26 March 1934.

Sir Frank Smith,
Secretary, The Royal Society,
Burlington House,
Piccadilly.

Dear Sir Frank,

I am now returning again Wilks' paper
with the more detailed criticism, which the Society expects of
a referee who is to fill up the standard form, but which I
had thought would be unnecessary, in view the misapprehension
under which the work proposed for publication was evidently
undertaken. The emendations suggested by the author do not meet
the criticism first made, as I had hoped they would, for I do
not doubt that the author has brains enough to see the mistake
with which he fell. As he is now inclined, evidently, to lay
stress on his work as a contribution to experimental design, it
will be necessary to deal also with this aspect.

Yours sincerely,

P.S. I am returning the copy of the author's remarks with
his letter.
REPORT.

The author's preliminary statement shows that he is ill acquainted with what the analysis of variance is, and what it is used for. Under (a) it may be noted that the hypothesis to be tested is seldom or never that of the homogeneity of all the observations, but that of the inefficacy of the treatments, which the experiment is designed to test, or of some specific components of the treatments. In well-designed experiments the other components of variation, such as those represented by rows and columns in a Latin Square, are deliberately made as great as possible in comparison with the residual error, and with the real errors affecting the treatments, so that homogeneity is neither attained in practice, nor desired.

The mean and standard deviation of the hypothetical normal population are not in any case "given", but are estimated from the data, and eliminated from the tests of significance.

(b) The sentence beginning "Furthermore" is obscure and seems to be self-contradicting.

(c) It is not clear whether the author means by "under the assumption" (1) that the test is valid subject to the truth of an assumption, or (2) that the test is designed to examine the truth of the assumption. As far as can judge the distinction between the hypothesis which a test of significance
is designed to test, and a postulate upon which the mathematical analysis depends for its validity, is obscure throughout the paper, although this distinction is vital in any attempt to discuss the validity of tests of significance.

This confusion seems to infect (d) Estimates of variances due to special factors are not a usual part of the analysis of variance procedure, since they are without objective meaning in most experimental work.

Page 3. The statement, now proposed as an amendment, that the independance of sums of squares in the randomised block and Latin Square arrangements has not been explicitly demonstrated, is a claim that cannot be defended. It has been explicitly demonstrated for regression in general, of which these arrangements offer a deliberately simplified special case. This seems to be now admitted in this country, and the inability of the author to recognise the wide scope of the regression procedure, however genuine this inability may be, can scarcely be held to support his claim to be making an original contribution to the subject. In this respect the publication of the paper would be unfair to other junior mathematician's, who have read the literature with more penetration, and in consequence do not offer for publication
proofs of what they recognise to be already established. Its publication would also give the impression that elaborate and cumbersome methods of proof were necessary to establish points which really belong to the elements of the subject.

Page 4. The claim put forward in the author's proposed amendment to have contributed to the problem of "designing lay-outs" misses the whole point of this problem. (1) His paper has no reference to the physical conditions of the experiment, in the laboratory or the field. Thus the author speaks of his row-column scheme as equivalent to the randomised block method of design, although neither the process of randomisation, which is essential for the validity of the tests of significance, nor the process of forming topographical blocks of plots, which the precision of the method, could be deduced from his schema, which is merely a two way classification of $r \times s$ values, irrespective of what these represent. (2) On the other side, the purely mathematical aspects of the problem of experimental design are ignored, in so far as these depend on the possibility of partitioning the smaller numbers, and their squares and products, in especially fruitful ways—e.g., considerable use has been made in practice of the fact that 9 objects may be partitioned into sets of 3 in 4 mutually orthogonal ways (i.e., a 3x3 Graeco-Latin Square exists), but 36 objects cannot
be partitioned in sets of 6 in more than 3 mutually orthogonal ways. Infact, both the practical and the combinatorial aspect of experimental design are ignored.

The belief that the hypothesis tested is that of the homogeneity of all the observations, is evidently responsible for the erroneous analysis, on page 27, of the sub-divided Latin-Square, discussed in section 6; and prevents the author from being able to explain why the ordinary analysis given on page 28 is the only possible correct procedure.

The same consideration explains why the arrangement proposed in section 7 would be useless in practice, and is, as far as I know never used.

On page 17 there is a remark, which I had not previously noticed, pertaining to be a criticism of my use of the Analysis of Covariance. I find that the somewhat inexact test criticised does not occur in my book on Statistical Methods, to which the author refers, and I cannot recall having used it in any paper of mine in which the Analysis of Covariance is used. If the paper were to be published, I hope the author may be induced to the phrase "as Fisher does", by specifying the particular case in which, in his opinion, the test of significance I have used is inexact. I should then be in a position to judge among the possibilities - (1) that I had
blundered, and used an incorrect method, (2) that I had used a sufficiently exact approximation, or (3) that the author had understood the correct process, or had been discussing a process ascribed to me at second hand, without any direct knowledge of my actual procedure. At present I think the last must be the true explanation.

This point is relevant to another issue. The author's method enables him to say that the process, which he ascribes to me is inexact. The complexity of his analytic approach, however, seems to preclude him from seeing what actually is wrong with the method, and how it can be put right. He has no clue as to whether the test as it stands is a good approximation to an exact test, or indeed as to whether an exact test exists. Yet none of these questions is difficult to answer. I cannot help feeling, of the paper as a whole, that its aim has been so much misdirected from the first that the author would, in a few years' time, be glad that it should not be now published, since, as far as it goes, the proofs it provides amount to no more than that certain widely used experimental arrangements possess the properties that they were originally designed to possess, and which have never been, save by temporary and accidental confusion, in the least doubt. However if other Fellows of the Society greatly wish for its publication, in
spite of this opinion, I would raise no further objection, provided it is immediately followed by a demonstration that the cases here considered fall into the general class of simple regressions, and are demonstrable individually by a very simple application of the general proof.