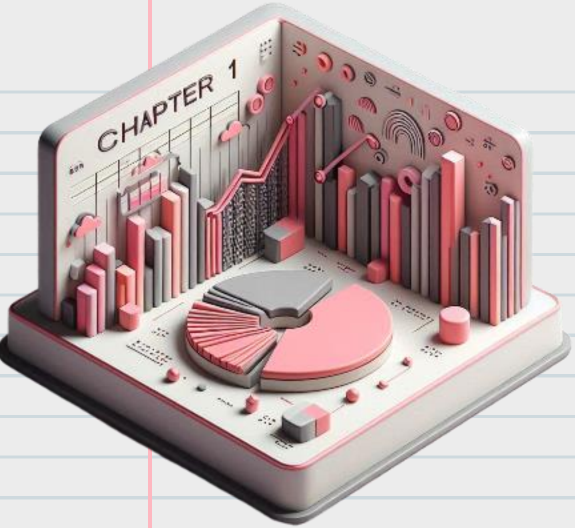


Part A : Statistics

Chapter 2: One-variable statistical series

01 Count, Frequency, Percentage:

02 Graphical representations

03 Position characteristics

04 Dispersion characteristics

Don't forget

- V. Qualitative
  - Ordinal
  - Nominal
- V. Quantitative
  - Discrete
  - Continuous

01 Count, Frequency, Percentage:

Partial Count (ni)

Corresponds to the number of individuals in the class or the count of how many times that value appears.

ni : the number of individuals who have the same xi, is called partial count of xi.

Example 11 :

A survey carried out in a village concerns the number of dependent children per family. We note X the number of children, the results are given by this table:

xi	0	1	2	3	4	5	6
ni(Count)	33	17	43	62	28	38	29

We have: P: all families, i: a family, X : Number of children per family

In this example, 28 is the number of families that have 4 children.

xi	0	1	2	3	4	5	6
ni(Count)	33	17	43	62	28	38	29

Cumulative Count (N)

- This is the total number of individuals in the population, it corresponds to the sum of all partial counts.

$$N = n1 + n2 + n3 + \cdots .. + ni = \sum_{i=1}^{k} ni$$

Example :

In the previous example, we have $N = 250$

x_i	0	1	2	3	4	5	6	N
n_i (Count)	33	17	43	62	28	38	29	250

Partial Frequency (fi)

It represents the ratio of the count (ni) by the total count (N).

$$fi = \frac{ni}{N} \quad (0 < fi \leq 1)$$

Example:

We take the previous example,

xi	0	1	2	3	4	5	6	N
ni (Count)	33	17	43	62	28	38	29	250
fi (Frequency)	$\frac{33}{250}$ 0,132	$\frac{17}{250}$ 0,068	$\frac{43}{250}$ 0,172	$\frac{62}{250}$ 0,248	$\frac{28}{250}$ 0,112	$\frac{38}{250}$ 0,152	$\frac{29}{250}$ 0,116	$\frac{250}{250}$ 1

Cumulative Frequency (Fi)

For each value xi, we set by definition

$$Fi = f1 + f2 + f3 + \cdots \ldots \ldots + fi$$

$$\sum_{i=1}^{k} fi = \sum_{i=1}^{k} \frac{ni}{N} = \frac{1}{N} \sum_{i=1}^{k} ni = 1$$

For i = 1, The cumulative frequency of order 1 is F1 = f1.

For i = 2, The cumulative frequency increasing of order 2 is F2 = f1 + f2

.

Example:

We take the previous example,

xi	0	1	2	3	4	5	6	N
ni (Count)	33	17	43	62	28	38	29	250
fi (Frequency)	$\frac{33}{250}$ 0,132	$\frac{17}{250}$ 0,068	$\frac{43}{250}$ 0,172	$\frac{62}{250}$ 0,248	$\frac{28}{250}$ 0,112	$\frac{38}{250}$ 0,152	$\frac{29}{250}$ 0,116	$\frac{250}{250}$ 1
$F_i=\sum f_i$	0,132	0,2	0,372	0,62	0,732	0,884	1	

Percentage

We can replace fi by **fi × 100** which then represents a percentage.

Ex: $f1=0,132$

$f1\%=0,132*100=13,2\%$

Comprehension exercise

The number of rooms in two hundred apartments with four rooms has been recorded. The following table was obtained:

- Calculate frequencies and cumulative frequencies.

xi	1	2	3	4	Total
ni	50	67	70	13	N=200
fi	0,25	0,335	0,35	0,065	F=1
fi%	25	33,5	35	6,5	100
Fi =∑fi	0,25	0,585	0,935	1	

•Statistical tables:

The statistical table of a discrete variable will be:

The statistical table of a continuous variable will be:

Xi (Variable)	x1	x2	x3	x4	x5
ni (Counts)	n1	n2	n3	n4	n5
fi (Frequencies)	f1	f2	f3	f4	f5

Remark: By convention, the upper bound of a class is excluded from this class, except for the last class it is always closed.

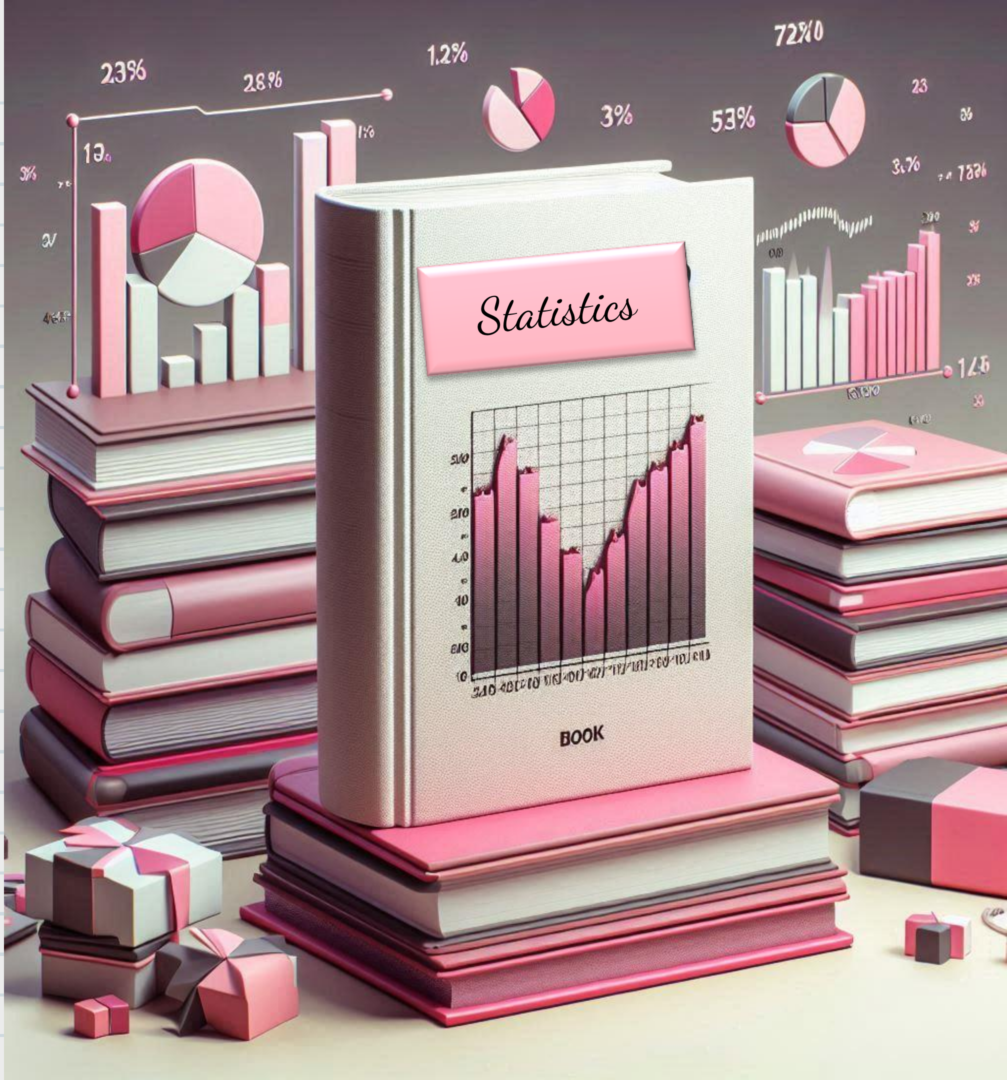
Classes	[a-b[	[b-c[	[c-d[	[d-e[	[e-f]
ni (Counts)	n1	n2	n3	n4	n5
fi (Frequencies)	f1	f2	f3	f4	f5

Don't forget

- Count
- Frequency
- Percentage

02

Graphical Representations



1. Case of a qualitative variable

a. Pie (circle) chart:

A pie chart, sometimes called a circle chart, is a way of summarizing a set of nominal data or displaying the different values of a given variable (e.g. percentage distribution). In This type of chart, the modalities are represented by an angular sector of a disk (or half-disk), the angle of which is proportional to the count or frequency. So, we make a cross product to know the angle of each sector. In the case of a complete disk, we have the following correspondences:

1. Case of a qualitative variable

a. Pie (circle) chart

N(100%)---------- 360°

ni (n%) --------- θi

$$\theta i = \frac{ni}{N} * 360° = CP * ni = fi * 360°$$

We can use the coefficient of proportionality: $CP = \frac{360}{N}$

Variable	x_1	x_2	x_3	x_4	Total	Coefficient of proportionality $CP=360/N$
Count	n_1	n_2	n_3	n_4	N	
frequency	$f_1= n_1/N$	$f_2= n_2/N$	$f_3= n_3/N$	$F_4= n_4/N$	1	
Angles	$\theta_1=n_1*CP$ $=f_1*360°$	$\theta_2=n_2*CP$ $=f_2*360°$	$\theta_3=n_3*CP$ $=f_3*360°$	$\theta_4=n_4*CP$ $=f_4*360°$	360°	

1. Case of a qualitative variable

a. Pie (circle) chart:

Example 13 :

The study regime was studied on a sample of 200 pupils from a high school, the results obtained are as follows:

Make the Graphical Representation of this data in pie chart using frequencies.

Education regime	External	Internal	Half boarder
Number of pupils ni	70	50	80
Frequency fi	70/200=0.35	50/200=0.25	80/200=0.4

1. Case of a qualitative variable

a. Pie (circle) chart:

Example 13 :

To make the pie chart of this data using frequencies, we must calculate the frequencies and the different angles θi of each sector.

n --------- 360°

ni --------- di

$\theta_i = n_i/N*360° f_i*360°$

$\theta_1 = f_1*360° = 0.35*360° = 126°$, $\theta_2 = f_2*360° = 0.25*360° = 90°$,

$\theta_3 = f_3*360° = 0.4*360° = 144$

Education regime	External	Internal	Half boarder
Number of pupils ni	70	50	80
Frequency fi	70/200=0.35	50/200=0.25	80/200=0.4

1. Case of a qualitative variable

a. Pie (circle) chart:

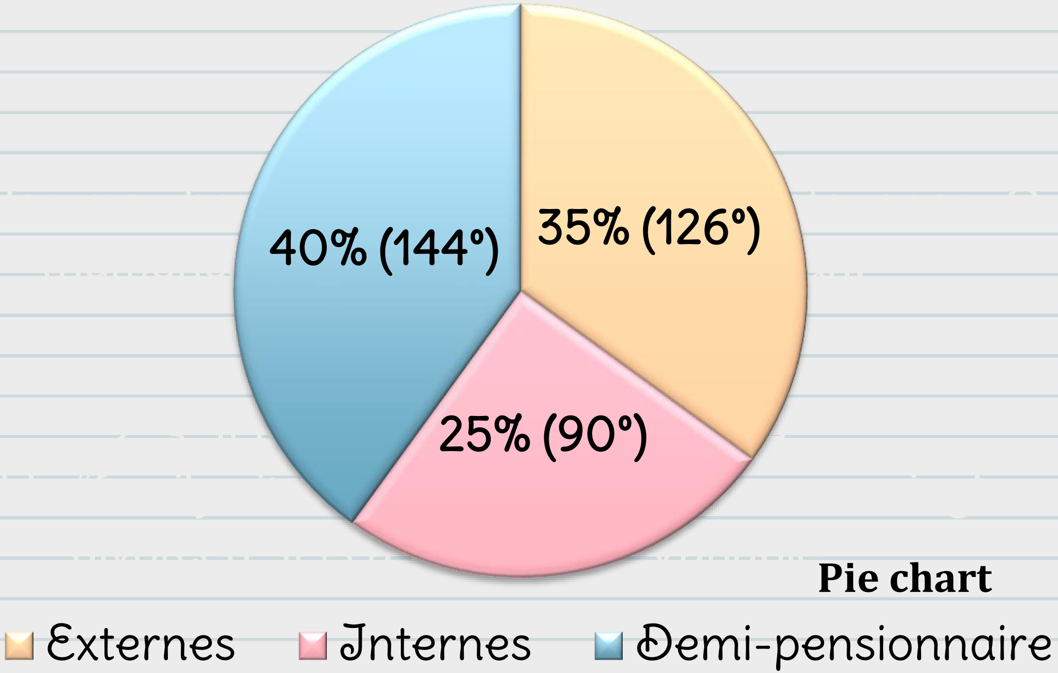


Figure. 2.1: Pie chart shown the study regime of 200 pupils from a high school

1. Case of a qualitative variable

b. Organ pipes:

We put the modalities on the x-axis, arbitrarily. We carry rectangles on the y-axis (vertically or horizontally) whose length is proportional to the counts, or frequencies, of each modality.

1. Case of a qualitative variable

Example 14 :

A survey conducted in 2015 in Algiers on the distribution of blood groups yielded the following results:

Blood groups	A	B	AB	O
Counts	219	123	78	242

1. Case of a qualitative variable

b. Organ pipes

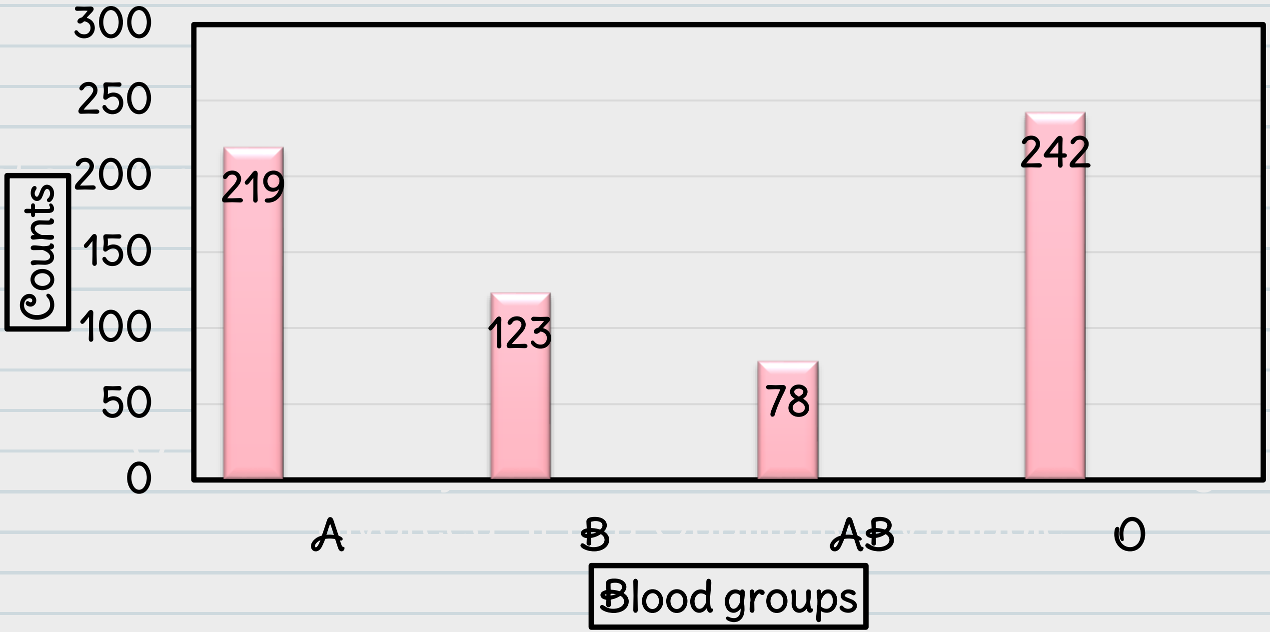


Figure. 2.2: Organ pipes shown the distribution of blood groups.

1. Case of a qualitative variable

c. Band chart:

It consists to represent each modality in the same vertical band by a slice whose height corresponds to its frequency percentage.

1. Case of a qualitative variable

b. Band chart

Example 15 :

The sales made by a car manufacturing company during the years 2000 and 2004 are as follows:

Vehicles	2000		2004	
	ni	fi	ni	fi
Two doors	800	800/5000=0.16	1600	0.23
Four doors	1500	0.3	2000	0.28
Five seats	1700	0.34	900	0.13
Lux Model	1000	0.2	2500	0.36
total	5000	1	7000	1

1. Case of a qualitative variable

b. Band chart

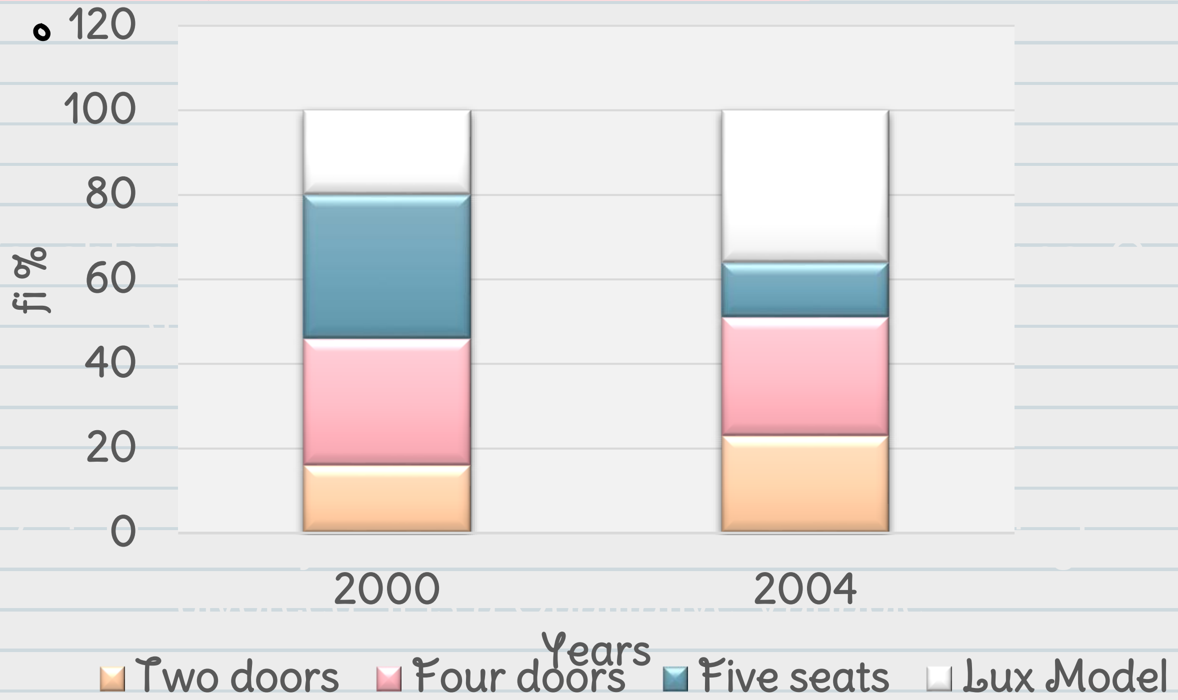
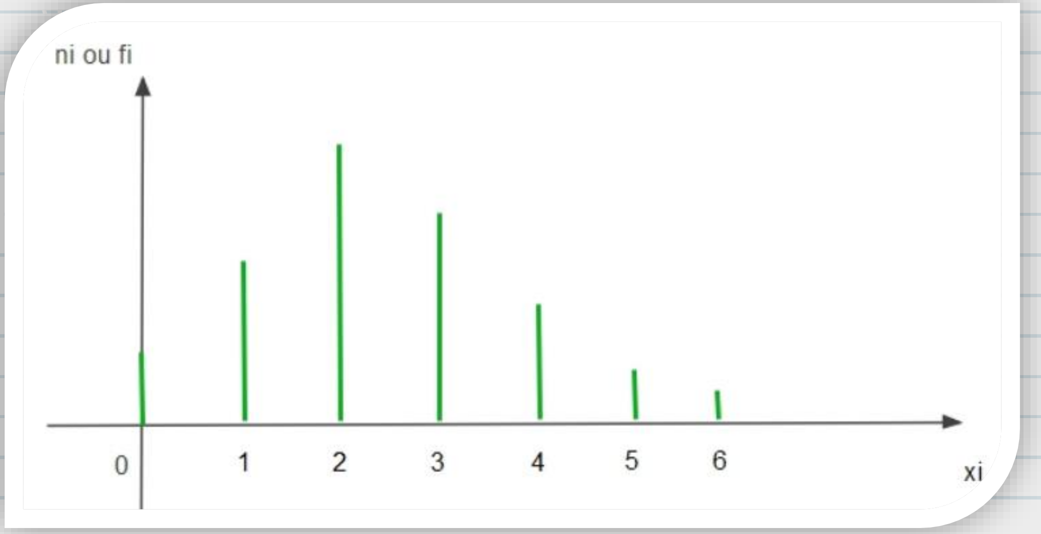


Figure. 2.3: Band chart shown the sales made by a car manufacturing company.

2. Case of a quantitative variable

a. Discrete variable:

Bar chart: We associate a vertical segment whose height is proportional to the known value (count or frequency).



2. Case of a quantitative variable

a. Discrete variable

Example 16 :

Here is the distribution of marks of a test marked out of 5 for a class.

Marks	**1**	**2**	**3**	**4**	**5**
Count	2	4	3	7	8

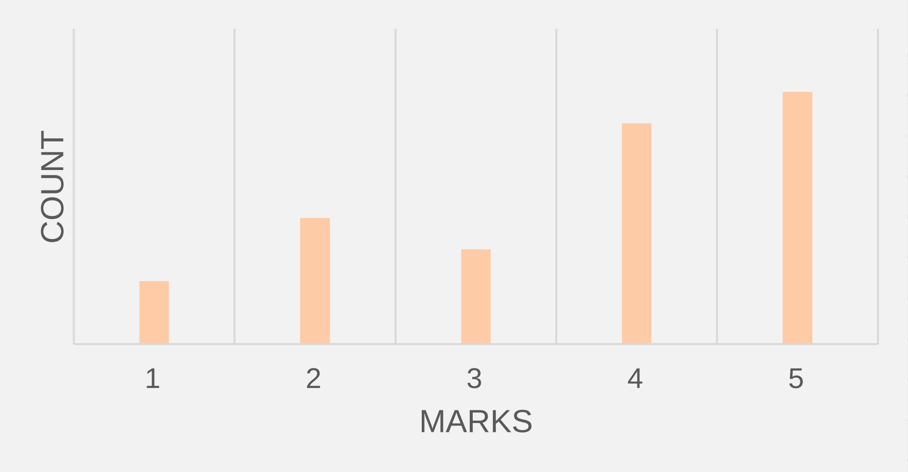


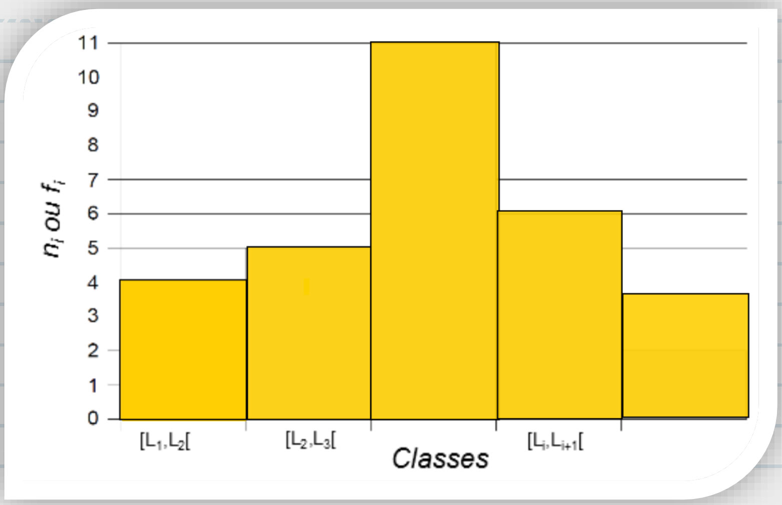
Figure 2.4: Bar chart shown the marks of a test marked out

2. Case of a quantitative variable

a. Continuous variable:

Histogram:

A histogram is a pictorial representation of the numerical data with rectangular bars whose bases are the classes, and whose areas are equal to the class frequencies (or counts).



2. Case of a quantitative variable – b, Continuous variable

Histogram

a, Equal Class sizes:

The classes will be represented on the x-axis, and above each of them, we draw a rectangle whose area is proportional to the associated frequency fi or the count ni.

- A class is a bounding interval $[b_i, b_{i+1}[$
- The class center is: $C_i = \frac{b_i + b_{i+1}}{2}$
- The class size is: $a_i = b_{i+1} - b_i$

2. Case of a quantitative variable

b, Continuous variable

Histogram Equal Class sizes

We consider the age of the inhabitants of a neighborhood.

Classes	[10-15[	[15-20[	[20-25[	[25-30[	[30-35]
ni (Counts)	20	10	5	15	10
fi (Frequencies)	0.33	0.16	0.08	0.25	0.16
ai (Class size)	5	5	5	5	5

2. Case of a quantitative variable

b, Continuous variable

Histogram Equal Class sizes

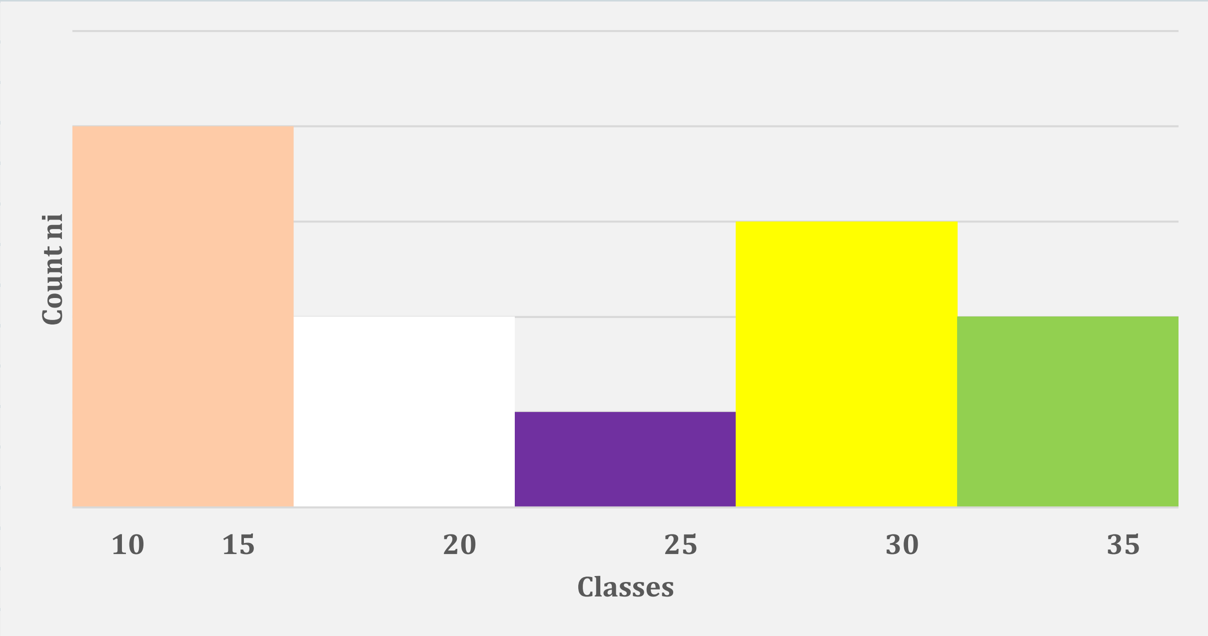


Figure. 2.5: Histogram shown the age of the inhabitants in a neighborhood.

b, Unequal Class sizes:

The classes will be represented on the x-axis, and above each of them, we draw a rectangle whose area is proportional to

- The rectified frequency fi' or the rectified count ni'.
- The density of frequency or the density of counts.

❖ rectified count (ni') and frequency (fi') are computed by using the following formula:

$$\text{Rectified frequency (or count)} = \frac{\text{Minimum class size} * \text{Frequency (or count) of the class}}{\text{Class size}}$$

2. Case of a quantitative variable

b, Continuous variable

Histogram

b, Unequal Class sizes:

- The density of the count (di): it is used for a count histogram, the height of the rectangle corresponding to class i is therefore given by: $di = \frac{ni}{ai}$
- The density of the frequency (di): it is used for a frequency histogram; it is given by: $di = \frac{fi}{ai}$

2. Case of a quantitative variable

b, Continuous variable

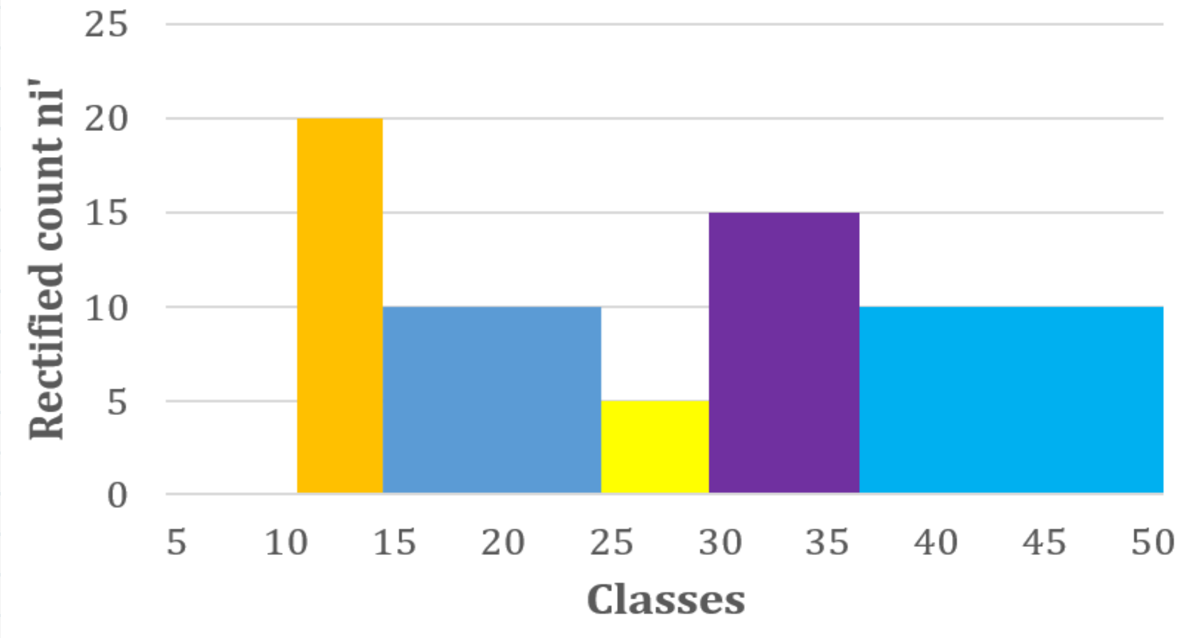
Histogram unequal Class sizes

We consider the age of the inhabitants of a neighborhood.

Classes	[11-15[	[15-25[	[25-30[	[30-37[	[37-50]
ni (Counts)	**20**	10	5	15	10
fi (Frequencies)	**0.33**	0.16	0.08	0.25	0.16
ai (Class size)	**4**	**10**	**5**	**7**	**13**
ni' (Rectified ni)	$\frac{4*20}{4} = 20$	$\frac{4*10}{10} = 4$	$\frac{4*5}{5} = 4$	$\frac{4*15}{7} = 8.5$	$\frac{4*10}{13} = 3$

2. Case of a quantitative variable -- b, Continuous variable

Histogram unequal Class sizes



2. Case of a quantitative variable

c, Sturges Rule: (Transform the discrete quantitative variable into a continuous quantitative variable)

In case we have more than 20 modalities, we should use too many bins in the graphical representations, we may just be visualizing the noise in a dataset. By chance, we can use a method known as Sturges' Rule to determine number of classes to use in a histogram or frequency distribution table. Class formation involves transforming data into continuous distribution.

c, Sturges Rule:

For grouping data (for quantitative variable), we should respect these points:

- ❑ The number of classes should be between five and twenty classes.
- ❑ Each piece of data must belong to one, and only one, class.
- ❑ Whenever feasible, all classes should have the same width (class size).

c, Sturges Rule:

Sturge's Rule: $k = 1 + 3.322 \log N$

k: Number of Classes (Rounded to the nearest integer)

N: Total number of observations

The classes are chosen of equal class sizes interval: $a = \frac{E}{k}$

a: class size

E: range

E= Xmax - Xmin (Difference Between Largest and Smallest Observation)

c, Sturges Rule:

Example 19:

Train classes for these observations:

153	165	160	150	159	151	163
160	158	149	154	153	163	140
158	150	158	155	163	159	157
162	160	152	164	158	153	162
166	162	165	157	174	158	171
162	155	156	159	162	152	158
164	164	162	158	156	171	164
158						

c, Sturges Rule:

Sturges Rule: k=1+3.322 logN=1 + (3,322 log 50) = 6,64= about 7 classes.

a: class size, $a = \frac{E}{k}$

Ɛ: Range, Ɛ= Xmax -Xmin = 174 –140 = 34

$$a = \frac{E}{k} = \frac{34}{6.64} = 5.12 \sim 5$$

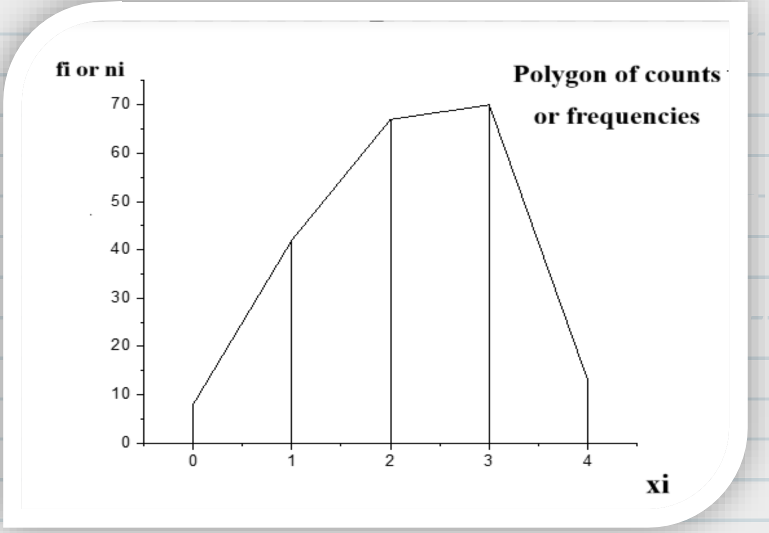
Classes	[140-145[	[145-150[	[150-155[	[155-160[	[160-165]	[165-170[	[170-175[
ni	1	1	9	17	16	3	3

Polygon

The polygon (counts or frequencies) is made from the bar chart or histogram. It is a broken line which joins the upper points of the bars or the midpoints of the upper sides of the successive rectangles of the histogram.

Polygon

For the discrete variable, the polygon of counts or frequencies starts at the top of the stick of the first value (x1) and ends with the top of the stick of the last value (xf).

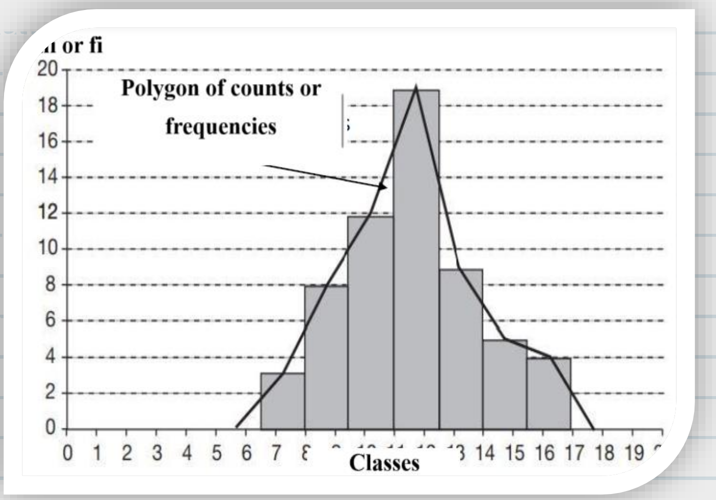


Polygon for Discrete variable

Polygon

- For the continuous variable, the polygon of counts or frequencies begins with:

- **Equal Class sizes:** We start and end the polygon with (a/2) and the widths of the bars are equal.
- **Unequal Class sizes:** We start the polygon with (a1/2) and end it with (af/2) and the widths of the bars are unequal.



Polygon for countinuous variable

Cumulative Curves

a, Discrete variable :

For a discrete variable the cumulative curve of frequencies or counts is represented by steps (stairs) curve. It's a continuous curve on the left.

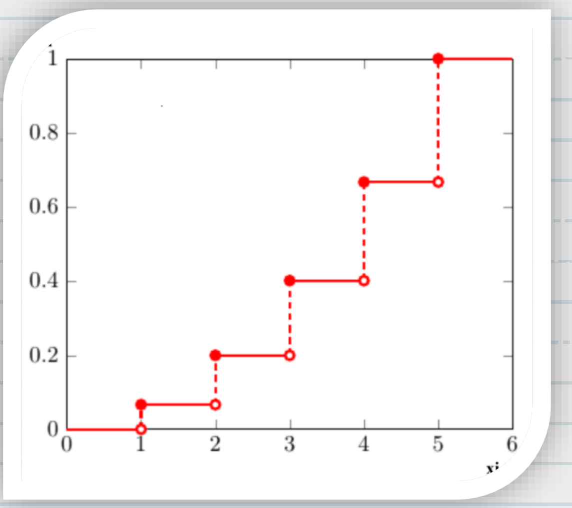


Figure. 2.7: Increasing cumulative curve for discrete variable.

Cumulative Curves

a, Discrete variable

Example 20:

Length (m) X	71	74	77	80	83	Total
Counts (ni)	6	17	41	27	9	
Cumulative Count (Ni)						
Frequencies (fi)						
Cumulative frequencies (Fi)						

Cumulative Curves

a, Discrete variable

Example 20:

Length (m) X	71	74	77	80	83	Total
Counts (ni)	6	17	41	27	9	100
Cumulative Count (Ni)	6	23	64	91	100	
Frequencies (fi)	0.06	0.17	0.41	0.27	0.09	1
Cumulative frequencies (Fi)	0.06	0.23	0.64	0.91	1	

Cumulative Curves

a, Discrete variable

Example 20:

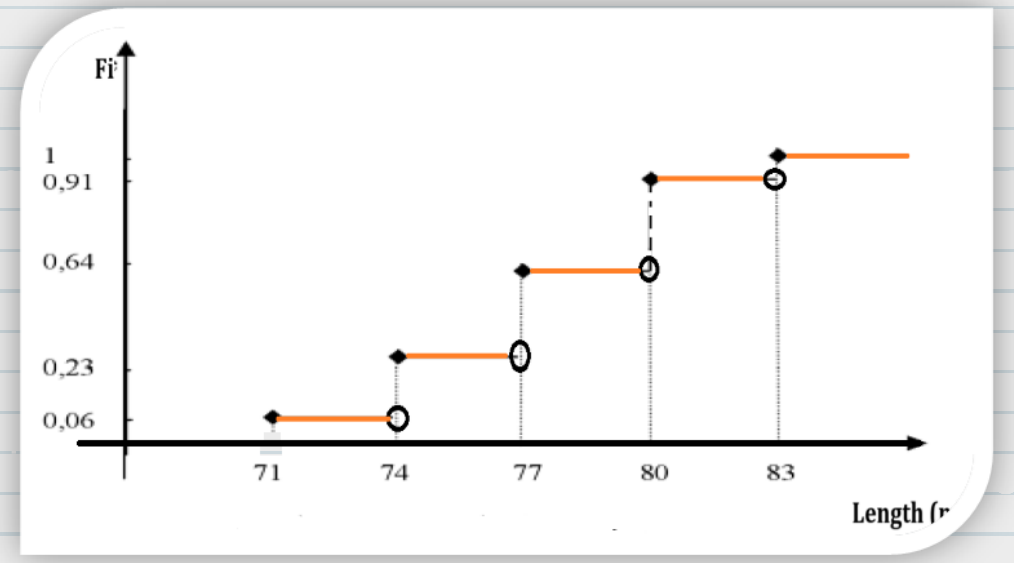


Figure. 2.8: Jncreasing frequencies cumulative curve for the variable length.

Cumulative Curves

b, Continuous variable :

The cumulative curve of a continuous variable is obtained by plotting the points whose x-coordinates represent the upper bound of each class and the y-coordinates the corresponding cumulative frequencies or counts, and then connecting these points by line segments.

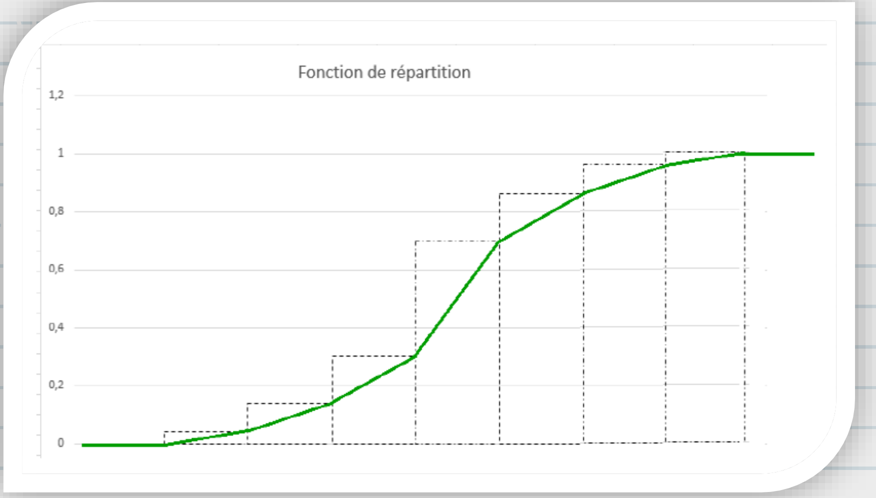


Figure. 2.9: Increasing cumulative curve for continuous variable.

Cumulative Curves

b, Continuous variable :

The cumulative count curve can be thought of as the graph of a function, called the cumulative count function and designated by N(x), defined on R and with values in the interval [0, n].

$$N : \mathbb{R} \to [0, n], \quad x \mapsto N(x)$$

Similarly, the cumulative frequency curve can be thought of as the graph of a function, called the cumulative frequency function and denoted by F(x), defined on R and with values in the interval [0, 1].:

$$F : \mathbb{R} \to [0, 1], \quad x \mapsto F(x)$$

Cumulative Curves

b, Continuous variable :

Example 21:

We consider the age of the inhabitants of a neighborhood.

Classes	[11-15[	[15-25[	[25-30[	[30-37[	[37-50]	Total
ni (Counts)	20	10	5	15	10	60
Ni (Cumulative Count)						
fi (Frequencies)						
Fi						

Cumulative Curves

b, Continuous variable :

Example 21:

We consider the age of the inhabitants of a neighborhood.

Classes	[11-15[	[15-25[	[25-30[	[30-37[	[37-50]	Total
ni (Counts)	20	10	5	15	10	60
Ni (Cumulative Count)	20	30	35	50	60	
fi (Frequencies)	0.34	0.17	0.08	0.25	0.16	1
Fi	0.34	0.51	0.59	0.84	1	
F(x) 0	0.34	0.51	0.59	0.84	1	

Cumulative Curves

b, Continuous variable :

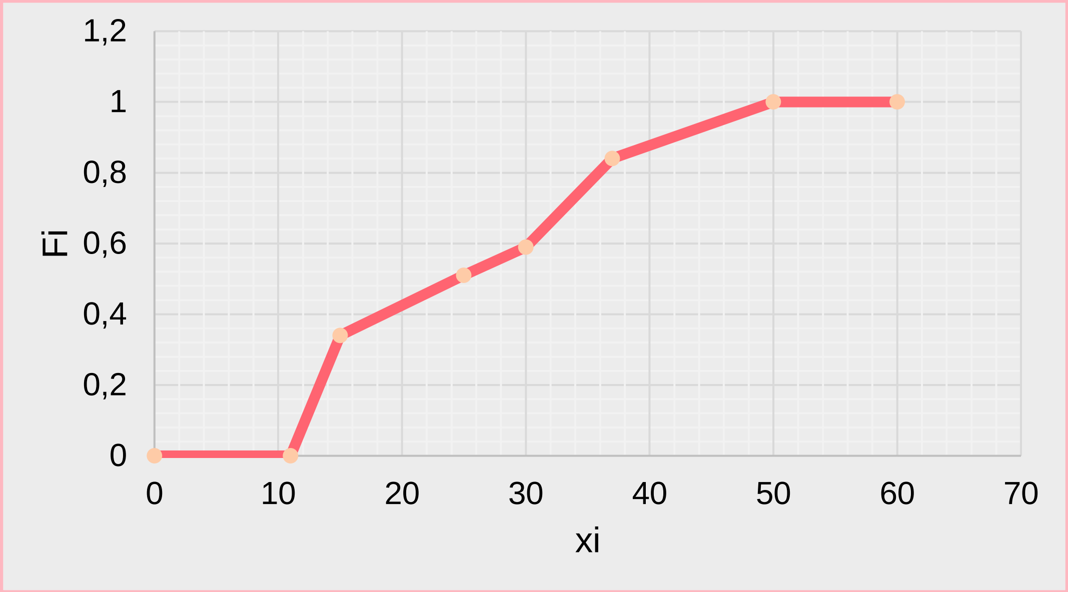


Figure. 2.10: Increasing frequencies cumulative curve.