Chemometrics and Intelligent Laboratory Systems, 9 (1990) 127–141
Elsevier Science Publishers B.V., Amsterdam

Multivariate Analysis of Variance (MANOVA)

LARS STÅHLE *

Department of Pharmacology, Karolinska Institute, Box 60 400, S-104 01 Stockholm (Sweden)

SVANTE WOLD

Research Group for Chemometrics, Department of Organic Chemistry, Umeå University, S-901 87 Umeå (Sweden)

(Received 9 October 1989; accepted 4 June 1990)

CONTENTS

Abstract 128
1 Introduction 128
1.1 Formulation of hypotheses, experimental design 129
2 Notation and organization of data 129
3 The one-factor MANOVA 131
3.1 An intuitive geometrical approach 131
3.2 An example using the geometrical approach 131
3.3 Covariance matrices 132
3.4 The mathematical model 132
3.5 Test statistics 132
3.6 Interpretation and further analysis 133
3.7 Assumptions, properties and limitations 133
4 Crossed two-factor MANOVA 133
4.1 The mathematical model 134
4.2 Tests for interaction and factors 134
4.3 Assumptions, properties and limitations 134
5 Classification 134
5.1 Discriminant analysis 135
5.2 SIMCA and K nearest neighbours 135
6 Partial least squares analysis 136
6.1 Geometry and mathematics of PLS 136
6.2 Design of analysis 137
6.3 Test statistics 137
6.4 Properties and limitations 137
7 Discussion 138
8 Acknowledgements 138
Appendix A 139
Appendix B 140

0169-7439/90/$03.50 © 1990 – Elsevier Science Publishers B.V.

Appendix C 140
References 140

ABSTRACT

Ståhle, L. and Wold, S., 1990. Multivariate analysis of variance (MANOVA). *Chemometrics and Intelligent Laboratory Systems*, 9: 127–141.

In this tutorial we illustrate the practical use of multivariate analysis of variance (MANOVA). MANOVA concerns the situation where several response variables, e.g. the high-performance liquid chromatographic retention times of a number of compounds, have been measured in a set of experiments in which one or several factors (treatments) have been changed (e.g. solvent, stationary phase). The experiment is repeated a number of times for each combination of factors. MANOVA is then used to test whether the changes in the factors have any effect on the response variables.

The mathematical models underlying one-factor MANOVA and crossed two-factor MANOVA are discussed in some detail. Hypothesis tests based on generalization of the univariate F-test are discussed and compared. Follow up, using Hotelling's T^2-test, univariate ANOVA and discriminant variate analysis, is described. The assumptions of MANOVA are discussed in some detail. Alternative approaches are discussed, in particular partial least squares analysis (PLS) corresponding to MANOVA is put forward as a useful method for situations in which the assumptions of MANOVA are not fulfilled.

1 INTRODUCTION

In a previous tutorial in this journal we reviewed the analysis of variance (ANOVA) for the case were one response variable is measured and the effect of one or more factors on this variable is assessed [1]. A typical chemical example is a study of the effects of various catalysts on the yield of a chemical reaction. While the design of experiments and investigations discussed in that paper remains appropriate for a great deal of scientific activity in chemistry, it is not common that only one response variable is measured. One may, for instance, measure the yield as well as the amount of a carcinogenic by-product. Under such circumstances, multivariate methods are called for.

The use of ANOVA and multivariate analysis of variance (MANOVA) is to perform a number of experiments for each treatment (factor level), e.g. for each catalyst, and then compare the fit of two models: (I) a separate mean for each treatment, and (II) a global mean for all treatments (a pooled mean). If model I is significantly better than model II it is concluded that the treatment has an effect, i.e. that the choice of catalyst does indeed influence the yield of the main product (and/or the amount of by-product) [1].

Briefly, MANOVA is the multivariate counterpart of ANOVA under circumstances when several response variables have been investigated with respect to the factors. It is the purpose of this tutorial to provide an introduction to MANOVA and to share our experience with this methodology, its capacity and its limitations. More elaborate texts on the statistics and mathematics of MANOVA can be found in refs. 2–5. Familiarity with our ANOVA tutorial (or with ANOVA in general) has been assumed, to avoid the need for repetition. We will give formulae in two forms: as summation formulae and in terms of matrix algebra. The former assumes only a limited mathematical background on the part of the reader but the formulae become somewhat lengthy. Matrix notation is compact and easy to handle but assumes a familiarity with linear algebra, an introduction to which can be found in refs. 6 and 7. Some aspects of MANOVA can only be treated by means of linear algebra.

There has been some recent progress in the field of multivariate analysis of data of the MANOVA type. Since the authors are involved in the development of partial least squares (PLS)

TABLE 1

Simulation data for the effect of factory outlet on the concentration of chlorophenol and PCB

Three random samples were taken from each of the two factories and the control site.

	Sample	Chlorophenol	PCB
Factory 1 (A)	1	1.10	0.28
	2	1.12	0.28
	3	1.13	0.31
Control site (B)	1	1.12	0.17
	2	1.13	0.15
	3	1.14	0.19
Factory 2 (C)	1	1.20	0.27
	2	1.22	0.29
	3	1.23	0.32

analysis for this purpose, we will also discuss this method [8–10].

Two examples will be used to illustrate the use of MANOVA and PLS. The first example is a simulated set of environmental pollution data in which sediment samples were taken close to two factories and from one control site. The concentration of chlorophenol and PBC were measured (Table 1). The objects were randomly sampled sediments, three for each site. In the second example, which has toxicological background, the influence of dithiocarbamates on the toxicity of lead was investigated using a so-called crossed design (Table 2). Here, the number of variables is close to the number of objects. This example was chosen to illustrate some of the limitations of MANOVA, in which case PLS may offer a solution to the problem.

1.1 Formulation of hypotheses, experimental design

When confronted with the literature on MANOVA, one is struck by the multitude of approaches that can be taken [2–5]. Much discussion is centered around the problem of analyzing and interpreting a significant MANOVA (see Sections 3.6 and 4.3). Our standpoint is that much (or perhaps all) of the confusion can be avoided if (a) we distinguish between model and analysis, and (b) the researcher decides in advance what scientific hypotheses should be tested. Given that sufficient time has been spent on planning and design of a research project it is possible to avoid post hoc hypotheses formulation (regarding the project). Hence, we strongly recommend texts such as that of Box et al. [11].

Usually, the initial models used in ANOVA and MANOVA are linear and additive. The first hypothesis tested is that of no effect of the treatment (null hypothesis), i.e. that all the runs essentially give the same resulting values of the response variables.

2 NOTATION AND ORGANIZATION OF DATA

The data may be regarded as forming a table in which each row corresponds to an object and each column corresponds to a measured variable. Thus, in example 1 (environmental data) the rows (objects) correspond to sediment samples and the two columns to the concentration of chlorophenol and PCB respectively. This table (matrix) is denoted **X** with the elements x_{im}. The indices l and m ranging over l, $m = 1 \ldots p$ will be used to indicate variables. Since ANOVA and MANOVA both involve a subdivision of the objects into groups (depending on their 'treatment', e.g. factory or control site) the index i for objects is split into two or more indices. In the one-factor MANOVA we use indices i (object within a group, the sediment samples from a given site) and j (group, e.g. site). In the crossed two-factor classification i, j and k are used where j and k index the two factors. Index i is in the range $i = 1 \ldots n_j$ (or n_{jk} in the two-factor case, etc.). The total number of objects is

$$N = \sum_{j=1}^{J} n_j \tag{1}$$

where J is the number of groups in the one-factor classification. In the two-factor case n_{jk} is summed over $j = 1 \ldots J$ and $k = 1 \ldots K$ etc.

Because of the linear additive models used in MANOVA, various averages (mean values) play a central role. We use the dot notation [1] to denote means, e.g.

$$x_{\cdot jm} = \sum_{i=1}^{n_j} x_{ijm}/n_j \tag{2}$$

TABLE 2

Raw data for the effect of lead, disulfiram or combined lead + disulfiram treatment compared to control animals (rats)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
Control group													
1013	31	42	56	73	26	15	13	24	901	250	3661	438	1841
1007	29	51	74	82	63	62	21	48	900	161	5102	410	3366
1417	29	38	75	107	10	13	23	35	769	164	3235	953	1799
1841	41	57	121	114	13	15	24	39	1136	147	6412	1086	3618
Disulfiram													
998	38	48	61	52	44	30	12	25	1198	101	2967	1524	1874
1604	43	81	74	65	54	33	15	26	874	144	3292	288	1659
765	21	43	43	51	91	134	7	20	558	199	6351	890	2148
1494	34	68	135	103	24	17	21	37	918	248	6484	1158	3458
Lead group													
1656	23	54	80	72	61	25	10	17	906	425	4869	945	3051
1521	35	50	118	148	24	13	11	18	813	419	4016	635	2049
1722	41	49	92	147	12	10	14	22	1050	405	3345	963	2704
2028	39	62	96	104	15	8	29	47	729	564	3619	1113	3184
Disulfiram + lead group													
1105	37	53	104	88	74	52	14	21	851	377	6110	2577	3683
2052	44	80	156	106	67	31	13	22	864	356	6875	1539	6965
1607	37	50	145	133	66	52	24	40	918	394	6070	2317	4054
1296	30	45	121	112	68	33	11	21	770	514	5719	2340	2947

We also have the following important means (see Appendix A for computational details): in (2) $x_{\cdot jm}$ is the mean of the jth group for the mth variable. The total mean for the mth variable is denoted $x_{\cdot\cdot m}$ and the factor mean is $x_{\cdot j\cdot m}$ in the two-factor MANOVA.

The sample estimates of variance and covariance will be denoted var(m) and cov(l, m) (where cov(l, m) = cov(m, l) and var(m) = cov(m, m)). Computational details are given in Appendix A.

Standard matrix notation is used, with matrices symbolised by capital boldface (e.g. **X** for the data matrix). Unless otherwise specified, vectors are column vectors denoted by lower-case italic boldface letters (e.g. the vector of group means $\boldsymbol{a}$ for a given variable). The transpose is denoted by a prime, e.g. $\boldsymbol{a}'$ (which is a row vector). The inverse of a matrix **W** is denoted $\mathbf{W}^{-1}$. The eigenvalues of a square matrix are denoted l_1, $l_2 \ldots l_p$, ranged in order of magnitude, the largest being l_1.

3 THE ONE-FACTOR MANOVA

3.1 An intuitive geometrical approach

To understand the idea behind MANOVA the following intuitive picture of the data may be useful. Let there be three groups of objects ($J = 3$, sites) and assume that two variables (concentrations of chlorophenol and PCB) are recorded on each object ($p = 2$). Disregard the number of objects in each group and instead think of each group as a data scatter within an ellipse (i.e. a sample from a bivariate normal distribution indicated by a confidence interval). Depending upon how much overlap there is between the ellipses (groups) it is more or less likely that they really differ (Fig. 1). If the distances between the mean points (centroids) are large compared to the variation within the groups (also taking the orientation of the ellipse into account) there is a good reason to believe that there is a true difference between some of the groups. Thus, the null hypothesis of equal treatment effects is rejected. In order to make probabilistic statements of this kind more precise (reject the null hypothesis at a certain level of probability), we need to formalize the shape

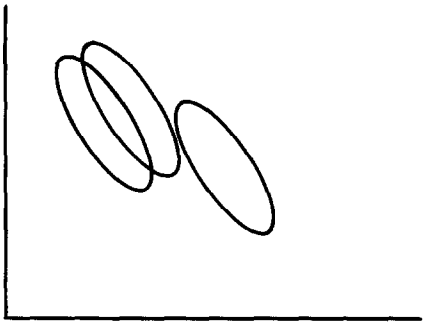


Fig. 1. Bivariate scatters (ellipses) from three groups of objects.

and size of the dispersion ellipse, the distances between controids and the relation between the two. It should be noted that in Fig. 1 all ellipses have the same orientation and are of equal size. This illustrates one assumption of MANOVA; that of equal dispersion (size and shape) within the groups.

3.2 An example using the geometrical approach

As in one-factor ANOVA [1], the way to construct a statistical test of the null hypothesis, that all groups are drawn from a population with the same centroid, is to compare the within-group variation with the between-groups variation. In fact, we shall base the test statistics for MANOVA given in Section 3.5 on the same kind of ratio between the between-groups dispersion and the within-groups dispersion as in ANOVA. Using the same type of illustration as above, analysis of the data of example 1 can be represented geometrically as comparing the within-group size of disper-

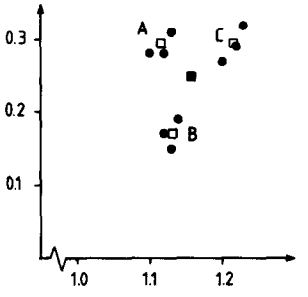


Fig. 2. Bivariate scatter plot of the data in Table 1. The individual points are closed circles and the mean point within each group (open) and the total mean (closed) are indicated as squares.

sion ('mean' size of the dispersion ellipses) with the between-groups size of dispersion (Fig. 2). An impression of the latter can be obtained by the dispersion of the group centroids around the total centroid. Hence, what is needed are multivariate measurements of dispersion.

3.3 Covariance matrices

In MANOVA the variance of each variable is not a sufficient measure of the variation. The possibility of covariation must be taken into account. The p variances and $p(p-1)/2$ covariances within (**W**) and between (**B**) groups are calculated as shown in Appendix A. As in ANOVA it is easy to show that the total sum of squares and cross-products is decomposed as

$$\mathrm{SSQ}CP_T(l,m) = \mathrm{SSQ}CP_W(l,m) + \mathrm{SSQ}CP_B(l,m) \quad (3)$$

These square and symmetrical matrices of size $(p \times p)$ are denoted as **W** with elements $\mathrm{SSQ}CP_W(l,m)$ and **B** with elements $\mathrm{SSQ}CP_B(l,m)$. Hence, the matrix containing the total (**T**) sums of squares and cross-products is

$$\mathbf{T} = \mathbf{W} + \mathbf{B} \quad (4)$$

The matrices for example 1 are

$$\mathbf{B} = \begin{vmatrix} 0.018 & 0.009 \\ 0.009 & 0.030 \end{vmatrix} \qquad \mathbf{W} = \begin{vmatrix} 0.001 & 0.001 \\ 0.001 & 0.003 \end{vmatrix}$$

3.4 The mathematical model

The ith observation in the jth group on the mth variable will be modelled additively in the same way as in ANOVA

$$x_{ijm} = \mu_m + \alpha_{jm} + e_{ijm} \quad (5)$$

where μ_m is the grand mean of the mth variable, α_{jm} is the effect of the jth treatment on the mth variable and e_{ijm} is the error term. This error term is assumed to have a multinormal distribution ($\mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ i.e. its expected value is 0 for each variable (**0**) and the dispersion around **0** is determined by the covariance matrix $\boldsymbol{\Sigma}$. In matrix notation the model is

$$\mathbf{x}'_{ij} = \boldsymbol{\mu}' + \boldsymbol{\alpha}'_j + \mathbf{e}'_{ij} \quad (6)$$

3.5 Test statistics

As in ANOVA, the null hypothesis of MANOVA is that there is no treatment effect, i.e. $\alpha_{jm} = 0$ for all j and m (in matrix notation $\boldsymbol{\alpha}_j = \mathbf{0}$). In analogy with ANOVA a ratio is formed between the between-group dispersion and the within-group dispersion. However, in MANOVA the dispersion appears in matrix form. We thus define this ratio as

$$\mathbf{R} = \mathbf{B}\mathbf{W}^{-1} \quad (7)$$

To provide an overall test of significance some function of $\mathbf{BW}^{-1}$ must be taken. Four functions, all based on the eigenvalues of this matrix, are quite frequently employed [5]: Wilk's lambda L, the Pillai–Bartlett trace P, the greatest characteristic root statistic of Roy R, and the Hotelling–Lawley trace H.

A practical problem for the user of MANOVA is that these four test statistics do not always agree. In fact, their power is different under various conditions [5,12,13]. When differences between groups are concentrated along a single dimension (e.g. along one response variable) their order of power is $R > H > L > P$ while group differences which have a diffuse spread are most powerfully detected in the order $P > L > H > R$. Departures from the assumptions of equal covariance matrices (see section 3.7) also affect the four test statistics differently. With respect to type I errors (i.e. false positives) P is apparently the most robust [12]. Transformations are available to convert L and P into F distributed test statistics (see Appendix A).

We illustrate the use of the test statistics by example 1 using the matrices

$$\mathbf{W}^{-1} = \begin{vmatrix} 2143 & -1071 \\ -1071 & 911 \end{vmatrix}$$

$$\mathbf{B}\mathbf{W}^{-1} = \begin{vmatrix} 27.81 & -10.37 \\ -11.55 & 16.88 \end{vmatrix}$$

to calculate the four test statistics:

$L = 0.0025 \quad (F_{4,10} = 47.2)$
$P = 1.88 \quad (F_{4,12} = 47.8)$
$H = 44.69$
$R = 34.58$

The fact that the two F transformations differ slightly illustrates that the test statistics do have somewhat different properties.

3.6 Interpretation and further analysis

Faced with a significant MANOVA one usually wishes to analyse subhypotheses, which may be the consequences of the way the study has been designed. In most studies particular pairwise comparisons between the groups will be investigated. Hotelling's T^2 test is appropriate for a multivariate pairwise test. The pooled within-group dispersion can be used as an estimate of the variance–covariance matrix **S**

$$\mathbf{S} = \mathbf{W}/(N-J) \tag{8}$$

A T^2 test between the first and the second groups is formulated as

$$T^2 = (n_1 n_2)(\boldsymbol{x}_{\cdot 1} - \boldsymbol{x}_{\cdot 2})' \mathbf{S}^{-1} (\boldsymbol{x}_{\cdot 1} - \boldsymbol{x}_{\cdot 2}) / (n_1 + n_2) \tag{9}$$

The T^2 statistic can be transformed to an F distributed variate

$$F = (n_1 + n_2 - p - 1) T^2 / [(n_1 + n_2 - 2) p] \tag{10}$$

Since there is at present no straight forward and easily available method corresponding to univariate multiple comparison procedures (see ref. 1), the easiest way to check for the risk of making type I errors is to divide the α level by the number of comparisons (Bonferroni procedure [5]). For example, with five groups, one of which is a control group, four pair wise comparisons give an α level of $0.05/4 = 0.0125$. We note that the power of this method declines with the number of comparisons. Further analysis within the pairwise comparison can be made by constructing confidence intervals for each variable based on the T^2 statistic [2,4].

The eigenvectors of $\mathbf{BW}^{-1}$ can also be used to plot the data along so-called discriminant (or canonical) variates. The first discriminant variate is the linear combination of the measured variables that best separates the groups. The second discriminant variate is the linear combination that best separates the groups in a direction orthogonal to the first discriminant variate. A hypothetical example is shown in Fig. 3. This topic is further discussed in Section 5.

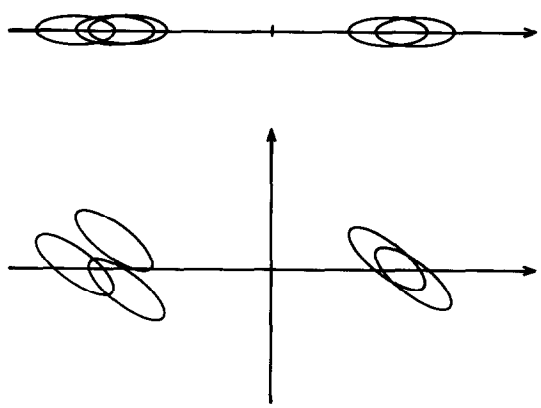


Fig. 3. The first two discriminant variates plotted for hypothetical data. The upper diagram uses only the first discriminant variate and seems to discriminate between two clusters of groups. The lower diagram uses to discriminant variates and shows further separation between the groups.

3.7 Assumptions, properties and limitations

MANOVA rests, in principle, on the assumptions that the objects are independent and that the covariance matrix **W** for the residuals is the same for all groups. The latter corresponds to the assumption of homoscedasticity for ANOVA. The distributions of the test statistics given in tables are all based on a multinormal $\mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ distribution of the residual covariance matrix. A mathematical requirement is that **W** is invertible. If all these requirements are fulfilled, and if the number of objects considerably exceeds the number of variables, the method apparently works well.

The power of the four test statistics is not only influenced by the way groups differ (see Section 3.5), but also by departures from the abovementioned assumptions. Inequality of the covariance matrices may seriously affect the power of all test statistics, although the Pillai–Bartlett trace is claimed to be less sensitive [12].

4 CROSSED TWO-FACTOR MANOVA

The crossed MANOVA is used to analyze designs in which two different kinds of treatments

are given, such as sampling site and season (winter/summer) in an environmental analysis problem. These two factors can be varied independently and all combinations of sites and seasons are possible (at least in principle). As in two-factor ANOVA there is the possibility that the two factors interact [1] i.e. that a particular combination of site and season can produce a special effect on the concentrations of the analytes of interest.

The avoid difficulties we assume in the following that the number of objects is exactly the same in all treatment groups ($n_{jk} = n$ for all $j = \ldots J$ and $k = 1 \ldots K$). Tests for so-called unbalanced designs do exist, however.

4.1 The mathematical model

As in the crossed two-factor univariate ANOVA there are two 'competing' models, one with an interaction term (δ) and one without interaction, containing only additive factor effects (α and β).

$$x'_{ijk} = \mu' + \alpha'_j + \beta'_k + e'_{ijk} \tag{11}$$

$$x'_{ijk} = \mu' + \alpha'_j + \beta'_k + \delta'_{jk} + e'_{ijk} \tag{12}$$

The choice between the two models is made by means of a hypothesis test of the significance of the interaction term.

4.2 Tests for interaction and factors

In much the same way as for the one-factor lay-out, the matrices **W** and $\mathbf{W}^{-1}$ are calculated. Covariance matrices corresponding to **B** in the one-factor MANOVA are calculated. They are the matrix of the first factor (**A**), the matrix of the second factor (**B**) and the matrix of the interaction between the factors (**D**).

Test statistics are calculated in the same way as for the one-factor MANOVA from the matrices $\mathbf{AW}^{-1}$, $\mathbf{BW}^{-1}$ and $\mathbf{DW}^{-1}$.

We illustrate this by the toxicity data in Table 2. For simplicity four variables have been chosen: x_1, x_2, x_4 and x_{11}. Calculations of the matrices are shown in Appendix B. From these we calculate the test statistics

interaction: $L = 0.6739$ ($F_{4,9} = 1.09$), $R = 0.484$

disulfiram: $L = 0.5550$ ($F_{4,9} = 1.80$), $R = 0.802$

lead: $L = 0.1411$ ($F_{4,9} = 13.70$), $R = 6.090$

We note how the degrees of freedom from the crossed two-factor design are transformed into the F approximation of Wilk's lambda L as shown in Appendix B.

4.3 Assumptions, properties and limitations

The same general assumptions are made for the two-way crossed design as for the one-factor MANOVA. In addition, hypothesis testing of the interaction term must preceed testing of main effects, just as in univariate ANOVA. However, it should be noted that while in ANOVA the presence of an interaction can be described as a difficulty for the continuation of the analysis (main effects), the situation is more severe in MANOVA. This is so simply for the reason that the interaction may involve some variables (or a combination of variables) while the main effects are seen in other variables. The probability of finding a significant interaction does, of course, increase with the number of variables, not least because of the fact that a larger span of the treatment effects are covered and, hence, the chances that nonlinear behaviour shows up are increased. In our example it turns out that there is a strong interaction between lead and disulfiram (see Section 5) but this is not detected by MANOVA. The reason for this is that not all variables can be included in the MANOVA due to the fact that, with 14 variables, 16 objects and 4 treatment groups, the matrix **W** is not of full rank and cannot be inverted.

5 CLASSIFICATION

A subject closely related to MANOVA is that of classification and discriminant analysis. In the

statistical literature the most commonly discussed method is the so-called discriminant analysis, while, in chemometrics, the K nearest neighbours method and SIMCA (soft independent modeling of class analogy) [14] are often used. The main reason is that the latter two methods are applicable to sets of data with many variables and few objects.

5.1 Discriminant analysis

Discriminant analysis is, in effect, a combination of MANOVA with the discriminant variate plots described in Section 3.6. The first step in discriminant analysis is to test the hypothesis that the preconceived groups differ (significantly) with respect to the variables measured. Unless this can be stated with some degree of confidence, there is no point in persuing the classification process. There are two ways to continue the analysis (usually both ways are investigated): (1) to determine in which way the groups differ or (2) to make class models.

One can use the discriminant variate plots to examine in which way the groups differ from one another, and what groups differ. It is possible to formally test how many discriminant variates will significantly contribute to a description of the differences between the groups [4]. This can be visually understood by taking Fig. 3 as an example. Assume that three variables have been measured but that only two contribute to a description of the differences between the five groups. The formal procedure [4] will then tell us that only two discriminant variates are significant. It is then said that the dimensionality of the group differences is 2.

Class modelling can be performed in at least two ways. One is by calculating the covariance matrix for each group and forming a confidence region around the mean of 95%, for instance. The second is to exploit an assumption made in the MANOVA, that of equal covariance matrices, in order to pool the data to calculate the common covariance matrix which is then used to form the confidence region around the mean. The latter method is more efficient, and is indeed necessary whenever the number of objects in a group is small compared to the number of variables ($n_j < p + 2$). Another method commonly employed to avoid this problem is to use principal component analysis to reduce the number of variables by discarding the components with the smallest variance. This procedure will ensure that the mathematical procedure necessary to calculate the confidence region, inversion of the covariance matrix, will be numerically possible. This is further discussed in Section 6.

To test whether a new object belongs to any of the previously modelled groups, one simply measures the distance from the mean of the group to the new object and relates it to the confidence region. The distance thus obtained is the so-called Mahalanobis distances [4]. The formal procedure used to test for group membership is a chi-square test.

5.2 SIMCA and K nearest neighbours

Conceptually the simplest method is probably the K nearest neighbours (KNN) test in which a new object is classified on the basis of the distance to its neighbours in the measurement space. A necessary step in KNN is to normalize the variables so that the distance becomes a meaningful concept.

SIMCA is conceptually similar to the modelling procedure of discriminant analysis. However, instead of using the confidence regions of discriminant analysis, a principal component model is calculated for each preconcieved group. The tests used for group (class) separation in SIMCA are directly transferable to the MANOVA situation. One would then test the fit of all training set objects to a single PC model versus the fit to separate models for each group by means of an approximate F test

$$F = \frac{\sum_{i=1}^{N}\sum_{k=1}^{p} e_{ik}^2}{\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sum_{k=1}^{pw_j} e_{ijk}^2} \qquad (13)$$

The terms e_{ik} and e_{ijk} in eq. (13) are the residual errors using the single overall PC model in the

numerator and the groupwise calculated PC models in the denominator. A difficulty with SIMCA is the choice of the appropriate degrees of freedom to be used in the F test. This problem has not yet been quite satisfactorily solved, but at present

$$(N-A-1)(p-A)/2 \tag{14}$$

is used as the degrees of freedom for the numerator and

$$\sum_{j=1}^{J}(n_j-A_j-1)(p-A_j)/2 \tag{15}$$

is used for the denominator. The number of PC components in the PC models are A and A_j, respectively. An alternative is to use a test based on cross-validation, but the distribution of that test statistic remains to be studied.

When the number of objects (N) is large in relation to the number of variables (p) and the variables are independent, the inverse of the covariance matrix exists and linear discriminant analysis is often employed. As mentioned previously, the test for class separation becomes exactly that of one-way MANOVA. With increasing collinearity in the measured variables, the PLS version of discriminant analysis [9,10] can be used instead with the test statistics discussed in the next section.

6 PARTIAL LEAST SQUARES ANALYSIS

Partial least squares analysis (PLS) has emerged during the last decade as a distribution-free regression method designed to handle situations with collinearity in **W** and/or $p > N$ in cases when methods using the inverse of **W** are numerically (and statistically) unstable, or simple mathematically impossible, respectively. PLS has been reviewed in detail elsewhere [8,15–18] and therefore only points relevant to the MANOVA discussion will be introduced here.

Notice that we make a change in notation below. This is motivated by the fact that the standard notation used in MANOVA is different from the standard notation of PLS. To facilitate further studies by readers familiar with one method we considered this choice better than the use of one notation for both methods.

The integers J and K denote the number of variables in two matrices **X** and **Y**. The indices j and k are used correspondingly. The number of objects is denoted as before by n and index i for objects.

6.1 Geometry and mathematics of PLS

While MANOVA works under the assumption that the residual covariance matrix is invertible and, hence, that the elements of the covariance matrix can be estimated, this assumption is explicitly avoided in PLS. In PLS, one dimension is calculated at a time and its significance is assessed, thus keeping the problem of collinearity under control. Typical illustrations of one- and two-dimensional PLS models are given in Fig. 4. Geometrically, PLS dimensions can be said to resemble the discriminant variates discussed in Sections 3.6 and 5.1 in the sense that one dimension is calculated at a time.

Mathematically, PLS gives the solution to the problem of finding the linear combination for each of two blocks of variables which maximizes

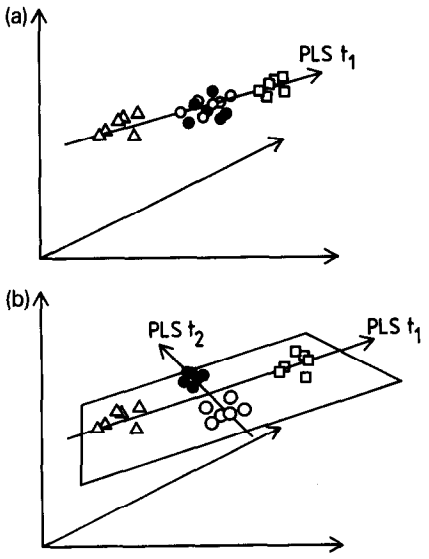


Fig. 4. Illustration of (a) a one-dimensional PLS model and (b) a two-dimensional PLS model. The measurement space is three-dimensional.

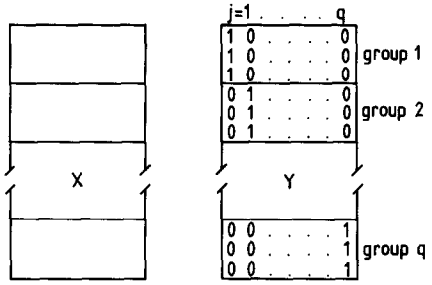


Fig. 5. Illustration of the matrices used for the PLS analogue of MANOVA.

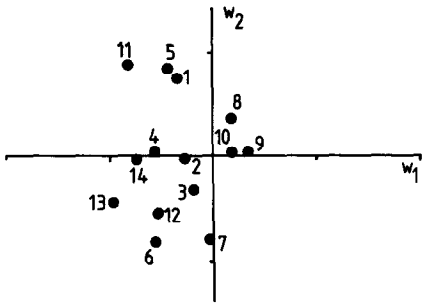


Fig. 7. PLS weight plot for the data in Table 2. The numbers refer to the variable numbers in Table 2.

the covariance between the two linear combinations (also called scores). The scores are calculated as shown in Appendix C in whole also the notation is explained.

6.2 Design of analysis

The simplest design of the PLS version of MANOVA is obtained with the observed data in **X** and a so-called design matrix in **Y**. The design matrix has as many columns (K) as there are treatment groups. Each column (variable) is a dummy variable of type $0-1$. Thus objects belonging to the kth group will get a 1 in the kth column and 0 in the others. The arrangement is illustrated in Fig. 5. In this way the design is balanced. The usual practice is to use the one-factor type of analysis. To illustrate the methodology we use the data in Table 2 (lead/disulfiram experiment). As can be seen in Fig. 6 there is a nice separation between the combined treatment group and the other groups in the two-dimensional score plot. The importance of the variables is shown in Fig. 7. Notice that directions in Fig. 6 and Fig. 7 correspond to one another.

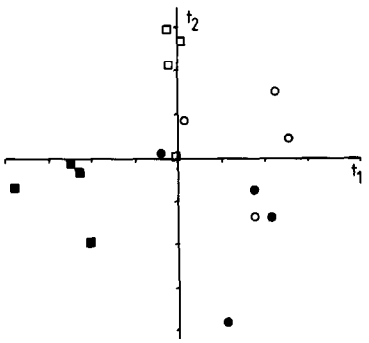


Fig. 6. PLS score-plot for the data in Table 2. Controls (open circles), disulfiram (closed circles), lead (open squares) and lead + disulfiram (closed squares).

6.3 Test statistics

Hypothesis testing with PLS is usually performed by means of cross validation [8,10]. This technique has been described in some detail elsewhere and it suffices to point out the following properties of cross validation. Cross validation simulates the predictive properties of the model by deleting part of the data, developing the model for the remaining data and then predict the ones deleted. This is repeated a number of times until each element has been deleted once and once only.

The test statistic calculated in cross-validation is the prediction error divided by the residual standard error (CVD/SD). Like any other test statistic, CVD/SD is a random variable and, as such, it has a probability distribution which depends upon the distribution of the residuals of the recorded variables. The distribution is not known as an analytic expressions but simulation studies have been performed [10] providing guidelines for probabilistic decision making.

6.4 Properties and limitations

PLS is a least squares method and not a maximum likelihood method. It is therefore nonparametric in the estimation of the model parameters (weights, loadings and scores). The hypothesis testing is, as always, based on the distribution of a

random variable and is therefore a function of the underlying distribution of the data. The cross validation test used here is rather insensitive to departures from normality (Ståhle, unpublished simulation data) and the distribution and 5% limits given in tables are calculated from simulations using the normal distribution [10,19]. Theory and experience show that PLS works well, regardless of the number of relevant variables. Like any method, a small number of objects will reduce the certainty of the conclusions.

Like all data analytical methods, PLS works best when the data are symmetrically distributed (as for MANOVA) transformations might be helpful, see ref. 1. Furthermore, PLS is scale sensitive. The usual practice is to normalise the data to zero mean and unit variance (this was done in the analysis of the data in Table 2, plotted in Fig. 6). Other scalings may be worthwhile, such as block scaling, which is used when there are blocks of variables. In each block the variables may be regarded as measuring the same characteristic and the whole block is therefore given a total unit variance. The usefulness of this approach is easy to see if e.g. molecules are characterised in various ways, for example UV absorption at different wavelengths. Whether 10 or 100 wavelengths are used will certainly influence the outcome of the PLS analysis since in the latter case they will account for a large proportion of the covariance in **X**.

The effects of heteroscedasticity have not been investigated. Moderate differences in the group size has been found not to influence the distribution of the cross-validation based test statistic in the two-class case [10].

7 DISCUSSION

When the assumptions for MANOVA are fulfilled and there are no interactions present, this technique works satisfactorily for testing equality or difference in the means between groups. For multivariate data, it can almost always be regarded as a better choice than the corresponding univariate techniques. This statement is not uncontroversial since it has been suggested that there are several situations in which the power of MANOVA is inferior to ANOVA of one variable at a time [20,21]. This, however, is accompanied by the risks of overlooking true differences in combinations of variables and an increased risk for type I errors (false positives). Thus, we advocate MANOVA over ANOVA for multivariate data, although the latter may be used as a descriptive complement to MANOVA.

When the assumptions are not fulfilled the situation is not as simple as with univariate ANOVA (which is fairly robust). Different number of objects or a large number of recorded variables ($p > N/4$) may hamper the performance of MANOVA. We illustrate this by running a MANOVA on the lead/disulfiram data from Table 2 using 12 variables (excluding variables 6 and 7) with the result that NONE of the tested effects is significant.

For such situations alternative methods should be used. We have found PLS to work satisfactorily and the PLS analogue to MANOVA has been run routinely for analyzing experimental data in our laboratories. In combination with cross validation, probabilistic statement can be made, and hence hypotheses can be tested. A direct comparison between MANOVA and PLS on simulated data has unfortunately not been published. Important questions regarding the relative power and sensitivity to distribution and configuration of the data should be addressed in such a study.

Finally it should be noted that hypothesis testing is only a small part of a data analysis. Choice of model and estimation of parameters and confidence regions are usually of greater importance. Compared to regression methods, the scope of MANOVA is limited, which explains why the former are more frequently employed. This should be born in mind when choosing statistical methods for the analysis of multivariate data.

8 ACKNOWLEDGEMENTS

The present work was supported by grants from the Swedish Natural Science Research Council, the Swedish Medical Research Council Grant No. 09069, the Karolinska Institute and the Swedish Physicians Association.

APPENDIX A

This appendix contains formulae for the means and variance–covariance matrices of prime importance, as well as formulae for the test statistics used in hypothesis testing, together with some F transforms.

The mean within the jth group for the mth variable is

$$x_{\cdot jm} = \sum_{i=1}^{n_j} x_{ijm}/n_j \tag{A1}$$

The total mean for the mth variable is

$$x_{\cdot\cdot m} = \sum_{j=1}^{J}\sum_{i=1}^{n_j} x_{ijm}/\sum_{j=1}^{J} n_j \tag{A2}$$

In the two-factor classification the factor means are calculated as

$$x_{\cdot j\cdot m} = \sum_{k=1}^{K}\sum_{i=1}^{n_{jk}} x_{ijkm}/\sum_{k=1}^{K} n_{jk} \tag{A3}$$

In eq. (A3) $x_{\cdot j\cdot m}$ is the mean of the jth treatment on the first factor measuring the mth variable.

The within (**W**) group variance–covariance is calculated as follows:

$$\mathrm{cov}_W(l,m) = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(x_{ijl}-x_{\cdot jl})(x_{ijm}-x_{\cdot jm}) \Big/ \sum_{j=1}^{J}(n_j-1) \tag{A4}$$

The between (**B**) group covariances are

$$\mathrm{cov}_B(l,m) = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(x_{\cdot jl}-x_{\cdot\cdot l})(x_{\cdot jm}-x_{\cdot\cdot m}) /(J-1) \tag{A5}$$

The sums of squares in ANOVA become, in MANOVA, the sums of squares and cross products and are

$$\mathrm{SSQ}CP_W(l,m) = \mathrm{cov}_W(l,m)\sum_{j=1}^{J}(n_j-1) \tag{A6}$$

$$\mathrm{SSQ}CP_B(l,m) = \mathrm{cov}_B(l,m)(J-1) \tag{A7}$$

The four test statistics in MANOVA are calculated as functions of the determinants or the eigenvectors of **W**. They are defined as follows:

$$L = \prod_{j=1}^{p} 1/(1+l_j) \tag{A8}$$

Originally L was defined as a ratio between two determinants:

$$L = |\mathbf{W}|/|\mathbf{T}| \tag{A9}$$

The Pillai–Bartlett trace P is defined as

$$P = \sum_{j=1}^{p} l_j/(1+l_j) \tag{A10}$$

The largest eigenvalue of $\mathbf{BW}^{-1}$ defines the greatest characteristic root statistic of Roy

$$R = l_1/(1+l_1) \tag{A11}$$

The 5% and 1% points of the distribution of R are given as charts in ref. 2. The Hotelling–Lawley trace is

$$H = \sum_{j=1}^{p} l_j \tag{A12}$$

There are transforms of some of the test statistics which are approximately (in some instances exactly) F distributed. For Wilk's lambda we have

$$F = (1-L^{1/s})[rs - p(J-1)/2 + 1] \Big/ [L^{1/s}p(J-1)] \tag{A13}$$

where

$$r = N - 1 - (p+J)/2 \tag{A14}$$

and

$$s = \sqrt{\left[\left(p^2(J-1)^2-4\right)/\left(p^2+(J-1)^2-5\right)\right]} \tag{A15}$$

This F-test is made with $p(J-1)$ and $rs - p(J-1)/2 + 1$ degrees of freedom (note that in ref. 5 there is a typographical error in the formula for the present eq. (A15)). The Pillai–Bartlett trace can be transformed by

$$F = (N-J-p+r)P/[s(r-P)] \tag{A16}$$

In eq. (A16) $r = \mathrm{rank}(\mathbf{BW}^{-1})$ (which is practice is $\min(J-1,\ p)$) and $s = \max(J-1,\ p)$. This F-test has rs and $r(N-J-p+r)$ degrees of freedom.

APPENDIX B

This appendix contains various formulae used in crossed MANOVA and some of the matrices calculated in the lead/disulfiram toxicity example.

For the first factor matrix (**A**) the elements are

$$\mathrm{SSQ}CP_A(l,m) = nK \sum_{j=1}^{J} (x_{\cdot j\cdot l} - x_{\cdots l}) \times (x_{\cdot j\cdot m} - x_{\cdots m}) \quad \text{(B1)}$$

For the second factor the matrix (**B**) elements are

$$\mathrm{SSQ}CP_B(l,m) = nJ \sum_{k=1}^{K} (x_{\cdot\cdot kl} - x_{\cdots l}) \times (x_{\cdot\cdot km} - x_{\cdots m}) \quad \text{(B2)}$$

The elements of the interactions (**D**) matrix are

$$\mathrm{SSQ}CP_D(l,m) = n \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{\cdot jkl} - x_{\cdot j\cdot l} - x_{\cdot\cdot kl} + x_{\cdots l}) \times (x_{\cdot jkm} - x_{\cdot j\cdot m} - x_{\cdot\cdot km} + x_{\cdots m}) \quad \text{(B3)}$$

For the lead/disulfiram example we get the following matrices

$$\mathbf{A} = \begin{vmatrix} 1.03\mathrm{E}{+5} & -1.28\mathrm{E}{+3} & -1.02\mathrm{E}{+4} & 1.62\mathrm{E}{+4} \\ -1.28\mathrm{E}{+3} & 1.60\mathrm{E}{+1} & 1.27\mathrm{E}{+2} & -2.02\mathrm{E}{+2} \\ -1.02\mathrm{E}{+4} & 1.27\mathrm{E}{+2} & 1.01\mathrm{E}{+3} & -1.60\mathrm{E}{+3} \\ 1.62\mathrm{E}{+4} & -2.02\mathrm{E}{+2} & -1.60\mathrm{E}{+3} & 2.55\mathrm{E}{+3} \end{vmatrix}$$

$$\mathbf{B} = \begin{vmatrix} 5.07\mathrm{E}{+5} & 3.56\mathrm{E}{+3} & 4.86\mathrm{E}{+4} & 3.63\mathrm{E}{+4} \\ 3.56\mathrm{E}{+3} & 2.50\mathrm{E}{+1} & 3.41\mathrm{E}{+2} & 2.55\mathrm{E}{+3} \\ 4.86\mathrm{E}{+4} & 3.41\mathrm{E}{+2} & 4.66\mathrm{E}{+3} & 3.48\mathrm{E}{+4} \\ 3.63\mathrm{E}{+4} & 2.55\mathrm{E}{+2} & 3.48\mathrm{E}{+4} & 2.60\mathrm{E}{+5} \end{vmatrix}$$

$$\mathbf{D} = \begin{vmatrix} 1.27\mathrm{E}{+4} & -1.13\mathrm{E}{+2} & -4.30\mathrm{E}{+3} & 3.99\mathrm{E}{+3} \\ -1.13\mathrm{E}{+2} & 1.00\mathrm{E}{+0} & 3.83\mathrm{E}{+1} & -3.55\mathrm{E}{+1} \\ -4.30\mathrm{E}{+3} & 3.83\mathrm{E}{+1} & 1.46\mathrm{E}{+3} & -1.36\mathrm{E}{+3} \\ 3.99\mathrm{E}{+3} & -3.55\mathrm{E}{+1} & -1.36\mathrm{E}{+3} & 1.26\mathrm{E}{+3} \end{vmatrix}$$

$$\mathbf{W} = \begin{vmatrix} 1.46\mathrm{E}{+6} & 2.15\mathrm{E}{+4} & 8.85\mathrm{E}{+4} & -1.78\mathrm{E}{+4} \\ 2.15\mathrm{E}{+4} & 6.58\mathrm{E}{+2} & 1.19\mathrm{E}{+3} & -1.76\mathrm{E}{+3} \\ 8.85\mathrm{E}{+4} & 1.19\mathrm{E}{+3} & 9.49\mathrm{E}{+3} & 3.10\mathrm{E}{+1} \\ -1.78\mathrm{E}{+4} & -1.76\mathrm{E}{+3} & 3.10\mathrm{E}{+1} & 5.06\mathrm{E}{+4} \end{vmatrix}$$

The residual degrees of freedom are

$$df_w = JK(n-1) \quad \text{(B4)}$$

while the hypothesis degrees of freedom are

$$df_a = J - 1 \quad \text{(B5)}$$

$$df_b = K - 1 \quad \text{(B6)}$$

$$df_d = (J-1)(K-1) \quad \text{(B7)}$$

The formulae (A13), (A14) and (A15) can thus be generalized (using the interaction as an example) to give

$$F = (1 - L^{1/s})(rs - pdf_d/2 + 1)/(L^{1/s} pdf_d) \quad \text{(B8)}$$

$$r = nJK - 1 - (p + df_d + 1)/2 \quad \text{(B9)}$$

$$s = \sqrt{[(p^2 df_d^2 - 4)/(p^2 + df_d^2 - 5)]} \quad \text{(B10)}$$

To calculate the F approximation for the main treatment effects df_a and df_b are substituted for df_d.

APPENDIX C

This appendix summarizes the most important steps in the PLS algorithm. Full descriptions of PLS from various aspects are given elsewhere [8–10,15–19].

The scores t_i (forming the vector $\boldsymbol{t}$) are calculated using weight coefficients, denoted $\boldsymbol{w}$ for the predictor block of variables (X) and $\boldsymbol{q}$ for the block of predicted variables (**Y**). The **X** score for the ith object is denoted t_i and is calculated as

$$t_i = \sum_{j=1}^{J} x_{ij} w_j \quad \text{(C1)}$$

Similarily the **Y** score u_i is calculated as

$$u_i = \sum_{k=1}^{K} y_{ik} q_k \quad \text{(C2)}$$

The regression coefficient between the score vectors $\boldsymbol{t}$ and $\boldsymbol{u}$ is denoted d. Another set of coefficients $\boldsymbol{p}$ are called loadings and are used to calculate residuals:

$$\mathbf{E} = \mathbf{X} - \boldsymbol{tp}' \quad \text{(C3)}$$

$$\mathbf{F} = \mathbf{Y} - d\boldsymbol{tq}' \quad \text{(C4)}$$

The residual matrices **E** and **F** are then substituted for **X** and **Y** in the calculation of the second PLS dimension.

REFERENCES

1 L. Ståhle and S. Wold, Analysis of variance (ANOVA), *Chemometrics and Intelligent Laboratory Systems*, 6 (1989) 259–272.

2 D.F. Morrison, *Multivariate Statistical Methods,* McGraw-Hill, New York, 1967.
3 W.W. Cooley and P.R. Lohnes, *Multivariate Data Analysis,* Wiley, New York, 1971.
4 K.V. Mardia, J.T. Kent and J.M. Bibby, *Multivariate Analysis,* Academic Press, London, 1979.
5 J.H. Bray and S.E. Maxwell, *Multivariate Analysis of Variance,* Sage University Press, Beverley Hills, CA, 1985.
6 G. Stephenson, *An Introduction to Matrices, Sets and Groups for Science Students*, Dover, New York, 1965.
7 H. Anton *Elementary Linear Algebra,* Wiley, New York, 1987.
8 S. Wold, A. Ruhe, H. Wold and W.J. Dunn III, The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, *SIAM Journal of Scientific Statistics and Computing,* 5 (1984) 735–743.
9 M. Sjöström, S. Wold and B. Söderström, PLS discriminant plots, in E.S. Gelsema and L.N. Kanal (Editors), *Pattern Recognition in Practice II,* Elsevier, Amsterdam, 1986, pp. 461–470.
10 L. Ståhle and S. Wold, Partial least squares analysis with cross-validation for the two-class problem: a Monte Carlo study, *Journal of Chemometrics*, 1 (1987) 185–196.
11 G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for Experimenters*, Wiley, New York, 1978.
12 C.H. Olson, On chosing a test-statistic in multivariate analysis of variance, *Psychology Bulletin*, 83 (1976) 579–586.
13 J. Stevens, Comment on Olson: Choosing a test statistic in multivariate analysis of variance, *Psychology Bulletin*, 86 (1979) 355–360.
14 C. Albano, W. Dunn III, U. Edlund, E. Johansson, B. Norden, M. Sjöström and S. Wold, Four levels of pattern recognition, *Analytica Chimica Acta*, 103 (1978) 429–442.
15 H. Wold, Soft modeling: the basic design and some extensions, in K.G. Jöreskog and H. Wold (Editors), *Systems under Indirect Observation*, North Holland, Amsterdam, 1982, pp. 1–54.
16 H. Martens, Multivariate calibration, *Thesis*, Technical University of Norway, Trondheim, 1985.
17 A. Lorber, L.E. Wangen and B.R. Kowalski, A theoretical foundation for the PLS algorithm, *Journal of Chemometrics*, 1 (1987) 19–31.
18 A. Höskuldsson, PLS regression methods, *Journal of Chemometrics*, 2 (1988) 211–220.
19 L. Ståhle and S. Wold, Multivariate data analysis and experimental design in biomedical research, in G.P. Ellis and G.B. West (Editors), *Progress in Medicinal Chemistry,* Vol. 25, Elsevier, Amsterdam, 1988, pp. 291–338.
20 T.J. Hummel and J.R. Sligo, Empirical comparison of univariate and multivariate analysis of variance procedures, *Psychology Bulletin,* 76 (1971) 49–57.
21 P.H. Ramsey, Empirical power of procedures for comparing two groups on p variables, *Journal of Educational Statistics,* 7 (1982) 139–156.