



statistics

Lecture _3

First stage

Dr.worood basim

Data Presentation

After the data collection stage is completed, the data are presented using several methods, including those explained below:

1. Reports

According to this method, data are presented in the form of a report that includes the nature and characteristics of the data. It may also include a summary of the data through tabulated tables based on a specific criterion, which may be temporal, spatial, quantitative, or descriptive.

A. Time Tabulation:

According to this type, data are presented in a table that takes into account the time basis in which the data were affected, as shown in Example (21) below:

Example (21): The table below shows the sales value of one of the cotton industry companies during the years 1996–2000.

Year	Sales Value (thousand dinars)
1996	2000
1997	1500
1998	2900
1999	3000
2000	4000

B. Spatial (Geographical) Tabulation

In this type, the spatial or geographical basis is used to summarize and present data, as shown in Example (22) below:

Example (22): The following table shows the number of units sold of a type of industrial product by governorates in Iraq in the year 2012.

Governorate	Number of Units
Baghdad	120000
Mosul	57000
Basra	41000
Kut (Wasit)	100000
Diyala	10000
Al-Qadisiyah	42000

C. Quantitative Tabulation

In this type, the quantitative basis is used to summarize and present data, as shown in Example (23) below:

Example (23): The following table shows the number of units produced of plastic products in Iraq during the first month of the year 2000, distributed according to factory size (average number of workers).

Average Number of Workers	Number of Units Produced
10	4000
20	5000
30	9000
40	17500

D. Classification by Attributes

In this type, the descriptive (qualitative) basis (attributes) is used to summarize and present data, as illustrated in the following example:

Example (24): The following table shows the values of industrial products in Iraq distributed according to the type of industry.

Type of Industry	Value of Products (thousand dinars)
Extractive	10000
Transformative	100000
Oil and Gas	700000

2. Frequency Distribution Table

A. Definition of the Frequency Distribution Table:

It is an organized summary of the data of a random variable in a structured table consisting of two columns:

- Classes
- Frequencies

A class is defined as a group or a specified interval:

- First: Beginning of the class: the value representing the lower limit (LL: Lower Limit) of the class.
- Second: End of the class: the value representing the upper limit (UL: Upper Limit) of the class.

Example of a class: 4 – 7

Its beginning is the lower limit (LL), which is the value 4.

It has an upper limit (UL), which is the value 7.

Third: Class width or class interval:

It is the difference between the upper limit and the lower limit of the class plus 1, if the values are integers. It is denoted by the symbol w , that is:

$$w = UL - LL + 1 \quad \dots (14)$$

Example: The class width or interval for the class 4–7 is:

$$w = 7 - 4 + 1 = 4$$

Fourth: Class midpoint:

It is the value of the random variable that represents the midpoint between the lower and upper limits of the class, that is:

$$X = \frac{LL+UL}{2} \quad \dots (15)$$

Example: The midpoint of the class 4–7 can be calculated using equation (15):

$$X = \frac{4 + 7}{2} = 5.5$$

Note that class intervals may be equal in width or may be unequal.

Frequency: It represents the number of observations whose values fall within the lower and upper limits of the class.

B. Constructing a frequency distribution table

When we want to construct a frequency distribution table to represent the distribution of the values of a random variable, the following is required:

First: Calculate the total range (Range):

It is the difference between the largest and smallest values of the variable plus one (to ensure all values are included in the frequency distribution table). That is, if the values are:

$$x_1, x_2, \dots, x_n$$

Then the range is:

$$R = \text{Largest } x - \text{Smallest } x + 1 \quad \dots (1)$$

Second: Choosing the number of classes:*

The number of classes that can properly represent the distribution of the values of a random variable can be selected. This number is often taken to be not less than 5 classes and not more than 15 classes.

The choice of the number of classes depends on the researcher's desire to obtain accurate or approximate information. The greater the desire for accuracy, the larger the number of classes. On the other hand, if the researcher prefers approximate rather than detailed results, the number of classes can be reduced.

It is also possible to follow one of the methods that use mathematical formulas to determine the number of classes, which are:

1. Yule's Method:

In this method, the number of classes is determined using the following formula:

$$k = (2.5) * \sqrt[4]{n} \quad \dots (2)$$

Where:

* (k): number of classes

* (n): number of observations (values of the random variable)

The result can be rounded to the nearest larger integer.

2. Sturges' Method:

In this method, the number of classes is determined using the following formula:

$$k = 1 + (3.322) \log(n) \quad \dots (3)$$

Third: Determining the length or width of a class:

The class width is determined based on the range (R) and the number of classes (k) obtained in steps (1) and (2). The class length (L) is calculated using the following formula:

$$L = \frac{R}{k} \quad \dots (4)$$

Ex 1/ The following data represents the number of orange trees for 60 farming families.

Required: Construct a frequency distribution table .

160, 76, 80, 120, 132, 82, 90, 60, 68, 72, 150, 142, 150, 164, 88, 90, 98, 101, 103, 110, 119, 116, 120, 126, 109, 114, 122, 111, 116, 90, 78, 93, 95, 98, 104, 120, 113, 121, 119, 125, 126, 130, 131, 136, 118, 120, 142, 150, 154, 122, 123, 139, 120, 156, 154, 136, 137, 110, 136, 149.

First: Calculating the Range according to formula (1). After determining the largest value (which is 164 here) and the smallest value (which is 60), we get:

$$R = 164 - 60 + 1 = 105$$

Second: Choosing the Number of Classes:

The number of classes k can be selected by the researcher in a way that gives a proper representation of the data, as previously explained, where: $5 \leq k \leq 15$ Or by using Yule's formula (17):

$$k = (2.5)^{\sqrt[4]{60}} = 2.5(2.783) = 6.958 \cong 7$$

Third : Calculating the length of a class according to the formula. Thus, we can

$$w = \frac{R}{k} = \frac{105}{7} = 15$$

Class sequence	Lower limit	Upper limit	class limit
1	60	74	60-74
2	75	89	75-89
3	90	104	90-104
4	105	119	105-119
5	120	134	120-134
6	135	149	135-149
7	150	164	150-164

Classes	classes	fi	midpoint
1	60-74	3	67
2	75-89	5	82
3	90-104	10	97
4	105-119	11	112
5	120-134	15	127
6	135-149	8	142
7	150-164	8	157
sum		60	

Ex 2/ The following data represents the weights (kg) of a sample of (60) students from a college. This data should be entered into a frequency distribution table.

101	65.3	70.5	83	94	55	46	47.8	62.3	77.2
61.3	62.9	66.5	66.9	68.3	70	73.2	73.8	60.2	82.4
58.2	59.9	54.4	58.7	52.6	51.8	74.1	80.2	75.1	78.3
58.5	62	66.3	65	67.1	88.2	89.1	80.1	76.1	76.3
81.5	96.3	61.1	65	63.1	64.9	66.2	65.3	72.1	73.6

$$1) T.R = XL - XS + 1 = 101 - 46 + 1 = 56$$

$$2) M = 1 + 3.322 \log n = 1 + 3.322 \log (50) = 6.644 \cong 7$$

$$3) L = \frac{T.R}{M} = \frac{56}{7} = 8$$

Class sequence	Lower limit	classes	fi	midpoint
1	46	46-	4	50
2	54	54-	9	58
3	62	62-	15	66
4	70	70-	11	74
5	78	78-	6	82
6	86	86-	2	90
7	94	94-102	3	98
sum			50	

3. Graphical presentation

1. *Graphs drawn using Cartesian coordinates*

- Line graph
- Frequency polygon
- Histogram
- Bar graph

Line Graph

For example find Line Graph

Year	MMR
1960	50
1970	45
1980	26
1990	15
2000	12

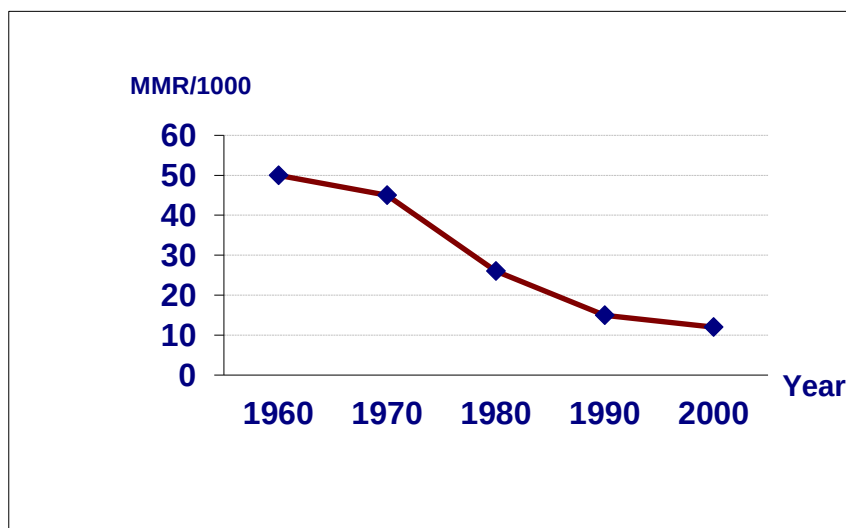


Figure (1): Maternal mortality rate of (country),
1960-2000

Bar chart

Marital status	f	$f\% = \left(\frac{f}{n}\right)*100$
single	12	32.43%
marital	20	54.05%
divorced	3	8.11%
widowed	2	5.41%
sum	37	100%

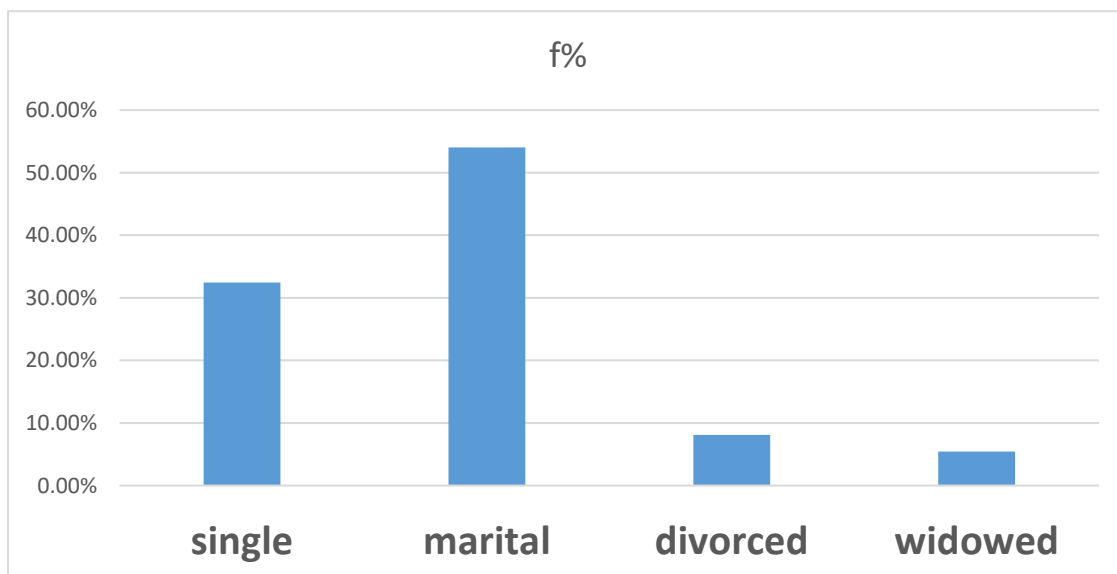


Figure (2): Marital status

Histogram

Distribution of a group of cholera patients by age

Age (years)	Frequency	%
25-	3	14.3
30-	5	23.8
40-	7	33.3
45-	4	19.0
60-65	2	9.5
Total	21	100

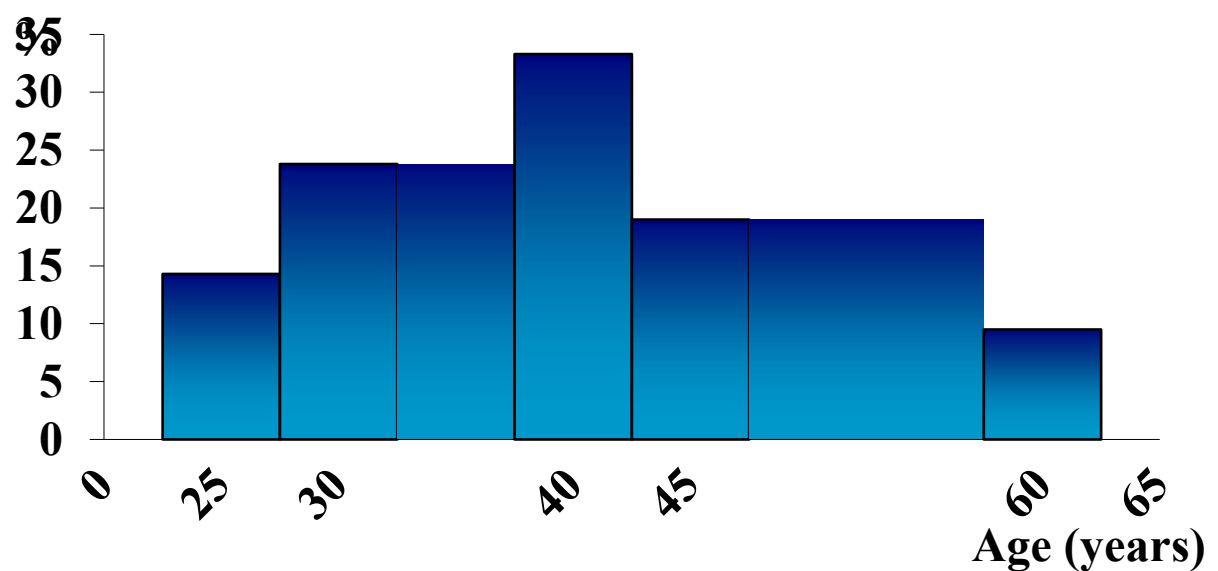


Figure (3): cholera patients by age

Frequency polygon

Age (years)	Sex		Mid-point of interval
	Males	Females	
20 -	3 (12%)	2 (10%)	$(20+30) / 2 = 25$
30 -	9 (36%)	6 (30%)	$(30+40) / 2 = 35$
40-	7 (8%)	5 (25%)	$(40+50) / 2 = 45$
50 -	4 (16%)	3 (15%)	$(50+60) / 2 = 55$
60 - 70	2 (8%)	4 (20%)	$(60+70) / 2 = 65$
Total	25(100%)	20(100%)	

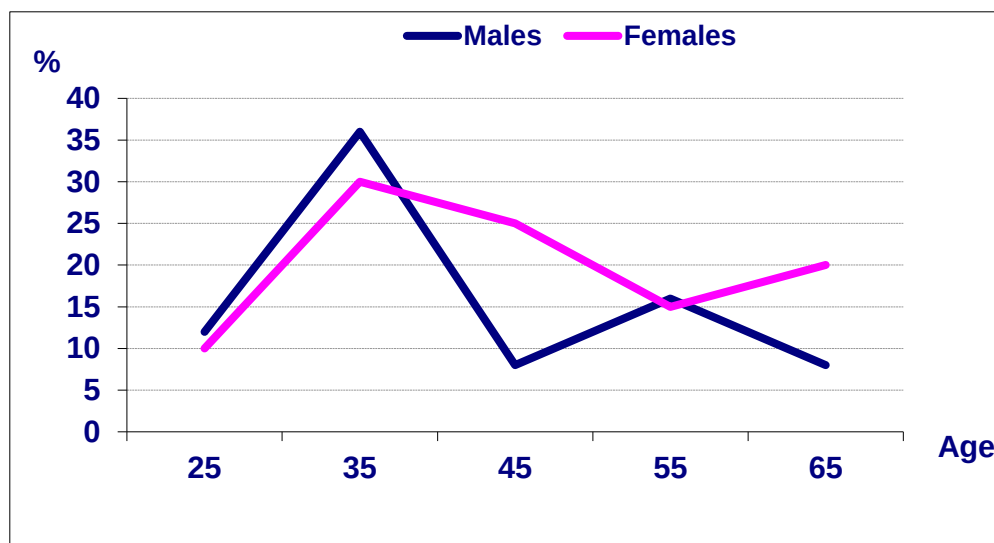


Figure (4): Distribution of 45 patients at (place) , in (time) by age and sex