A METHOD FOR ANALYSIS OF RESPONSE CURVE

By

D.S. TONK AND UMED SINGH

Haryana Agricultural University, Hissar

(Received : August, 1979)

SUMMARY

Experimenters often make several observations on a given experimental unit. If these observations can be associated with some continuous variable, they collectively form a curve. Various methods in literature are studied in addition to the proposed method for the analysis of response curve data.

The procedure proposed here combines the ANOVA model and the modified principal component analysis. It develops statistics which describe the level and shape of the curves. These statistics are analyzed to determine the effects of the treatments on the curve. One example is presented to illustrate this method of analysis and interpretation. The results are compared to those obtained using the other existing methods of analysis.

1. INTRODUCTION

In biological investigations the growth of an animal (or plant) or part of an animal (or plant) is often the subject of study and the experimenters often make several observations on a given experimental unit. If these observations can be associated with some continuous variable, such as time or temperature, they collectively form a curve. In many situations, it is important to determine the effects of different experimental conditions on such curves. Curves are characterized by two attributes, namely (*i*) the level of the curve and (*ii*) the shape of the curve. Several authors have presented methods which are applicable to data of this type.

Wishart [8] recommended that a general model (linear, quadratic, exponential, etc.) be fitted to each curve separately producing a set of estimated parameters for each experimental unit. These parameters are analysed statistically to determine the effects of the experimental conditions (treatments) on the curve. Box [1] suggested the conventional analysis of variance for the analysis of growth curves. The analysis was performed on successive differences. If certain assumptions were not valid, he suggested a multivariate analysis. Other multivariate procedures have been presented by Danford *et al.* [4] and Cole and Grizzle [3]. Church [2] presented a method where by a principal component analysis is used to transform the curve into orthogonal components, called derived responses, which reflect the characteristics of the curves. These derived responses are analysed to test the significance of the treatments on the curves. Snee [7] presented a method which combines the analysis of variance model suggested by Box and the principal component analysis proposed by Church. It develops statistics which describe the level and shape of the curves. These statistics are analysed to determine the effects of the treatments on the curve.

Some problems arise, however, with the above procedures. Wishart's procedure requires a general model for some data sets; on the other hand, such a model may contain many parameters, complicating the overall interpretation of the experiment. The assumptions required for the valid use of analysis of variance are some times not satisfied; while the multivariate procedures are less powerful than the univariate analysis of variance. To test the multivariate assumption of homogeneous co-variance matrices, the number of samples per treatment (n) must be greater than the number of measurements (p) per experimental unit. Thus, if the number of measurements is large, the amount of replication needed for a valid multivariate analysis becomes prohibitive, especially where the cost per experimental unit is high. Although the multivariate procedures provide sophisticated tests of significance for differences among the curves, they are difficult for scientists to understand and interpret and do not provide any insight as to the nature in which the curves are different. The principal component analysis sometimes yields derived responses whose physical meaning is difficult to determine. Snee's method overcomes most of the difficulties encountered by other methods but the derived responses it yields may not be uncorrelated.

The method presented in this paper is a modified form of the method poposed by Snee and it results into derived responses which are uncorrelated. It also develops statistics which describe the level and shape of the curves. These statistics are analysed to determine the effect of the treatments on the curve. An example is presented to illustrate this method of analysis and interpretation. The results are compared to those obtained using the other method of analysis.

2. Proposed Method

Consider the situation in which an experimenter has made p equi-spaced measurements on each of k treatments or experimental conditions which collectively form a curve and under i-th treatment let there are N_i experimental units (individuals) or sampling units so that the total number of experimental units are $N=N\Sigma N_i$, $i=1, 2, \ldots, k$. The observations in this set up are described by the mathematical model

$$y_{irj}=\mu_c+\alpha_i+\beta_j+(\alpha\beta)_{ij}+e_{irj}; \qquad \ldots(2.1)$$

$$i=1, 2, \ldots, k$$

$$r=1, 2, \ldots, N_i$$

$$j=1, 2, \ldots, p.$$

where μ_c is a constant

α_i is the i-th treatment effect,

β_j is the j-th measurement effect,

$(\alpha\beta)_{ij}$ is the ij-th treatment $\times$ measurement interaction effect,

and

e_{irj} is the error component.

Differences in the parameters in this model describe the effect of the treatments on the curve. A significant treatment effect indicates the treatments (α_i) have an effect on the level of the curve where

$$\bar{y}_{ir}=\frac{1}{p}\sum_{j=1}^{p} y_{irj} \qquad \ldots(2.2)$$

defines the level of the curve. A significant Treatment $\times$ measurement interaction $[(\alpha\beta)_{ij}]$ indicates the treatments have a significant effect on the Shape of the curve. This is equivalent to saying the treatment curves are not parallel. Thus, the curves are evaluated according to two criteria—level and shape.

The assumptions for this model are :

1. The errors of prediction are normally distributed with mean zero, and
2. The $(p\times p)$ measurement covariance matrix Σ is equal for all experimental units.

If Y_{ir} denotes the observation p-vector or r-th individual under i-th treatment, then

$$Y_{ir}=(y_{ir1}, y_{ir2}, \ldots, y_{irp})'$$

and $$E(Y_{ir})=\mu_i=(\mu_{i1}, \mu_{i2}, \ldots, \mu_{ip})'$$

Also $\text{Var}\,(Y_{ir})=\underset{p\times p}{\Sigma}$

Defining a $(p\times p)$ matrix $E(e_{uv})$,

where $e_{uv}=(uv)$-th element of the E-matrix

$$=\sum_{i=1}^{k}\left[\sum_{r=1}^{N_i} y_{iru}\, y_{irv}-\frac{1}{N_i}\, T_{iu}T_{iv}\right]$$

T_{iu}=sum of all observations on the u-th response in the presence of the i-th treatment.

$$=\sum_{r=1}^{N_i} y_{iru}$$

and $$T_{iv}=\sum_{r=1}^{N_i} y_{irv},\ u,\ v=1,\ 2,\ \ldots,\ p.$$

Clearly an unbiased estimate of (uv)-th element of Σ is given by

$$\frac{1}{N-K}e_{uv}$$

i.e. $$\hat{\Sigma}=\frac{1}{N-K}E \qquad \ldots(2.3)$$

Our interest is to derive a new set of variables such that the covariance matrix of the new p-vector of variables is a diagonal one.

The covariance matrix Σ can be expressed as

$$\Sigma=BD\lambda \qquad \ldots(2.4)$$

where $D\lambda$ is a diagonal matrix assumed to be of full rank and $\underset{p\times p}{B}$ is a symmetric matrix and of full rank. It is known from linear algebra that there exists a non-singular matrix P such that

$$P'\Sigma P=D\lambda \qquad \ldots(2.5)$$

Consider the transformation

$$\underset{p\times 1}{Z}=\underset{p\times p}{P'}\ \underset{p\times 1}{Y}, \qquad \ldots(2.6)$$

where $\underset{p\times 1}{Y}$ is a observation p-vector on an individual.

Obviously

$$\operatorname{Var}\underset{p\times 1}{(Z)} = P'\Sigma P = D\lambda \qquad \ldots(2.7)$$

Most of the time it is not possible to know Σ in advance but in most situations a consistent estimate Σ based on large sample size is available. In such situations we find $\underset{p\times p}{M}$ such that

$$M'\hat{\Sigma}M = D'_\lambda \qquad \ldots(2.8)$$

Here $\underset{p\times p}{M}$ is a matrix of eigen vector of $\hat{\Sigma}$ or say E (since the eigenvectors of E are the same as of $\hat{\Sigma}$). D'_λ is a diagonal matrix. Without loss of generalization we assume that the diagonal elements in D'_λ are in descending order.

Finally we consider

$$\underset{p\times 1}{Z^*} = \underset{p\times p}{M'}\ \underset{p\times 1}{Y}$$

$$= \begin{pmatrix} M_1' \\ M_2' \\ \vdots \\ M_p' \end{pmatrix} Y$$

$$= \begin{pmatrix} M_1' Y \\ M_2' Y \\ \vdots \\ M_p' Y \end{pmatrix} = \begin{pmatrix} Z_1^* \\ Z_2^* \\ \vdots \\ Z_p^* \end{pmatrix} \qquad \ldots(2.9)$$

$M_1, M_2, \ldots, M_p$ are the p eigenvectors of $\hat{\Sigma}$ and have the property that

$$\left.\begin{aligned} M_i\, M_j &= 0 \text{ for } i\neq j \\ M_i\, M_i &= 1 \end{aligned}\right\} \qquad \ldots(2.10)$$

Due to property (2.10) z_i^* and z_j^* for $i\neq j$ are uncorrelated. Note that corresponding to each observation vector Y_p we get Z_p^* vector of transformed variables whose components are uncorrelated. But it will be worthwhile in our analysis to consider only w ($w<p$) components of $\underset{p\times 1}{Z^*}$ which account for most of the variability. The smaller the number w the greater is the efficiency of transformation (2.9).

The derived responses by this method are given by

$$Z^{*}_{irj'},\ M'_{j'},\ Y_{ir}, \qquad \ldots(2.11)$$

$$j'=1, 2, \ldots, w.$$

Centering the observation vectors, the derived responses can be expressed as

$$Z^{**}_{irj'} = M'_{j'}\ (Y_{ir} - \bar{Y}),\ \ j'=1, 2, \ldots, w. \qquad \ldots(2.12)$$

where $\bar{Y}$ is the mean observation vector.

Vector of derived responses is given by

$$Z^{**}_{ir} = (z^{**}_{ir1},\ z^{**}_{ir2},\ \ldots,\ z^{**}_{irw}).'$$

Illustration

A experiment was conducted at Haryana Agricultural University, Hissar to find out the storage behaviour of various cultivars of muskmelon. Seven important commercial varieties *viz.* Arkjeet, Hara Madhu, Pb. Sunheri, Pb. Hybrid, DPM, Sel-I, and S-445 were assessed. Varieties were packed in bambo baskets using paper cutting as packing material and kept at room temperature. The loss in weight of the experimental units was observed at two days interval. The four time points in the experiment were 0th, 2nd, 4th and 6th day. Two units of each variety were stored giving two repeatitions.

The experimental data given in Table 1 are analysed by various existing methods and by the method proposed here.

3.1. Analysis by Wishart Method

Orthogonal quadratic polynomials are fitted to each experimental unit. This produces a set of parameter estimates $(\hat{\mu}, \hat{\alpha}, \hat{\beta})$ corresponding to each experimental unit. In order to study the significance of the level and shape of the curve, the parameter estimates are analysed separately. The univariate analysis of variance is performed on each parameter estimate and is presented in Table 2.

The ANOVA table indicates that the varieties are found to be differing significantly in storage period for the level of the curve. The linear and quadratic components so of the shape *i.e.* $\hat{\alpha}$ and $\hat{\beta}$ are found to be non-significant statistically.

TABLE 1

Evaluation of different Muskmelon cultivars for their storage behaviour
(weight loss in gms at selected abscissas)

Treatment (curve)	*Selected abscissas*							
	1		*2*		*3*		*4*	
	R_I	R_{II}	R_I	R_{II}	R_I	R_{II}	R_I	R_{II}
1. Arkajeet	1500	1500	1485	1491	1460	1465	1445	1432
2. Hara Madhu	2600	2350	2510	2225	2330	2080	2080	2040
3. Pb. Sunheri	3680	3700	3525	2950	3450	2830	3270	2750
4. Pb. Hybrid	1400	880	980	825	915	755	880	730
5. DPM	670	810	610	750	570	690	545	680
6. Sel-1	3910	3121	3685	2960	3380	2735	3230	2660
7. S-445	2730	2510	2580	2375	2400	2200	1960	2150
Average	2355.71	2124.29	2196.43	1939.43	2072.14	1822.14	1915.71	1774.43

TABLE 2

ANOVA of $\hat{\mu}$, $\hat{\alpha}$ and $\hat{\beta}$

Parameter	*Source*	*d.f.*	*M.S.*	*F*
	Treatments	6	2167846	39.00*
$\hat{\mu}$	Error	7	55590	
	Total	13	1030478	
	Treatments	6	2907	2.50
$\hat{\alpha}$	Error	7	1163	
	Total	13	1968	
	Treatments	6	2740	0.76
$\hat{\beta}$	Error	7	3619	
	Total	13	3213	

3.2. Analysis by Box Method

The analysis of variance for the data by Box method is presented in Table 3.

TABLE 3

Univariate ANOVA

Source	*d.f.*	*M.S.*	*F*
Period	3	401173 3	0.50*
Treatments	6	8671386	39.00*
Individual within Treatments	7	222361	
Period × Treatment Interaction	18	24250	1.84
Individuals × Period within Treatment	21	13151	
Total	55	1009110	

Univariate analysis of variance table shows that the observations at different points of time as well as the variety effects are found differing highly significantly but the interaction between time and varieties is non-significant. The interpretations here can also be made in terms of level and shape-the two criteria of the curve since Greenhouse and Geisser [6] define level as the treatment main effects and shape as the (time × treatment) interaction.

3.3. Analysis by Church Method

The sum of squares and products matrix S is computed for the given set of experimental data in Table 1. The characteristic values and corresponding characteristic vectors (L_1, L_2, L_3, L_4) are computed on a computer and are presented here in Table 4.

TABLE 4

Principal component analysis

Selected abscissas (responses)	Components eigenvectors)			
	L_1	L_2	L_3	L_4
1	0.5479	—0.8311	0.0919	0.0260
2	0.5115	0.2567	—0.5591	—0.5999
3	0.4824	0.3148	—0.2494	0.7785
4	0.4533	0.3799	0.7853	—0.1829
Eigenvalues or variance explained	53848115	313806	118007	17596
Percentage of total variability	99.17	0.58	0.22	0.03

The results in Table 4 above indicate that only one linear combination will account for most of the variability. The elements of the first vector are all positive and are approximately of the same magnitude indicating that this vector reflects the differences in the level of the curves. Finally, the analysis of variance on derived response z_1 is performed and is reported in Table 5.

TABLE 5

ANOVA of the derived response z_1 (level)

Source	d.f.	M.S.	F
Treatment	6	8714778	39.10*
Error	7	222778	
Total	13	4142163	

It is seen from Table 5 that treatments are having a significant effect on the level of the curves.

3.4. Analysis by Snee Method

The sum of squares and products matrix E of the error elements is computed for the experimental data. The characteristic values and corresponding characteristic vectors (M_1, M_2, M_3, M_4) of matrix E are presented here in Table 6.

TABLE 6

Modified principal component analysis

Selected abscissas (responses)	*Components (eigenvectors)*			
	M_1	M_2	M_3	M_4
1	0.4798	—0.8586	0.1677	0.0663
2	0.5613	0.1851	—0.3757	—0.7158
3	0.5328	0.2747	—0.3932	0.6972
4	0.4134	0.3911	0.8222	—0.0063
Eigenvalues or variance explained	1575867.30	201257.17	54502.30	1075.99
Percentage of total variability	85.99	10.98	2.97	0.06

The analysis results reported in Table 6 indicate that only two linear combinations will account for most of the variability in the data. The elements of first vector are all positive and are of the same magnitude indicating, this vector is reflecting the differences in the level of the curves. The second vector have positive and negative elements both which indicate that this reflect differences in the shape of the curves.

Computing the values of $\bar{y}_{ir}$ and derived responses $\tilde{z}_1$ and $\tilde{z}_2$ for each of the experimental unit, the analysis of variance on derived response $\tilde{z}_2$ and mean value $\bar{y}_{ir}$ is performed. The results are presented in Table 7.

There is no need to analyze the $\tilde{z}_1$ derived response since this reflect the level of the curve and analysis for level has already been presented. In Table 7 the statistics $\bar{y}_{ir}$ describe the level of the curve and derived response $\tilde{z}_2$ describe the shape of the curve. The analysis of variance results reported show that treatments have a large effect on the level of the curves. The treatments have a non-significant effect on the shape of the curves.

TABLE 7

ANOVA of the derived response $\tilde{z}_2$ (shape) and mean value $\bar{y}_{ir}$ (level)

Source	*d.f.*	y_{ir} *(level)*		$\tilde{z}_2$ *(shape)*	
		M.S.	*F*	*M.S.*	*F*
Treatment	6	2167846	39.00*	53141[illegible]	1.59
Error	7	55590		28751	
Total	13	1030478		40008	

3.5. Analysis of Proposed Method :

The analysis results in Table 6 show that only two linear combinations will account for most of the variability. So, the derived responses z_1^{**} and z_2^{**} are computed for each of the individual by using the derived responses vector equation (2.12). Finally, these derived responses are analysed statistically and analysis is presented in Table 8.

TABLE 8

ANOVA of the derived responses z_1^{**} (level) and z_2^{**}(shape)

Source	*d.f.*	z_1^{**} (level)		z_2^{**} (shape)	
		M.S.	*F*	*M.S.*	$F\alpha$
Treatment	6	8611716	38.25*	57450	2.00
Error	7	225124		28751	
Total	13	4095859		41997	

The results reported in Table 8 show that treatments are observed to be largely effecting the level of the curves, which is described by the derived response z_1^{**}. Here the level of the curve can also be studied using the statistic $\bar{y}_{ir.}$ as in the Table 7. The differences in the shape of the curves are found to be non-significant, statistically.

It is seen that slight voilation of the assumption doesn't adversely effect the analysis procedures as is clear from the interpretations of the results of an experimental data set by different methods.

The analysis of regression coefficients is probably the most popular method of analysis response curve data. This method works best when the curves can be described by a simple model with two or three coefficients and it requires the assumption $\Sigma = I\sigma^2$. It is observed that in most cases $\Sigma \neq I\sigma^2$. However, it is not always possible to find such a model. In these situations, the modified principal component analysis alongwith ANOVA is recommended. For some data sets two or more regression coefficients may be required to describe the differences in the shape of the curve, while the same differences can be described by only one or two derived responses. Methcds available in literature do not yield uncorrelated estimates of the parameters, however, the proposed me hod yields derived responses which are uncorrelated.

The principal component analysis proposed by Church (1966) sometimes yields derived responses whose physical meaning is difficult to determine. A detailed discussicn of shortcoming of performing the principal component analysis has been presented by Gollob [5].

The procedure proposed here combines the ANOVA model and the modified principal component analysis. It develcps statistics which are uncorrelated and describe the level and shape of the curves.

REFERENCES

[1] Box, G.E.P. (1950). : Problems in the analysis of growth and wear curves. *Biometrics* **6**, 362-369.

[2] Church, A. Jr. (1966) : Analysis of data when the response is a curve. *Technometrics* **8**, 229-246.

[3] Cole, J.W.L. and Grizzle, J.E. (1966) : Application of multivariate analysis of variance to repeated measurements experiments. *Biometrics* **22**, 810-828.

[4] Danford, M.B., Hughes, H.M., and McNEE, R.C. (1960) : On the analysis of repeated-measurements experiments. *Biometrics* **16**, 547-665.

[5] Gollob, H.F. (1968) : Confounding of sources of variation in factor-analytic techniques. *Psychological Bulletin* **70**, 330-344.

[6] Greenhouse, S.W. Geisser, S. (1959) : On methods in the analysis of profile data. *Psychometrika* **24**, 95-112.

[7] Snee, R.D. (1972) : On the analysis of response curve data. *Technometrics* **14**, 47-62.

[8] Wishart, J. (1938) : Growth-rate determinations in nutrition studies with the bacon pig and their analysis. *Biometrika* **30**, 16-28.