Errors of Observation in Physical Measurements

Sixteen years ago in the course of an investigation of Upper Air Relationships [1] I had occasion to examine the question of correlation between errors of observation. In this connexion I found that precise knowledge regarding the standard error of even simple length measurements were completely lacking. In recent years the statistical theory of errors has received some attention; but the actual nature of errors of observation in physical measurements has not yet been properly studied.

This question is, however, becoming of increasing importance in view of the fact that in recent years results of even certain precision physical measurements performed independently by different observers have shown significant differences among themselves. The situation may be stated a little more precisely. If m_1 be the mean value of a set of n_1 measurements of any physical quantity carried out by one observer, m_2 the mean value of a similar set of n_2 measurements of the same physical quantity carried out independently by a second observer, and s_1 and s the respective observed standard deviations, then the corresponding standard error* of m_1 and m_2 are given by e_1 and e_2 respectively, where $e_1 = \frac{(s_1)}{\sqrt{(n_1)}}$ and $e_2 = \frac{(s_2)}{\sqrt{(n_2)}}$. If $d = (m_1 - m_2)$ be the difference between the mean values based on the two independent sets of observations, the probable error of d is given by the well-known formula $\sqrt{(e_1^2 + e_2^2)}$ Dividing "d" by this probable error we get the deviation in terms of standard errors the probability of occurrence of which can be easily obtained from tables of the probability integral.† If the deviation is large and the probability of occurrence is small, the difference $d = m_1 - m_2$ between the two results must be treated as significant.

Such significant differences may arise from a variety of causes. There may be of course systematic differences in physical conditions under which the observations were made. Assuming, however, that such differences in physical conditions have been eliminated, if the significant differences are still found to persist, as they appear to do in many cases, it becomes necessary to study more closely the nature of "errors of observations."

Eight or nine years ago I collected some material for this purpose. The length of several brass rods (or rather of the distance between two fine marks on each rod) of various sizes were measured a large number times (of the order of 500 or more) by different observers. Originally it was the intention to study the relation between the mean value (m) and the corresponding standard deviation (s), in particular to test whether the standard deviation was independent of the mean value, that is, whether the error of measurement remained sensibly constant for different values of the length measured, or whether s/m, that is, the proportional error was constant, or whether any other empirical relationship could be traced between s and m.

The preliminary examination of the material, however, clearly showed that, under the conditions of these particular experiments, the errors deviated very markedly from the normal (Gauss-Laplacian) distribution. There was unmistakable evidence of a particular kind of bias. Each observer tended to conform much more closely to particular values of the observations than one would expect from normal theory. Further examination showed that something like the following was probably happening: An observer gets scale readings of the two end-points (or marks on the brass rod) as say 4.36 cm. and 2.31 cm. which yield an observation of 2.05 cm. for the length. This knowledge sets up a bias (of which the observer is usually unconscious) in favour of the particular length 2.05 cm. Subsequently, for example, if the observer gets a reading of say 13.23 cm. for one end-point, there is a tendency for him to register the reading for the other end-point by (unconsciously) adding 2.05 cm. to 13.23 cm. and getting 15.28 cm. or some reading very near this value.

As already stated, the readings for the same observer naturally clustered round the biassed value far more closely than one would expect from the normal theory, so that the calculated standard errors were small. Consequently the mean values obtained by different observers in most cases differed very significantly which, of course, merely showed the presence of marked "personal equations."

A possible way of eliminating the above kind of bias suggested to me by making use of non-uniform scales in place of ordinary standard scales. A number of non-uniform scales were constructed in the laboratory, the necessary apparatus for which was borrowed from the Bengal Engineering College, Shibpur, through the courtesy of Prof. Amaresh Chakravarti. In the non-uniform scales the centimetre and millimetre marks were varied at random but the variations were kept so small that they could not be at once detected by merely looking at the scale. Preliminary experiments with these non-uniform scales along with an ordinary uniform scale were started some time ago under the general supervision of Subhendu Sekhar Bose. These experiments also showed that a marked bias was set up when working with the uniform scale.

In view of the intrinsic interest of the problem we started a new series of observations in April, 1939 under the supervision of Mr Susobhan Datta, M.Sc., P.R.S. A preliminary analysis of the material has confirmed the emergence of biassed values when using the uniform scale. The use of non-uniform scales, of course, together with calibration tables, on the other hand, gives better results; so that we reach the apparently paradoxical conclusion, namely, that non-uniform or 'inaccurate' scales give more reliable results than uniform or accurate scales. The preliminary analysis also definitely indicates that the distribution of errors, even when measurements are made with non-uniform scales with a view to eliminate personal bias, shows marked skewness indicating a significant deviation from the normal Gauss-Laplacian curve.

The material collected so far is being analysed in detail and will be published in due course. Experimental

*It will be remembered that the "probable errors" e_1' and e_2' are given by 0.6745 (e_1) and 0.6745 (e_2) respectively. In modern statistical practice "standard errors" are usually used instead of "probable errors."

†That is, on the assumption that s_1 and s_2 are the "true" values of the respective standard deviations, which is more or less justified when n_1 and n_2 are both large.

work is also proceeding and arrangements have been made to study in the same way results of observations with more precise measuring instruments of different types. After clearing up some of the obscure points discussed above, we also hope to be able to go back to our original problem, namely, the study of the relation, if any, between the size of the object and the standard error of the measurements.

Statistical Laboratory
Presidency College
Calcutta, 27-11-39. P. C. Mahalanobis

Memoirs of the Indian Meteorological Department, Vol. XXIV, Part II, 1923.