TABLE 5

Comparison of Two Proprietary Powder Detergents

| Detergent | Type | Fabric | Terg-O-Tometer reflectance vs standard (100%)[a] Clay | Spangler | Bundle % preference (UV light) |
|---|---|---|---|---|---|
| E | STPP built/anionic powder with bleach and enzymes | Cotton | 100 | 104 | 53 |
| | | PE/cotton | 99 | 102 | 42 |
| | | Overall | | | 48 |
| F | STPP built/anionic powder with enzymes | Cotton | 99 | 103 | 47 |
| | | PE/cotton | 101 | 99 | 58 |
| | | Overall | | | 52 |

[a]AATCC standard with 50% STPP and 17% anionic.

visual preferences of detergents E and F were plotted on a weekly basis as shown in Figure 3, the two products were essentially equivalent in performance during the first nine weeks of the test period with no preference trends developing for either detergent. The preference for detergent F in the tenth week may indicate that the test should be extended for a few more weeks to determine if the preference was just noise or the beginning of a significant performance trend. Once again, this demonstrates the necessity to evaluate all of the data available.

Bundle testing provides a realistic evaluation of performance that bridges the gap between laboratory screening tests and extended consumer testing. The bundle test has proven to be an invaluable cost effective tool in the new product evaluation cycle. Information obtained from the bundle test often shows the necessity for additional laboratory development work before proceeding on to more extended and expensive consumer testing.

# Predicting a bundle test winner

*The following, based on a talk given by Paul X. Riccobono, was prepared by Riccobono and Richard Polanski, both of Colgate Palmolive Co., Piscataway, New Jersey.*

The ultimate objective of product development is introducing profitable new products into the marketplace. Usually it is known in the early stages of development whether a product is new or not. Profitability is not as easy to determine and is dependent on a number of interrelated factors, not the least of which is consumer acceptance. This is particularly true in the laundry products area, where the consumer's ability to discern a point of difference in performance between a new or improved product entry and a product already in the marketplace at times rests on rather tenuous differences in sensory perception.

Unfortunately, the ability to accurately predict by quantitative laboratory tests the effects of new laundry detergent compositions on consumer perception has remained largely an unfulfilled goal of research workers in this area. Usually, the evaluation of a new laundry formulation involves a progression of testing methodologies, from simple laboratory determinations of detergency utilizing the Terg-O-Tometer, to complex consumer-testing involving hundreds of participants. The ultimate are the sales tests, in which entire cities or regions of the country are involved (1-3). At some point in the process, the new product must be taken out of the laboratory and handed over to consumers for their judgment. The decision to consumer-test a new product is a critical point in the product development process for it involves large sums of money and considerable amounts of time and human effort. At Colgate, the decision to consumer-test a new laundry product is often made only after results of a bundle test have been evaluated and factored into the decision-making process.

In the arsenal of laboratory test methods available today, the bundle test is generally acknowledged by the detergent industry as closest to typical consumer response (4). Formalized as an ASTM method in 1972 (5), it today is the principle "bridge" between the closely controlled laboratory-testing of the formulation chemist and the variability of the real world.

Useful as it is, the bundle test is a rather long and tedious procedure. A decision to proceed with the bundle testing of a formulation—which typically takes six to eight weeks—is itself a decision of some significance. Thus, the ability to accurately predict the outcome (i.e., the visually preferred product) of a bundle test in one or two days would be of considerable value to a product development staff. It is this problem which is the subject of this paper.

Our approach was to construct via multiple regression analysis a mathematical model that would predict the visually preferred product of a two-product bundle test. The equation we derived to determine the visual preference ratio (VPR) was based on Terg-O-Tometer studies of laboratory-soiled swatches which were performed concurrently with the bundle test and using the same test product. The reflectance data accumulated in the Terg-O-Tometer studies along with the visual preference votes for the items obtained after the last cycle of the bundle tests served as the database for the regression analysis. Bluing and brightening terms derived from data obtained from clean cotton percale swatches were also included in our regressions. The equation we derived was then used to predict the results of bundle tests not used in the data-base for the model.

TABLE 1

Soiled Swatches Used in Terg-O-Tometer Studies

| Fabric | Soil type | Source[a] |
|---|---|---|
| Polyester/cotton (65/35) | Piscataway clay | Laboratory prepared |
| Polyester single knit | Sebum—particulate | Laboratory prepared |
| Nylon | Mineral oil + carbon black | Testfabrics, Inc.[b] |
| Cotton | Mineral oil + carbon black | Testfabrics, Inc. |
| Cotton (EMPA 101) | Oil and particulate | Testfabrics, Inc. |
| Cotton | Piscataway clay | Laboratory prepared |
| Polyester/cotton (50/50) | Piscataway clay | Laboratory prepared |

[a]Laboratory prepared swatches are described in the Experimental Details section.
[b]Testfabrics, Inc., Middlesex, N.J.

**Experimental details**

Our model is based on the results of 137 bundle tests run over a period of five years. The procedure outlined is the basic procedure used for all of the bundle tests involved in our study, with occasional minor variations which would not affect the outcome of the test.

Instrumentally matched items (using a Gardner XL-865 reflectance colorimeter) were provided to a Colgate family who used the items in their usual fashion for one week. The items were then returned to the laboratory for weekly laundering under controlled conditions. Prior to laundering, the matched items were separated into two parts, A and B, according to a predesigned code. The two groups of items were washed separately under identical conditions of water temperature (120°F) and water hardness (150 ppm) in matched washing machines of the same manufacturer and model using the detergent products under evaluation.

Included in each wash cycle were a group of matched clean swatches which were instrumentally evaluated for redeposition and brightener effects after five cycles. Items in Group A were always washed with detergent A and items in Group B with detergent B. If a commercial product or products were being evaluated, the manufacturer's recommended dose was used.

A total of 17 gallons of water were used for the launderings. After washing and drying, the laundered items were recombined and the bundle was ready for another cycle of use. After one prewash and five cycles of use and laundering, the items were evaluated instrumentally and visually.

A number of soiled swatches representing typical soil/fabric combinations were run in parallel Terg-O-Tometer studies with the products being studied in the bundle tests (Table 1). Reflectance measurements on the soiled swatches before and after six cycles of washing in the Terg-O-Tometer were obtained. These values along with values obtained from the cotton percale clean swatches were used in the regression analysis as the explanatory (independent) variables.

**Visual examination**

Evaluations were conducted under controlled lighting conditions with 10 panelists obtained from the Colgate R&D staff. Panelists were prescreened for visual impairment. The items were examined under a Macbeth lighting system both under incandescent light and simulated north daylight. The judges examined the bundle, one pair of items at a time, and indicated a preference for one or the other item in each set. No preference votes were also recorded and these votes were split equally between the products for the final analysis. The data from the 10-member visual panel were analyzed statistically for significance of the resulting preference ratio after the votes were tallied. The final preference ratio for white items only was used in our regression analysis as the response (dependent) variable. A total of 300 votes was obtained for each bundle test (15 white items evaluated under two lighting conditions by 10 panelists).

**Instrumental evaluation**

All bundle test items were measured for reflectance using a Gardner XL-865 large-area reflectance colorimeter. This instrument was designed for the Colgate-Palmolive Co. by the Gardner/Neotec Instrument Divison. It has a 6.75 inch in diameter aperture suitable for instrumental measurements of garments or household items such as towels. The reflectometer is interfaced to a TRS-80 Model II computer. A special software program was designed for matching the items used in the bundle test. The differences allowed between

**TABLE 2**

**Parameter Estimates**

| | Model | | | | |
|---|---|---|---|---|---|
| Swatch | All observations | Nonphosphate products | Phosphate products | Anionic products | Nonionic products |
| Bluing | 7.2 ± 0.6 | 5.4 ± 1.0 | 8.8 ± 1.0 | 8.1 ± 1.1 | 7.3 ± 0.8 |
| Brightening | −5.2 ± 0.7 | −3.0 ± 1.2 | −7.1 ± 1.0 | −6.9 ± 1.3 | −4.5 ± 0.8 |
| Sebum + particulate on dacron single knit | 1.0 ± 0.4 | 0.7 ± 0.5 | 0.9 ± 0.6 | 2.1 ± 0.5 | |
| Testfabrics nylon | 0.4 ± 0.4 | 0.4 ± 0.4 | | | |
| EMPA 101 | 0.4 ± 0.3 | | 1.4 ± 0.5 | 1.3 ± 0.4 | |
| Clay on polyester/ cotton (65/35) | 1.8 ± 0.9 | 1.3 ± 0.9 | 6.4 ± 4.3 | | 3.9 ± 1.1 |
| Testfabric cotton | | | | −1.6 ± 0.8 | |
| Clay on cotton | | | | | 5.8 ± 2.5 |
| Clay on 50/50 polyester/cotton | | | | | −2.6 ± 1.4 |

**TABLE 3**

**Terminology and Notation for Regression**

**Dependent (response) variable**

$Y^a$ = Total visual preference votes for white items washed in Product A + 1/2 of the no preference votes

$Y^b$ = Same as above for Product B

**Independent (explanatory) variables**

Stained swatches

$X_1^a$ = Final Rd value for soiled swatch 1 washed in Product A

$X_1^b$ = Same as above for Product B

Bluing swatches

$X_1^a$ = $\Delta$ b-1 for cotton percale swatch washed in Product A (Note: One is subtracted from $\Delta$ b to keep all be values less than zero)

$X_1^b$ = Same as above for Product B

Brightener swatches

$X_1^a$ = $\Delta$ Rb for cotton percale swatch

$X_1^b$ = Same as above for Product B

**Parameters**

$B_1, \ldots, B_n$ = parameters for each swatch, depending on its number.

matched pairs of items are: Rd $\leqslant$ 0.5 units; a $\leqslant$ 0.3 units; b $\leqslant$ 0.3 units; Rb $\leqslant$ 5 units. Rd is the whiteness scale and is related to L by the equation: $L = 10 \sqrt{Rd}$. The a and b values are the values for the red-green and blue-yellow components, respectively. Rb values are a measure of the fluorescent effects of brightener.

Reflectance values for soiled swatches were obtained on a Gardner XL-805 reflectance colorimeter.

### Determination of visual preference ratio

This procedure is described in three parts. Part 1 describes how bluing (b) and brightening (Rb) values are obtained from clean swatches, Part 2 is a description of how whiteness values (Rd) are obtained from soiled swatches and Part 3 details the preparation of soiled swatches not commercially available.

The equipment used includes:

- Reflectance colorimeter—capable of obtaining fluorescence reflectance.
- One pair of washing machines of the same manufacturer and model number.
- One pair of dryers of the same manufacturer and model number.
- Two instrumentally matched sets of (14″ × 15″) cotton percale swatches (seven swatches for each product).
- One set of stained swatches (3″ × 5″) chosen from the following: Piscataway clay (or any local clay that is uniform and abundant), on 65/35 polyester/cotton; Spangler sebum and particulate, on polyester single knot; Testfabrics Inc. soils, on Testfabric nylon, Testfabric cotton and EMPA 101 from Testfabrics Inc., Middlesex, New Jersey.
- New white items (for ballast): terry hand towels, cotton t-shirts, easy care t-shirts, clean swatches.

### Part 1

Match the seven pairs of cotton percale swatches instrumentally with the reflectance colorimeter. We consider swatches matched when: RD $\leqslant$ 0.5, a $\leqslant$ 0.3, b $\leqslant$ 0.3 and Rb $\leqslant$ 5.0. Label each set of swatches with product code.

Add each set of percale swatches to a ballast load (simulating a bundle test load, approximately 4.5 pounds) consisting of the following clean white items: three terry hand towels, three cotton t-shirts, three easy care t-shirts and 3″ × 5″ clean swatches consisting of three nylon, three polyester doubleknit, three polyester/cotton (65/35) and six polyester single-knit swatches.

Wash each load with the appropriate product using matched pairs of washing machines. Wash under the conditions of temperature, hardness and concentration that would be used if the products were being bundle-tested. Dryer dry all items for each load in separate, matched dryers.

The wash/dry cycle is repeated five additional times. The fluorescence reflectance values (Rb) and the blueness (b) values for the clean swatches are obtained after the sixth wash/dry cycle.

**Part 2**

Depending on product builder type, prepare and label replicate soiled and clean swatches in the following manner:

- Phosphate-built products—five EMPA 101, five sebum/particulate on polyester single knit and five Piscataway clay on polyester/cotton (65/35), as well as two 14″ × 15″ clean swatches (1 cotton percale and 1 banlon).
- Nonphosphate products—five sebum/particulate on polyester single knit, five Piscataway clay on polyester/cotton (65/35), five Testfabrics nylon and clean swatches as described for phosphate-built products.
- For those infrequent comparisons of nonphosphate versus phosphate products, use the soiled swatches listed in Table 2 under the heading "All observations" and clean swatches as above, then proceed with the next step. Wash the swatches and ballast load in the appropriate products as described in steps 2-4 in Part 1. Obtain average Rd reflectance values for each set of the soiled swatches. Calculate VPR using the appropriate Rd reflectance values for soiled swatches and Rb and b values for cotton percale swatches (from Part 1) substituted for the X terms in the following equation:

$$Y^a/Y^b = e\,[B_1 \log (X^a_1/X^b_1) + \ldots + B_n \log (X^a_n/X^b_n)] \quad [1]$$

**Part 3**

To prepare clay swatches, local clay is sieved through a #200 sieve. Then, 200 g of clay are mixed with 800 ml of deionized water in a blender at high speed. The dispersion can be applied to the cloth by padding or with a doctor blade. One liter of the clay dispersion is enough to coat 16 ft$^2$ of Testfabrics Inc. (Style 7435) cotton.

For the sebum/particulate mixture, 100 g of Spangler sebum (6) is mixed with 4 g of particulate mixture and 1000 ml of hot water (120°F) in a blender at high speed. The mixture is filtered warm through glass wool and applied warm to the fabric. This mixture is enough to coat 16 ft$^2$ of style 730 polyester single knit fabric obtained from Testfabrics Inc.

The particulates consist of a uniform mixture of 86 g kaolin clay, 8 g carbon black, 4 g black iron oxide and 2 g yellow iron oxide.

**TABLE 4**

**Coefficient of Determination for Regressions**

| Model | Number of bundle tests in data base (N) | Coefficient of determination ($R^2$) |
|---|---|---|
| All observations | 137 | 0.64 |
| Nonphosphate products | 73 | 0.72 |
| Phosphate products | 38 | 0.68 |
| Anionic surfactant products | 66 | 0.63 |
| Nonionic surfactant products | 47 | 0.73 |

**Visual preference ratio**

Regression is used to study the relationship between a response, such as the voter preference in a bundle test, and a number of explanatory variables, such as soil removal, bluing and brightening. The ultimate objective is to generate a regression equation such that the voter preference predicted by the explanatory variables will be close to the observed preference. Given the values of the explanatory variables, we can use this equation to predict the results of future bundle tests. By a stepwise regression, we can choose the explanatory variables that are important in predicting the response and discard those variables that are unimportant or redundant.

In our regression, the observed bundle test vote ratio, splitting tie votes between the two products, is used as the response variable. The explanatory variables are functions of the instrumental readings of the soiled swatches (Table 1) and the cotton percale swatch used to obtain bluing and brightener values. The terminology and notation for the regression are shown in Table 3.

For our regression, we chose a logarithmic model to place greater emphasis on changes in the vote ratio near the 1/1 region. Our experience shows it is far more common to obtain vote ratios near 1 to 1 or breakeven than to have huge differences such as 8 or 9 to one. The change from a 2/1 to a 3/1 vote ratio is much more dramatic in the logarithmic model than a change from 8/1 to 9/1. Linear models do not have this property (7).

A stepwise regression using weighted least squares was used to select the swatches that best predicted the log (bundle test preference ratio). A weight of 1/(1/A wins + 1/B wins), which is 1 over the variance of the log ratio, was used in our calculations.

In addition to a regression utilizing the data for all 137 bundle tests in our data-base, we decided to run regressions on various subsets of the data-base. Our objective was to determine if by separating the bundle tests into the types of product being tested, we would find better correlations with the observed preference.

Regressions, therefore, were run based on bundle tests of different builder type such as phosphate-built products, and nonphosphate products and surfactant type such as nonionic and anionic surfactants.

The results are shown in Table 4, which shows the coefficient of determination for the regressions, and Table 2, which gives the parameter estimates and their standard errors.

The coefficient of determination ($R^2$) takes on values between 0 and 1. The higher the $R^2$, the better the regression explains the response variable. In our case, $R^2$ varied from 0.63 to 0.73 (Table 4). These are reasonably high values considering the variability inherent in the bundle test. In our model, the variability in VPR is due to differences in the products used. In the regression with all 137 bundle tests, 64% (i.e., $R^2 = 0.64$) of this variation is explained by the differences observed in the soiled, bluing and brightener swatches. The remaining 36% is accountable to noise and due to variations in experimental conditions and panelist preferences.

Based on the small spread in the $R^2$ values obtained, the model chosen to calculate the VPR can be based either on the builder system or the surfactant type of the products being evaluated.

The parameter estimates shown in Table 2 are the coefficients of the regression. The log ratio of the reflectance values for product A to product B for each component (i.e., soiled swatches and bluing and brightening swatch) in the regression is multiplied by the appropriate coefficient and is summed. The anti-logarithm of this sum is the predicted VPR.

It is interesting to note that the coefficient for the brightener term is negative. This would suggest that an increase in brightening would have a negative effect on voter preference. This, however, is not the case. In practice, enhanced brightener levels increase both the fluorescence value (Rb) and the bluing component (b) of the reflected light. Increasing these values for a product should, within limits, act to increase voter preference for the product (8). A balance in the contributions of Rb and b to enhanced voter preference is reflected in the negative coefficient of the Rb term. The contribution of the improvement in b values more than compensates for the negative effect on the equation of the increase in Rb. The net effect is an increase in visual preference for the product with the greater brightener value.

The parameter estimates and the X′X matrix were used in a Basic program for prediction and prediction intervals (9). From this, we were able to develop the upper and lower prediction intervals of the VPR as well as the predicted value. In our case, there are 300 votes in a bundle test and an observed ratio of 1.3 is needed to assure significance (10). Thus, a VPR is considered a significant win (preference) for product A if the lower prediction interval (and by necessity, the VPR) exceeds a value of 1.3.

TABLE 5

**Bundle Tests Used for Verification of the Model**

| Bundle test | Surfactant type[a] | Builder type[a] | Model used for VPR |
|---|---|---|---|
| A | Nonionic | Nonphosphate | Nonphosphate |
| B | Anionic | Phosphate | Phosphate |
| C | Mixed | Phosphate | Phosphate |
| D | Mixed | Phosphate | Phosphate |
| E | Anionic | Nonphosphate | Nonphosphate |
| F | A = Nonionic<br>B = Mixed | Nonphosphate | Nonphosphate |
| G | A = Nonionic<br>B = Mixed | Nonphosphate | Nonphosphate |
| H | Nonionic | A = Phosphate<br>B = Nonphosphate | All data |
| I | Nonionic | A = Nonphosphate<br>B = Mixed | All data |
| J | Nonionic | Phosphate | Phosphate |
| K | Mixed | Phosphate | Phosphate |

[a]Applies to both Products A and B unless otherwise specified.

TABLE 6

**Comparison of VPR With Actual Vote Preference**

| Bundle test | Actual visual preference ratio | VPR Predicted value | VPR Limits Upper | VPR Limits Lower |
|---|---|---|---|---|
| A | −1.0 | 1.1 | 2.2 | −1.9 |
| B | 1.7 | −1.0 | −2.3 | 2.1 |
| C | 2.0 | 1.2 | 2.9 | −1.9 |
| D | 3.5 | 4.4 | 13.2 | 1.5 |
| E | 1.3 | 1.3 | 2.6 | −1.6 |
| F | −6.2 | −12.0 | −58.8 | −2.7 |
| G | −11.3 | −11.1 | −45.5 | −2.7 |
| H | −1.2 | 1.1 | 2.3 | −2.1 |
| I | 1.8 | −1.0 | −2.3 | +2.1 |
| J | 1.4 | 1.5 | 3.2 | −1.5 |
| K | 5.9 | 1.4 | 3.1 | −1.5 |

**Testing the VPR model**

In order to determine how closely our model predicts the results of visual preference voting, we checked our model in a series of 11 bundle tests that were not part of the data-base used to generate our equations.

Table 5 shows the types of detergent systems used for our verification and the equations used to obtain the VPR. As can be seen from this table, the equations chosen are those based on builder type rather than surfactant system. Either approach is valid and our choice was based on the fact that builder types are usually either all phosphate or all nonphosphate whereas the more frequently encountered mixed surfactant systems can lead to confusion in the choice of model to use.

The actual voter preference ratio for the 11 bundle tests as well as the predicted VPR and the upper and lower limits for the VPR are summarized in Table 6 and shown graphically in Figure 1. A number of observations can be made based on the data presented in Table 6:

- There are eight bundle tests where the absolute value of the observed visual preference ratio exceeds 1.3 and therefore a significant preference for one of the products exists.
- VPR correctly predicts five of the eight significant preference ratios (D,F,G,J and K). Four of the eight are predicted correctly in direction and extent of preference (i.e., moderate, strong) whereas in one case—bundle test K—VPR predicts a moderately significant preference for product A (VPR = 1.4) while a strong preference (observed vote ratio = 5.9) is actually observed. Note: Preference ratios >1.3 to 2.3 are moderate; those >2.3 are considered strong.
- Three of the eight significant bundle tests (B, C and I) were predicted to be breakevens by VPR, whereas moderate preferences were the case.
- VPR did not predict significance where none occurred nor did VPR predict a significant preference for the wrong product.

In all cases but one (K), the actual visual preference ratio fell between the upper and lower limits of the VPR. The reasons for the rather atypical behavior displayed by bundle test K are not yet known, but may be due to normal statistical variation.

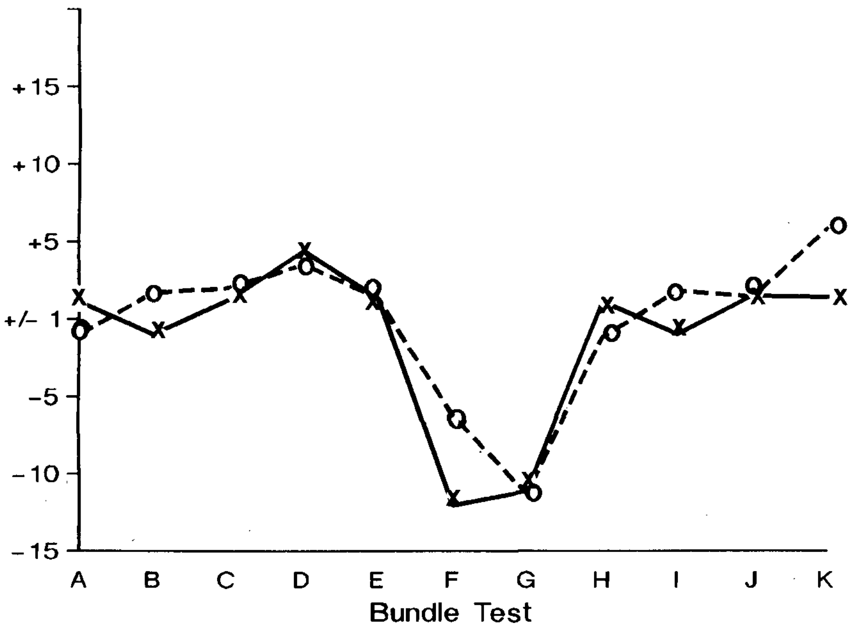


**FIG. 1. Observed and predicted visual preference ratios (VPR). O- - -O, observed; X—— X, predicted.**

- The reliability of VPR is somewhat lower for bundle tests where the observed vote preference shows a slight or marginal preference for one of the products. Therefore, a prediction of breakeven by VPR does not necessarily exclude a slight preference for one of the products.

The objective of our work was to provide a tool that will quantitatively predict the outcome (i.e., "winner") of a bundle test by relatively rapid laboratory methods. The VPR provides such a tool in that it has been demonstrated to be a reliable predictor of the extent and direction of bundle test visual preference.

## ACKNOWLEDGMENTS

The authors express their gratitude to Janet Leonardis and William Gross for their many invaluable contributions to this work.

## REFERENCES

1. Blake, J.R., in *Detergents In Depth*, The Soap and Detergent Association, 1975, p. 32.
2. Krussmann, H., *J. Am. Oil Chem. Soc. 55*:175 (1978).
3. Neiditch, O.W., *Soap, Cosmet. Chem. Spec.*, January 1975, p.44.
4. Pacheco, J.F., et. al., *Ibid.*, June 1974, p. 29.
5. *Annual Book of ASTM Standards*, Section 15, 1987, p. 588, ASTM D2960-84, Controlled Laundering Test Using Naturally Soiled Fabric and Household Appliances.
6. Spangler, W.G., H.D. Cross III and B.R. Schaafsma, *J. Am. Oil Chem. Soc. 42*:723 (1965).
7. Haberman, S.J., *Analysis of Quantitative Data*, Academic Press Inc., New York, 1978, p. 292.
8. Hunter, R.T., *J. Am. Oil Chem. Soc. 45*:362 (1968).
9. Draper et. al., in *Applied Regression Analysis, 2nd Edition*, John Wiley and Sons, New York, 1981, p. 80.
10. Haberman, S.J., *Ibid.*, p. 292.