Box G E P. Non-normality and tests on variances. *Biometrika* **40**:318-35, 1953.
[Imperial Chemical Industries, Dyestuffs Division Headquarters, Blackley, Manchester, England and University of North Carolina, Raleigh, NC]

The effect of particular kinds of non-normal error distributions on tests on means (analysis of variance tests) and tests on variances (Bartlett's test) are contrasted. It is found that the former are far more robust than the latter. [The *Science Citation Index*® *(SCI*®*)* and the *Social Sciences Citation Index*® *(SSCI*®*)* indicate that this paper has been cited over 190 times since 1961.]

George E.P. Box
Department of Statistics
University of Wisconsin
Madison, WI 53706

December 10, 1981

"In 1947, when I was studying statistics at University College, London, I got a summer job as a vacation student with Imperial Chemical Industries at their research center near Manchester in the north of England. I was assigned to a laboratory which tested rubber and plastic materials. Some of the tests involved repeated flexing of the material until it broke, others tested wearing qualities by measuring weight loss after successive periods of abrasion. Typically, a single machine tested simultaneously a small, fixed number of samples (for example, four) so allowing their direct comparison.

"My wartime experience had made me very familiar with experimental designs introduced by Fisher, Yates, and Youden for efficiently comparing groups of materials when only a limited number of direct comparisons was possible.[1,2] However, I was doubtful as to the validity of my statistical analyses. These assumed *normally* and *independently* distributed errors. But it was obvious that breaking strength (which was critically determined by the weakest link and might therefore be expected to have a distribution like an extreme value from a normal distribution) was unlikely itself to be normal, and I found this view was supported when I plotted distributions of past data. Furthermore, *successive* measurements of wear made on the same samples would surely be autocorrelated, not independent. There were suggestions in the literature that moderate violations of the basic assumptions might not matter too much, but I continued to be concerned about them.

"That summer the company offered me a job and I went to work for them the following year, and stayed until 1956. During this period I spent a good deal of time studying these problems, particularly while I was on leave for a year at the University of North Carolina at Raleigh, and I published papers setting out my conclusions. This particular paper contrasted the extreme sensitivity to non-normality of tests to compare variances with the much lesser sensitivity of tests to compare means. I think that one reason it has been referred to so often is that in this paper the word *robust* is used for the first time to mean *insensitive to violation of assumptions*. Also, for the analysis of variance, it criticizes the use of Bartlett's test as a preliminary test of variance homogeneity, in these words which have often been quoted: 'To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port.'

"I remember being disappointed on receiving the referee's report from *Biometrika*, which was highly disapproving and opposed publication. However, on looking again into the envelope, I found a handwritten note from E.S. Pearson, the editor. He said, 'You will see that the referee does not like your paper but I do and subject to mild revision I am going to publish it anyway.' I have found it invariably true that papers containing fresh ideas are the most difficult to publish. P.J. Bickel has published a recent review in this field."[3]

1. **Fisher R A.** *The design of experiments.* Edinburgh: Oliver & Boyd, 1942. 236 p.
2. **Yates F.** *The design and analysis of factorial experiments.* Harpenden, England: Imperial Bureau of Soil Science, 1937. 96 p.
3. **Bickel P J.** Another look at robustness: a review of reviews and some new developments. *Scand. J. Statist.* **3**:145-58, 1976.