Colloquium Biometricum, 37
2007, str. 95–104

RELATIVE POTENCY OF PREPARATIONS BASED ON PRINCIPAL COMPONENTS ANALYSIS OF MULTIVARIATE RESPONSES

Joanna Olejnik, Zofia Hanusz

Department of Applied Mathematics, Agricultural University
Akademicka 13, 20-950 Lublin
e-mails: joanna.olejnik@.ar.lublin.pl, zofia.hanusz@ar.lublin.pl

Summary

Estimation of a test preparation efficiency in relation to a standard preparation in multivariate bioassays is presented in this paper. This estimation is provided if null hypotheses about parallelism and the relative potency fail to reject. However, when many traits in the response for dosages of preparation are considered, both hypotheses can be rejected. Then we cannot estimate the relative potency or we can choose the traits in the response which are the most important for the experimenter responsible for acceptation of the hypotheses. In order to reduce dimensionality of the responses, using principal components analysis is proposed in this paper. Principal components related to the greatest eigenvalues of a sample correlation matrix are taken into account. Test functions for the transformed responses are given. Theoretical results presented in the paper are applied to estimate the efficiency of two forms of nitrogen fertilizer in tobacco and winter wheat experiments.

Key words and phrases: relative potency, multivariate bioassays, principal components analysis, parallel-line assay, multivariate observations, completely randomized block design, testing hypotheses

Classification AMS 2000: 62H15, 62H12

1. Introduction

In the multivariate biological experiments (bioassays), to estimate an effect of a test preparation in relation to an effect of a standard preparation, a relative potency can be used. The relative potency of two preparations is defined as a ratio of doses of preparations giving the same responses, i.e. $\rho = \frac{u_T}{u_S}$, where u_S and u_T are the doses of a standard and a test preparations, respectively, producing similar responses. The estimate of the potency specifies the dose of the test preparation which should be administered to produce similar response by unit dose of the standard preparation. In estimation of the relative potency we test a hypothesis about the same vectors of slopes for the preparations (parallel-line assay) and a hypothesis describing the relation between the relative potency and model parameters are tested. The relative potency can be estimated if both hypotheses are true. However, in practice, for responses having many measurable traits, both hypotheses are frequently rejected. In such a case we are not able to estimate the relative potency or we can choose some traits in responses, which are the most important for the experimenters. Sometimes this technique allows to estimate the relative potency of preparations. In this paper we propose to use spectral decomposition of correlation matrix in order to find out the most important combinations of responses to reduce a dimensionality. This method will be applied to experimental data sets considered in Hanusz et al. (2003) and Hanusz, Rutkowska (2006). In these papers, the problem of estimation of a relative efficiency of two forms of nitrogen fertilizer in tobacco and winter wheat experiments were considered. Several doses of both preparations were administered to plots forming complete block designs. The influence of the nitrogen fertilizer was measured by six and nine traits, respectively. The goal of the papers was to estimate the dose of nitrogen administered in different way during plants vegetation with respect to nitrogen administered in the classical way. To get answer, the method of multivariate bioassay was used. However, for the data sets obtained in experiments, for all traits considered in the response, the estimation of the relative potency can not be provided. In such cases, estimation of the relative potency of preparations can be done for some traits in the response. In the present paper we apply the spectral decomposition of correlation matrix to choose the most important combinations of all traits.

2. The linear model of responses in multivariate parallel-line bioassay

Let us assume that an effect of a test preparation T is compared to an effect of a standard preparation *S*. Let ν_i denote the number of different doses of the *i*th preparation $(i = S,T)$. Let u_{ij} denote the *j*th dose of *i*th preparation, where $i = S,T$; $j = 1,2,\cdots,\nu_i$. Next, we assume that doses of both preparations are administered to experimental plots forming complete randomized block design with *b* blocks. Let us assume that the effect of preparations can be measured by *p* measurable traits. Let $\mathbf{y}_{ijk}$ denote a $(p\times 1)$ response vector for the *j*th dose of the *i*th preparation applied in the *k*th block, where $i = S,T$; $j = 1,2,\cdots,\nu_i$; $k = 1,2,\cdots,b$. Each response can be described as follows

$$\mathbf{y}_{ijk} = \boldsymbol{\tau}_k + \boldsymbol{\alpha}_i + x_{ij}\,\boldsymbol{\beta}_i + \mathbf{e}_{ijk}, \tag{2.1}$$

where $x_{ij} = \log u_{ij}$, $\boldsymbol{\tau}_k$, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ denote $(p\times 1)$ vectors of block effects, intercepts and slopes, respectively. We assume that *p* dimensional response vectors in (2.1) are independently normally distributed, i.e. $\mathbf{y}_{ijk} \sim N_p\left(\boldsymbol{\tau}_k + \boldsymbol{\alpha}_i + x_{ij}\boldsymbol{\beta}_i, \boldsymbol{\Sigma}\right)$. Then, a general linear model of all observations has the following form

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}, \tag{2.2}$$

where $\mathbf{Y}$ is an $(n\times p)$ matrix of responses described in (2.1), allocated as rows, $\mathbf{X} = [\mathbf{D}_{\boldsymbol{\tau}}, \boldsymbol{\Delta}_{\boldsymbol{\alpha}}, \boldsymbol{\Delta}_{\boldsymbol{\beta}}]$, $\mathbf{B} = [\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta}]'$, $\mathbf{D}_{\boldsymbol{\tau}} = \mathbf{1}_v \otimes \mathbf{I}_b$ $(v = v_S + v_T)$ is a matrix connected with block effects, $\mathbf{1}_n$ and $\mathbf{I}_n$ denote vector of ones and identity matrix of the size *n*, respectively, $\otimes$ is Kronecker product of matrices, $\boldsymbol{\Delta}_{\boldsymbol{\alpha}} = diag\left(\mathbf{1}_{n_S}, \mathbf{1}_{n_T}\right)$ is $(n\times 2)$ block diagonal matrix with the vectors $\mathbf{1}_{n_i}$ on diagonal, $n_i = b\cdot v_i$ $(i = S,T)$, $n = n_S + n_T$, $\boldsymbol{\Delta}_{\boldsymbol{\beta}} = diag\left(\mathbf{x}_S, \mathbf{x}_T\right)$ is $(n\times 2)$ block diagonal matrix with the $(n_i\times 1)$ vectors $\mathbf{x}_i$, including logarithms of all applied in experiment doses for the *i*th preparation, and $\boldsymbol{\tau} = [\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots, \boldsymbol{\tau}_b]$, $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_S, \boldsymbol{\alpha}_T]$, $\boldsymbol{\beta} = [\boldsymbol{\beta}_S, \boldsymbol{\beta}_T]$ are matrices of block effects, intercepts and slopes, respectively. Matrix of observation, $\mathbf{Y}$, has multivariate normal distribution, $\mathbf{Y} \sim N_{n,p}\left(\mathbf{XB}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)$.

Maximum likelihood estimates of parameter matrix and covariance matrix have the forms (see Muirhaed, 1982; Krzyśko, 2000)

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}, \qquad \mathbf{S} = \hat{\mathbf{\Sigma}} = \frac{1}{n}\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right).$$

To reduce the column rank of **Y** we define a new matrix using principal components analysis. As traits in the response matrix **Y** have different units of measure, we use spectral decomposition of correlation matrix for the traits. Let $\mathbf{R} = \mathbf{D}^{-\frac{1}{2}}\mathbf{S}\mathbf{D}^{-\frac{1}{2}}$ denote the correlation matrix, where $\mathbf{D}^{-\frac{1}{2}} = diag^{-1}\left(\sqrt{s_{11}}, \sqrt{s_{22}}, \cdots, \sqrt{s_{pp}}\right)$ and s_{jj} $(j = 1, \cdots, p)$ are diagonal elements of **S**. Let λ_1, λ_2, $\cdots, \lambda_p$ denote eigenvalues of **R**, in non increasing order, and $\mathbf{h}_1$, $\mathbf{h}_2$, $\cdots, \mathbf{h}_p$ - the corresponding orthonormal eigenvectors. Let us define the matrix $\mathbf{H}_t$, which contains *t* eigenvectors corresponding to the *t* largest eigenvalues, i.e. $\mathbf{H}_t = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_t]$. Then $\mathbf{Z} = \mathbf{Y}\mathbf{D}^{-\frac{1}{2}}\mathbf{H}_t$ has $N_{n,t}\left(\mathbf{X}\mathbf{\Theta}, \mathbf{I}_n \otimes \tilde{\mathbf{\Sigma}}\right)$, where $\mathbf{\Theta} = \mathbf{B}\mathbf{D}^{-\frac{1}{2}}\mathbf{H}_t$, $\tilde{\mathbf{\Sigma}} = \mathbf{H}_t'\mathbf{D}^{-\frac{1}{2}}\mathbf{\Sigma}\mathbf{D}^{-\frac{1}{2}}\mathbf{H}_t$.

3. Estimation of a relative potency in multivariate bioassay

In multivariate bioassays, estimation of the relative potency of two preparations is connected with testing a hypothesis about parallelism and a hypothesis about the log relative potency. In this section we give test function for both hypotheses.

3.1. Testing hypothesis about parallelism

Test preparation could be compared to a standard one by the relative potency if they have similar impact on the response. This similarity takes place if vectors of slopes for the standard and the test preparations are not significantly different. Parallelism of the responses can be formulated by a hypothesis $H_{\boldsymbol{\beta}}^0$ of the form

$$H_{\boldsymbol{\beta}}^0 : \boldsymbol{\beta}_T - \boldsymbol{\beta}_S = \mathbf{0}_p,$$

where $\mathbf{0}_p$ denotes null vector of the size *p*, or as follows

$$H_{\boldsymbol{\beta}}^{0} : \mathbf{c}'\boldsymbol{\Theta} = \mathbf{0}_t,$$

where $\mathbf{c}' = [\mathbf{0}_b', \mathbf{0}_2', \mathbf{c}_{\beta}']$, $\mathbf{c}_{\beta}' = [1, -1]$.

The hypothesis $H_{\boldsymbol{\beta}}^{0}$ can be tested, versus the alternative $H_{\boldsymbol{\beta}}^{1} : \boldsymbol{\beta}_T - \boldsymbol{\beta}_S \neq \mathbf{0}$, using Wilks' lambda test of the form (Meisner et al., 1986):

$$\Lambda = \frac{1}{1 + \left(\mathbf{c}'\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{c}\right)^{-1}\left(\mathbf{c}'\hat{\boldsymbol{\Theta}}\right)\hat{\tilde{\boldsymbol{\Sigma}}}^{-1}\left(\mathbf{c}'\hat{\boldsymbol{\Theta}}\right)' / n}, \tag{3.1}$$

where $\hat{\boldsymbol{\Theta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'\mathbf{Z}$, $\hat{\tilde{\boldsymbol{\Sigma}}} = \frac{1}{n}\left(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\Theta}}\right)'\left(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\Theta}}\right)$.

The hypothesis $H_{\boldsymbol{\beta}}^{0}$ is rejected if $-\left(n - r(\mathbf{X}) - \frac{t-1}{2}\right) ln\Lambda > \chi_{t-1,\alpha}^{2}$, where $r(\mathbf{X})$ denotes the rank of the matrix **X** and α is a known significant level. The effect of a test preparation is similar to the effect of a standard preparation if $H_{\boldsymbol{\beta}}^{0}$ is not rejected.

3.2. Testing hypothesis about the relative potencies

Let us suppose that the hypothesis $H_{\boldsymbol{\beta}}^{0} : \boldsymbol{\beta}_T - \boldsymbol{\beta}_S = \mathbf{0}_p$ failed to reject. Then, the matrices **X** and **B** in model (2.2) take new forms, namely, $\tilde{\mathbf{X}} = [\mathbf{D}_{\tau}, \boldsymbol{\Delta}_{\alpha}, \mathbf{x}]$, $\mathbf{x} = [\mathbf{x}_S', \mathbf{x}_S']'$, and instead of the last two rows in **B** we put the only one vector $\boldsymbol{\beta}$, the same vector of slopes for the standard and the test preparations. To estimate the potency of the test preparation relative to the standard preparation, the following hypothesis should be true

$$H_{\mu}^{0} : \boldsymbol{\alpha}_T - \boldsymbol{\alpha}_S = \mu\boldsymbol{\beta},$$

where μ denotes the logarithm of potency of the test preparation to the standard preparation, $\mu = \log\rho$. To test the hypothesis H_{μ}^{0}, the Wilks' lambda statistic could be also used

$$\Lambda(\mu)=\frac{1}{1+\left[\mathbf{c}_{\mu}'\left(\tilde{\tilde{\mathbf{X}}}'\tilde{\tilde{\mathbf{X}}}\right)^{-}\mathbf{c}_{\mu}\right]^{-1}\left(\mathbf{c}_{\mu}'\hat{\hat{\boldsymbol{\Theta}}}\right)\hat{\hat{\boldsymbol{\Sigma}}}^{-1}\left(\mathbf{c}_{\mu}'\hat{\hat{\boldsymbol{\Theta}}}\right)'/n}, \tag{3.2}$$

where $\mathbf{c}_{\mu}'=[\mathbf{0}_b',-1,1,-\mu]$, $\hat{\hat{\boldsymbol{\Theta}}}=\left(\tilde{\tilde{\mathbf{X}}}'\tilde{\tilde{\mathbf{X}}}\right)^{-}\tilde{\tilde{\mathbf{X}}}'\mathbf{Z}$. The hypothesis H_{μ}^{0} is rejected if $-\left(n-r(\tilde{\tilde{\mathbf{X}}})-\frac{t-1}{2}-\frac{min\Lambda(\mu)}{1-min\Lambda(\mu)}\right)ln\Lambda(\hat{\mu})>\chi_{t-1,\alpha}^{2}$.

Maximum likelihood estimator of μ is defined as $\hat{\mu}$ which maximizes $\Lambda(\mu)$ and for which the hypothesis H_{μ}^{0} is not rejected.

4. Numerical illustration of the results

We apply the estimation of the relative potency of preparations, based on principal components, presented in the previous section, to assess the efficiency of nitrogen forms applied in two different experiments. The first one was considered in Hanusz et al. (2003) and the second one in Hanusz, Rutkowska (2006). Both experiments were performed in randomized complete block designs with four blocks.

Table1. Correlation matrix, eigenvalues, proportions and cumulative proportions

Correlation matrix	Eigenvalue	Proportion	Cumulative
$\mathbf{R}=\begin{bmatrix} 1.000 & -.005 & .082 & .058 & -.120 & -.104 \\ -.005 & 1.000 & -.042 & -.200 & .739 & .719 \\ .082 & -.042 & 1.000 & -.249 & .206 & .104 \\ .058 & -.200 & -.249 & 1.000 & -.441 & -.004 \\ -.120 & .739 & .206 & -.441 & 1.000 & .899 \\ -.104 & .719 & .104 & -.004 & .899 & 1.000 \end{bmatrix}$	$\lambda_1 \cong 2.698$	0.450	0.450
	$\lambda_2 \cong 1.229$	0.205	0.655
	$\lambda_3 \cong 1.017$	0.169	0.824
	$\lambda_4 \cong 0.783$	0.130	0.954
	$\lambda_5 \cong 0.272$	0.045	0.999
	$\lambda_6 \cong 0.0005$	0.00008	$\cong 1$

***Example* 1.** In the tobacco experiment (see Hanusz et al., 2003), both nitrogen fertilizers were applied in doses: 7.5, 15, 22.5, 30, 45 $\mathrm{kg\,N\cdot ha^{-1}}$ in the form of lime saltpeter or ammonium saltpeter. In the response vector, six traits were measured: y_1 - yield of tobacco leaves, y_2 - a height of tobacco plant, y_3 - a total number of leaves, y_4 - a certain index of the middle tobacco leaves, y_5 - a wide of tobacco leaves, y_6 - a long of tobacco leaves. For the data set obtained in 1998 we get the eigenvalues of correlation matrix shown in Table 1.
The results in Table 1, show that the first four principal components describe more then 95% of response variation while three principal components explain more then 82%. Taking into account the eigenvectors corresponding to these eigenvalues, we get the results for testing the hypotheses $H_{\boldsymbol{\beta}}^0$, H_{μ}^0 using (3.1) and (3.2) and the estimate of the relative potency shown in Table 2.

Table2. Test functions and the relative potency estimates for chosen principal components

Number of principal components	$H_{\boldsymbol{\beta}}^0 : \boldsymbol{\beta}_S = \boldsymbol{\beta}_T$		$H_{\mu}^0 : \boldsymbol{\alpha}_S - \boldsymbol{\alpha}_T = \mu\boldsymbol{\beta}$		Potency estimate
	χ_0^2	*p-value*	χ_0^2	*p-value*	$\hat{\rho}$
4	1.813	0.612	2.353	0.503	2.44
3	1.827	0.401	2.342	0.310	2.47

For three and four principal components both hypotheses are not rejected on significant level 0.05 so we can estimate the relative potency. The estimates in both cases are very similar and are almost the same and equal about two and the half. It means that to get the same effect in tobacco yielding the dose of the test fertilizer should be twice and half bigger than the dose of the standard fertilizer.

***Example* 2.** In the experiment on winter wheat (see Hanusz, Rutkowska, 2006) doses of a standard form of nitrogen: 90, 130, 170, 210, and 250 $\mathrm{kg\,N\cdot ha^{-1}}$ and for a test: 130, 170, 210, 250 and 270 $\mathrm{kg\,N\cdot ha^{-1}}$ were administered, respectively. Eight traits in the response vector were measured: x_1 - chlorophyll in ear emergence, x_2 - chlorophyll in florescence, x_3 - yield of grains, x_4 - collection of zinc, x_5 - % N, x_6 - weight of 1000 grains, x_7 - number of ears, x_8 - number of grains per ear. In the experiment basic nitrogen fertilization in different doses and different period of plants' vegetation was applied. For the obtained data set, we get the eigenvalues of correlation matrix shown in Table 3.

Table 3. Correlation matrix, eigenvalues, proportions and cumulative proportions

Correlation matrix	Eigenvalue	Proportion	Cumulative
$\mathbf{R} = \begin{bmatrix} 1.00 & .71 & .46 & .47 & .16 & .42 & .06 & .01 \\ .71 & 1.00 & .59 & .49 & -.13 & .49 & .02 & .09 \\ .46 & .59 & 1.00 & .92 & -.14 & .58 & .24 & .38 \\ .47 & .49 & .92 & 1.00 & .22 & .57 & .19 & .29 \\ .16 & -.13 & -.14 & .22 & 1.00 & .01 & -.24 & -.24 \\ .42 & .49 & .58 & .57 & .01 & 1.00 & .09 & .08 \\ .06 & .02 & .24 & .19 & -.24 & .09 & 1.00 & .07 \\ .01 & .09 & .38 & .29 & -.24 & .08 & .07 & 1.00 \end{bmatrix}$	$\lambda_1 \cong 3.399$	0.425	0.425
	$\lambda_2 \cong 1.446$	0.181	0.606
	$\lambda_3 \cong 0.980$	0.122	0.728
	$\lambda_4 \cong 0.920$	0.115	0.843
	$\lambda_5 \cong 0.626$	0.078	0.921
	$\lambda_6 \cong 0.418$	0.052	0.973
	$\lambda_7 \cong 0.198$	0.025	0.998
	$\lambda_8 \cong 0.014$	0.002	1

The results presented in Table 3 show that the first five eigenvalues explain more then 92% of response variation but four or three principal components describe more then 84% or 72%, respectively. Taking into account the eigenvectors corresponding to these eigenvalues, we get the results of testing hypotheses $H_{\boldsymbol{\beta}}^0$, H_{μ}^0 and the estimate of the relative potency shown in Table 4.

Table 4. Test functions and the relative potency estimates for chosen principal components

Number of principal components	$H_{\boldsymbol{\beta}}^0 : \boldsymbol{\beta}_S = \boldsymbol{\beta}_T$		$H_{\mu}^0 : \boldsymbol{\alpha}_S - \boldsymbol{\alpha}_T = \mu\boldsymbol{\beta}$		Potency estimate
	χ_0^2	*p-value*	χ_0^2	*p-value*	$\hat{\rho}$
5	8.438	0.079	6.101	0.192	0.81
4	8.548	0.036	2.433	0.487	0.87
3	7.320	0.026	0.947	0.623	0.86

For all five considered principal components, the hypotheses $H_{\boldsymbol{\beta}}^0$ and H_{μ}^0 are not rejected on significant level 0.05 so we are able to estimate the relative efficiency of two forms of nitrogen fertilizers. Estimate of the potency in these cases are a little bit smaller than 0.9. We can conclude that both fertilizers are almost of the same potency. We can take 0.9 of the unit dose of the standard fertilizer to get the same responses.

5. Conclusions

In the paper we have shown that principal components analysis can be applied to reduce the dimension of responses in the multivariate bioassays. Test functions for testing the hypothesis about a parallel-line assay and the hypothesis about the relative potency of preparations were modified for transformed matrix of the data. Using tests given in the paper the estimation of the relative efficiencies of different form of nitrogen fertilizers in tobacco and winter wheat experiments was done.

References

Hanusz Z., Kowalczyk-Juśko A., Olejnik J. (2003). Estimation f relative potency of two nitrogen fertilizers in analysis o tobacco yielding. *Fragmenta Agronomica* 4(80), 32-42.

Hanusz Z., Rutkowska A. (2006). Comparison of several test preparations to one standard. *Biometrical Letters* 43 (2), 99-107.

Krzyśko M. (2000). *Wielowymiarowa analiza statystyczna*. UAM Poznań.

Meisner M., Kushner H.B., Laska E.M. (1986). Combining multivariate bioassay. *Biometrics* 42, 421-427.

Muirhead R.J. (1982). *Aspects of Multivariate Statistical Theory*. J. Wiley&Sons, New York.

WZGLĘDNA EFEKTYWNOŚĆ PREPARATÓW OPARTA NA ANALIZIE SKŁADOWYCH GŁÓWNYCH WIELOWYMIAROWYCH OBSERWACJI

Streszczenie

W pracy przedstawiona została metoda estymacji efektywności preparatu testowego względem standardowego dla wielowymiarowych doświadczeń biologicznych. Estymacja ta związana jest z testowaniem dwóch hipotez o równoległości prostych regresyjnych oraz zależności względnej efektywności od parametrów modelu. W sytuacji, gdy w doświadczeniu badanych jest wiele cech określających reakcję na stosowane dawki preparatów, nie zawsze powyższe hipotezy są prawdziwe. W efekcie tego, nie możemy oszacować względnej efektywności lub szacujemy ją dla wybranych cech w wektorze obserwacji, które dla eksperymentatora są najważniejsze, i dla któ-

rych obydwie hipotezy nie zostają odrzucone. W pracy prezentujemy podejście oparte na analizie składowych głównych. Spośród wszystkich składowych głównych proponuje się wybranie kilku z nich, odpowiadających najwyższym wartościom własnym w rozkładzie spektralnym macierzy korelacji. Zastosowanie tej metody zilustrowano dwoma przykładami liczbowymi z doświadczeń, w których badano wpływ różnych form nawożenia azotem na plonowanie tytoniu i pszenicy ozimej. Doświadczenia te zostały przeprowadzone w układzie zrandomizowanych bloków kompletnych.

Słowa kluczowe: względna efektywność preparatów**,** wielowymiarowe doświadczenia biologiczne, analiza składowych głównych, doświadczenia liniowo równoległe, obserwacje wielowymiarowe, bloki kompletnie zrandomizowane, testowanie hipotez

Klasyfikacja AMS 2000: 62H15, 62H12