Colloquium Biometricum 39
2009, 69–79

TWO–WAY PROFILE ANALYSIS WITH INTERACTION

Wojciech Zieliński

Department of Econometrics and Statistics
Warsaw University of Life Sciences
Nowoursynowska 159, PL-02-787 Warszawa
e-mail: wojtek.zielinski@statystyka.info

Summary

In the paper the model of a two–way profile analysis is considered. There are given tests of hypotheses verified in such the model.

Key words and phrases: profile analysis, multivariate statistical analysis, two-way experiment

Classification AMS 2000: 62H15, 62J10

1. Introduction

In many experiments measurements are provided repeatedly on the same units. For example, the contents of a microelement is measured on the same plant repeatedly during its growing, psychological tests are provided on the same person in some time distances etc. As a result of observations for each unit a vector of measurements is obtained. It is clear that observations provided on the unit are not independent. There are at least two ways of analyzing results of such data. The first way consists in estimating the dependence within a vector of observations and then methods of growth curve analysis or time series analysis are used. Those methods usually requires a sufficiently large number of observations per unit. In the second way the methods of multivariate linear models methods are applied, especially profile analysis. Here by profile is understood the plot of results obtained on the same unit. Suppose now that observed units are divided into disjoint groups. By a group profile is meant the plot of mean values of individual profiles of units in a group. The aim of the profile analysis is to investigate interrelations between group profiles.

The literature on the multivariate statistical models is rather plenteous (see References). Some books and articles on the subject are given in bibliography. Statistical profile analysis as a part of multivariate linear models is considered in all cited papers. Unfortunately, only so called one-way profile analysis is considered, but nether two–way. As because two factor experiments are frequently provided in agriculture there is a need of tools for two–way profile analysis. Tests for hypotheses verified in two–way case are explicitly shown in the third section of the paper. In the second section of the paper main ideas of one--way profile analysis are recalled.

2. Some remarks on a one-way profile analysis

Suppose that *N* objects are divided into *k* groups of sizes $N_1, \ldots, N_k$ respectively. For each object we have a vector of *p* observations (for example, the same variate is measured *p* times). Observations may be modeled in the following way:

$$\mathbf{y}_{ij} = \boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_{ij},\ j = 1, \ldots, N_i,\ i = 1, \ldots, k,$$

where $\mathbf{y}_{ij} = [y_{ij1}, \ldots, y_{ijp}]'$ is the vector of observations on the *j*-th unit in the *i*-th group, $\boldsymbol{\xi}_i = [\xi_{i1}, \ldots, \xi_{ip}]'$ is the mean vector for *i*-th group and $\boldsymbol{\varepsilon}_{ij}$ is the *p*-vector of random errors. The model may be written in the terms of a single observation as

$$y_{ijh} = \xi_{ih} + \varepsilon_{ijh},\ i = 1, \ldots, k,\ j = 1, \ldots, N_i,\ h = 1, \ldots, p.$$

Vectors $\boldsymbol{\xi}_1, \ldots \boldsymbol{\xi}_k$ may be presented graphically as profiles (see Figures A, B, C for $k = 3$ and $p = 4$). The main aim of the one-way profile analysis is to verify three hypotheses on profiles. The first one is the question if profiles are **parallel** (Fig. A.). This hypothesis is written in the form:

$$\begin{bmatrix} \xi_{11} - \xi_{12} \\ \vdots \\ \xi_{1,p-1} - \xi_{1p} \end{bmatrix} = \cdots = \begin{bmatrix} \xi_{k1} - \xi_{k2} \\ \vdots \\ \xi_{k,p-1} - \xi_{kp} \end{bmatrix}$$

or equivalently

$$\xi_{ih} - \xi_{i,h+1} = const \quad \forall h \quad (i = 1, \ldots, k).$$

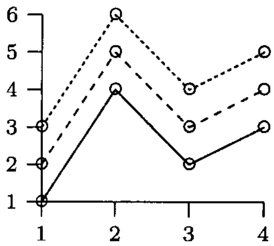

Fig. A
Parallel profiles

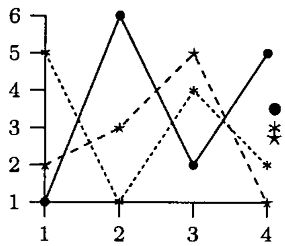

Fig. B Nonequal
levels of groups

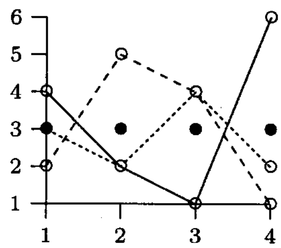

Fig. C
Equal response means

The second question is on **equal level** of groups: are the mean values of the observed p variables the same or not (different symbols on the right side of the Fig. B):

$$\sum_{h=1}^{p} \xi_{1h} = \cdots = \sum_{h=1}^{p} \xi_{kh} \quad \text{or} \quad \sum_{h=1}^{p} \xi_{ih} = const \; (i = 1, \ldots, k).$$

The third question (**equal response means**) is if the overall means in groups are the same or not (bullets in the Fig. C):

$$\sum_{i=1}^{k} \xi_{i1} = \cdots = \sum_{i=1}^{k} \xi_{ip} \quad \text{or} \quad \sum_{i=1}^{k} \xi_{ih} = const \quad \forall h .$$

The tests for all above hypotheses may be found in (for example) Morrison (1967).

The problems of comparison of profiles in a two–way experiment are analogous. Details are given in the next section.

3. Two–way profile analysis

Consider the following linear model for observations

$$y_{ijlh} = \mu_h + \alpha_{ih} + \tau_{jh} + \eta_{ijh} + \varepsilon_{ijlh},$$

for $i = 1,\ldots,k;\ j = 1,\ldots,c;\ l = 1,\ldots,m$ and $h = 1,\ldots,p;\ p > 1$. Here k is the number of levels of the first treatment, c is the number of levels of the second treatment, m is the number of independent vectors of observations in each combination and $p > 1$ is the number of random variables observed per cell, i.e. the dimension of observed vector. For convenience the overall number of observations we denote by $N = kcm$. In the model, μ_h is a general mean of hth response, α_{ih} is an effect of ith level of the first treatment on hth response, τ_{jh} is an effect of jth level of the second treatment on hth response, η_{ijh} is an effect of interaction of ith and jth levels of treatments on hth response and ε_{ijlh} is an usual normal random variable term.

In matrix notation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{Y}_{N\times p}$ is the matrix of observations, $\boldsymbol{\varepsilon}_{N\times p}$ is the matrix of random errors,

$$\mathbf{X} = [\mathbf{1}_N \,\vdots\, \mathbf{1}_c \otimes \mathbf{I}_k \otimes \mathbf{1}_m \,\vdots\, \mathbf{I}_c \otimes \mathbf{1}_{km} \,\vdots\, \mathbf{I}_{kc} \otimes \mathbf{1}_m],$$

and

$$\beta = \begin{bmatrix} \mu_1 & \alpha_{11} & \cdots & \alpha_{k1} & \tau_{11} & \cdots & \tau_{c1} & \eta_{111} & \cdots & \eta_{kc1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mu_p & \alpha_{1p} & \cdots & \alpha_{kp} & \tau_{1p} & \cdots & \tau_{cp} & \eta_{11p} & \cdots & \eta_{kcp} \end{bmatrix}'.$$

Here $\mathbf{1}_z$ is the z-vector of ones, $\mathbf{I}_z$ is the identity $(z \times z)$-matrix. In what follows the null z-vector and null $(z_1 \times z_2)$-matrix will be exploit and will be denoted by $\mathbf{0}_z$ and $\mathbf{0}_{z_1\times z_2}$ respectively.

Rank r of matrix $\mathbf{X}$ is of course kc.

Below there are given tests for all hypothesis verified in the two–way profile analysis. Because all hypothesis are of the form (A.1) (see Appendix) in all tests the statistic (A.2) is applied. Hence there will be only given explicit formulae for matrices (A.3) and (A.4). Ranks *g* and *u* of the appropriate matrices will be given as well.

Test of the parallelism of profiles of all treatments

$$H : \mu_{ijh} - \mu_{ij,h+1} = const \ \ \forall h \left(i = 1,\ldots,k;\ j = 1,\ldots,c\right).$$

Matrix notation

$$H : \mathbf{C\beta M} = \mathbf{0},$$

$$\mathbf{C} = \mathbf{C}_1\left[\mathbf{1}_{kc} \vdots \mathbf{1}_c \otimes \mathbf{I}_k \vdots \mathbf{I}_c \times \mathbf{1}_k \vdots \mathbf{I}_{kc}\right]$$

with

$$\mathbf{C}_1 = [\mathbf{I}_{kc-1} \vdots -\mathbf{1}_{kc-1}],\ \mathbf{M} = [\mathbf{I}_{p-1} \vdots -\mathbf{1}_{p-1}]' \quad .$$

Here $g = kc - 1$ and $u = p - 1$.

Test of equality of general means

$$H : \mu_h = const \ \ \forall h .$$

Matrix notation

$$H : \mathbf{C\beta M} = \mathbf{0},$$

$$\mathbf{C} = [1 \vdots \mathbf{0}_{k+c+kc}],\ \mathbf{M} = [\mathbf{I}_{p-1} \vdots -\mathbf{1}_{p-1}]'.$$

Here $g = 1$ and $u = p - 1$.

Test of the parallelism of profiles of the first treatment

$$H : \alpha_{ih} - \alpha_{i,h+1} = const \ \ \forall h \left(i = 1,\ldots,k\right).$$

Matrix notation

$$H:\mathbf{C}\boldsymbol{\beta}\mathbf{M}=\mathbf{0},$$

$$\mathbf{C}=[\mathbf{0}_{k-1}\vdots\mathbf{C}_1\vdots\mathbf{0}_{(k-1)\times(k+1)c}]$$

with

$$\mathbf{C}_1=[\mathbf{I}_{k-1}\vdots-\mathbf{1}_{k-1}],\ \mathbf{M}=[\mathbf{I}_{p-1}\vdots-\mathbf{1}_{p-1}]'.$$

Here $g=k-1$ and $u=p-1$.

Test of the parallelism of profiles of the second treatment

$$H:\tau_{jh}-\tau_{j,h+1}=const\ \ \forall h\ (j=1,\ldots,c).$$

Matrix notation

$$H:\mathbf{C}\boldsymbol{\beta}\mathbf{M}=\mathbf{0},$$

$$\mathbf{C}=[\mathbf{0}_{c-1}\vdots\mathbf{0}_{(c-1)\times k}\vdots\mathbf{C}_1\vdots\mathbf{0}_{(c-1)\times kc}]$$

with

$$\mathbf{C}_1=[\mathbf{I}_{c-1}\vdots-\mathbf{1}_{c-1}],\ \mathbf{M}=[\mathbf{I}_{p-1}\vdots-\mathbf{1}_{p-1}]'.$$

Here $g=c-1$ and $u=p-1$.

Test of the parallelism of profiles of interactions

$$H:\eta_{ijh}-\eta_{ij,h+1}=const\ \ \forall h\ (i=1,\ldots,k\,;\,j=1,\ldots,c).$$

Matrix notation

$$H:\mathbf{C}\boldsymbol{\beta}\mathbf{M}=\mathbf{0},$$

$$\mathbf{C}=[\mathbf{0}_{(k-1)(c-1)}\vdots\mathbf{0}_{(k-1)(c-1)\times(k+c)}\vdots\mathbf{I}_{k-1}\otimes\mathbf{C}_1\vdots-\mathbf{1}_{(k-1)}\otimes\mathbf{C}_1]$$

with

$$\mathbf{C}_1=[\mathbf{I}_{c-1}\vdots-\mathbf{1}_{c-1}],\ \mathbf{M}=[\mathbf{I}_{p-1}\vdots-\mathbf{1}_{p-1}]'.$$

Here $g=(k-1)(c-1)$ and $u=p-1$.

Test of the equality of levels of all treatments

$$H:\sum_{h=1}^{p}\mu_{ijh}=const \quad (i=1,\dots,k;\ j=1,\dots,c).$$

Matrix notation

$$H:\mathbf{C}\boldsymbol{\beta}\mathbf{M}=\mathbf{0},$$

$$\mathbf{C}=\mathbf{C}_1\left[\mathbf{1}_{kc}\vdots\mathbf{1}_c\otimes\mathbf{I}_k\vdots\mathbf{I}_c\times\mathbf{1}_k\vdots\mathbf{I}_{kc}\right]$$

with

$$\mathbf{C}_1=\left[\mathbf{I}_{kc-1}\vdots-\mathbf{1}_{kc-1}\right],\ \mathbf{M}=\mathbf{1}_p\ .$$

Here $g=kc-1$ and $u=1$.

Test of the equality of levels of the first treatment

$$H:\sum_{h=1}^{p}\alpha_{ih}=const \quad (i=1,\dots,k).$$

Matrix notation

$$H:\mathbf{C}\boldsymbol{\beta}\mathbf{M}=\mathbf{0},$$

$$\mathbf{C}=[\mathbf{0}_{k-1}\vdots\mathbf{C}_1\vdots\mathbf{0}_{(k-1)\times(k+1)c}]$$

with

$$\mathbf{C}_1=\left[\mathbf{I}_{k-1}\vdots-\mathbf{1}_{k-1}\right],\ \mathbf{M}=\mathbf{1}_p\,.$$

Here $g=k-1$ and $u=1$.

Test of the equality of levels of the second treatment

$$H:\sum_{h=1}^{p}\tau_{jh}=const \quad (j=1,\dots,c).$$

Matrix notation

$$H : \mathbf{C}\boldsymbol{\beta}\mathbf{M} = \mathbf{0},$$

$$\mathbf{C} = [\mathbf{0}_{c-1} \vdots \mathbf{0}_{(c-1)\times k} \vdots \mathbf{C}_1 \vdots \mathbf{0}_{(c-1)\times kc}]$$

with

$$\mathbf{C}_1 = [\mathbf{I}_{c-1} \vdots -\mathbf{1}_{c-1}],\ \mathbf{M} = \mathbf{1}_p .$$

Here $g = c-1$ and $u = 1$.

Test of the equality of levels of interactions

$$H : \sum_{h=1}^{p} \eta_{ijh} = const \quad (i = 1,\ldots,k\ ;\ j = 1,\ldots,c).$$

Matrix notation

$$H : \mathbf{C}\boldsymbol{\beta}\mathbf{M} = \mathbf{0},$$

$$\mathbf{C} = [\mathbf{0}_{(k-1)(c-1)} \vdots \mathbf{0}_{(k-1)(c-1)\times(k+c)} \vdots \mathbf{I}_{k-1} \otimes \mathbf{C}_1 \vdots -\mathbf{1}_{(k-1)} \otimes \mathbf{C}_1]$$

with

$$\mathbf{C}_1 = [\mathbf{I}_{c-1} \vdots -\mathbf{1}_{c-1}],\ \mathbf{M} = \mathbf{1}_p .$$

Here $g = (k-1)(c-1)$ and $u = 1$.

Test of the equality of response means of all treatments

$$H : \sum_{i=1}^{k}\sum_{j=1}^{c} \mu_{ijh} = const \ \ \forall h \ .$$

Matrix notation

$$H : \mathbf{C}\boldsymbol{\beta}\mathbf{M} = \mathbf{0},$$

$$\mathbf{C} = [\mathbf{1}'_{1+k+c+kc}],\ \mathbf{M} = [\mathbf{I}_{p-1} \vdots -\mathbf{1}_{p-1}]'.$$

Here $g = 1$ and $u = p-1$.

Test of the equality of response effect of the *i*-th level of the first treatment

$$H : \alpha_{ih} = const \quad \forall h.$$

Matrix notation

$$H : \mathbf{C}\boldsymbol{\beta}\mathbf{M} = \mathbf{0},$$

$$\mathbf{C} = [0 \vdots \mathbf{e}'_{ki} \vdots \mathbf{0}'_{c} \vdots \mathbf{0}'_{kc}]$$

with

$$\mathbf{e}_{ki} = [e_{1i},\ldots,e_{ki}]',\ e_{mi} = \begin{cases} 1 & \text{if } m = i \\ 0 & \text{if } m \neq i \end{cases},\quad \mathbf{M} = [\mathbf{I}_{p-1} \vdots -\mathbf{1}_{p-1}]'.$$

Here $g = 1$ and $u = p-1$.

Similarly for levels of the second treatment and interactions.

4. Concluding remarks

In the paper there are given exact formulae for testing hypothesis in a two-way profile analysis. Using given formulae it is quite easy to do analysis numerically with the aid of the standard mathematical packages as Mathematica, Mathlab and other. Unfortunately, any of known to the author statistical packages do not preform profile analysis.

References

Anderson T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd. ed. John Wiley & Sons.

Cole J.W. L., Grizzle J. E. (1966). Applications of multivariate analysis of variance to repeated measurements experiments. *Biometrics* 22, 810–828.

Harris R. J. (1975). *A Primer of Multivariate Statistics*. Academic Press, Inc.

Johnson R. A., Wichern D. W. (1992). *Applied Multivariate Statistical Analysis*. 3rd. ed. Prentice-Hall, Inc.

Khatri C. G. (1966). A note on a MANOVA model applied to problems in growth curve. *Annals of the Inst. of Statistical Mathem*. 18, 75–86.

Morrison D. F. (1976). *Multivariate Statistical Methods*. 2nd. ed. McGraw-Hill (polish translation PWN 1990).

Potthoff R. F., Roy S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 313–325.

Rao C. R. (1965). *Linear Statistical Inference and Its Applications*. John Wiley & Sons. (polish translation PWN 1982).

Appendix

Multivariate linear model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\mathbf{Y}_{N\times p}$ – matrix of observations, $\mathbf{X}_{N\times q}$ – matrix of experiment, $\boldsymbol{\beta}_{q\times p}$ – matrix of unknown coefficients, $\boldsymbol{\varepsilon}_{N\times p}$ – matrix of random errors. It is assumed that $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1 \vdots \cdots \vdots \boldsymbol{\varepsilon}_N]'$, where $\boldsymbol{\varepsilon}_i$ is a p-vector distributed as $N_p(\mathbf{0}_p, \boldsymbol{\Sigma})$ for $i = 1,\ldots,N$, $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\varepsilon}_j$ are independent for $i \neq j$.

Test for hypothesis

$$H_0 : \mathbf{C}_{g\times q}\boldsymbol{\beta}_{q\times p}\mathbf{M}_{p\times u} = \mathbf{0} \tag{A.1}$$

$r = rank(\mathbf{X})$, $g = rank(\mathbf{C})$, $u = rank(\mathbf{M})$, $s = \min(g, r)$ is based on the statistic (see for example Morrison 1976)

$$F = \frac{w}{t} tr\mathbf{H}\mathbf{E}^{-1}, \tag{A.2}$$

where $t = |g - u|$, $w = N - r - u$ and

$$\mathbf{H} = \mathbf{M}'\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'\left(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'\right)^{-}\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}\mathbf{M} \tag{A.3}$$

$$\mathbf{E} = \mathbf{M}'\mathbf{Y}'\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\right)\mathbf{Y}\mathbf{M}\,. \tag{A.4}$$

Hypothesis H_0 is not rejected at α level if $F < F(\alpha; t, w)$, $F(\alpha; t, w)$ being a critical value of the F distribution with (t, w) degrees of freedom.

Useful matrix in the two-way model:

$$\mathbf{X}^{-} = \frac{1}{(k+1)(c+1)m}[\mathbf{1}_N \vdots \mathbf{B}_1 \vdots \mathbf{B}_2 \vdots \mathbf{B}_3]',$$

where

$$\mathbf{B}_1 = \mathbf{1}_c \otimes \left((k+1)\mathbf{I}_k - \mathbf{1}_k\mathbf{1}_k'\right) \otimes \mathbf{1}_m,$$
$$\mathbf{B}_2 = \left((c+1)\mathbf{I}_k - \mathbf{1}_c\mathbf{1}_c'\right) \otimes \mathbf{1}_{km},$$
$$\mathbf{B}_3 = \left((c+1)\mathbf{I}_k - \mathbf{1}_c\mathbf{1}_c'\right) \otimes \left((k+1)\mathbf{I}_k - \mathbf{1}_k\mathbf{1}_k'\right) \otimes \mathbf{1}_m\,.$$

Of course $(\mathbf{X}'\mathbf{X})^{-} = \mathbf{X}^{-}(\mathbf{X}^{-})'$. Now, matrix $\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ as well as other matrices may be easy obtained.

DWUCZYNNIKOWA ANALIZA PROFILOWA Z INTERAKCJAMI

Streszczenie

W pracy rozważany jest model dwuczynnikowej z interakcjami analizy profilowej. Podane są testy hipotez weryfikowanych w takim modelu.

Słowa kluczowe: analiza profilowa, wielowymiarowa analiza statystyczna, doświadczenie dwu-czynnikowe

Klasyfikacja AMS 2000: 62H15, 62J10