**Histograms:**

  A histogram is a chart that plots the distribution of a numeric variable's values as a series of bars.

Process to plot the histograms:

* Sort the numbers
* Create the Bins (Bins means number of group)
* Calculate the bin size (Group size)

Let us take the below example to understand better.

Ages = {10,12,14,18,24,26,32,35,36,37,40,41,42,43,50,51,65,68,78,90,95,100}

Age’s column is continuous variable.

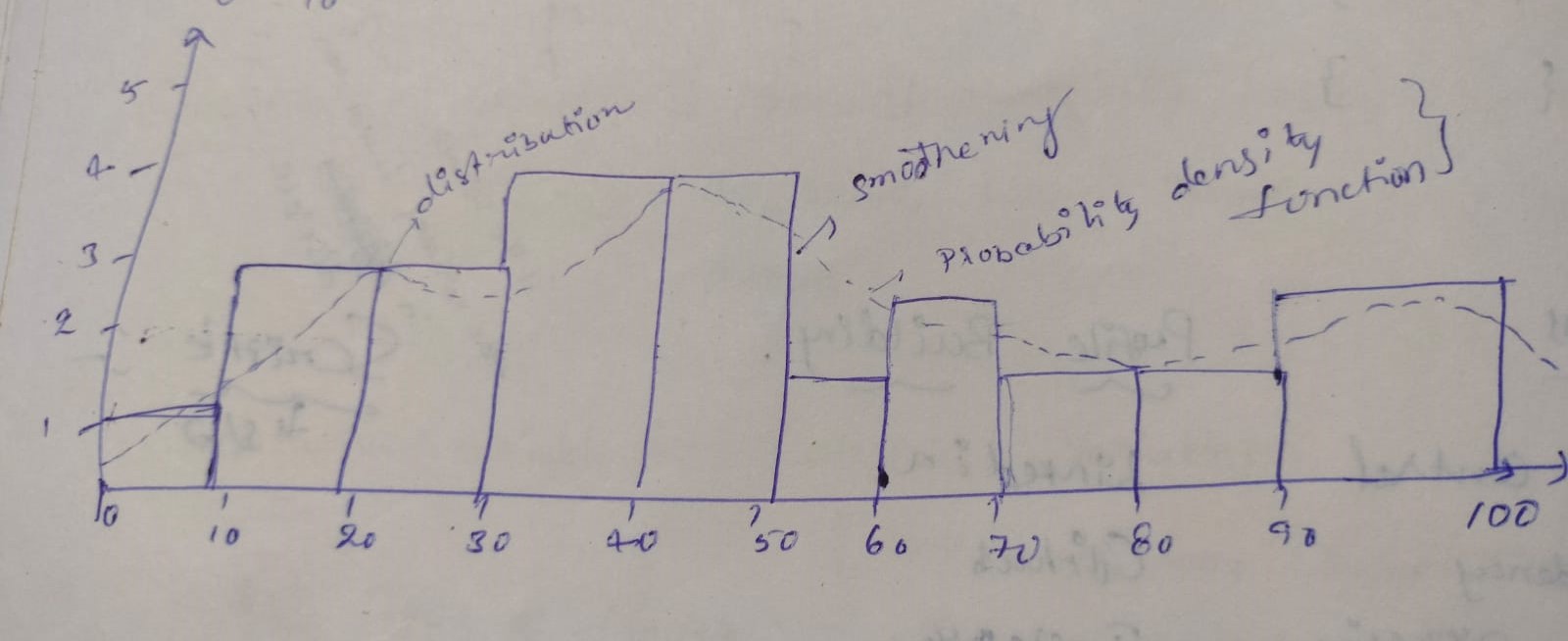
If we want to plot histogram with bin of 10.

Now let us calculate the bin size.

Max value in the dataset-min value in the dataset /number of bins = 100-/10 = 10

Here Bin size is 10.

Histogram is as follows:



Here x-axis represents the bin size intervals, y-axis represents the frequency.

|  |  |
| --- | --- |
| Intervals | Frequency |
| 0-10 | 1 |
| 10-20 | 3 |
| 20-30 | 3 |
| 30-40 | 4 |
| 40-50 | 4 |
| 50-60` | 2 |
| 60-70 | 2 |
| 70-80 | 1 |
| 80-90 | 1 |
| 90-100 | 2 |

For suppose if the numbers of bins are 20. Then we will get the bin size as **100-0/20=5**

**Probability density function (PDF) is used for continuous variable.**

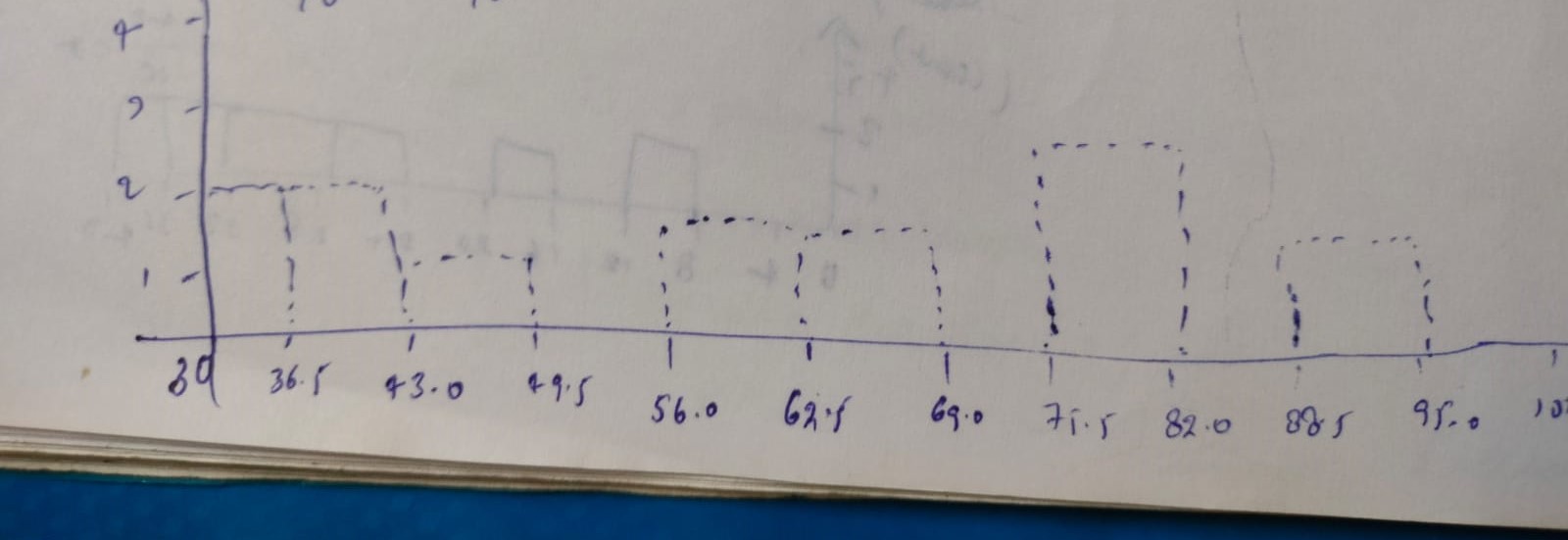
Another scenario to calculate the bin size where we have given the numerical dataset along with number of bins.

Now let us take the example of weights:

Weights = {30,35,38,42,46,58,59,62,63,68,75,77,80,90,95} and bin should be 10.

**Bin size: max-min/bins (95-30/10 = 6.5)**

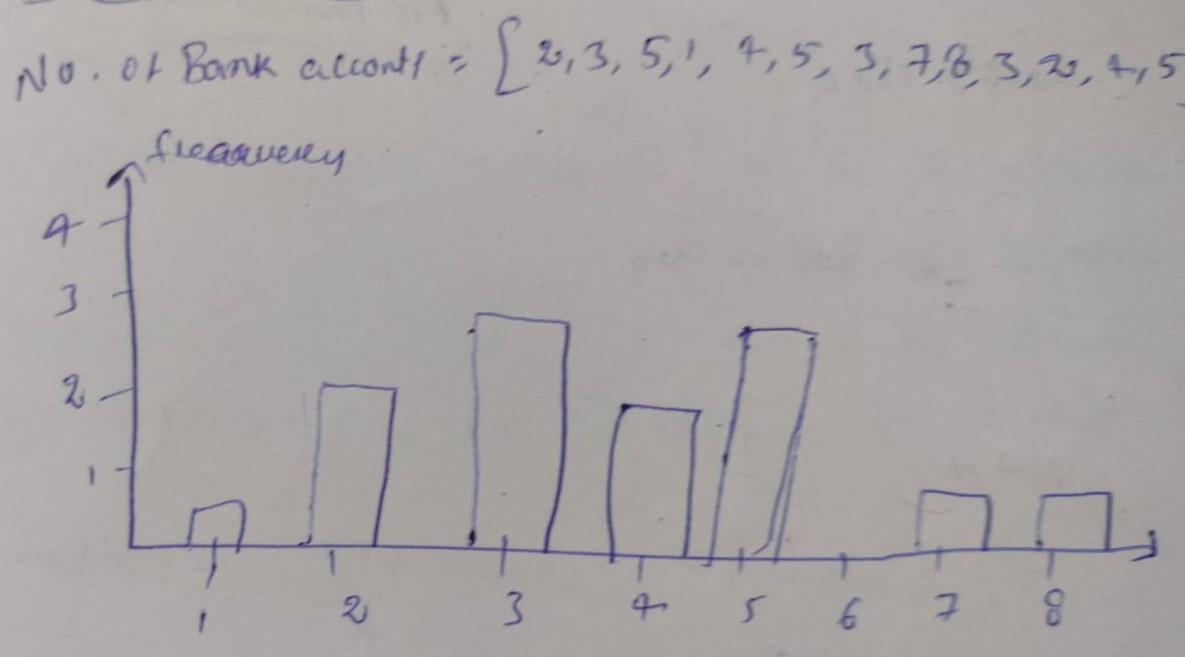
The histogram for this dataset is as follows:



Explanation: If we observe carefully, here we are starting the x-axis from number 30, with the bin size of 6.5.

Let us take another example of **discrete variable histogram:**

**Ex: Number of Bank accounts =[2,3,5,1,4,5,3,7,8,3,2,4,5]**



**Note:** For discrete variables, we use (**Probability Mass Function**) PMF