Biostatistics

Section I

Prepared by M. Nozad H. Mahmood

What is Statistics ?

- *Statistics* **is a group of methods used to collect, organize, present, Analyze, and interpret data and to make decisions.**

Examples of Some Statistics

- **The number of educated people in Sulaimaniya City.**

- **The number of homeless in the United States of America.**

- **The number of cars in Iraq.**

- **The cost of health care services in Iraq**

What is Bio-Statistics ?

- *Bio-Statistics* is a group of methods used to collect, analyze, present, and interpret of biological, medical, and public health data to make decisions.

Examples of Some Biostatistics Data

- **The number of diagnostic tests in the hospital**

- **The number of hepatitis cases in the region**

- **The number of cancer deaths in Iraq**

- **The number of infectious diseases in Kurdistan Region of Iraq**

What are The Types of Statistics ?

Descriptive Statistics

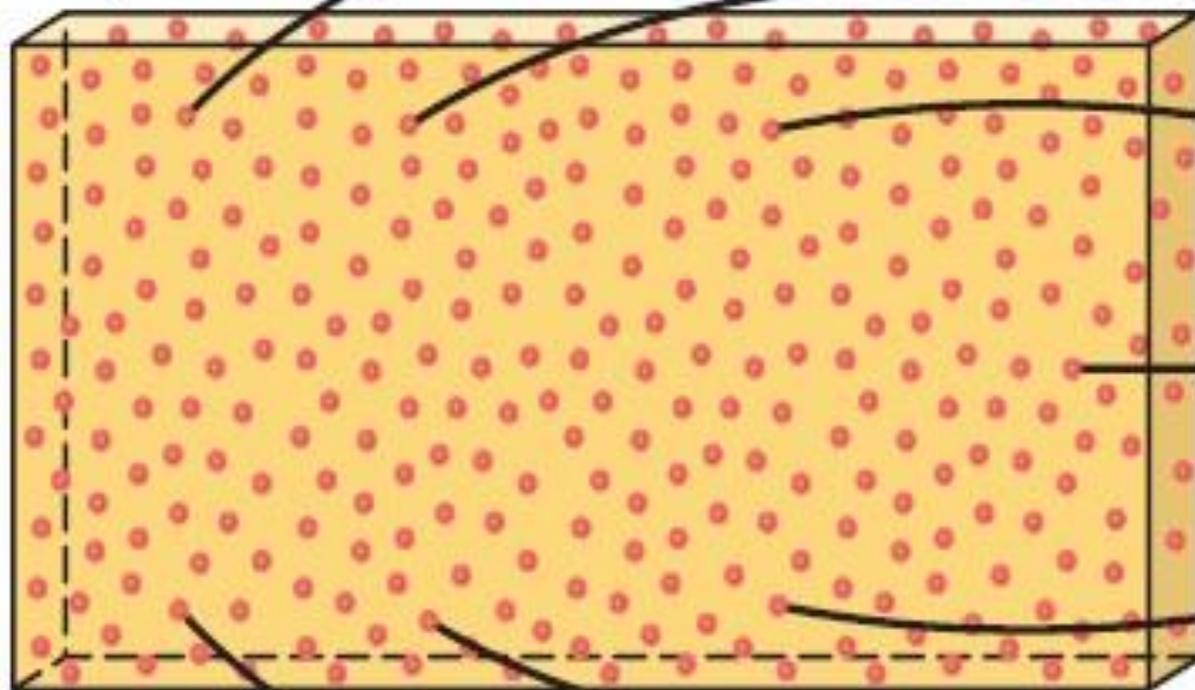*Descriptive statistics* consists of methods for organizing, displaying, and describing data by using statistical tables, graphs, and summary measures

Population / consists of all elements, individuals, items or objects whose characteristics are being studied
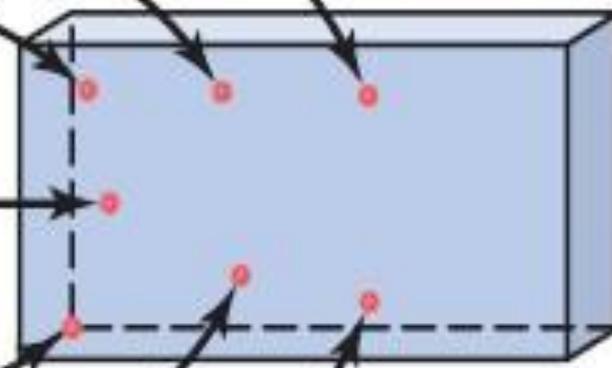
Sample / A portion of the population selected for study

Sample Survey / The technique of collecting information from a portion of population

What are The Types of Statistics ?

Inferential Statistics

Inferential statistics consists of methods that use sample results to help make decisions or predictions about a population.

Example of Inferential Statistics

- **The starting salary of a college graduate. We may select a sample of 200 recent college graduates.**

- **Look at to the starting salaries to make a decision based on this information.**

What are The Types of Data?

- **Quantitative data** / It can be measured numerically

Types of quantitative variables

1- Continuous data / Any numerical values over a certain interval or intervals (Should has a decimal)

E.g. (Length, age, height, weight, time, and blood cholesterol)

What are The Types of Data?

- **2- Discrete data** /   Only a certain numerical values with no intermediate values (Never has a decimal).

- **E.g. (Number of cars, houses, daily number covid-19 deaths)**

What are The Types of Data?

- **Qualitative data** / **It can not be measured numerically**

Types of qualitative variables

1- Nominal data / **It is another name for a categorical data**

E.g. (Colors, type of disease, type of cars, blood groups)

What are The Types of Data?

2- Ordinal data/ **They have ordered categories**

E.g. (Severity of disease, socioeconomic class, level of education)

Low – Middle - High

Mild – Moderate - Severe

What is The Meaning of Variable ?

- **It is a characteristic under study that assumes different values (scores of students in biostatistics) for different elements (students in department of medical laboratory analysis)**
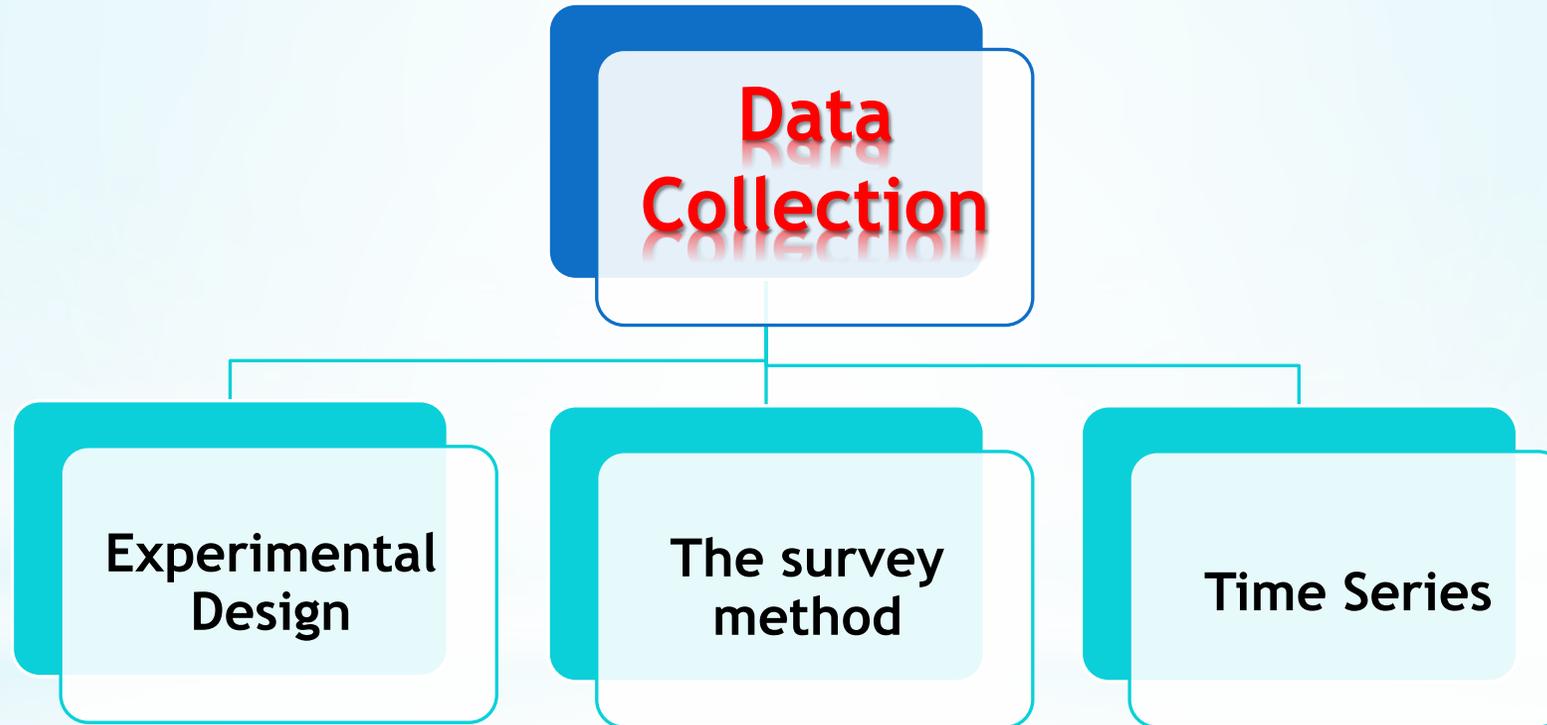
What are The Types of Variables ?

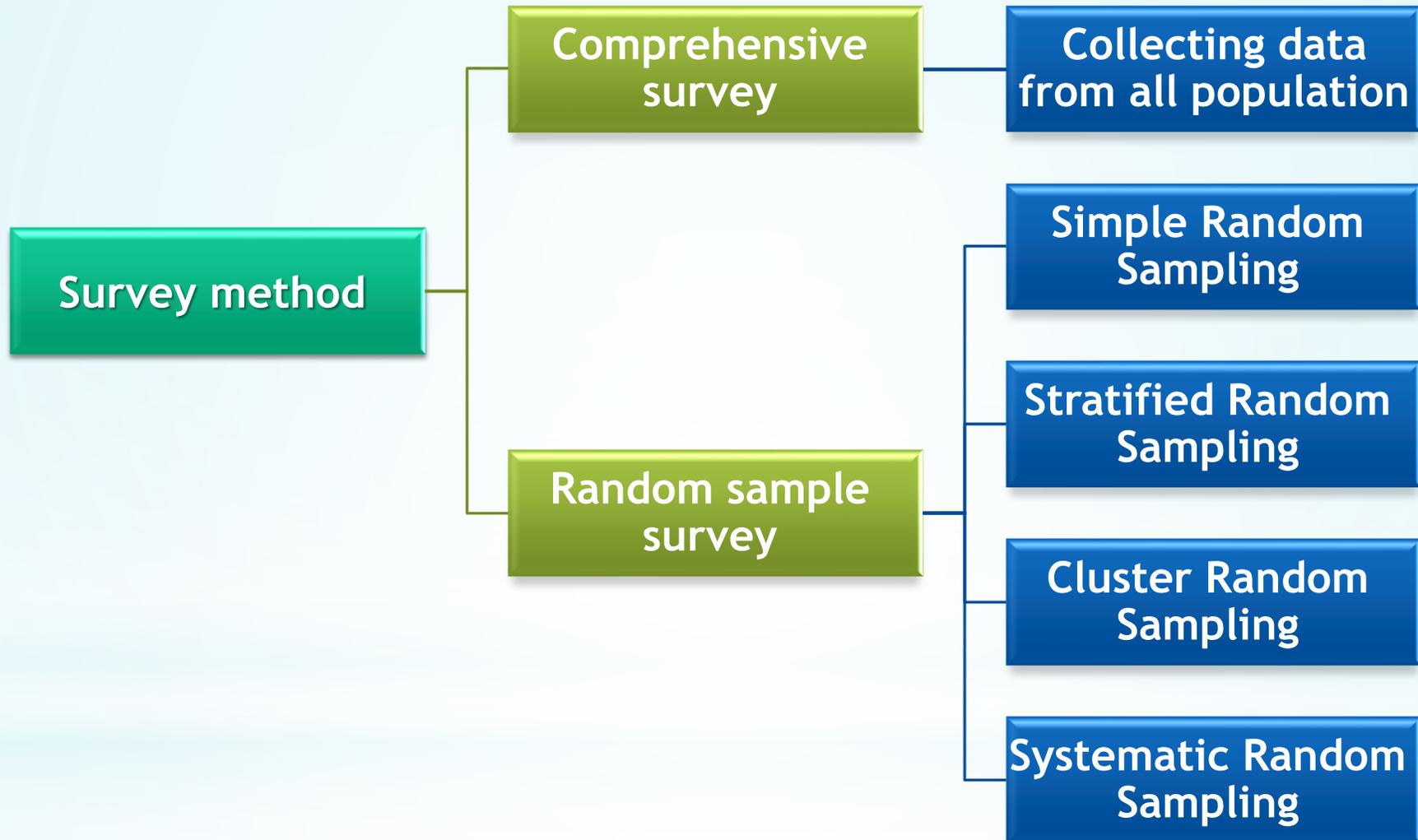**3- Binary variables** / **They have only two categories**

E.g. (Gender : Male and Female)

Question answer (True, False)

Question (Yes, No)
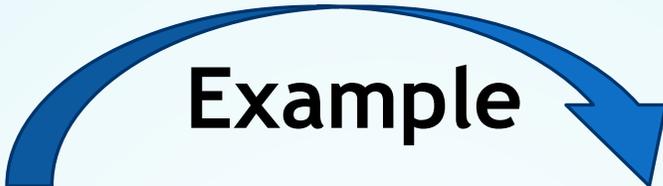
each member of the population has an **equal chance** to select

We give a single number for each member, then we randomly selected a group of numbers manually or one by one

Example

Suppose 3rd Stage of MLA department has 80 students, and we need to select a sample size of 20 students to ask them about the feedback of Biostatistics class.

Assign a number from 1 to 80 to each students, then randomly selected 20 numbers.

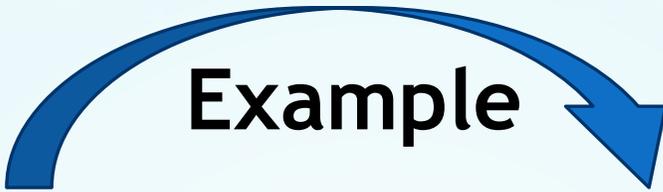You can either chose in one time or one by one.

We divide the entire population into different **smaller groups or strata**, then randomly selects a simple random sampling from the different strata.

Stratified Sample

formula for choosing sample size of each strata.

$$Sample\ size * \frac{Strata\ size}{Population\ size}$$

Example

To know the student academic performance in Statistics subject, we chose a stratified sample size of 80 students out of 800 students from Administration and Financial Sciences college

Divided the population(College of Administration and Financial Sciences) into departments, then we selected a simple random sampling from each department according to the number of students
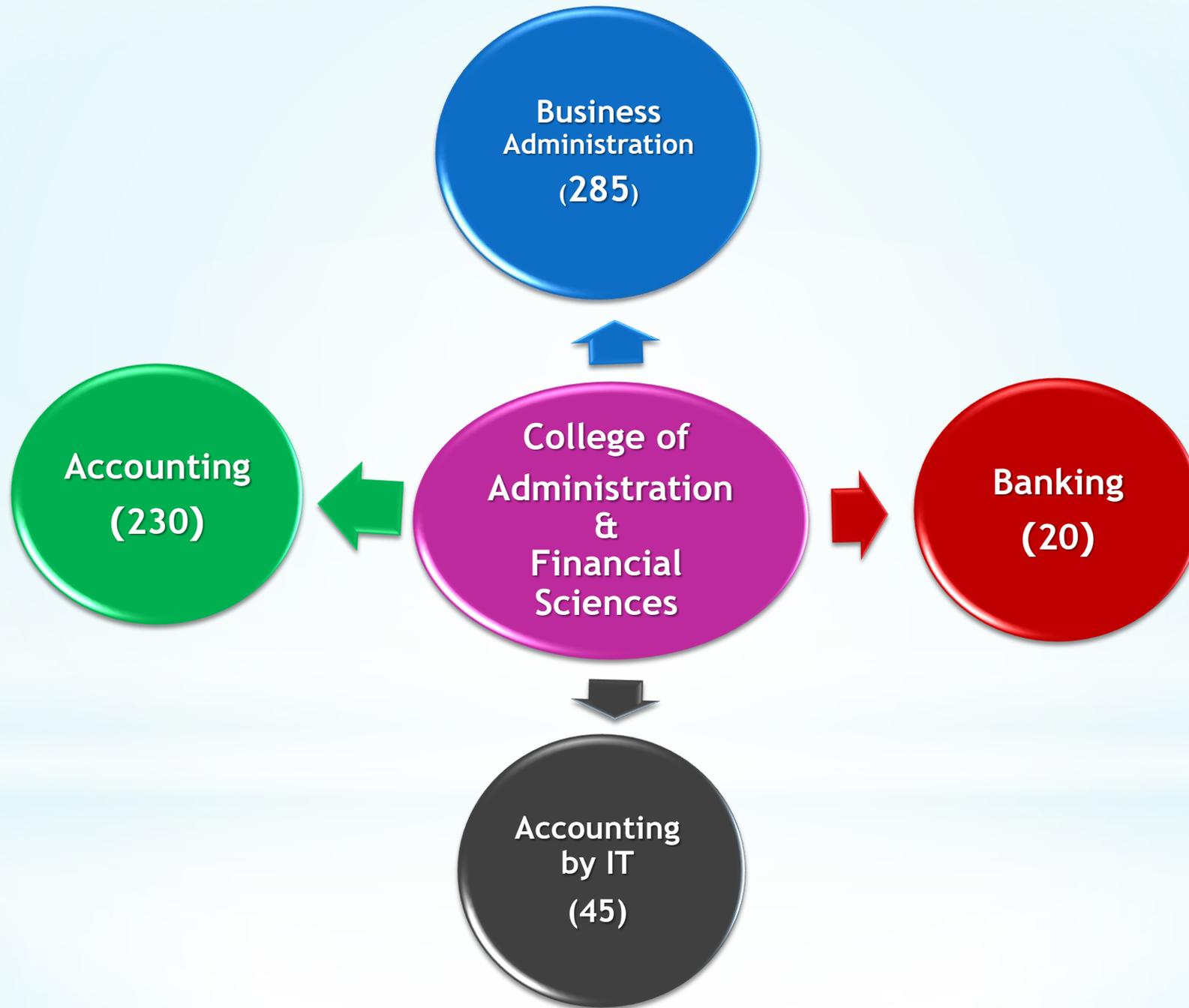
Population size $= 285 + 230 + 45 + 20 =$ 580 *Students*

The total sample size we needed is 80 students

$$Sample\ Size * \frac{Strata\ Size}{Population\ Size}$$

- Sample size selected from Business Administration Department:

$$80 * \frac{285}{580} = 39.31 \approx \mathbf{39}$$
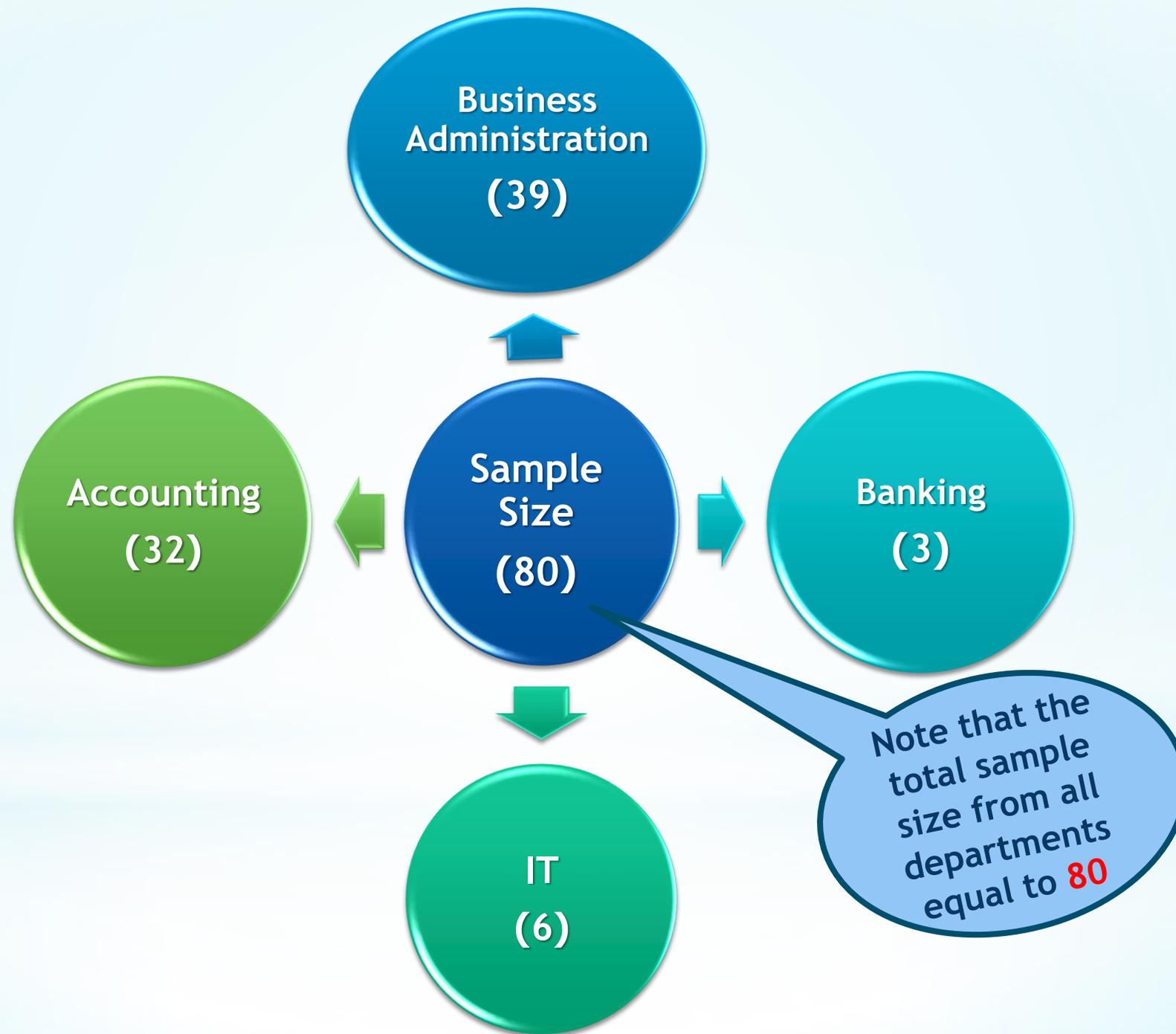
- Sample size selected from Accounting Department:

$$80 * \frac{230}{580} = 31.72 \approx \mathbf{32}$$

- Sample size selected from IT Department:

$$80 * \frac{45}{580} = 6.21 \approx \mathbf{6}$$

- Sample size selected from Banking Department:

$$80 * \frac{20}{580} = 2.76 \approx \mathbf{3}$$

Cluster sampling is accomplished by dividing the population into groups -- usually geographically. These groups are called clusters. The clusters are randomly selected, and each element in the selected clusters are used.

Cluster Sampling

This is the single-stage cluster sampling
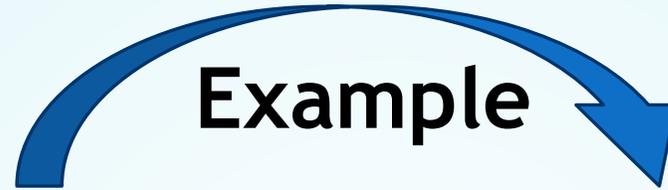
Example

If we want to know Performance schools in Sulaimany City

We divide Sulaimany in several areas, each area is a cluster

We want to randomly chose two areas, then we use all schools in selected area. For examples we selected Central and Southern areas
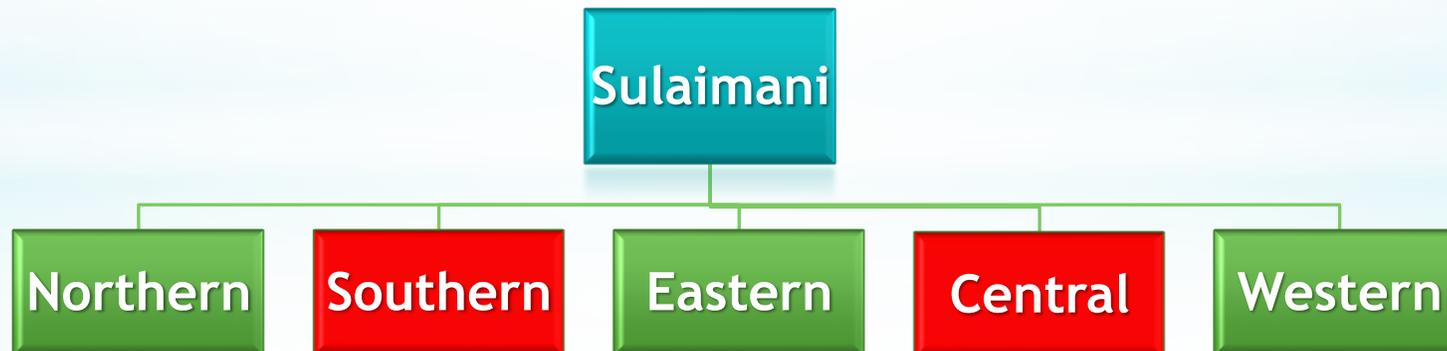
We divide the population in to groups, and the number of groups should be equal to the sample size. Then we select an element from first group.

Systematic Sampling

If the selected element from the first group, is the fourth element, then from all the remaining groups we should chose the fourth element.
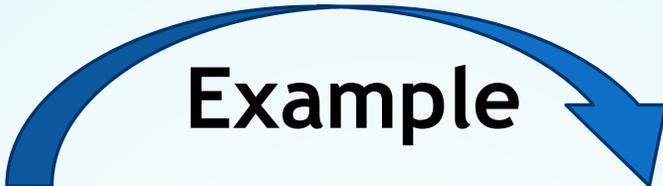
Example

A Furniture factory produces 100 pieces per day, we wana chose 10 pieces of them to test the quality of the products. How can we select the sample size by systematic sampling?

Divided the total products into 10 groups after given the number for each elements.
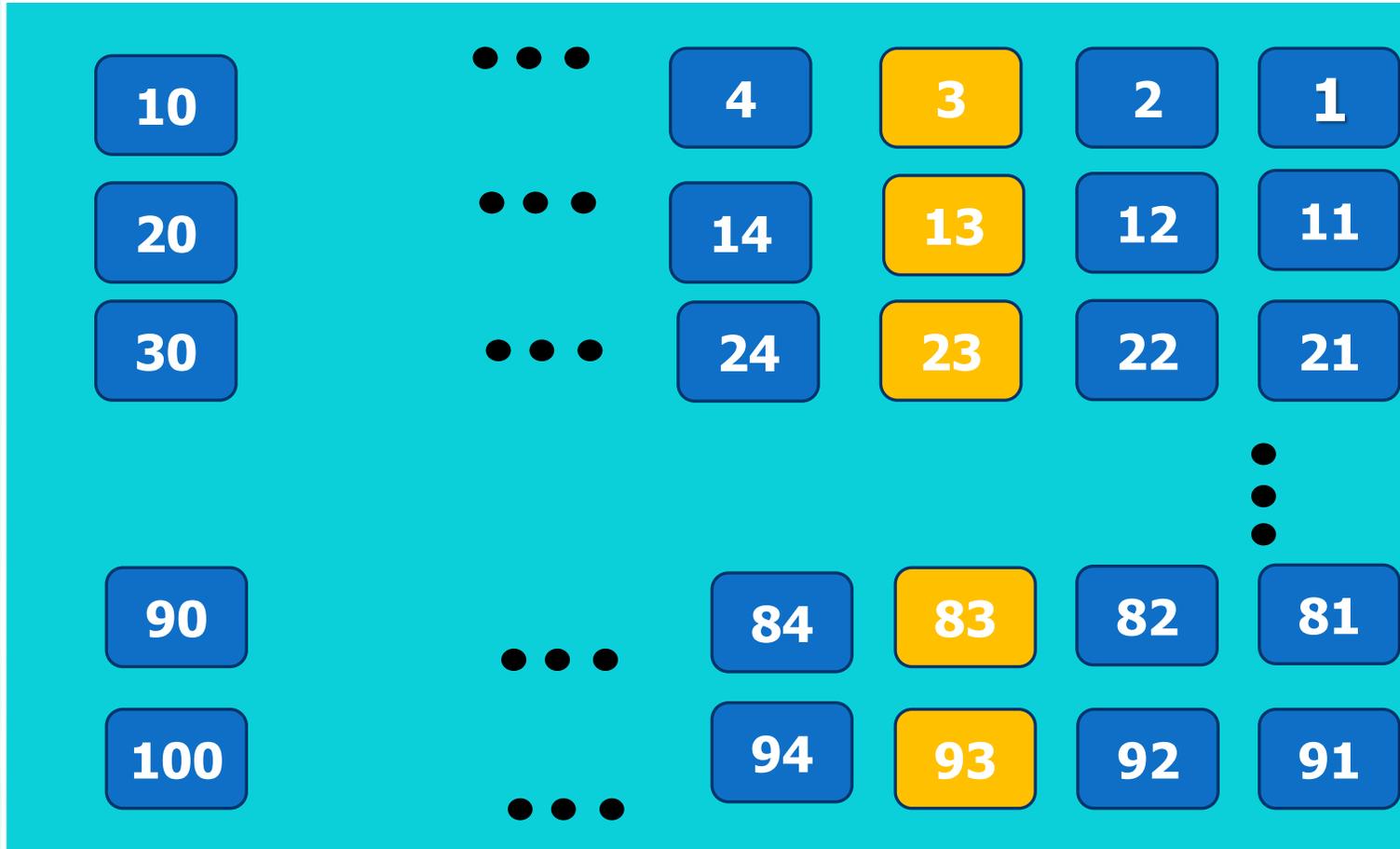
Types of Data

Examples	Quantitative Data		Qualitative Data	
	Discrete	Continuous	Nominal	Ordinal
Number of Children				
Speed of a Car				
Hair Color				
English Language Level				
Passport Numbers				
Room Size				
Academic Certificates				
Price				
Blood Pressure				
Gender				
Date				
Income and Salary				
Number of Hotels in Sulaimany				
Car Numbers				

Organize and Present Data

Organize & Present Data

Graphic

Frequency Distribution

IQ Test

■ High ■ Moderate ■ Low
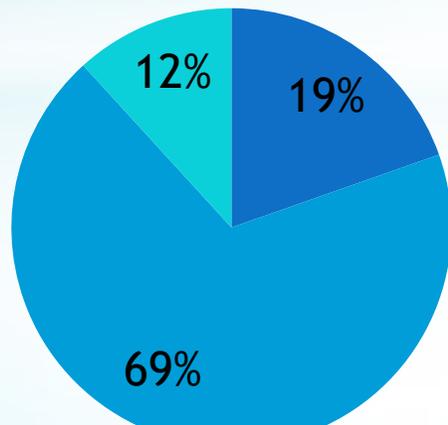
Daily Absence	Frequency
0	5
1	10
2	5
3	6

It is a primary data collected from the sources	The Raw Data
Data that has been organized in the **frequency distribution tables**	Classified Data
It is a table that shows "**classes**" or "**intervals**" of data entries with a count of the number of entries in each class	Frequency Distribution Table
Frequency of each category to the total frequencies $$p_i = \frac{f_i}{\sum f}$$	Relative Frequency
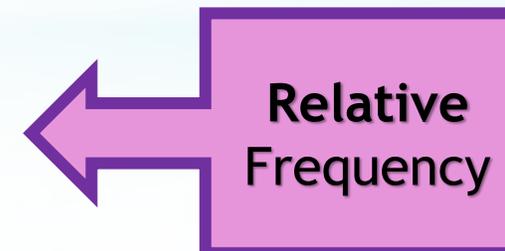
General Table of Frequency Distribution

Classes	Frequency f_i	Relative Frequency $p_i = \dfrac{f_i}{\sum f}$	Ratio $p_i\%$
Class (1)	f_1	$p_1 = \dfrac{f_1}{\sum f}$	$p_1 \times 100$
Class (2)	f_2	$p_2 = \dfrac{f_2}{\sum f}$	$p_2 \times 100$
Class (3)	f_3	$p_3 = \dfrac{f_3}{\sum f}$	$p_3 \times 100$
. . .	. . .	. . .	. . .
Class(k)	f_k	$p_k = \dfrac{f_k}{\sum f}$	$p_k \times 100$
Total	$\sum f_i$	1	100%

Example of Qualitative Data

The following table shows the qualifications of the employees of a company

PhD	BSc	BSc	MSc	PhD	MSc	MSc
BSc	BSc	Dip	BSc	BSc	BSc	High School
BSc	BSc	MSc	MSc	Dip	MSc	MSc
BSc	MSc	BSc	MSc	MSc	BSc	BSc
Dip	BSc	BSc	MSc	BSc	MSc	MSc

Form Frequency distribution table?

Qualification Classes	Tally Mark	f_i	p_i	$\% \, p_i$
PhD	‖	2	$\frac{2}{35} = 0.057$	0.057×100=5.7%
MSc	‖‖ ‖‖ ‖‖	13	$0.371 = \frac{13}{35}$	0.371×100=37.1
BSc	‖‖ ‖‖ ‖‖	16	0.457	45.7 %
Dip	‖‖	3	0.086	8.6 %
High School		1	0.029	2.9 %
Total		**35**	**1**	**100 %**

Example of Quantitative Data/Discrete

The following table shows the absence of 30 students in one month in KTI

2	0	1	0	3	0	0	3	0	2	2	1	0	0	1
2	4	0	4	2	1	0	1	0	0	2	0	1	3	2

Number of absence (Classes)	Tally Mark	Frequency	f_i	$\% f_i$
0	IIII IIII II	12	0.400	40%
1	IIII I	6	0.200	20%
2	IIII II	7	0.233	23.3%
3	III	3	0.100	10%
4	II	2	0.067	6.7%
Total		30	1	100 %

The following table shows the daily wages in dollars for 50 people in a company.

53	47	36	40	55	75	43	46	21	10
32	66	56	46	35	47	48	52	41	30
22	27	25	57	15	37	21	63	61	62
32	54	42	35	49	39	31	45	72	50
44	65	18	79	23	48	51	32	44	42

Steps to form the frequency distribution table

1) Calculate the Range (R) of the data, which is deference between the max and the min value of the data:

$$R = \text{max} - \text{min} = 79 - 10 = 69$$

2) Calculate the number of classes (K):

$$k = 1 + 3.3 \times \log n \qquad n = \text{number of data}$$

$$k = 1 + 3.3 \times \log 50 = 6.61 \approx 7$$

3) Calculate the length of classes(h):

$$h = \frac{R}{k} = \frac{69}{7} = 9.86$$

Suppose that the Midterm exam score of 50 students in a Biostatistics class were recorded bellow

50	29	72	7	36	30	19	67	46	42
41	41	59	17	44	30	56	37	31	39
18	56	11	77	69	21	54	33	54	44
40	34	56	22	39	80	23	39	31	83
78	17	73	7	28	62	29	51	53	20

Construct Frequency Distribution Table from the above data?

$R = Max - Min$
$= 83 - 7 = 76$

$K = 1 + 3.3 *log\ 50 = 6.6 \cong 7$

$H = R/K = 76 / 7 = 10.86 \cong 11$

Classes	Frequency f_i	Relative Frequency p_i	Percentage of p_i
7 – 18	5	0.1	10 %
18 – 29	7	0.14	14 %
29 – 40	13	0.26	26 %
40 – 51	8	0.16	16 %
51 – 62	8	0.16	16 %
62 – 73	4	0.08	8 %
73 – 84	5	0.1	10 %
Total	**50**	**1**	**100 %**

Graphical representation of data

It is the best useful graphic to represent the nominal data

Pie

1) Find the relative frequency p_i of the data

2) Drawing radius using the ruler

3) Determine the angle of each class using the law

Angle circular sector = $p_i \times 360°$

Steps

Example

In order to know the proportion of Kurdish nationality in Diyarbakir city, we collected a data of (1250) doctors in all the hospitals of the city. Then we got the following frequency distribution table:

Nationality Classes	Number of Doctors	Relative Frequency
Kurdish	900	0.72
Turkish	250	0.2
Others	Others	0.08
Total	1250	1

Angle circular sector for class Kurdish:

$$0.72 \times 360° \approx 259°$$
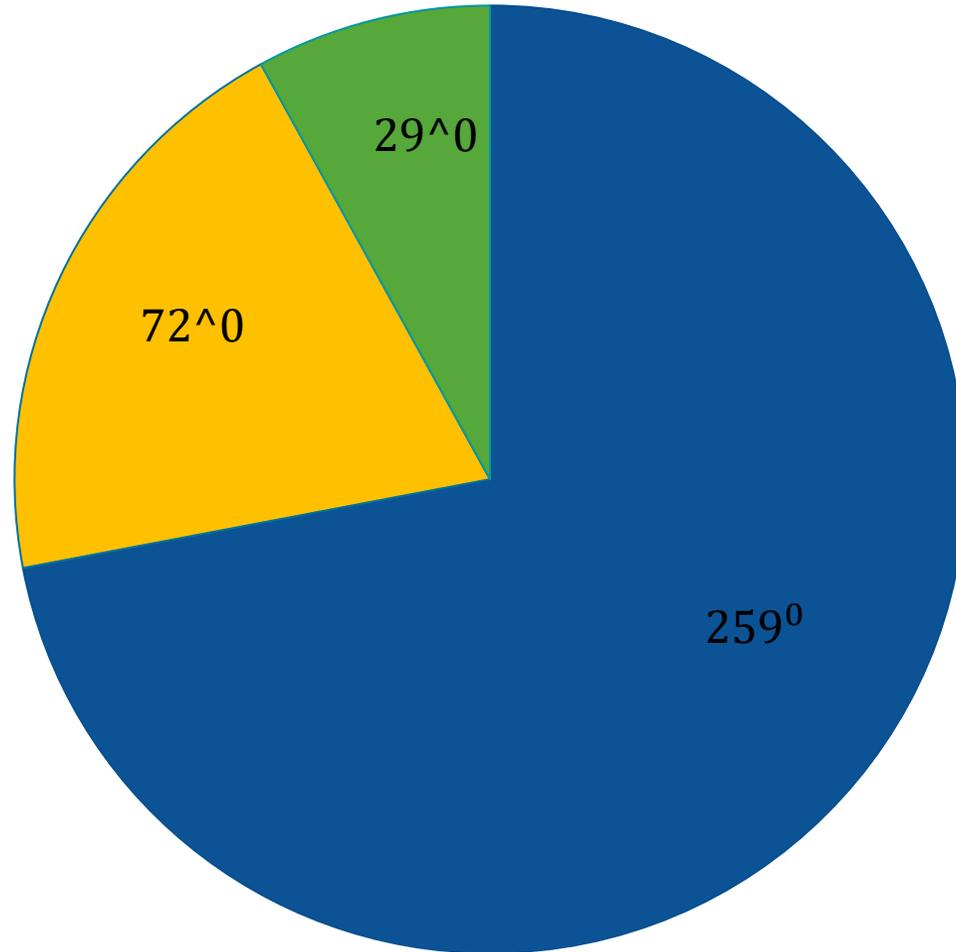
Angle circular sector for class Turkish:

$$0.2 \times 360° \approx 72°$$

Angle circular sector for class Others:

$$0.08 \times 360° \approx 29°$$

Example 2

The following table shows some of the important industries in a country by millions of dollars
Find the angle circular sector for (food) Industry?

Industries	production values
Mineral	500
Engineering	450
Food	300
Textile	250
Σ	1500

Angle circular sector $= p_i \times 360°$

$$Relative\ Frequency\ (p_i) = \frac{Frequency}{Total\ Frequency} = \frac{f_i}{\sum f}$$

$$p_i\ of\ food = \frac{300}{1500} = 0.2$$

Angle circular sector for class food:
$$p_i \times 360°$$

$$0.2 \times 360° = 72°$$

Graphical representation of Data

**It is the best graphic to represent qualitative data(Ordinal)and quantitative data(Discrete)**

Bar Chart

1) Calculate Frequency Distribution

2) Drawing two axis, the horizontal represents classes, and the vertical represents frequency f_i

3) Drawing columns with equal distance, and length depends on the number of frequency.

Steps

Example1

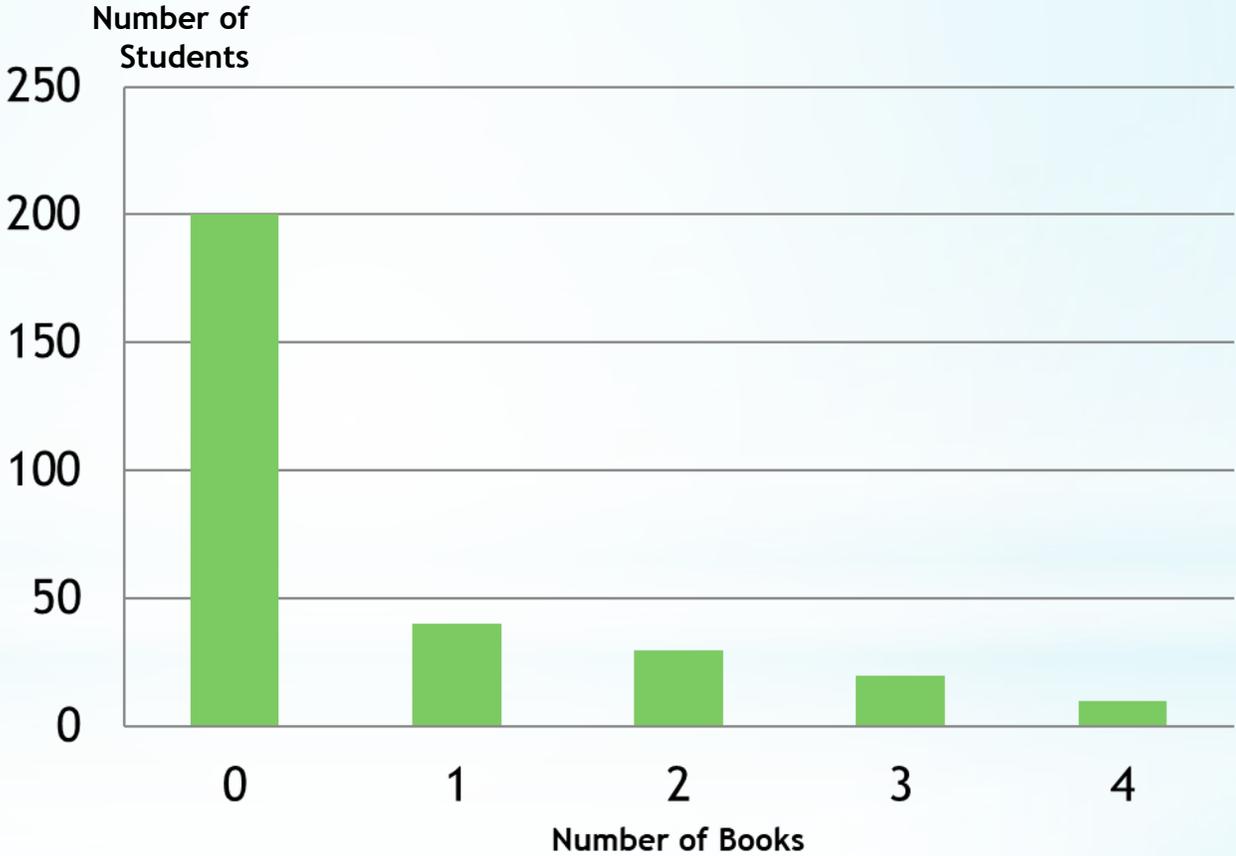The following table represents the frequency distribution of 4500 students studying abroad:

Classes	Classes
PhD	200
MSc	300
BSc	1000
Dip	2500
High School	500
Total	4500
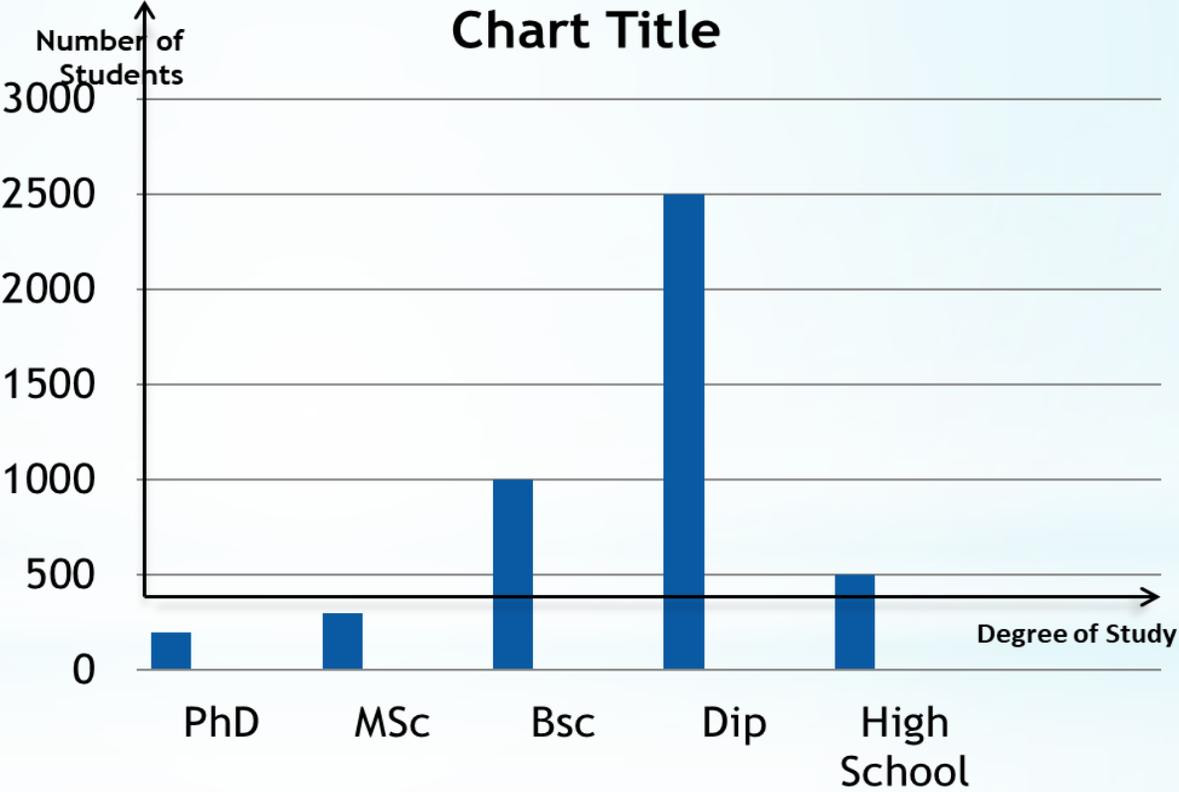
Example2

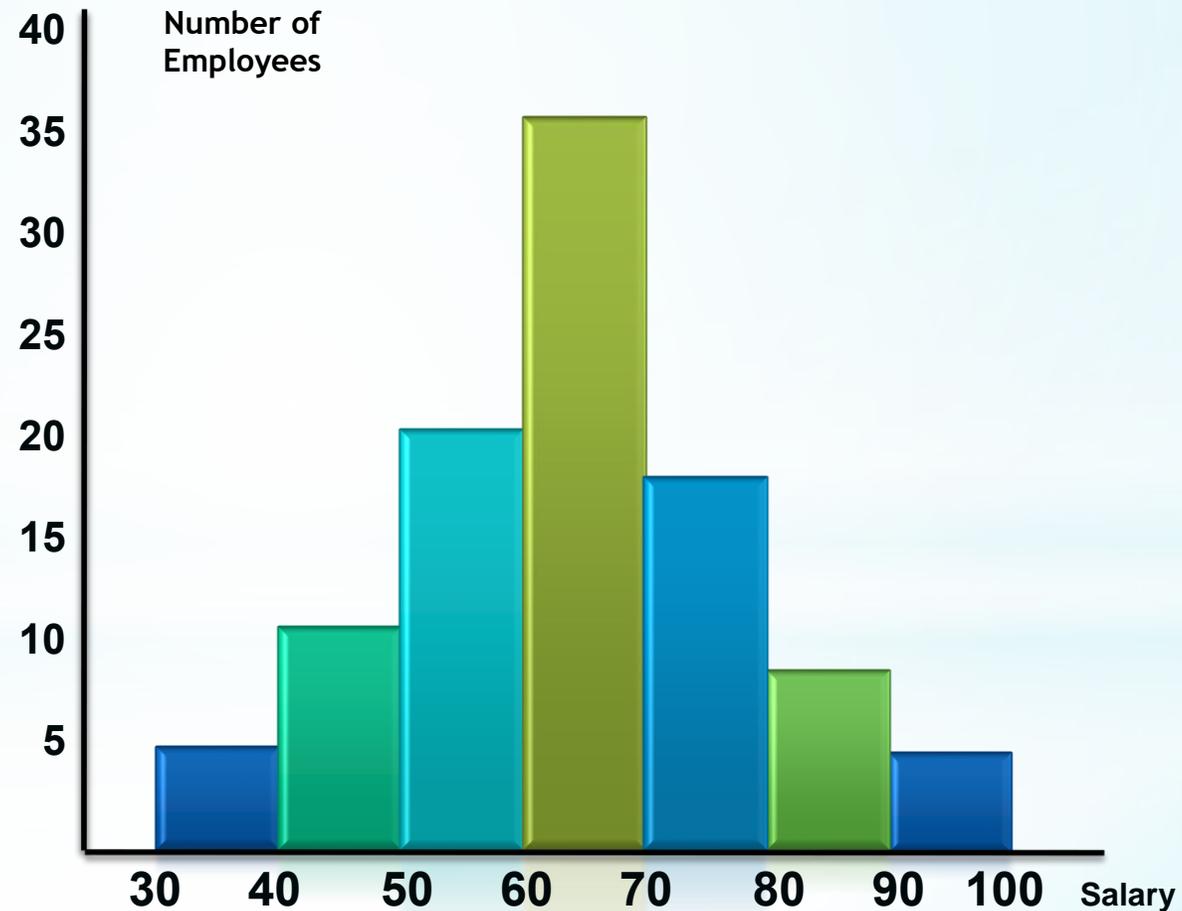| Number of books (Classes)	Number of books (Classes)
0	200
1	40
2	30
3	20
4	10
Total	300

Histograms are a special form of bar chart where the data represent continuous rather than discrete categories

Histogram

The following table shows salaries of a sample of 100 employees in a company

Salary (Classes)	Number of Employees
30-	4
40-	11
50-	20
60-	36
70-	17
80-	8
90-	4
Σ	100

Example

The following table shows salaries of a sample of 100 employees in a company
Line graphs are usually used to show time series data

Curved Frequency

classes	Frequency
30-40	4
40-50	11
50-60	20
60-70	36
70-80	17
80-90	8
90-100	4
Total	100
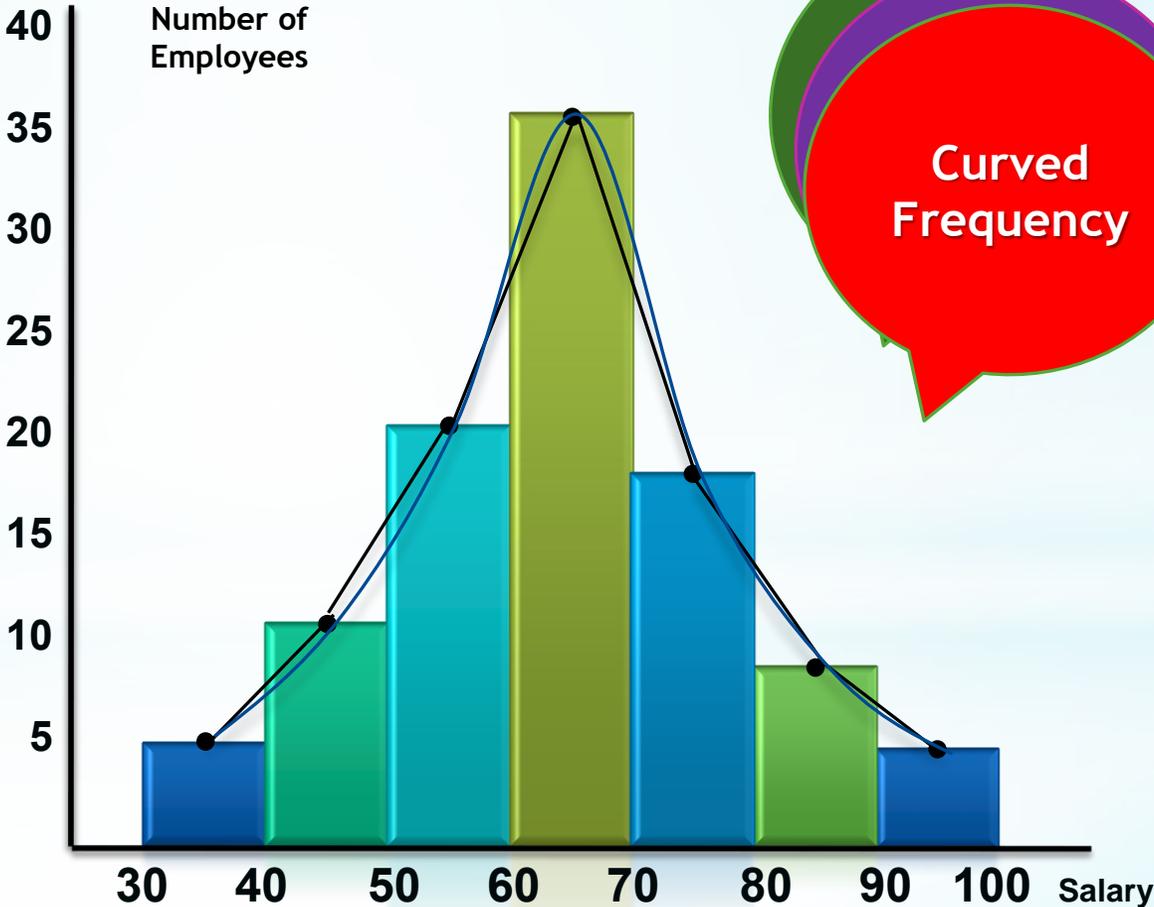
Good Luck for All Students

- **Please do not hesitate to contact me if you have any questions.**

- **M. Nozad H. Mahmood**

- **Email. nawzad@knights.ucf.edu**

- **Email. nozad.mahmood@sulicihan.edu.krd**