Class 4
Machine Learning With Python

Grouped Shortcut Method

It is a Short-cut Method for calculating Mean, This method is useful because it can also used for **Open ended** classes also because we do not need to class marks for calculating the **mean.**

$$\text{Mean}(\bar{x}) = a + h \times \frac{\sum f_i d_i}{\sum f_i}$$

Here,
a = assumed mean = 45
h = width of class = 10
d_i = $\frac{(x_i - a)}{h}$
di = we also can say that we just giving code as **0** to the class to which we assuming as **Assumed mean class interval (a),** for above classes of **0** we mark code as **-1,-2,-3** and for lower classes we mark code as 1,2,3 simultaneously.

Marks obtained	Number of students(f_i)	Mid-point (x_i)	$x_i - a = (x_i - 45)$	$d_i = \frac{x_i - a}{h}$	$f_i d_i$
10 – 20	2	$\frac{10+20}{2}$ = 15	15 – 45 = -30	$\frac{-30}{10}$ = -3	2 × -3 = -6
20 – 30	3	15 + 10 = 25	25 – 45 = -20	-2	3 × -2 = -6
30 – 40	8	35	35 – 45 = -10	-1	8 × -1 = -8
40 – 50	14	45	45 – 45 = 0	0	14 × 0 = 0
50 – 60	8	55	55 – 45 = 10	1	8 × 1 = 8
60 – 70	3	65	65 – 45 = 20	2	3 × 2 = 6
70 – 80	2	75	75 – 45 = 30	3	2 × 3 = 6
$\sum f_i$	40				$\sum f_i d_i = 0$

Mean($\bar{x}$) = 45 + 10 x (0/40)
$\bar{x}$ = 45 + 0
$\bar{x}$= 45

ADVANTAGE AND DISADVANTAGES OF MEAN

Advantages

- It is rigidly defined.
- It is easy to calculate and simple to follow.
- It is based on all the observations.
- It is determined for almost every kind of data.
- It is finite and not indefinite.
- It is readily put to algebraic treatment.
- It is least affected by fluctuations of sampling.
- All data is used.
- Easy to understand.
- The concept is familiar and more universally accepted in data science.
- Help in the comparison of 2 or more datasets or 2 or more span of data.

Disadvantages

- The arithmetic mean is highly affected by extreme values.
- It cannot average the ratios and percentages properly.
- It is not an appropriate average for highly skewed distributions.
- It cannot be computed accurately if any item is missing and needs missing value treatment.
- Difficult to calculate for open ended classes.
- The mean sometimes does not coincide with any of the observed value.
- Large data problem.
- Time consuming.
- Problem of percentage and ratio.

Weighted Arithmetic Mean

When calculating the arithmetic mean, the importance of all the items are considered to be equal. However, there may be situations in which all the items under considerations are not of equal importance. For example, when we want to find the average number of marks per students in different subjects like mathematics, statistics, physics and biology. These subjects do not have equal importance. Thus, ***the arithmetic mean computed by considering the relative importance of each item is called the weighted arithmetic mean.***

The weighted arithmetic mean is computed by using the following formula:

$$\overline{X}_w = \frac{\sum wx}{\sum w}$$

$\overline{x}$ = stands for weighted arithmetic mean
x = stands for values of the items and
w = stands for the weight of the item

Example: A student obtained the marks 40, 50, 60, 80, and 45 in math, statistics, physics, chemistry and biology respectively. Assuming weights 5, 2, 4, 3, and 1 respectively for the above mentioned subjects, find the weighted arithmetic mean per subject.

Solution:

Subject	Mark Obtained x	Weight w	wx
Math	40	5	200
Statistics	50	2	100
Physics	60	4	240
Chemistry	80	3	240
Biology	45	1	45
Total		$\sum w = 15$	$\sum wx = 825$

Now we will find the weighted arithmetic mean as:

$$\overline{X}_w = \frac{\sum wx}{\sum w} = \frac{825}{15} = 55 \text{ marks/subject.}$$

Ques : - In what scenario Arithmetic Mean = Weighted Arithmetic Mean.?

Ans :- Arithmetic Mean is equivalent to weighted Arithmetic Mean when the weight are equal.