



لجان الترقيات

STATISTICS



MORPHINE ACADEMY

MORPHINE
ACADEMY

Pharmaceutical Statistics

Lecture 2

Types of Data → في علم الإحصاء مهم نعرف نوع البيانات إلى عنا حتى نقرر نعالجها ونقدمها بطريقة صحيحة

&

**Graphical Presentation
of Data**

Prepared and presented by
Dr. Muna Oqal

Types of Data

- Nominal data

- Example 2: Race

- 0 -- did not answer

- 1 -- black

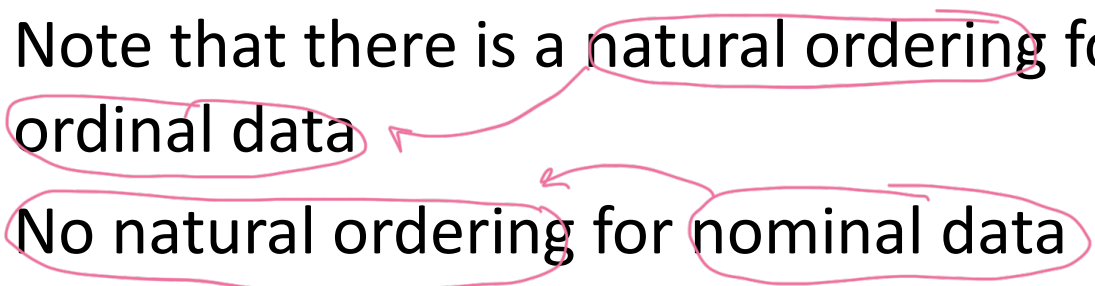
- 2 -- white

- 3 -- Asian

- 4 -- other

تصنيف بدون ترتيب مايقدر
فني و صفة اظهره الجدل

Types of Data

- Note that there is a natural ordering for the ordinal data
 - No natural ordering for nominal data
- 

Types of Data

Quantitative
الكمية

الرقم
المتقطع

- **Discrete data**

- Numbers represent measurable quantities
- Can take on only specified values that differ by fixed amounts
- Often count data (e.g., number of hospitalizations)

عدد حالات دخول المستشفى

Types of Data

- Discrete data
- Example – number of risk factors
 - 0 -- no risk factors
 - 1 -- one risk factor
 - 2 -- two risk factors
 - 3 -- three risk factors
 - 4 -- four risk factors

qualitative

Types of Data

متنوع

- Continuous data

- Represents measurable quantities
- Not restricted to specified values
- Fractional values possible

كمية قابلة
للقياس

عاري يكون
كسر / فاصلة عشرية

مثال درجة
الحرارة 15.5

New York Heart Association Functional Class

← لقسّموا مرضى فشل القلب إلى 4 درجات حسب الأعراض و الجهد المبذول

- Measures extent of heart failure
 - 1: no symptoms or limitations
 - 2: mild symptoms, slight limitations
 - 3: marked limitation in activities due to symptoms
 - 4: severe limitations, symptoms even at rest

Poll

- What type of scale is the NYHA functional class?
 - Nominal
 - **Ordinal** ترتیبی
 - Discrete
 - Continuous

طرق تمثيل
البيانات

عرض البيانات

Data Presentation

- Nominal data

كيف طرق عرض المعلومات الاسمية

- Limited options for presenting data

- Counts in each category

- Histograms  مخطط أعمدة

- Relative frequencies $\frac{\text{كيفية}}{\text{احصها}} = \frac{\text{عدد التكرارات}}{\text{العدد الكلي}}$

- Relative frequency diagram

- Pie chart مخطط دائري



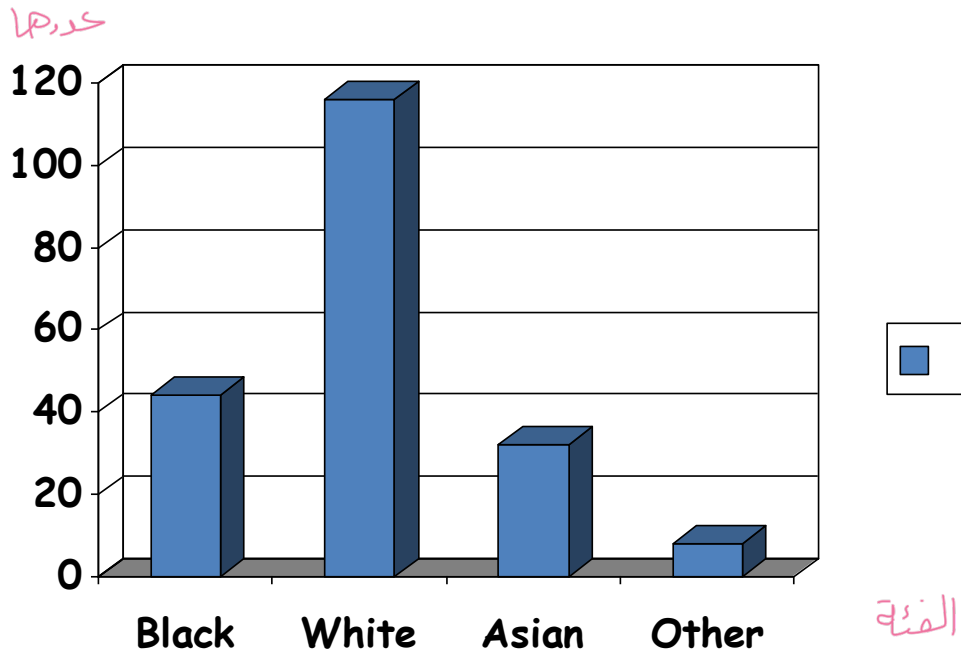
Nominal Data

الفئة

عددها

<u>Race</u>	<u>N</u>
Black	44
White	116
Asian	32
Other	8

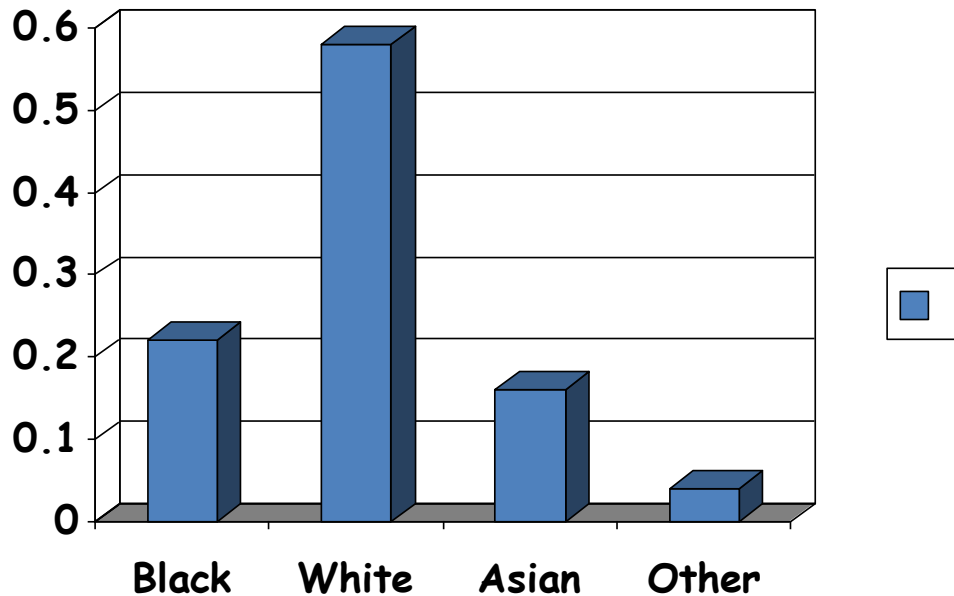
Histogram



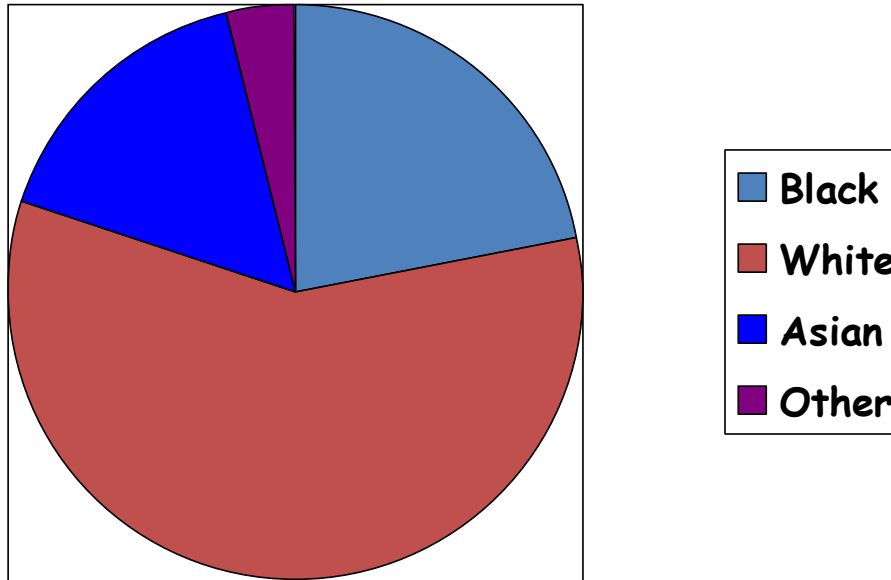
Nominal Data

<u>Race</u>	<u>N</u>	Relative <u>frequency</u>
Black	$\frac{44}{200} \times 100\%$	.22
White	$\frac{116}{200} \times 100\%$	.58
Asian	$\frac{32}{200} \times 100\%$	.16
Other	8	.04
Total	200	1.00

Relative Frequency Diagram



Pie Chart



Data Presentation

- **Ordinal data**

- All the tools for nominal data
- Additional options for presenting data
 - Cumulative frequencies
 - Cumulative frequency polygons
 - Percentiles

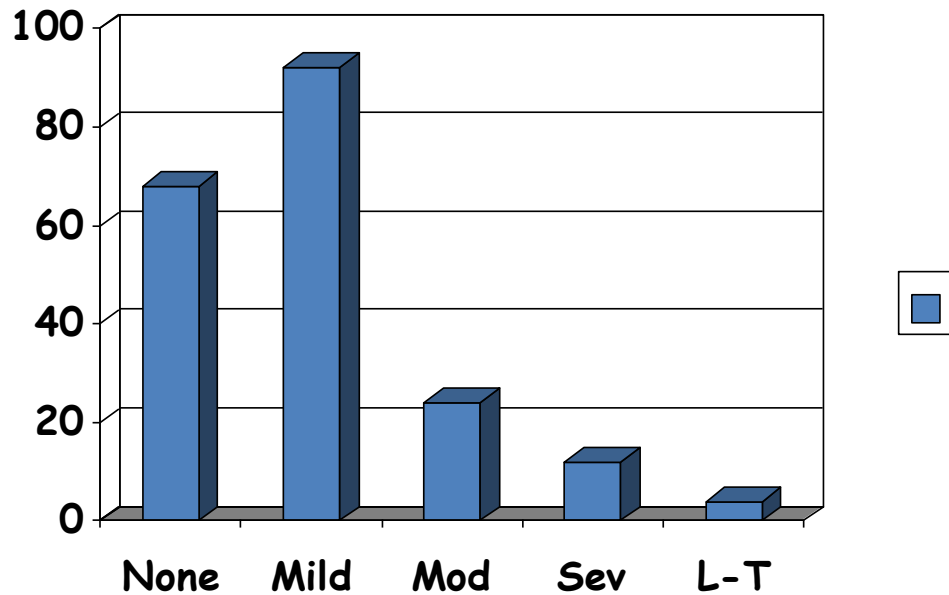
إلى قبل وزيادة
عليهم هضول



Ordinal Data

<u>Toxicity</u>	<u>N</u>
None	68
Mild	92
Moderate	24
Severe	12
Life-threatening	4

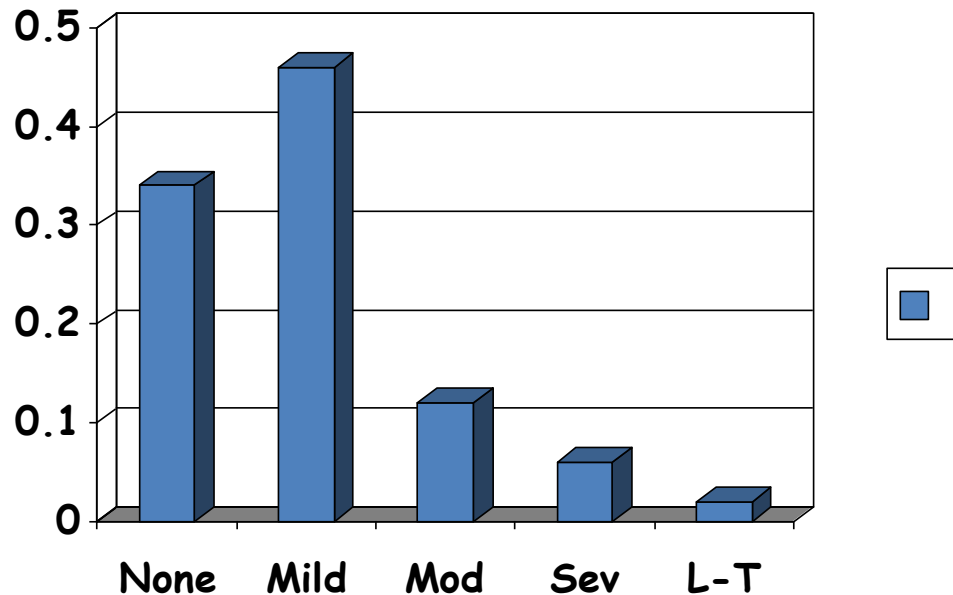
Histogram



Ordinal Data

<u>Toxicity</u>	<u>N</u>	Relative <u>Freq</u>
None	68	.34
Mild	92	.46
Moderate	24	.12
Severe	12	.06
Life-threatening	4	.02

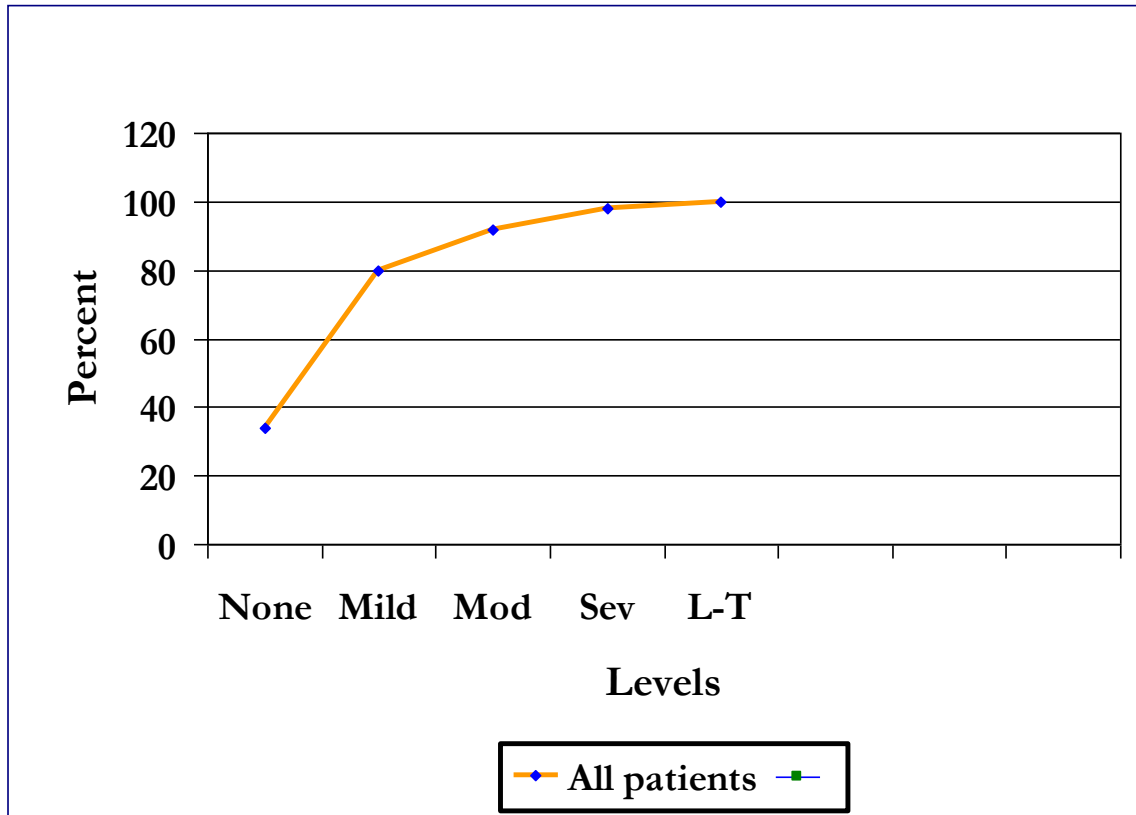
Relative Frequency Diagram



Ordinal Data

<u>Toxicity</u>	<u>N</u>	Relative <u>Freq</u>	<u>Cumulative Freq</u> (التكرار التراكمي)
None	68	.34	.34
Mild	92	.46	.80
Moderate	24	.12	.92
Severe	12	.06	.98
Life-threatening	4	.02	1.00

Cumulative Frequency Polygon



Ordinal Data

- Since the data are ordered^{ترتيب}, we can also use percentiles^{نسبة ١٠٠}
- Percentiles divide a distribution into equal or ordered parts
- The **median**^{المتوسط} (50th percentile)^{المتوسط} divides a population into 2 equal-size parts (more on medians in a couple minutes)

Data Presentation

- Discrete data
 - All the tools for ordinal data
 - Values of the measure have numerical meaning
 - Additional options for presenting data
 - means

لنقصد
المتوسط (means)
لأنها بيانات رقمية

Discrete Data

Number of <u>Risk</u> <u>Factors</u>	<u>N</u>
0	62
1	86
2	32
3	14
4	6

Discrete Data

- The values of the observations now represent actual numeric values
- We can calculate a mean (simple arithmetic average)
- Mean = sum of observations / N

Mean $\rightarrow \frac{\text{مجموع القيم}}{\text{عددهم}}$

Median $\rightarrow$ القيمة التي تنقسم البيانات
لثلاثين

Mode $\rightarrow$ القيمة الأكثر
تكراراً

Range $\rightarrow$ الفرق بين الأعلى والأدنى

SD $\rightarrow$ يظهر مدى تشتت البيانات
حول المتوسط

Discrete Data

عدد مراتب دخول المريض للمستشفى :-

عدد المرضى	مراتب الدخول
62	0
86	1
32	2
14	3
6	4

- In our example,

$$\text{Mean} = [(0 \times 62) + (1 \times 86) + (2 \times 32) + (3 \times 14) + (4 \times 6)] / 200$$

← المتوسط
مجموع
عدد

عدد

$$= [0 + 86 + 64 + 42 + 24] / 200$$

$$= 216 / 200$$

$$= 1.08 \text{ risk factors}$$

Data Presentation

- Continuous data
 - Wide variety of options
 - numerical
 - graphical
 - Important to consider both central tendency and dispersion
 - Examining the distribution is important
 - Divided into (interval & ratio)

Continuous Data

- Data on fluoride levels (1st 15 values)

0.079	0.146	0.112
0.071	1.335	0.072
0.224	0.553	0.071
0.159	0.415	0.119
2.467	0.288	0.154

Measures of Central Tendency

- We can calculate the sample mean as before, i.e., add all the observations together and divide by N, the number of observations
 - Mean = 0.697

Measures of Central Tendency

- To calculate the median, we order observations from smallest to largest.
- If N is ^{فردی} odd, define $j=(N+1)/2$. The j th ordered observation is the median
- If N is ^{زوج} even, $j=N/2$. The median is the average of ordered observations j and $j+1$.

از اعداد
ترتیب
محاسبه
median

Measures of Central Tendency

- Suppose we have 9 observations:

2.60, 2.75, 2.89, 4.05, 2.25,

2.68, 3.00, 4.02, 2.85

- The mean is ^{متوسط} $27.09 / 9 = 3.01$

- The median is 2.85

^{القيمة المتوسطة}

Measures of Central Tendency

- Suppose the 8th observation were 40.02 instead of 4.02

2.60, 2.75, 2.89, 4.05, 2.25,

2.68, 3.00, 40.02, 2.85

- The mean is now $63.09 / 9 = 7.01$
- The median is 2.85

Measures of Dispersion

- It is also important to know something about the variability of the data
- Common measures include:
 - **Range:** difference between the largest and smallest observations
 - **Interquartile range:** difference between the 25th and 75th percentiles
 - **Standard deviation:** square root of the variance

جذر تنوع

Standard Deviation

x	$x - \text{mean}$
2.60	-0.41
2.75	-0.26
2.89	-0.12
4.05	1.04
2.25	-0.76
2.68	-0.33
3.00	-0.01
4.02	1.01
2.85	-0.16
total	0.00

Standard Deviation

$$S.D = \sqrt{\frac{(x - \text{mean})^2}{N - 1}}$$

x	$x - \text{mean}$	$(x - \text{mean})^2$
2.60	-0.41	0.1681
2.75	-0.26	0.0676
2.89	-0.12	0.0144
4.05	1.04	1.0816
2.25	-0.76	0.5776
2.68	-0.33	0.1089
3.00	-0.01	0.0001
4.02	1.01	1.0201
2.85	-0.16	0.0256
total	0.00	3.0640

Standard Deviation

- Variance is the sum of the squared deviations divided by (N-1)
 - Variance = $3.0640 / 8 = 0.383$
- Standard deviation is the square root of the variance
 - s.d. = 0.619

Data Presentation Methods

- Presenting only the mean does not inform the reader of the diversity in the set of values from the sample. Thus, standard deviation (SD) is calculated.
- SD is presented with the mean value of the sample (e.g., 145 mg/dL ±15, where the former number is the mean and the latter number is the SD).

Data Organization

هاي البيانات غير مرتبة

- Measurements that have not been organized, summarized or otherwise manipulated are called raw data.
- Unless the number of observations is extremely small, it will be unlikely that these raw data will impart much information until they have been put into some kind of order.
- Always it is easier to analyze organized Data

Table 1: Raw data of cholesterol lowering effect of a drug given to 156 subjects

17	-12	25	-37	-29	-39
-22	0	-22	-63	34	-31
-64	-12	-49	5	-8	33
-50	-7	16	-11	-38	-17
0	-9	-21	1	2	-30
-32	-34	-14	-18	5	6
24	-6	14	10	-41	-66
-25	-12	14	10	-41	-66
-31	35	21	-19	-27	17
-6	-17	-6	1	-28	40
-31	17	-54	-27	-16	16
-44	10	-3	-3	5	6
-19	9	-10	-20	-9	-8
-10	-11	11	-39	19	-32
4	-15	-18	35	6	20
26	24	-27	-19	6	-60
27	23	-22	-1	12	-27
-13	-39	39	-34	-97	-26
38	14	-47	8	26	-15
-62	12	-53	11	21	-47
-54	-11	-5	0	55	34
-69	-11	-44	20	-50	19
0	-25	-24	-4	14	2
-34	26	-23	-71	-58	9
9	2	-2	-58	13	14
17	-13	-22	-3	-17	1

Ordered Array

- The preparation of the ordered array is the first step in organizing data.

- An ordered array is a listing of the values of a collection (either population or sample) from the smallest value to the largest value.

- The ordered array enables one to determine quickly the value of the smallest measurement, the value of the largest measurement and the general trends in the data.

مرتبة من الجهتين للكميات

Table 2. Ordered array of data reported in Table 1.

-97	-38	-21	-9	5	17
-71	-37	-20	-8	5	17
-69	-34	-19	-8	6	17
-66	-34	-19	-7	6	19
-66	-34	-19	-6	6	19
-64	-32	-18	-6	6	20
-63	-32	-18	-6	8	20
-62	-31	-17	-5	9	21
-60	-31	-17	-4	9	21
-58	-31	-17	-3	9	23
-58	-30	-16	-3	10	24
-54	-29	-15	-3	10	24
-54	-28	-15	-2	10	25
-53	-27	-14	-1	11	26
-50	-27	-13	0	11	26
-50	-27	-13	0	12	26
-49	-27	-12	0	12	27
-47	-26	-12	0	13	33
-47	-25	-12	1	14	34
-44	-25	-11	1	14	34
-44	-24	-11	1	14	35
-41	-23	-11	2	14	35
-41	-22	-11	2	14	38
-39	-22	-10	2	16	39
-39	-22	-10	4	16	40
-39	-22	-9	5	17	55

Grouped Data

The frequency distribution

- Although a set of observation can be made more comprehensible and meaningful by means of an ordered array, further useful summarization may be achieved by grouping the data.
- To group a set of observations, we select a set of non-overlapping intervals such that each value in the data set of observations can be placed in one, and only one, interval.
غير متداخلة يعني القيمة الواحدة تقع في فتره واحدة
- These intervals are usually referred to as Class Intervals.
مترددات الفترات
- Usually class intervals are ordered from smallest to largest.
- Interval width = $19 - 10 + 1 = 10$ or $20 - 10 = 10$
- Total range = $69 - 10 + 1 = 60$ or $10 * 6 = 60$

Class Interval	Frequency
10-19	4
20-29	66
30-39	47
40-49	36
50-59	12
60-69	4
Total	169

Grouped data

R.F = $\frac{\text{عدد النسبة}}{\text{Total}}$

The relative frequency distribution

- It may be useful sometimes to know the proportion rather than the number, of values falling between a particular class interval.
- We obtain this information by dividing the number of values in the particular class interval by the total number of values.
- We refer to the proportion of values falling within a class interval as the relative frequency of values in that interval.
- We may sum (cumulate) the frequencies and relative frequencies to facilitate obtaining information regarding frequency or relative frequency of values within two or more contiguous class intervals.

Total # of values higher than 10 and lower than 19, inclusive

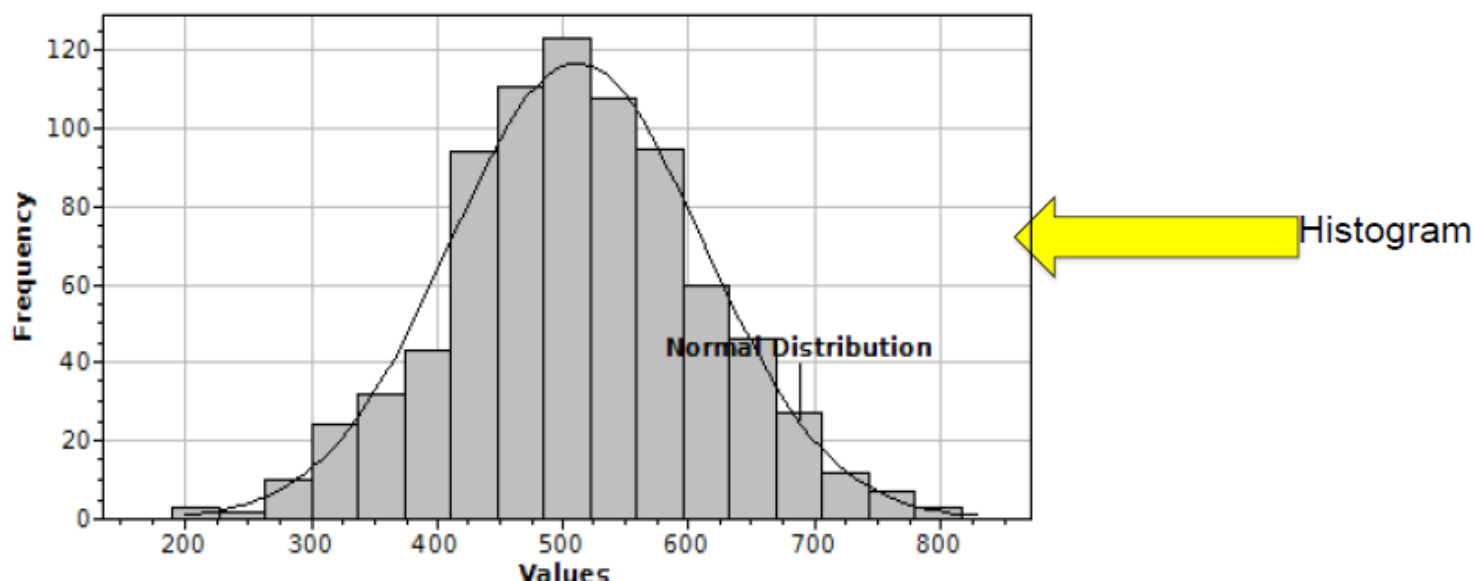
Class Interval	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
10-19	4	0.0237	4	0.0237
20-29	66	0.3905	70	0.4142
30-39	47	0.2781	117	0.6923
40-49	36	0.2130	153	0.9053
50-59	12	0.0710	165	0.9763
60-69	4	0.0237	169	1.0000
Total	169	1.0000		

the ratio of the frequency of a particular event in a statistical experiment to the total frequency.

طریقہ علیہ
ما (طرز)

Histogram

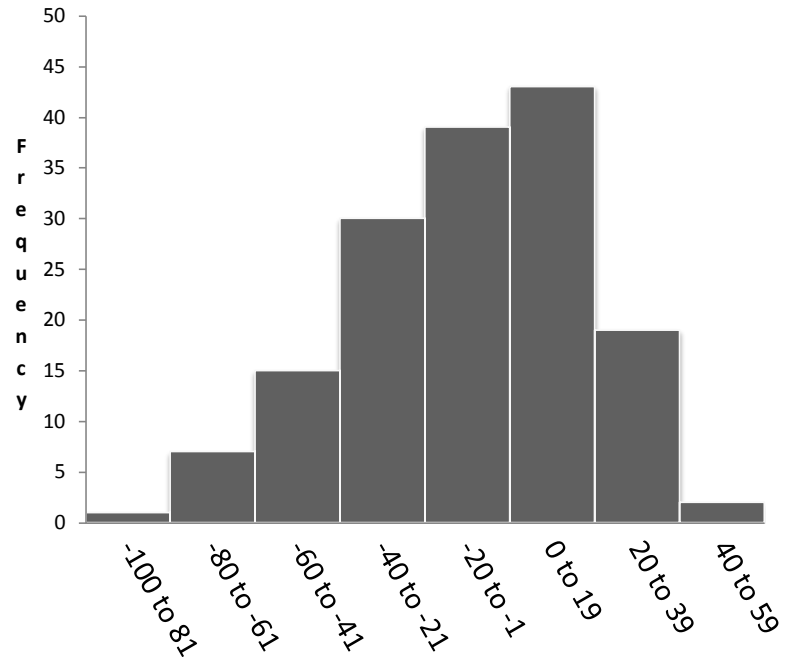
- We may display a frequency distribution (or a relative frequency distribution) graphically in the form of a histogram, which is a special type of bar graphs.
- When we construct a histogram, the **variable** under consideration are represented by the horizontal (**x**) axis, while the **frequency** (or relative frequency) of occurrence is the (**y**) axis.



Construction of Frequency Distribution Table

According to the previous slide,
Data in Table 1 can be put into
frequency distribution table as
follows

Interval	Frequency
-100 to -81	1
-80 to -61	7
-60 to -41	15
-40 to -21	30
-20 to -1	39
0 to 19	43
20 to 39	19
40 to 59	2

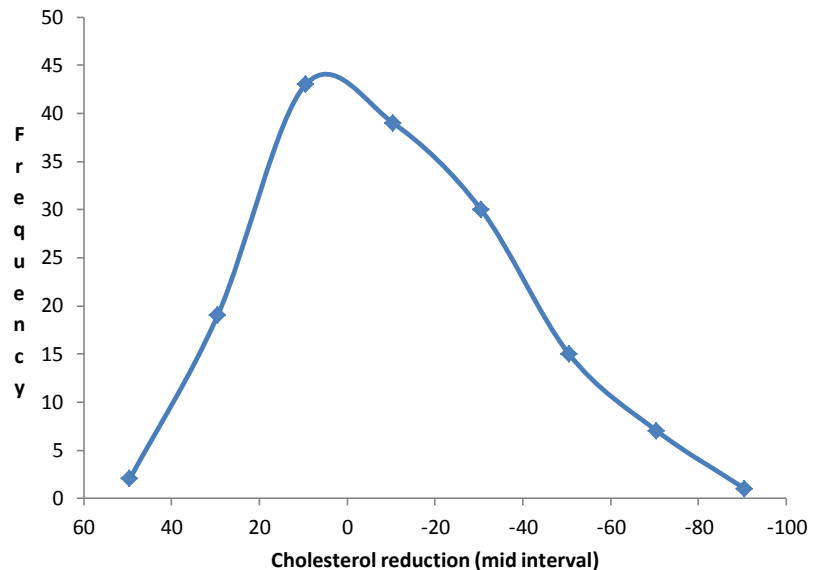


Frequency Distribution Curve

A plot of frequency versus mid interval size taken as the average of upper and lower limits for each interval. For the data in Table 1 the following Table is plotted

Mid interval	Frequency
-90.5	1
-70.5	7
-50.5	15
-30.5	30
-10.5	39
9.5	43
29.5	19
49.5	2

156



Review

Measures of central tendency

- **Mean:** average of the data points

e.g. Values are (2,4,6,10,12,16) → Mean = $\frac{\text{مجموع}}{\text{عدد}}$
 $((2+4+6+10+12+14)/6)=8$

- **Mode:** is most frequently occurring data point

e.g. values are (3,6,6,8,6,5) → mode = 6

العدد الأكثر تكراراً

Measures of central tendency

- **Median:** midpoint of the data (point at which the data lie 50% above and below)

e.g1. (5,6,10,15,14,16,3) → Ascending order
(3,5,6,10,14,15,16)

No. of values (n) = 7 (odd of number)

Median = $(n+1)/2$ = $(7+1)/2 = 4^{\text{th}}$ value = 10

محاسبه
Median

Measures of central tendency

- e.g2. (5,6,10,15,14,16,3,2) → Ascending order
(2,3,5,6,10,14,15,16)

No. of values (n) = 8 (even of number)

Median = Average of ($n/2$ value and $((n/2) + 1)$ value)

Median = Average of the 4th value and the 5th value

= Average of 6 & 10 = 8

Data Presentation Methods

Type of Data	Mode	Median	Mean	Range	SD
Nominal	X				
Ordinal	X	X		X	
Interval and Ratio	X	X	X	X	X