# Assessing interaction effects in Latin square-type designs

Rajeev KOHLI *

Latin, Graeco-Latin and hyper-Graeco-Latin squares are experimental designs in which all main effects are confounded with interaction effects involving two or more experimental factors. Most marketing research experiments using these designs blindly test for main effects without establishing that interaction effects are indeed not significant. This paper first shows how the presence of significant interaction effects can distort the results of experiments using Latin square-type designs. It then presents three procedures that test the assumption of insignificant interaction effects in these designs and discusses the conditions under which each method is best employed. The unique feature common to all three procedures is that they utilize the experimental data itself to test the validity of the additivity assumption. Finally, a new procedure is presented for replicated Graeco-Latin and hyper-Graeco Latin squares that unconfounds a single, major main effect from all second order interaction effects, and estimates one two-way interaction effect involving the major experimental factor of interest. Applications illustrating the usefulness of all four procedures are presented.

## 1. Introduction

Latin, Graeco-Latin and hyper-Graeco-Latin squares are related experimental designs using three, four and five experimental factors, respectively. Collectively, the three designs are referred to as Latin square-type designs in this paper. Holland and Cravens (1972: 272) and Peterson (1982: 601) note that these designs have been used more often than any others in marketing. Applications of Latin square-type designs to marketing and consumer research are described by Sawyer (1973), Edell and Staelin (1983), Mitchell and Olson (1981), Moore, Hausknecht and Thamodaran (1986), Sheluga, Jaccard and Jacoby (1979), Currim and Sarin (1983), Brunk and Federer (1953), Cox (1964), Hoofnagle (1963), and Smith, Clement and Hoofnagle (1956).

A major reason for the popularity of Latin square-type designs is that they use a small number of treatments (e.g., products, advertisements, prices) and experimental units (e.g., test cities, supermarkets, households), both of which can be expensive in many marketing experiments. However, a consequence of their economy is that these designs confound all main effects with interaction effects among two or more experimental factors. Latin square-type designs are therefore recommended when the main effect of only *one* experimental factor is of interest. The remaining factors are assumed to not interact with each other or with the principal experimental factor. However, in practice, interaction effects *can* be significant in many experiments using these designs. For example, among the previously noted studies, Currim and Sarin (1983) employ *product attributes*, and Edell and Staelin (1983) use *ad structure, ad content* and *product class* as factors in a Latin square; Mitchell and Olson (1981) employ *brand name / ad content* and *number of repetitions* as two of three factors in a Latin square; and Moore, Hausknecht and Thamodaran (1986) use *ads* and *speed of exposure* as two of four factors in a Graeco-

* Rajeev Kohli is Assistant Professor of Marketing, Joseph M. Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, USA.
The author thanks Rajiv Grover, Ajay Kohli, Vijay Mahajan, C.W. Park, Robert Small and Jerry Wind for helpful comments and suggestions. This research was supported by the Center for Marketing Strategy Research at the Wharton School, University of Pennsylvania.

Latin square. Even in studies in which only one main effect is of interest, the assumption that all interaction effects are insignificant is not so clear cut. For example, Sheluga, Jaccard and Jacoby (1979) use *subjects* and *order of testing* as factors in a Latin square in which the main effect of *preference scaling tasks* is of interest. Because it is possible for the effect of order of testing to vary differently across both tasks and subjects, the significance of interaction effects should not be assumed in such experiments. Rather, it should be *verified*, if possible from the *experimental data itself*. Hypotheses of interest should be tested only after confirming their interaction effects are not significant. Alternatively, if only one main effect is of interest and, as in many cases, a Latin square-type design is replicated, it is useful to perform the replication over a set of different squares across which the principal main effect of interest is not confounded with at least second order interaction effects, which generally are considered to be the interaction effects most likely to be significant.

The purpose of this paper is to describe a procedure by which these objectives can be attained. First, three methods for testing interaction effects in Latin-square-type designs are described, the conditions under which each is appropriate are discussed, and examples illustrating their usefulness are presented. Next, a procedure that ensures that a single, major main effect is not confounded with any second order interaction effects in Graeco-Latin and hyper-Graeco-Latin squares (but not in Latin squares) is described. A secondary benefit of this procedure is that it permits estimation of any *one* two-way interaction effect involving the main experimental factor of interest. This is a new procedure which, to our knowledge, has not been previously described in either the statistics or the marketing literatures.

## 2. Tests for interaction effects in Latin square-type designs

Before discussing the proposed testing procedures, consider an example that illustrates the erroneous conclusion that can result from using a Latin square-type design when interaction effects are significant. Suppose that a test marketing experiment is conducted using *products* ($T_i$), *prices* ($P_j$) and *ad budget* ($A_k$) as experimental factors, each at three levels. Assume, first, that the factors are used in a full-factorial design. The 27 combinations of test products, prices, and ad budgets are randomly assigned to, and simultaneously implemented in, two test cities each (i.e., 54 test-cities are employed in this hypothetical example). Per-capita sales over a three month period are recorded for each city (table 1a), their ANOVA (table 1b) indicating that the main effects of all three factors are significant, as are three of the four interaction effects. [1]

Now consider the nine italicized values in table 1a, which correspond to observations for a Latin square (table 2a). A researcher may elect to use the Latin square in an actual test-marketing experiment because it uses far fewer cities than the full-factorial design. But will the Latin square provide useful information about the effects of the three factors on sales? An ANOVA for the Latin square data (table 2b) suggests that it will not: all three main effects now appear insignificant, suggesting that none of the factors affect sales. Clearly the confounding of the main effects with the significant interaction effects seriously distorts the results of the Latin-square experiment. Even from a predictive standpoint, a comparison of the $\omega^2$ statistic (Hays (1963)), which measures the variation in the

[1] The data in table 1 are in fact for a brick-manufacturing experiment reported by Youden and Hunter (1955: 402, table 2). The current test-marketing context is assumed for expository purposes.

Table 1
Per capita sales ( × 100) and ANOVA for test marketing example employing a 3 × 3 × 3 factorial design.

(a) *Per capita sales data for test marketing experiment: 3 × 3 × 3 factorial design (2 observations per cell).*

| Advertising | Price | Test product | | |
|---|---|---|---|---|
| | | $T_1$ | $T_2$ | $T_3$ |
| $A_1$ | $P_1$ | 9.45, 9.61 | *9.33*, 9.68 | 9.62, 9.50 |
| | $P_2$ | 9.69, 9.60 | 9.44, 8.82 | *9.42*, 9.58 [a] |
| | $P_3$ | *9.64*, 9.64 | 9.49, 9.64 | 9.65, 9.74 |
| $A_2$ | $P_1$ | 9.05, 8.97 | 9.69, 9.27 | *9.08*, 8.92 |
| | $P_2$ | *9.36*, 9.46 | 9.25, 9.85 | 8.92, 9.04 |
| | $P_3$ | 9.40, 9.24 | *9.05*, 9.43 [a] | 9.50, 9.17 |
| $A_3$ | $P_1$ | *8.42*, 8.45 [a] | 8.48, 8.72 | 8.51, 8.81 |
| | $P_2$ | 8.68, 7.90 | *9.81*, 9.89 | 8.72, 8.79 |
| | $P_3$ | 8.45, 8.80 | 9.93, 10.20 | *8.90*, 9.02 |

(b) *ANOVA of 3 × 3 × 3 factorial design for test marketing experiment*

| Source | df | SS | MS | $F$ | $\omega^2$ |
|---|---|---|---|---|---|
| Test products | 2 | 1.2022 | 0.6011 | 13.183 ** | 0.089 |
| Prices | 2 | 0.7892 | 0.3946 | 8.654 ** | 0.056 |
| Ad. budget | 2 | 3.4130 | 1.7065 | 37.423 ** | 0.130 |
| Test products × prices | 4 | 0.1864 | 0.0466 | 0.978 | |
| Price × ad. budget | 4 | 2.7336 | 0.6834 | 14.987 ** | |
| Test products × prices × ad. budget | 8 | 1.7360 | 0.2170 | 4.759 ** | |
| Error | 27 | 1.2312 | 0.0456 | | |
| Total | 53 | 12.0108 | | | |

[a] Duplicate observations in the partially replicated Latin square of table 4.
** $p < 0.05$.

dependent variable accounted for by a factor, suggests that the main effects in the full-factorial design explain a much larger fraction of variance than the confounded main effects in the Latin square.

As the example illustrates, the validity if the results obtained from a Latin square-type experiment can be seriously hampered when interaction effects are significant. The effect of such confounding can be less extreme in other experiments, but the types of errors remain the same: significant main effects can appear insignificant, insignificant main effects can appear significant, and the predictive accuracy of the model can be reduced.

(a) *Latin square*

| | $C_1$ | $C_2$ | ... | $C_a$ |
|---|---|---|---|---|
| $A_1$ | $B_1$ | $B_2$ | ... | $B_a$ |
| $A_2$ | $B_2$ | $B_3$ | ... | $B_1$ |
| ⋮ | ⋮ | ⋮ | ⋮ | |
| $A_a$ | $B_a$ | $B_1$ | ... | $B_{a-1}$ |

(b) *Partially replicated Latin square*

| | $C_1$ | $C_2$ | ... | $C_a$ |
|---|---|---|---|---|
| $A_1$ | $B_1B_1'$ | $B_2$ | ... | $B_a$ |
| $A_2$ | $B_2$ | $B_3B_3'$ | ... | $B_1$ |
| ⋮ | ⋮ | ⋮ | ⋮ | |
| $A_a$ | $B_a$ | $B_1$ | ... | $B_{a-1}B_{a-1}'$ |

Fig. 1. Latin square and partially replicated Latin square.

## *2.1. Tukey's test for non-additivity in Latin square-type designs*

One method for assessing if there are significant interaction effects in a Latin

Table 2
Per capita sales ( × 100) and ANOVA for test marketing example employing a 3 × 3 Latin square.

(a) *Per capita sales ( × 100) for Latin square experiment* [a]

| Ad. Budget | Test product | | | | | |
|---|---|---|---|---|---|---|
| | $T_1$ | | $T_2$ | | $T_3$ | |
| | Prices | Sales | Prices | Sales | Prices | Sales |
| $A_1$ | $P_3$ | 9.64 | $P_1$ | 9.33 | $P_2$ | 9.42 |
| $A_2$ | $P_2$ | 9.36 | $P_3$ | 9.05 | $P_1$ | 9.08 |
| $A_3$ | $P_1$ | 8.42 | $P_2$ | 9.81 | $P_3$ | 8.90 |

(b) *ANOVA for Latin square experiment*

| Source | df | SS | MS | $F$ | $\omega^2$ |
|---|---|---|---|---|---|
| Test products | 2 | 0.1352 | 0.0676 | 0.2984 | 0 [b] |
| Prices | 2 | 0.7892 | 0.3946 | 0.6218 | 0 [b] |
| Ad. budget | 2 | 0.5196 | 0.2598 | 1.1505 | 0.042 |
| Residual | 2 | 0.4516 | 0.2258 | | |
| Total | 8 | 1.3872 | | | |

[a] Observations for the Latin square correspond to the italicized values in table 1.
[b] $\omega^2 = 0$ if the $F$-statistic is less than 1 (Hays (1963)).

Table 3
Analysis of MBA-student preferences for job profiles generated by a $4 \times 4$ Latin square.

(a) *Preference data*

| Type of job | Salary | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $32,000 | | $28,000 | | $25,000 | | $23,000 | |
| | Location | Rating | Location | Rating | Location | Rating | Location | Rating |
| Marketing | East Coast | 1.000 | Midwest | 0.667 | South | 0.594 | West Coast | 0.000 |
| Consulting | Midwest | 0.842 | South | 0.225 | West Coast | 0.625 | East Coast | 0.613 |
| Finance | South | 0.773 | West Coast | 0.553 | East Coast | 0.225 | MidWest | 0.531 |
| Accounting | West Coast | 0.553 | East Coast | 0.649 | MidWest | 0.030 | South | 0.735 |

(b) *Analysis of variance*

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Location | 3 | 0.018 | 0.0060 | 0.052 |
| Salary | 3 | 0.392 | 0.1307 | 1.122 |
| Type of job | 3 | 0.080 | 0.0270 | 0.232 |
| Residual | 6 | 0.699 | 0.1165 | |
| Total | 15 | 1.189 | | |

(c) *Analysis of variance following partitioning of residual for the non-additivity test*

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Location | 3 | 0.018 | 0.0060 | 0.10 |
| Salary | 3 | 0.392 | 0.1307 | 2.182 |
| Type of job | 3 | 0.080 | 0.0270 | 0.451 |
| Non-additivity | 1 | 0.3997 | 0.3997 | 6.673 ** |
| Error | 5 | 0.2993 | 0.0599 | |
| Total | 15 | 1.189 | | |

** $p < 0.05$.

square-type experiment is Tukey's non-additivity test. Consider the Latin square in fig. 1a. The experimental factors are denoted A, B and C, and their respective levels are denoted $A_i$, $B_j$ and $C_k$. Each factor appears at $a$ levels. The main effects model for the Latin square is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \quad (1)$$

where $y_{ijk}$ is the response variable, $\mu$ is the mean effect of treatments, $\alpha_i$, $\beta_j$, $\gamma_k$ are the effects associated with levels $A_i$, $B_j$, $C_k$, respectively, and $\epsilon_{ijk}$ is a normally distributed random error term with mean zero and variance $\sigma^2$. A standard ANOVA of the Latin square partitions the $y_{ijk}$ sums of squares into main effects and residual components. Tukey's procedure further partitions the residual sums of squares into non-additivity and error components (see Appendix A for details). The ratio of non-additivity- and error mean-squares is used to test if interaction effects are significant. If they are, the Latin square experiment is unlikely to yield meaningful results.

As an example, consider the data in table 3a regarding an MBA student's preferences for multi-attribute job descriptions. As in an earlier study by Currim and Sarin (1983), *salary, location* and *type of job* were used as factors in a $4 \times 4$ Latin square. An MBA

Table 4
Per capita sales and ANOVA for test marketing example employing a 3 × 3 Latin square.

(a) *Partially-replicated Latin-square*

| Ad. budget | Test product | | | | | |
|---|---|---|---|---|---|---|
| | $T_1$ | | $T_2$ | | $T_3$ | |
| | Price | Sales | Price | Sales | Price | Sales |
| $A_1$ | $P_3$ | 9.64 | $P_1$ | 9.33 | $P_2$ | 9.42, 9.58 [a] |
| $A_2$ | $P_2$ | 9.36 | $P_3$ | 9.05, 9.43 [a] | $P_1$ | 9.08 |
| $A_3$ | $P_1$ | 8.42, 8.45 [a] | $P_2$ | 9.81 | $P_3$ | 8.90 |

(b) *ANOVA for partially-replicated Latin-square*

| Source | df | SS | MS | *F* |
|---|---|---|---|---|
| Ad. budget | 2 | 0.7175 | 0.3588 | 12.59 ** |
| Test product corrected for ad. budget | 2 | 0.2737 | 0.1369 | 4.80 |
| Price corrected for test product and prices | 2 | 0.6663 | 0.3331 | 11.69 ** |
| Interaction | 2 | 0.4083 | 0.2042 | 7.165 ** |
| Error | 3 | 0.08545 | 0.0285 | |
| Total | 11 | 2.1512 | | |

[a] Duplicates correspond to observations market by a [a] in table 1.
** $p < 0.10$.

student evaluation the 16 job profiles according to Currim and Sarin's procedure, which combines conjoint analysis and von Neumann–Morgenstern utility theory. Table 3b presents the results of a standard ANOVA, which indicates that none of the job attributes significantly affect the student's job preferences. However, a partitioning of the residual into non-additivity and error components (table 3c) suggests the non-additivity component *is* significant, and that the conclusion of significant main effects may be erroneous. Thus, using a Latin square to generate job profiles is inappropriate here, and it is not possible to conclude whether or not the attributes significantly affect the student's job preferences.

As the example suggests, some of Currim and Sarin's subjects could also have had preferences that are in fact affected by salary, location and type of job, but which *appear* not to be because a Latin square was used to generate job profiles. An application of the non-additivity test would have been useful to identify subjects whose preferences are not appropriately modeled by the Latin square. Additional data from these subjects could then have been obtained to unconfound interaction effects, starting with two-way interactions. The result would have been a more valid assessment of individual risk attitudes and their classification into risk categories (segments), and a more accurate assessment of the predictive accuracy of the proposed methodology.

Observe that the main benefit of Tukey's test is that it uses no additional observations beyond those collected for the Latin square. However, if there are only a few residual degrees of freedom (e.g., in a 3 × 3 Latin square, 4 × 4 Graeco-Latin square, and 5 × 5 hyper-Graeco-Latin square), the further loss of 1 degree of freedom for Tukey's non-additivity term test can significantly reduce the precision of the statistical tests. In such cases, the replication methods described below are more appropriate.

*2.2. Testing interaction effects via partial or complete replication*

Partial replication (Youden and Hunter (1955)) or complete replication can also be used to test the additivity assumption in Latin square-type designs. In *partial replication*, duplicate observations are obtained in '$a$' cells of an '$a \times a$' Latin square-type design. A duplicate occurs in exactly one cell in each row and each column. For convenience, the cells in which the duplicates occur can be shown as lying along the main diagonal (fig. 1b). Randomizing the rows and columns changes the pattern of duplicated cells.

The duplicate observations provide an estimate of error variance, the within cells variation being entirely due to random fluctuations in responses to identical treatments. Interaction sums of squares are estimated as the difference between the sums of squares *unexplained* by the experimental factors, and the error sums of squares.

To illustrate, consider the previously discussed test-marketing example. Augment the nine original observations of the Latin square with three duplicate observations along the diagonal (table 4a). The duplicates are taken from the full-factorial experiment and correspond to the observations marked with a [a] in

Table 5
Fully replicated Latin square design and test for interaction effects = Hoofnagle's (1963) apple sales data. [a]

(a) *Experimental design: Apple sales by treatment and time periods in nine retail food stores*

| Store | Experiment period | | | | | |
|---|---|---|---|---|---|---|
| | Oct.–Nov. 11 | | Nov. 13–Nov. 25 | | Nov. 27–Dec. 9 | |
| | Treatment | Sales | Treatment | Sales | Treatment | Sales |
| | | Pounds | | Pounds | | Pounds |
| *First replication:* | | | | | | |
| Store 1 | A | 779 | B | 496 | C | 424 |
| Store 2 | B | 312 | C | 314 | A | 238 |
| Store 3 | C | 803 | A | 599 | B | 314 |
| *Second replication:* | | | | | | |
| Store 4 | A | 703 | B | 416 | C | 319 |
| Store 5 | B | 376 | C | 458 | A | 276 |
| Store 6 | C | 623 | A | 397 | B | 556 |
| *Third replication:* | | | | | | |
| Store 7 | A | 557 | B | 382 | C | 346 |
| Store 8 | B | 313 | C | 489 | A | 396 |
| Store 9 | C | 170 | A | 211 | B | 85 |

(b) *ANOVA for fully replicated Latin square including test for interaction effects*

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Stores | 8 | 448,801 | 56,100.13 | 6.169 ** |
| Time periods | 2 | 157,245.5 | 78,627.25 | 8.646 ** |
| Treatments | 2 | 49,976.3 | 24,988.15 | 2.7480 * |
| Interaction | 2 | 51,029 | 25,514.5 | 2.805 * |
| Error | 12 | 109,134.2 | 9,094.52 | |
| Total | 26 | | 816,195 | |

[a] Treatments represent color ranges of apples.
* $p < 0.10$.
** $p < 0.025$.

table 1a. Table 4b presents the ANOVA for the partially replicated Latin square. The large value of the interaction mean squares relative to the error mean squares gives warning that the main effects are confounded with significant interaction effects (note that even with 3 df for error, the $F$-value for interaction effects is 7.165, which is significant at less than the 10% level). Thus, if the experiment were conducted using the Latin square, the sales data from three additional cities (i.e., the duplicates) could be used to at least caution the researcher that the main effects are not what they appear to be, and that inappropriate strategic conclusions can be drawn from the experiment.

A *completely replicated* Latin square-type design has $r$ ($\geq 2$) observations in *each* cell. Like partial replication, it is recommended over Tukey's procedure when there are only a few residual degrees of freedom. However, when resources permit, complete replication should be preferred for two reasons. First, because each cell has the same number of observations, all main effects are orthogonal. Second, the error variance is based on a significantly larger number of degrees of freedom, and hence the statistical tests have greater precision.

As for partial replication, the within-cells variability in a fully-replicated design provides an estimate of the error variance. An estimate of interaction sums of squares is therefore obtained as the difference between the sums of squared unexplained by the main effects and the sums of squares due to error.

As an example of the complete replication procedure, consider the data in table 5a from Hoofnagle (1963: 155). A $3 \times 3$ Latin square is used to study the effect of color of apples on their sales. *Test stores, time periods* and *color of apples* are used as design factors. The Latin square is replicated three times (i.e., $r = 3$), different supermarkets being used in each replication. During each time period, apples of a specific color are simultaneously marketed in three stores, one store per Latin square. All stores carry each color range in one of the three time periods.

An ANOVA of the sales data is reported in table 5b. Unlike the original analysis by Hoofnagle, it uses the replicates to test if the main effects are confounded with significant interaction effects. The analysis indicates that interaction mean squares are nearly three times the error mean square (the corresponding $F$-value is marginally significant at the 10% level). It is therefore possible that significant interaction effects confound the main effects, and that sales are *not* actually affected by the color of apples.

### *2.3. Rotation of Graeco-Latin and hyper-Graeco-Latin squares*

The preceding procedures only test for the *presence* of significant interaction effects in Latin square-type designs. However, they do not *unconfound* main and interaction effects. A procedure is now presented to ensure that a *single*, major main effect of interest is not be confounded with *any* second-order interaction effect in Graeco-Latin or hyper-Graeco-Latin squares, and to also permit the estimation of any *one* second-order interaction effect involving the major factor of interest. The procedure is not applicable to Latin squares, because in this case it results in a full-factorial design.

The proposed procedure utilizes an initial '$a \times a$' Graeco-Latin (hyper-Graeco-Latin) square to generate a set of $a$ Graeco-Latin (hyper-Graeco-Latin) squares. Only the 'new' squares are employed in the experiment, the initial square serving as a 'seed' for generating these squares. To facilitate exposition, we develop the procedure via an example for a $5 \times 5$ Graeco-Latin square. Subsequently, we generalize the procedure to Graeco-Latin squares of any size, and then to hyper-Graeco-Latin squares.

(a) *'Seed' Graeco-Latin square ($S_0$)*

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ |
|---|---|---|---|---|---|
| $A_1$ | $B_1D_1$ | $B_2D_2$ | $B_3D_3$ | $B_4D_4$ | $B_5D_5$ |
| $A_2$ | $B_2D_3$ | $B_3D_4$ | $B_4D_5$ | $B_5D_1$ | $B_1D_2$ |
| $A_3$ | $B_3D_5$ | $B_4D_1$ | $B_5D_2$ | $B_1D_3$ | $B_2D_4$ |
| $A_4$ | $B_4D_2$ | $B_5D_3$ | $B_1D_4$ | $B_2D_5$ | $B_3D_1$ |
| $A_5$ | $B_5D_4$ | $B_1D_5$ | $B_2D_1$ | $B_3D_2$ | $B_4D_5$ |

(b) *Graeco-Latin squares generated by 'rotating' $S_0$*

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | |
|---|---|---|---|---|---|---|
| $A_1$ | $B_1D_1$ | $B_2D_2$ | $B_3D_3$ | $B_4D_4$ | $B_5D_5$ | |
| $A_2$ | $B_2D_3$ | $B_3D_4$ | $B_4D_5$ | $B_5D_1$ | $B_1D_2$ | |
| $A_3$ | $B_3D_5$ | $B_4D_1$ | $B_5D_2$ | $B_1D_3$ | $B_2D_4$ | ($S_1$) |
| $A_4$ | $B_4D_2$ | $B_5D_3$ | $B_1D_4$ | $B_2D_5$ | $B_3D_1$ | |
| $A_5$ | $B_5D_4$ | $B_1D_5$ | $B_2D_1$ | $B_3D_2$ | $B_4D_5$ | |

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | |
|---|---|---|---|---|---|---|
| $A_2$ | $B_2D_3$ | $B_3D_4$ | $B_4D_5$ | $B_5D_1$ | $B_1D_2$ | |
| $A_3$ | $B_3D_5$ | $B_4D_1$ | $B_5D_2$ | $B_1D_3$ | $B_2D_4$ | |
| $A_4$ | $B_4D_2$ | $B_5D_3$ | $B_1D_4$ | $B_2D_5$ | $B_3D_1$ | ($S_2$) |
| $A_5$ | $B_5D_4$ | $B_1D_5$ | $B_2D_1$ | $B_3D_2$ | $B_4D_5$ | |
| $A_1$ | $B_1D_1$ | $B_2D_2$ | $B_3D_3$ | $B_4D_4$ | $B_5D_5$ | |
| ⋮ | | | ⋮ | | ⋮ | ⋮ |

| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | |
|---|---|---|---|---|---|---|
| $A_5$ | $B_5D_4$ | $B_1D_5$ | $B_2D_1$ | $B_3D_2$ | $B_4D_5$ | |
| $A_1$ | $B_1D_1$ | $B_2D_2$ | $B_3D_3$ | $B_4D_4$ | $B_5D_5$ | |
| $A_2$ | $B_2D_3$ | $B_3D_4$ | $B_4D_5$ | $B_5D_1$ | $B_1D_2$ | ($S_5$) |
| $A_3$ | $B_3D_5$ | $B_4D_1$ | $B_5D_2$ | $B_1D_3$ | $B_2D_4$ | |
| $A_4$ | $B_4D_2$ | $B_5D_3$ | $B_1D_4$ | $B_2D_5$ | $B_3D_1$ | |

Fig. 2. 'Rotation' procedure for generating five Graeco-Latin squares ($S_1$–$S_5$) from a 'seed' Graeco-Latin square ($S_0$).

Fig. 2a displays a $5 \times 5$ Graeco-Latin square ($S_0$) that is utilized as the seed for the five new Graeco-Latin squares ($S_1$–$S_5$) shown in fig. 2b. Let the $i$th row of $S_0$ be called *ordered array* $O_i$. For example, the following ordered array appears in the third row of $S_0$: $O_3 = B_3\ D_5\ B_4\ D_1\ B_5\ D_2\ B_1\ D_3\ B_2\ D_4$. In each of the five new Graeco-Latin squares, levels $A_1$–$A_5$ are assigned to rows 1–5, and levels $C_1$–$C_5$ are assigned to columns 1–5. Ordered arrays $O_1$–$O_5$ are assigned to the rows of $S_1$–$S_5$ in a manner to be described below. For the moment, note that once an ordered array is assigned to row $A_i$ of a Graeco-Latin square, its first term is written in the cell identified by row $A_i$-column $C_1$, its second element in the cell identified by row $A_i$-column $C_2$, etcetera. Thus, a Graeco-Latin square ($S_1$–$S_5$) is completely specified once each of its row is assigned an ordered array.

To assign ordered arrays to rows, write a $5 \times 5$ Latin square whose columns correspond to *Graeco-Latin squares* ($S_1$–$S_5$), rows to *rows of Graeco Latin squares* ($A_1$–$A_5$) and cell entries to *ordered arrays* ($O_1$–$O_5$):

| | Graeco-Latin Square | | | | |
|---|---|---|---|---|---|
| Row of Graeco-Latin Square | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ |
| $A_1$ | $O_1$ | $O_2$ | $O_3$ | $O_4$ | $O_5$ |
| $A_2$ | $O_2$ | $O_3$ | $O_4$ | $O_5$ | $O_1$ |
| $A_3$ | $O_3$ | $O_4$ | $O_5$ | $O_1$ | $O_2$ |
| $A_4$ | $O_4$ | $O_5$ | $O_1$ | $O_2$ | $O_3$ |
| $A_5$ | $O_5$ | $O_1$ | $O_2$ | $O_3$ | $O_4$ |

Each column of the Latin square identifies a new Graeco-Latin square ($S_1$–$S_5$). For example, $S_2$ is obtained by assigning ordered arrays $O_2$–$O_5$ to rows $A_1$–$A_4$, respectively, and by assigning ordered array $O_1$ to row $A_5$. Observe that each ordered array appears *exactly once* in each new Graeco-Latin square, so that Graeco-Latin squares $S_1$–$S_5$ differ only with respect to which ordered arrays are assigned to which rows. The Latin square only ensures that, across $S_1$–$S_5$, each ordered array is assigned exactly once to each row. Because the Latin square rotates, across $S_1$–$S_5$, the rows associated with an ordered array, we call this technique a 'rotation' procedure.

The procedure immediately generalizes to Graeco-Latin squares in which each factor has '$a$' levels. The ordered arrays of an initial ('seed') "$a \times a$" Graeco-Latin square are used to generate '$a$' new Graeco-Latin squares. The rows in which the ordered arrays appear in each Graeco-Latin square are, in turn, determined by an '$a \times a$' Latin square that employ *Graeco-Latin squares, ordered arrays* and *rows of Graeco-Latin squares* as factors.

While the ordered arrays in Graeco-Latin squares consist of combinations of levels of two factors (B and D), they consist of com-

Table 6. [a]

(a) *'Rotated' Graeco-Latin squares for car mileage study* [a]

| Drivers | Cars | | | Drivers | Cars | | | Drivers | Cars | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $C_3$ | | $C_1$ | $C_2$ | $C_3$ | | $C_1$ | $C_2$ | $C_3$ |
| $D_1$ | $G_1P_1$ | $G_3P_2$ | $G_2P_3$ | $D_1$ | $G_2P_3$ | $G_1P_1$ | $G_3P_2$ | $D_1$ | $G_3P_2$ | $G_2P_3$ | $G_1P_1$ |
| $D_2$ | $G_2P_2$ | $G_1P_3$ | $G_3P_1$ | $D_2$ | $G_3P_1$ | $G_2P_2$ | $G_1P_3$ | $D_2$ | $G_1P_3$ | $G_3P_1$ | $G_2P_2$ |
| $D_3$ | $G_3P_3$ | $G_2P_1$ | $G_1P_2$ | $D_3$ | $G_1P_2$ | $G_3P_3$ | $G_2P_1$ | $D_3$ | $G_2P_1$ | $G_1P_2$ | $G_3P_3$ |

(b) *ANOVA for 'rotated' Graeco-Latin square for car mileage study*

| Source | df | SS | MS | *F* |
|---|---|---|---|---|
| Cars | 2 | 72 | 36 | 10.75 ** |
| Gasoline grades | 2 | 18 | 9 | 2.687 |
| Drivers | 2 | 54 | 27 | 8.06 ** |
| Order-of-driving | 2 | 54 | 27 | 8.06 ** |
| Cars × Gasoline-grades | 4 | 36 | 18 | 5.373 ** |
| Error | 14 | 47 | 3.35 | |
| Total | 26 | 281 | | |

[a] Cell entries denote performance in miles per gallon.
** $p < 0.05$.

binations of levels of three factors (B, D and E) in hyper-Graeco-Latin squares. In all other respects, the procedure for generating the set of hyper-Graeco-Latin squares is the same as for Graeco-Latin squares.

For both Graeco-Latin and hyper-Graeco-Latin squares, the proposed procedure results is a set of squares across which the main effect of factor A is not confounded with any second order interaction effect. Additionally, the interaction effect between the A and C factors is also estimable, so that the design can also be used when this interaction effect is expected to be significant. Derivation of these properties of rotated Graeco-Latin and hyper-Graeco-Latin squares are presented in Appendix B.

As an example of the rotation procedure, suppose that three *cars*, three *drivers* and three *grades of gasoline* are used in a study comparing the fuel efficiency of three car models. A Graeco-Latin square is used for the experiment (table 6a), the rows ($D_i$) corresponding to drivers and the columns ($C_k$) to cars. $G_1$, $G_2$ and $G_3$ are the three grades of gasoline. Finally, $P_1$ (first), $P_2$ (second) and $P_3$ (third) are the three *positions* in the sequence in which the cars are tested by each driver. Since each of $P_1$, $P_2$ and $P_3$ appear exactly once in every column, the cars associated with $P_1$, $P_2$ and $P_3$ are tested first, second and third, respectively, by each driver. Thus, systematic differences among drivers, gasoline grades and order of testing are eliminated in the Graeco-Latin square if all interaction effects are assumed insignificant. However, this assumption is inappropriate because the interaction effect between cars and grades of gasoline can be significant if the fuel efficiency of cars varies differently by the types of gasoline.

The proposed rotation of Graeco-Latin squares is appropriate in this case because it unconfounds the main effect of cars from all second-order interaction effects, and also permits the interaction effect between cars and gasoline grades to be estimated. Note that the rotation does not involve any major additional expenditure, because the same drivers and cars are used across the rotations. Note also that, in a single $3 \times 3$ Graeco-Latin square, there are no degrees of freedom available for error, so that replications of the Graeco-Latin square are anyway required to estimate error variance. The rotation procedure is gainfully employed in this case because the only change across the Graeco-Latin squares it generates is in the order in which the three cars are tested by each driver, and the gasoline grade used in a car when it is tested by the same driver across replicates.

The three Graeco-Latin squares produced by the rotation procedure are shown in table 6a. Suppose that the fuel efficiency study yields the results shown, the response variable being noted in miles per gallon. An ANOVA of these data appears in table 6b. The interaction effect between cars and gasoline grades is significant, as is the main effect of cars. That is, some cars are more efficient than others, and the relative efficiency of cars also varies by the gasoline grade employed. Had a mere replication of the Graeco-Latin squares been performed instead of their rotation, the interaction effect between cars and gasoline grades would not have been detected. Further, this significant interaction would have confounded the main effects, which as a result would be either exaggerated or subdued.

## 3. Conclusion

To use Latin square-type designs, a researchers must assume that all interaction effects are insignificant. This paper discusses how Tukey's non-additivity test, and partial or complete replication can be used to assess the validity of the 'additivity' assumption. A 'rotation' procedure is then presented so that, across a set of Graeco-Latin (hyper-Graeco-Latin) squares, the main effect of a single, major factor is not confounded with any second order interaction effect. This procedure also permits estimation of any one second-order interaction effect involving the major experimental factor of interest.

The choice between Tukey's non-additivity test and the two replication based tests depends on the size of the design and the resources available to a researcher. If losing a single degree of freedom for non-additivity significantly reduces the precision of statistical tests, partial or complete replication procedures are to be preferred. Between the two replication procedures, complete replication is preferred because the effects are orthogonally estimated and there are more degrees of freedom for error. However, complete replication also requires more experimental units. Thus, a researcher must make a trade-off between the resources required for collecting more data and the precision of the results. The rotation procedure for Graeco-Latin and hyper-Graeco-Latin squares requires the greatest number of observations, and should be employed when second order interactions are likely to be significant, so that at least the main effect of the principal experimental factor can be guaranteed to be unconfounded. Moreover, this design permits estimation of any one, two-way interaction involving the main experimental factor, so that hypothesis regarding such an interaction, if of interest, can be tested. Of course, there are a variety of other fractional factorials that can be more suitable than a Latin square-type design in many experimental settings. In these cases, a researcher should investigate alternative designs rather than restrict attention to the seemingly simple family of Latin squares.

## Appendix A

*Tukey's non-additivity test for Latin square-type designs*

The non-additivity component for Tukey's test for an '$a \times a$' Latin Square is computed as follows. Let

$\bar{Y}$ = mean of all $a^2$ responses,
$Y_i$ = mean for $A_i$ (row $i$),

$Y_j =$ mean for $B_j$,
$Y_k =$ mean for $C_k$ (column $k$),
where $A$, $B$ and $C$ are the experimental factors (fig. 1a). The ANOVA model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \tag{A.1}$$

leads to the following parameter estimates:

$$\hat{\alpha}_i = Y_i - \overline{Y}, \quad \hat{\beta}_j = Y_j - \overline{Y}, \quad \hat{\gamma}_k = Y_k - \overline{Y}, \tag{A.2}$$

and the fitted (predicted) values

$$\hat{Y}_{ijk} = \overline{Y} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k. \tag{A.3}$$

Let

$$Z_{ijk} = \hat{Y}^2_{ijk}. \tag{A.4}$$

Use $Z_{ijk}$ as a dependent variable in a main effects model in which in the $A$, $B$ and $C$ factors are again the independent variables; i.e.,

$$Z_{ijk} = \Phi + \delta_i + \theta_j + \rho_k + \Omega_{ijk}, \tag{A.5}$$

where $\Phi$ is the mean effect of treatments, $\delta_i$, $\theta_j$, $\rho_k$ are the effects associated with $A_i$, $B_j$, $C_k$, respectively, and $\Omega_{ijk}$ is a normally distributed error term. The ANOVA model (A.5) leads to the following parameter estimates:

$$\hat{\delta}_i = Z_i - \overline{Z}, \quad \hat{\theta}_j = Z_j - \overline{Z}, \quad \hat{\rho}_k = Z_k - \overline{Z}, \tag{A.6}$$

and the fitted values

$$\hat{Z}_{ijk} = \overline{Z} + \hat{\delta}_i + \hat{\theta}_j + \hat{\rho}_k. \tag{A.7}$$

In expressions (A.6) and (A.7), $\overline{Z}$ is the grand mean of the $Z$-variable, and $Z_i$, $Z_j$ $Z_k$ are the means of the $Z$-values associated with $A_i$, $B_j$, $C_k$, respectively. The non-additivity sums of squares are shown by Tukey (1955) to be

$$SS_{\text{nonadd}} = \frac{\Sigma_{i,j}\left[\left(Z_{ijk} - \hat{Z}_{ijk}\right)\left(Y_{ijk} - \hat{Y}_{ijk}\right)\right]^2}{\Sigma_{i,j}\left(Z_{ijk} - \hat{Z}_{ijk}\right)^2}. \tag{A.8}$$

The error sums of squares are obtained as the difference between the residual sums of squares (i.e., the sums of squares unexplained by the main effects of $A$, $B$ and $C$) and the non-additivity sums of squares. The non-additivity mean squares ($MS_{\text{nonadd}}$) are based on one degree of freedom (i.e., $MS_{\text{nonadd}} = SS_{\text{nonadd}}$), and the error mean squares (MSE) on $(a-1)(a-2)-1$ degrees of freedom. The statistic ($MS_{\text{nonadd}}$/MSE) has an $F$-distribution, a significant value of the statistic indicating that significant interactions do indeed confound the main effects in the Latin square.

Expression (A.8) can be generalized to test for interaction effects in Graeco-Latin and hyper-Graeco-Latin squares. A Graeco-Latin square employs a fourth factor ($D$) in addition to the three factors ($A$, $B$, $C$) in a Latin square. Therefore the main-effects models (A.1) and (A.5) include additional parameters for the effects of factor $D$. The estimates of these parameters contribute to $\hat{Y}$ and $\hat{Z}$, which in turn enter into the computation of the non-additivity sums of squares (A.8). Similarly, a hyper-Graeco-Latin square employs a fifth factor ($E$) in addition to the four factors ($A$, $B$, $C$, $D$) in a Graeco-Latin square. Therefore the main-effects models (A.1) and (A.5) include additional parameters for the effects of factor $E$, the estimates of these effects contribute to $\hat{Y}$ and $\hat{Z}$, which in turn enter into the computation of the non-additivity sums of squares.

## Appendix B

### *Confounding patterns in rotated Graeco-Latin and hyper-Graeco-Latin squares*

We begin by deriving the results for an '$a \times a$' Graeco-Latin square in which $A$, $B$, $C$ and $D$ are the experimental factors. First write the full, fixed effects model containing terms for all main effects and interaction effects for the four factors:

$$\begin{aligned} y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} \\ + (\alpha\delta)_{il} + (\beta\gamma)_{jk} + (\beta\delta)_{jl} + (\gamma\delta)_{kl} \\ + (\alpha\beta\gamma)_{ijk} + (\alpha\beta\delta)_{ijl} + (\alpha\gamma\delta)_{ikl} \\ + (\beta\gamma\delta)_{jkl} + (\alpha\beta\gamma\delta)_{ijkl} + \epsilon_{ijkl}. \end{aligned} \quad \text{(B.1)}$$

In expression (B.1), $y_{ijkl}$ is the response variable; $\mu$ is the mean response; $\alpha_i$, $\beta_i$, $\gamma_k$, $\delta_l$, are the main effects associated with $A_i$, $B_j$, $C_k$, $D_l$, respectively; the bracketed terms with two, three and four elements are the second, third, and fourth order interactions, respectively, among some or all of factors $A$, $B$, $C$, and $D$; and $\epsilon_{ijkl}$ is a normally distributed, random error term.

The following standard assumptions are associated with the linear-model (B.1):

$$\begin{aligned} &\Sigma_i \alpha_i = \Sigma_j \beta_j = \Sigma_k \gamma_k = \Sigma_l \delta_l = 0, \\ &\Sigma_i (\alpha\beta)_{ij} = \Sigma_j (\alpha\beta)_{ij} = \Sigma_i (\alpha\gamma)_{ik} \\ &\quad = \ldots \Sigma_k (\gamma\delta)_{kl} = \Sigma_l (\gamma\delta)_{kl} = 0, \\ &\Sigma_i (\alpha\beta\gamma)_{ijk} = \Sigma_j (\alpha\beta\gamma)_{ijk} = \Sigma_k (\alpha\beta\gamma)_{ijk} \\ &\quad = \ldots \Sigma_j (\beta\gamma\delta)_{jkl} \\ &\quad = \Sigma_k (\beta\gamma\delta)_{jkl} = \Sigma_l (\beta\gamma\delta)_{jkl} = 0, \\ &\Sigma_i (\alpha\beta\gamma\delta)_{ijkl} = \Sigma_j (\alpha\beta\gamma\delta)_{ijkl} = \Sigma_k (\alpha\beta\gamma\delta)_{ijkl} \\ &\quad = \Sigma_l (\alpha\beta\gamma\delta)_{ijkl} = 0. \end{aligned} \quad \text{(B.2)}$$

Define the following totals across Graeco-Latin squares:

- $G$ = grand total of observations across Graeco-Latin squares,
- $T_i$ = total of observations in which level $A_i$ of factor $A$ appears,
- $T_j$ = total of observations in which level $B_j$ of factor $B$ appears,
- $T_k$ = total of observations in which level $C_k$ of factor $C$ appears,
- $T_l$ = total of observations in which level $D_l$ of factor $D$ appears.

It follows from (B.1) and (B.2) that

$$\begin{aligned} T_i &= a^2[\mu + \alpha_i] + (\beta\gamma\delta)\text{-terms} \\ &\quad + (\alpha\beta\gamma\delta)\text{-terms} + \text{error}, \\ T_j &= a^2[\mu + \beta_j] + (\gamma\delta)\text{-terms} \\ &\quad + (\beta\gamma\delta)\text{-terms} + \text{error}, \\ T_k &= a^2[\mu + \tau_k] + (\gamma\delta)\text{-terms} \\ &\quad + (\beta\gamma\delta)\text{-terms} + \text{error}, \\ T_l &= a^2[\mu + \delta_l] + (\beta\gamma)\text{-terms} \\ &\quad + (\beta\gamma\delta)\text{-terms} + \text{error}. \end{aligned} \quad \text{(B.3)}$$

Observe that $T_i$ involves no second-order-interaction terms. Consequently, assuming all third and fourth order interaction effects are insignificant, the expected value of $T_i$ is

$$\mathrm{E}[T_i] = a^2[\mu + \alpha_i],$$

and therefore the estimate of $\alpha_i$ (the main effect for level $A_i$) is

$$\hat{\alpha}_i = [T_i/a^2] - \hat{\mu}, \quad \text{(B.4)}$$

where $\hat{\mu} = G/a^3$ is the estimate of the mean effect. Note that because $T_i$ does not contain any second-order interaction terms, the main effect of factor $A$ is not confounded with any second order interaction effect. Further, the expressions for $T_j$, $T_k$ and $T_l$ do not involve the $(\alpha\tau)$-terms, which are associated with the $A \times C$ interaction effect. Because all other interaction effects are assumed to be insignificant, the expected values of these terms are

$$\begin{aligned} \mathrm{E}(T_j) &= a^2[\mu + \beta_j], \\ \mathrm{E}(T_k) &= a^2[\mu + \gamma_k], \\ \mathrm{E}(T_l) &= a^2[\mu + \delta_l], \end{aligned} \quad \text{(B.5)}$$

and the main effects estimates

$$\begin{aligned} \hat{\beta}_j &= [T_j/a^2] - \hat{\mu}, \\ \hat{\gamma}_k &= [T_k/a^2] - \hat{\mu}, \\ \hat{\delta}_l &= [T_l/a^2] - \hat{\mu}, \end{aligned} \quad \text{(B.6)}$$

for $B_j$, $C_k$ and $D_l$, respectively.

As previously noted, the $A \times C$ interaction effect (i.e., the $\alpha\gamma$-terms) are also estimable in the above design. To see this, consider the sum across Graeco-Latin squares of the '$a$' terms in the $ik$th cell (i.e., the cell identified by row $i$, column $k$). This sum ($T_{ik}$) is

$$T_{ik} = a\left[\mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik}\right] + (\beta\delta) \text{ terms} + (\alpha\beta\delta) \text{ terms} + (\alpha\beta\gamma\delta) \text{ terms} + \text{error}. \quad (B.7)$$

Since only the main effects and the $A \times C$ interaction effects are assumed significant, the expected value of $T_{ik}$ is

$$\mathrm{E}\left[T_{ik}\right] = a\left[\mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik}\right]. \quad (B.8)$$

It follows from (B.6), (B.7) and (B.8) that

$$(\alpha\gamma)_{ik} = 1/a\left[T_{ij} - a\hat{\alpha}_i - a\hat{\gamma}_k + a\hat{\mu}\right], \quad (B.9)$$

which is the expression for the $ik$th term of the $A \times C$ interaction.

For hyper-Graeco-Latin squares, an analysis similar to that above shows that the main effects are not confounded with the $A \times C$ interaction terms in the set of hyper-Graeco-Latin squares produced by the rotation procedure. Further, the $A \times C$ interaction effect is also estimable across this set of hyper-Graeco-Latin squares.

## References


Box, George E.P., William G. Hunter and J. Stuart Hunter, 1978. Statistics for experimenters. New York: Wiley.

Brunk, Max E. and Walter T. Federer, 1953. Experimental designs and probability sampling in marketing research. American Statistical Association Journal 48 (September), 440–452.

Cox, Keith, 1964. The responsiveness of food sales to shelf space changes in supermarkets. Journal of Marketing Research vol. 1 (May), 63–67.

Currim, Imran S. and Rakesh K. Sarin, 1983. A procedure for measuring and estimating consumer preferences under uncertainty. Journal of Marketing Research 20 (August), 249–256.

Edell, Julie A. and Richard Staelin, 1983. The information processing of pictures in print advertisements. Journal of Consumer Research 10 (June), 45–61.

Hays, William L., 1963. Statistics for psychologists. New York: Holt, Rinehart and Winston.

Holland, Charles W. and David W. Cravens, 1973. Fractional factorial experimental designs in marketing research. Journal of Marketing Research 10 (August), 270–276.

Hoofnagle. William S., 1963. Experimental designs in measuring the effectiveness of promotion. Journal of Marketing Research 2 (May), 154–162.

Mitchell, Andrew A. and Jerry C. Olson, 1981. Are product attribute beliefs the only mediator of advertising effects on brand attitude?. Journal of Marketing Research 18 (August), 318–332.

Moore, Danny L., Douglas Hausknecht and Kanchana Thamodaran, 1986. Time compression, response opportunity, and persuasion. Journal of Consumer Research 13 (June), 85–99.

Peterson, Robert A., 1982. Marketing research. Plano, TX: Business Publications, Inc.

Sawyer, Alan G., 1973. The effects of repetition of refutational and supportive appeals. Journal of Marketing Research 10 (February), 23–33.

Sheluga David A., James Jaccard and Jacob Jacoby, 1979. Preference, search, and choice: An integrative approach. Journal of Consumer Research 6 (September), 166–176.

Smith, Hugh M., Wendell E. Clement and William S. Hoofnagle, 1956. Merchandising natural cheddar cheese in retail food stores. MRR-115 (Washington, DC: U.S. Department of Agriculture) April.

Tukey, John S., 1949. One degree of freedom for non-additivity. Biometrics 5 (September), 232–242.

Tukey, John S., 1955. Queries. Biometrics 11 (March), 111–113.

Youden, William J. and J. Stuart Hunter, 1955. Partially replicated Latin squares. Biometrics vol. 11 (December), 399–405.