

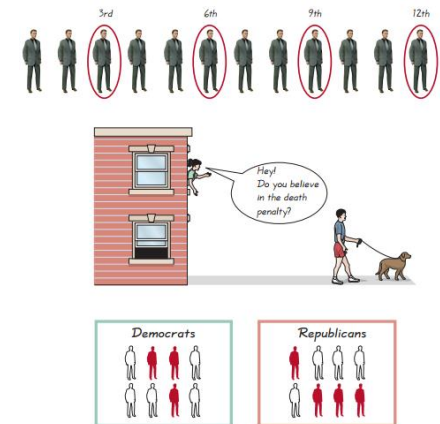
Title of Lecture

Biostatistics for Nursing

Lecturer:

Assistant Professor
Dr. Omiad Saber Abdullah Shwany
(PhD Statistics)

25-9-2021



Chapter One

Statistics

- Introduction:

The word "statistics" derived of the Latin word "status" or the Italian word "statista" or the German word "statistik" each of which means a "political state"

ان كلمة الاحصاء (Statistics) مشتقة من الكلمة اللاتينية "Status" أو من الكلمة الايطالية "statista" أو الكلمة الالمانية "statistik" وجميعها تعني فيما تعنيه حقائق ومعلومات عن الدولة

Political State

- Definitions:

Statistics: is concerned with scientific methods for collecting, organizing, summarizing, presenting, and analyzing data for specified phenomenon. As well as summarizing and interpreting the results and making reasonable decisions.

علم الاحصاء:- هو الطريقة العلمية التي تختص بجمع البيانات والحقائق عن ظاهرة او فرضية (ظواهر أو فرضيات) معينة وتنظيم وتبويب هذه البيانات والحقائق بالشكل الذي يسهل عملية تحليلها وتفسيرها ومن ثم استخلاص النتائج واتخاذ القرار على ضوء ذلك.

زانستی ئامار:- ئەو رېځگه زانستی يه که تايبەت مەندە به کوکردنه وهی داتا وراستی يه کان دەر باره ی دیاردهیهک یان گریمانیک (چەند دیاردهیهک یان چەند گریمانیک) ی دیاری کراو وەرپیک خستن وپۆلین کردنی ئەو زانیاریانە به شیوهیهک که بتوانریت به ئاسانی کردارەکه شی بکاته وه و تەفسیری بکات وه له ژیر روښنایی ئەنجامه کان بریاری له سەر بدریت.

Statistical method in scientific research:

1. Determine the problem.
2. Collecting the information and data for problem.
3. Organizing and presenting data in graphics and tabulations form.
4. Analysis of data.
5. Summarizing and interpreting the results.
6. Making decisions.

فيما يلي المراحل الرئيسية للطريقة الاحصائية في البحث العلمي:

1. تحديد مشكلة أو فرضية البحث أو الدراسة.
2. جمع البيانات والمعلومات عن ظاهرة أو الظواهر ذات العلاقة بالبحث أو الدراسة.
3. تصنيف البيانات وتبويبها وعرضها.
4. حساب المؤشرات الاحصائية كتقديرات لمعالم مجتمع البحث أو الدراسة.
5. تحليل معطيات الدراسة والتوصل للنتائج على ضوء فرضية أو فرضيات البحث أو الدراسة.
6. تفسير النتائج وعملية اتخاذ القرار بشأن فرضيات البحث.

Types of statistics:

1. Descriptive statistics: consists collecting, organizing, summarizing and presenting data.
2. Inferential statistics: consists generalizing from samples to populations. Consisting hypothesis testing, determining relationships among variables, and estimations.

وبشكل عام تنقسم الاحصاء الى فرعين رئيسيين هما:

1- الاحصاء الوصفي Descriptive statistics

2- الاحصاء الاستدلالي Inferential statistics

أسلوب تصميم البحوث :Research Design Method

تحديد الغرض من البحث

Determine the purpose of the research.

تحديد امكانية التنفيذ الفعلي للبحث

Determine the possibility of actual implementation of the research.

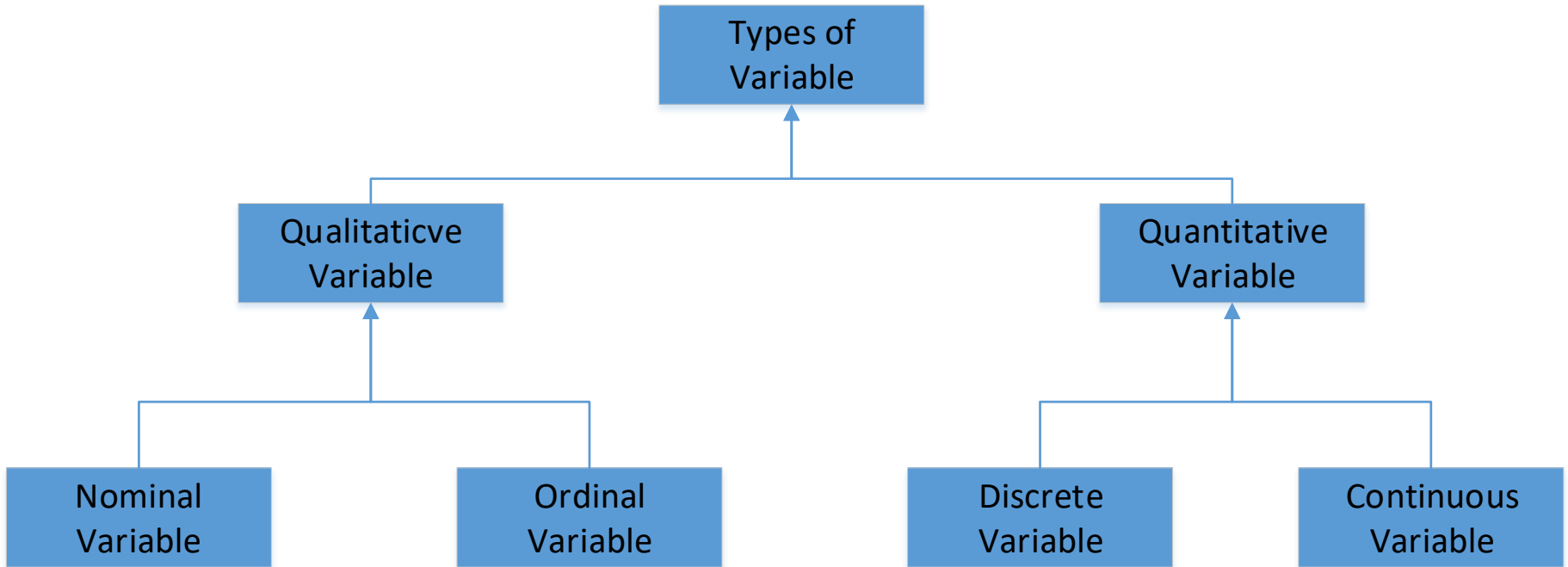
تحديد اطار البحث

Select the research frame

تحديد أسلوب جمع البيانات والمعلومات

Identify the method of data collection and information.

- **Data:** are a set of numbers. Those values represent by the thing called a variable, because the values vary from element to another. If the values are the same, the thing is called a constant.
- **Variable:** A Characteristic that varies from one element to another.
- **Random Variable:** A variable whose values are selected by chance.



Qualitative Variables: Variables which assume non-numerical values.

Example: gender (male or female).

Example: Course grades A, B, C, or D

Nominal Variable: There is no order or rank among characteristics (observations) of variable.

For example: gender (male or female).

Ordinal Variable: There is order or rank among characteristics of variable.

Example: Course grades A, B, C, or D

Example: Economic state (v. good, good, bad)

Quantitative Variables: Variables which assume numerical values.

For example: number of students

For example: age (54 , 65.6 , 43 ,...).

Discrete Variables: Variables which assume a finite or countable number of possible values. Usually obtained by counting.

For example: number of students

Continuous Variables: Variables which assume an infinite number of possible values. Usually obtained by measurement.

For example: age (54 , 65.6 , 43 , ...).

For example: height (154 , 165.6 , 143 , ...).

Or Variables which take real numbers.

$$-\infty < X < \infty$$

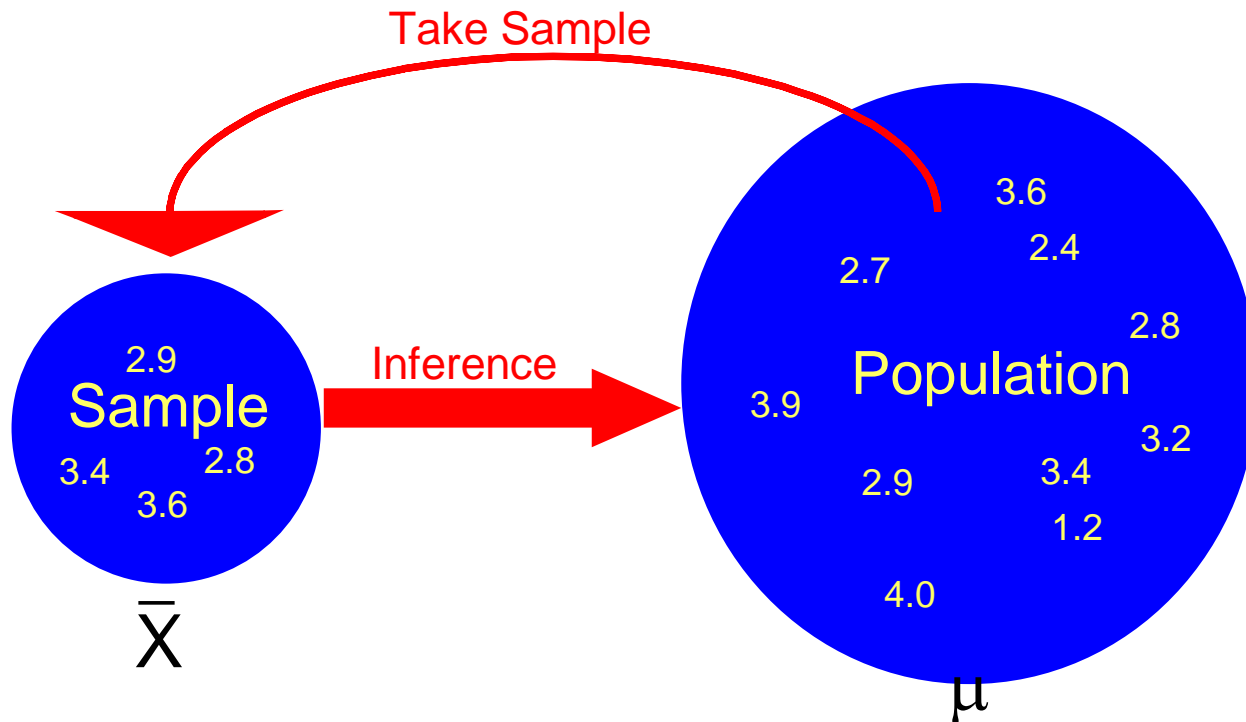
Sources of data:

1. Historical sources.
2. Field sources.

- Methods of Collection the data:

1. Census method: the collection of data from every elements in a population
2. Samples method: the collection of data from a set of elements in a population

- A population (Statistical population): is all the elements or values under a statistical study.
- Sample: is a part from the population.



- Homogenous population any population which every element has joint specified characteristics
- Non-homogenous population any population which every element hasn't joint specified characteristic

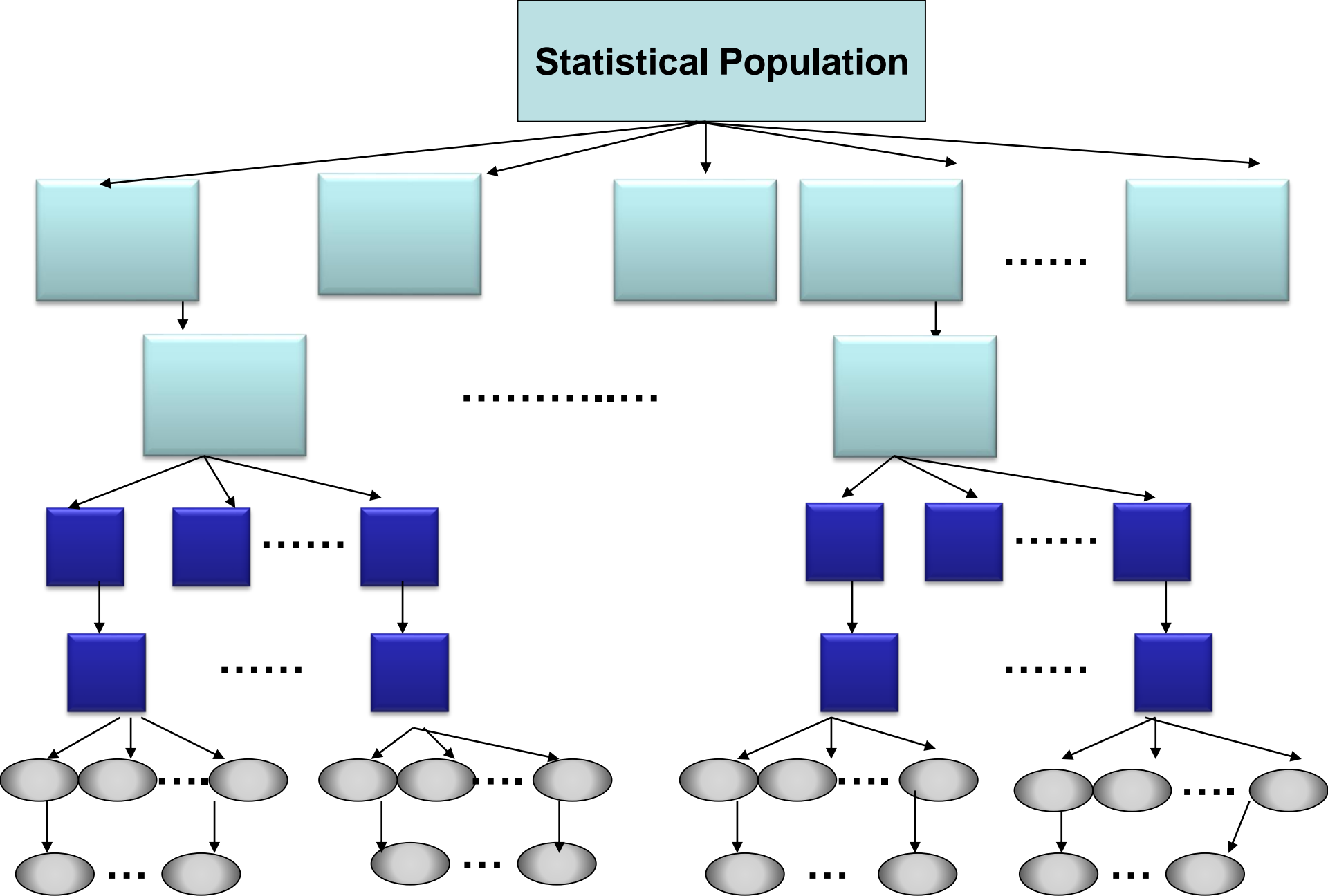
- Types of Samples:

1. Random Samples: A sample collected in such a way that every element in the population has the same chance of selection.
2. Non random Samples: A sample collected in such a way that every element in the population hasn't the same chance of selection.

- Types of Random Samples:
 1. Simple random sampling: is the simplest method of sampling. Uses these method when the population is homogenous. Every element in the population has an equal chance of selection. Some procedures uses for selecting a simple random sample:
 - a. Lottery method.
 - b. Random number table.

2. Stratified random sampling: Uses these method when the population is non-homogenous. The sample select by dividing the population into groups (strata) according to some characteristic and then taking samples from each group by using simple random sampling.
3. Systematic random sampling: are obtained by numbering each value in the population. Choose the first element randomly, then every k^{th} observation.
4. Multistage random sampling:

Statistical Population



- Types of Non-Random Samples:
 1. quota sample
 2. Purposive sample

Chapter Two

Data Presentation

- Introduction:

After collecting data, the next step should be express it in some form such as:

- 1- Frequency distribution (Tabular presentation)
- 2- Graphical presentation

- Ungrouped data: is the raw data which is collect and not grouped, such as: 20, 23, 21, 43, 32, 26.
- Grouped data: is the raw data which is collect and organizing in frequency table.

- Frequency distribution:
- Frequency Table: is a simple table has two columns, first column called classes and the second column called frequency.

Example (1)

Example (2)

Classes (gender)	frequency
male	20
female	30

Classes (age)	frequency
10 –	5
20 –	8
30 –	10
40 – 50	12

Frequency distribution for qualitative data :

Example: The following data represent eye's color for 10 students:

Blue , Brown , Black , Brown , Black , Brown , Black , Blue ,
Black , Black

Construct the frequency table.

Classes (eyes color)	Mark frequency	frequency
Blue	11	2
Brown	111	3
Black	1111	5
Total		10

Relative frequency distribution:

Relative frequency: The ratio of frequency of class to the total of frequencies.

$$f^* = \frac{f_i}{\sum_{i=1}^m f_i}$$

Example: Find relative frequency for this table

(eyes color)	Frequency (f_i)	Relative frequency (f^*)
Blue	2	2/10=0.2
Brown	3	3/10=0.3
Black	5	5/10=0.5
Total	10	1

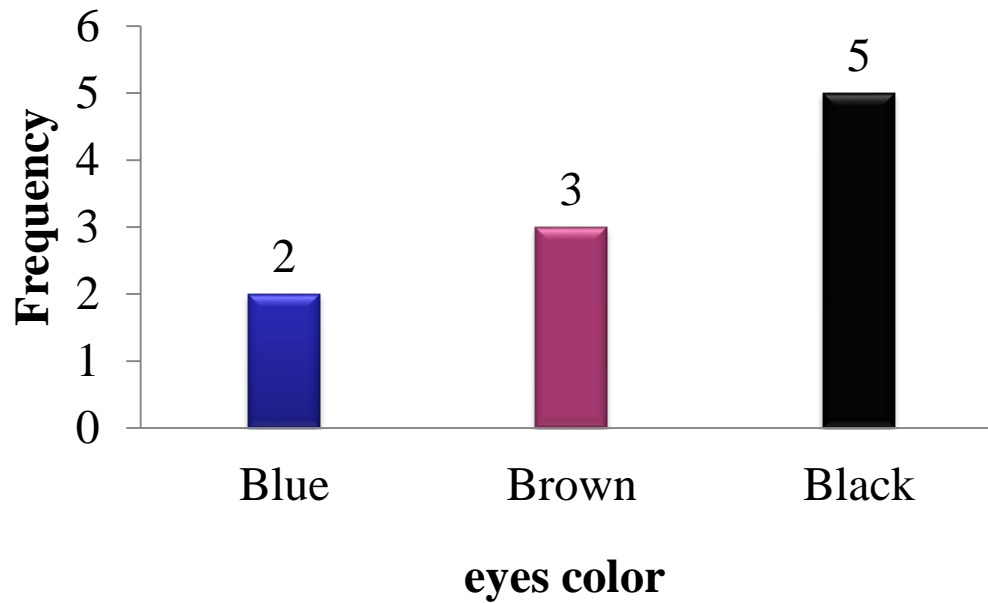
➤ Graphical presentation for qualitative data:

Graphical presentation: It is another method for presenting and summarizing grouped data

1- Bar chart: A graph that display the characteristics of variable on the horizontal axis and the frequency on the vertical axis.

Example (1):Present this data graphically (Bar chart):

Classes (eyes color)	frequency
Blue	2
Brown	3
Black	5
Total	10



2- Pie chart: is a circle divided to number of sectors which are representing characteristics of variable.

The angle of each sector

$$= \frac{f_i}{\sum_{i=1}^m f_i} * 360$$

Example (1): Present this data in Pie chart:

Classes (eyes color)	frequency
Blue	2
Brown	3
Black	5
Total	10

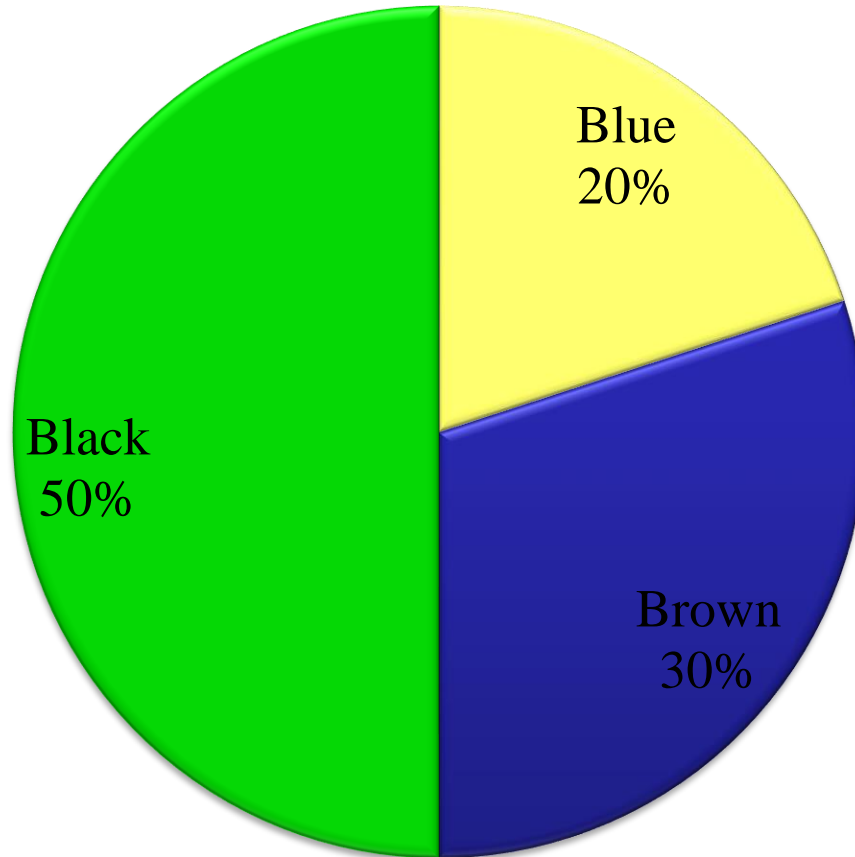
The angle of each sector $= \frac{f_i}{\sum_{i=1}^m f_i} * 360$

The angle of Blue = $2/10 * 360 = 72$

The angle of Brown = $3/10 * 360 = 108$

The angle of Black = $5/10 * 360 = 180$

Eyes Color

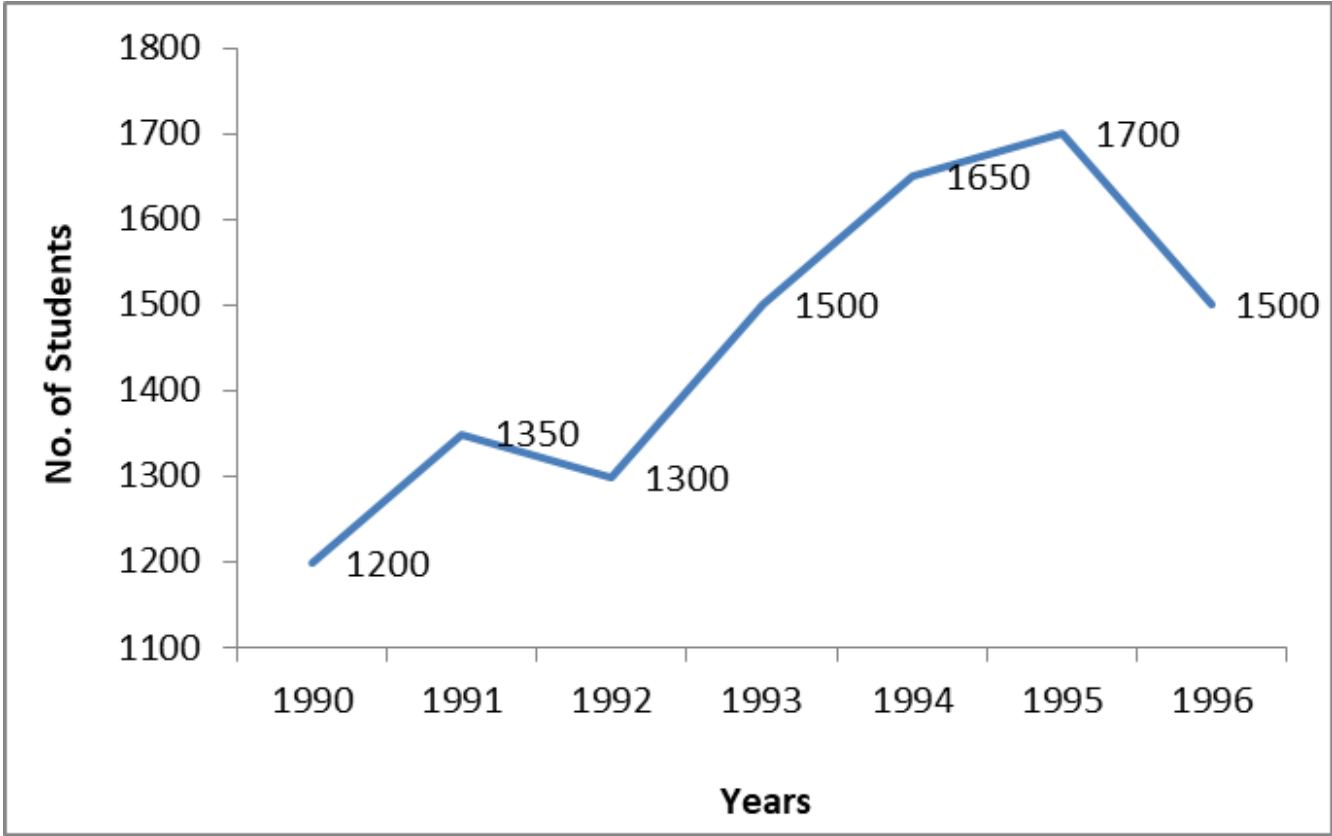


3- Line Chart: is a plot of the variable over time. It is using often for time series data. It is created by plotting the value of the variable on the vertical axis and the time periods on the horizontal axis.

Example: The following data represent number of students in Administration college

Year	1990	1991	1992	1993	1994	1995	1996
No. of student	1200	1350	1300	1500	1650	1700	1500

Present these data by line chart



Frequency distribution for quantitative data :
General rules for constructing frequency table

1- Find the Total Range

$$T.R = X_L - X_S + 1$$

2- Find the number of classes

$$m = 2.5 * \sqrt[4]{n}$$

3- find the length of classes

$$L = \frac{T.R}{m}$$

4- Writing class limits

a- For continuous variable

Lower Limit	Upper Limit less than
X_S	$X_S + L$
$X_S + L$	$X_S + 2L$
$X_S + 2L$	$X_S + 3L$
.	.
.	.
.	.
$X_S + (m-1)L$	$X_S + mL$

b- For discrete variable

Lower Limit	Upper Limit
X_S	$X_S + L - 1$
$X_S + L$	$X_S + 2L - 1$
$X_S + 2L$	$X_S + 3L - 1$
· · ·	· · ·
$X_S + (m-1)L$	$X_S + mL - 1$

5- Find frequencies of any classes

Example: The following data represent weights (30) children from a small town

20 , 10 , 8.6 , 7 , 5.9 , 6 , 12 , 8 , 14 , **5**

9 , 16.5 , 25 , 16 , 19 , 20.9 , 7.8 , 15 , 18 , 24

9 , 7 , 15 , 16.9 , 11 , 8 , **26** , 17.6 , 21 , 10

Construct the frequency table

1- Find the range

$$R = X_L - X_S + 1 = 26 - 5 + 1 = 22$$

2- Find the number of classes

$$m = 2.5 * \sqrt[4]{n}$$

$$m = 2.5 * \sqrt[4]{30} = 5.85 \cong 6$$

3- find the length of classes

$$L = \frac{T.R.}{m} = \frac{22}{6} = 3.66 \cong 4$$

4- Writing class limits

5- Find frequencies of any classes

Classes	Mark frequency	frequency
5 –	1111 1111	9
9 –	1111 1	6
13 –	1111 1	6
17 –	1111	5
21 –	11	2
25 – 29	11	2

Example (2): The following data represent persons number of family in a small town

3 , 5 , 8 , 10 , 12 , **2** , 7 , 5 , 11 , 9

5 , 6 , 13 , **15** , 12 , 4 , 5 , 4 , 3 , 5

Construct the frequency table

1- Find the range

$$T.R = X_L - X_S + 1 = 15 - 2 + 1 = 14$$

2- Find the number of classes

$$m = 2.5 * \sqrt[4]{n}$$

$$m = 2.5 * \sqrt[4]{20} = 5.2 \cong 5$$

3- find the length of classes

$$L = \frac{T.R}{m} = \frac{14}{5} = 2.8 \cong 3$$

4- Writing class limits

5- Find frequencies of any classes

Classes	Mark frequency	frequency
2 – 4	1111	5
5 – 7	1111 11	7
8 – 10	111	3
11 – 13	1111	4
14 – 16	1	1
Total		20

Center of class (Mid point): is the arithmetic mean for lower limit (L.L.) and upper limit (U.L.) of each class

$$X_i = \frac{L.L. + U.L.}{2}$$

Example: From the following frequency table: Find:

- 1- Relative frequency (f^*)
- 2- Percentage Relative frequency ($f^*\%$)
- 3- Mid points of the classes (X_i)

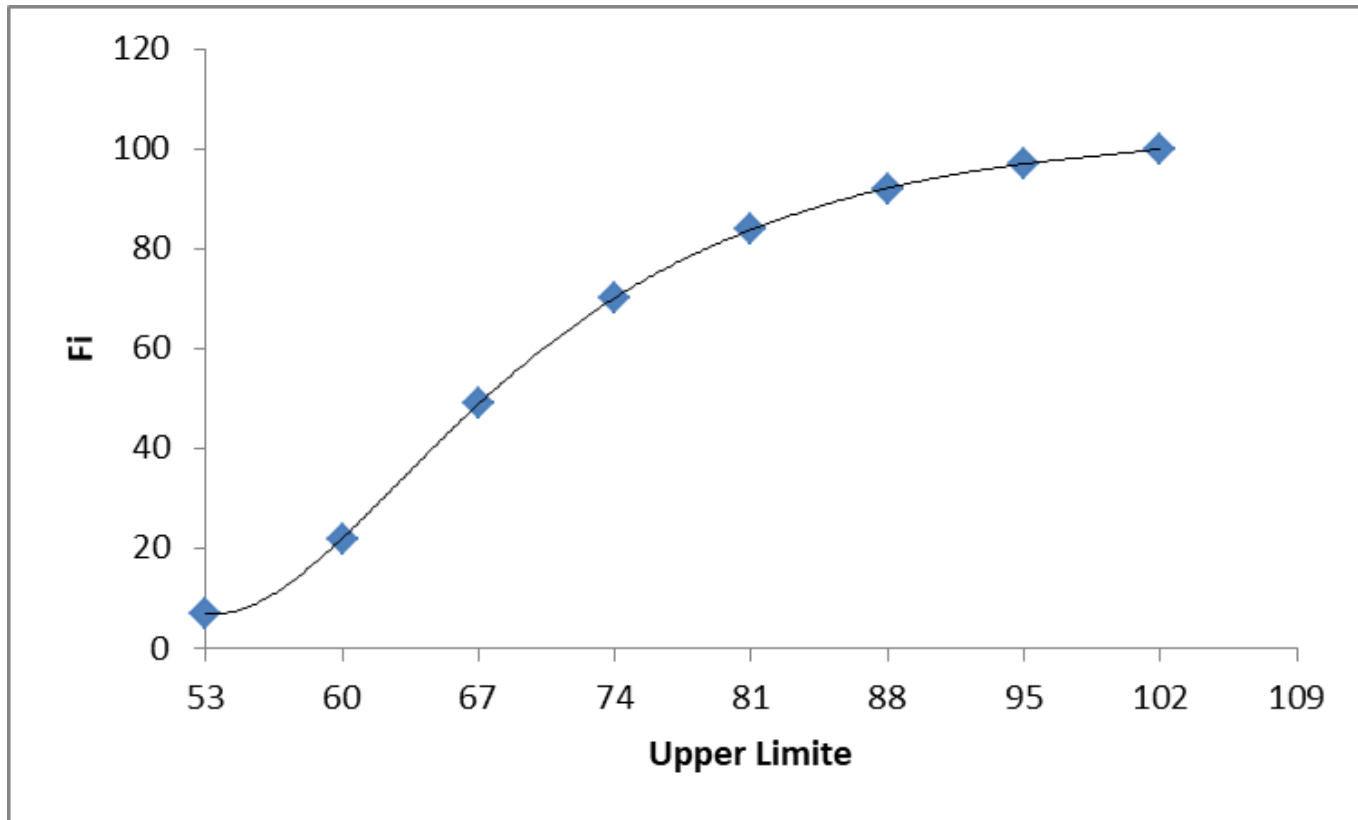
Classes	f_i	f^*	$f^*\%$	X_i
5 –	9	0.3	$0.3 * 100 = 30$	$(5+9)/2 = 7$
9 –	6	0.2	$0.2 * 100 = 20$	$(9+13)/2 = 11$
13 –	5	0.167	16.666	15
17 –	6	0.2	20	19
21 –	2	0.067	6.666	23
25 – 29	2	0.067	6.666	27
Total	30	1	100	

- Cumulative frequency distribution :
 - 1- Rising cumulative frequency distribution: is the distribution which displays frequencies cumulative which are computing it based on upper limits of classes.

Example: From the following frequency table: Find rising cumulative frequency distribution (F_i)

For continuous variable

Classes	f_i	Upper limits	F_i
46 –	7	53	7
53 –	15	60	22
60 –	27	67	49
67 –	21	74	70
74 –	14	81	84
81 –	8	88	92
88 –	5	95	97
95 – 102	3	102	100
Total	100		

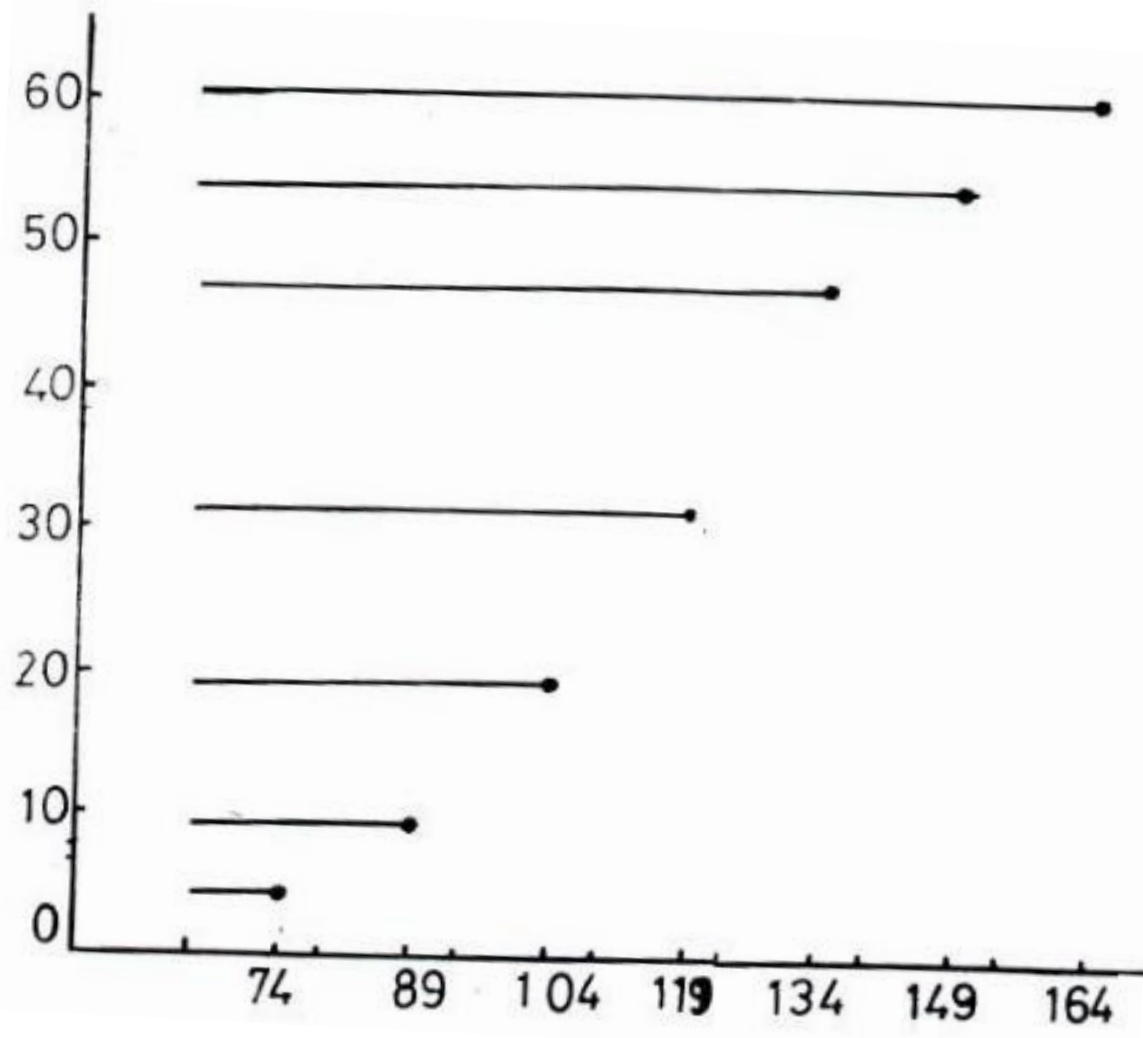


Example: From the following frequency table: Find rising cumulative frequency distribution (F_i)

For discrete variable

Classes	f_i	Upper limits	F_i
60 – 74	4	74	4
75 – 89	5	89	9
90 – 104	10	104	19
105 – 119	12	119	31
120 – 134	16	134	47
135 – 149	7	149	54
150 – 167	6	167	60
Total	60		

F_i



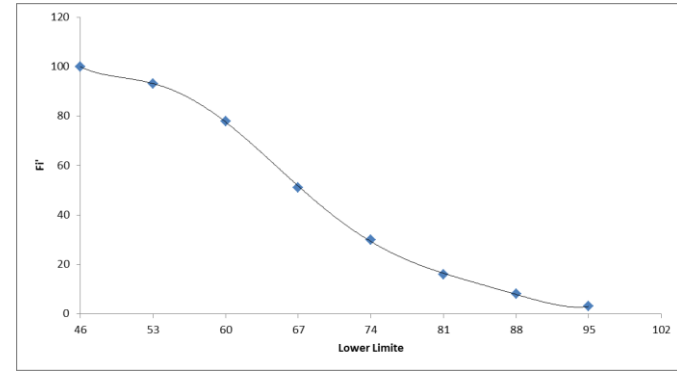
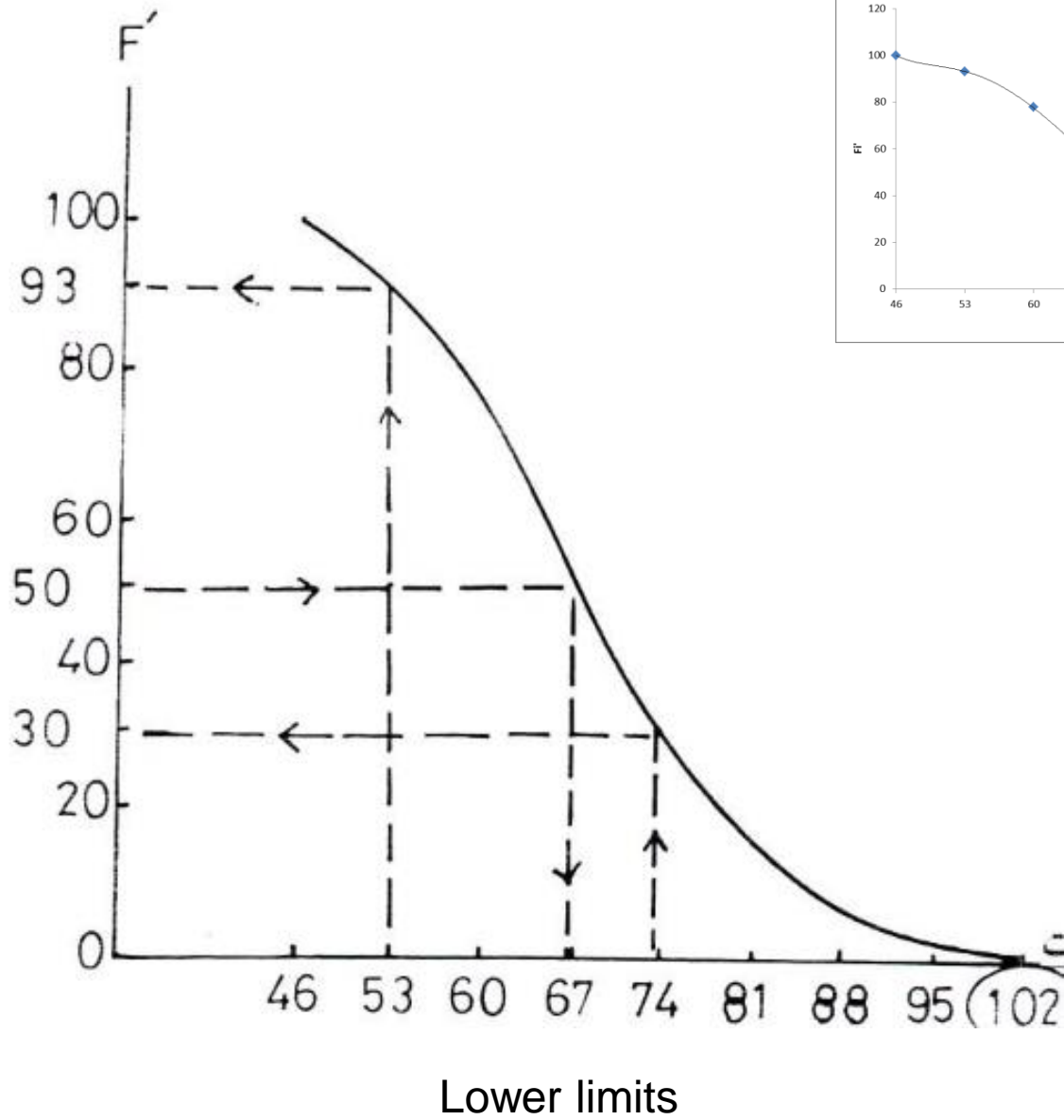
Upper limits

2- Descending cumulative frequency distribution: is the distribution which displays frequencies decrement which are computing it based on lower limits of classes.

Example: From the following frequency table: Find Descending cumulative frequency distribution (F_i')

For continuous variable

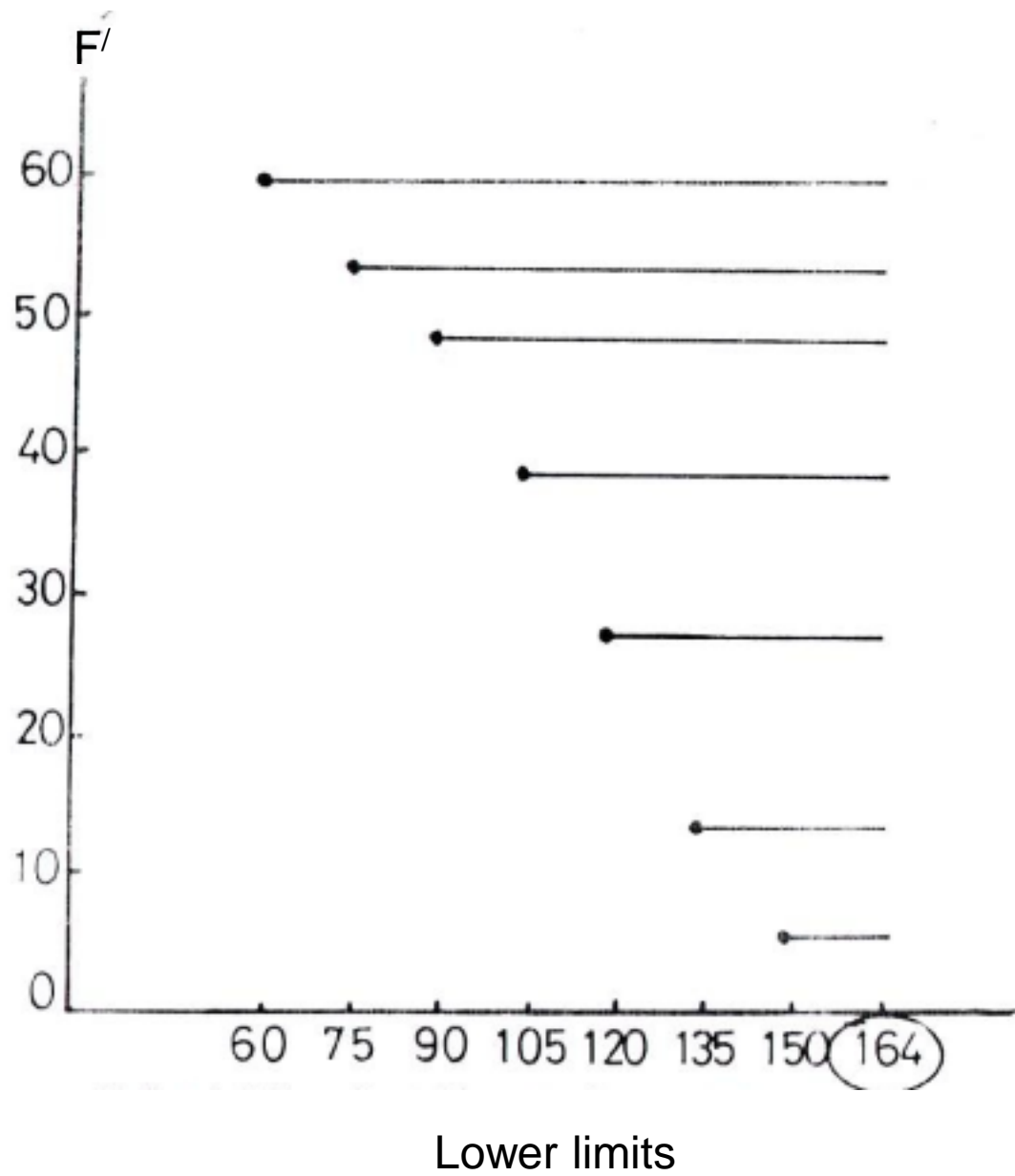
Classes	f_i	Lower limits	F_i'
46 –	7	46	100
53 –	15	53	93
60 –	27	60	78
67 –	21	67	51
74 –	14	74	30
81 –	8	81	16
88 –	5	88	8
95 – 102	3	95	3
Total	100		



Example: From the following frequency table: Find
Descending cumulative frequency distribution (F_i)

For discrete variable

Classes	f_i	Lower limits	F_i'
60 – 74	4	60	60
75 – 89	5	75	56
90 – 104	10	90	51
105 – 119	12	105	41
120 – 134	16	120	29
135 – 149	7	135	13
150 – 164	6	150	6
Total	60		



➤ Graphical presentation for Quantitative data:

1- Histogram

Example (1): Present this data graphically (Histogram):

Classes	f_i
2 – 4	5
5 – 7	7
8 – 10	3
11 – 13	4
14 – 16	1
Total	20

1- Histogram

Example (2): Present this data graphically (Histogram):

Classes	f_i
5 –	9
9 –	6
13 –	5
17 –	6
21 –	2
25 – 29	2
Total	30

➤ Graphical presentation for Quantitative data:

2- Polygon

Example (1): Present this data graphically (Polygon):

Classes	f_i
2 – 4	2
5 – 7	5
8 – 10	8
11 – 13	4
14 – 16	1
Total	20

2- Polygon

Example (2): Present this data graphically (Polygon):

Classes	f_i
5 –	9
9 –	6
13 –	5
17 –	6
21 –	2
25 – 29	2
Total	30

➤ Graphical presentation for Quantitative data:

3- Curve

Example (1): Present this data graphically (Curve):

Classes	f_i
2 – 4	5
5 – 7	7
8 – 10	3
11 – 13	4
14 – 16	1
Total	20

3- Curve

Example (2): Present this data graphically (Curve):

Classes	f_i
5 –	9
9 –	6
13 –	5
17 –	6
21 –	2
25 – 29	2
Total	30

Chapter Three

**Sigma Notation
(Summation Notation)** Σ

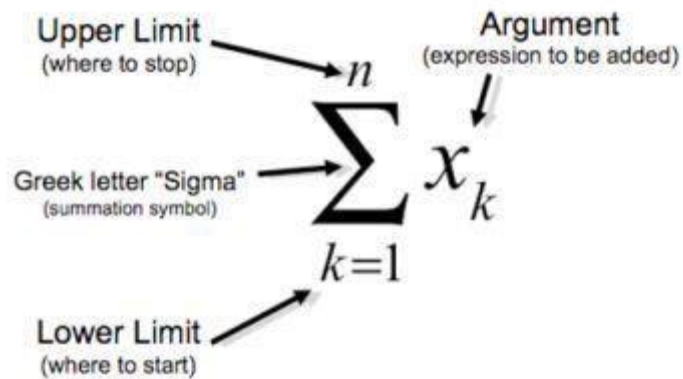
and

Pi Notation Π

1) Sigma Notation (Summation Notation) Σ

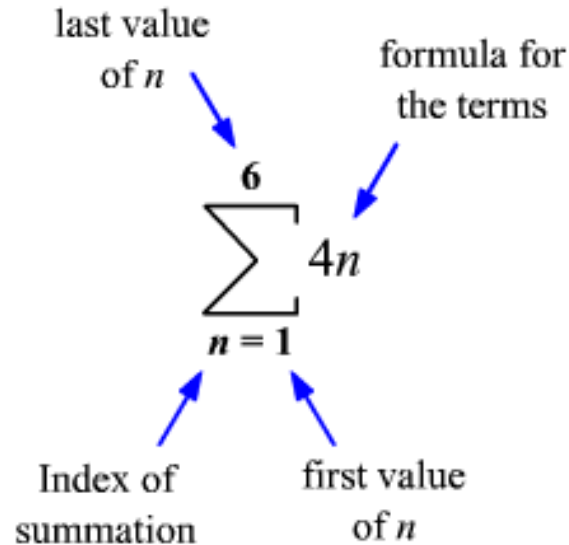
SIGMA NOTATION

(represents the sum of a sequence)



Sum of the values of X from X_1 through X_n

$$\sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n$$



Sigma Notation (Summation Notation) :

$$(1) \quad \sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n$$

$$(2) \quad \sum_{i=1}^n X_i^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

$$(3) \quad \left(\sum_{i=1}^n X_i \right)^2 = (X_1 + X_2 + \dots + X_n)^2$$

$$(4) \quad \sum_{i=1}^n \frac{1}{X_i} = \frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}$$

$$(5) \quad \frac{1}{\sum_{i=1}^n X_i} = \frac{1}{X_1 + X_2 + \dots + X_n}$$

Sigma Notation (Summation Notation) :

$$(6) \quad \sum_{i=1}^n \log_{10} X_i = \log_{10} X_1 + \log_{10} X_2 + \dots + \log_{10} X_n$$

$$(7) \quad \sqrt{\sum_{i=1}^n X_i} = \sqrt{X_1 + X_2 + \dots + X_n}$$

$$(8) \quad \sum_{i=1}^n \sqrt{X_i} = \sqrt{X_1} + \sqrt{X_2} + \dots + \sqrt{X_n}$$

Properties of Sigma Notation (Summation Notation) :

If X_i and Y_i are two Variables, and a, b are two constants n sample size then:

$$(1) \quad \sum_{i=1}^n a = na$$

$$(2) \quad \sum_{i=1}^n aX_i = a \sum_{i=1}^n X_i$$

$$(3) \quad \sum_{i=1}^n (X_i \pm a) = \sum_{i=1}^n X_i \pm na$$

$$(4) \quad \sum_{i=1}^n (aX_i \pm bY_i) = a \sum_{i=1}^n X_i \pm b \sum_{i=1}^n Y_i$$

$$(5) \quad \sum_{i=1}^n X_i = \sum_{i=1}^k X_i + \sum_{i=k+1}^n X_i \quad k < n$$

$$(6) \quad \sum_{i=1}^n X_i Y_i = X_1 Y_1 + X_2 Y_2 + \dots + X_n Y_n$$

Example// If we have the following data: $X_i: 2,4,5,8,3$ $Y_i: 1,2,3,2,3$ Find:

$$(1) \sum_{i=1}^n X_i$$

$$(2) \sum_{i=1}^n Y_i^2$$

$$(3) \sum_{i=1}^n 3X_i$$

$$(4) \sum_{i=1}^n (X_i - 4)$$

$$(5) \left(\sum_{i=1}^n X_i \right)^2$$

$$(6) \sum_{i=1}^n \frac{1}{X_i}$$

$$(7) \sum_{i=1}^n \sqrt{X_i}$$

$$(8) \sum_{i=1}^n \log_{10} X_i$$

$$(9) \sum_{i=1}^n X_i Y_i$$

$$(10) \sum_{i=1}^n (X_i - Y_i)$$

$$(11) \sum_{i=1}^n X_i Y_i^2$$

$$(12) \sum_{i=1}^n (X_i - 4)(Y_i - 3)$$

Pi Notation

π

$$\prod_{i=1}^n X_i = X_1 * X_2 * \dots * X_n$$



In Excel we can use the function :

=PRODUCT(A1:A10)

Properties of Pie Notation:

If X_i and Y_i are two Variables, and a, b are two constants n sample size then:

$$(1) \prod_{i=1}^n a = a^n$$

$$(6) \prod_{i=1}^n \sqrt{X_i} = \sqrt{\prod_{i=1}^n X_i}$$

$$(2) \prod_{i=1}^n aX_i = a^n \prod_{i=1}^n X_i$$

$$(7) \prod_{i=1}^n X_i^a = \left(\prod_{i=1}^n X_i \right)^a$$

$$(3) \prod_{i=1}^n abX_iY_i = (ab)^n \prod_{i=1}^n X_i \prod_{i=1}^n Y_i$$

$$(8) \log_{10} \prod_{i=1}^n X_i^{ai} = \sum_{i=1}^n ai \log_{10}(X_i)$$

$$(4) \prod_{i=1}^n \frac{1}{X_i} = \frac{1}{\prod_{i=1}^n X_i}$$

$$(9) \prod_{i=1}^n \frac{Y_i}{X_i} = \frac{\prod_{i=1}^n Y_i}{\prod_{i=1}^n X_i}$$

$$(5) \log_{10} \prod_{i=1}^n X_i = \sum_{i=1}^n \log_{10}(X_i)$$

Example//

If we have the following data: X_i : 1,2,3,2,4,3,5,3 Y_i : 2,1,3,3,5,2,4,2 Find:

$$(1) \prod_{i=1}^4 X_i = 12$$

$$(2) \prod_{i=2}^5 Y_i = 45$$

$$(3) \prod_{i=1}^4 5X_i = 7500$$

$$(4) \prod_{i=1}^7 -9Y_i = -3443737680$$

$$(5) \prod_{i=1}^8 X_i^{-1} = 0.000462962962962963$$

$$(6) \prod_{i=1}^n Y_i^{-2} = 0.000000482253086419753$$

$$(7) \prod_{i=1}^3 2X_i Y_i = 288$$

$$(8) \prod_{i=1}^5 (Y_i + 2) = 2100$$

$$(9) \prod_{i=1}^n \sqrt{3Y_i} = 3073.734$$

$$(10) \prod_{i=1}^n X_i^2 = 4665600$$

$$(11) \prod_{h=1}^3 h^2 = 36$$

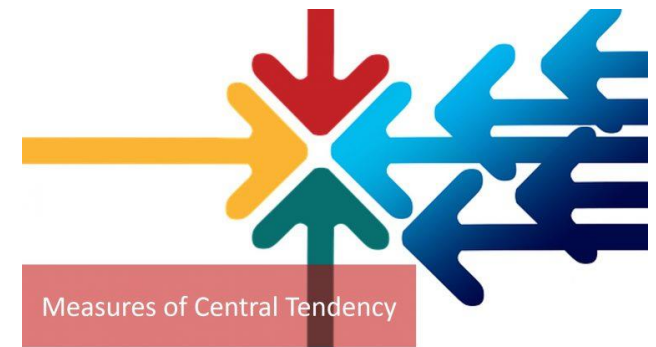
$$(12) \prod_{h=1}^5 \sum_{t=1}^6 ht^2 = 748838574120$$

Chapter Four

Measures of Central Tendency

A measure of central tendency is a single value that attempts (محاولات) 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. They are also classed as summary statistics. The mean (often called the average) is most likely the measure of central tendency that you are most familiar with, but there are others, such as the median and the mode.

The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. In the following sections, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.



1) Mean (Arithmetic Mean)

The mean (or average) is the most popular and well known measure of central tendency.

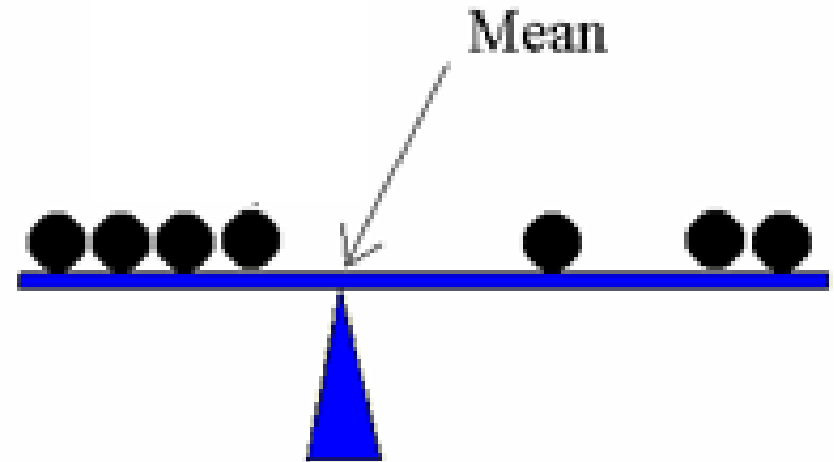
Sample

$$Mean = \bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

For ungrouped data

$$Mean = \bar{X} = \frac{\sum_{i=1}^m f_i X_i}{\sum_{i=1}^m f_i}$$

For grouped data



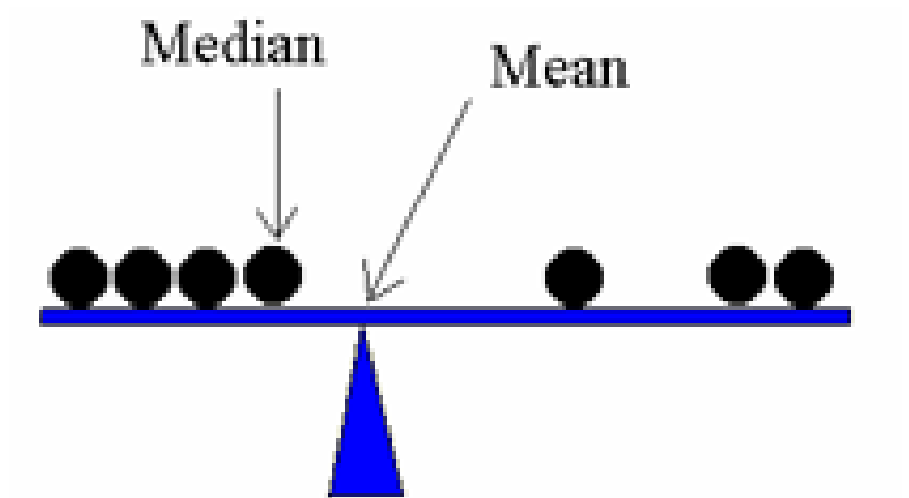
Weighted mean

Weighted mean is calculated when certain values in a data set are more important than the others. A weight w_i is attached to each of the values x_i to reflect this importance.

$$\text{Weighted Mean} = \bar{X}_W = \frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i}$$

2) Median

The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by outliers and skewed data. In order to calculate the median, suppose we have the data below:



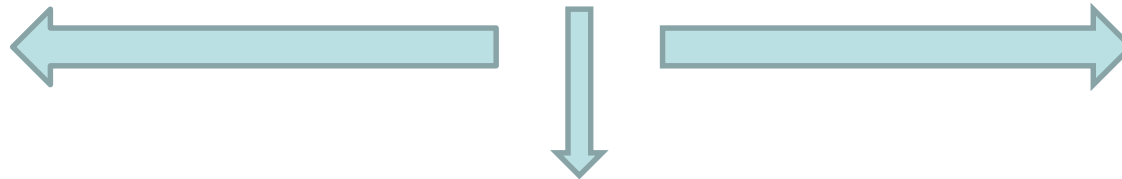
1- Ungrouped Data

Example1// from the following data find the median? **n is odd**

X_i	4	8	5	2	6	3	8
-------	---	---	---	---	---	---	---

Solution// the first stapes we sort or order the data from smallest to largest or from largest to smallest :

X_i	2	3	4	5	6	8	8
-------	---	---	---	---	---	---	---



The Median = 5 Unite


1- Ungrouped Data

Example2// from the following data find the median? **n is even**

X_i	4	8	5	6	3	8
-------	---	---	---	---	---	---

Solution// the first stapes we sort or order the data from smallest to largest or from largest to smallest :

X_i	3	4	5	6	8	8
-------	---	---	---	---	---	---



$$\text{The Median} = M_e = \frac{5+6}{2} = 5.5 \text{ Unite}$$

2- Grouped Data

Example 1 Discrete data// from the following information find the Median:

Classes	frequency	F_i
2-4	6	6
5-7	9	15
8-10	12	27
11-13	20	47
14-16	14	61
17-19	11	72
20-22	8	80
Total	80	

$$F_{k-1} < \frac{\sum_{i=1}^m f_i}{2} \leq F_k$$

$$F_{k-1} < \frac{\sum_{i=1}^m f_i}{2} \leq F_k$$

$$27 < \frac{80}{2} \leq 47$$

$$27 < 40 \leq 47 \quad \text{medianClass} = 4$$

Then:

$$\text{Median} = \frac{11+13}{2} = 12 \text{Unite}$$

Classes	f_i	X_i	F_i
51-55	6	53	6
56-60	25	58	31
61-65	7	63	38
66-70	4	68	42
71-75	3	73	45
76-80	2	78	47
81-85	1	83	48
	48		

2- Grouped Data

Example 2 Continuous data// from the following information find the Median:

Classes	frequency	F_i
10-	2	2
20-	4	6
30-	9	15
40-	20	35
50-	30	65
60-	11	76
70-80	6	82
Total	82	

$$F_{k-1} < \frac{\sum_{i=1}^m f_i}{2} \leq F_k$$

$$35 < \frac{82}{2} \leq 65$$

$$35 < 41 \leq 65$$

$medianClass = 5$

Then:

$$Median = M_e = L_k + \frac{h_k}{f_k} \left(\frac{\sum_{i=1}^m f_i}{2} - F_{k-1} \right)$$

$$M_e = L_5 + \frac{h_5}{f_5} \left(\frac{\sum_{i=1}^7 f_i}{2} - F_{5-1} \right)$$

Lower Limit of Median Class

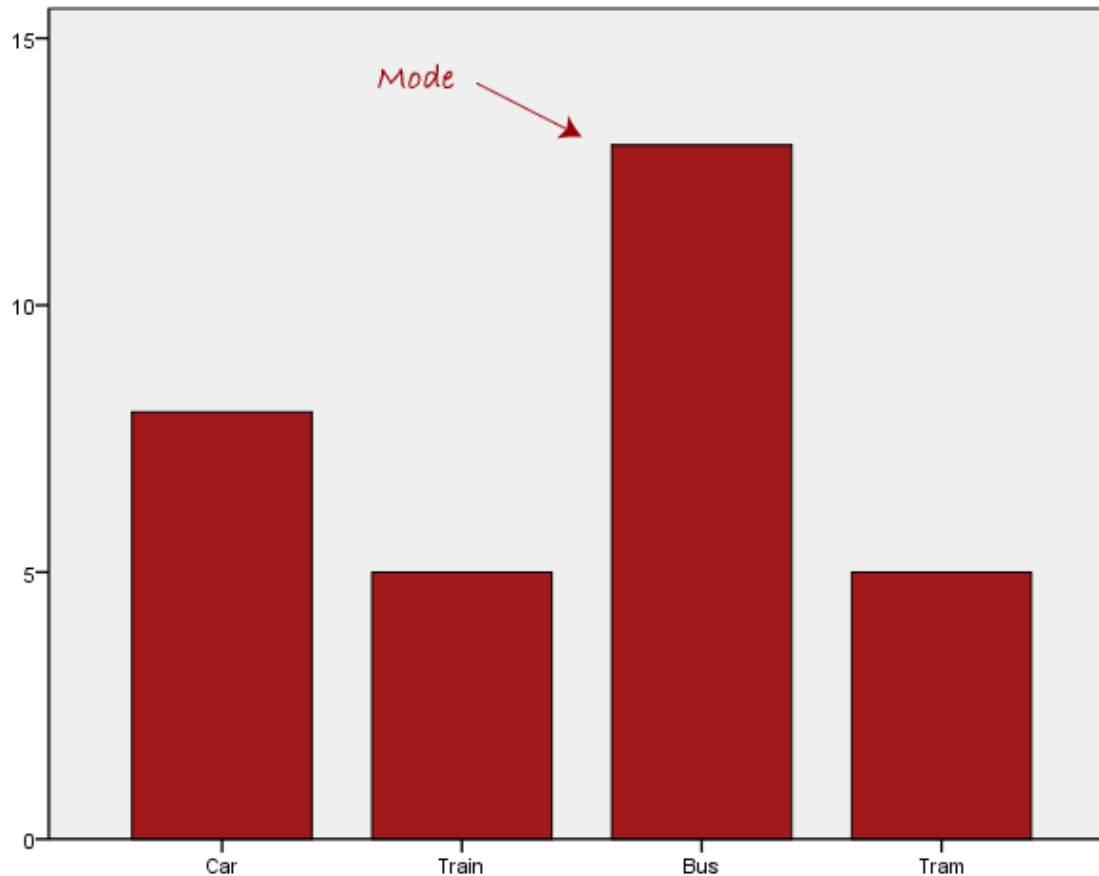
Length of Median Class

Frequency of Median Class

$$Median = M_e = 50 + \frac{10}{30} \left(\frac{82}{2} - 35 \right) = 52 \text{ Unite}$$

3) Mode

The mode is the most frequent score in our data set. On a histogram it represents the highest bar in a bar chart or histogram. You can, therefore, sometimes consider the mode as being the most popular option. An example of a mode is presented below:



Example// from the following information find the Mode:

Classes	fi
2-4	3
5-7	7
8-10	14
11-13	8
14-16	1

$$M_o = \frac{8+10}{2} = 9 \text{ Unit}$$

Example// from the following information find the Mode:

Classes	fi
50-	55
60-	85
70-	71
80-	15
90-100	4

$$M_o = L_k + \frac{(f_k - f_{k-1})}{(f_k - f_{k-1}) + (f_k - f_{k+1})} \times h_k$$

$$M_o = 60 + \frac{(85 - 55)}{(85 - 55) + (85 - 71)} \times 10 = 66.82 \text{ Marks}$$

Example// Find the Mode by the Chart:

type of Variable	Best measure of central tendency
Nominal	Mode
Ordinal	Median
Interval/Ratio (not skewed)	Mean
Interval/Ratio (skewed)	Median

4) Geometric Mean

It is defined as the arithmetic mean of the values taken on a log scale. It is also expressed as the n^{th} root of the product of an observation.

$$\text{Geometric Mean} = G = \sqrt[n]{\prod_{i=1}^n X_i}, \quad X_i > 0$$

• قليل التآثر بالقيم الشاذة والمتطرفة.

• يشمل جميع قيم المتغير، مما يجعله من المقاييس التي تعبر بشكل جد عن كل البيانات.

• في معدلات النمو والتطور، وجميع المعدلات التي تميل إلى التغير بنسب ثابتة.

هو أحد أشكال المتوسطات، ويستخدم بكثرة في دراسة المعدلات التي تميل إلى الزيادة بنسب ثابتة، كما يستخدم في دراسة السلاسل الزمنية القائمة على مقارنة معدل ما واختلافاته باختلاف الزمن.

5) Harmonic mean

It is the reciprocal of the arithmetic mean of the observations.

$$\text{Harmonic Mean} = H = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}, X_i \neq 0$$

يفضل استخدامه على باقي المتوسطات في حالة إيجاد معدل السرعات ومعدلات التغيير

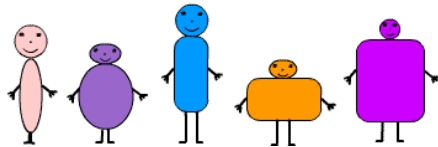
6) Quadratic mean

$$\text{Quadratic Mean} = Q = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n}} \quad , X_i \geq 0$$

Chapter Five

Measures of Variation

Variation is a way to show **how data is dispersed, or spread out**. Several measures of variation are used in statistics.



1) The Range

$$R = X_L - X_S$$

Example// from the following data find the range:

X_i : 3, 5, 8, 9, 10, 22, 15, 14, 1, 13

$$R = X_L - X_S$$

$$R = 22 - 1 = 21 \text{Unit}$$

2- Variance

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad , n > 30$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad , n \leq 30$$

3- Standard deviation

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} \quad , n > 30$$

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad , n \leq 30$$

Example// from the following data find (standard deviation and the Variance):

X_i
6
13
25
42
41
74
52
19
Mean=34

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
6	$(6-34)=-28$	784
13	-21	441
25	-9	81
42	8	64
41	7	49
74	40	1600
52	18	324
19	-15	225
Total		3568

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{(6-34)^2 + (13-34)^2 + \dots + (19-34)^2}{8-1}} = \sqrt{\frac{3568}{7}} = 22.577 \text{ Unit}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{(6-34)^2 + (13-34)^2 + \dots + (19-34)^2}{8-1} = \frac{3568}{7} = 509.714 \text{ Unit}^2$$

What is the Coefficient of Variation?

How to Find the Coefficient of Variation

$$\text{Coefficient of Variation} = \frac{\sigma}{\mu} \times 100\%$$

Where:

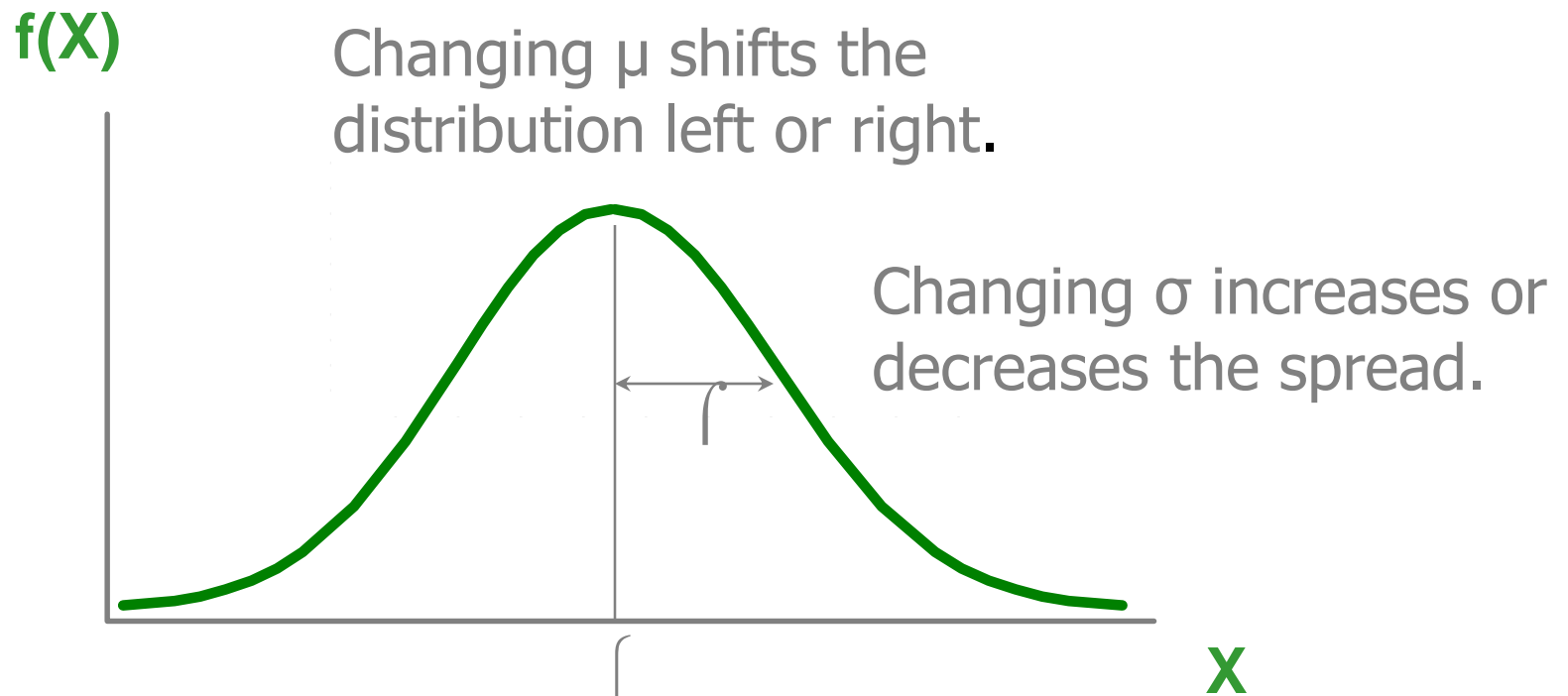
- σ - the standard deviation
- μ - the mean

$$CV = \frac{S}{\bar{X}} \times 100$$

Chapter Six: Standard Normal Distribution and Standard Score (z-score)

Examples of continuous probability distributions:

- The normal and standard normal



The Normal Distribution: as mathematical function (pdf)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Note constants:

$\pi=3.14159$

$e=2.71828$

This is a bell shaped curve with different centers and spreads depending on μ and σ

The Normal PDF

It's a probability function, so no matter what the values of μ and σ , must integrate to 1!

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

Normal distribution is defined by its mean and standard dev.

$$E(X)=\mu$$

$$\text{Var}(X)=\sigma^2$$

$$\text{Standard Deviation}(X)=\sigma$$

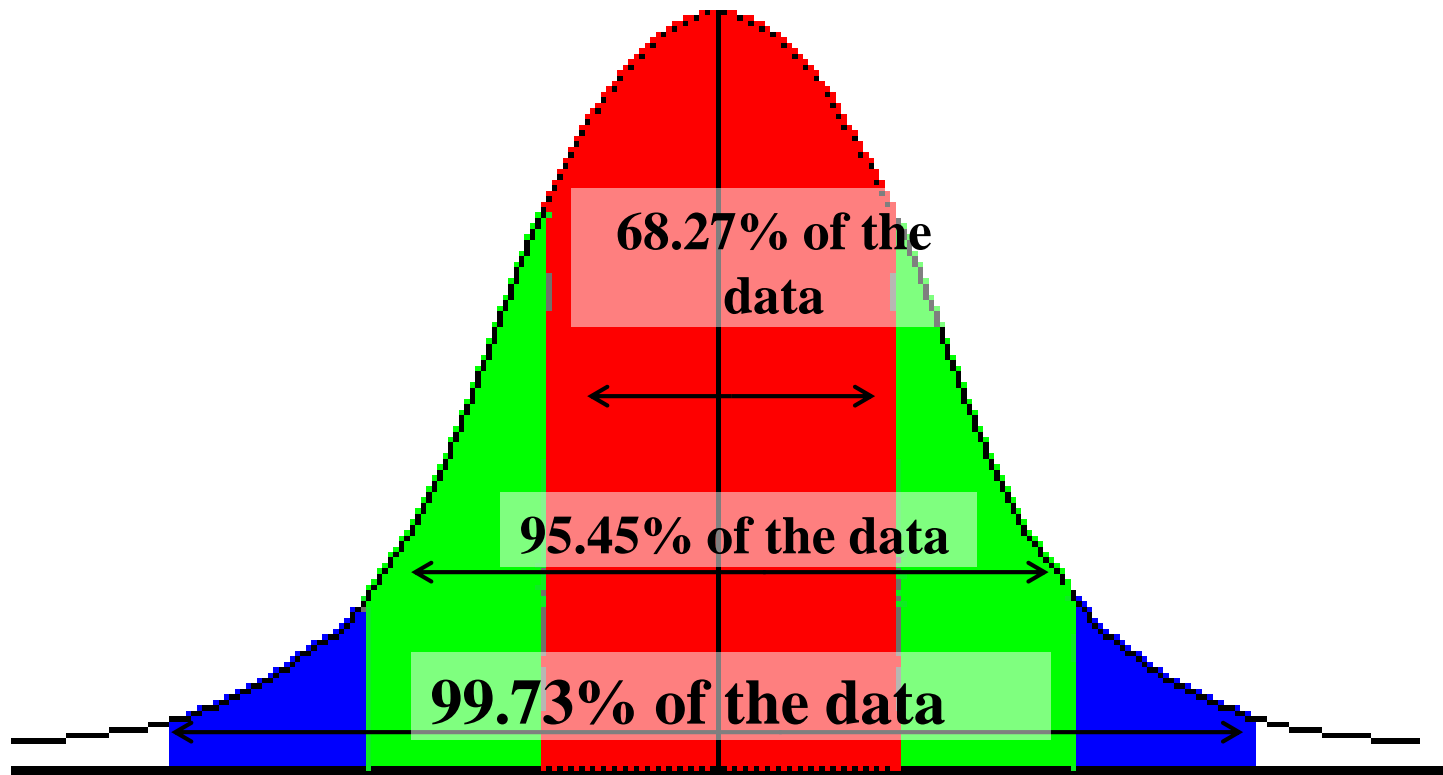
$$\int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$\int_{-\infty}^{+\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx) - \mu^2$$

**The beauty of the normal curve:

No matter what μ and σ are, the area between $\mu - \sigma$ and $\mu + \sigma$ is about 68.27%; the area between $\mu - 2\sigma$ and $\mu + 2\sigma$ is about 95.45%; and the area between $\mu - 3\sigma$ and $\mu + 3\sigma$ is about 99.73%. Almost all values fall within 3 standard deviations.

68.27 - 95.45 - 99.73 Rule



$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0.6827$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0.9545$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 0.9973$$

How good is rule for real data?

Check some example data: •

The mean of the weight of the women = 127.8 •

The standard deviation (SD) = 15.5 •

The Standard Normal (Z): “Universal Currency”

The formula for the standardized normal probability density function is

$$p(Z) = \frac{1}{(1)\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{Z-0}{1}\right)^2} = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(Z)^2}$$

The Standard Normal Distribution (Z)

All normal distributions can be converted into the standard normal curve by subtracting the mean and dividing by the standard deviation:

$$Z = \frac{X - \mu}{\sigma}$$

Somebody calculated all the integrals for the standard normal and put them in a table! So we never have to integrate!

Even better, computers now do all the integration.

Example

- For example: What's the probability of getting a math SAT score of 575 or less, $\mu=500$ and $\sigma=50$?

$$Z = \frac{575 - 500}{50} = 1.5$$

- i.e., A score of 575 is 1.5 standard deviations above the mean

$$\therefore P(X \leq 575) = \int_{200}^{575} \frac{1}{(50)\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-500}{50}\right)^2} dx \longrightarrow \int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2} dz$$

Yikes!

But to look up $Z=1.5$ in standard normal chart (or enter into SAS) \rightarrow no problem! = .9332

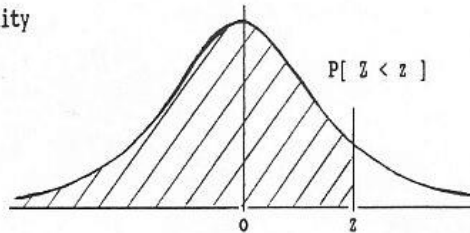
Looking up probabilities in the standard normal table

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[Z < z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2) dz$$



What is the area to the left of $Z=1.51$ in a standard normal curve?

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817

$Z=1.51$

Area is 93.45%

$Z=1.51$
103

$$P_r(a' \leq z \leq b') = P_r(z \leq b') - P_r(z \leq a')$$

$$z = \frac{x - \mu}{\sigma}, a' = \frac{a - \mu}{\sigma}, b' = \frac{b - \mu}{\sigma}$$