

NOTE ON DR. BERKSON'S CRITICISM OF  
TESTS OF SIGNIFICANCEBY R. A. FISHER  
*The Galton Laboratory*

WHEN, about eighteen years ago, I was looking for a good example to illustrate the test of straightness appropriate to regression lines, I felt myself particularly lucky to find one in which the departure from linearity was of a somewhat unusual kind, and in which a superficial examination by graphical methods, without submitting the subjective impression to any objective test, was likely to be misleading. The case chosen was from a comprehensive paper by A. H. Hersh on the influence of temperature on the number of eye facets developed in *Drosophila melanogaster* in a number of homozygous and heterozygous phases of the bar factor. Several of these phases showed to graphical inspection a remarkable discontinuity in the rate or direction of change in the neighborhood of 24° C. The heterozygote between the wild gene and that known as ultra-bar, on the other hand, gave frequencies which, if examined uncritically, or by methods capable of detecting only the obvious, would have failed to indicate that anything remarkable was happening at the critical temperature. It was, therefore, of some little interest to see whether a new criterion based on the rigorous solution of a problem of distribution then but recently cleared up, would enable the experimenter with such material to detect features of importance in his data, which, without such aid, would have escaped his notice.

Dr. Joseph Berkson writing, curiously enough, in 1942, when the advantages of objective tests are at least more widely appreciated than they were, refers to this example in the September number of this JOURNAL. He has drawn the graph. He has applied his statistical insight and his biological experience to its interpretation. He enunciates his conclusion that "on inspection it appears as straight a line as one can expect to find in biological material." The fact that an objective test had demonstrated that the departure from linearity was most decidedly significant is, in view of the confidence which Dr. Berkson places upon subjective impressions, taken to be evidence that the test of significance was misleading, and therefore worthless.

It is not my purpose to make Dr. Berkson seem ridiculous, nor, of course, to prevent him from providing innocent amusement. Had he looked up Hersh's original paper he would have been spared a blunder, but we should have lost an example of the dangers of authoritarian judgment, based on subjective impressions, which even at the present

date may be of value. Evidently, good biological data, examined by accurate methods, are capable of being much more informative than Dr. Berkson imagines. It is very well worth while to be reminded that general condemnations of "biological material" based on limited experience, as Dr. Berkson's judgment must be, may vastly underestimate the cogency of the evidence which careful and extensive work neatly provides.

A further confusion arises in Dr. Berkson's final comment on this example: "In this case my own judgment would be, not that the regression is non-linear, but that the temperature has varied during each or some of the experiments. At least that would explain the small *P*." As experimenters well know, one of the commonest uses for tests of significance is to detect errors of technique, or to confirm the belief that the technique has been adequate. Improbable as such an explanation seems in the case of Hersh's data, it is in general true, and important, that a discrepancy between observation and expectation may be due in fact principally to imperfections in the actual conduct of the work. If, for example, it could be supposed that the extensive series of observations at temperatures 23° and 25° C. had in error been completely interchanged, the evidence for non-linearity in the case of this genotype, though not of the others, would disappear. The fact that the test of significance has shown the series of facet numbers to be non-linearly related to the series of temperatures *recorded* would, if there really were reason to expect them to be linear on the true temperatures, have aided in the detection of such an error. What Dr. Berkson fails to realize is that the judgment, from inspection, that the line appears as straight as one ought to expect, would have given no aid whatever towards discovering the cause of any real anomaly, whatever the cause might be, because that judgment in effect denies the evidence that any real anomaly exists.