12-1969

The Mean in Statistics

Rex E. Babcock
Ouachita Baptist University

THE MEAN IN STATISTICS

Course Number

H71

by

Rex E. Babcock

Instructor

Miss Jones

December 15, 1969

THE MEAN IN STATISTICS

When one speaks of how statistics have been utilized in decades past there is much to say. However, it is even more astonishing when one speaks of the utilization of statistics today. The use of statistics is widespread virtually over all of the world today, and is put to use in such fields as surveys, census, psychometrics, and almost every business, vocation, and employment that may come to one's mind.[1] But I wish to confine myself to the mean in statistics; I will even make another boundary on myself by discussing only the arithmetic mean, geometric mean, harmonic mean, and quadratic mean. And the quadratic mean I will only discuss briefly, for it is not used to the extent as the other three.

Arithmetic Mean

When people see the words 'arithmetic mean' they have a great tendency to tell theirselves they do not have any acknowledgment of the words. But I am sure that most of them do know the meaning of the words. I think one of the better definitions I have found for them is "the arithmetic mean, which is sometimes described as a number that is typical of the whole group; that is, it is representative

[1]John G. Peatman, Introduction to Applied Statistics, Harper & Row Publishers, New York, New York, 1963, pp. 9-12.

of what is frequently called the central tendency."[2] And thus one may see that the arithmetic mean is nothing more than the average of a group of numbers.

If one would wish to find the arithmetic mean of a group of numbers he would do it as if he were finding the average; that is, divide the sum of the numbers by how many numbers there are. Thus the arithmetic mean of 36, 5, and 16 is 19 as seen in Table I. And

Table I

Σx	N
36	1
5	1
+16	+1
57	3

The arithmetic mean is equal to $57 \div 3 = 19$

following this on through one may come up with the equation:

$$\frac{\Sigma x}{N} = M$$

Σx is the sum of the numbers, N is the number or how many numbers there are, and M is the arithmetic mean.[3] Thus the definition and the equation for arithmetic mean has been given, and now one may go on to deviations from the arithmetic mean.

One law or property that the arithmetic mean has is that it is the same numerical distance from the sum of the numbers below

---

[2]George R. Davies, and Dali Yoder, Business Statistics, John Wiley & Sons Inc., New York, New York, 1949, p. 87.

[3]Ibid.

it and the sum of the numbers above it. Each separate distance fron a number to the mean is called the deviation from the mean, and is denoted by (d). We shall let the numbers greater than the arithmetic mean have a positive deviation and the numbers less than the arithmetic mean have a negative deviation. And when adding all the deviations together their sum should be zero. If their sum is not zero then the mean used in finding the deviations was not the actual arithmetic mean. For example, the deviation of the numbers 36, 5, and 16 where 19 is the arithmetic mean are as follows:

Table II

x	M	d
5	19	-14
16	19	- 3
+36	+19	+17
57	57	0

And from this another more specific definition of the arithmetic mean may be stated. This new definition would be that the arithmetic mean is the number if used to replace each number of a set of numbers the sum will be equal to the sum of the numbers in the set. In this case the set is 36, 5, and 16. And their sum is 57; when 19 is put in for each of the numbers the sum is also 57 so thus 19 is the actual arithmetic mean of this set.

A short-cut method of finding the arithmetic mean may be found by using the idea that the sum of the deviations must be zero for the arithmetic mean used to be the actual arithmetic mean. This short-cut method is started by roughly guessing what the mean is.

This first approximate value of the arithmetic mean (A) is used as the origin for finding the deviations of the numbers in the set. These deviations (d') found from the approximate arithmetic mean are averaged and then divied by N to find a correction factor (f). Then the actual arithmetic mean is found by adding the correction factor to the approximate arithmetic mean. For the number set 36, 5, and 16 one may guess the arithmetic mean is 21. So to find the actual arithmetic mean would be as such:

Table III

x	A	d'	N	
5	21	-16	1	$f = \frac{-6}{3} = -2$
16	21	- 5	1	
36	21	+15	+1	$M = 21 + (-2) = 19$
		- 6	3	

The equation for this method may be written as:

$$M = A + f$$
$$= A + (\Sigma d' \div N)$$

M is the actual arithmetic mean, A is the approximate arithmetic mean value, f is the correction factor, $\Sigma d'$ is the sum of the deviations found from A, and N is how many numbers there are in the set.

This short-cut method of finding the arithmetic mean may not seem like a shorter method as seen in Table III; however, when working with many numbers of fractions this method is somewhat faster and perhaps easier than the direct method as demonstrated in Table I. The short-cut method is also quite useful when working with data that are grouped in the form of a frequency distribution. The finding of this type of arithmetic meam in frequency distribution is exhibited in Table IV.

Table IV

G	m	F	d'	F·d'	
10-15	13	1	-10	-10	
15-20	17	5	- 6	-30	
20-25	23	10	0	0	
25-30	27	+14	+ 4	+56	
		N = 30		30 ⌊16	
				.53...	= f
				+23.00	= A
				23.53...	= M

Table V

G	m	F	m·F	
10-15	13	1	13	
15-20	17	5	285	
20-25	23	10	230	
25-30	27	+14	+378	
		N = 30	30 ⌊706	
			23.53....	= M

In explaining Table IV, I picked a college class that had an enrollment of 30 students, and then I counted the student attendance for 30 class meetings. The attendance was divided into four groups; between 10 and 15 students attending, between 15 and 20 students attending, between 20 and 25 students attending, and between 25 and 30 students attending. G denotes each of the four groups, F represents the frequency or how many times each of the four groups happened in the 30 class meetings, m is the mid-point of each individual group from which the d' is determined, and d' is the deviation from m to the approximate arithmetic mean (A). And in this case I picked A to equal 23. It is seen then that in Table IV the arithmetic mean is found by adding the products of F and d' then divide by the number of class sessions (N), and this gives the correction factor (f). The correction factor is then added to the approximated arithmetic mean which gives the actual arithmetic mean. And the equation that may be written to express this is:

$$M = \frac{\Sigma(F \cdot d')}{N} + A$$

Table V is finding the arithmetic mean for the same data only by the direct method. The products of the mid-points and frequencies are added together then divided by the number of class sessions. This method gives the same arithmetic mean as the short-cut method, and the equation that will express this direct method is:

$$M = \frac{\Sigma(m \cdot F)}{N} + A$$

Again in this case the direct method looks shorter than the short-cut method, but the more frequencies that are used in a problem of this sort the harder the direct method becomes and the easier the short-cut method becomes.

Geometric Mean

The geometric mean is very useful as is the arithmetic mean is in statistics, but it seems to be so much more troublesome than the arithmetic mean. I feel this trouble comes to the geometric mean because it is so closely related to logarithms, and as we all know logarithms 'ain't no fun.'

I would say there are many ways to approach the geometric mean, but basically "the geometric mean may be defined as the nth root of the product of n items or values."[4] The geometric mean may also be thought of as "calculated in a manner similar to that by which the arithmetic mean is discovered except that the measures are transferred to the geometric scale by using their logarithms instead of the measures themselves."[5] What this last statement is saying is that much or all of the multiplication and division that is needed to work with the geometric mean may be done by logarithms. The equation used to describe the geometric mean, where GM is the geometric mean, x is the variate, and n is the number of variates, is expressed as such:[6]

$$GM = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \ldots x_n}$$

If looking for ways to use the geometric mean in statistics two ways come to me quickly; when items are considered as factors or as ratios. A very simple example of this would be to find the geometric mean of 1, 4, and 16. This calculation is shown in Table VI-A using long division and multiplication while Table VI-B is using logarithms for its division and multiplication.

Table VI

A. GM of 1, 4, and 16 is by definition equal to $\sqrt[3]{1 \cdot 4 \cdot 16} = \sqrt[3]{64} = 4$

---

[4] Ya-Lun Chou, Applied Business and Economic Statistics, Holt, Rinehart, and Winston, New York, New York, 1964, p. 155.

[5] Davies, and Yoder, op. cit., pp. 94-95.

[6] A. C. Rosander, Elementary Principles of Statistics, D. Van Nostrand Company Inc., Princeton, New Jersey, 1957, p. 70.

Table VI

$$\begin{array}{rl} \text{B. } \log 1 = & .0000 \\ \log 4 = & .6021 \\ \log 16 = & 1.2041 \\ 3\,\overline{)\,1.8020} & \\ & 0.6020 \end{array}$$

antilog of 0.6020 = 4 = GM

One may ask how to put this to use. Taking the three numbers 1, 4, and 16, and use them to make a box. The volume of this box is 64 cu. units, and it may be said that the average diminsions of the box is 4 units by 4 units by 4 units. Now, it may be seen that the geometric mean is very closely related to the arithmetic mean, because the geometric mean is the number used to replace 1, 4, and 16 without changing the results of the problem.

A third way in which the geometric mean is utilized in statistics is when an average rate of successive increases and decreases is required. Say that the population of a city grew 40% in one decade, 10% in the next decade, but in the next decade it declined 10%, and the rate of increase per decade is required for the three decades. Let I stand for the initial population; so the population at the end of the three decades is equal to $I \cdot 1.40 \cdot 1.10 \cdot .90 = 1.386\ I$.

In this case one plus each rate $(1 + r)$ taking the rate with its algebraic sign, are factors, respectively, in the final products. The geometric mean of $(1 + r)$ is $\sqrt[3]{1.40 \cdot 1.10 \cdot .90} = \sqrt[3]{1.386} = 1.114$; this would be saying that the average rate of increase is 0.114 or 11.4%.

The geometric mean like the arithmetic mean may be used in ungrouped and grouped data; whereas in the grouped data frequency is used.

As a rule any roots passed the cube root should be handled with logarithms because of the actual math involved. I am sure one may see the obvious advantages he would have if he knows how to use logarithms when trying to find such values as $\sqrt[7]{34,663}$ and $\sqrt[9]{234,112}$.

In Table VII the ungrouped method is used to find the geometric mean while in Table VIII the grouped data method is used.

Table VII

x	N	log x
15	1	1.1761
21	1	1.3222
27	1	1.4314
31	1	1.4914
35	+1	+1.5441
	5	5 | 6.9652

log GM = 1.3930
GM = antilog 1.3930
GM = 24.7

$$\log GM = \frac{\Sigma(\log x)}{N} =$$

$$\frac{6.9652}{5} = 1.3930 \quad \text{Thus:}$$

GM = antilog 1.3930 = 24.7

Table VIII*

G	m	log m	F	F•log m
10-15	13	1.1139	1	1.1139
15-20	17	1.2304	5	6.1520
20-25	23	1.3636	10	13.6360
25-30	27	1.4314	+14	+20.0396
			N = 30	30 | 40.9415

log GM = 1.3647
GM = antilog 1.3647
GM = 23.2

*Data used is from Table IV.

$$\log GM = \frac{\Sigma(F \cdot \log m)}{N} =$$

$$\frac{40.9415}{30} = 1.3647 \quad \text{Thus:}$$

GM = antilog 1.3647 = 23.2

Harmonic Mean

The harmonic mean is not used nearly as much as the arithmetic mean or the geometric mean, however, it is still quite useful when dealing with problems that have to do with weighed averages. "The harmonic mean of a series of values is defined as the reciprocal of the arithmetic mean of the reciprocals of the several values."[7] And the equation where HM is the harmonic mean, n is the number of how many numbers there are, and x_i's are the numbers may be written as such:[8]

$$HM = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

A problem as such will make use of the harmonic mean. We have the information given that:

Date	Price Per Pound	Total Cost
November 4, 1969	$0.10	$20.00
November 5, 1969	$0.20	$40.00

These facts are regarding two purchases of a certain commodity, and the problem is that we want to know the average price in the two transactions. The natural weight for the price per pound is the number of pounds, since the number of pounds represents the number of times the price is spent and is, therefore, the frequency. For this reason, in order to secure an appropriate weighted average, it is necessary to determine the number of pounds purchased on each of the dates. This result is achieved by dividing $20.00 by $0.10 and $40.00 by $0.20, thus discovering 200 and 200 pounds, as the

---

[7] Ibid., p. 72.

[8] Ibid.

appropriate weights. The problem may then be restated with the purchases described as follows:

Date	Price Per Pound	Pounds	Total Cost
November 4, 1969	\$0.10	200	\$20.00
November 5, 1969	\$0.20	200	+\$40.00
		400	400) \$60.00
			\$0.15

Such a restatement makes clear the average price per pound as the total cost \$60.00, divided by the total number of pounds 400, or \$0.15 per pound.[9] Here again the problem is simple, but when problems such as these become harder the harmonic mean is most useful. In Table IX the same data is applied to the definition with the use of frequency.

Table IX

Date	x	$\frac{1}{x}$	F	$\frac{1}{x} \cdot F$
November 4, 1969	\$0.10	10	20	200
November 5, 1969	\$0.20	5	+40	+200
			60	400

HM of reciprocals = $400 \div 60 = 6.66\overline{6} \ldots$
HM = $1 \div 6.66\overline{6} \ldots$ = \$0.15 per pound

Quadratic Mean

The fourth and perhaps the least used of the four means is quadratic mean. It "is defined as the square root of the arithmetic mean of the squares of the observations (root-mean-squares)."[10]

---

[9] Davies, and Yoder, op. cit., pp. 93-94.

[10] Chou, op. cit., p. 164.

The equation for the quadratic mean may be written as such:

$$QM = \sqrt{\frac{\Sigma x^2}{N}}$$

Where $\frac{\Sigma x^2}{N}$ is the arithmetic meam squared, and QM is the quadratic mean.[11] This type of mean is used to average values, but it is also used to find standard deviations and to average standard deviations.

I have discussed the four basic types of means and have given some examples of their uses in statistics, but the uses of these four mathematical functions are virtuality unbounded in the field of statistics today. And I fell they shall become of even greater importance in the future in this field.

---

[11] *Ibid.*

BIBLIOGRAPHY

Blair, Myers, Elementary Statistics, Henry, Holt, and Company, New York, New York, 1944.

Burr, McGraw, Engineering Statistics and Quality Control, Hill Book Company Inc., New York, New York, 1953.

Butch, R. L. C., How To Read Statistics, The Bruce Publishing Company, Milwaukee, Wisconsin, 1949.

Chou, Ya-Lun, Applied Business and Economic Statistics, Holt, Rinehart, and Winston, New York, New York, 1964.

Cohen, Lillian, Statistical Methods for Social Scientists, Prentice-Hall Inc., New York, New York, 1954.

Cooke, Dennis H., Minimum Essentials of Statistics, The Macmillan Company, New York, New York, 1936.

Copeland, Melvin T., Business Statistics, Harvard University Press, London, 1921.

Cox, D. R., Planning of Experiments, John Wiley & Sons Inc., New York, New York, 1958.

Croxton, Frederick E., Elementary Statistics with Applications in Medicine, Prentice-Hall Inc., New York, New York, 1953.

Davies, George R., and Yoder, Dali, Business Statistics, John Wiley & Sons Inc., New York, New York, 1949.

Downie, R. W., Basic Statistical Methods, Heath, Harper, & Row Publishers, New York, New York, 1965.

Ekeblad, Frederick A., The Statistical Method in Business, John Wiley & Sons Inc., New York, New York, 1962.

Garrett, Henery E., Statistics in Psychology and Education, Longmans, Green, and Company, New York, New York, 1955.

Holzinger, Karl J., Statistical Methods for Students in Education, Ginn and Company, Dallas, Texas, 1928.

Huntsberger, David V., Elements of Statistical Inference, Allyn and Bacon Inc., Boston, Massachusetts, 1961.

Jerome, Harry, Statistical Method, Harper & Brothers Publishers, New York, New York, 1924.

Leach, H. W., and Beakley, George C., The Slide Rule, The Macmillan Company, New York, New York, 1968.

Macdonald, Marian E., Practical Statistics for Teachers, The Macmillan Company, New York, New York, 1935.

Nelson, Boyd L., Elements of Modern Statistics, Appeton-Century Crafts Inc., New York, New York, 1961.

Peatman, John G., Introduction to Applied Statistics, Harper & Row Publishers, New York, New York, 1963.

Richmand, Samuel B., Statistical Analysis, The Ronald Press Company Inc., New York, New York, 1964.

Rosander, A. C., Elementary Principles of Statistics, D. Van Nostrand Company Inc., Princeton, New Jersey, 1957.

Stockton, John R., An Introduction to Business Statistics, D. C. Heath and Company, Boston, Massachusetts, 1947.

Thurstone, L. L., The Fundamentals of Statistics, The Macmillan Company, New York, New York, 1925.