THE DUBLIN JOURNAL

OF

MEDICAL SCIENCE.

MAY 1, 1894.

PART I.

ORIGINAL COMMUNICATIONS.

ART. XIV.—*Arithmetic, Geometric, Harmonic, and Quadratic Means.* By SIR CHARLES A. CAMERON, F.R.C.S.I., &c.

It is frequently necessary in making a series of medical or other observations, to the results of which a numerical value can be assigned, to deduce a mean value for the whole series of observations, and to draw conclusions, one way or the other, based on this mean value of the series.

As the accuracy of the conclusions drawn depends in a very great measure on the accuracy of the observations, and on these again depend the mean value deduced, I shall point out those instances in which the application of this method is suited to the case under observation, and, when suited, the *modus operandi*, and the precautions to be taken in order to avoid error as much as possible.

In the first place, there are some statements of mean value from which no conclusions of any value can be drawn—as when, for example, it is stated that the average age of two persons is 40. Here the age of the younger may be any number of years from 1 to 40, and of the older from 79 to 40. All that we are sure of in this case is that one, at least, of the persons is 40 or more. So when there is a great divergence in the numbers, the average of which is being sought, so much the less must there be confidence placed in the mean value of these numbers.

To take another example—two parties of riflemen, consisting of three in each party compete; the one party scores 84, and the other 79. What conclusions can be justly drawn in this instance? Only one, from the amount of information thus supplied—viz., that the total scored by one party exceeded that scored by the other, the obvious conclusion; but there is no information as to the individual scores, and we are not justified in saying that there were better riflemen in the winning party than in the other, as these totals may have been made up as follow:—

$$84 = 30 + 28 + 26 \quad \text{and} \quad 79 = 35 + 33 + 11.$$

It is obvious that were the mean values given in each of these instances—viz., 1st party, 28; 2nd party, 26·3—no more information could be gained respecting the individual competitors than is given by the totals.

There is another case which occurs sometimes in the comparison of two series of observations, which may be illustrated by reference to two series of researches published by Boecker.[a] He gave to an individual, whose diet was in other respects exactly estimated, a certain quantity of sarsaparilla daily, and found that the quantities of urine passed were in cubic centimetres as follow:—

1467	1744	1665	1220	1161	1369
1675	2129	887	1643	934	2093

During a second series of 12 days the individual experimented on took, instead of the decoction of sarsaparilla, his dietary remaining the same, as much distilled water, when the quantities of urine were as follow:—

1263	1740	1538	1526	1387	1422
1754	1320	1809	2139	1574	1114

These numbers are assumed to represent accurately the quantity of urine excreted by the individual. The question that arises, therefore, is—has the decoction of sarsaparilla influenced the excretion of urine or not; and, if it have, whether to increase or diminish that excretion? Upon examining each series of numbers separately it may be noticed that there are very great fluctuations in the numbers themselves. There does not appear to be any steadily augmenting or diminishing agency at work influencing

[a] See Reil's Journal für Pharmacodynamick und Toxicologie. Vol. II., Parts I. and II. "On the Action of Sarsaparilla."

the excretion of the urine; but the numbers registered on each succeeding day may be either greater or less than on the previous. Under such circumstances great caution must be exercised in accepting the mean values of the two series as at all representing the influence of the application of the sarsaparilla or its withdrawal. This doubt as to the means taken by themselves, representing the true influence of the sarsaparilla, is increased when it is found that if the means be taken for the first eight days of each series they are negatived by the respective means of the whole twelve of each series; and that if we suppose the numbers obtained on the thirteenth day to be the same as those obtained on the twelfth, the relation of the two series to one another is again reversed. Such a result as this might have been expected from the first, when it was seen that there was no appreciable relation existing between the numbers registered on each succeeding day; and, therefore, the inference to be drawn in this case is that whatever the influence of the sarsaparilla might have been, its influence was so much counteracted or otherwise by other agencies at work as to render it impossible to draw any inference in the case under observation.

So far the examples given have been those in which the application of the arithmetic mean has not been suited to the case chosen, or only partially; but there are a great many cases in which, when used with caution, the arithmetic mean, or pure average, is a valuable adjunct in experimental research. Take one or two instances in which it is only possible to express a general result by giving the average of a long series of observations, and be it borne in mind that the more extended the series of observations the nearer we approach to a mean value, from which there is little or no divergence.

It is stated that the annual rainfall in the Thames basin is 26 inches. Now by this statement it is neither meant that 26 inches of rain fell on every portion of the floor of the basin, nor that the amount of rain which falls each year over this area is invariably the same. By this statement is meant, that if the total amount of rain which has fallen during many years on the portion of England drained by the Thames and its tributary streams had accumulated, none of it being lost by evaporation or percolation through the soil, and if the number representing the height at which that water would stand, supposing the whole area to be level, were divided by the number of years during which the observations had been made, then we should find that each division would be represented by a height of 26 inches. It is obvious that this method represents to the mind in the most concise manner the general character of the rainfall over this area, and that it also furnishes a ready and sufficiently accurate estimate upon which other calculations of a cognate character may be based.

So much having been said on the value of the arithmetic mean, when judiciously employed, it is as well to state now that besides the arithmetic mean there are various other means—such as the geometric, harmonic, and quadratic means; and perhaps the simplest method of explaining to the non-mathematical reader the differences that exist between each of these means will be to indicate the methods by which they can be discovered. Let it be required to find the arithmetic mean of the numbers 3, 6, 9, 4, 13. In order to do this these numbers must be added together, and the sum divided by the number of addenda, as, total of the numbers, 35, which divided by the number of addenda (5) is equal to 7. Seven is, therefore, the arithmetic mean of the numbers 3, 6, 9, 4, 13.

It will be seen that the mean relation that 7 bears to these numbers may be exemplified by showing that, if all the numbers less than 7, as 3, 6, 4, be subtracted from 7, and the remainders be added together, the sum of the remainders obtained will be equal to the sum of the remainders obtained when 7 is subtracted from each of the numbers greater than itself, as 9, 13. Thus, $4+1+3=2+6$. Therefore 7 bears a mean relation to the numbers 3, 6, 9, 4, 13.

Let it be required to find the geometrical mean of the numbers 3 and 12. To do this the numbers are to be multiplied together, and such a root extracted as is indicated by the number of separate multipliers, as $\sqrt{12\times 3}=\sqrt{36}=6$. This operation may be expressed in more general terms by stating—If n represents the number of individual factors, the geometric mean of which is required, the n^{th} root of the product of these factors is the geometric mean of the factors.

In order to show that 6 bears a mean relation to 12 and 3 it is necessary to show that 3 bears the same ratio to 6 as 6 does to 12.

$$3=\frac{6}{2},\quad 6=\frac{12}{2};\quad \text{or,}\quad \frac{3}{6}=\frac{6}{12}=\frac{1}{2};$$

or again, 6 is the double of 3, as 12 is the double of 6; or still again, 6 is as many times greater than 3 as 12 is greater than 6. Therefore 6 holds a mean ratio between 3 and 12.

$\frac{3}{12}$ represents the ratio of 3 to 12,

$\frac{12}{3}$ " " 12 to 3,

and $\frac{6}{6}=1$ is the geometric mean of these.

Let us take into consideration the geometric mean of three numbers—for instance, 4, 6, 9. The geometric mean

$$=\sqrt[3]{4\times6\times9}=\sqrt[3]{216}=6.$$

It so happens in this instance that the geometric mean of the three numbers is one of them—that one which bears the same relation to the first that the last, or greatest, does to it. But does it always happen that the geometric mean of three or more numbers holds an intermediate position between the greatest and least numbers in the series? Let us consider this point. As the extraction of a root is a kind of converse operation to that of the multiplication of the factors, from which the product has been obtained, and as the root is always such a number as, when multiplied as many times by itself as is indicated by the index of the root, will produce the product, it follows that the root must always be some one number, which, when multiplied a certain number of times by itself, will produce the number from which it has been extracted—or, in other words, the number obtained by the multiplication of the several factors. But as these factors may vary greatly in magnitude among themselves, and the root can be only a single number, it follows that this single number must be greater than the smallest factor, and less than the greatest; indeed, that it must hold an intermediate position amongst all the factors. If there be many large factors, the larger the mean will be; if there be more factors small numbers than the number of the large factors, the smaller the mean will be. This will probably be better illustrated by taking an example. The geometric mean of 4, 6, 9 is 6—*i.e.*, if we take the product of $4\times6\times9$ and extract the cube root we obtain 6. Now, if we take the product of $4\times7\times9$ we evidently obtain a greater product, and the cube root will in consequence be greater—*i.e.*, nearer to 9. If, on the other hand, we take the product of $4\times5\times9$ we evidently obtain a smaller product, and, on the extraction of the cube root, a smaller number, or one nearer to 4; but we could never obtain 4 unless all the numbers were 4—*i.e.*, $4 \times 4 \times 4$—and we could never obtain 9 unless all the numbers were 9—*i.e.*, $9 \times 9 \times 9$; and the greater the other numbers are the greater will be the geometric mean, and the less the less. Therefore the geometric mean must always hold a mean position in any series of numbers.

The geometric mean cannot differ, therefore, much from the arithmetic mean, and this may be easily seen by reference to the numbers 4, 6, 9, the arithmetic mean of which is $6\frac{1}{3}$, while, as already stated, its geometric mean is 6.

Few words only are required in explanation of the harmonic mean, as its application is inappropriate in the cases of which we are taking cognizance; but as we are discussing means we may as well refer to this one also.

Whenever the product of two numbers is unity these numbers are said to be reciprocally related to each other, or reciprocals; or, more concisely, it may be stated that the reciprocal value of a number may be found by dividing unity by that number. Thus a number, unity, and the reciprocal of that number are in geometric progression, while the reciprocals of numbers in harmonic progression are in arithmetic progression. The harmonic mean may, therefore, be found for a series of numbers by taking the reciprocal value of the arithmetic mean of the reciprocals of those numbers. To take an example—let it be required to find the harmonic mean of 3 and 12. The reciprocal values of these numbers are $\frac{1}{3}$ and $\frac{1}{12}$, and the arithmetic mean of $\frac{1}{3}$ and $\frac{1}{12}$ equals $\frac{\frac{1}{3}+\frac{1}{12}}{2}=\frac{4+1}{24}=\frac{5}{24}$, and the reciprocal of $\frac{5}{24}=\frac{24}{5}=4\frac{4}{5}$. Therefore, $4\frac{4}{5}$ is the harmonic mean of 3 and 12.

A fourth mean is the quadratic mean. This mean may be defined as equivalent to the square root of the arithmetic mean of the squares of the given numbers, and may be found in the following manner:—

Let it be required to find the quadratic mean of 3 and 12. According to the definition the quadratic mean of 3 and 12 is equivalent to the square root of the arithmetic mean of the squares of the given numbers—that is, the mean of 3^2 and 12^2. $3^2=9$, $12^2=144$; $144+9=153$, and the arithmetic mean is equal to $\frac{153}{2}=76{\cdot}5$, the square root of which is $8{\cdot}75$.

It will be fitting to institute in this place a comparison between the arithmetic and quadratic means, and to point out those cases in which the application of one is preferable to that of the other.

The arithmetic mean of 3 and 12 is $7\frac{1}{2}$.
,, ,, ,, 3, 9, 8 is $6\frac{2}{3}$.
,, quadratic ,, 3 and 12 is $8{\cdot}75 = 8\frac{3}{4}$.
,, ,, ,, 3, 9, 8 is 7·16.

It will be seen that in each of these examples the quadratic exceeds the arithmetic mean; and it may here be remarked that it can be exactly and mathematically demonstrated that the quadratic always to a certain extent exceeds the arithmetic mean, and that this excess is in proportion to the inequality of the given numbers. The nearer the given numbers approach equality the nearer does the quadratic mean converge to the arithmetic mean; and the greater the inequality of the numbers the greater the difference between the two means; and when the given numbers arrive at equality the two means are identical.

We will turn this peculiarity of the quadratic mean (which, however, is also shared in by the geometric and harmonic means) to use in the estimation of what may be called the probable error in a series of estimations of the value of a definite fixed quantity—such, for instance, as the percentage of carbon in a given substance.

It is well known to chemists that, however cautious and experienced they may be, and however accurate be the balances employed by them, in quantitative analysis there is always some error in their weighing, be that error small or great; and that when an estimate of great nicety has to be made it is necessary to frequently repeat the experiment, and on some occasions to employ various methods of analysis. It will be the object of this paragraph to point out a method of estimating the mean error in a series of observations.

One of the best methods is by what is called "the method of successive means." A large number of determinations having been made, in which it is presumed that the error is as likely to be on the one side of the normal as on the other, and that those on one side are as nearly equal as possible to those on the other, we take the means successively of the first two numbers, then of the first three, first four, and so on, in each case, to about four places of decimals, when we shall find that the successive means will coincide in their whole numbers, but that they will differ in their first decimal place. By continuing the process far enough the first decimal will agree, then the first two, then the first three, and ultimately the first four; but the observations should have been carried out to many hundreds in order to secure this result. As an example of this we will take a series of numbers from a paper by W. Kaupp, "On the Dependence of the Amount of Chloride of Sodium in the Urine upon that in the Food." The series is as follows:—

24·300	24·173	23·340	23.600	26·057	23·101
24·511	23·343	26·400	22·650	23·590	23·644

Although this is not a case of the measurement of a definite fixed quantity, it will furnish us with an example of the accuracy of the mean as indicating the probable value of a fixed quantity. By taking the successive means we get—

As the mean of	the first	2	days	24·2365
"	"	3	"	23·9377
"	"	4	"	23·8533
"	"	5	"	24·2940
"	"	6	"	24·0950
"	"	7	"	24·1544
"	"	8	"	24·0530
"	"	9	"	24·3138
"	"	10	"	24·1474
"	"	11	"	24·0967
"	"	12	"	24·0590

It will be seen that after the third mean all those following agree in the whole number, 24. This number may be taken as certain. The numbers differ in their decimals, but in no place is the difference greater than ·4. We may, therefore, assume that the mean lies between 24+·4 and 24—·4; *i.e.*, if we take 24 as the round number the limit of uncertainty will be ·4. Taking the arithmetic mean of the first decimals as the basis for a general mean, the successive means, leaving out all but first decimals, are 24·2, 23·9, 23·9, 24·3, 24·1, 24·2, 24·1, 24·3, 24·1, 24·1, 24·1, consequently the arithmetic mean of the series = 24·1, and the greatest divergence from 24·1 is only ·3; 24·1 is the value of the mean, with an uncertainty = ·3.