DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 20(13), 2165-2174 (1994)



# MULTIVARIATE ANALYSES OF A PRODUCTION FORMULATION OPTIMIZATION EXPERIMENT

N. R. Bohidar
Philadelphia College of Pharmacy and Science
Villanova University
1530 Bridal Path Road. Lansdale, PA. 19446
and
Norman R. Bohidar
University of Washington, Seattle, Washington

## ABSTRACT

Three prominent multivariate statistical analyses, canonical correlation analysis (CCA), principal component analysis (PCA) and CAS-Regression analysis (CAS-R) are appropriately applied to the formulation optimization data associated with Product-T for determining a set of key excipient/process variables and a set of key response variables to be used in monitoring the future performance of the optimizated formula. CCA which considers both sets of variables simultaneously in a single analysis, successfully delineated two key parameters, one for each set. PCA which considers only the response variables concurred with the CCA results and CAS-R which considers each response variable separately also concurred. Even though CCA is a predominant technique, adjunct results of PCA and CAS-R could be supplemented for a comprehensive interpretation. It is recommended that all three analyses be carried out and interpreted appropriately.

## INTRODUCTION

Presently, formulation optimization experiments are conducted routinely by the research division of many pharmaceutical companies. Invariably, a well-defined statistical design, such as fractional factorial central composite design (1,2), is used to successfully explore the experimental region with a limited number of experimental points. The results of such experiments are subjected to a comprehensive optimization analysis, such as, M-SOOP-GRID-SEARCH procedure (1,2) with SAS-PROC-RSREG (3,4), to determine the optimum formulation containing the optimum levels of the excipient/process factors (X) which optimize simultaneously a selected set of several response variables (Y) considered in the experiment for a targeted drug product. Having obtained the final optimization results and a considerable small-scale processing experience, the technology transfer process between the research division and the production division is initiated. Predominant in this process is the establishment of a monitoring system (1,2) consisting of one or two key excipient/process factors and of one or two key product response variables for monitoring and controlling the future performance of the optimum formulation under the production environment in a most cost-and-time effective manner. The identification of these key variables, process (X) as well as product (Y), can effectively be accomplished by a multivariate statistical analysis known as, canonical correlation analysis (CCA) (5) which indeed has the capability of analyzing simultaneously all the formulation data, comprised of a set of several X-variables and a set of several Y-variables associated with a set of n formulations. Another multivariate statistical method known as, principal component analysis (PCA), (6) primarily applied to a set of Y-variables, accomplishes the task of determining the key response variables. A conditional multivariate analysis known as CAS-Regression (CAS-R) (implying, combination of All-Possible-Regression and Stepwise-Regression) (7) has the capability of detecting the key X-variables for each Y-variable <u>separately</u>. It should be noted here that PCA and CAS-R do not consider all the formulation data simultaneously. It should also be noted here that the selection of key control variables for the monitoring system does not any way imply that the other variables considered in the experiment are not important. They all play an <u>important</u> role. Because of the premise on which the M-SOOP-GRID-SEARCH procedure (1) has been established, the contribution of <u>all</u> variables, response as well as process, associated with the system become essential for accomplishing the final optimization solution.

The primary purpose of this paper is to apply all three multivariate statistical analyses mentioned above to the same formulation optimization data (Product-T data (1,2)), for the purpose of comparison and confirmation of the results of the three analyses, and for determining the key parameters of the system.

## DESCRIPTION OF PRODUCT-T FORMULATION DATA

Product-T (1,2) a pioneer product, has been in the market for a long time and it is now decided to streamline the dissolution profile of the dosage form (tablet) based on the same dissolution specifications as those currently applied to the recent products, by changing from an acid-based dissolution to a water-based one, basket to paddle, 150 rpm to 50 rpm and 70% in 45 min. to 90% in 30 min. An optimization experiment is conducted within the confines of the compendial, regulatory and in-house constraint limits originally established. The experimental design is a five-factor half-fractional factorial, orthogonal, central, composite second order design with a center point, yielding 27 formulations [$1/2(2^5) + (2 \times 5) + 1$]. The five excipient variables with their respective ranges are as follows: $X_1$ = CAB - 0 - Sil(mg.)(0.001 - 2.48), $X_2$ = Encompress/Lactose ration(mg./mg.)(0.98 - 6.68), $X_3$ = Starch disintegrant(mg.)(0.001 - 12.88), $X_4$ = Stearic acid(mg.)(0.001 - 6.80) and $X_5$ = Magnesium stearate(mg.) (0.35 - 1.75). The eight product variables (Y) with their respective ranges (experimentally observed) across the 27 formulations are, as follows: $Y_1$ = Content uniformity (%) (93.5 - 102.7), $Y_2$ = Hardness(tablet breaking strength) (Kg.)(3.69-5.73), $Y_3$ = Dissolution (15 min.)(%)(44 - 86), $Y_4$ = Dissolution (30 min.)(%)(78 - 91), $Y_5$ = Dissolution (45 min.)(%)(83 - 97), $Y_6$ = Content uniformity standard deviation(%)(0.63-5.60), $Y_7$ = Disintegration (min.)(3.3 - 13.3) and $Y_8$ = Weight uniformity (%)(1.53 - 5.85). Based on the 5 X-variables, 8 Y-variables and 27 formulations, CCA, PCA and CAS-R analyses are accomplished.

## THEORY

### Canonical Correlation Analysis (CCA)(5):

A comprehensive depiction of the theory has been presented in detail in reference (5). Only a succinct description is considered here. Let the linear function of the p X-variables be $\Sigma a_i X_i = A'X = W$ and of the t Y-variables be $\Sigma b_i Y_i = B'Y = Z$ ($p \leq t$). The variances of $A'X$ and $B'Y$ are $A'S_{xx}A$ and $B'S_{yy}B$ respectively and the covariance is $A'S_{xy}B$ where, $S_{xx}$, $S_{yy}$ and $S_{xy}$ are the variance-covariance matrices of X's, Y's and (X,Y)'s. The correlation between the two linear functions is

$$R_{wz} = [A'S_{xy}B]/[A'S_{xx}A]^{1/2}[B'S_{yy}B]^{1/2}.$$

For CCA one needs to determine those values of $a_i$'s and $b_i$'s which maximize $R_{wz}$, under the constraint $A'S_{xx}A = 1$ and $B'S_{yy}B = 1$. The Lagrange multiplier function has the following structure,

$$L = A'S_{xy}B - 1/2\theta_1(A'S_{xx}A - 1) - 1/2\theta_2(B'S_{yy}B - 1)$$

By taking the partial derivatives of L with respect to A, B, $\theta_1$ and $\theta_2$, and setting the derivative to zero, one obtains,

$$(S_{xy}S_{yy}^{-1}S_{yx} - \theta^2 S_{xx})A = 0$$

whose solutions are the eigen values and eigen vectors of the determinantal equation $\det[S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx} - \theta^2 I] = 0$, denoted by $\theta_1$, $\theta_2$, -- $\theta_p$ and $A_1$, $A_2$, ---$A_p$ respectively. The positive square root of the first eigen value, $\theta_1$ constitutes the maximum correlation between the two sets, and the elements of the first eigen vector $A_1$ provide the canonical coefficients, whose magnitudes determine the relative weights of the X-variables. The solution of the following equation provides the same set of eigen values $\theta_1$, $\theta_2$, ---$\theta_p$ and but a different set of eigen vectors, $B_1$, $B_2$ ---$B_p$,

$$(S_{yx}S_{xx}^{-1}S_{xy} - \theta^2 S_{yy})B = 0$$

and the elements of $B_1$, the canonical coefficients, provide the relative weights for the Y-variables. If one wishes to use the correlation matrices, $R_{xx}$, $R_{yy}$ and $R_{xy}$, then,

$$R_{wz} = C'R_{xy}D/[C'R_{xx}C]^{1/2}[D'R_{yy}D]^{1/2}$$

where, C and D provide the standardized canonical coefficients, however the p eigen values as before remain invariant. The correlation matrices are used for the analysis of this set of data.

<u>Principal Component Analysis (PCA-COV) (1,6)</u>:

A detailed description of the analysis has been presented in reference (6). Only a brief introduction will be provided here. PCA-COV considers only one set of variables at a time (generally the dependent variables) unlike CCA which considers two sets of variables simultaneously. Let $B'Y$ be the linear function of the t Y-variables and let the variance of $B'Y$ be denoted by $B'S_{yy}B$, where $S_{yy}$ is the variance-covariance matrix of the Y-variables. For PCA-COV, one needs to determine those values of B-vector which maximize $B'S_{yy}B$ under the constraint $B'B = 1$. The Lagrange multiplier function here has the following structure,

$$L = B'S_{yy}B - \theta(B'B - 1).$$

By taking the partial derivatives of L with respect to B and $\theta_1$ one obtains, by setting it to zero,

$$(B'S_{yy}B - \theta)B = 0,$$

whose solutions are the eigen values and eigen vectors of the determinantal equation $\det[B'S_{yy}B - \theta I] = 0$, denoted by $\theta_1$, $\theta_2$, -- $\theta_t$ and $B_1$, $B_2$, ---$B_t$ respectively. The first eigen value, $\theta_1$, constitutes the maximum variance of the linear function $B'Y$ and $B_1$ provides the principal coefficients of the function, whose relative magnitudes determine the relative weights of the Y-variables.

<u>Principal Component Analysis (PCA-CORR)(8)</u>:

In this analysis one uses the correlation matrix $R_{yy}$ instead of the variance-covariance matrix, $S_{yy}$. The steps of the derivation are essentially identical, and the eigen values and eigen vectors derived from the equation,

$$(G'R_{yy}G - \theta{*}I)G = 0$$

are <u>not</u> invariant. $\theta_1^*$ provides the maximum variance, and eigen vector G provides the standardized weights associated with the Y-variables. PCA-CORR is generally used as an exploratory analysis of the data since the standardized coefficients sometimes reveal latent structures among the variables. The purpose and interpretations of the two analyses, PCA-COV and PCA-CORR, are drastically different.

CAS-Regression Analysis (CAS-R)(1,7):

All Possible Regression (APR)(7): Consider a regression of K process variables (X) on a single response variable (Y) denoted by the model, $Y = XB + E$ where matrix X, vector Y and vector E have (nx(k + 1)),(n x 1) and (n x 1) dimensions respectively. The Gauss-Markoff Least Squares procedure yields, the estimates of the regression coefficients $B^* = (X'X)^{-1}X'Y$, regression sum of squares = $[B^{*\prime}X'Y - R(b_o)]$ and total sum of squares = $[Y'Y - R(b_o)]$, where, $R(b_o) = nY*^2$ ($Y*$ = mean of Y's). Now $R^2 = [B^{*\prime}X'Y - R(b_o)]/[Y'Y - R(b_o)]$. APR provides and examines $(2^K - 1)$ $R^2$-values generated by considering one X-variable, two X-variables,--- and k X-variables at a time. The smallest set of X-variables which attains the highest possible $R^2$-value is the set with the most important X-variables.

Stepwise Regression (SWR)(1,7):

SWR selects that X-variable (say $X_4$) which has the highest correlation with the Y-variable and then conducts a F-test using

$$F = [n-k-1][B^{*\prime}X'Y - R(b_o)]/[Y'Y - B^{*\prime}X'Y][k].$$

If the test is significant, it proceeds to select next that X-variable (say $X_2$) which has the highest partial correlation (conditioned on $X_4$) and then proceeds to test using a sequential F-test, where for example, the partial correlation,

$$R_{2y.4} = [R_{2y} - R_{24}R_{4y}]/[(1 - R^2_{24})(1 - R^2_{4y})]^{1/2}$$

The method progresses in this stepwise manner by using correlation, partial correlation, F-test and sequential F test at each step to arrive at the set of important X-variables.

## RESULTS, DISCUSSION AND INTERPRETATION

Canonical Correlation Analysis (CCA)(5):

A canonical correlation analysis is performed for the Product-T formulation optimization experiment (1,2,5). Simultaneous consideration of all thirteen variables (5X-and 8Y-variables) is the most attractive feature of this procedure enabling one to draw appropriate multivariate simultaneous statistical inferences.

The first canonical correlation is the maximum correlation between the variables of the two sets. In this case, the magnitudes of the eigen values are: $\theta_1^2$ = 0.9162, $\theta_2^2$ = 0.5910, $\theta_3^2$ = 0.4466, $\theta_4^2$ = 0.1923 and $\theta_5^2$ = 0.0115 and their respective canonical correlations are: $R_{c1} = 0.9572 = (0.9162)^{1/2}$, $R_{c2} = 0.7688$, $R_{c3} = 0.6683$, $R_{c4}$ =

0.4386 and $R_{c5}$ = 0.1071. Here the maximum correlation ($R_{c1}$ = 0.9572) is indeed the maximum since the highest bivariate correlation between X and Y variables is only 0.8181 (see Table-I, intersection of $X_3$-row and $Y_3$-column). This confirms not only the theoretical foundation of the method but also the appropriateness of the multivariate analysis, CCA. It is found that $R_{c1}$ = 0.9572 is statistically highly significant based on the Wilk's Lambda test ($p$ = .0004), Hotelling-Lawley trace test ($p$ = .0001), Roy's greatest root test ($p$ = .0001) and Pillai's trace test ($p$ = .0189). Furthermore, the other four canonical correlations are statistically not significant ($p > 0.05$) with $p$ = .27, .63, .93 and .99, respectively based on Wilk's test. Based on the significant canonical correlation, $R_{c1}$, the two canonical functions have the following form, derived from the eigen value, $\theta_1^2 = 0.9162$:

$$W = -0.3404X_1 - 0.2809X_2 + 0.8970X_3 + 0.0380X_4 - 0.0478X_5 \text{ and}$$

$$Z = 0.0312Y_1 + 0.2496Y_2 + 1.0288Y_3 - 0.1855Y_4 + 0.0638Y_5 - 0.0822Y_6 - 0.1382Y_7 - 0.1475Y_8$$

where, X and Y must be expressed in their standardized forms and the numerical values represent their respective canonical coefficients. Note that the regular correlation between the two canonical variates W and Z, $R_{wz}$ = 0.9572.

The magnitude of the first eigen value, which determines the extent to which the above two functions account for the thirteen variables, is an impressive 91.6% ($\theta_1^2$ x 100 = 91.6%), indicating adequate representation of the thirteen variables by the two canonical functions and assuring high potentiality for accurate predictability. The regression equation of Z on W is Z = 0.9572W with an $R^2$-value of 0.9162, which is very high.

The next step is to determine the relative contribution of each variable, to rank order the variables and to delineate the most important variables in each set by examining the absolute magnitudes of the canonical coefficients associated with the variables. Consider first the X-set. The absolute values of the coefficients are, $C_1$ = 0.3404, $C_2$ = 0.2809, $C_3$ = 0.8970, $C_4$ = 0.0380 and $C_5$ = 0.0478 indicating that $X_3$ is the highest contributor of the set. Expressed as a percentage of the total absolute weight ($100C_i/\Sigma C_i$), one has, $X_1$ = 21.2%, $X_2$ = 17.5%, $X_3$ = 55.9%, $X_4$ = 2.4% and $X_5$ = 3.0%. This shows that starch disintegrant turns out to be the most important key excipient variable in this study. It should also be noted that the bivariate correlations between $X_3$ and the two canonical variates are very high, corr($X_3$,W) = 0.8924 and corr($X_3$,Z) = 0.8542. Table I shows that corr($X_3$,$Y_3$) = 0.8181 and corr($X_3$,$Y_4$) = 0.5075, where corr stands for correlation. Note that these bivariate correlations are all highly statistically significant (p

TABLE-I

BIVARIATE CORRELATION VALUES BETWEEN VARIABLES

| | $Y_1$ | $Y_2$ | $Y_3$ | $Y_4$ |
|---|---|---|---|---|
| $X_1$ | 0.3305 | 0.2393 | -0.3251 | -0.4271 |
| $X_2$ | -0.0588 | -0.0247 | -0.2252 | -0.3413 |
| $X_3$ | 0.0906 | 0.2059 | 0.8181 | 0.5075 |
| $X_4$ | -0.2451 | 0.0381 | 0.1190 | 0.1741 |
| $X_5$ | 0.1380 | 0.1086 | 0.0238 | -0.1826 |
| | $Y_5$ | $Y_6$ | $Y_7$ | $Y_8$ |
| $X_1$ | -0.4005 | 0.1257 | 0.4692 | 0.3008 |
| $X_2$ | -0.2646 | 0.1555 | 0.2131 | 0.2546 |
| $X_3$ | 0.3870 | 0.1383 | -0.3864 | 0.0973 |
| $X_4$ | 0.3232 | 0.1912 | 0.2540 | 0.1787 |
| $X_5$ | -0.1192 | 0.1889 | 0.2154 | 0.3330 |

$< 0.01$). Now consider the Y-set. The absolute magnitudes of the canonical coefficients are, $d_1 = 0.0312$, $d_2 = 0.2496$. $d_3 = 1.0288$, $d_4 = 0.1855$, $d_5 = 0.0638$, $d_6 = 0.0822$, $d_7 = 0.1382$ and $d_8 = 0.1475$. Clearly $Y_3$ stands out as the highest contributor to the set. Expressed as a percentage of the total absolute weight ($100d_i/\Sigma d_i$), one has $Y_1 = 1.6\%$, $Y_2 = 12.9$, $Y_3 = 53.4\%$. $Y_4 = 9.6\%$, $Y_5 = 3.3\%$, $Y_6 = 4.3\%$, $Y_7 = 7.2\%$ and $Y_8 = 7.7\%$, indicating clearly that $Y_3$ is sharing more than half of the total absolute weight. This shows that dissolution (15 min.) is the most important key response parameter in the study. It should also be noted that the bivariate correlations between $Y_3$ and the two canonical variates are very high, corr($Y_3$,Z) = 0.9519, and corr($Y_3$,W) = 0.9111. Table I shows that corr($Y_3$,$X_3$) = 0.8181. The bivariate correlations are all highly statistically significant ($p < 0.01$). It should be clearly recognized that starch disintegrant and dissolution (15 min.) have emerged as the two most important key parameters in this study. CCA has accomplished the explicit delineation of the two important key variables, one for each set.

Principal Component Analysis (PCA)(6):

As noted in the theory section, PCA has two distinct forms, (i) PCA-COV, which uses the variance-covariance matrix to extract the required eigen values and eigen vectors and (ii) PCA-CORR, which uses the bivariate correlation matrix to achieve the same. However, it is important to note that the objectives, results and interpretation of these two procedures are drastically different. The differences are elaborated in the following. Consider PCA-COV first. This procedure attempts to detect those response variables whose values significantly vary from formulation to formulation. The selected variable would be expected to produce a broad and significant change for a minor streamline change in the levels of the process/excipient variable. This quality is highly desirable in a monitoring system. The variation across formulations is the prime consideration here. For practical considerations, this variation must be expressed in the original response unit. Note that this is a special application of PCA-COV appropriate for formulation studies only. Now consider PCA-CORR. This procedure attempts to detect redundancy among response variables, by detecting the presence of very high correlations (0.90 or above) among the variables. If a set of variables are moderately correlated (say .20 to .80) PCA-CORR would generally consider most or all of them to be important. If they are orthogonal (zero or near zero correlation) it will declare that everyone is important. If they are all highly correlated, it will show that only one variable is sufficient. PCA-CORR also has the property of revealing some latent structures existing among groups of response variables (which may or may not be meaningful). Indeed this is the primary purpose for which PCA-CORR and multivariate "factor" analysis are used in other fields. In short, PCA-COV deals with characterization of formulation variation whereas, PCA-CORR deals with correlations, redundancy and latent structures(6,8).

The results of both the analyses, PCA-COV and PCA-CORR, are presented in the following, with the full understanding that, PCA-COV is a confirmatory analysis with direct applications and PCA-CORR is an exploratory analysis of the data.

PCA-COV(6): For the eight Y-variables involved in this analysis, PCA-COV yields the following eight eigen values, $\theta_1 = 148.5$, $\theta_2 = 15.8$, $\theta_3 = 7.5$, $\theta_4 = 3.9$, $\theta_5 = 1.3$, $\theta_6 = 0.58$, $\theta_7 = 0.47$ and $\theta_8 = 0.13$. It is noted that $\theta_1$ alone accounted for 83.4% of the total variation associated with the original variables. Therefore it is considered here to examine only the first principal component (first eigen vector). The magnitudes of the 8 elements are: $b_1 = -0.0157$, $b_2 = -0.0042$, $b_3 = 0.8885$, $b_4 = 0.3982$, $b_5 = 0.1811$, $b_6 = 0.0171$, $b_7 = -0.1364$ and $b_8 = 0.0068$.

Expressed as a percentage of the total absolute weight ($100b_i/\Sigma b_i$), one obtains, $Y_1$ = 0.96%, $Y_2$ = 0.26%, $Y_3$ = 53.91%. $Y_4$ = 24.16%, $Y_5$ = 10.99%, $Y_6$ = 1.04%, $Y_7$ = 8.27% and $Y_8$ = 0.41%, indicating clearly that $Y_3$ is the dominant variable with more than half of the total absolute weight. In other words, PCA-COV has determined that dissolution (15 min.) is the key response parameter, a result which fully concurrs with that of CCA.

PCA-CORR(6): Here the eight eigen values are: $\theta_1 = 3.47$, $\theta_2 = 2.5$, $\theta_3 = 0.81$, $\theta_4 = 0.63$, $\theta_5 = 0.26$, $\theta_6 = 0.21$, $\theta_7 = 0.08$ and $\theta_8 = 0.05$. It would take as many as three eigen values prior to reaching 84.6% of the total "correlation" (trace=8) of the system. The elements of only the first principal component is discussed here. The magnitudes are, $g_1 = 0.3339$, $g_2 = 0.2822$, $g_3 = -0.3307$, $g_4 = -0.4186$, $g_5 = -0.4234$, $g_6 = 0.2068$, $g_7 = 0.4626$ and $g_8 = 0.2971$. Expressed as a percentage of the total absolute weight ($100g_i/\Sigma g_i$), one obtains, $Y_1$ = 12.12%, $Y_2$ = 10.24%, $Y_3$ = 12.01%. $Y_4$ = 15.19%, $Y_5$ = 15.37%, $Y_6$ = 7.54%, $Y_7$ = 16.79% and $Y_8$ = 10.78%, showing that all eight variables are essentially equally weighted. (Note that out of 28 correlations, 12 values are below 0.30, 13 values are between 0.31-0.69 and only three values between 0.7 and 0.9). Here the highest weight is given to $Y_7$ because it is moderately correlated (0.38-0.58) with all other seven variables. The only feature that can be considered as a latent structure is that the principal component separates the variables into two groups, $Y_3$, $Y_4$ and $Y_5$ as one and the rest as the other, which can be meaningfully interpreted as a contrast between the dissolution-related variables and non-dissolution related variables (see the signs). Note that PCA-CORR results would be better interpretable if the response variables emanate from a single formulation (one treatment group) rather than from a group of several formulations, as in this case.

CAS-R (7): In this analysis each response variable has been analyzed separately. Only the salient features of the results are presented. For APR, any response variable with a $R^2$-value below 0.50 is not presented, and for SWR, the significance level is set at 0.01 or below for the F-tests of X-variables. For $Y_3$, APR shows a $R^2$-value of 0.84 with $X_1$, $X_2$, $X_3$ and $X_4$ in the regression model. However, SWR shows that the F-test p-values of $X_3$ and $X_1$ are 0.0001 and 0.0026. So the min-central subset (7) is indeed $X_3$ (starch disintegrant), which concurs with the result of CCA. For $Y_4$, APR shows a $R^2$-value of 0.59 with $X_1$, $X_2$, $X_3$ and $X_4$ in the model. However, SWR shows that the F-test p-values for $X_3$ and $X_1$ are 0.0069 and 0.0100. So the min-central subset is $X_3$ again. Note that $X_1$ does have a significant effect on $Y_3$ and $Y_4$, only when they are considered individually. In a multivariate set-up, however, this may not be the case because of simultaneous considerations of all variables. It should be noted that the statistical findings applies only to the system considered.

In conclusion, it should be emphasized that, simultaneous analysis of all variables with CCA should be the prime consideration. Adjunct results from the PCA and CAS-R analyses should provide necessary supplemental information. The results of all three analyses must be considered simultaneously for a comprehensive interpretation and for appropriate pharmaceutic decisions.



## ACKNOWLEDGEMENT

Deep gratitudes are due to Mrs. Barbara J. Tomlinson for her excellent talent in word-processing this manuscript with utmost rapidity and quality.

## REFERENCES

1. N. R. Bohidar and K. E. Peace, Pharmaceutical Formulation Development. Chapter IV. "Biopharmaceutical Statistics in Drug Development." Marcel Dekker, Inc. New York, N.Y. 149-229 (1988)
2. N. R. Bohidar, Application of Optimization Techniques in Pharmaceutical Formulation-An Overview. Proceedings of the American Statistical Association. Biopharmaceutical Section. 6-13 (1984).
3. N.R. Bohidar, Pharmaceutical Formulation Optimization Using SAS. Drug Development and Industrial Pharmacy, Vol.17, No.3, 421-441 (1991).
4. SAS Institute Inc. "SAS User's Guide: Statistics" Version 5 Edition. SAS Institute, Inc. Cary, NC(1985)
5. N. R. Bohidar and N. R. Bohidar, Canonical Correlation Analysis of Formulation Optimization Experiments. Drug development and Industrial Pharmacy. Submitted for Publication (1993).
6. N. R. Bohidar, F. A. Restaino and J. B. Schwartz, Selecting Key Parameters in Pharmaceutical Formulations by Principal Component Analysis. J. Pharm. Sci, Vol. 64, No. 6, 966-969 (1975).
7. N. R. Bohidar, F. A. Restaino and J. B. Schwartz, Selecting Key Pharmaceutical Formulation Factors by Regression Analysis. Drug Development and Industrial Pharmacy. Vol. 5, No. 2, 175-216 (1979).
8. T. W. Anderson, "An Introduction to Multivariate Statistical Analysis." John Wiley and Sons, New York, N.Y. (1958)