Psychological cautions in the use of statistics / by Charles S. Myers.

Separatabdruck aus
„Zeitschrift für angewandte Psychologie“,
herausgegeben von William Stern und Otto Lipmann
Bd. 36 (1930), Heft 1 und 2
Verlag von Johann Ambrosius Barth in Leipzig

Psychological cautions in the use of statistics

By

Dr. Charles S. Myers (London)

My object in writing this paper is to direct attention to certain dangers which appear to me to arise in the routine mechanical application of statistical methods to psychological problems. I will start with an extreme, and perhaps at first sight ridiculous, case by supposing that we have a series of mental measurements which have been obtained from a single individual by submitting him repeatedly, day after day, to precisely the same experimental conditions, e. g. to the same mental test. The results may vary from day to day owing to varying conditions of health, attention, etc. But let us assume that the chief cause of their daily variation is ascribable to the effects of practice, and that for our present purpose other causes are relatively insignificant. Now of what use is it, from the psychological standpoint, to determine the average of this series of data and to measure their average deviation from the average? Neither the average value of these measurements nor their average deviation from the average has any psychological significance.

The same holds for any measure of the reliability of a series of measurements, except in so far as it is introduced merely to avoid undue errors of observation. The very presence of the effects of practice prevents any constancy of the results. But let us suppose that the subject has already attained full practice. Of what use is it to determine a coefficient of reliability now? Is it our aim to obtain the same unvarying numerical data by exposing the subject to the same unvarying experimental conditions? I do not hesitate to say that this aim is psychologically absurd. Nothing can keep the internal, the psycho-physiological, conditions of a subject invariable. The very fact that the sub

ject has already been subjected to a given experiment changes his response to it on future occasions.

The statistician's attitude has been well expressed by one of themselves — "We rejoice in numbers and figures for their own sake". Not so the psychologist, who rejoices in them only for the value of each number and figure considered individually. Sir FRANCIS GALTON once told me that he saw little objection to fairly wide errors of observation if only the number of observations became sufficiently large, — because the errors on one side or the other of the true value tended to balance one another in the long run. I do not think that this is necessarily true even for errors of observation; some cause may weight these errors to one side. But I am convinced that it is untrue for variations due to the different responses of the subject at different times

We must realise that the psychologist is never dealing merely with the exercise of the special processes which his experiment or test is designed to call into play. These are invariably influenced by other even irrelevant processes and by the varying personality of the individual. Indeed a test which produces remarkable uniformity of result time after time is, to my mind, psychologically suspect. It is either too coarse in character to register fine variations, or it produces such complexity of variation that by mutual interaction and neutralisation a false constancy and hence a false feeling of confidence are engendered.

It may be objected that no sensible person would think of taking the average, or the variability, or the reliability of a series of measurements unless they obeyed a normal — or a skew — distribution. The statistician's view is that the average represents the true or most probable value and that deviations from it are due to the play of innumerable 'accidental' conditions. The psychologist's attitude is quite a different one. His very aim is to investigate the nature of those so-called 'accidental' conditions. For him the average is often a meaningless hotch-potch, concealing a number of most important differences.

It may happen that the effects of any given condition act in one direction in the case of some subjects, in the opposite direction in the case of others, while some subjects may not be affected by it at all. If we merely take an average of the total effects on different subjects, we shall come to the conclusion that the given condition produces no change whatever. This is a

fallacy which not uncommonly vitiates such mass experiments. I might quote various instances of published work claiming to prove, for example, that the menstrual period has no influence on the mental and muscular efficiency of women, or that the administration of ultra-violet rays has no influence on the output of industrial workers. I mention these instances because I have myself been concerned in investigations which disprove them.[1] In some women the menstrual period has a favourable, in others an unfavourable influence on their mental and muscular efficiency, whilst yet in others it produces no ascertainable effect whatever. So it is with ultra-violet rays. Nevertheless, if we group all the individuals together and average their data, we shall reach a result which completely obscures their individual differences.

Similar erroneous results are met with in other industrial fields of psychological experiment. The wellknown 'saddle-back' form of work curve, rising to a maximum near the mid-period and falling away from it at the beginning and end of the spell of work, has general reference only to the total output of a group of workers. It has only an industrial, a social, value. If the workers are individually studied, their work curves will be found to be of very varying form; and the form of each worker's curve often remains fairly constant for, and peculiar to, that particular worker. The same holds for determination of the most favourable length of rest pause and of the point at which it should be introduced during the work-spell.

I come now to the psychological dangers of employing the usual statistical methods of linear correlation. I will take a purely imaginary case — the correlation of general intelligence with sensitivity to pain. It is quite conceivable (this is here a mere assumption) that no correlation whatever exists between the latter and moderate ranges of intelligence (or that there may be actually a negative correlation between them), whereas with high or low degrees of intelligence, the correlation of intelligence with sensitivity to pain may be positive. In other words beyond a certain critical point, as it were, — or below another critical point —, a given psycho-physical process or sum-total of pro-

[1] "Two Contributions to the Study of the Menstrual Cycle". *Report Industrial Fatigue Research Board* **45**; "The Influence of Ultra-Violet Rays on Industrial Output". *Journal of the National Iustitute of Industrial Psychology* **4**, p. 144.

cesses (here intelligence) may exercise quite a different effect from that exercised at other values. It may often be psychologically quite unsound, as is so often done, to assume strictly linear correlation; and at the same time, owing to dearth of data, it may be difficult statistically to disprove the assumption. Beyond (or below) a given threshold, a sudden jump or change may occur, totally altering the play of the ability under consideration. There may thus be no psychological meaning in the ordinary correlation-coefficient of the product moment formulae. In any case, for the vocational purposes of industrial psychology the coefficient is unnecessary in correlating success in one or more vocational selection tests with success in factory work. All that is required of selection tests is that they serve to eliminate those persons who are likely to prove inefficient workers. And all that the vocational expert needs to do is to draw a line somewhere low in the order of ranking of the tested individuals which he receives from the factory supervisors, and to observe whether his tests would have excluded those who fall below this pass level. He is not concerned with the closeness of correlation throughout the entire range of subjects tested.

When a number of measurements are carried out (a) on one individual or on a group of individuals, and when those measurements are repeated (b) on another individual or on another group of individuals, the question arises whether the difference obtained between the two averages (a) and (b) has any statistical significance. It is commonly assumed to have such significance, if it exceeds about four times the probable error of the difference. Frequently this condition is not fulfilled, because of the wide scatter of the individual measurements about their average. Indeed a wide scatter is unfortunately characteristic of attempts at psychological (and of any biological) measurement. To satisfy statistical requirements, the difference between two biometric averages must often be so great (unless the number of observations is unusually large) that it is apparent to the naked eye and only needs measurement to carry this conviction to some one else or to formulate it in terms of precision.

But too commonly the experimental psychologist who has recourse to statistical methods assumes that because the difference between two averages is not statistically significant it may be summarily dismissed from further consideration. Let us, however,

suppose that we have to consider not one but a series of such differences, each of which by reason of its relatively high probable error, fails to reach statistical significance. We are then fairly justified in accepting the reality of these differences, if they relate to the same change of condition; yet how commonly one finds the occasion for such acceptance neglected!

Nothing is more important than that the experimental psychologist should be well grounded in the theory and practice of statistical measurement. But at the same time nothing is more important than that he should know when and how to use this statistical knowledge and skill, employing them not merely mechanically and mathematically but with due regard to psychological considerations. It was Huxley, I think, who expressed his conviction that the special value of the study of metaphysics for the scientist lay in the clearer knowledge it gave him of the line of demarcation between natural science and metaphysics. The same may have ultimately to be said of the value of statistical knowledge to the psychologist, unless he uses statistical methods with fuller consideration of their psychological dangers and implications.