

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



السلام عليكم ورحمة الله وبركاته

# Biostatistics

L VII

**PROF. DR. WAQAR AL-KUBAISY**

**6 -8-2024**

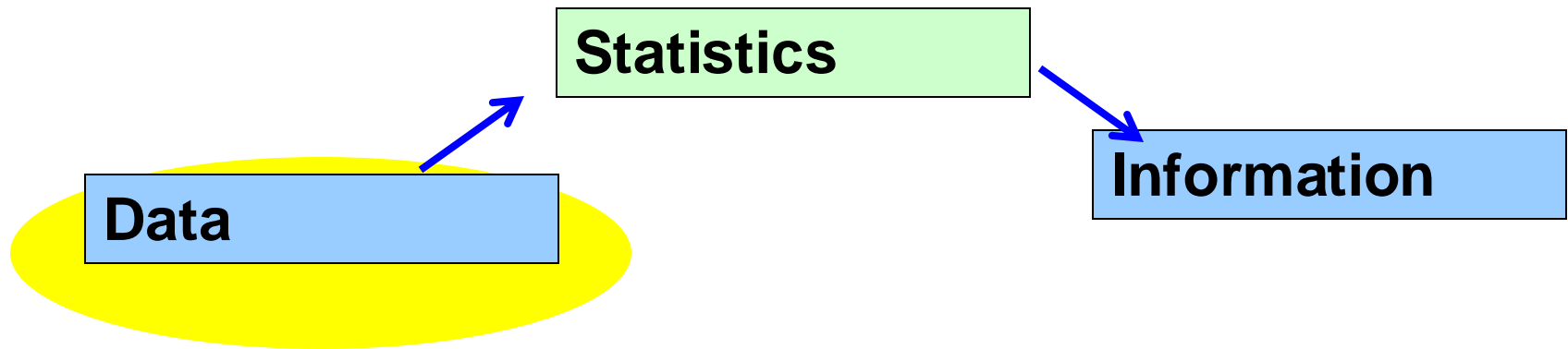
# **SAMPLING METHODS**

## **Part 1**

## **LEARNING OBJECTIVES**

- **Learn the reasons for sampling**
- **Distinguish between probability & non probability sampling**
- **Develop an understanding about different sampling methods**

**Statistics** is a tool for converting **data** into **information**:



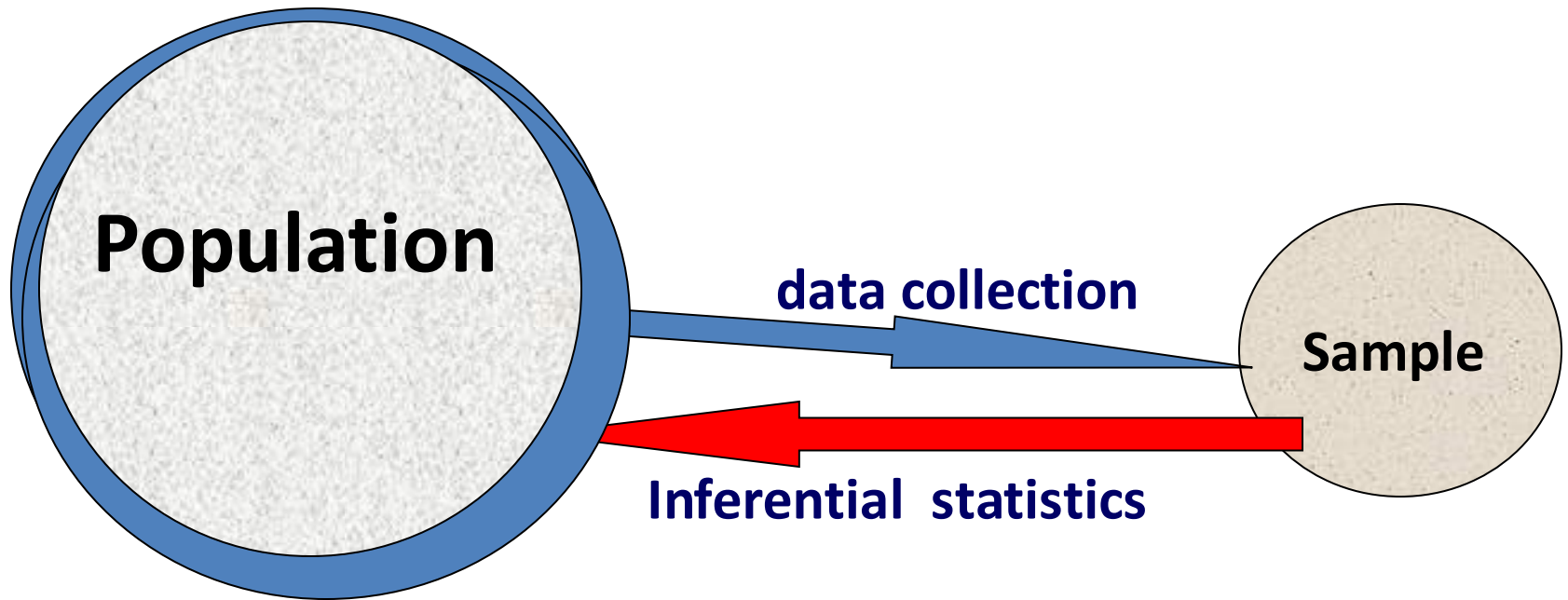
where does **data** come from?

How is it **gathered**?

How do we ensure its **accurate**?

Is the data **reliable**?

Is it **Representative** of the **population** from which **it was drawn**?



a **sound** generalized **information** about the **population** from which the **sample** has been **drown**, depending on evidence of this sample

# Population & Sampling

## Unobserved & Observed

Generally any set of observed data is a **part of** a large aggregate of potential, but **unobserved data**,

the observed data called **sample**  
whereas

the unobserved large group is called a **population.**

## Population or " universe "

- ❑ A population can be defined as **all** people or items with the **characteristic, one wishes** to understand/study .
  
- ❑ Is a group of **entities** , having some **quantifiable characteristic in common**, for which we have an interest at a particular time ..
  
- ❖ The group (unit) may be **people, machines, animal, bacteria ....Or it is not limited to population**
  
- ❖ They may be **finite** or **infinite** in No.
  
- ❖ *The quantifiable variable* or characteristic may **be** **continuous** or a **discrete** variable ..



❑ **Population also** described **as a set of data** consist of all **hypothetically possible observation of a given phenomenon .**

❑ **Therefore population is; a full set of individuals to whom we limit any discussion or inference .**

❑ **When might you sample the entire population?**

- ❖ •When your population is **very small**
- ❖ •When you have **extensive**( large) **resources**
- ❖ •When you **don't** expect a very high **response**
- ❖ •Data gathered from entire population : **Census**

## Cont. ...SAMPLES

*The first questions that the worker must ask himself are :*

- What data do I need ??
  - Can I investigate the problem by mean of sample ?
  - If so what is the **sample size** should be representative
  - How could we chose the sample ?
- ❑ •Data gathered in experiments and observational studies come from samples.



# SAMPLES

- ❖ • A sample is a **subset** of population
  - It is a subset of population, that had been chosen from population under study, in a way that it **should be**
  - **representative to whole population.**
- ❖ • A sample is “a smaller (**but hopefully representative**) collection of units from a population, used to determine
- ❖ **truths about that population”**
- ❖ Sample is of interest not in its own right, but for **what it tells the investigator about the population .**
- ❖ **Therefore care** must be taken to ensure that the sample is **truly represents** the **population** about which information is required



- ❖ Its some finite No. of the unit from population of individual .
- ❖ It is part of population .

- ❖ **The main objective** of most statistical or studies
- ❖ is to make **sound generalization** of information on the basis of sample about the **population** from which the sample comes .

□ This one is achieved through

- ❖ **choosing** the **sample** from the population under study in **a way that** it should be **representative** to whole population

□ **Making inference** from a **sample** to a **population** is called as ***statistical inference*** .

## Cont. ...SAMPLES

❖ **Sample is a set of data that consist of only a part of these observation (population) .**

### • **Why sample?**

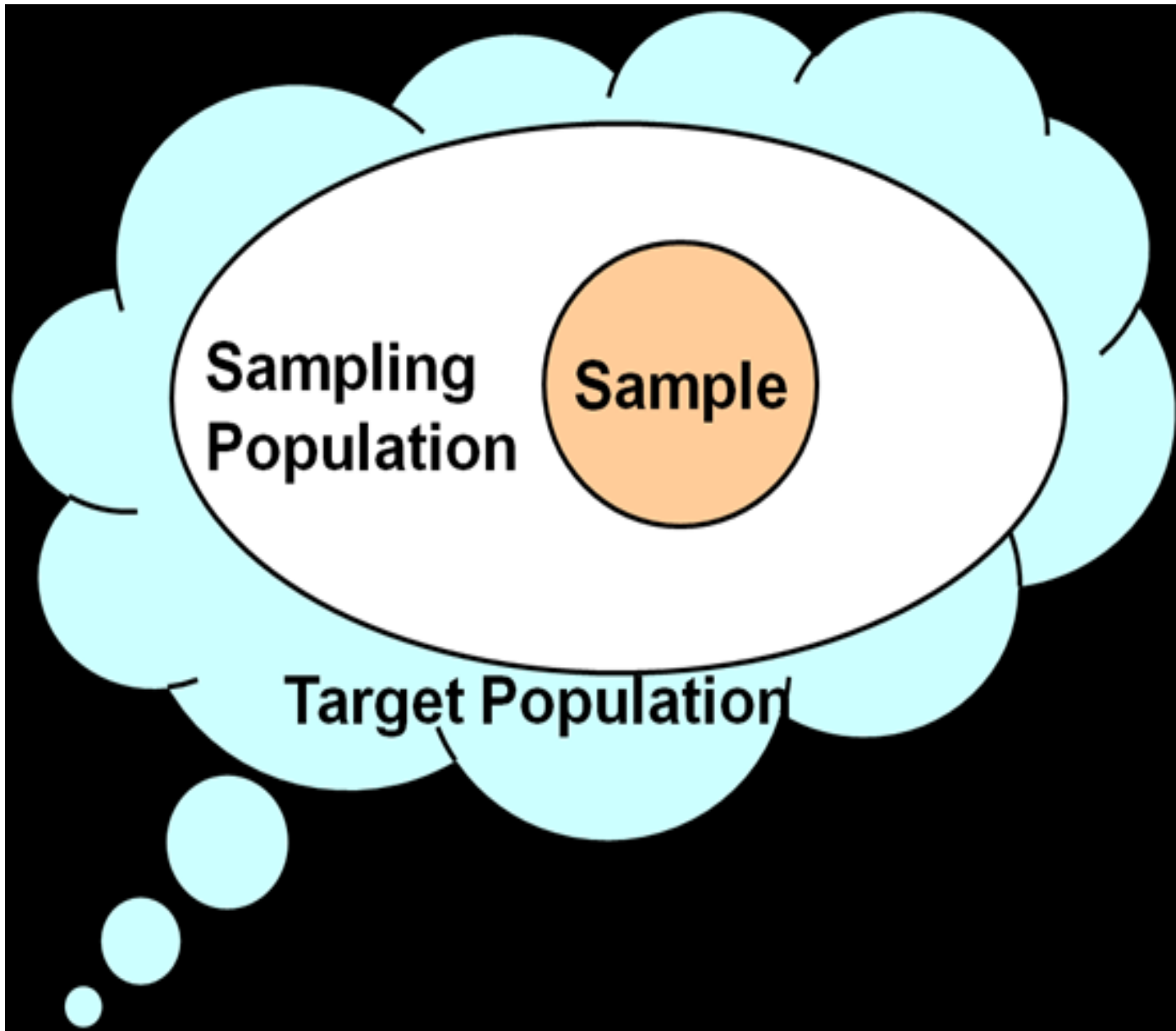
**Resources (time, money) and workload**

**Gives results with known accuracy that can be calculated mathematically**

• **The sampling frame** is the list from which the potential respondents are drawn

## **SAMPLING FRAME**

- **sampling frame which has the property that we can identify every single element and include any in our sample**
- **The sampling frame must be representative of the population**



# Why Do We Do Sampling

	Population	Sample
<b>Size</b>	impossible & impractical	Possible & practical
<b>Cost</b>	High	Less
<b>Observed</b>	Not all be observed	All be observed
<b>Staff</b>	Large size	Smaller
<b>Time</b>	More	Less
<b>Effort</b>	More	Less
<b>Accurate</b>	Less	More
		<ul style="list-style-type: none"> <li>• Work Lighter</li> </ul>
		<ul style="list-style-type: none"> <li>• Uniform way</li> </ul>
		<ul style="list-style-type: none"> <li>• Highly skilled</li> </ul>
		More precision (well trained)



## SAMPLING

❑ **A sample should be representative of the population**

❑ **3 factors that influence sample representativeness**

- Sampling procedure
- Sample size
- Participation (response)

### Sample Size N

It is the No. of individuals that collected in the sample, denoted by **N**.

OpenEpi for sample size calculation

# SAMPLING

- **Non-probability:**

–one in which, the **judgment** of the experimenter, the **methods** in which the data are collected, or other **factors** could affect the results of the sample **BIAS**

- **Probability:**

The **chance** of selection of each item of the population **is known** before the sample is picked **NO BIAS**

# TYPES OF SAMPLING

## ☐ Probability (Random) Samples

1. Simple random sample
2. Systematic random sample
3. Stratified random sample
4. Cluster sample
5. Multistage sample
6. Multiphase sample

## ☐ Non-Probability Samples

- 1) Convenience sample
- 2) Purposive sample
- 3) Quota

# Random Sample

- ❖ This is the most popular one,
- ❖ it is most commonly used in survey and research,
- ❖ it is a sample drawn from a population or unit in such a way that;
- ❖ every member of the population
- ❖ has the same probability of selection.
- ❖ Every member in the population having independent
- ❖ and equal chance of appearing in the sample.

*Therefore, sample now*

- **FREE OF BIAS** and it is representative to the whole population
- Random Sample is the simplest and the best known way to avoid bias and to be representative to the population

## PROBABILITY SAMPLING Random Sample

- **Probability sampling includes:**

1. Simple Random Sampling,
2. Systematic Random Sampling,
3. Stratified Random Sampling,
4. Cluster Sampling
5. Multistage Sampling.
6. Multiphase sampling

## Simple R.S.

- ❖ *By using Random Digit*
- ✓ **Identify** the **population size**, and
- ✓ **give No.** for each one of population.
- ✓ **Identify** the **sample size**
- ✓ **Chose first No.** blindly from the random digit .
- ✓ **Decide going vertically or horizontally .**
- ✓ **Chose second, third, fourth..... No.**
- ✓ **Collect the sample size .**

### Ignore :

Repeated No.

No. larger than population size .

11164	36318	75061	37674	26320	75100	10431	20418	19228	91792
21215	91791	76831	58678	87054	31687	93205	43685	19732	08468
10438	44482	66558	37649	08882	90870	12462	41810	01806	02977
36792	26236	33266	66583	60881	97395	20461	36742	02852	50564
73944	04773	12032	51414	82384	38370	00249	80709	72605	67497
49563	12872	14063	93104	78483	72717	68714	18048	25005	04151
64208	48237	41701	73117	33242	42314	83049	21933	92813	04763
51486	72875	38605	29341	80749	80151	33835	52602	79147	08868
99756	26360	64516	17971	48478	09610	04638	17141	09227	10606
71325	55217	13015	72907	00431	45117	33827	92873	02953	85474
65285	97198	12138	53010	94601	15838	16805	61004	43516	17020
17264	57327	38224	29301	31381	38109	34976	65692	98566	29550
95639	99754	31199	92558	68368	04985	51092	37780	40261	14479
61555	76404	86210	11808	12841	45147	97438	60022	12645	62000
78137	98768	04689	87130	79225	08153	84967	64539	79493	74917
62490	99215	84987	28759	19177	14733	24550	28067	68894	38490
24216	63444	21283	07044	92729	37284	13211	37485	10415	36457
16975	95428	33226	55903	31605	43817	22250	03918	46999	98501
59138	39542	71168	57609	91510	77904	74244	50940	31553	62562
29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
29621	66583	62966	12468	20245	14015	04014	35713	03980	03024
12639	75291	71020	17265	41598	64074	64629	63293	53307	48766
14544	37134	54714	02401	63228	26831	19386	15457	17999	18306
83403	88827	09834	11333	68431	31706	26652	04711	34593	22561
67642	05204	30697	44806	96989	68403	85621	45556	35434	09532
64041	99011	14610	40273	09482	62864	01573	82274	81446	32477
17048	94523	97444	59904	16936	39384	97551	09620	63932	03091
93039	89416	52795	10631	09728	68202	20963	02477	55494	39563
82244	34392	96607	17220	51984	10753	76272	50985	97593	34320
96990	55244	70693	25255	40029	23289	48819	07159	60172	81697
09119	74803	97303	88701	51380	73143	98251	78635	27556	20712
57666	41204	47589	78364	38266	94393	70713	53388	79865	92069
46492	61594	26729	58272	81754	14648	77210	12923	53712	87771
08433	19172	08320	20839	13715	10597	17234	39355	74816	03363
10011	75004	86054	41190	10061	19660	03500	68412	57812	57929
92420	65431	16530	05547	10683	88102	30176	84750	10115	69220
35542	55865	07304	47010	43233	57022	52161	82976	47981	46588
86595	26247	18552	29491	33712	32285	64844	69395	41387	87195
72115	34985	58036	99137	47482	06204	24138	24272	16196	04393
07428	58863	96023	88936	51343	70958	96768	74317	27176	29600
35379	27922	28906	55013	26937	48174	04197	36074	65315	12537
10982	22807	10920	26299	23593	64629	57801	10437	43965	15344
90127	33341	77806	12446	15444	49244	47277	11346	15884	28131
63002	12990	23510	68774	48983	20481	59815	67248	17076	78910
40779	86382	48454	65269	91239	45989	45389	54847	77919	41105
43216	12608	18167	84631	94058	82458	15139	76856	86019	47928
96167	64375	74108	93643	09204	98855	59051	56492	11933	64958
70975	62693	35684	72607	23026	37004	32989	24843	01128	74658

49563	12872	14063	93104	78483	72717	68714	18048	25005	04151
64208	48237	41701	73117	33242	42314	83049	21933	92813	04763
51486	72875	38605	29341	80749	80151	33835	52602	79147	08868
99756	26360	64516	17971	48478	09610	04638	17141	09227	10606
71325	55217	13015	72907	00431	45117	33827	92873	02953	85474
65285	97198	12138	53010	94601	15838	16805	61004	43516	17020
17264	57327	38224	29301	31381	38109	34976	65692	98566	29550
95639	99754	31199	92558	68368	04985	51092	37780	40261	14479
61555	76404	86210	11808	12841	45147	97438	60022	12645	62000
78137	98768	04689	87130	79225	08153	84967	64539	79493	74917
62490	99215	84987	28759	19177	14733	24550	28067	68894	38490
24216	63444	21283	07044	92729	37284	13211	37485	10415	36457
16975	95428	33226	55903	31605	43817	22250	03918	46999	98501
59138	39542	71168	57609	91510	77904	74244	50940	31553	62562
29478	59652	50414	31966	87912	87154	12944	49862	96566	48825
96155	95009	27429	72918	08457	78134	48407	26061	58754	05326
29621	66583	62966	12468	20245	14015	04014	35713	03980	03024
12639	75291	71020	17265	41598	64074	64629	63293	53307	48766
14544	37134	54714	02401	63228	26831	19386	15457	17999	18306
83403	88827	09834	11333	68431	31706	26652	04711	34593	22561

67642	05204	30697	44806	96989	68403	85621	45556	35434	09532
64041	99011	14610	40273	09482	62864	01573	82274	81446	32477
17048	94523	97444	59904	16936	39384	97551	09620	63932	03091
93039	89416	52795	10631	09728	68202	20963	02477	55494	39563
82244	34392	96607	17220	51984	10753	76272	50985	97593	34320
96990	55244	70693	25255	40029	23289	48819	07159	60172	81697
09119	74803	97303	88701	51380	73143	98251	78635	27556	20712
57666	41204	47589	78364	38266	94393	70713	53388	79865	92069
46492	61594	26729	58272	81754	14648	77210	12923	53712	87771
08433	19172	08320	20839	13715	10597	17234	39355	74816	03363
10011	75004	86054	41190	10061	19660	03500	68412	57812	57929
92420	65431	16530	05547	10683	88102	30176	84750	10115	69220
35542	55865	07304	47010	43233	57022	52161	82976	47981	46588
86595	26247	18552	29491	33712	32285	64844	69395	41387	87195
72115	34985	58036	99137	47482	06204	24138	24272	16196	04393
07428	58863	96023	88936	51343	70958	96768	74317	27176	29600
35379	27922	28906	55013	26937	48174	04197	36074	65315	12537
10982	22807	10920	26299	23593	64629	57801	10437	43965	15344
90127	33341	77806	12446	15444	49244	47277	11346	15884	28131
63002	12990	23510	68774	48983	20481	59815	67248	17076	78910
40779	86382	48454	65269	91239	45989	45389	54847	77919	41105
43216	12608	18167	84631	94058	82458	15139	76856	86019	47928
96167	64375	74108	93643	09204	98855	59051	56492	11933	64958
70975	62693	35684	72607	23026	37004	32989	24843	01128	74658
85812	61875	23570	75754	29090	40264	80399	47254	40135	69916



## Cont. ... SIMPLE RANDOM SAMPLING

- **Applicable** when population is small, homogeneous & readily available
- All subsets of the frame are given an equal probability. Each element of the frame thus has an equal probability of selection.
- It provides for greatest number of possible samples. This is done by assigning a number to each unit in the sampling frame.
- A table of random number or lottery system is used to determine which units are to be selected.

- **Advantages**

- Estimates are easy to calculate.
- Simple

- **Disadvantages**

- If sampling frame large, this method is impracticable.
- Need complete sampling frame.
- Minority subgroups of interest in population may not be present in sample in sufficient numbers for study.

# How to draw a simple random sample example

The following example describes the drawing of a sample of 20 individuals from a population of 80 using random number tables

Give every individual a number (80 individuals)

Use five-digit random numbers table (these may be found in most statistics textbooks) – the table opposite shows only the last 2 digits from the left of the 5 digit random numbers from a 5 digit random table, this is to avoid cluttering

Close your eyes and put your pen on one of the numbers in the random number table, for example 45. This will be your starting point

Start with the selected number and choose a direction (up, down, left or right)

Record the numbers that appear in the table, moving in the chosen direction until you have selected 20 numbers which lie between 1 and 80

Any numbers above 80, and numbers which have already been selected are ignored

Graphics by R. Vijayan

Random number table				
25	19	64	82	84
23	02	41	46	01
55	85	66	96	28
68	45	19	69	59
69	31	46	29	85
37	31	61	28	98
66	42	19	24	91
33	65	78	12	35
76	32	06	19	35
43	33	42	02	59
28	31	93	43	94
97	19	21	53	20

# SYSTEMATIC SAMPLING

- Sampling with system.
- By using **predefine** system :
  - ❖ Identify population size .
  - ❖ Identify sample size .
  - ❖ Identify **predefine** system we need 10th 8th every **kth** element
- In this case,  $k = (\text{population size} / \text{sample size})$ .
- ❖ Chose first No. By using **random digit** .
- It is **important** that the starting point is not automatically the first in the list, but is instead randomly chosen from
  - ❖ **within** the first to the **kth** element in the list.
  - ❖ Use **predefine** system to collect 2nd 3rd .... K No.
  - ❖ then selecting elements at **regular intervals** through that ordered list.
  - ❖ Collect the sample size

# SYSTEMATIC SAMPLING

•All elements have the same probability of selection (in the example given, one in ten). It is not 'simple random sampling' because different subsets of the same size have different selection probabilities - e.g. the set  $\{4,14,24,\dots,994\}$  has a one-in-ten probability of selection, but the set  $\{4,13,24,34,\dots\}$  has zero probability of selection



# 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

# SYSTEMATIC SAMPLING

- **ADVANTAGES:**

- Sample easy to select
- Suitable sampling frame can be identified easily
- Sample evenly spread over entire reference population

- **DISADVANTAGES:**

- Sample may be biased if hidden periodicity in population coincides with that of selection.
- Difficult to assess precision of estimate from one survey.

*Thank You*



# STRATIFIED SAMPLING

## By using well define stratum

- Where population enclose, a number of distinct categories, the frame can be organized into separate "strata."
- " Each stratum is then sampled as an independent sub-population, out of which individual elements can be randomly selected.
- Every unit in a stratum has same chance of being selected.
- Using same sampling fraction for all strata ensures proportionate representation in the sample.
- Adequate representation of minority subgroups of interest can be ensured by stratification & varying sampling fraction between strata as required.
- Finally, since each stratum is treated as an independent population, different sampling approaches can be applied to different strata.

