International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 7 ǁ 2021 ǁ PP. 01-06

Kandari's Formula: a short and easy way to calculate mean of grouped-data

Vaibhav Kandari
The TonsBridge School, Dehradun, Uttarakhand, India
Email: Contact.vaibhavkandari@gmail.com

Abstract
In this research paper, I showcase a formula that I have discovered and named as Kandari's formula. The complete derivation of the formula has been shown in this paper. Kandari's formula is a shorter and an easier way to calculate mean of grouped-data, as compared to the traditional methods. The problem with the traditionally known methods, like Step-deviation method and Assumed mean method, which promise to make complex calculation easy, is that they are quite lengthy. Due to this drawback, one may not be able to use those methods mentally (without pen and paper) to find the mean of the grouped-data involving large numbers. But, Kandari's formula eliminates this drawback. Kandari's formula promises to remove useless complex calculation and shorten the whole process.

Keywords: *Calculating mean of grouped data, Big values.*

Date of Submission: 29-06-2021 Date of acceptance: 13-07-2021

I. INTRODUCTION

The mean (or average) of observations, as we know, is the sum of the values of all the observations divided by the total number of observations. Arithmetic mean is a special case of power mean and is one of the Pythagorean means [1]. We can calculate the mean of grouped and ungrouped data. In this research paper, we will only take grouped data into consideration and not the ungrouped data. In most of our real life situations, data is usually so large that to make a meaningful study it needs to be condensed as grouped data. So, we need to convert given ungrouped data into grouped data and devise some method to find its mean. There are three methods to calculate mean of grouped data, Direct method, Assumed Mean Method and Step-Deviation Method. Direct method is used when the given data is small. Assumed mean method is used for intermediate data and Step deviation method is used for huge data. Step deviation method helps you to make the division and multiplication easier, than the Assumed mean method does. But, these methods cannot be used to calculate mean of grouped data in head (without pen and paper), as they include calculating many columns and values, which must be remembered as they are further used to calculate the mean, the end result. Also, these methods are quite lengthy. This is the situation where Kandari's formula is needed. Kandari's formula is easy to understand, short and very easy to execute, even in head, as compared to the other methods of calculating mean of grouped data. Kandari's formula is also a lot faster and easier in case of huge values, as compared to Step deviation method.

II. SYMBOLS USED

1. $i = i$ is a serial number for class intervals; the value of i varies as we go down through the number of class intervals.
    a. Minimum value of $i = 1$ and maximum value of $i =$ number of class intervals
2. $a =$ assumed mean
3. $k =$ it is that value of i, which defines the position of the class interval whose class-mark is taken as assumed mean
4. $L_i =$ lower limit of class interval i
5. $U_i =$ upper-limit of the class interval i
6. $L_a =$ lower limit of the class interval whose class mark is taken as the assumed mean (a)
7. $U_a =$ upper limit of the class-interval whose lower limit is L_a
8. $h =$ class size
9. $F_i =$ It is the frequency of class interval i. Here, it is also denoted as f_i
10. $x_i =$ class mark of class interval i
11. $d_i =$ deviation of the class interval i
12. $u_i =$ it is the step-deviation of class interval i

13. $\bar{x} =$ it is the mean
14. $\bar{u} = \frac{\Sigma f_i u_i}{\Sigma f_i}$

III. DERIVATION OF KANDARI'S FORMULA

We know that, $d_i = x_i - a$

$$=> \quad d_i = \frac{L_i + U_i}{2} - \frac{L_a + U_a}{2}$$

$$= \frac{L_i + (L_i + h)}{2} - \frac{L_a + (L_a + h)}{2}$$

$$= \frac{2L_i + h}{2} - \frac{2L_a + h}{2}$$

$$= \frac{2L_i + h - (2L_a + h)}{2}$$

$$= \frac{2L_i - 2L_a}{2}$$

$$= L_i - L_a$$

Hence, $d_i = L_i - L_a$

Now, we have, $d_i = L_i - L_a$

$$= L_i - [L_i + (k - i)h]$$

$$= L_i - L_i - (k - i)h$$

$$= -(k - i)h$$

$$\therefore d_i = (i - k)h$$

We also know that, $u_i = \frac{x_i - a}{h} = \frac{d_i}{h}$

$$=> \quad u_i = \frac{(i - k)h}{h} = i - k$$

$$\therefore u_i = i - k$$

This gives us, $F_i u_i = F_i(i - k)$

Since, $F_i u_i = F_i(i - k)$

$$=> \sum_{i=1}^{n} F_i u_i = \sum_{i=1}^{n} F_i(i - k)$$

$$=> \sum_{i=1}^{n} F_i u_i = \sum_{i=1}^{n} (F_i i - F_i k)$$

$$=> \sum_{i=1}^{n} F_i u_i = \sum_{i=1}^{n} F_i i - \sum_{i=1}^{n} F_i k$$

$$=> \sum_{i=1}^{n} F_i U_i = [F_1(1) + F_2(2) + F_3(3) + \cdots + F_n(n)] - [F_1(k) + F_2(k) + F_3(k) + \cdots + F_n(k)]$$

$$= [F_1(1) + F_2(2) + F_3(3) + \cdots + F_n(n)] - k[F_1 + F_2 + F_3 + \cdots + F_n]$$

$$= [F_1(1) + F_2(2) + F_3(3) + \cdots + F_n(n)] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= \left[\sum_{i=1}^{n} F_i + F_2(1) + F_3(2) + \cdots + F_n(n-1)\right] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= \left[\sum_{i=1}^{n} F_1 + \sum_{i=2}^{n} F_i + F_3(2) + \cdots + F_n(n-2)\right] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= \left[\sum_{i=1}^{n} F_1 + \sum_{i=2}^{n} F_i + \sum_{i=3}^{n} F_i + F_4(1) + \cdots + F_n(n-3)\right] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= \left[\sum_{i=1}^{n} F_i + \sum_{i=2}^{n} F_i + \sum_{i=3}^{n} F_i + \sum_{i=4}^{n} F_i + \cdots + \sum_{i=n}^{n} F_i\right] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= \left[\sum_{i=1}^{n} F_i + \left(\sum_{i=1}^{n} F_i - \sum_{i=1}^{1} F_i\right) + \left(\sum_{i=1}^{n} F_i - \sum_{i=1}^{2} F_i\right) + \left(\sum_{i=1}^{n} F_i - \sum_{i=1}^{3} F_i\right) + \cdots + \left(\sum_{i=1}^{n} F_i - \sum_{i=1}^{n-1} F_i\right)\right] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= \left[n\left(\sum_{i=1}^{n} F_i\right) - \sum_{i=1}^{1} F_i - \sum_{i=1}^{2} F_i - \sum_{i=1}^{3} F_i - \cdots - \sum_{i=1}^{n-1} F_i \right] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= \left[n\left(\sum_{i=1}^{n} F_i\right) - \left(\sum_{i=1}^{1} F_i + \sum_{i=1}^{2} F_i + \sum_{i=1}^{3} F_i + \cdots + \sum_{i=1}^{n-1} F_i\right) \right] - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= n\left(\sum_{i=1}^{n} F_i\right) - \left(\sum_{i=1}^{1} F_i + \sum_{i=1}^{2} F_i + \sum_{i=1}^{3} F_i + \cdots + \sum_{i=1}^{n-1} F_i\right) - k\left(\sum_{i=1}^{n} F_i\right)$$

$$= n\left(\sum_{i=1}^{n} F_i\right) - k\left(\sum_{i=1}^{n} F_i\right) - \left(\sum_{i=1}^{1} F_i + \sum_{i=1}^{2} F_i + \sum_{i=1}^{3} F_i + \cdots + \sum_{i=1}^{n-1} F_i\right)$$

$$= (n-k)\left(\sum_{i=1}^{n} F_i\right) - \left(\sum_{i=1}^{1} F_i + \sum_{i=1}^{2} F_i + \sum_{i=1}^{3} F_i + \cdots + \sum_{i=1}^{n-1} F_i\right)$$

$$= (n-k)\left(\sum_{i=1}^{n} F_i\right) - \left[(n-1)F_1 + (n-2)F_2 + (n-3)F_3 + \cdots + \big(n-(n-1)\big)F_{n-1}\right]$$

$$= (n-k)\left(\sum_{i=1}^{n} F_i\right) - \left[(n-1)F_1 + (n-2)F_2 + (n-3)F_3 + \cdots + F_{n-1}\right]$$

$$= (n-k)\left(\sum_{i=1}^{n} F_i\right) - \left\{(n-1)\left(\sum_{i=1}^{n} F_i\right) - \left[(1)F_2 + (2)F_3 + \cdots + (n-3)F_{n-2} + (n-2)F_{n-1} + (n-1)F_n\right]\right\}$$

$$= (n-k)\left(\sum_{i=1}^{n} F_i\right) - (n-1)\left(\sum_{i=1}^{n} F_i\right) + \left[(1)F_2 + (2)F_3 + \cdots + (n-3)F_{n-2} + (n-2)F_{n-1} + (n-1)F_n\right]$$

$$= (1-k)\left(\sum_{i=1}^{n} F_i\right) + \left[(1)F_2 + (2)F_3 + \cdots + (n-3)F_{n-2} + (n-2)F_{n-1} + (n-1)F_n\right]$$

$$= \left[(1)F_2 + (2)F_3 + \cdots + (n-3)F_{n-2} + (n-2)F_{n-1} + (n-1)F_n\right] - (k-1)\left(\sum_{i=1}^{n} F_i\right)$$

$$= \sum_{i=2}^{n} (i-1)F_i - (k-1)\left(\sum_{i=1}^{n} F_i\right)$$

$$\therefore \sum_{i=1}^{n} F_i u_i = \sum_{i=2}^{n} (i-1)F_i - (k-1)\left(\sum_{i=1}^{n} F_i\right)$$

Now, we know that, $\bar{x} = a + h\bar{u}$

$$=> \quad \bar{x} = a + h\left(\frac{\sum_{i=1}^{n} F_i u_i}{\sum_{i=1}^{n} F_i}\right)$$

$$= a + h\left(\frac{\sum_{i=2}^{n}(i-1)F_i - (k-1)(\sum_{i=1}^{n} F_i)}{\sum_{i=1}^{n} F_i}\right)$$

$$= \left(\frac{U_a + L_a}{2}\right) + h\left(\frac{\sum_{i=2}^{n}(i-1)F_i - (k-1)(\sum_{i=1}^{n} F_i)}{\sum_{i=1}^{n} F_i}\right)$$

$$= \left(\frac{U_a + L_a}{2}\right) + \frac{h[\,\sum_{i=2}^{n}(i-1)F_i - (k-1)(\sum_{i=1}^{n} F_i)]}{\sum_{i=1}^{n} F_i}$$

$$= \frac{(U_a + L_a)(\sum_{i=1}^{n} F_i) + 2h[\,\sum_{i=2}^{n}(i-1)F_i - (k-1)(\sum_{i=1}^{n} F_i)]}{2(\sum_{i=1}^{n} F_i)}$$

$$= \frac{U_a(\sum_{i=1}^{n} F_i) + L_a(\sum_{i=1}^{n} F_i) + 2h[\,\sum_{i=2}^{n}(i-1)F_i\,] - 2h(k-1)(\sum_{i=1}^{n} F_i)}{2(\sum_{i=1}^{n} F_i)}$$

$$= \frac{U_a(\sum_{i=1}^{n} F_i) + L_a(\sum_{i=1}^{n} F_i) + 2h[\,\sum_{i=2}^{n}(i-1)F_i\,] - 2(U_a - L_a)(k-1)(\sum_{i=1}^{n} F_i)}{2(\sum_{i=1}^{n} F_i)}$$

$$= \frac{U_a(\sum_{i=1}^{n} F_i) + L_a(\sum_{i=1}^{n} F_i) + 2h[\,\sum_{i=2}^{n}(i-1)F_i\,] - 2k(U_a - L_a)(\sum_{i=1}^{n} F_i) + 2(U_a - L_a)(\sum_{i=1}^{n} F_i)}{2(\sum_{i=1}^{n} F_i)}$$

$$= \frac{U_a(\sum_{i=1}^{n} F_i) + L_a(\sum_{i=1}^{n} F_i) + 2h[\,\sum_{i=2}^{n}(i-1)F_i\,] - 2k(U_a - L_a)(\sum_{i=1}^{n} F_i) + 2U_a(\sum_{i=1}^{n} F_i) - 2L_a(\sum_{i=1}^{n} F_i)}{2(\sum_{i=1}^{n} F_i)}$$

$$= \frac{2h[\sum_{i=2}^{n}(i-1)F_i\,] + 2U_a(\sum_{i=1}^{n}F_i) - 2k(U_a - L_a)(\sum_{i=1}^{n}F_i) + U_a(\sum_{i=1}^{n}F_i) - 2L_a(\sum_{i=1}^{n}F_i) + L_a(\sum_{i=1}^{n}F_i)}{2(\sum_{i=1}^{n}F_i)}$$

$$= \frac{2h[\sum_{i=2}^{n}(i-1)F_i\,] + 2U_a(\sum_{i=1}^{n}F_i) - 2k(U_a - L_a)(\sum_{i=1}^{n}F_i) + U_a(\sum_{i=1}^{n}F_i) - L_a(\sum_{i=1}^{n}F_i)}{2(\sum_{i=1}^{n}F_i)}$$

$$= \frac{2h[\sum_{i=2}^{n}(i-1)F_i\,] + 2U_a(\sum_{i=1}^{n}F_i) - 2k(U_a - L_a)(\sum_{i=1}^{n}F_i) + (U_a - L_a)(\sum_{i=1}^{n}F_i)}{2(\sum_{i=1}^{n}F_i)}$$

$$= \frac{2h[\sum_{i=2}^{n}(i-1)F_i\,] + 2U_a(\sum_{i=1}^{n}F_i) - (U_a - L_a)(\sum_{i=1}^{n}F_i)[2k-1]}{2(\sum_{i=1}^{n}F_i)}$$

$$= \frac{2h[\sum_{i=2}^{n}(i-1)F_i\,] + 2U_a(\sum_{i=1}^{n}F_i)}{2(\sum_{i=1}^{n}F_i)} - \frac{(U_a - L_a)(\sum_{i=1}^{n}F_i)[2k-1]}{2(\sum_{i=1}^{n}F_i)}$$

$$= \frac{h[\sum_{i=2}^{n}(i-1)F_i\,] + U_a(\sum_{i=1}^{n}F_i)}{\sum_{i=1}^{n}F_i} - \frac{(U_a - L_a)[2k-1]}{2}$$

$$= \frac{h[\sum_{i=2}^{n}(i-1)F_i\,] + U_a(\sum_{i=1}^{n}F_i)}{\sum_{i=1}^{n}F_i} - \frac{h[2k-1]}{2}$$

$$= \frac{h[\sum_{i=2}^{n}(i-1)F_i\,]}{\sum_{i=1}^{n}F_i} + \frac{U_a(\sum_{i=1}^{n}F_i)}{\sum_{i=1}^{n}F_i} - \frac{h[2k-1]}{2}$$

$$= \frac{h[\sum_{i=2}^{n}(i-1)F_i\,]}{\sum_{i=1}^{n}F_i} + U_a - \frac{h(2k-1)}{2}$$

$$= h\left(\frac{\sum_{i=2}^{n}(i-1)F_i}{\sum_{i=1}^{n}F_i} - \frac{2k-1}{2}\right) + U_a$$

$$\therefore\ \bar{x} = h\left(\frac{\sum_{i=2}^{n}(i-1)F_i}{\sum_{i=1}^{n}F_i} - \frac{2k-1}{2}\right) + U_a$$

I have named the above formula as, Kandari's Formula, as no other name related to the formula clicked my mind, and so I just named it the way many other discoveries have been named, i.e. naming on the basis of the name of the person who devised it.

Kandari's Formula: $\bar{x} = h\left(\frac{\sum_{i=2}^{n}(i-1)F_i}{\sum_{i=1}^{n}F_i} - \frac{2k-1}{2}\right) + U_a$

IV. EXAMPLES OF KANDARI'S FORMULA

Each example is first solved by the traditional method and then by Kandari's formula.

Example 1: Consider the following distribution of daily wages of 50 workers of a factory. Find the mean daily wages of the workers of the factory by using an appropriate method.

Daily wages	No. of workers
100-120	12
120-140	14
140-160	8
160-180	6
180-200	10

Solution (i): In this case, we can use step-deviation method to make calculations easy. Here, $a = 50$ and $h = 20$.

Serial Number	Class interval	Frequency (f_i)	Class marks (x_i)	$u_i = \frac{x_i - a}{h}$	$f_i u_i$
1	100-120	12	110	-2	-24
2	120-140	14	130	-1	-14
3	140-160	8	$150 = a$	0	0
4	160-180	6	170	1	6
5	180-200	10	190	2	20
		$\sum f_i = 50$			$\sum f_i u_i = -12$

∴ Mean, $\bar{x} = a + h\left(\frac{\sum f_i u_i}{\sum f_i}\right)$

$$= 150 + 20\left(\frac{-12}{50}\right) = 150 - \frac{240}{50} = 150 - 4.8 = 145.2$$

Hence, mean daily wages of the workers is 145.2.

Let us now solve the same question using Kandari's Formula.

Solution (ii): Let $U_a = 160, k = 3$ and $h = 20$.

Note: In the traditional methods of calculating mean of grouped-data, we choose any class mark as assumed mean, randomly. Similarly, while using Kandari's formula, we choose any class mark as assumed mean, but we do not use class mark in calculation, while using Kandari's formula. But, what we do use is the upper limit of the class interval whose class mark is chosen as assumed mean and the serial number of the same class interval. Since Assumed mean is chosen randomly, therefore, we can randomly choose the class interval and take its serial number too, into consideration, while using Kandari's formula.

Serial Number (i)	Class interval	Frequency (f_i)	$(i-1)f_i$
1	100-120	12	(Not required)
2	120-140	14	14
3	140-160	8	16
4	160-180	6	18
5	180-200	10	40
		$\sum f_i = 50$	$\sum_{i=2}^{n}(i-1)F_i = 88$

Mean, $\bar{x} = h\left(\frac{\sum_{i=2}^{n}(i-1)f_i}{\sum_{i=1}^{n} f_i} - \frac{2k-1}{2}\right) + U_a$

$$= 20\left(\frac{88}{50} - \frac{2(3)-1}{2}\right) + 160 = 20(1.76 - 2.5) + 160 = 20(-0.74) + 160 = -14.8 + 160 = 145.2$$

It is clearly visible that Kandari's formula is shorter than the Step-deviation method and easy to follow.

Example 2: The table below gives the percentage distribution of female teachers in the primary schools of rural areas of various states of a country. Find the mean percentage of female teachers.

Percentage of female teachers	Number of states
15-25	6
25-35	11
35-45	7
45-55	4
55-65	4
65-75	2
75-85	1

Solution (i): In this case, we shall use Assumed mean method. Here, $a = 50$ and $h = 10$.

Serial Number	Percentage of female teachers	Number of states (f_i)	Class marks (x_i)	$d_i = x_i - a$	$f_i d_i$
1	15-25	6	20	-30	-180
2	25-35	11	30	-20	-220
3	35-45	7	40	-10	-70
4	45-55	4	50 = a	0	0
5	55-65	4	60	10	40
6	65-75	2	70	20	40
7	75-85	1	80	30	30
		$\sum f_i = 35$			$\sum f_i d_i = -360$

∴ Mean, $\bar{x} = a + \left(\frac{\sum f_i d_i}{\sum f_i}\right)$

$$= 50 + \left(\frac{-360}{35}\right) = 50 - \frac{360}{35} = 50 - 10.29 = 39.71$$

Therefore, the mean percentage of female teachers in the primary schools of rural areas is 39.71.
Solution (ii): Let $U_a = 55,\ k = 4$ and $h = 10$.

Serial Number	Percentage of female teachers	Number of states (f_i)	$(i-1)f_i$
1	15-25	6	(Not required)
2	25-35	11	11
3	35-45	7	14
4	45-55	4	12
5	55-65	4	16
6	65-75	2	10
7	75-85	1	6
		$\sum f_i = 35$	$\sum_{i=2}^{n}(i-1)F_i = 69$

Mean, $\bar{x} = h\left(\frac{\sum_{i=2}^{n}(i-1)f_i}{\sum_{i=1}^{n} f_i} - \frac{2k-1}{2}\right) + U_a$

$$= 10\left(\frac{69}{35} - \frac{2(4)-1}{2}\right) + 55 = 10(1.971 - 3.5) + 55 = 10(-1.529) + 55 = -15.29 + 55 = 39.71$$

It is clearly visible that Kandari's formula is shorter than the Assumed mean method and easy to follow.

V. CONCLUSION

It is well known that, mental calculation helps us improve our cognitive functions [2]. Due to this reason many people, even in the age of computers, keep improving their mental math skills. Kandari's formula enables people to calculate the mean of grouped data, comprising of huge values, in their head. Kandari's formula is faster and easier than any other traditional method of calculating mean of grouped-data. Kandari's formula just requires one column to be calculated, which is much easier and faster as compared to other traditional methods for huge values, where you are required to calculate more than two columns. Now, as there is always some space for improvement in everything, therefore, if you think that there are some essential improvements that could be made, then do explore and work on them, as it is possible that you may come up with a new property or formula or something extraordinary, which could be of great importance to the mathematical world. Embrace these four words: Imperfection, Progression and Indifference, observance.

REFERENCES

[1]. Weisstein, Eric W. "Arithmetic Mean." From MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/ArithmeticMean.html

[2]. Uchida, S., & Kawashima, R. (2008). Reading and solving arithmetic problems improves cognitive functions of normal aged people: a randomized controlled study. Age (Dordrecht, Netherlands), 30(1), 21–29. https://doi.org/10.1007/s11357-007-9044-x