Letter to the Editor.

IN Journal du Conseil, Vol. IX. No. 1 (1934) reviews have been given by H. J. B.-W. of the various papers in Hvalrådets Skr. No. 7. In addition to accounts of the papers some critical remarks appear, and as regards "A mathematical method for the study of growth" and "The optimum catch" the criticism is of such a nature that it necessitates some reply. As far as I understand the reason for the critique of our methods, I think it is due to the fact that H. J. B.-W. represents another opinion as to the employment of statistical methods in biological investigations from that held by the authors of the papers mentioned. To make my ideas clearer, I will in the following confine myself to the consideration of an example.

Suppose that there is given a frequency series — for instance a series resulting from length measurements of fish —:

x :	x_0	x_1	x_2	....	x_n
$H(x)$:	$H(x_0)$	$H(x_1)$	$H(x_2)$	....	$H(x_n)$

in which x denotes the values of the quantitative sign (length of the fish) and $H(x)$ the number of individuals with this value. Very often the investigator must try to characterise this series by some few figures, as it is in itself unsurveyable. As a rule the arithmetical mean is calculated:

$$m = \frac{\Sigma x \cdot H(x)}{\Sigma H(x)}$$

The significance of this quantity is easily understood by everybody, without any deeper insight into mathematical statistics. In the science of statistics, however, the mean is taken as a preliminary summary expression of the series as a whole, and besides some other quantities are used for characterising the series. As a rule some other symmetrical functions of the observations are used (moments, seminvariants, factorial sums etc.). The main point to which I should like to call attention is that in mathematical statistics the arithmetical mean is but one of the many quantities which are used for the purpose of characterising the frequency series, although it is the most important one.

From both a statistical and a biological standpoint one must reject the opinion advanced by H. J. Buchanan-Wollaston that the shape of the frequency series (or curve) is a function of the value of the mean. He says: "The shape is merely a function of the position of the mean." (Journal du Conseil, Vol. VIII, No. 1, p. 9, 1933). It is obvious that the value of the mean depends on the shape of the frequency series (or curve) and its position on the value-scale of the quantitative sign.

In many cases the arithmetical mean forms a sufficient basis for the elucidation and solution of the problems in question. In other cases, however, it is not sufficient, and a more complete characterisation of the series is necessary. As mentioned above we can calculate some other symmetrical functions — for instance the seminvariants — for this purpose. But here the theoretical frequency functions also come into use. If, for instance, the series has an appearance of approximating to the exponential frequency function:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}}$$

it is completely characterised by only two figures m and σ, the mean and the standard deviation respectively.

The theory of frequency functions is one of the most fundamental and difficult objects of the mathematical statistics. Professor Karl Pearson's system of frequency functions (or curves) is well known (see for instance: W. Palin Elderton: Frequency-Curves and Correlation, London 1906). Some mathematicians keep aloof from Prof. Pearson's treatment of these statistical phenomena (see for instance Prof. J. F. Steffensen: Matematisk Iagttagelseslære, Kjøbenhavn 1923). They argue that only those functions which are founded on the theory of observations can be considered as frequency functions. Some other researchers have, however, advanced farther on the way of Prof. Pearson. Their task is to find methods by means of which the "best-fitting curve" to fit to statistical data can be obtained. If we consider the mathematical functions found in this way as formulae of interpolation, nothing can be argued against the methods. But I do not understand the purpose for biological investigations of finding such algebraical functions as they cannot, of course, give any deeper insight into the problems than the empirical series itself or its graphical interpretation. Evidently, no logical or causal connection exists between the empirical series and the formula of interpolation. The modern methods of finding the "best-fitting curve" to fit to statistical data have some interest for mathematical algebra, but no importance for the biological investigations.

It is a matter of fact that in very few cases can it be understood why a frequency series can be approximately substituted by a certain function, and the frequency function must therefore be considered as a formula of interpolation. This fact thus applies to the agreement which very frequently appears between the exponential function and the statistical series obtained as results of measurements, enumerations of organs etc. The question arises, therefore, whether the agreement with the exponential function is of greater importance than the agreement with any other algebraical function. It is of no interest to discuss this question in a theoretical way. Experience, however, cannot be neglected. It shows that the agreement with the exponential function very frequently appears if the original material is analysed to a certain extent. On the other hand it is not to be expected that agreement with the exponential law appears without such sub-division because the original material is as a rule composed in a very complicated way of various groups. According to experience, however, the fact that a statistical series resulting from length measurements, enumerations of organs etc. does not agree with the exponential function, raises the question whether sub-division may result in frequency series which agree well with the exponential function. As such a means for statistical analysis I think the exponential frequency function (in other cases other frequency functions) has its main importance.

The exponential function is symmetrical. Consequently skew frequency series do not agree with this law. But skewness is just one of the characters which most strongly invites to sub-division of the material. The reason why I made some remarks on skewness in my paper is the necessity of emphasizing that, when a skew frequency series appears, sub-division will as a rule result in symmetrical series with different means. This is a fact of experience and has no bearing upon dogmatical speculations.

As regards the remarks of H. J. B.-W. on Prof. Guldberg's approximation coefficient, I will confine myself to saying that the decision whether the empirical series appears to approximate to the binomial frequency function or not, is, of course, left to personal judgment. The approximation coefficients are the mathematical criteria of the correctness of our hypothesis. The very same applies as well to the method of Prof. Pearson. I cannot understand that the decision by approximation coefficients is a matter of difficulty. In some cases the coefficients vary between the limits of 0.7 and 1.5, in other cases they take the values from 0 to 5, 6 or 7. Moreover, it must be mentioned that the approximation coefficients ought not to be used alone. Both Prof. Guldberg and I have, in addition, investigated whether calculated and empirical frequency numbers agree with each other.

If we go into statistical philosophy, which seems to be of great interest of "modern statisticians", I think that the philosophical foundation of the exponential frequency function is as well cleared up as can be expected. It is proved that a quantity (A) which is the sum of a number of elementary quantities:

$$A = \frac{a_1}{n} + \frac{a_2}{n} + \frac{a_3}{n} + \ldots\ldots\ldots + \frac{a_k}{n}$$

takes the form of the exponential frequency function if the number of terms of this sum is sufficiently great.[1]) If we imagine that each of these terms represents an elementary cause which within a group of individuals varies relatively unessentially, it will be understood that these individuals are distributed in accordance with the exponential function on the values of the quantitative sign A. Or in general, if the observed quantity can be taken as a sum of terms, the frequency distribution tends to take the form of the exponential frequency function even if the single terms do not follow this law. All biological quantities may be taken as such a sum or at any rate as due to a large number of elementary causes. "Thus the deviation of the actual chest measure of an individual from the average may be regarded as the sum of a very great number of very small deviations (positive or negative) due to the separate factors in the heredity and environment of the individual." (Whittaker and Robinson, p. 167).[2]) This philosophical hypothesis gives, in my opinion, an excellent explanation of the frequent occurrence of exponential series in the biological investigations.

If we now consider the mathematical methods for the treatment of the sigmoid growth, the very same discussion as above can be applied. Of course, not so extensive experiences are available as in the case of frequency distributions resulting from length measurements, enumerations of organs etc. As shown in my paper, however, growth can be studied in the same way as frequency series, i. e. as frequency distributions according to birth- (or death-) number. This is the general part of the method. The examples examined in my paper, appeared to be in good accordance with the exponential frequency function. Whether the agreement with this law is a general rule or not, only experience can prove. A theoretical discussion of this fact is of no value. The examples examined by me, were taken from essentially different types of organisms (yeast cells, herring, students), and there are thus good reasons for the belief that the exponential function in the treatment of the sigmoid growth should very frequently come into question. With regard to skewness or symmetry of growth, a theoretical discussion cannot, of course, lead to profitable conclusions. But to my thinking it must be considered as very remarkable if all the samples examined by me represent the exceptional case and not the rule. Moreover, experience shows in this case that skewness can be due to inhomogeneity. Future experience only can give a correct answer to these questions.

---

[1]) See about this matter:

[1]) M. W. Croftons: On the Proof of the Law of Errors of Observations, Lond. Trans. 159 (1869).

[2]) F. Y. Edgeworth: Law of Error, Cambridge, Phil. Trans. XX, 1904.

[2]) E. T. Whittaker and G. Robinson: The Calculus of Observations, London 1924.

When we are considering the growth-studies also, we are confronted with the very same dogmatical speculations which are so characteristic of the "modern statisticians". The fact that agreement with a certain mathematical function appears, has been taken as a proof of causal connection, and the sigmoid growth has been considered as a phenomenon homologous with certain autocatalytic chemical processes.

The method used for calculating the stock of whales (in "The optimum catch"), is of course subject to many errors. But this method possesses the advantage over other methods that the size of the stock and the fluctuations in it from year to year becomes not only known but understood as the result of known factors. I do not understand the purpose for biological investigations of substituting the curve of catches by the algebraical function $y = e^{a+bx+cx^2+\cdots}$ (which always appear in "modern statistical speculations") the coefficients of which mean nothing and cannot be understood.

Finally, I should like to add only that it must be remembered that we are dealing with biological problems, and that mathematical statistics in this connection are but a means towards the description and analysis of the empirical material. In his book "The Emperor's New Clothes" Prof. Johan Hjort writes: "When biology applies the statistical method, it starts from observations in Nature; it first obtains the necessary figures for the mathematical treatment of its subject by measuring or counting the things it is studying, and then draws its conclusions as to the variations within the material thus obtained. The next step is to look for factors which might influence the observed variations. It is in the cases when no such factors can be found that we hear the familiar statement that "chance" is the only explanation of the variation. But this is obviously to substitute the mind's own instrument for reality, and such statements are quickly forgotten as soon as the influence of factors in the environment is proved, as has been happening more than once in the ten or fifteen years."

Per Ottestad.