The Annals of Statistics
1981, Vol. 9, No. 1, 1–2

EGON S. PEARSON (AUGUST 11, 1895–JUNE 12, 1980). AN APPRECIATION

By Jerzy Neyman

University of California, Berkeley

Egon S. Pearson died on June 12, 1980. *Sit ei terra levis*!

I feel indebted to Egon S. Pearson for several periods of joint research activity, the memories of which are very dear to me. The initiative for cooperative studies was Egon's. Also, at least during the early stages, he was the leader. Our cooperative work was conducted through voluminous correspondence and at sporadic get-togethers, some in England, others in France and some others in Poland. This cooperation continued over the decade 1928–1938. It resulted in ten coauthored papers that appeared in five different periodicals, four of them published in England and one in Poland. In 1966 all these papers were reproduced in the volume *Joint Statistical Papers of J. Neyman and E. S. Pearson* [University of California Press, Berkeley and Los Angeles].

My joint work with Egon was concerned with the problems of testing hypotheses. One aspect of this work was philosophical, partly related to the dispute between the French probabilists Joseph Bertrand and Émile Borel. In a sense, Bertrand was pessimistic and thought that no test based on probability can give reliable answer to a specific question as to whether an observed phenomenon is due to some specific cause or simply to chance. Contrary to this, Borel was optimistic. He thought that reliable probabilistic tests of hypotheses are possible, provided one selects an appropriate criterion. Also, his interpretation of the problem was different from that of Bertrand. According to Borel, the reliability of probabilistic test criteria must be measured by the long-range frequency of correct conclusions that a given criterion can provide for judging the hypotheses tested. The reader will notice that this statement of Borel, somewhat vague as it is, represents the basic idea of the modern theory of testing hypotheses. The decade-long joint work of Egon and myself was aimed at building a mathematical theory of tests the use of which could minimize the frequency of erroneous conclusions regarding the hypotheses considered, the hypotheses that we termed statistical hypotheses.

Our first idea (or shall I call it "discovery?") was that in testing a statistical hypothesis, say H, one can commit errors of two different kinds: one can reject H when it is true ("error of first kind") and one can accept H when it is false ("error of second kind"). Through the consideration of a number of particular cases, including the familiar "Student's hypothesis," we noticed that [particularly when H is "simple"] the control of the errors of the first kind is easy and could be achieved in an infinity of ways. One has only to use the hypothesis H to deduce the distribution of the contemplated possible test criterion. Furthermore, with particular reference to Student's problem, it was found that an infinite class of different test criteria, say T, exists, all having the same distribution as the familiar Student's test, say t. Finally, it was found that each T satisfied the inquality $|tT| \leqslant 1$, that implies possible sharp contradictions between the criteria T and the then well established Student's test criterion t.

This result brings to the fore the control of the errors of the second kind which implies the necessity of considering the set of all admissible hypotheses including all those alternative to the hypothesis tested H.

Our first attempt at the solution of the difficulty was based on the dictum of Laplace "Le calcul des probabilités n'est au fond que le bon sens réduit au calcul." The "good common sense" suggested to us the formulation of the principles of the "likelihood ratio test," that we denoted by λ, more or less

Received July 28, 1980.
AMS 1970 subject classification. Primary 01A70.

> *If in a given case there exists an admissible hypothesis H' alternative to the tested H, that implies the probability of the facts observed that is much larger than that implied by H, then the truth of H deserves considerable skepticism.*

Thus, the recommended test criterion: λ.

The reader will notice that the recommendation of the likelihood ratio λ is somewhat dogmatic, the "good common sense" notwithstanding. In consequence, there followed a number of studies intended to determine what we now call the performance characteristics of the λ tests in a variety of cases. These studies are akin to quite a few performed in modern times. For example, see Berkson (1980).

Our further studies with Egon aimed at a nondogmatic theory of testing statistical hypotheses. They culminated in our paper of 1933 that provided the basic theory. Subsequent papers were concerned with existence or otherwise of uniformly most powerful tests, of similar regions, of uniformly most powerful unbiased tests, etc.

The memories of cooperation with Egon during the decade 1928–1938 are very dear to me.

REFERENCES

BERKSON, JOSEPH (1980). Minimum chi-square, not maximum likelihood! *Ann. Statist.* **8** 457–487.

NEYMAN, J. and PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc.* **231** 289–337.

STATISTICAL LABORATORY
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720