	56	96	97	65	20	68	68	60	97	90	46	63	37	10	34	41	64	85	01
09	35	89	97	97	10	00	76	39	82	49	94	15	89	60	65	57	03	91	68
73	81	11	08	52	73	64	85	22	72	85	16	15	97	76	28	41	95	00	33
49	69	80	41	46	62	26	32	58	16	88	76	54	32	06	37	46	45	28	95
64	60	49	70	33	73	71	57	83	26	19	25	86	21	64	60	11	01	86	70
93	05	36	44	59	19	99	51	54	21	37	48	18	60	22	92	68	34	39	02
39	88	11	26	68	92	81	14	12	16	37	64	61	48	21	69	77	76	33	00
89	34	19	12	83	76	35	11	96	53	04	76	63	10	93	68	52	42	73	20
77	29	03	26	45	36	15	17	27	28	79	58	38	98	73	52	63	72	48	41
86	75	51	29	70	78	24	78	94	78	64	17	32	23	95	52	87	79	14	30
95	98	77	51	14	65	76	49	42	36	11	33	23	89	32	01	60	48	91	44
22	09	01	14	04	96	97	56	92	52	83	44	45	08	72	78	10	36	26	70
30	49	36	23	36	81	11	76	91	08	67	60	01	15	64	77	21	33	72	29
77	59	88	92	17	75	04	47	18	02	94	84	71	44	87	63	06	04	49	33
03	50	80	26	74	74	18	85	92	20	64	39	98	68	29	26	90	14	77	36
46	32	79	69	41	06	26	04	47	24	67	10	66	69	21	55	66	63	48	47
65	73	98	08	05	96	92	27	22	86	54	87	95	87	40	27	09	97	47	21
68	82	77	73	08	37	28	47	73	49	10	65	53	48	87	74	02	99	52	86
93	98	12	19	82	69	61	08	00	42	88	83	70	85	08	48	74	94	88	61
61	27	39	16	42	17	89	81	27	44	12	33	43	24	92	41	55	13	45	01
54	74	04	79	72	61	21	87	23	83	96	56	97	63	67	02	67	30	36	89
28	00	40	86	92	97	06	22	37	37	83	00	97	17	08	06	43	95	76	84
61	78	71	16	41	01	69	63	35	96	60	65	09	44	93	42	72	11	22	85
68	60	92	99	60	97	53	55	34	61	43	40	77	96	19	87	63	49	22	47
21	76	13	39	25	89	91	38	25	19	44	33	11	36	72	21	40	90	76	95
73	59	53	04	35	13	12	31	88	70	05	40	43	42	47	17	03	86	14	10
85	68	66	48	05	24	28	97	84	84	91	65	62	83	89	68	07	51	01	02
60	30	10	46	44	34	19	56	00	83	20	53	53	05	29	03	47	55	23	26
44	63	80	62	80	80	99	43	33	87	70	52	51	62	02	12	02	90	44	44
89	38	13	68	31	31	97	15	35	67	23	74	76	96	62	82	62	19	65	58
55	20	77	12	79	81	42	15	30	67	88	83	69	08	99	82	20	39	92	40
67	40	42	16	46	06	60	74	61	22	95	47	24	62	81	06	19	67	15	06
57	19	76	98	65	64	55	28	34	03	58	62	35	22	67	40	04	88	17	59
21	72	97	04	82	62	09	54	35	17	22	73	35	72	53	65	95	48	55	12
46	89	95	61	31	77	14	14	24	14	91	58	76	56	19	33	98	67	09	04
99	73	85	64	96	58	61	65	60	83	62	10	87	00	82	63	39	90	83	17
85	52	98	27	40	33	09	59	80	17	22	06	84	03	41	48	76	07	26	69
50	12	17	86	50	57	91	28	42	29	83	87	00	87	93	52	53	47	08	65
92	84	02	93	44	36	93	19	08	54	76	62	31	65	94	68	38	04	62	31
69	74	30	25	68	65	19	77	57	05	71	56	91	30	16	66	70	48	78	65
51	69	76	00	20	92	58	21	24	33	74	08	66	90	61	89	56	83	39	58
27	25	81	29	75	02	85	09	58	89	77	83	03	40	21	14	45	90	54	01
44	03	62	96	68	65	24	57	44	43	07	72	59	16	04	94	23	36	55	85
40	59	49	20	48	63	35	74	33	12	96	25	59	35	07	45	80	97	19	90
92	91	07	14	82	22	50	70	75	15	69	71	31	20	60	06	99	56	57	74



No.	Name	Sex	Age
1	Ayad Ali	M	22
2	Basam Amin	M	24
3	Huda Hamid	F	21
4	Safia Khedher	F	23
5	Lukman Mosa	M	19
6	Ahmad Khalil	M	24
7			
8			
9			
10			

No.	Name	Sex	Age
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

Simple random sampling (SRS)(Cont.)

Advantages:

- SRS is practicable for small, geographically compact populations

- SRS often used for final selection of study units in multi-stage sampling

2. Systematic sampling

May be more convenient than SRS

1. Units must be in some sort of sequence

2. Divide the study population size (N) by the required sample size (n) to determine interval 'k' (round k to nearest whole number) **N/n=k**

   e.g. population size (1000)/sample size (50)=20 (k)

3. Choose a random number between 1 and k (20) and select that unit

   e.g. 17 (between 1 and 20)

2. Systematic sampling(Cont.)

4. Select every k^{th} (e.g. 20^{th}) subsequent unit

   e.g. 17,37,57,77,97,117,137,157,177, etc till 1000) = 50

2. Systematic sampling(Cont.) Another example

Faculty of Nursing (Population or students 'N'= 40)

If we need 10 students (n=10)

K=N/n

K=40/10 =4

Choose a random number between 1 and 4 (e.g. 1)

Select 1

Select every k^{th} (4^{th})

1, 5,9,13,17,21,25,29,33,37.

3. Cluster sampling (Two-stage or multi-stage)

- It is not always feasible to draw a SRS or a systematic sample
  - A sampling frame may not exist (Erbil population)
  - The effort to create one may be too great

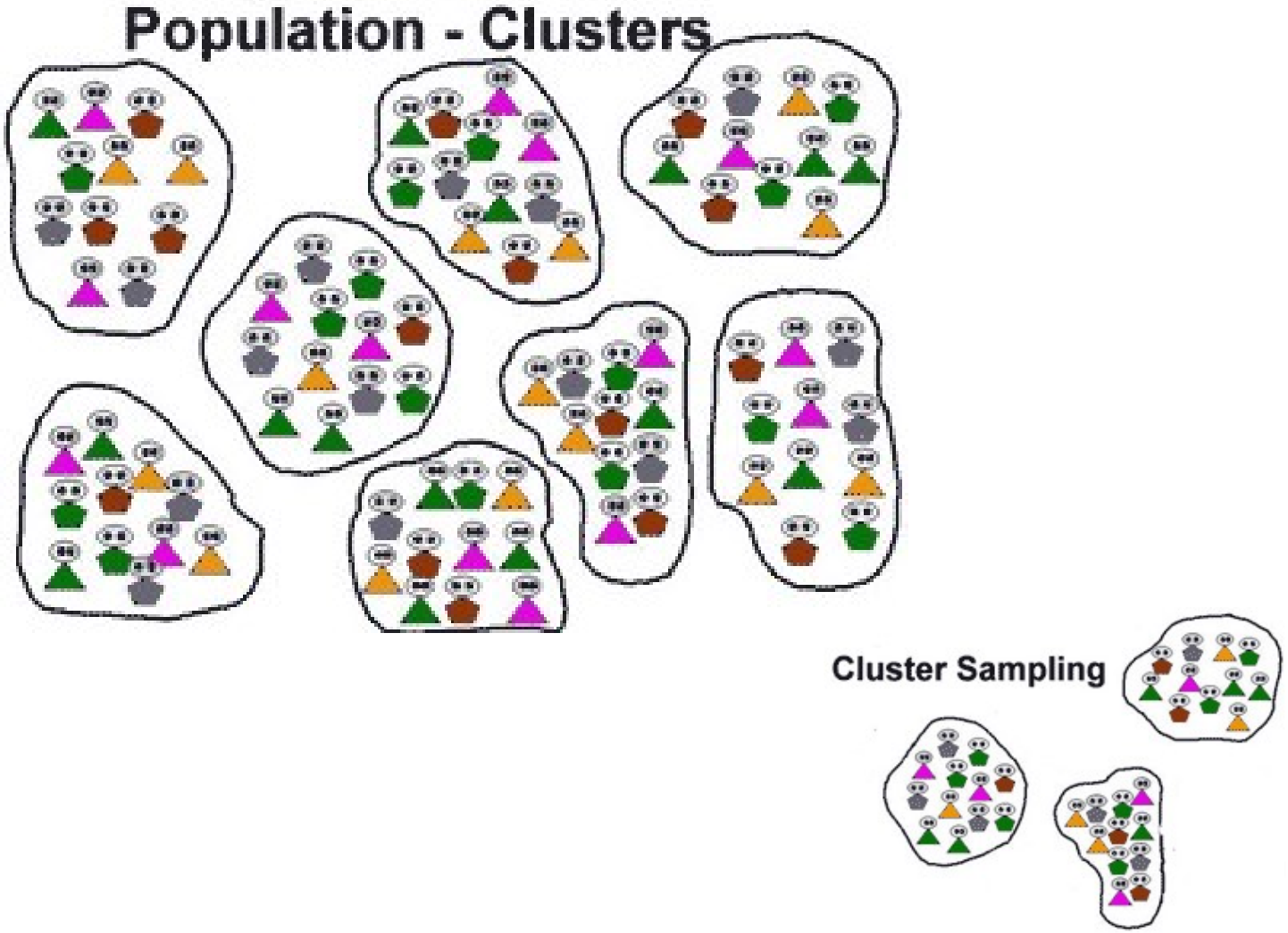
- It results from 2 stage process. The population is divided into clusters, and a subset of the clusters is randomly selected.

3. Cluster sampling(Cont.)

- Clusters are commonly based on

geographic areas or districts.

- Most populations exist in groups or ‘clusters’ e.g. villages, quarter, compounds, health clinics
- It may be more practical to take a sample of clusters first, and then sample units within clusters.


Population - Clusters
Cluster Sampling

3. Cluster sampling (Cont.)

1. Draw up a list of all first-stage units

   (called the 'primary sampling units' or PSUs) e.g. villages

2. Select a random sample of PSUs (villages)

3. For each PSU selected, draw up a sampling frame of second-stage units (e.g. households in each selected village)

4. Select a random sample of the second stage units (households)

- Reduced cost, if collecting data in-person
- Less precise estimates
- Can extend to 3 or more stages (Multi-stage)

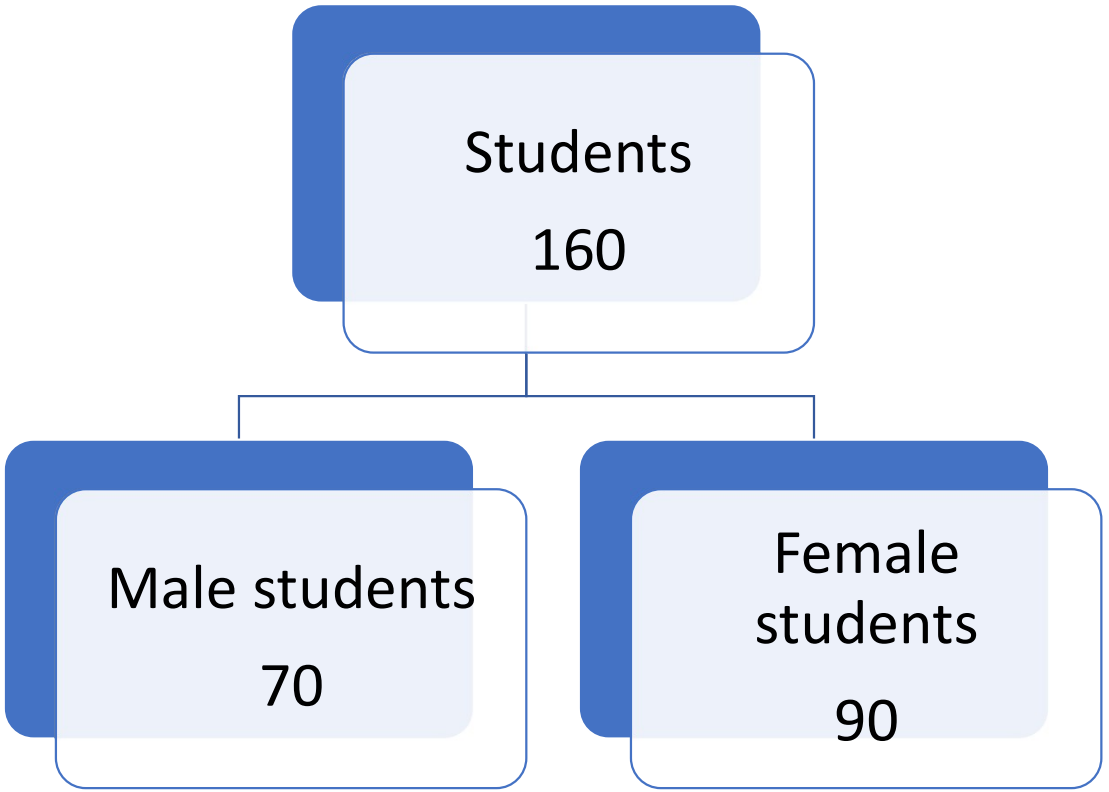
4. Stratified random sampling

- We may want independent results for different sub-populations or ‘strata’ (e.g. male and female, students of medicine and nursing)

- Simple random samples are taken from each ‘stratum’

Advantages

- Can ensure samples are large enough to achieve desired precision for each sub-population
- Can use different sample designs in different strata
- May be easier for administrative reasons



Select 20 by SRS or systematic sampling

Select 20 by SRS or systematic sampling

Non-probability sampling

- Not all the individuals in the population have the same chance to be included in the sample.

- Needs less time, money, and personnel.

- Subjects are usually selected on the basis of their accessibility or by the purposive personal judgment of the researcher

Non-probability sampling(Cont.)

1. Convenience sampling.

2. Judgmental sampling.

3. Quota sampling.

4.Snowball sampling

1.Convenience sampling

- Most common of all sampling techniques.
- Select the most easily obtained participants.
- Subjects are selected because they are

accessible and easy to recruit.

- Easiest, cheapest and least time consuming sampling technique. e.g. a busy physician who wants to make a study on 50 patients attending the outpatient Clinic.

2.Judgmental sampling

- Select participants based on personal belief about which are representative (e.g. one quarter in Erbil city).
- Subjects are chosen to be part of the sample with a specific purpose in mind.
- The researcher believes that some subjects are more fit for the research compared to other individuals.

3. Quota sampling

- Equal or proportionate representation of subjects depending on which trait is considered as basis of the quota.

- The bases of the quota are usually age, gender, education, race, religion and socioeconomic status.

4.Snowball sampling

- Usually done when there is a very small population size.

- The researcher asks the initial subject to identify another potential subject who also meets the criteria of the research.

Errors in samples (Cont.)

- When we take a sample, it is unlikely that our results will be exactly equal to the results for the whole population. Our results will be subject to errors.

- We can reduce the errors to acceptable level by increasing the sample size and careful design of the sampling procedure

Sources of error in samples

- Sample size (n)
- Sampling design
- Non-response
- Measurement
  - observer (or interviewer)
  - participant (or respondent)
  - instrument (e.g. questionnaire)
  - mode of administration

Summary
Population and sample

The concepts of:

- Study population
- Reasons for sampling
- Sampling frame
- Sample

References

- Essential Medical Statistics, by Betty Kirkwood & Jonathan Sterne
  (Published by Blackwell)

  Statistics Without Tears, a Primer for Non-mathematicians, by Derek Rowntree
  (Published by Penguin)

THANK YOU