THE INTERPRETATION OF CERTAIN REGRESSION METHODS AND THEIR USE IN BIOLOGICAL AND INDUSTRIAL RESEARCH[1]

By C. Eisenhart

1. **Introduction.** Just as the scientific theorist depends upon the research worker for the facts upon which to build his theory, so does the practical man rely upon empirical relationships to help him estimate (or predict) the value of one quantity from that of another. Sometimes he is interested in assessing the value of some quantity which it is impracticable or impossible to observe directly in a given instance, the estimation being performed with the aid of a previously established relationship between the quantity whose value is sought and another whose value can be determined directly. In other instances he wishes to make use of the relationship existing between two or more quantities to help him adopt a course of action which has a good chance of leading him to a desired result. An example is that of a manufacturer who wishes to exercise control at various stages of a manufacturing process so as to produce a product whose quality lies within a specified range.

In appealing to the interests of the practical man, proponents of statistical methods have often illustrated their writings with beautiful examples of the power of this implement of research, without adequately discussing the abstract ideas that underlie the methods they have promoted—ideas essential to correct statistical thinking. The result has been that to many research workers certain problems with similar objectives appear amenable to identical statistical solution, when in fact intrinsic differences exist which alter considerably the details of their solution.

Such misinformation is particularly prevalent among those whose knowledge of the mathematics of correlation, and of curve fitting, has been derived from the treatment in elementary statistics courses of problems in which no one of the variables stands out from the rest as being *the* dependent variable, with its values determined (not exactly, but within limits) from the values that happen to be assumed by the other variables in the data under investigation. In elementary courses the usual procedure in such cases is to *take* one of the variables as the dependent variable, and then *consider* the others as independent variables. Furthermore, the curve-fitting procedure usually adopted depends on the additional assumption that the values of the independent variables are known exactly (without error)—an assumption often passed by without mention, and one that

[1] Revised from an expository paper presented, under different title, to the American Statistical Association, at Detroit, December 29, 1938, at the invitation of the program committee of the Biometrics Section.