A MULTIVARIATE ANALYSIS OF AN EXPERIMENTAL DESIGN INVOLVING A COMPLETE SET OF LATIN SQUARES

MAN-WAI CHAN*

INTERNATIONAL BUSINESS MACHINES CORPORATION, SAN JOSE, CALIFORNIA

A model for an experimental design involving a complete set of Latin squares for testing the homogeneity of treatment effects was constructed and analyzed by Gourlay. In his analysis, however, if one or both of the preliminary F-tests are significant, the analysis cannot differentiate. He then suggests the use of a less desirable test which is biased and has fewer degrees of freedom, regardless of the number of replications (the d.f. cannot be increased by increasing the replications). Further, when heterogeneity of variance occurs, Gourlay's test procedures are in general invalid. The present paper reviews Gourlay's analysis and proposes a modified test procedure.

1. Introduction

In many psychological experiments, individual variations among the subjects may affect the observations (e.g., test scores) to such an extent that it is difficult to discern significant differences between treatments (the variables tested) by means of the usual completely randomized design. In other words, differences between treatments are masked by differences between subjects. In such cases, it would hardly be practical to increase the block size and replication sufficiently to obtain a test with reasonable power for testing the homogeneity of treatment effects. To achieve economy in experiments of this type, and, in addition, to eliminate "order of presentation" effects, experimenters often employ a Latin-square design. In a Latin-square design, each subject serves as a block and receives a sequence of tests employing the different treatments to be compared. To form a complete set of Latin squares, subjects are often divided at random into as many groups as there are possible orders of presentation of the treatments, so that the order of presentation to each group is unique.

Several complications arise in applying the Latin-square design. First, for a small number of treatments, the number of degrees of freedom for error is too small. This difficulty, however, can easily be overcome by replicating the square and pooling the error mean square. More serious complications are: (i) the observations within a block, since they are obtained from the

*My thanks are due to Drs. R. S. Hirsch, R. L. Erdmann, and R. M. Simons of IBM Corporation for many stimulating discussions, and to Professor D. Teichroew of Stanford University for permission to refer to his paper [6] and his assistance. I also wish to thank B. A. Snyder for correcting many linguistic mistakes.
same subjects, are likely to be correlated; and (ii) the earlier treatments may affect, or interact with, those that follow. This second kind of interaction, producing the "residual effects" described by some writers, is classified as Type B interaction by Gourlay [3]. These two complications and the interaction between main effects (called Type A interaction by Gourlay) are the main reasons why the Latin-square design has long been criticized (see, for example, McNemar [5] and Chew [2]).

A comparatively large amount of work has been done on the problem of employing a Latin-square design when the interactions mentioned above may exist in the experiment. Among those works, the one by Gourlay [3] gives as a major result that, under certain conditions as stated in the assumptions of his model, for a valid application of Latin-square techniques it is not necessary for all interactions to be zero. His model incorporates a parameter representing interactions, and the design is replicated so that the parameters can be tested. The $F$-test bias, measured by the $B$-ratio, can be estimated from the table of expected mean squares.

Since Gourlay's model can be generalized to a design with a complete set of Greco-Latin squares of larger dimensions, and can widely and favorably be adapted for testing the homogeneity of treatment effects for many psychological experiments, it deserves further consideration and analysis. The purpose of this paper is twofold. (i) To provide a more desirable test for testing main effects when the result of one or both of the preliminary $F$-tests in Gourlay's analysis is significant—more desirable in the sense that it is unbiased and in general more powerful when the number of replications is greater than or equal to the number of cells in the Latin squares. (ii) To provide a valid test procedure when heterogeneity of variance occurs and Gourlay's test procedures are therefore invalid.

2. Gourlay's Model and His Analysis

Gourlay's model, shown below, was developed for the case of two $3 \times 3$ Latin squares corresponding to the six possible orders of treatment $ABC$, $BCA$, $CAB$, $ACB$, $BAC$, $CBA$.

These two squares are replicated $n$ times, i.e., each of the six groups consists of $n$ individuals chosen at random from a total of $6n$ individuals to whom the treatments are applied. Each individual goes through a sequence (row) of three treatments. This design is a complete set of permutations in the sense that each treatment in its ordinal position is preceded twice by each of the other two treatments in the hope that the residual effects can be eliminated in the analysis.

Each of the $n$ observations (scores) in any one of the eighteen cells has the structure

\[ Y_{ij(k)} = \mu + \pi_i + Q_j + T_{(k)} + N_{ij} + \xi_{ij} + \epsilon_{ij(k)} , \]