Partners Universal International Research Journal (PUIRJ)
Volume: 01 Issue: 03 | July-September 2022 www.puirj.com

Integral Valued Numerical Data: Measure of Central Tendency

Dhritikesh Chakrabarty

Independent Researcher, Ex Associate Professor of Statistics, Handique Girls' College, Assam, India

---***---------------------------------------

Abstract - *In the situation where the data set to be studied consists of integral valued numerical observations, the existing measures of central tendency may fail to provide the value, which is a valid one, of central tendency since the values provided by them are not bound to be integers. A valid measure of the same, in the case of such data, has been derived in this study. This article describes the derivation of the measure and its application in the data on number of rainy days at New Delhi.*

Keywords: Data, integral valued observations, central tendency, measure.

1. INTRODUCTION

The following abbreviations have been used in this article:

Table - 1

Abbreviation	Meaning
AM	Arithmetic Mean
AGM	Arithmetic-Geometric Mean
AGHM	Arithmetic-Geometric-Harmonic Mean
AHM	Arithmetic- Harmonic Mean
ANoRD	Annual Number of Rainy Days
CT	Central Tendency
GHM	Geometric-Harmonic Mean
GM	Geometric Mean
HM	Harmonic Mean
MoCT	Measure of Central Tendency
MMoCT	Mathematical Measure of Central Tendency
NoRD	Number of Rainy Days

Average is a characteristic of a list of numbers/numerical values while central tendency [19 , 20 , 25] is that of data on a variable.

Measure of central tendency [26] is based on measure of average [3 , 4 , 5 , 6 ,7 , 8 , 10 , 18].

The three common MMoCT of numerical data are AM, GM and HM, which are also known as Pythagorean Means [2].

Recently, some more MMoCT of data have been developed which are respectively AGM [1 , 9 , 14 , 17 , 20], AHM [10 , 11 , 14 , 15 , 16 , 17 , 20 , 21], GHM [12 , 14 , 17 , 20 , 24] and AGHM [13 , 14 , 17 , 20]. However, data to be studied may be of different types. There are many cases where numerical data to be studied are of integral valued so that their central value is also an integral value. In such situation, the above measures fail to provide the central value, which is a valid one, of central tendency since the values provided by them are not bound to be integers. One MoCT has therefore been developed, in this study, which is valid in such situation. This article describes the derivation of the measure and its application in the data on number of rainy days at New Delhi.

2. DERIVATION OF THE MEASURE

Let

$x_1, x_2, \ldots\ldots\ldots, x_N$

be N observations which are integral values observed on an integral valued random variable X, whose central value (central tendency) is μ.

Then each observation x_i is composed of two components; one is μ and the other is ε_i (say).

This means, the observations follow the mathematical model

x_i = μ + ε_i, (i = 1,2,..........,N) (1)

where

$\varepsilon_1, \varepsilon_2, \ldots\ldots\ldots, \varepsilon_N$

are the respective error components of the N observations which assume positive integral values and negative integral values in (– ∞ , ∞) in random order.

Also,

P (ε_i assumes positive value) = P (ε_i assumes negative value)

which means,

P (ε_i = n)= P(ε_i = –n), n is integer (2)

Thus,

$$E(\varepsilon_i) = \sum_{n=-\infty}^{\infty} n\, P(\varepsilon_i = n)$$

$$= \sum_{n=-\infty}^{-1} n\, P(\varepsilon_i = n) + 0\, P(\varepsilon_i = 0) + \sum_{n=1}^{\infty} n\, P(\varepsilon_i = n)$$

$$= \sum_{n=1}^{\infty} (-n)\, P(\varepsilon_i = -n) + \sum_{n=1}^{\infty} n\, P(\varepsilon_i = n)$$

$$= \sum_{n=1}^{\infty} (-n+n)\, P(\varepsilon_i = n)$$, since P(ε_i = n)=P(ε_i = – n)

Therefore,

E (ε_i)= 0 (3)

Now from equation (1),

A (x : N) = μ + A (ε : N)

where

$$A(x:N) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\& \; A(\varepsilon:N) = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i$$

By law of large numbers [22 , 23],

$A(\varepsilon:N) = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i$ converges in probability to 0 as N → ∝ (4)

which implies,

$A(x:N) = \frac{1}{N}\sum_{i=1}^{N} x_i$ converges in probability to μ as N → ∝ (5)

This result can also be obtained as a consequence of the law of large numbers together with the fact that

E(x_i)= μ ,by equations(1)&(3) (6)

Since $A(x:N) = \frac{1}{N}\sum_{i=1}^{N} x_i$ converges to μ as the data size N → ∝ ,

therefore for data of finite size, the integral value nearest to A(x : N) will be the value of μ i.e. the value of central tendency of the observations.

3. NUMERICAL EXAMPLE

Data on number of rainy days at New Delhi corresponding to different months during the period from 1969 to 2001 (years have been written in the table as 69 , 70 , 71, ..., 99, 100, 101), as shown in Table - 2, have been collected from Indian Meteorological Department

It is observed that the period June – September is the most rainy period in New Delhi compared to the other two periods October – January & February – May. Accordingly, data corresponding to these three periods, shown in Table - 3 which have extracted from the data in Table - 2, have been considered for study.

Computational values A (x : N) for different values of N have been presented in Table - 4, Table - 5, Table - 6 & Table - 7 respectively.

Table - 2: NoRD

Year → Month↓	69	70	71	72	73	74
Jan	0	3	1	1	2	0
Feb	1	4	1	1	1	0
Mar	2	2	0	0	0	1
Apr	0	0	1	2	0	0
May	3	1	3	0	2	0
June	5	5	4	2	2	4
July	11	5	6	10	8	14
Aug	6	11	13	13	18	5
Sept	6	9	6	3	3	1
Oct	0	1	1	1	2	3
Nov	0	0	0	2	0	0
Dec	0	0	0	0	1	1

Table - 2 Continued: NoRD

Year→ Month↓	75	76	77	78	79	80
Jan	2	0	4	0	4	1
Feb	1	2	0	2	4	1
Mar	1	2	0	3	2	3
Apr	0	1	5	1	1	0
May	0	3	3	0	2	2
June	6	6	8	5	4	3
July	11	11	21	17	11	18
Aug	14	17	4	14	7	7
Sept	11	2	11	5	3	4
Oct	3	0	2	0	1	2
Nov	0	0	0	0	0	0
Dec	0	0	2	1	2	1

Table - 2 Continued: NoRD

Year→ Month↓	81	82	83	84	85	86
Jan	2	2	3	0	0	1
Feb	3	2	2	1	0	2
Mar	3	5	2	0	0	2
Apr	0	4	6	1	3	1
May	0	4	3	1	0	2
June	2	4	3	4	5	3
July	12	7	9	11	17	4
Aug	4	13	10	12	10	7
Sept	2	0	10	4	5	4
Oct	0	4	1	0	5	0
Nov	4	1	0	0	0	0
Dec	0	2	2	0	6	2

Table - 2 Continued: NoRD

Year→ Month↓	87	88	89	90	91	92
Jan	1	0	2	0	0	2
Feb	2	3	1	5	2	
Mar	1	4	3	1	1	
Apr	1	1	0	1	0	
May	5	1	0	0	0	0
June	3	8	4	4	1	1

July	5	14	6	11	5	11
Aug	5	13	6	8	11	
Sept	3	5	4	9	3	
Oct	0	0	1	0	0	
Nov	0	0	1	2	2	
Dec	1	1	1	1	3	

Table - 2 Continued: NoRD

Year→ Mont↓	93	94	95	96	97	98
Jan	2	2	2	1	2	0
Feb	1	1	4	2	0	2
Mar	1	0	3	0	2	3
Apr	1	2	0	1	2	1
May	1	3	0	2	4	2
June	5	6	3	6	6	5
July	9	18	4	13	8	10
Aug	5	10	18	14	11	12
Sept	10	0	5	9	3	7
Oct	0	0	0	1	4	3
Nov	0	0	0	0	1	1
Dec	0	0	0	0	4	0

Table - 2 Continued: NoRD

Year → Month↓	99	100	101
Jan	3	3	1
Feb	0	2	1
Mar	0	2	1
Apr	0	0	2
May	2	2	6
June	5	7	9
July	7	11	8
Aug	3	9	6
Sept	4	2	2
Oct	2	0	1

Table – 3:

Year	**NoRD during**			
	The year	**June – Sept**	**Oct – Jan**	**Feb – May**
69	34	28	0	6
70	41	30	4	7
71	36	29	2	5
72	35	28	4	3
73	39	31	5	3
74	29	24	4	1
75	49	42	5	2
76	44	36	0	8
77	60	44	8	8
78	48	41	1	6
79	41	25	7	9
80	42	32	4	6
81	32	20	6	6
82	48	24	9	15
83	51	32	6	13
84	34	31	0	3
85	51	37	11	3
86	28	18	3	7
87	27	16	2	9
88	50	40	1	9

89	29	20	5	4
90	42	32	3	7
91	28	20	5	3
92	---	---	--	---
93	35	29	2	4
94	42	34	2	6
95	39	30	2	7
96	49	42	2	5
97	47	28	11	8
98	46	34	4	8
99	26	19	5	2
100	38	29	3	6
101	37	25	2	10

Table – 4:

(1) Serial No(N)	**(2) NoRD (x_i) during The Year**	**(3) Value of A(x : N) = $\frac{1}{N}\sum_{i=1}^{N} x_i$**
1	34	34
2	41	37.5
3	36	37
4	35	36.5
5	39	37
6	29	35.6666666666666
7	49	37.57142857142857
8	44	38.375
9	60	40.77777777777777
10	48	41.5
11	41	41.45454545454545
12	42	41.5
13	32	40.76923076923076
14	48	41.28571428571428
15	51	41.93333333333333
16	34	41.4375
17	51	42
18	28	41.22222222222222
19	27	40.47368421052631
20	50	40.95
21	29	40.38095238095238
22	42	40.45454545454545
23	28	39.91304347826086
24	35	39.70833333333333
25	42	39.8
26	39	39.76923076923076
27	49	38.67857142857142
28	47	40.35714285714285
29	46	40.55172413793103
30	26	40.0666666666666
31	38	40
32	37	39.90625

Table – 5:

(1) Serial No(N)	**(2) NoRD (x_i) during June – Sept**	**(3) Value of A(x : N) = $\frac{1}{N}\sum_{i=1}^{N} x_i$**
1	28	28
2	30	29
3	29	29
4	28	28.75

5	31	29.2
6	24	28.333333333333333
7	42	30.28571428571428
8	36	31
9	44	32.44444444444444
10	41	33.3
11	25	32.54545454545454
12	32	32.5
13	20	31.53846153846153
14	24	31
15	32	31.06666666666666
16	31	31.0625
17	37	31.41176470588235
18	18	30.66666666666666
19	16	29.89473684210526
20	40	30.4
21	20	29.9047619047619047
22	32	30
23	20	29.5652173913043478
24	29	29.5416666666666666
25	34	29.72
26	30	29.7307692307692307
27	42	30.1851851851851851
28	28	30.1071428571428571
29	34	30.2413793103448275
30	19	29.8666666666666666
31	29	29.8387096774193548
32	25	29.6875

Table – 6:

(1) **Serial No (N)**	**(2)** **NoRD (x_i) during Oct – Jan**	**(3)** **Value of A (x : N) =** $\frac{1}{N}\sum_{i=1}^{N} x_i$
1	0	0
2	4	4
3	2	3.5
4	4	2.5
5	5	3
6	4	3.16666666666666666
7	5	3.42857142857142857
8	0	3
9	8	3.55555555555555555
10	1	3.3
11	7	3.63636363636363636
12	4	3.66666666666666666
13	6	3.84615384615384615
14	9	4.21428571428571428
15	6	4.33333333333333333
16	0	4.0625
17	11	4.47058823529411764
18	3	4.38888888888888888
19	2	4.26315789473684210
20	1	4.1
21	5	4.14285714285714285
22	3	4.09090909090909090
23	5	4.13043478260869565
24	2	4.04166666666666666
25	2	3.96
26	2	3.88461538461538461

27	2	3.814814814814814 81
28	11	4.0714285714285714 2
29	4	4.068965517241379 31
30	5	4.1
31	3	4.064516129032258 06
32	2	4

Table – 7:

(1) Serial No(N)	(2) NoRD (x_i) during Feb – May	(3) Value of A (x : N) = $\frac{1}{N}\sum_{i=1}^{N} x_i$
1	6	6
2	7	6.5
3	5	6
4	3	5.25
5	3	4.8
6	1	4.1666666666666666
7	2	3.857142857142857 14
8	8	4.375
9	8	4.7777777777777777
10	6	4.9
11	9	5.2727272727272727 2
12	6	5.3333333333333333 3
13	6	5.384615384615384 61
14	15	6.0714285714285714 2
15	13	6.5333333333333333 3
16	3	6.3125
17	3	6.117647058823529 41
18	7	6.1666666666666666 6
19	9	6.315789473684210 52
20	9	6.45
21	4	6.3333333333333333 3
22	7	6.3636363636363636 3
23	3	6.217391304347826 08
24	4	6.125
25	6	6.12
26	7	6.153846153846153 84
27	5	6.1111111111111111
28	8	6.1785714285714285 7
29	8	6.241379310344827 58
30	2	6.1
31	6	6.096774193548387 09
32	10	6.21875

4. FINDINGS AND CONCLUSION

It is found that the computed values in column (3) of Table - 4 shows that the trend of arithmetic mean of NoRD during the year is to converge to a value whose nearest integral value is 40. Accordingly, the central value of ANoRD during at New Delhi is 40 days.

Similarly, computed values in column (3) of Table - 5 show that the arithmetic mean of NoRD during the period June – Sept converge to a value whose nearest integral value is 30. Accordingly, the CT of NoRD at New Delhi during the period June – Sept can be regarded as 30 days.

Similarly, from the computed values in column (3) of each of the two tables namely Table - 6 & Table - 7 it is obtained that the CT of NoRD at New Delhi during the period October – January is 4 days and during the period February – May is 6 days.

It has already been shown in section 2 that A(x : N) given by

A(x : N) = $\frac{1}{N}\sum_{i=1}^{N} x_i$

is to converge to a an integral value which is the central value of the observations.

However, the value of A (x : N) for sample of finite size (though large) may not be an integral value. In his case, the integral value nearest to A(x : N) will be central value of the observations.

REFERENCES

[1] David A. Cox (2004): "The Arithmetic-Geometric Mean of Gauss", In J. L. Berggren; Jonathan M. Borwein; Peter Borwein (eds.). Pi: A Source Book. Springer. p. 481. ISBN 978-0-387-20571-7, (first published in L'Enseignement Mathématique, t. 30 (1984), 275 – 330).

[2] David W. Cantrell (……): "Pythagorean Means", Math World.

[3] Dhritikesh Chakrabarty (2018): "Derivation of Some Formulations of Average from One Technique of Construction of Mean", American Journal of Mathematical and Computational Sciences, 3(3), 62 – 68. http://www.aascit.org/journal/ajmcs .

[4] Dhritikesh Chakrabarty (2018): "One Generalized Definition of Average: Derivation of Formulations of Various Means", JECET, Section C, (E-ISSN: 2278 – 179 X),7(3), 212 – 225.www.jecet.org.

[5] Dhritikesh Chakrabarty (2018): "General Technique of Defining Average", NaSAEAST-2018, Abstract ID: CMAST_NaSAEAST-1801(I). https://www.researchgate.net/profile/Dhritikesh_Chakrabarty/stats .

[6] Dhritikesh Chakrabarty (2019): "One General Method of Defining Average: Derivation of Definitions/Formulations of Various Means", JECET, Section C, (E-ISSN : 2278 – 179 X), 8(4), 327 – 338. www.jecet.org .

[7] Dhritikesh Chakrabarty (2019): "A General Method of Defining Average of Function of a Set of Values", ABJMI {ISSN (Print) : 0975-7139, ISSN (Online) : 2394-9309}, 11(2), 269 – 284. www.abjni.com .

[8] Dhritikesh Chakrabarty (2020): "Definition / Formulation of Average from First Principle", JECET, Section C, (E-ISSN : 2278 – 179 X), 9(2), 151 – 163. www.jecet.org .

[9] Dhritikesh Chakrabarty (2020): "AGM: A Technique of Determining the Value of Parameter from Observed Data Containing Itself and Random Error", JECET, Section C, (E-ISSN : 2278 – 179 X), 9(3), 473 – 486. [DOI: 10.24214/jecet.C.9.3.47386]. www.jecet.org .

[10] Dhritikesh Chakrabarty (2020): "AHM: A Measure of the Value of Parameter μ of the Model X = μ +ε", IJARSET, (ISSN : 2350 – 0328), 7(10), 15268 – 15276. www.ijarset.com .

[11] Dhritikesh Chakrabarty (2020): "Arithmetic-Harmonic Mean: Evaluation of Parameter from Observed Data Containing Itself and Random Error", IJEAR (ISSN : 2395 – 0064), 7(1), 29 – 45. http://eses.net.in/online_journal.html .

[12] Dhritikesh Chakrabarty (2020): "Determination of the Value of Parameter μ of the Model X = μ + ε by GHM", IJARSET, (ISSN : 2350 – 0328), 7(11), 15801 – 15810. www.ijarset.com.

[13] Dhritikesh Chakrabarty (2020): "AGHM as A Tool of Evaluating the Parameter from Observed Data Containing Itself and Random Error", IJEAE (ISSN : 2395 – 0064), 7(2), 05 – 23. http://eses.net.in/online_journal.html .

[14] Dhritikesh Chakrabarty (2021): "AGM, AHM, GHM & AGHM: Evaluation of Parameter μ of the Model X = μ + ε ",IJARSET, (ISSN : 2350 – 0328), 8(2), 16691 – 16699. www.ijarset.com

[15] Dhritikesh Chakrabarty (2021): "AHM as A Measure of Central Tendency of Sex Ratio", BBIJ, (ISSN : 2350 – 0328), 10(2), 50 – 57. DOI: 10.15406/bbij.2021.10.00330 . http://medcraveonline.com .

[16] Dhritikesh Chakrabarty (2021): "Arithmetic-Harmonic Mean: A Measure of Central Tendency of Ratio-Type Data", IJARSET, (ISSN : 2350 – 0328), 8(5), 17324 – 17333. www.ijarset.com .

[17] Dhritikesh Chakrabarty (2021): "Four Formulations of Average Derived from Pythagorean Means", IJMTT (ISSN: 2231 – 5373), 67(6), 97 – 118. doi:10.14445/22315373/IJMTT-V67I6P512. http://www.ijmttjournal.org .

[18] Dhritikesh Chakrabarty (2021): "Recent Development on General Method of Defining Average: A Brief Outline", IJARSET, (ISSN : 2350 – 0328), 8(8), 17947 – 17955 www.ijarset.com .

[19] Dhritikesh Chakrabarty (2021): "Model Describing Central Tendency of Data", IJARSET, (ISSN : 2350 – 0328), 8(9), 18193 – 18201. www.ijarset.com .

[20] Dhritikesh Chakrabarty (2021): "Measuremental Data: Seven Measures of Central Tendency", IJEAR (ISSN : 2395 – 0064), 8(1), 15 – 24. www.eses.net.in .

[21] Foster D. M. E. and Phillips G. M. (1984): "The Arithmetic-Harmonic Mean", Journal of American Mathematical Society, 42(165), 183-191.

[22] Sen P. K , Singer J. M. (1993).: "Large sample methods in statistics", Chapman & Hall.

[23] Seneta Eugene (2013): "A Tercentenary history of the Law of Large Numbers",

[24] Weisstein, Eric W. ( ): "Harmonic-Geometric Mean", From MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/Harmonic-GeometricMean.html .

[25] Weisberg H. F. (1992): "Central Tendency and Variability", Sage University Paper Series on Quantitative Applications in the Social Sciences, ISBN 0-8039-4007-6 p.2.
[26] Williams R. B. G. (1984): "Measures of Central Tendency", Introduction to Statistics for Geographers and Earth Scientist, Soft cover ISBN978-0-333-35275-5, eBook ISBN978-1-349-06815-9 , Palgrave, London, 51 – 60.