DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 17(7), 959-974 (1991)

KEY VARIABLES IN DOSAGE FORM DESIGN

T. Schofield, J. F. Bavitz, C. M. Lei, L. Oppenheimer, and P. K. Shiromani

Merck Sharp & Dohme Research Laboratories
West Point, PA 19486

ABSTRACT

The study described herein was undertaken to learn more about the relationships between tablet properties and formulation/process variables. Five independent variables were varied according to a central composite star design, and a number of properties have been evaluated. A discussion of various statistical procedures/issues have been examined for exploratory studies such as this one.

INTRODUCTION

The techniques of optimization in pharmaceutical dosage form are well documented[1,2,3,4]. Schwartz, et al.[5,6] developed a technique whereby a formulation with optimum properties could be obtained through computer assisted data analysis. Sequential Prediction Analysis[7] has been suggested for optimizing multiple potency systems coupled with invariant tablet weight.

In the optimization of a drug delivery system, several independent formulation and/or process variables are studied and

several tablet properties are measured. Since a large number of variables is generally involved, it becomes relevant to study the interrelation within these two groups of variables, and to determine the key variables which distinguish formulations. This enhancement of the classical optimization technique is characterized in Bohidar, et.al.[8,9].

In this paper, we elaborate on these dimension reduction techniques, and consider their application to process optimization experiments performed in these laboratories.

STATISTICAL METHODS

A. Examining the relationship between tablet properties.

Two interrelated statistical techniques for studying the relationship between dependent variables (here tablet properties) are (1) correlation analysis, and (2) principal component analysis. These are described herein.

A correlation matrix examines the relationship between all pairs of variables (tablet properties). The Pearson[10] correlation coefficient measures the "linear" relationship between two variables. An alternative measure which is less sensitive to aberrant results, and does not require that the underlying relationship be linear, is the Spearman correlation coefficient[10]. The Spearman correlation coefficient is calculated as is the Pearson, with the computations now performed on "ranks" rather than the actual values.

One examines the Spearman correlation matrix (i) to see if any pair of variables have correlations close to ±1 (indicating strong positive/negative association), or (ii) to see if any

variables have correlations close to zero (no association) for all other variables. If two or more variables have correlations close to ±1 then they may be assumed to be measuring similar tablet properties and are somewhat redundant. One would then tend to analyze only one of the several variables, thereby, reducing the multiplicity of conclusions obtained from the several. If one of the variables was uncorrelated with all other variables, this would indicate that it is providing information that was not readily available from any of the other variables, and thus would certainly be included in the set of tablet properties examined.

Prinicipal component analysis[11] is a technique by which we try to capture most of the correlation structure among several dependent variables into a parsimonious set of variables, called principal components. This is similar to correlation analysis, however principal component analysis allows one to examine all the variables simultaneously, and not just two at a time. The analysis first provides the linear combination of all the variables that describes the largest proportion of the correlation structure among all the dependent variables, called the first principal component. Subsequent principal components are linear combinations that account for the highest proportion of the remaining correlation structure. The first principal component does not imply redundancy, however. The last principal component does that.

Principal component analysis can be performed on the covariance matrix (and, thereby, variances) for a set of observations on several tablet properties[9], but we would not

recommend this when the scales of the several variables are inconsistent (e.g., granulation surface area is in 100's, disintegration time is in 1's). Consistent with correlation analysis, principal component analysis will be performed on the Spearman correlation matrix (and, thereby, the ranks of the actual data).

A second order response surface of the form

$$Y = A_0 + A_1X_1 + \ldots + A_5X_5 + A_{11}X_1^2 + \ldots + A_{55}X_5^2 + A_{12}X_1X_2 + \ldots + A_{45}X_4X_5$$

was fit to data for each response variable (tablet property). Both the "full model" containing 21 parameters and a "reduced model," resulting from backward elimination[13] were estimated.

For each fitted model, several summary statistics and diagnostic measures were computed in order to assist in comparing these models and to evaluate the adequacy of fit of each model: (1) an adjusted R^2 for goodness of fit, (2) the model F-value for adequacy of fit, and (3) Cook's D statistic for influential observations. These are described herein.

The coefficient of determination (R^2) measures the proportion of the total variation about the mean explained by the fitted model. The model F-value tests hypothesis that all regression parameters are zero. When the number of parameters to be estimated gets close to the total number of experiments, the value of R^2 can be misleading, however. One might conclude on the basis of R^2 alone that any "reduced model" is inferior to the "full model" simply because it has fewer parameters, and,

thereby, a lower R^2 "moderates" this algebraic relationship, by adjusting for the number of parameters in the fitted model.

Single observations that have a large impact on the resulting prediction equation are called "influential observations." A measure of influence, called Cook's D, can be calculated, and observations with large values of that measure can be tagged for close scrutiny. When a model has been fit and large values of Cook's D have been observed among the data, the model has been refit without the influential observations. The effects of elimination of those observations can then be appraised via other summary measures from the regression.

EXPERIMENTAL

The experimental design employed was a five-factor, orthogonal, central composite, second order design. The design has been employed many times in these laboratories and is described in the literature[5]. Independent formulation and process variables and dependent tablet properties are listed in Table I. Through combinations of the independent formulation and process variables, as shown in Table II, 27 formulations were manufactured. The translation of the statistical coding into physical units is given in Table III. Each formulation consisted of a batch of 6,000 tablets. Wet granulation was carried out in a high intensity granulator[1] at a low impeller (500 rpm) and chopper (1000 rpm) speed. Granulations were tabletted on a rotary press[2]. Dissolution rates were determined in a 900 ml, 0.1M hydrochloric acid with USP apparatus 2 at 50 rpm. Tablet disintegration time, breaking strength and friability were

TABLE I

Independent Variables

X1 - Quantity of Granulating Water, (i.e., water required for an equivalent starch paste)

X2 - Time for Granulation

X3 - Screen Size for Dry Grinding

X4 - Quantity of Magnesium Stearate

X5 - Magnitude of Compression Pressure

Dependent Variables

Y1 - Granulation Surface Area Expressed as cm^2/g

Y2 - Mean Dissolution Rate of 4 Tablets Expresses as % of Label Dissolved in 30 minutes

Y3 - Mean Disintegration Time of 6 Tablets expressed in Minutes, Without Discs

Y4 - Mean Hardness of 10 Tablets Expressed in Kp, using tensile strength modified jaws

measured with commonly employed equipment. Experimental results are presented in Table IV.

RESULTS AND DISCUSSION

A. Examining the relationship between tablet properties.

Table V contains the Spearman correlation matrix for the four most important tablet properties. We notice that three of the variables (Y1,Y2,Y3) seem to be intercorrelated, Y2 (disolution rate) being positively correlated with Y1 (granulation surface area) and Y3 (disintegration time) negatively correlated with Y1 and Y2. Furthermore, Y4 (hardness) is uncorrelated with any of the other tablet properties.

TABLE II

Experimental Design

Trial	Exp. No.	X1	X2	X3	X4	X5
1	23	-1	-1	-1	-1	+1
2	18	+1	-1	-1	-1	-1
3	16	-1	+1	-1	-1	-1
4	08	+1	+1	-1	-1	+1
5	20	-1	-1	+1	-1	-1
6	25	+1	-1	+1	-1	+1
7	06	-1	+1	+1	-1	+1
8	07	+1	+1	+1	-1	-1
9	05	-1	-1	-1	+1	-1
10	09	+1	-1	-1	+1	+1
11	26	-1	+1	-1	+1	+1
12	15	+1	+1	-1	+1	-1
13	19	-1	-1	+1	+1	+1
14	12	+1	-1	+1	+1	-1
15	27	-1	+1	+1	+1	-1
16	03	+1	+1	+1	+1	+1
17	14	-1.547	0	0	0	0
18	13	+1.547	0	0	0	0
19	24	0	-1.547	0	0	0
20	22	0	+1.547	0	0	0
21	04	0	0	-1.547	0	0
22	17	0	0	+1.547	0	0
23	01	0	0	0	-1.547	0
24	11	0	0	0	+1.547	0
25	10	0	0	0	0	-1.547
26	21	0	0	0	0	+1.547
27	02	0	0	0	0	0

This structure is born out in a principal component analysis of these same variables. Table VI contains the proportion of "explained correlation" for the first 3 principal components, as well as the coefficients of each linear combination. 57% of the total "correlation structure" can be explained by a linear combination of the form

$$PC_1 = 0.60*Y1 + 0.53*Y2 - 0.62*Y3 + 0.06*Y4$$

TABLE III

Independent Variables	-1.547 eu	-1 eu	0 eu	+1 eu	+1.547 eu
X1 = Quantity of Granulating Water (i.e. Water required for an equivalent starch paste); 1 eu = 0.5%	7.626%	7.90%	8.4%	8.9%	9.17%
X2 = Time for Granulation; i eu = 1 min	1.45 mins	2.0 mins	3.0 mins	4.0 mins	4.55 min
X3 = Screen Size for Dry Grinding; 1 eu = 0.015"	0.04" (No.1A)	0.048" (No.1B)	0.063" (No.2)	0.078" (No.2AA)	0.086" (No.2A=0.093"
X4 = Quantity of Magnesium Stearate: 1 eu = 0.55 mg	0.85 mg	1.15 mg	1.7 mg	2.25 mg	2.55 mg
X5 = Magnitude of Compression Pressure 1 eu = 0.5 tons	1.2 tons	1.5 tons	2 tons	2.5 tons	2.8 tons

This is consistent with the correlations observed between tablet properties. In fact, the second principal component seems to be associated with Y4 alone, and describes an additional 26% of the "correlation structure."

We will consider the utility of these results in subsequent discussion.

B. Examining the relationship between each tablet property and the process/formulation variables.

Results of the "full model" with all 21 parameters are presented in Table VII. The dependent variable under

TABLE IV

Experimental Results

Trial	Y1	Y2	Y3	Y4
1	931.57	96	6.38	13.9
2	1250.27	92	3.25	10.8
3	427.63	96	8.50	10.5
4	734.89	101	5.50	11.5
5	896.55	97	2.53	10.5
6	1205.09	98	3.13	9.22
7	312.46	92	15.72	11.8
8	622.85	98	4.25	9.0
9	884.22	100	3.13	9.9
10	1329.12	100	4.67	11.9
11	529.74	64	26.9	9.8
12	600.26	94	25.0	8.9
13	848.70	86	10.55	12.4
14	1319.13	100	2.17	9.2
15	384.77	62	121.04	8.5
16	569.86	66	28.22	10.6
17	329.26	90	13.22	9.3
18	1146.19	99	2.88	10.4
19	1059.36	98	2.83	10.3
20	453.30	98	18.16	10.1
21	599.85	98	10.47	10.1
22	615.03	98	9.92	10.1
23	645.06	94	4.70	12.0
24	671.01	90	15.3	9.7
25	678.03	98	3.08	9.2
26	520.76	93	14.13	12.2
27	570.85	100	10.33	10.1

TABLE V

Spearman Correlation Matrix

	Y1	Y2	Y3	Y4
Y1	1	0.5	-0.78	0.14
Y2		1.0	-0.66	-0.05
Y3			1.00	-0.03
Y4				1.00

TABLE VI

	Principal Components		
	1	2	3
Y1	0.60	0.14	-0.55
Y2	0.53	-0.21	0.78
Y3	-0.62	0.04	0.18
Y4	0.06	0.94	0.24
Proportion Explained	0.57	0.26	0.12
Cumulative Proportion	0.57	0.83	0.95

consideration is the disintegration time - Y4. Terms were sequentially deleted from the model when the coefficient P-value was greater than 0.25. Results of this "backward elimination" are also presented in Table VII.

Although R^2 is often used as a measure of model adequacy, it is not useful in this content, the reason being that R^2 cannot decrease by definition, as additional terms are added to the model. Thus, if R^2 is the sole criterion for model selection, there will be a tendency to add terms that should not be included in the model. The adjusted R^2 is a better measure of model utility since it takes into account not only how much variability the model explains but also how many variables were utilized in the process. We notice for results in Table VII that while R^2 for the "full model" (R^2=0.93) is greater than that for the "reduced model", (R^2=0.91), the adjusted R^2 for the "reduced

TABLE VII

Backward Elimination

	Full Model	Backward Elimination	Without Obs. 15
Intercept	-11.3 (.99)	-403 (.02)	45.5 (.00)
X1	-46.4 (.86)		
X2	64.8 (.32)	83.1 (.10)	-9.9 (.02)
X3	4933 (.28)	6578 (.05)	-442 (.00)
X4	88.4 (.45)	120.5 (.17)	-51.5 (.00)
X5	-98.0 (.46)		29.5 (.02)
X1X1	7.0 (.65)	5.1 (.02)	
X2X2	2.8 (.47)		0.7 (.21)
X3X3	11773 (.49)		
X4X4	8.5 (.51)		
X5X5	7.9 (.60)		
X1X2	-12.5 (.09)	-12.5 (.03)	
X1X3	-880 (.07)	-855 (.02)	
X1X4	-19.2 (.14)	-19.3 (.06)	4.1 (.02)
X1X5	20.6 (.14)	13.6 (.00)	-5.0 (.00)
X2X3	426.6 (.08)	426.6 (.03)	
X2X4	18.4 (.02)	18.4 (.00)	6.7 (.00)
X2X5	-12.0 (.10)	-12.6 (.03)	
X3X4	759.7 (.09)	759.7 (.03)	
X3X5	-633 (.17)	-690 (.06)	222.4 (.00)
X4X5	-21.2 (.11)	-22.3 (.03)	2.1 (.25)
Model F	4.20*	8.37**	44.00**
R2	0.93	0.91	0.97
Adjusted R2	0.71	0.8	0.95
Obs.(Cook's D)	15 (6.6)	15 (6.59)	
Actual		10.3	10.3
Predicted (CI)		12.3 (-9.7,35.6)	9.6 (5.5,13.7)

model" (adjusted R^2=0.80) dominates that for the "full model" (adjusted R^2=0.71).

A projection of the response surface estimated from "backward elimination" is depicted in Figure 1. Expected disintegration time (Y3) versus granulation time (X2) x screen size (X3) is

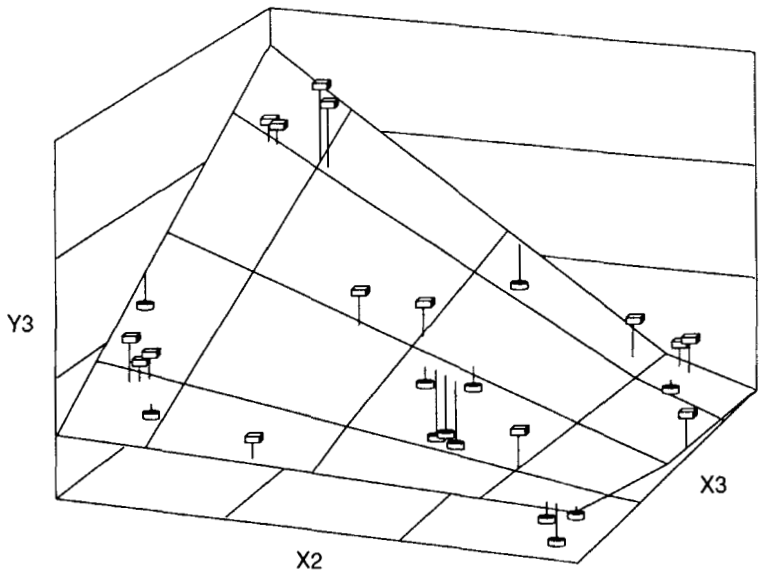

FIGURE 1

represented by the surface in the figure, while observed disintegration time for the 27 experimental points have been adjusted to that subspace. Experimental observations that have been underestimated by the model are depicted as boxes, while experimental observations that have been overestimated by the model are represented by cylinders. "Needles" have been used to connect the observations to the response surface. A longer needle represents a larger deviation (residual) between observed and expected disintegration time.

It is noted in the "backward elimination" that observation 15, (disintegration time of about 121 minutes), is associated with a large "Cook's D." This is a diagnostic for influential observations, and is usually indicative of a large residual.

Other experiment observations have, however, larger absolute residuals than observation 15 (see Figure 1). In addition to having a large (absolute) residual, observation 15 resides on an extreme of the design space. It is a well known statistical fact that in regression analysis, observations with extreme x-values have the most "influence" on the fitted model. "Cook's D", therefore, combines the attributes of (1) large absolute residual with (2) extreme design point, to identify points which may be outliers among the observations.

Observation 15 was eliminated from the analysis, and a new "reduced model" was fitted to the remaining observations. The fit to that model is represented by the response surface in Figure 2, with adjusted observations and residuals. The elimination of observation 15 has had a profound effect on the response surface fit (at least in the X2 x X3 dimension). A better fit to the data without observation 15 is further exemplified in the large value for the adjusted R^2 (adjusted R^2 = 0.95). None of the remaining observations appear to be outliers relative to the new "reduced model."

The value of the methods demonstrated above will now be discussed.

The goal of correlation analysis of the dependent variables is to understand the pairwise interrelationship among tablet properties. This can help in the analysis of obtained results, in assisting the practitioner in focusing on a single or a few tablet properties, while ignoring redundant dependent variables. Furthermore, confirmed association among several tablet properties

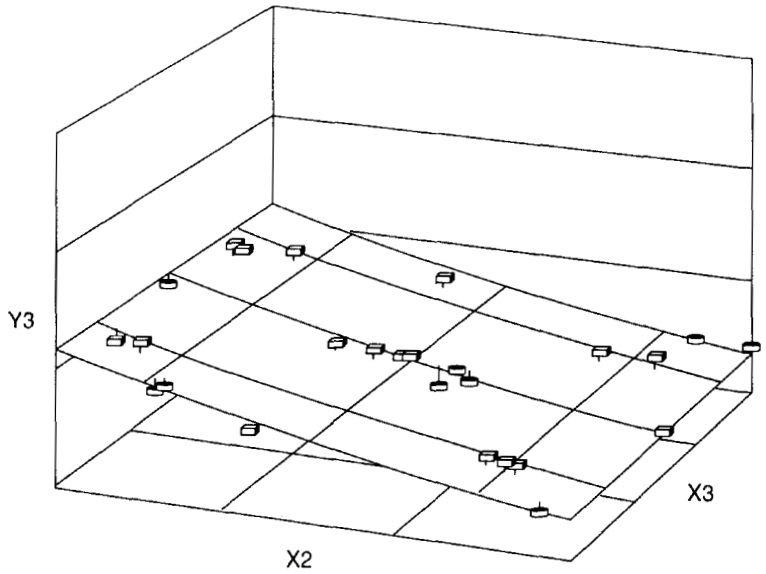

FIGURE 2

might aid in the reduction of experimental data collected, by collecting data on only a few of the nonredundant vaiables, saving on analytical time.

The goal of principal component analysis is to derive a more parsimonious set of dependent variables, as a linear combination(s) of the tablet properties. If only one or two of these linear combination (principal components) account for a substantial proportion of the correlation structure among the dependent variables, then these variables could be analyzed in lieu of the actual tablet properties. This was not done in this paper, and will be the subject of a future analysis.

The utility of stepwise regression methods is very clear in an application like optimization experiments, where very few

experimental observations are expected to support a large multi-dimensional response surface model. Unfortunately, the choice of technique is subjective, and two different stepwise procedures can give very different results. It is our experience that several options should be investigated, and the results compared. Very often, through a thorough investigation of residuals, and utilizing current state of the art graphics packages, an appropriate model can be found.

We try to emphasize appropriate use of "goodness of fit" diagnostics, including an adjusted R^2 and tests for "influential observations." These are but a few of the tools available for performing an adequate evaluation of the experimental data contained in an optimization experiment.

FOOTNOTES

[1]BPMC 10 Liter Granulator - Baker Perkins Chemical Machinery Limited, Stoke-on-Trent, England

[2]Manesty Betapress - Manesty Machines Limited, Liverpool, England.

REFERENCES

1. E. M. Scattergood, J. B. Schwartz, M. O. Villarejos, W. J. McAleer and M. R. Hilleman, Drug Develop. and Ind. Pharm., 9, 745 (1983).

2. M. L. Cotton and G. R. B. Down, J. Chromatography, 259, 17 (1983).

3. D. A. Doornbos, Pharm. Weekblad (Sci. Ed.), 3, 549 (1981).

4. J. B. Schwartz, J. Soc. Cosmet. Chem., 32, 287 (1981).

5. J. B. Schwartz, J. R. Flamholz and R. H. Press, J. Pharm. Sci., 62, 1165 (1973).

6. J. B. Schwartz, J. R. Flamholz and R. H. Press, ibid., 62, 1518 (1973).

7. N. R. Bohidar, J. F. Bavitz, and P. K. Shiromani, Drug Develop. and Ind. Pharm., 12, 1503 (1986).

8. N. R. Bohidar, F. A. Restaino, and J. B. Schwartz, ibid., 5, 175 (1979).

9. N. R. Bohidar, F. A. Restaino, and J. B. Schwartz, J. Pharm. Sci., 64, 966 (1975).

10. G. W. Snedecor and W. G. Cochron, "Statistical Methods" Seventh Edition, The Iowa State University Press, Ames, Iowa, USA (1980).

11. D. F. Morrison, "Multivariate Statistical Methods", Mc-Graw-Hill Book Company, New York, (1967).

12. N. Draper and H. Smith, "Applied Regression Analysis", 2nd Edition, John Wiley & Sons, Inc., New York, (1981).