Northern Technical University
Technical college of Kirkuk
Surveying Tech. Eng. Dept.


2nd Stage
lecturer:
Sumaya Falih Hasan
Engineering Statistics

Topic 1

Email: Sumaya.h.falih@ntu.edu.iq

# outline

**What is Statistics?**

**Types of Statistics**

**Importance of Statistics**

**Frequency**

**Example**

**Relative frequency and percentage frequency-**

**Example**

**Homework**

# What is Statistics?

**Definition of Statistics?**

Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting and presenting empirical data. Statistics is a highly interdisciplinary field; research in statistics finds applicability in virtually all scientific fields and research questions in the various scientific fields motivate the development of new statistical methods and theory. In developing methods and studying the theory that underlies the methods statisticians draw on a variety of mathematical and computational tools.

# Types of Statistics

**Types of Statistics**

Statistics can be categorized into 2 types:

1. **Descriptive Statistics:** It is used for summarizing observations etc.

2. **Inferential Statistics:** It is used for interpreting the meaning of the descriptive stats.

# Importance of Statistics

**Importance of Statistics**

1. Statistics makes the work simple & provides a clear picture on the work we do on daily basis.

2. The statistical methods helps us to research on different streams such as medicine, economics, business, social science and so on.

3. Statistics provides us different types of organized data with the help of graphs, diagrams and charts.

4. Statistics comes handy while we do critical analysis.

# Frequency

**What Is Frequency Distribution?**

A frequency distribution is a representation, either in a graphical or tabular format, that displays the number of observations within a given interval. The interval size depends on the data being analyzed and the goals of the analyst. The intervals must be mutually exclusive and exhaustive. Frequency distributions are typically used within a statistical context. Generally, frequency distribution can be associated with the charting of a normal distribution.

The *frequency* (**f**) of a particular observation is the number of times the observation occurs in the data. The *distribution* of a variable is the pattern of frequencies of the observation. Frequency distributions are portrayed as frequency tables, histograms, or polygons.

**Range= Lowest value- highest value**

**Interval width (length interval) = Range ÷ Number of categories**

# Example

**Example – Constructing a frequency distribution table for large numbers of observations**

Thirty AA batteries were tested to determine how long they would last. The results, to the nearest minute, were recorded as follows:

423, 369, 387, 411, 393, 394, 371, 377, 389, 409, 392, 408, 431, 401, 363, 391, 405, 382, 400, 381, 399, 415, 428, 422, 396, 372, 410, 419, 386, 390

**Answer**

**Range= Lowest value- highest value**

**Interval width (length interval) = Range ÷ Number of categories(number of class)**

**Number of categories(number of class)=8**

The lowest value is 363 and the highest is 431.

**Range= 431-363=68**

**Interval width (length interval) =68/8=8.5 ≈ 9**

The completed frequency distribution table should look like this:

| Class interval | record | Frequency(f) |
|---|---|---|
| 363—372 | 𝍸 | 4 |
| 373—381 | 𝍷 | 2 |
| 382—390 | 𝍸 | 5 |
| 391—399 | 𝍸 | 5 |
| 400—408 | 𝍸 | 4 |
| 409—417 | 𝍸 | 4 |
| 418—426 | 𝍷 | 3 |
| 427—435 | 𝍷 | 3 |
| ∑ | 30 | 30 |

# Relative frequency and percentage frequency

**Relative frequency and percentage frequency**

An analyst studying these data might want to know not only how long batteries last, but also what proportion of the batteries falls into each class interval of battery life.

This *relative frequency* of a particular observation or class interval is found by dividing the frequency (**f**) by the number of observations (**n**): that is, (**f ÷ n**). Thus:

**Relative frequency = frequency ÷ number of observations**

The *percentage frequency* is found by multiplying each relative frequency value by 100. Thus:

**Percentage frequency = relative frequency X 100 = f ÷ n X 100**

| Class interval | Frequency(f) | Relative frequency | Percentage frequency |
|---|---|---|---|
| 363—372 | 4 | 4/30=0.1333 | 13.333 |
| 373—381 | 2 | 0.0666 | 6.666 |
| 382—390 | 5 | 0.1666 | 16.666 |
| 391—399 | 5 | 0.1666 | 16.666 |
| 400—408 | 4 | 0.1333 | 13.333 |
| 409—417 | 4 | 0.1333 | 13.333 |
| 418—426 | 3 | 0.1 | 10 |
| 427—435 | 3 | 0.1 | 10 |
| ∑ | 30 | 0.9997≈1 | 99.997≈100 |

**Homework :-**

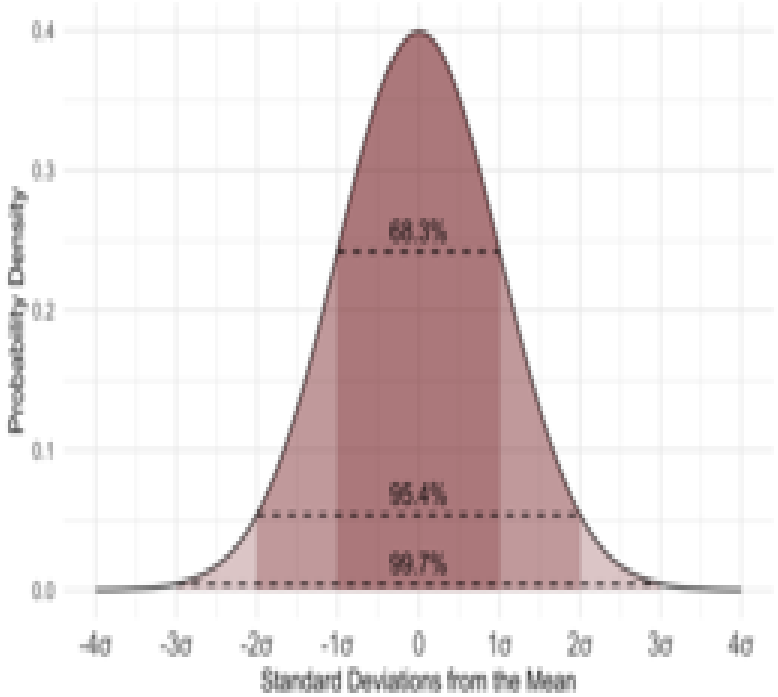
The marks obtained by 40 students of class V in an examination are given below

16, 17, 18, 3, 7, 23, 18, 13, 10, 21, 7, 1, 13, 21, 13, 15, 19, 24, 16, 2, 23, 5, 12, 18, 8, 12, 6, 8, 16, 5, 3, 5, 0, 7, 9, 12, 20, 10, 2, 23

**Constructing a frequency distribution table when the (number of class)=5**

**Then calculate the  Relative and percentage frequency.**

**Northern Technical University**
**Technical college of Kirkuk**
**Surveying Tech. Eng. Dept.**




**2nd Stage**
**lecturer:**
**Sumaya Falih Hasan**
**Engineering Statistics**

**Topic 2**

**Email: Sumaya.h.falih@ntu.edu.iq**

# outline

**Constructing a frequency distribution table**

**1-Relative frequency &Percentage frequency**

**2-Middle interval**

**3-Exact interval**

**4-Frequency curve**

**Example**

**Homework**

# 1-Relative frequency &Percentage frequency

**Relative frequency and percentage frequency**

An analyst studying these data might want to know not only how long batteries last, but also what proportion of the batteries falls into each class interval of battery life.

This *relative frequency* of a particular observation or class interval is found by dividing the frequency (**f**) by the number of observations (**n**): that is, (**f ÷ n**). Thus:

**Relative frequency = frequency ÷ number of observations**

The *percentage frequency* is found by multiplying each relative frequency value by 100. Thus:

**Percentage frequency = relative frequency X 100 = f ÷ n X 100**

# 2-Middle interval

**Middle interval = (Lowest value + highest value)/2 for each class**

| Class interval | f | Mid. interval |
|---|---|---|
| **360–369** | 2 | 364.5 |
| **370–379** | 3 | 374.5 |

**Middle interval= 360+369/2=364.5 for class 1**

**Middle interval= 370+379/2=374.5 for class 2**

# 3-Exact interval

**Exact interval = lower interval-0.5 ---- high interval+0.5**

| Class interval | Exact interval |
|---|---|
| **360–369** | 359.5–369.5 |
| **370–379** | 369.5–379.5 |

**Exact interval = 360-0.5---- 369+0.5=359.5–369.5 for class 1**

**Exact interval = 370-0.5 ---- 379+0.5=369.5–379.5 for class 2**

# 4-Frequency curve

**Frequency curve:**

A frequency-curve is a smooth curve for which the total area is taken to be unity.

**Types of frequency curves**

There exist four types of frequency-curves namely

(1) Bell-shaped curve

(2) U-shaped curve

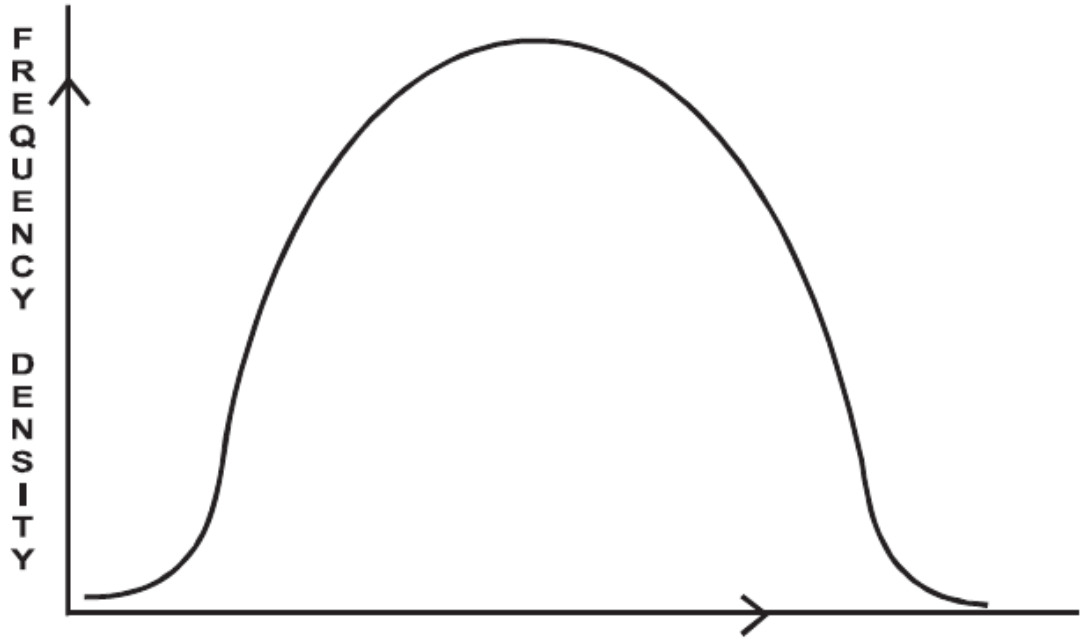
(3) J-shaped curve

(4) Mixed curve.

# Types of frequency curves

**1-Bell - shaped curve**

**Most of the commonly used distributions provide bell-shaped curve, which, as suggested by the name, looks almost like a bell.**

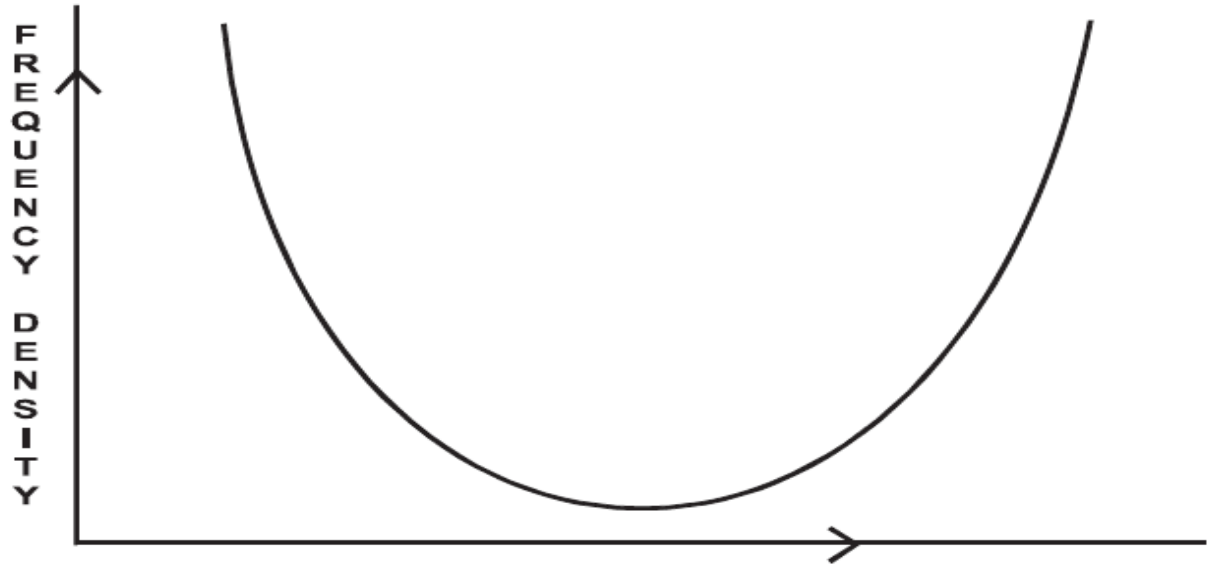



Bell-shaped curve

# Types of frequency curves

**2-U - shaped curve**

**For a U-shaped curve, the frequency is minimum near the central part and the frequency slowly but steadily reaches its maximum at the two extremities.**

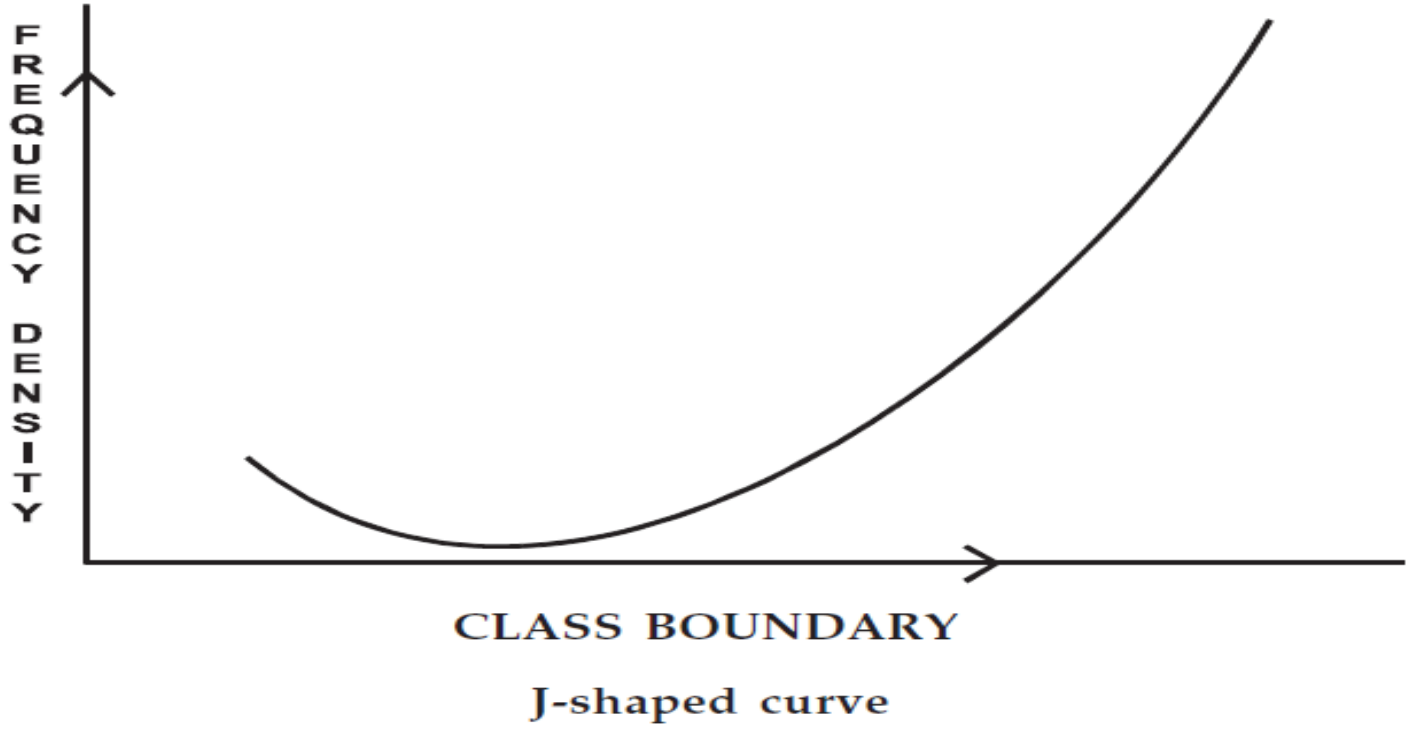



U-shaped curve

# Types of frequency curves

**3- J - shaped curve**

**The J-shaped curve starts with a minimum frequency and then gradually reaches its maximum frequency at the other extremity.**

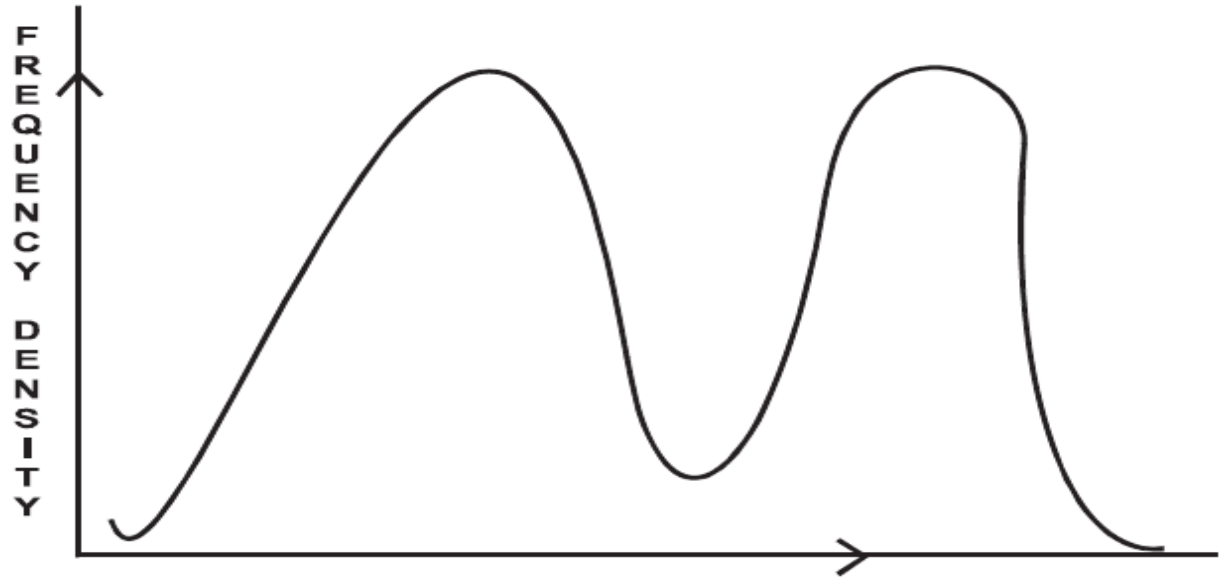



J-shaped curve

# Types of frequency curves

**4-Mixed curve**

**Lastly, we may have a combination of these frequency- curves, known as mixed curve.**

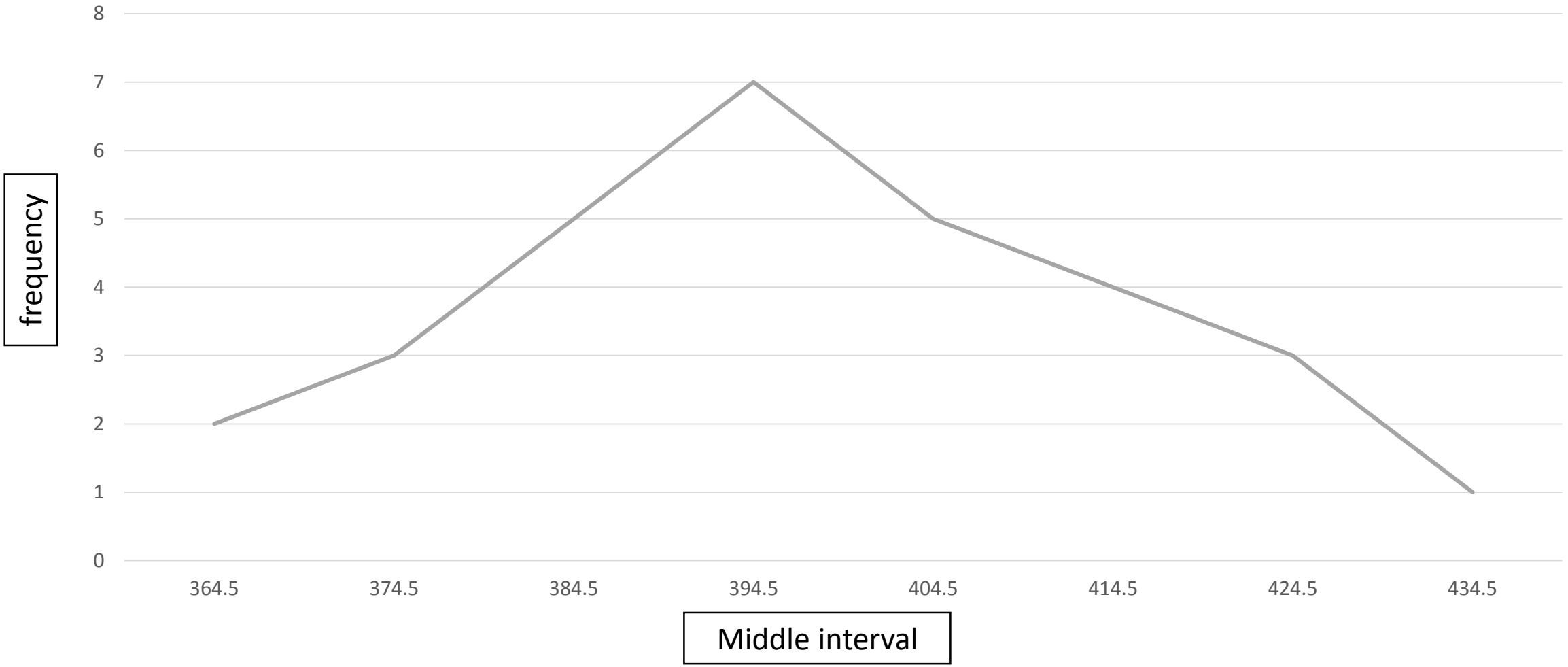



Mixed Curve

# Example

Complete the previous example

| Class interval | f | Mid. interval | Exact interval |
|---|---|---|---|
| **360–369** | 2 | 364.5 | 359.5–369.5 |
| **370–379** | 3 | 374.5 | 369.5–379.5 |
| **380–389** | 5 | 384.5 | 379.5–389.5 |
| **390–399** | 7 | 394.5 | 389.5–399.5 |
| **400–409** | 5 | 404.5 | 399.5–409.5 |
| **410–419** | 4 | 414.5 | 409.5–419.5 |
| **420–429** | 3 | 424.5 | 419.5–429.5 |
| **430–439** | 1 | 434.5 | 429.5–439.5 |
| **Total** | 30 | | |


frequency
8
7
6
5
4
3
2
1
0
364.5
374.5
384.5
394.5
404.5
414.5
424.5
434.5
Middle interval

## Homework :-

For the following shown below, Find

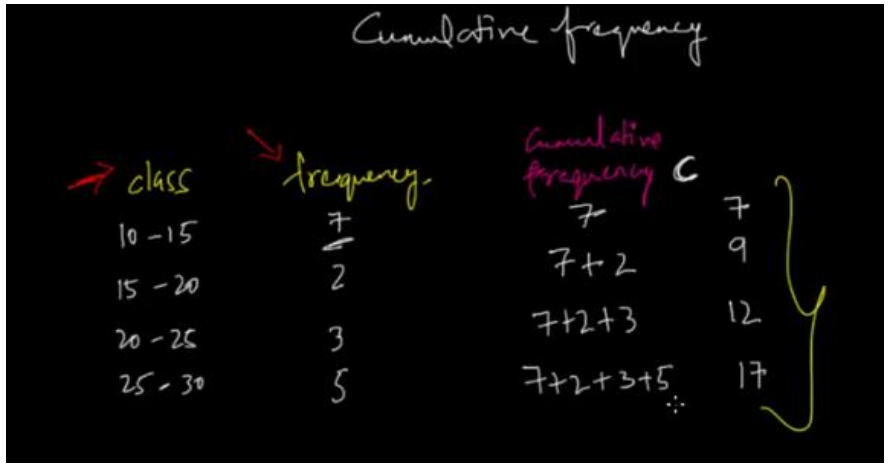
1-Relative frequency &Percentage frequency

2-Middle interval

3-Exact interval

| Interval | Frequency |
|---|---|
| 19----21 | 2 |
| 22----24 | 4 |
| 25----27 | 6 |
| 28----30 | 8 |
| 31----33 | 10 |
| 34----36 | 12 |

**Northern Technical University**
**Technical college of Kirkuk**
**Surveying Tech. Eng. Dept.**




**2nd Stage**
**lecturer:**
**Sumaya Falih Hasan**
**Engineering Statistics**

**Topic 3**

**Email: Sumaya.h.falih@ntu.edu.iq**

# outline

**Constructing a cumulative frequency distribution table**

**1-*Cumulative frequency***

**2- Cumulative Frequency Graph**

**3-Upper cumulative frequency**

**4-Lower cumulative frequency**

**Data display**

**Homework**

***Cumulative frequency***

*Cumulative frequency*

is used to determine the number of observations that lie above (or below) a particular value in a data set. The cumulative frequency is calculated using a frequency distribution table, which can be constructed from stem and leaf plots or directly from the data.

The cumulative frequency is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors. The last value will always be equal to the total for all observations, since all frequencies will already have been added to the previous total

**What is a Cumulative Frequency Graph?**

A Cumulative Frequency Graph is a graph plotted from a cumulative frequency table. A cumulative frequency graph is also called cumulative frequency curve.

# Upper cumulative frequency(U.C.F)

**Example :**

**The marks of 200 students in a test were recorded and shown by the following frequency distribution.**

| Marks % | Number of Students |
|---|---|
| 10 - 19 | 7 |
| 20 - 29 | 11 |
| 30 - 39 | 20 |
| 40 - 49 | 46 |
| 50 - 59 | 57 |
| 60 - 69 | 37 |
| 70 - 79 | 15 |
| 80 - 89 | 7 |

**A** Construct the upper cumulative frequency table.

Also answer the following.

1- How many students obtained less than 50 marks?

2- How many students obtained at least 60 marks?

**Solution:**

| Class Interval | Frequency | Exact interval | U.C.F | Cumulative Frequency | F(cumulative) |
|---|---|---|---|---|---|
| 10—19 | 7 | 9.5—19.5 | ≤19.5 | 7 | 7 |
| 20—29 | 11 | 19.5—29.5 | ≤29.5 | 7+11=18 | 18 |
| 30—39 | 20 | 29.5—39.5 | ≤39.5 | 7+11+20=38 | 38 |
| 40—49 | 46 | 39.5 — 49.5 | ≤49.5 | 7+11+20+46=84 | 84 |
| 50—59 | 57 | 49.5 —59.5 | ≤59.5 | 7+11+20+46+57=141 | 141 |
| 60—69 | 37 | 59.5—69.5 | ≤69.5 | 7+11+20+46+57+37=178 | 178 |
| 70—79 | 15 | 69.5—79.5 | ≤79.5 | 7+11+20+46+57+37+15=193 | 193 |
| 80—89 | 7 | 79.5 — 89.5 | ≤89.5 | 7+11+20+46+57+37+15+7=**200** | 200 |
| ∑ | **200** | | | | |

1- The number of students obtaining less than 50 marks

= the cumulative frequency of the **class interval 40 - 49 = 84.**

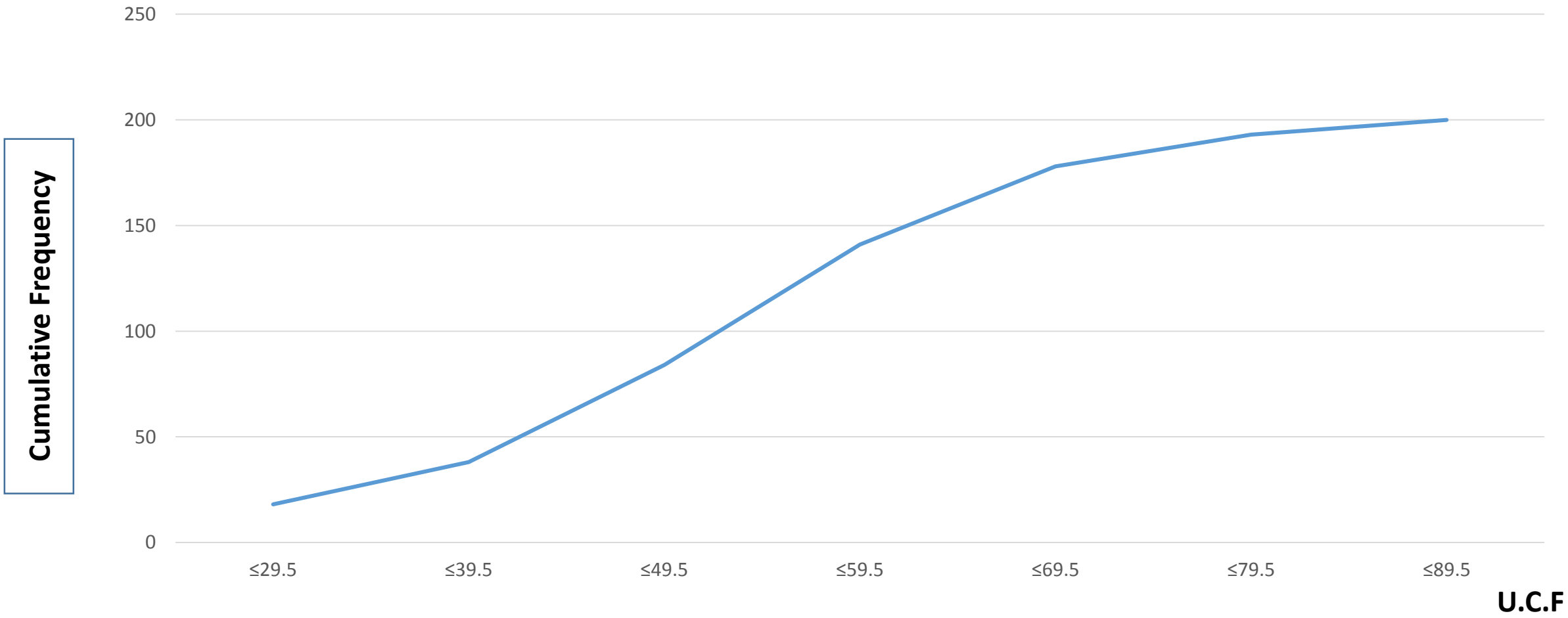
2- The number of students obtaining at least 60 marks

= total number of students - the number of students getting less than or equal to 59

= **200 - 141**

= **59**.

B Draw a cumulative frequency graph for the frequency table.

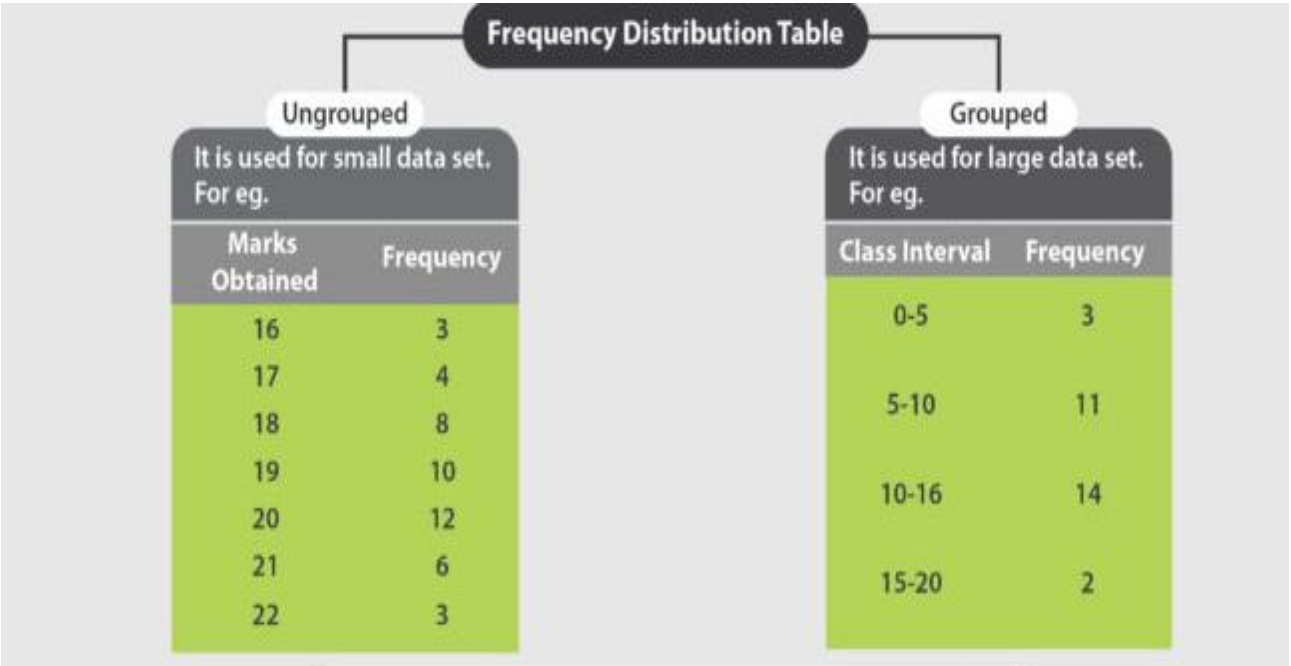


C Construct the Lower cumulative frequency(L.C.F)table.

| Class Interval | Frequency | Exact interval | L.C.F | Cumulative Frequency |
|---|---|---|---|---|
| 10—19 | 7 | 9.5—19.5 | ≥9.5 | 200 |
| 20—29 | 11 | 19.5—29.5 | ≥19.5 | 200-7=193 |
| 30—39 | 20 | 29.5—39.5 | ≥29.5 | 200-7-11=182 |
| 40—49 | 46 | 39.5 — 49.5 | ≥39.5 | 200-7-11-20=162 |
| 50—59 | 57 | 49.5 —59.5 | ≥49.5 | 200-7-11-20-46=116 |
| 60—69 | 37 | 59.5—69.5 | ≥59.5 | 200-7-11-20-46-57=59 |
| 70—79 | 15 | 69.5—79.5 | ≥69.5 | 200-7-11-20-46-57-37=22 |
| 80—89 | 7 | 79.5 — 89.5 | ≥79.5 | 200-7-11-20-46-57-37-15=7 |
| ∑ | **200** | | | |

# Data display

It is the different values of that quantity represented together in a set. It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information. The source of data ( primary data, secondary data) is also an important factor.

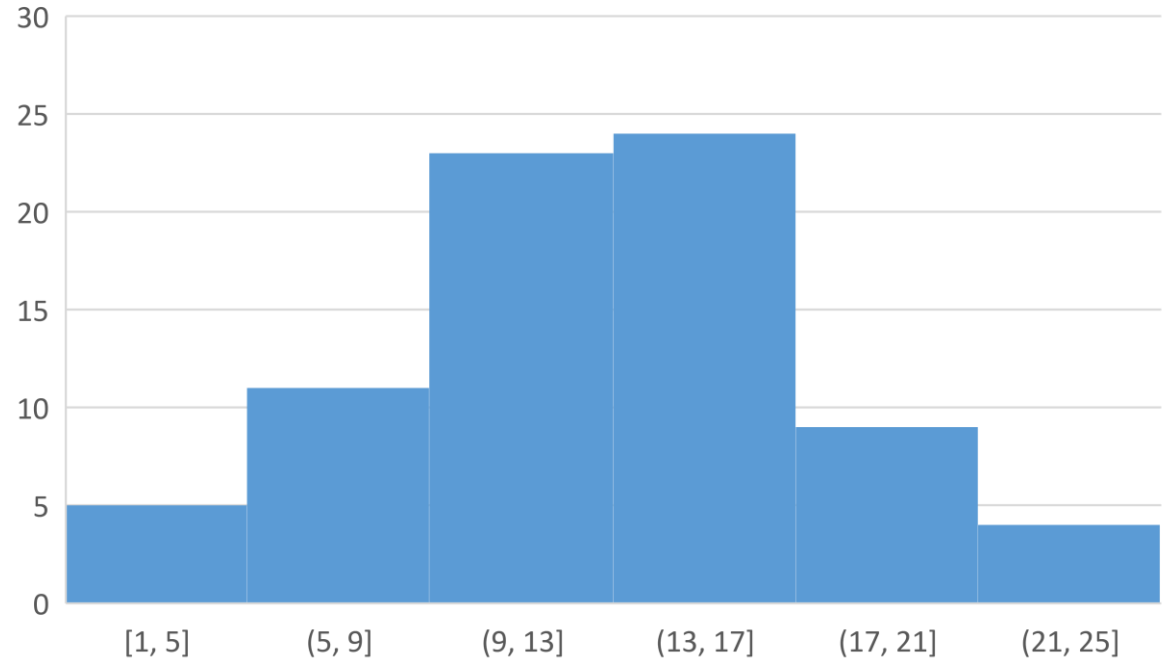


## graphical representation of frequency distribution table

1-histogram

X axis represent the exact interval

Y axis represent the frequency or cumulative frequency

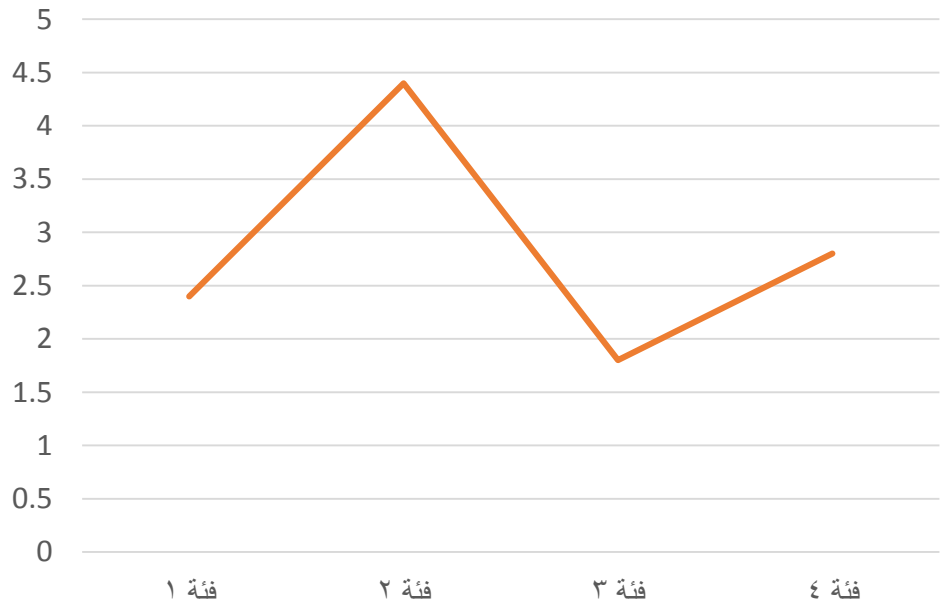


**graphical representation of frequency distribution table**

2-frequency polygons

X axis represent the middle class

Y axis represent the frequency or cumulative frequency

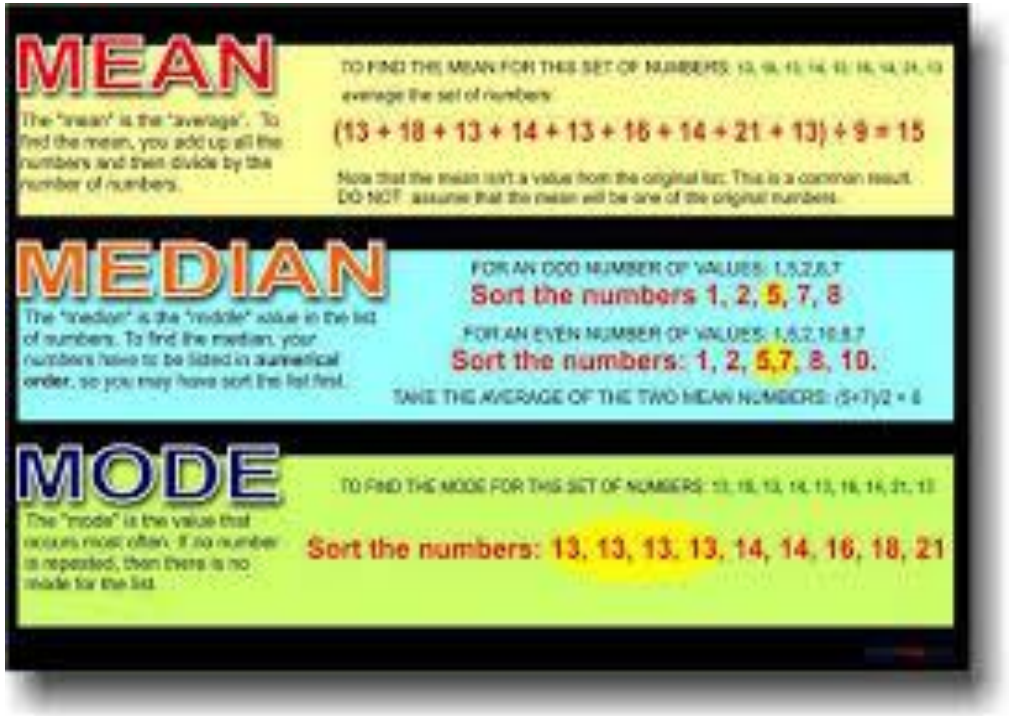


# Homework3

**Example:** For the following shown below, Find
1- upper cumulative frequency
2-lower cumulative frequency
3-display (or draw)  the data of upper cumulative frequency by two methods

| Length (x mm) | Frequency |
|---|---|
| 11 – 15 | 2 |
| 16 – 20 | 4 |
| 21 – 25 | 8 |
| 25 – 30 | 14 |
| 31 – 35 | 6 |
| 36 – 40 | 4 |
| 41 – 45 | 2 |

**Northern Technical University**
**Technical college of Kirkuk**
**Surveying Tech. Eng. Dept.**




**2nd Stage**
**lecturer:**
**Sumaya Falih Hasan**
**Engineering Statistics**

**Topic 4-1**

**Email: Sumaya.h.falih@ntu.edu.iq**

# outline

**measure of central tendency**

**A- Mean** **B-Median** **C-mode**

**A- Mean**

**The Mean of the Individual Data (Not Classified)**

**The Mean of the Classified Data:**

- **Arithmetic mean by general rule**
- **Arithmetic mean by central premise**
- **Arithmetic mean by condensed distraction**
- **Weighted Mean**
- **Harmonic mean**
- **Geometric mean**
- **Homework**

# measure of central tendency

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data. As such, measures of central tendency are sometimes called measures of central location.

It’s a numerical measures used to measure the data concentrate position or cluster. In general the data is tend to be concentrated around a particular values.

These values is so-called central tendency measures These measures are used to summarize data numerically, where these data values are considered to be a typical or ideal.

# A- Mean ($\bar{X}$)

The mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data. The mean is equal to the sum of all the values in the data set divided by the number of values in the data set.

**The Mean of the Individual Data (Not Classified)**

$$\text{Mean} = \frac{Sum of all observations}{number of observations}$$

Thus, the mean of n observation $x_1, x_2, \ldots, x_n$, is given by

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \sum_{i=1}^{n} x_i$$

**Example:**

Find the average (arithmetic mean) of the following

views, which is a weights (kg) for a group of seven people:

25,30, 40, 45, 35, 55 and 50.

$$\bar{X} = \frac{\sum_i^n x_i}{n} = \frac{25+30+40+45+35+55+50}{7} = 40$$

## The Mean of the Classified Data:

### 1-Arithmetic mean by general rule

**Input Data :**
Data set x = 1, 2, 4, 5, 8
Data set y = 5, 20, 40, 80, 100
Total number of elements = 5

**Objective :**
Find what is group arithmetic mean for given input data?

**Formula :**

$$\text{Group Arithmetic Mean } \bar{x} = \frac{\sum xy}{\sum y}$$

**Solution :**

$$\bar{x} = \frac{(1 \times 5) + (2 \times 20) + (4 \times 40) + (5 \times 80) + (8 \times 100)}{5 + 20 + 40 + 80 + 100}$$

$$= \frac{5 + 40 + 160 + 400 + 800}{245}$$

$$= \frac{1405}{245}$$

Group Arithmetic Mean = 5.7347

$$\bar{X} = \frac{\sum_k^n x_k f_k}{n} = \frac{\sum_k^n x_k f_k}{\sum_k^n f_k} = \frac{x_1 f_1 + x_2 f_2 + \ldots + x_k f_k}{f_k + f_k + \ldots + f_k}$$

**Example:**
Find the mean of the hemoglobin level in the blood sample for 50 people.

| CLASS INTERVAL | Freq.(f) | X(period central) Middle | X*f |
|---|---|---|---|
| 11–13 | 3 | (11+13)/2=12 | 12*3=36 |
| 14–15 | 5 | (14+15)/2=14.5 | 14.5*5=72.5 |
| 16–17 | 15 | (16+17)/2=16.5 | 16.5*15=247.5 |
| 18–19 | 16 | 18.5 | 296 |
| 20–21 | 10 | 20.5 | 205 |
| 22–23 | 1 | 22.5 | 22.5 |
| ∑ | **50** | | **879.5** |

Average (arithmetic mean ) for the hemoglobin level is

$$\mathbf{A.M}=\bar{X}=\frac{\sum(X*f)}{\sum f}$$

$$=\frac{879.5}{50}=17.59$$

## 2-Arithmetic mean by central premise

$$\text{A.M}=\bar{X} = C.P + \frac{\sum(D.V * f)}{\sum f}$$

$$D.V = X - C.P$$

**Where: C.P=central premise**

**D.V=deviation value about arithmetic mean**

**f =frequency**

**X=mid. interval or class**

**EX: find the arithmetic mean by central premise for this data assume C.P=50**

| Class interval | f | Mid. Interval (x) | D.V= (X-C.P) | D.V*f |
|---|---|---|---|---|
| 22–26 | 9 | (22+26)/2=24 | (24-50)=-26 | (-26*9)=-234 |
| 27–31 | 3 | (27+31)/2=29 | (29-50)=-21 | (-21*3)=-63 |
| 32–36 | 10 | (32+36)/2=34 | (34-50)=-16 | (-16*10)=-160 |
| 37–41 | 8 | 39 | -11 | -88 |
| 42–46 | 12 | 44 | -6 | -72 |
| 47–51 | 8 | 49 | -1 | -8 |
| ∑ | 50 | | | -625 |

$$A.M = C.P + \frac{\sum (D.V * f)}{\sum f}$$

$$A.M = 50 + \frac{-625}{50}$$

A.M=$\bar{X}$ =37.5

## 3-Arithmetic mean by condensed distraction

$$\text{A.M}=\bar{X} = C.P + \frac{\sum \left( (\frac{DV}{L}) * f \right)}{\sum f} * L$$

$$D.V = X - C.P$$

L=high interval - lower interval +1

C.P=central premise

D.V=deviation value about arithmetic mean

L=length interval

f =frequency

X=mid. interval or class

**EX: find the arithmetic mean by condensed distraction for this data assume C.P=50**

| Class interval | f | Mid. Interval (x) | D.V= (X-C.P) | $\frac{DV}{L}$ | $\left(\frac{DV}{L}\right) * f$ |
|---|---|---|---|---|---|
| 22–26 | 9 | 24 | (24-50)= -26 | $\frac{-26}{5}$=-5.2 | (-5.2*9)=-46.8 |
| 27–31 | 3 | 29 | (29-50)=-21 | $\frac{-21}{5}$=-4.2 | (-4.2*3)=-12.6 |
| 32–36 | 10 | 34 | (34-50)=-16 | $\frac{-16}{5}$=-3.2 | (-3.2*10)=-32 |
| 37–41 | 8 | 39 | (39-50)=-11 | -2.2 | -17.6 |
| 42–46 | 12 | 44 | (44-50)=-6 | -1.2 | -14.4 |
| 47–51 | 8 | 49 | (49-50)=-1 | -0.2 | -1.6 |
| Total | 50 | | | | -125 |

$$A.M = C.P + \frac{\sum \left(\left(\frac{DV}{L}\right) * f\right)}{\sum f} * L$$

$$A.M = 50 + \frac{-125}{50} * 5$$

Length interval = L=26-22+1=5

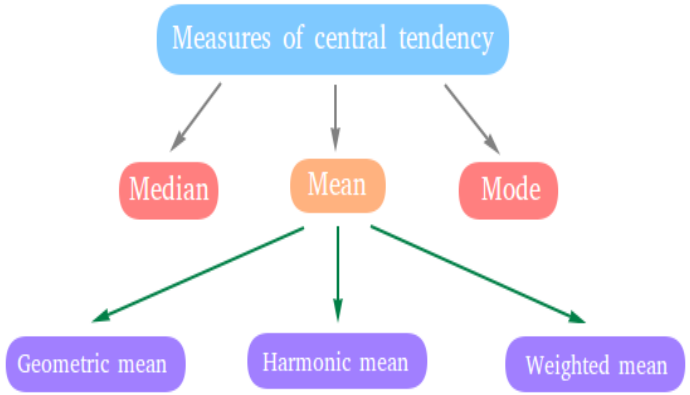
A.M=$\bar{X}$ =37.5

## Homework4

find the arithmetic mean for this data by three methods:
**assume C.P=40**

| Class interval | f |
|---|---|
| 70–74 | 8 |
| 75–79 | 8 |
| 80–84 | 5 |
| 85–89 | 11 |
| 90–94 | 9 |
| 95–99 | 5 |
| Total | |

**Northern Technical University**
**Technical college of Kirkuk**
**Surveying Tech. Eng. Dept.**




**2nd Stage**
**lecturer:**
**Sumaya Falih Hasan**
**Engineering Statistics**

**Topic 4-2**

**Email: Sumaya.h.falih@ntu.edu.iq**

# outline

**measure of central tendency**

**A- Mean**

**Types of mean**

**Weighted Mean**

**Harmonic mean**

**Geometric mean**

**Homework**

## Weighted Arithmetic Mean: ➢

When the values are not of equal importance, we assign them certain numerical values to express their relative importance. These numerical values are called weights.

**If $X_1, X_2, \ldots\ldots, X_k$ have weights $W_1, W_2, \ldots\ldots, W_3$,** then **the weighted arithmetic mean** or **the weighted mean** which is denoted as $\bar{X}_w$, is calculated by the following formula;

$$\bar{X}_w = \frac{W_1X_1 + W_2X_2 + \cdots\cdots + W_kX_k}{W_1 + W_2 + \cdots\cdots + W_k} = \frac{\sum WX}{\sum W}$$

Thus the mean of grouped data may be regarded as the weighted mean of the values of the values

**$X_1, X_2, \ldots\ldots, X_k$ whose weights are the respective class frequencies $f_1, f_2, \ldots\ldots, f_k$**

**For Example:**

The marks obtained by a student in **English, Arabic and Sciences were 70, 76, and 82** respectively. Find the appropriate average if weights of **5, 4 and 3** are assigned to these subjects.

We will use the **weighted mean, the weights attached to the marks being 5, 4 and 3. Thus,**

$$\bar{x}_w = \frac{\sum wx}{\sum w}$$

$$= \frac{5(70) + 4(76) + 3(82)}{5 + 4 + 3}$$

$$= \frac{350 + 304 + 246}{12}$$

$$= 75$$

$$A.M = \frac{x1w1 + x2w2 + \cdots xnwn}{w1 + w2 + \cdots wn}$$

**X=view**

**W=frequency**

**Geometric Mean:** ➢

The geometric mean, G, of a set of *n* positive values $X_1$, $X_2$, ……, $X_n$ is the nth root of the product of the values. Mathematically the formula for geometric mean will be as follows;

$$G = \sqrt[n]{X_1, X_2, \dots\dots, X_n}$$

$$= (X_1, X_2, \dots\dots, X_n)^{1/n}$$

**For Example:**

The geometric mean of the values 2, 4 and 8 will be

$$G = \sqrt[3]{2 \times 4 \times 8}$$

$$= \sqrt[3]{64}$$

$$= 4$$

**Geometric Mean for Grouped Data:**

When the data have been arranged into a frequency distribution, each of the original observation in a class is assumed to have a value equal to its class marks. Suppose $X_1, X_2, \ldots\ldots, X_k$ represents the class marks in a frequency distribution with $f_1, f_2, \ldots\ldots, f_k$ as the corresponding class frequencies, where $f_1 + f_2 + \ldots\ldots\ldots + f_k = \sum f = n$. **since** $X_1$ **occurs** $f_1$ **times,** $X_2$ **occurs** $f_2$ **times,**$\ldots\ldots\ldots$**,** $X_k$ **occurs** $f_k$ times, then the formula for the geometric mean will be as;

$$G = \sqrt[n]{X_1^{f_1} \cdot X_2^{f_2} \cdot \ldots\ldots \cdot X_k^{f_k}}$$

Where, $n = \sum f$. This is sometimes called **the weighted geometric mean** with weights $f_1, f_2, \ldots\ldots, f_k$.

**For Example:**

The geometric mean for the following distribution tables, by using the basic formula is as follows;

| x | f |
|---|---|
| 1 | 2 |
| 2 | 3 |
| 3 | 4 |
| 4 | 1 |
| Total | 10 |

$$G = \sqrt[n]{X_1^{f_1} \cdot X_2^{f_2} \cdot \ldots\ldots \cdot X_k^{f_k}}$$

$$G = (X_1^{f_1} \cdot X_2^{f_2} \cdot \ldots\ldots \cdot X_k^{f_k})^{1/n}$$

$$= [(1)^2 \cdot (2)^3 \cdot (3)^4 \cdot (4)^1]^{1/10}$$

$$= (2592)^{1/10}$$

$$= 2.1946$$

## Harmonic Mean: ➢

The harmonic mean, **H, of a set of n values $X_1$, $X_2$, ……, $X_n$** is the reciprocal of the arithmetic mean of the reciprocals of the values. Mathematically, the formula for harmonic mean will be as follows

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots\cdots + \frac{1}{X_n}}$$

$$= \frac{n}{\sum\left(\frac{1}{X}\right)}$$

**For Example:**

The harmonic mean of the value **3, 5, 6, 6, 7, 10 and 12** will be as follows;

$$H = \frac{7}{\frac{1}{3}+\frac{1}{5}+\frac{1}{6}+\frac{1}{6}+\frac{1}{7}+\frac{1}{10}+\frac{1}{12}} = \frac{7}{1.1929} = 5.87$$

**Harmonic Mean for Grouped Data:** •

Suppose $X_1, X_2, \ldots\ldots, X_k$ represents the class marks in a frequency distribution • with $f_1, f_2, \ldots\ldots, f_k$ as the corresponding class frequencies, where $f_1 + f_2 + \ldots\ldots\ldots + f_k = \sum f = n$. Then the reciprocals of the class marks will be

The formula for calculating harmonic mean for grouped data will be as follows;

$$H = \frac{\sum f}{\sum f\left(\frac{1}{X}\right)}$$

**Where, $n = \sum f$. This is sometimes called the weighted geometric mean with weights $f_1, f_2, \ldots\ldots, f_k$.**

Example:

The harmonic mean of the frequency distribution of weights of students at a university, is calculated by using the following

| Weight (lb) | Frequency ($f$) | Class Mark (X) MID INTERVAL | 1/x | F*(1/x) |
|---|---|---|---|---|
| 110 – 119 | 1 | 114.5 | 0.008734 | 0.00873 |
| 120 – 129 | 4 | 124.5 | 0.008032 | 0.03212 |
| 130 – 139 | 17 | 134.5 | 0.007435 | 0.12639 |
| 140 – 149 | 28 | 144.5 | 0.00692 | 0.19377 |
| 150 – 159 | 25 | 154.5 | 0.006472 | 0.16181 |
| | $\sum f = 75$ | | | $\sum f*(1/x)=0.52282$ |

$$H = \frac{\sum f}{\sum f\left(\frac{1}{X}\right)}$$

***H=75/0.52282***
***=143.452***

# Homework :-

For the following shown below, Find

**Weighted Mean**

**Harmonic mean**

**Geometric mean**

| Interval | Frequency |
|---|---|
| 19----21 | 2 |
| 22----24 | 4 |
| 25----27 | 6 |
| 28----30 | 8 |
| 31----33 | 10 |
| 34----36 | 12 |