# Introduction to Sampling Theory

## Lecture 5
## Simple Random Sampling


**Shalabh**

**Department of Mathematics and Statistics**

**Indian Institute of Technology Kanpur**

**Slides can be downloaded from**

**http://home.iitk.ac.in/~shalab/sp**

# Probability of Selection of a Unit

**Let the size of the population is *N*.**

**One out of *N* sampling unit is to be chosen.**

## SRSWOR

**The probability of drawing a sampling unit =** $\frac{1}{N}$

## SRSWR

**The probability of drawing a sampling unit =** $\frac{1}{N}$

# Probability of Selection of a Unit by SRSWOR or SRSWR



**Probability of drawing ball 1= 1/10**


**Probability of drawing ball 2= 1/10**

**…**


**Probability of drawing ball 10= 1/10**

# Proof: Probability of Selection of a Unit: SRSWOR

**Let $A_i$ : Event that a particular $j^{th}$ unit is not selected at the $i^{th}$ draw.**

**The probability of selecting, say, $j^{th}$ unit at $k^{th}$ draw is**

$$P(\text{selection of } u_j \text{ at } k^{th} \text{ draw})$$

$$= P(A_1 \cap A_2 \cap \ldots \cap A_{k-1} \cap \overline{A}_k)$$

$$= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)\ldots P(A_{k-1} \mid A_1, A_2 \ldots A_{k-2})P(\overline{A}_k \mid A_1A_2 \ldots A_{k-1})$$

$$= \left(1-\frac{1}{N}\right)\left(1-\frac{1}{N-1}\right)\left(1-\frac{1}{N-2}\right)\ldots\left(1-\frac{1}{N-k+2}\right)\frac{1}{N-k+1}$$

$$= \frac{N-1}{N} \cdot \frac{N-2}{N-1} \ldots \frac{N-k+1}{N-k+2} \cdot \frac{1}{N-k+1}$$

$$= \frac{1}{N}.$$

# Probability of Selection of a Sample

**Let the size of the population is *N*.**

**Let the size of the sample is *n*.**

**A sample of *n* sampling units out of *N* sampling units is to be chosen.**

# Probability of Selection of a Sample

## SRSWOR

**Total number of combinations to choose *n* sampling units out of *N***

**sampling unit =** $\binom{N}{n}$

**The probability of drawing a sample =** $\frac{1}{\binom{N}{n}}$

# Probability of Selection of a Sample

## SRSWOR

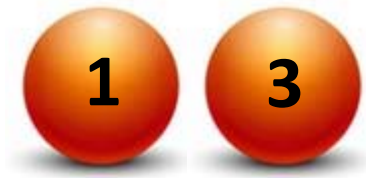
Suppose $N = 3$, $n = 2$


Total samples = $\binom{3}{2} = 3$

Sample 1


Sample 2


Sample 3




Probability of drawing a sample = $\frac{1}{3}$

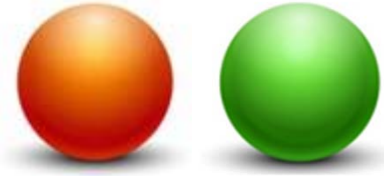
# Probability of Selection of a Sample

## SRSWOR

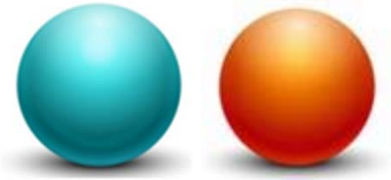
**Suppose *N* =3, *n* = 2**

**Total samples =** $\binom{3}{2} = 3$

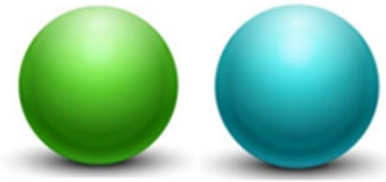
**Sample 1**



**Sample 2**



**Sample 3**



**Probability of drawing a sample =** $\frac{1}{3}$

# Proof: Probability of Selection of a Sample: SRSWOR

**A unit can be selected at any one of the *n* draws.**

**Let $u_i$ be the $i^{th}$ unit selected in the sample.**

**This unit can be selected in the sample either at first draw, second draw, …, or $n^{th}$ draw.**

**Let $P_j(i)$ denotes the probability of selection of $u_i$ at the $j^{th}$ draw, $j$ = 1,2,…,$n$. Then**

$$P_j(i) = P_1(i) + P_2(i) + \ldots + P_n(i)$$

$$= \frac{1}{N} + \frac{1}{N} + \ldots + \frac{1}{N} \quad (n \text{ times})$$

$$= \frac{n}{N}.$$

# Proof: Probability of Selection of a Sample: SRSWOR

Let $u_1, u_2,\ldots,u_n$ are the $n$ unit selected in the sample.

The probability of their selection is

$$P(u_1, u_2,\ldots, u_n) = P(u_1).\ P(u_2).\ .\ .P(u_n)$$

When the first unit is to be selected, then there are $n$ units left to be selected in the sample from the population of $N$ units.

So $P(u_1)= \frac{n}{N}$

# Proof: Probability of Selection of a Sample: SRSWOR

**When the second unit is to be selected, then there are ($n$ – 1) units left to be selected in the sample from the population of ($N$ – 1) units.**

**So $P(u_2)=\frac{n-1}{N-1}$**

**When the third unit is to be selected, then there are ($n$ – 2) units left to be selected in the sample from the population of ($N$ – 2) units and so on.**

**So $P(u_3)=\frac{n-2}{N-2}$**

**And so on, $P(u_n)=\frac{1}{N-(n-1)}$**

# Proof: Probability of Selection of a Sample: SRSWOR

**Thus probability of their selection is**

$$P(u_1, u_2, \ldots, u_n) = P(u_1).P(u_2)\ldots P(u_n)$$

$$= \frac{n}{N} \cdot \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} \cdots \frac{1}{N-n+1}$$

$$= \frac{1}{\binom{N}{n}}.$$

# Probability of Selection of a Sample

## SRSWR

**Total number of combinations to choose *n* sampling units out of *N* sampling unit = $N^n$**

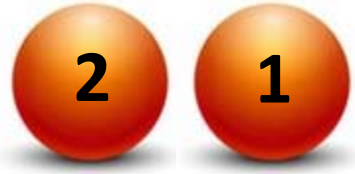
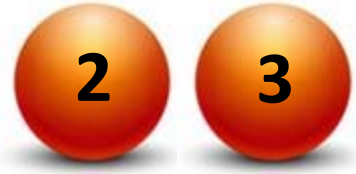
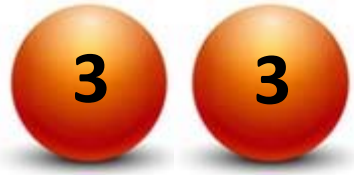
**The probability of drawing a sample = $\frac{1}{N^n}$**

# Probability of Selection of a Sample

## SRSWR

**Suppose $N$ =3,** 


**Total samples $N$=3, $n$=2, $N^n = 3^2 = 9$**

| | | | |
|---|---|---|---|
| **Sample 1**   | **Sample 4**   | **Sample 7**   |
| **Sample 2**   | **Sample 5**   | **Sample 8**   |
| **Sample 3**   | **Sample 6**   | **Sample 9**   |

**Probability of drawing a sample =** $\frac{1}{9}$

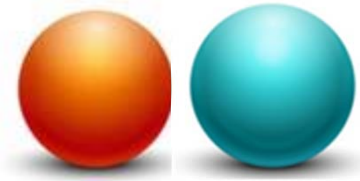
# Probability of Selection of a Sample

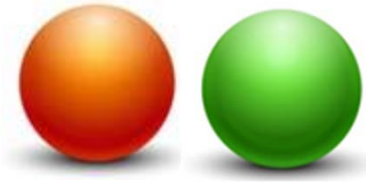
## SRSWR

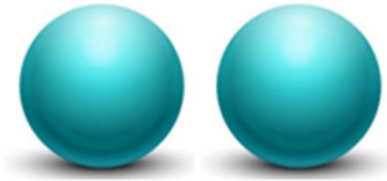
**Suppose $N$ =3,** 

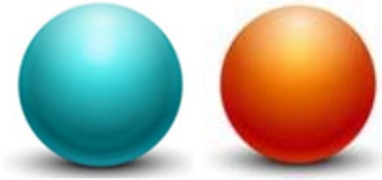
**Total samples $N$=3, $n$=2, $N^n = 3^2 = 9$**

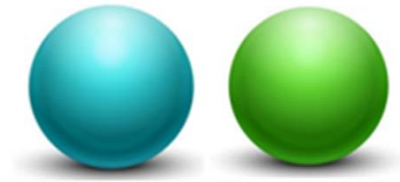
**Sample 1** 

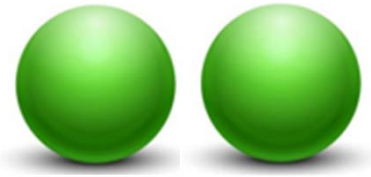
**Sample 2** 

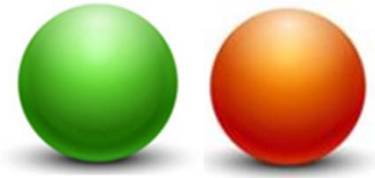
**Sample 3** 

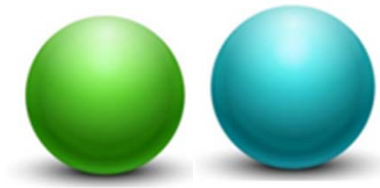
**Sample 4** 

**Sample 5** 

**Sample 6** 

**Sample 7** 

**Sample 8** 

**Sample 9** 

**Probability of drawing a sample = $\frac{1}{9}$**

# Proof: Probability of Selection of a Sample: SRSWR

**Let $u_i$ be the $i^{th}$ unit selected in the sample.**

**This unit can be selected in the sample either at 1st draw, 2nd draw, …, or $n^{th}$ draw.**

**At any stage, there are always $N$ units in the population in case of SRSWR, so the**

**probability of selection of $u_i$ at any stage = $1/N$ for all $i = 1,2,\ldots,n$.**

# Proof: Probability of Selection of a Sample: SRSWR

**Then the probability of selection of *n* units $u_1$, $u_2$,…,$u_n$ in the sample is**

$$P(u_1,u_2,\ldots,u_n) = P(u_1)\ .\ P(u_2)\ldots P(u_n)$$

$$= \frac{1}{N}\ .\ \frac{1}{N}\ldots\frac{1}{N}$$

$$= \frac{1}{N^n}.$$

# Notations:

**Following notations will be used:**

***N*** **: Number of sampling units in the population (Population size).**

***n*** **: Number of sampling units in the sample (Sample size)**

***Y*** **: The characteristic under consideration**

$Y_i$ **: Value of characteristic for the** $i^{th}$ **unit of the population**

**(*i* = 1,2,… 2,…,*N*)**

$y_i$ **: Value of the characteristic for the** $i^{th}$ **unit of the sample**

**(*i* = 1, 2,…,*n*)**

# Notations: Example

**$Y$: Height of students in a class**

**$N$ = 10 : Number of students in the class (Population size)**

**$n$ = 3 : Number of students in the sample (Sample size)**

**$Y_i$ : Height of $i^{th}$ student in the population**

# Example

**$Y$: Height of students in a class**

**$N$ = 10 : Number of students in the class (Population size)**

**$n$ = 3 : Number of students in the sample (Sample size)**

| Name of Student | $Y_i$ = Height of students (in Centimeters) |
|---|---|
| A | $Y_1$= 151 |
| B | $Y_2$= 152 |
| C | $Y_3$ = 153 |
| D | $Y_4$= 154 |
| E | $Y_5$ = 155 |
| F | $Y_6$= 156 |
| G | $Y_7$ = 157 |
| H | $Y_8$= 158 |
| I | $Y_9$ = 159 |
| J | $Y_{10}$= 160 |

# Notations: Example

**Suppose**

$Y_1$ = 151 cms., $Y_2$ = 152 cms., $Y_3$ = 153 cms., $Y_4$ = 154 cms., $Y_5$ = 155 cms.,

$Y_6$ = 156 cms., $Y_7$ = 157 cms., $Y_8$ = 158 cms., $Y_9$ = 159 cms., $Y_{10}$ = 160 cms.,

$y_i$ : Height of $i^{th}$ student in the sample

Selected sample = $3^{rd}$ , $7^{th}$ and $9^{th}$ student

$y_1 = Y_3$ = 153 cms., $y_2 = Y_7$ = 157 cms., $y_3 = Y_9$ = 159 cms.

## Drawing of sample

**Suppose we want to select the name of student or Height of the student.**

**The data in R will usually be given in a data frame, CSV file or any other format.**

**Suppose the data is stored in a data frame `heightdata` by using the following commands:**

```
height=c(151,152,153,154,155,156,157,158,159,160)
name=c("A","B","C","D","E","F","G","H","I","J")
heightdata=data.frame(name,height)
```

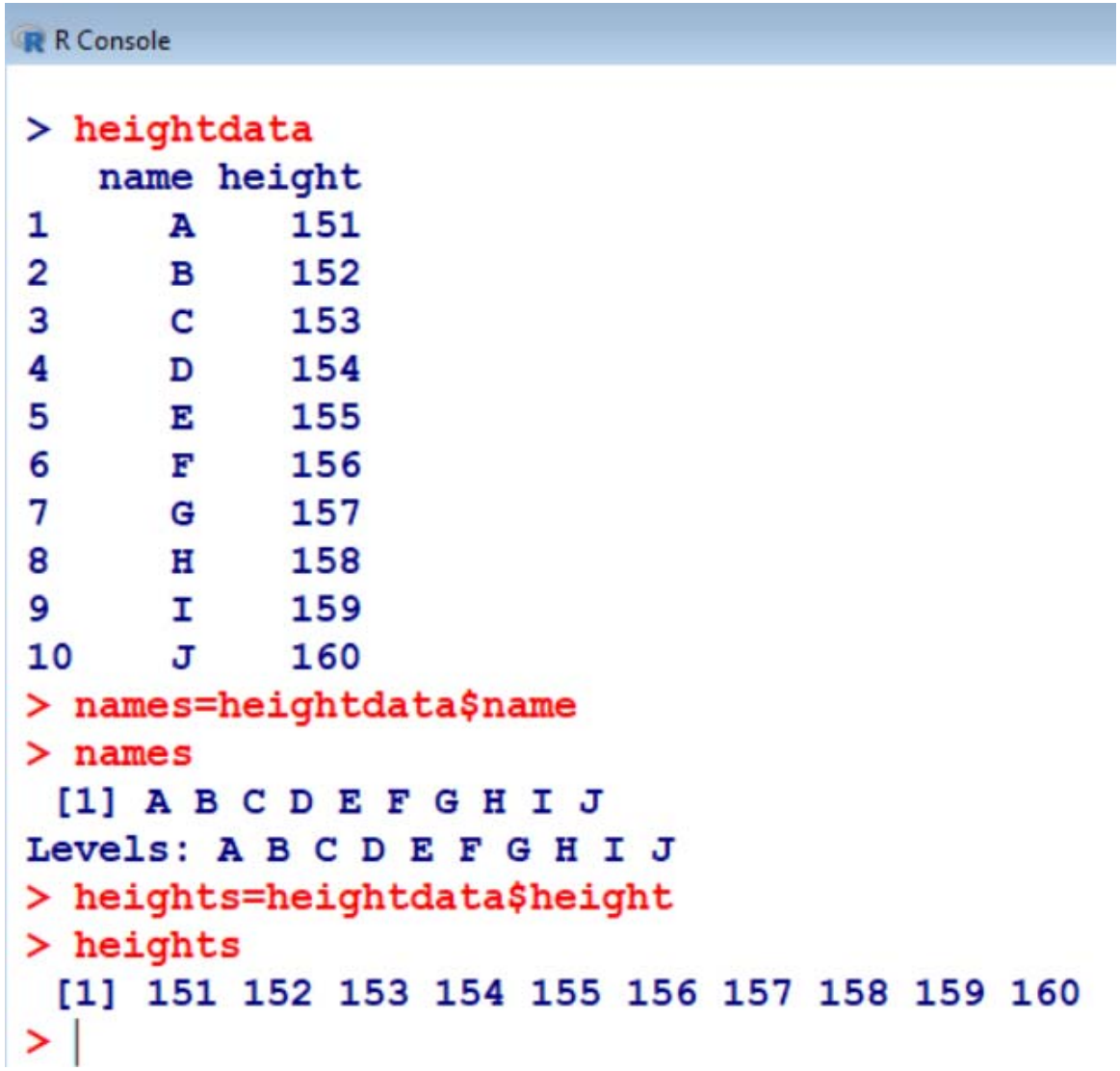
# Drawing of sample using R

```
> heightdata
   name height
1     A    151
2     B    152
3     C    153
4     D    154
5     E    155
6     F    156
7     G    157
8     H    158
9     I    159
10    J    160

> names=heightdata$name
> names
 [1] A B C D E F G H I J
Levels: A B C D E F G H I J

> heights=heightdata$height
> heights
 [1] 151 152 153 154 155 156 157 158 159 160
```

# Drawing of sample using R



# Drawing of sample using R : SRSWOR

**Suppose we want this sample in terms of names of persons.**

```
sample(names, size=5, replace = FALSE)
```

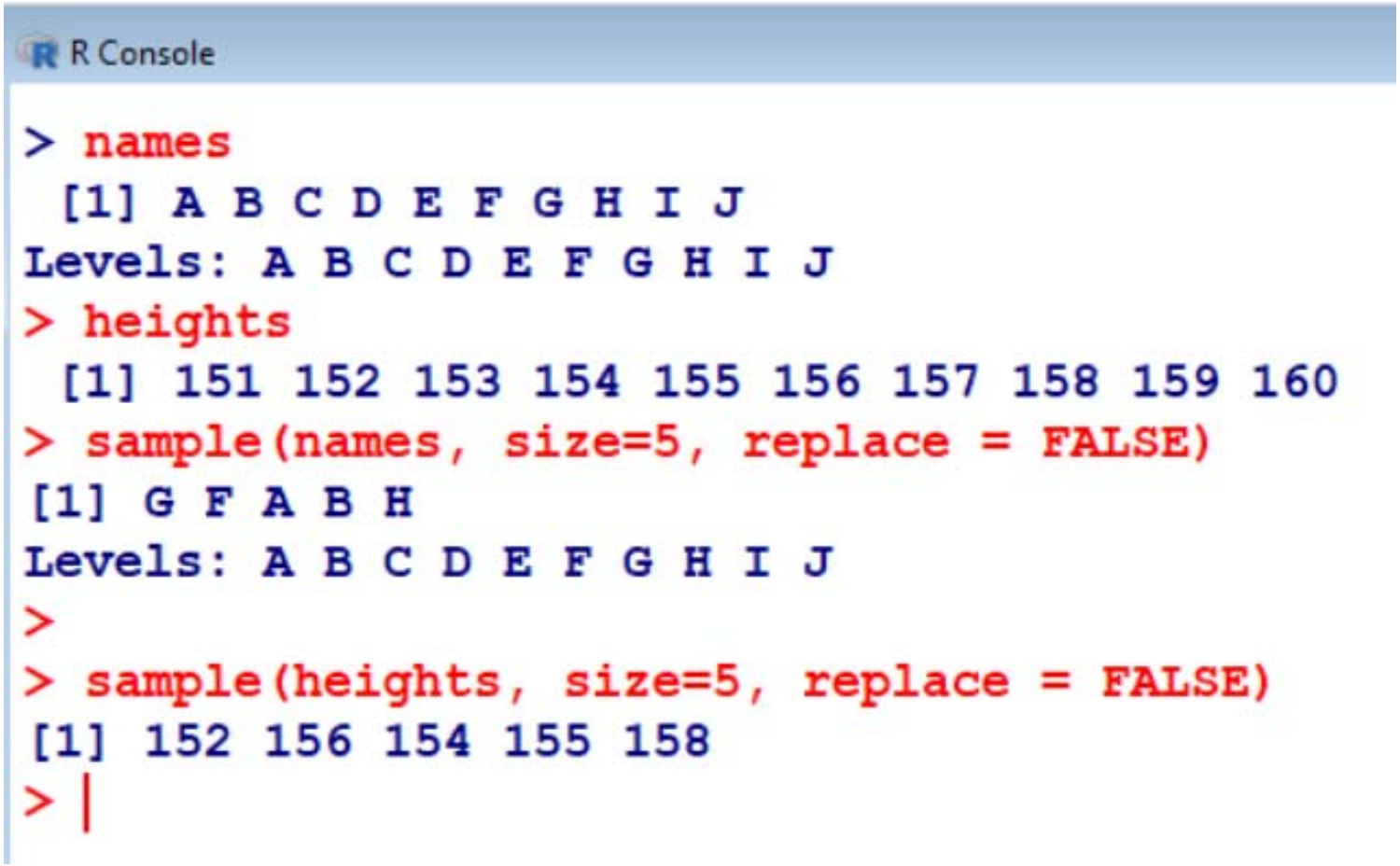
```
> sample(names, size=5, replace = FALSE)
[1] G F A B H
Levels: A B C D E F G H I J
```

**Suppose we want this sample in terms of heights of persons.**

```
sample(heights, size=5, replace = FALSE)
```

```
> sample(heights, size=5, replace = FALSE)
[1] 152 156 154 155 158
```

# Drawing of sample using R : SRSWOR



# Drawing of sample using R : SRSWR

**Suppose we want this sample in terms of names of persons.**

**Sample of size 5**

```
> sample(names, size=5, replace = TRUE)
[1] F F I E A
Levels: A B C D E F G H I J
```

**Sample of size 8**

```
> sample(names, size=8, replace = TRUE)
[1] C C D D J H G E
Levels: A B C D E F G H I J
```

# Estimation of population mean: Notations

$Y_1, Y_2, \ldots, Y_N$: **Population**

$y_1, y_2, \ldots, y_n$: **Sample**

$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ : **Population mean**

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ : **Sample mean**