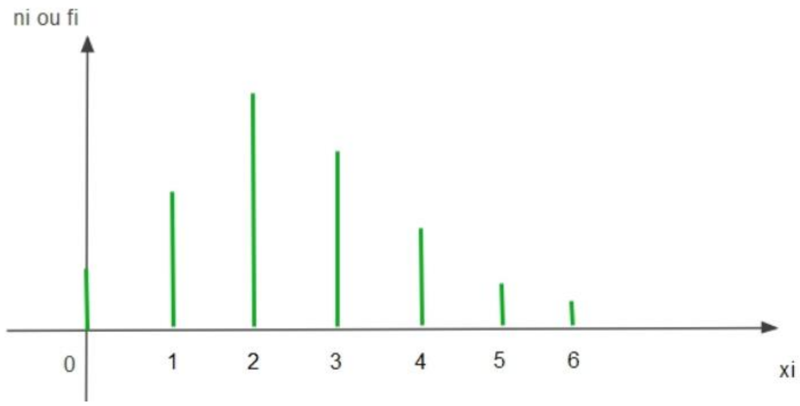
2 Case of a quantitative variable:

a. Discrete variable:

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



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

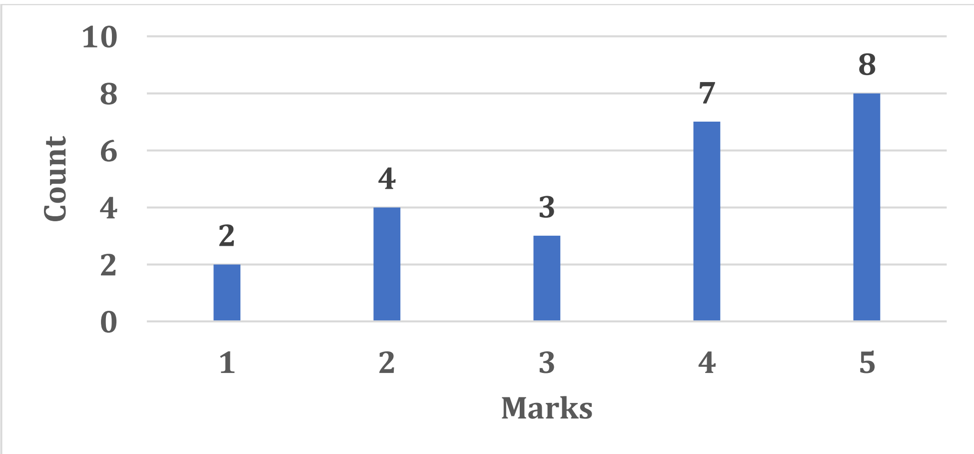


Figure. 2.4: Bar chart shown the marks of a test marked out for a class.

b. Continuous variable:

- **Histogram:**

A histogram is a pictorial representation of the numerical data with rectangular bars.

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: $Ci = \frac{bi+ bi+1}{2}$
- The class size is: $a_i = b_{i+1} - b_i$

Example 17:

We consider the age of the inhabitants in 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

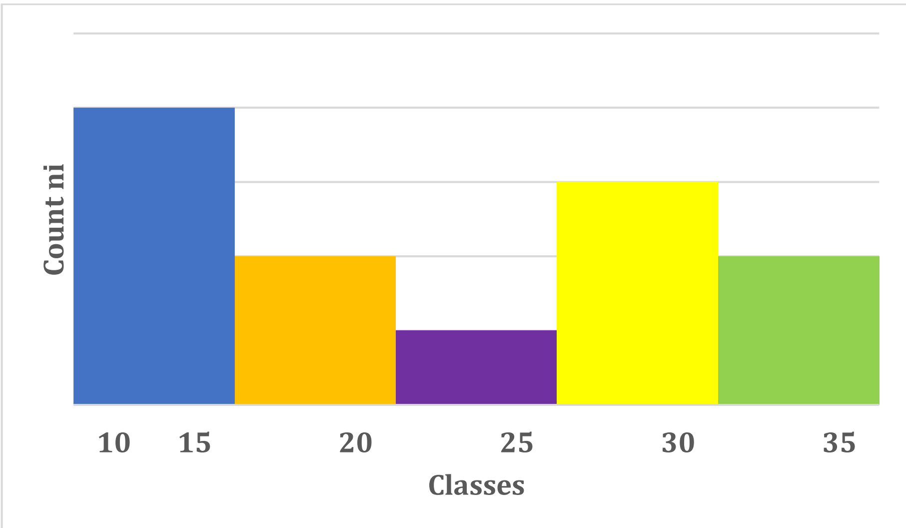


Figure. 2.5: Histogram shown the age of the inhabitants in a neighborhood.

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:

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

- 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: $di = \frac{fi}{ai}$

Example 18:

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)	$\frac{4*20}{4} = 20$	$\frac{4*10}{10} = 4$	$\frac{4*5}{5} = 4$	$\frac{4*15}{7} = 8.5$	$\frac{4*10}{13} = 3$

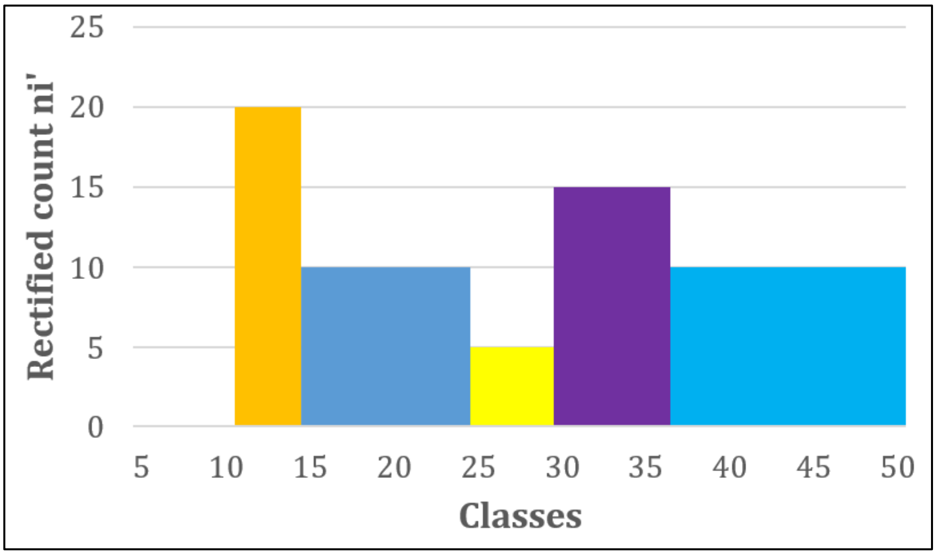


Figure. 2.6: Histogram shown the age of the inhabitants in a neighborhood.

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.

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).

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)

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						

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

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

E: Range, E= 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