Statistical Science
1990, Vol. 5, No. 4, 463–464

Introductory Remarks on Neyman (1923)

T. P. Speed

The translation which follows was taken from a paper constituting part of Neyman's doctoral thesis, submitted to the University of Warsaw in 1924. It is based on research he carried out at the Agricultural Institute of Bydgoszcz (formerly Bromberg) during the period 1921 to 1922 (see Reid, 1982, for further details). Its primary interest lies in the clear formulation of Neyman's model for comparing the yields of a number of varieties, each sown in a number of plots in a field. A short summary of the contents of this paper can be found in Scheffé (1956, page 269), where the model is said to refer to "the completely randomized experiment with zero technical errors." Although sampling notions play a key role in the paper, the notion of randomization is not mentioned.

Part 1 of the paper reviews the basic notions of mathematical statistics: the Gaussian law, expected values, variances, independence, Tchebychev's inequality, the weak law of large numbers, the central limit theorem for averages of independent and identically distributed observations with finite variance, estimation of means and variances when sampling with and without replacement and posterior probability intervals obtained via Bayes' theorem (using a uniform prior).

Perhaps the most interesting section of Part 1 is Section 5, entitled *Definition of True Value*. Here Neyman is concerned with the relationship between scientific (including mathematical) concepts and empirical notions. The former he viewed as fictions which have value when properly defined, and he regarded it as important to distinguish between scientific and everyday uses of the same word or phrase. He says that he would like to define the true value of a crop yield, for example, to be the common expectation of a sequence of independent and identically distributed random variables corresponding to different measurements of that yield. However, he points out, this definition does not help much, as it then depends on the definition of expectation. Neyman's solution was to define true yield to be a number that possesses one property also possessed by expectation, namely that it serves as a location parameter for the normal approximation to the distribution of the mean of the abovementioned sequence of random variables. This permits approximate probability statements to be made about the true yield—by Bayes' theorem!—and he then moves on to a discussion of these notions in the context of agricultural experiments. See page 470 for an illustration of Neyman's use of this definition.

Part 2 begins with a brief discussion of covariance, correlation and regression, in the context of a finite population. All of this was material familiar to statisticians at that time. Neyman referred to Czuber's *Wahrscheinlichkeitsrechnung* (1914) and *Theorie der Beobachtungsfehler* (1891), Markov's 1913 *Calculus of Probability* and a paper by Bernstein published in the Kharkov Mathematical Society Journal. Pearson's *The Grammar of Science* (1900) was also cited.

Neyman's original ideas on the application of statistics (or probability theory) to agricultural experimentation are contained in the portion of the paper translated. Their most well-known exposition is, of course, in the paper Neyman (1935, with the cooperation of K. Iwaskiewicz and St. Kolodzieczyk), on the occasion of the bitter clash with R. A. Fisher. However, the presentation in the later paper contrasts dramatically with that translated below. In 1923 Fisher's theory of designed experimentation, including his z-test and randomization, lay in the future, as did the Neyman–Pearson theory of testing hypotheses; the notion of technical error is also absent from the earlier discussion. By contrast, the 1935 paper discusses only randomized block and Latin square designs, includes technical errors, explicitly evaluates means and mean squares with respect to randomization (permutation) distributions and pays attention to the type 1 and type 2 errors of the associated tests. Thus the emphasis shifts attention away from what was perhaps Neyman's greatest contribution to designed experimentation: his explicit use of hypothetical responses corresponding to what would have been observed, had the treatment allocation been different. For an insight into Neyman's own view of this work, see Reid (1982, pages 45–49). It seems possible that those present at Neyman's 1935 oral presentation of his ideas missed this basic point at first hearing or reading, hidden as it was in a broad critique of the then popular methods of Fisher, and that this explains, at least in part, the strong reaction to Neyman's criticisms. Later writers on the design and analysis of experiments Kempthorne (1952), Cox (1958), and others, e.g., Hodges and Lehmann (1970,

---

T. P. Speed is Professor and Chair, Department of Statistics, University of California, Berkeley, California 94720.

Section 9.4) have clearly found the model of pedagogical value.

In view of the fact that Scheffé (1956) described the model in this paper as corresponding to the "completely randomized experiment," the reader may wonder if there is in this paper any explicit reference to randomization. We could find none. In our view, there is an implicit assumption of randomization, in that the plots which are assigned to particular treatments are supposed—for the purposes of calculating means and variances—to be simple random samples satisfying the prescribed constraints. But implicit is not explicit: randomization as a physical act, and later as a basis for analysis, was yet to be introduced by Fisher. Fisher's priority is explicitly recognized in Neyman (1935, page 131), where Neyman, discussing the conditions under which an observed difference between two treatment means can be regarded as an unbiased estimate of their true difference, writes

> The difficulty has been overcome by the device proposed by R. A. Fisher, which consists in making the η's random variables with mean equal to zero. For this purpose the plots within each block are randomly distributed among the different objects [treatments].

Why publish a translation of Neyman's introduction of this model just now? In our view, it is the model for designed experiments and observational studies that permits most (all?) of the important issues associated with their analysis and interpretation to be clearly defined, discussed and elucidated. Apart from the earlier writers cited, we feel that this has been convincingly demonstrated most recently in a series of papers by Rubin, beginning with Rubin (1974, 1977, 1978); see especially Holland and Rubin (1983, 1988) and papers by Rosenbaum, alone or jointly with Rubin, cited in Holland (1986). The desirability of drawing attention to this fine body of work and placing it more clearly in the tradition from which it arose would seem to justify carrying out the translation. We hope that the paper will also have some appeal to those interested in the evolution of the statistical notions associated with experiments and other studies.

ACKNOWLEDGMENTS

We would like to thank Professor J. Baksalary for supplying a photocopy of the paper and Professors D. Freedman, E. Lehmann and S. Stigler for their encouragement and helpful comments in relation to this project.

REFERENCES

COX, D. R. (1958). *The Planning of Experiments.* Wiley, New York.

CZUBER, E. (1891). *Theorie der Beobachtungsfehler.* B. G. Teubner, Leipzig.

CZUBER, E. (1914). *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung Statistik und Lebensversicherung,* 3rd ed. B. G. Teubner, Leipzig.

HODGES, J. L. and LEHMANN, E. (1970). *Basic Concepts of Probability and Statistics,* 2nd ed. Holden-Day, San Francisco.

HOLLAND, P. W. (1986). Statistics and causal inference (with discussion). *J. Amer. Statist. Assoc.* **81** 945–970.

HOLLAND, P. W. and RUBIN, D. B. (1983). On Lord's paradox. In *Principals of Modern Psychological Measurement* (H. Wainer and S. Messick, eds.) 3–25. Erlbaum, Hillsdale, N.J.

HOLLAND, P. W. and RUBIN, D. B. (1988). Causal inference in retrospective studies. *Evaluation Rev.* **12** 203–231.

KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments.* Wiley, New York.

MARKOV, A. A. (1913). *Calculus of Probability.* St. Petersburg. (In Russian.)

NEYMAN, J. with the cooperation of K. Iwaskiewicz and St. Kolodziejczyk (1935). Statistical problems in agricultural experimentation (with discussion). *Suppl. J. Roy. Statist. Soc. Ser. B* **2** 107–180.

PEARSON, K. (1900). *The Grammar of Science,* 2nd ed. Adam and Charles Black, London.

REID, C. (1982). *Neyman: From Life.* Springer, New York.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychology* **66** 688–701.

RUBIN, D. B. (1977). Assignment of treatment group on the basis of a covariate. *J. Educ. Statist.* **2** 1–26.

RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58.

SCHEFFÉ, H. (1956). Alternative models for the analysis of variance. *Ann. Math. Statist.* **27** 251–271.