MULTIVARIATE STATISTICAL ANALYSIS: A GEOMETRIC PERSPECTIVE

Tyurin, Y. N.

February 3, 2009

Contents

Introduction

Linear statistical analysis, and the least squares method specifically, achieved their modern complete form in the language of linear algebra, that is in the language of geometry. In this article we will show that *multivariate* linear statistical analysis in the language of geometry can be stated just as beautifully and clearly. In order to do this, the standard methods of linear algebra must be expanded. The first part of this article introduces this generalization of linear algebra. The second part introduces the theory of multivariate statistical analysis in the first part's language. We believe that until now multivariate statistical analysis, though explained in dozens of textbooks, has not had adequate forms of expression.

Multivariate observations are the observations of several random quantities in one random experiment. We shall further record multivariate observations as columns. We commonly provide multivariate observations with indices. In the simple case, natural numbers serve as the indices. (This can be the numbers of observations in the order they were recorded). For independent evenly distributed observations this is a fitting way to organize information. If the distributions of observation depend on one or more factors, the values or combinations of values of these factors can serve as indices. Commonly the levels of factors are numbered. In that case the index is the set of numbers. So, in a two-factor scheme (classification by two traits) pairs of natural numbers serve as indices.

We shall call the set of observations, provided with indices and so organized, an *array*.

For theoretical analysis the linear numeration of data is most convenient. Further we will be holding to this system. When analyzing examples we will return, if needed, to the natural indexing of data.

In univariate statistical analysis the numeration of data allows recording as rows. In the multivariate case the entirety of the enumerated data (that is arrays) can also be examined as a row of columns. In many cases (but not always) such an array can be treated as a matrix.

Arrays of one form naturally form a vector space under the operation of addition and multiplication by numbers. For the purposes of statistical analysis this vector space is given a scalar product. In one dimensional analysis, if the observations are independent and have equivalent dispersions, then the most fitting scalar product is the euclidean product. In more detail: let the observations have an index α; let arrays $\mathbf{T}_X$ and $\mathbf{T}_Y$ be composed of the one-dimensional elements X_α, Y_α. Then the euclidean scalar product of arrays $\mathbf{T}_X$ and $\mathbf{T}_Y$ is

$$\langle \mathbf{T}_X,\ \mathbf{T}_Y \rangle = \sum_\alpha X_\alpha Y_\alpha, \tag{0.1}$$

where the index of summation goes through all possible values. We shall record multivariate observations as columns. In the multivariate case, the elements X_α, Y_α are columns. For arrays composed of columns, let us accept the following definition of the scalar product of arrays $\mathbf{T}_X$ and $\mathbf{T}_Y$:

$$\langle \mathbf{T}_X,\ \mathbf{T}_Y \rangle = \sum_\alpha X_\alpha Y_\alpha^T. \tag{0.2}$$

The scalar product (0.2) is a square matrix. Therefore, for arrays composed of columns, square matrices of the corresponding dimensions must play the role of scalars. With the help of the scalar product (0.2) and its consequences, this article develops a theory of multivariate statistical analysis, parallel to existing well-known univariate theory.

1 Modules of Arrays Over a Ring of Matrices

1.1 Space of Arrays

In the introduction we agreed to hold to a linear order of indexation for simplicity's sake. However, all the introduced theorems need only trivial changes to apply to arrays with a different indexation.

Let us consider a p-dimensional array with n elements,

$$\mathbf{T} := \{X_i \mid i = \overline{1, n}\}, \tag{1.1}$$

where $X_1, X_2, \ldots, X_n$ are p-dimensional vector-columns. Arrays of this nature form a linear space with addition and multiplication by numbers.

1. Addition:

$$\{X_i \mid i = \overline{1, n}\} + \{Y_i \mid i = \overline{1, n}\} = \{X_i + Y_i \mid i = \overline{1, n}\}.$$

2. Multiplication by numbers: let λ be a number; then

$$\lambda\{X_i \mid i = \overline{1, n}\} = \{\lambda X_i \mid i = \overline{1, n}\}.$$

In addition, we will be examining the element-by-element multiplication of arrays by square matrices of the appropriate dimensions.

3. Left Multiplication by a Matrix: let K be a square matrix of dimensions $p \times p$. Suppose

$$K\{X_i \mid i = \overline{1, n}\} = \{K X_i \mid i = \overline{1, n}\}. \tag{1.2}$$

Note that the multiplication of an array by a number can be examined as a special case of multiplication by a square matrix. Specifically: multiplication by the number λ is multiplication by the matrix λI, where I is the identity matrix of dimensions $p \times p$.

4. Right Multiplication by matrices: let $Q = \|q_{ij} \mid i = \overline{1,n}, j = \overline{1,n}\|$ — a square n by n matrix. Let us define the right multiplication of array $\mathbf{T}$ (1.1) by matrix Q as

$$\{X_i \mid i = \overline{1,n}\}Q = \{\sum_{j=1}^{n} X_j q_{ij} \mid i = \overline{1,n}\}. \tag{1.3}$$

It is clear that the product $\mathbf{T}Q$ is defined by the common matrix multiplication method of "row by column" with the difference that elements of a row of $\mathbf{T}$ (array $\mathbf{T}$) are not numbers but columns $X_1, \ldots, X_n$.

5. Let us define the **inner product** in array space. For it's properties we shall call it the scalar product (or, generalized scalar product). In more detail: let

$$\mathbf{T} = \{X_i \mid i = \overline{1,n}\}, \quad \mathbf{R} = \{Y_i \mid i = \overline{1,n}\}.$$

Definition 1. *The Scalar (generalized scalar) product of arrays* $\mathbf{T}$ *and* $\mathbf{R}$ *is defined as*

$$\langle \mathbf{T},\ \mathbf{R} \rangle = \sum_{i=1}^{n} X_i Y_i^T. \tag{1.4}$$

The result of the product is a square p by p matrix. The scalar product is not commutative:

$$\langle \mathbf{R},\ \mathbf{T} \rangle = \langle \mathbf{T},\ \mathbf{R} \rangle^T.$$

6. The **Scalar square** of array

$$\langle \mathbf{T},\ \mathbf{T} \rangle = \sum_{i=1}^{n} X_i X_i^T. \tag{1.5}$$

is a symmetric and non-negatively defined $(p \times p)$ matrix. For the representation of the scalar square, we shall use the traditional symbol of absolute value: $\langle \mathbf{T},\ \mathbf{T} \rangle = |\mathbf{T}|^2$. In our case, $|\mathbf{T}|$ is the so-called matrix module. [7]

7. The Properties of the scalar product in array spaces are similar to the properties of the traditional scalar product in euclidean vector spaces. If $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$ are arrays in general form, then

$\langle \mathbf{T}_1 + \mathbf{T}_2,\ \mathbf{T}_3 \rangle = \langle \mathbf{T}_1,\ \mathbf{T}_3 \rangle + \langle \mathbf{T}_2,\ \mathbf{T}_3 \rangle$;
$\langle K\mathbf{T}_1,\ \mathbf{T}_2 \rangle = K\langle \mathbf{T}_1,\ \mathbf{T}_2 \rangle$ where K is a square $(p \times p)$ matrix;
$\langle \mathbf{T}_1,\ \mathbf{T}_1 \rangle \succcurlyeq 0$ in the sense of the comparison of square symmetrical matrices;
$\langle \mathbf{T}_1,\ \mathbf{T}_1 \rangle = 0$ iff $\mathbf{T}_1 = 0$.

8. We say that array $\mathbf{T}$ is **orthogonal** to array $\mathbf{R}$, if

$$\langle \mathbf{T},\ \mathbf{R} \rangle = 0.$$

Note that if $\langle \mathbf{T},\ \mathbf{R}\rangle = 0$, then also $\langle \mathbf{R},\ \mathbf{T}\rangle = 0$. Therefore the property of orthogonality of arrays is reciprocal. The orthogonality of arrays $\mathbf{T}$ and $\mathbf{R}$ shall be denoted as $\mathbf{T} \perp \mathbf{R}$.

9. Notice a Pythagorean theorem: if arrays $\mathbf{T}$ and $\mathbf{R}$ are orthogonal, then

$$\langle \mathbf{T}+\mathbf{R},\ \mathbf{T}+\mathbf{R}\rangle = \langle \mathbf{T},\ \mathbf{T}\rangle + \langle \mathbf{R},\ \mathbf{R}\rangle. \tag{1.6}$$

We note again that the result of a scalar product of two arrays is a $(p \times p)$ matrix, therefore in array spaces square matrices of corresponding dimensions should play the role of scalars. In particular, left multiplication by a $(p \times p)$ matrix shall be understood as multiplication by a scalar, and array $k\mathbf{T}$ shall be understood as proportional to array $\mathbf{T}$.

Together with arrays of the form (1.1) we shall consider one-to-one corresponding matrices

$$\mathcal{X} = ||X_1, X_2, \ldots, X_n||. \tag{1.7}$$

Matrix (1.7) is a matrix with p rows and n columns.

Notation. Matrices with p rows and n columns shall be called $(p \times n)$ matrices. The set of $(p\times n)$ matrices we shall call $\mathbb{R}^p_n$. Matrices of dimensions $(p \times 1)$ we shall call p-columns, or simply columns. The set of p-columns we represent as $\mathbb{R}^p_1$. Matrices $(1 \times n)$ we shall call n-rows, or simply rows. The set of n-rows we represent as $\mathbb{R}^1_n$.

Many operations with arrays can be carried out in their matrix forms. For instance, the addition of arrays is equivalent to the addition of their corresponding matrices; left multiplication by a square $(p \times p)$ matrix k is equivalent to the matrix product $k\mathcal{X}$; right multiplication by matrix Q is equivalent to the matrix product $\mathcal{X}Q$; the scalar product of arrays

$$\mathbf{T}_X = \{X_i \mid i = \overline{1,n}\}, \quad \mathbf{T}_Y = \{Y_i \mid i = \overline{1,n}\}$$

is equal to the product of their equivalent matrices $\mathcal{X}$ and $\mathcal{Y}$:

$$\langle \mathbf{T}_X,\ \mathbf{T}_Y\rangle = \mathcal{X}\mathcal{Y}^T. \tag{1.8}$$

We show, for instance, that array $\mathbf{T}Q$ corresponds to matrix $\mathcal{X}Q$. Here $\mathbf{T}$ is the arbitrary array of form (1.1) and $\mathcal{X}$ is the corresponding $(p \times n)$ matrix (1.7). Let $Q = \{q_{\alpha\beta} \mid \alpha, \beta = \overline{1,n}\}$ be a $(n \times n)$ matrix (with numerical elements $q_{\alpha\beta}$).

Proposition 1. *Matrix $\mathcal{X}Q$ corresponds to array $\mathbf{T}Q$.*

Proof. Elements of array $\mathbf{T}$, being columns of matrix $\mathcal{X}$, must be represented in detailed notation. Let

$$X_j = (x_{1j}, x_{2j}, \ldots, x_{pj})^T,\ j = \overline{1,n}.$$

In this notation,

$$\mathcal{X}Q = \| \sum_{j=1}^{n} x_{ij} q_{jk} \mid i = \overline{1,p}, k = \overline{1,n} \|. \tag{1.9}$$

The array

$$\mathbf{T}_Y = \{Y_k \mid k = \overline{1,n}\},$$

corresponds to matrix $\mathcal{X}Q$ where

$$Y_k = (y_{1k}, y_{2k}, \ldots, y_{pk})^T,$$

and

$$y_{ik} = \sum_{j=1}^{n} x_{ij} q_{jk},$$

by (1.9). Array $\mathbf{T}Q$, by definition (1.3), is equal to

$$\mathbf{T}Q = \{X_i \mid i = \overline{1,n}\}Q = \{\sum_{j=1}^{n} X_j q_{kj} \mid k = \overline{1,n}\} = \{Z_k \mid k = \overline{1,n}\},$$

where p-row

$$\begin{aligned} Z_k &= \sum_{j=1}^{n} X_j q_{kj} = \sum_{j=1}^{n} (x_{1j}, \ldots, x_{pj})^T q_{kj} = \\ &= \left(\sum_{j=1}^{n} x_{1j} q_{kj}, \sum_{j=1}^{n} x_{2j} q_{kj}, \ldots, \sum_{j=1}^{n} x_{pj} q_{kj} \right)^T . \end{aligned} \tag{1.10}$$

Comparing expressions (1.9) and (1.10), we see the equality of their elements. □

Thus in a tensor product $\mathbb{R}_n^p \otimes \mathbb{R}_n^1$ we introduced the structure of a module over the ring of square matrices supplied with an inner product, which we called a scalar product.

1.2 Linear Transformations

Many concepts of classical linear algebra transfer to array space almost automatically, with the natural expansion of the field of scalars to the ring of square matrices. For instance, the transformation $f(\cdot)$ of array space (1.1) onto itself is called *linear* if for any array $\mathbf{T}_1$ and $\mathbf{T}_2$ and for any $(p \times p)$ matrix k_1 and k_2

$$f(K_1 \mathbf{T}_1 + K_2 \mathbf{T}_2) = K_1 f(\mathbf{T}_1) + K_2 f(\mathbf{T}_2). \tag{1.11}$$

Linear transformations in array space are performed by right multiplication by square matrices. Let Q be an arbitrary $(n \times n)$ matrix, $\mathbf{T}$ be an arbitrary array (1.1). That transformation

$$f(\mathbf{T}) = \mathbf{T}Q$$

is linear in the sense of (1.11), directly follows from the definition (1.3). That there are no other linear transformations follows from their absence even in the case $p = 1$. (As we know, all linear transformations in vector spaces of rows are performed by right multiplication by square $(n \times n)$ matrices.)

Note that the matrix form (1.7) of representing an array is fitting also for the representation of linear transformations: matrix $\mathcal{X}Q$ (the product of matrices $\mathcal{X}$ and Q) coincides with the matrix form of an array (1.3)

$$\mathbf{T}Q = \{X_i \mid i = \overline{1,n}\}Q = \{\sum_{j=1}^{n} X_j q_{ij} \mid i = \overline{1,n}\}.$$

We shall call a linear transformation of array space onto itself *orthogonal* if this transformation preserves the scalar product. It means that for any arrays $\mathbf{T}$ and $\mathbf{R}$

$$\langle \mathbf{T}Q,\ \mathbf{R}Q \rangle = \langle \mathbf{T},\ \mathbf{R} \rangle.$$

It is easy to see that orthogonal transformations are performed by right multiplication by orthogonal matrices. Indeed,

$$\langle \mathbf{T}Q,\ \mathbf{R}Q \rangle = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} X_j q_{ij} \right) \left(\sum_{l=1}^{n} Y_l q_{il} \right)^T =$$

$$= \sum_{j=1}^{n} \sum_{q=1}^{n} X_j Y_l^T \left(\sum_{i=1}^{n} q_{ij} q_{il} \right) = \sum_{j=1}^{n} X_j Y_j^T,$$

since matrix Q is orthogonal and therefore

$$\sum_{i=1}^{n} q_{ij} q_{il} = \delta_{jl} \quad \text{(Kronecker symbol)}.$$

1.3 Generating Bases and Coordinates

Let $\alpha \in \mathbb{R}_1^p, x \in \mathbb{R}_n^1, \alpha x \in \mathbb{R}_n^p$. Here αx denotes the product of matrices α and x. The matrices of form αx plays a special role in array spaces.

Let n-rows $e_1, e_2, \ldots, e_n \in \mathbb{R}_n^1$ form the basis of the space $\mathbb{R}_n^1$. Let $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}_1^p$ be arbitrary p-columns. Let us consider $(p \times n)$-matrices $\alpha_1 e_1, \alpha_2 e_2, \ldots, \alpha_n e_n$.

Theorem 1. *Any $(p \times n)$-matrix $\mathcal{X}$ (1.7) can be represented as*

$$\mathcal{X} = \sum_{i=1}^{n} \alpha_i e_i \tag{1.12}$$

for some choice of $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}_1^p$ uniquely.

Proof. Let us define $(n \times n)$-matrix E formed by n-rows $e_1, e_2, \ldots, e_n$. Let us also introduce a $(p \times n)$-matrix A formed by p-columns $\alpha_1, \alpha_2, \ldots, \alpha_n$. With matrices A and E the sum (1.12) can be represented as

$$\sum_{i=1}^{n} \alpha_i e_i = AE.$$

Here are some calculations to confirm that assertion. Let $\alpha_i = (\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_{pi})^T$, $e_i = (e_{i1}, e_{i2}, \ldots, e_{in})$.

$$\sum_{i=1}^{n} \alpha_i e_i = \sum_{i=1}^{n} (\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_{pi})^T (e_{i1}, e_{i2}, \ldots, e_{in}).$$

The element at (k, l)-position of each product $\alpha_i e_i$, $i = 1, \ldots, n$, is in essence $\alpha_{ki} e_{il}$. Their total sum, which is the element of matrix $\sum_{i=1}^{n} \alpha_i e_i$, is $\sum_{i=1}^{n} \alpha_{ki} e_{il}$.

The element at (k, l)-position of matrix AE (calculated by the row by column rule) is

$$\sum_{i=1}^{n} \alpha_{ki} e_{il}.$$

The calculated results are equal.

The theorem shall be proven if we show that the equation

$$\mathcal{X} = AE \tag{1.13}$$

has a unique solution relative to the $(p \times n)$-matrix A. Since the $(n \times n)$-matrix E is invertible, the solution is obvious:

$$A = \mathcal{X} E^{-1}. \tag{1.14}$$

□

The theorem allows us to say that the basis of $\mathbb{R}_n^1$ generates the space $\mathbb{R}_n^p$ (using the above method). Thus, the bases in $\mathbb{R}_n^1$ shall be called *generating bases* in relation to $\mathbb{R}_n^p$. The p-columns $\alpha_1, \alpha_2, \ldots, \alpha_n$ from (1.12) can be understood as the coordinates of $\mathcal{X}$ in the generating basis $e_1, e_2, \ldots, e_n$. For the canonical basis of the space $\mathbb{R}_n^1$ (where e_i is an n-row, in which the ith element is one, and the others are zero) coordinates $\mathcal{X}$ relative to this basis are p-columns $X_1, \ldots, X_n \in \mathbb{R}_n^p$, which form the matrix $\mathcal{X}$.

The coordinates of the $(p \times n)$-matrix $\mathcal{X}$ in two different generating bases are connected by a linear transformation. For example, let n-rows $f_1, \dots, f_n$ form the basis in $\mathbb{R}^1_n$. Let F be an $(n \times n)$-matrix composed of these n-rows. By Theorem 1 there exists a unique set of p-columns $\beta_1, \beta_2, \dots, \beta_n$ that are coordinates of $\mathcal{X}$ relative to the generating basis $f_1, \dots, f_n$. Matrices $B = ||\beta_1, \dots, \beta_n||$ and F are connected to the $(p \times n)$-matrix $\mathcal{X}$ by the equivalence

$$\mathcal{X} = BF. \tag{1.15}$$

With (1.13) this gives

$$BF = AE.$$

Therefore,

$$B = AEF^{-1}, \quad A = BFE^{-1}.$$

Corollary 1. *If the generating bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are orthogonal, then the transformation of the coordinates of an array in one basis to the coordinates of it in another is performed through multiplication by an orthogonal matrix.*

Let us consider an arbitrary orthogonal basis $e_1, \dots, e_n$ in $\mathbb{R}^p_n$. For arbitrary $(p \times n)$-matrices $\mathcal{X}$ and $\mathcal{Y}$ we have the decompositions of (1.12) with respect to this basis:

$$\mathcal{X} = \sum_{i=1}^{n} \alpha_i e_i, \quad \mathcal{Y} = \sum_{i=1}^{n} \gamma_i e_i.$$

We can express the scalar product of $\mathcal{X}$ and $\mathcal{Y}$ through their coordinates. It is easy to see that

$$\langle \mathbf{T}_X, \ \mathbf{T}_Y \rangle = \mathcal{X}\mathcal{Y}^T = \sum_{i=1}^{n} \alpha_i \gamma_i^T. \tag{1.16}$$

Corollary 2. *In an orthogonal basis, the scalar product of two arrays is equal to the sum of the paired product of the coordinates.*

Proof. Indeed,

$$\mathcal{X}\mathcal{Y}^T = \langle \sum_{i=1}^{n} \alpha_i e_i, \ \sum_{j=1}^{n} \gamma_j e_j \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i e_i e_j^T \gamma_j^T = \sum_{i=1}^{n} \alpha_i \gamma_i^T,$$

since for the orthogonal basis $e_i e_j^T = \delta_{ij}$. □

Therefore the scalar square of $\mathbf{T}_X$ equals

$$|\mathbf{T}_X|^2 = \mathcal{X}\mathcal{X}^T = \sum_{i=1}^{n} \alpha_i \alpha_i^T.$$

We can conclude from here that the squared length of an array is equal to the sum of its squared coordinates in an orthogonal basis, as for the squared euclidean length of a vector.

1.4 Submodules

We define a Submodule in array space (1.1) (or the space of corresponding matrices (1.7)) to be a set which is closed under linear operations: addition and multiplication by scalars. Remember that multiplication by scalars means left multiplication by $(p \times p)$-matrices. For clarity, we shall discuss arrays in their matrix forms in future.

Definition 2. *The set $\mathcal{L} \subset \mathbb{R}_n^p$ we shall define to be the submodule of space $\mathbb{R}_n^p$, if for any $\mathcal{X}_1, \mathcal{X}_2 \in \mathcal{L}$*

$$K_1\mathcal{X}_1 + K_2\mathcal{X}_2 \in \mathcal{L} \tag{1.17}$$

with arbitrary $(p \times p)$-matrices K_1, K_2.

Theorem 2. *Any submodule $\mathcal{L}$, $\mathcal{L} \subset \mathbb{R}_n^p$, is formed by some linearly independent system of n-rows. The number of elements in this system is uniquely determined by $\mathcal{L}$. This number may be called the dimension of the linear subspace $\mathcal{L}$.*

Proof. Let $\mathcal{X} \in \mathcal{L}$. The set of $(p \times n)$-matrices of the form $K\mathcal{X}$ (where K is an arbitrary $(p \times p)$-matrix) forms a submodule. Let us label it as $\mathcal{L}(\mathcal{X})$ Let $x_1, \ldots, x_p$ be n-rows of the $(p \times n)$-matrix $\mathcal{X}$. Let us choose from among these n-rows a maximal linear independent subsystem, such as $y_1, \ldots, y_k$. It is obvious that

$$\mathcal{L}(\mathcal{X}) = \{\mathcal{Y} \mid \mathcal{Y} = \sum_{i=1}^{k} \beta_i y_i,\ \beta_1, \ldots, \beta_k \in \mathbb{R}_1^p\}.$$

If $\mathcal{L}(\mathcal{X}) = \mathcal{L}$, then $y_1, \ldots, y_k$ form a generating basis for $\mathcal{L} \subset \mathbb{R}_n^p$. If $\mathcal{L}(\mathcal{X}) \neq \mathcal{L}$, then let us find in $\mathcal{L}$ an element, say $\mathcal{Z}$, that does not belong to $\mathcal{L}(\mathcal{X})$. Let us expand the system $y_1, \ldots, y_k$ with n-rows $z_1, \ldots,\ z_p$ of $(p \times n)$-matrix $\mathcal{Z}$. Then we find in this set of n-rows the maximal linearly independent subsystem, and repeat. At some point the process ends. □

Corollary 3. *Any submodule $\mathcal{L} \subset \mathbb{R}_n^p$ can be expressed as the sum of one-dimensional submodules $\mathcal{L}_i \subset \mathbb{R}_n^p$:*

$$\mathcal{L} = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \ldots \oplus \mathcal{L}_l, \tag{1.18}$$

where

$$\mathcal{L}_i = \{\mathcal{X} \mid \mathcal{X} = \alpha y_i, \alpha \in \mathbb{R}_1^p\}$$

for some $y_i \in \mathbb{R}_n^1$. The number l is the same in any representation (1.18) *of $\mathcal{L}$. This number can be called the dimension of subspace $\mathcal{L}$: $l = \dim \mathcal{L}$.*

Note. One can choose an orthogonal linearly independent system of n-rows that generates $\mathcal{L}$. For proof, it is sufficient to note that the generating system can be transformed into an orthogonal one by the process of orthogonalization.

Theorem 2 establishes the one-to-one correspondence between linear subspaces of vector space $\mathbb{R}_n^1$ and the submodules of the matrix space $\mathbb{R}_n^p$. Let us state this as

Corollary 4. *Each linear subspace L in the space of n-rows $\mathbb{R}_n^1$ corresponds to some submodule $\mathcal{L}$ in the space of $(p \times n)$-matrices $\mathbb{R}_n^p$. The dimensions of the linear subspace L and the submodule $\mathcal{L}$ coincide.*

In this manner, the space $\mathbb{R}_n^p$ (and the corresponding array space) and the space $\mathbb{R}_n^1$ have an equal "supply" of linear subspaces and submodules. This leads to significant consequences for multivariate statistical analysis.

Definition 3. *An orthogonal compliment of the submodule $\mathcal{L}$ with respect to the whole space is said to be*

$$\mathcal{L}^{\perp} = \{\mathcal{X} \mid \mathcal{X} \in \mathbb{R}_n^p,\ \langle \mathcal{X},\ \mathcal{Y} \rangle = 0,\ \forall\ \mathcal{Y} \in \mathcal{L}\}. \tag{1.19}$$

s It is easy to see that $\mathcal{L}^{\perp}$ is a submodule and that

$$\mathcal{L} \oplus \mathcal{L}^{\perp} = \mathbb{R}_n^p, \quad \dim \mathcal{L}^{\perp} = n - \dim \mathcal{L}.$$

1.5 Projections onto Submodules

Let us consider array space (1.1) with the introduced scalar product (1.4). Let $\mathcal{L}$ be the submodule (1.17). Let us call *the projection of array* $\mathbf{T}$ *onto a linear subspace* $\mathcal{L}$ the point of $\mathcal{L}$ that is closest to $\mathbf{T}$ in the sense of comparing scalar squares (1.5).

Let us say it in details. Let array $\mathbf{R}$ pass through the set $\mathcal{L}$. We shall call the point $\mathbf{R}^0 \in \mathcal{L}$ *closest to* $\mathbf{T}$ if for any $\mathbf{R} \in \mathcal{L}$

$$\langle \mathbf{T} - \mathbf{R}^0,\ \mathbf{T} - \mathbf{R}^0 \rangle \preccurlyeq \langle \mathbf{T} - \mathbf{R},\ \mathbf{T} - \mathbf{R} \rangle.$$

Note that $\langle \mathbf{T} - \mathbf{R},\ \mathbf{T} - \mathbf{R} \rangle$ is the function of $\mathbf{R}$ with values in the set of $(p \times p)$-matrices. The existence of a minimal element in the set of matrices (generated by $\mathbf{R} \in \mathcal{L}$) is not obvious and is not provided naturally. So the existence of $\operatorname{proj}_{\mathcal{L}} \mathbf{T}$ needs to be proved. We state this result in the following theorem.

Theorem 3. *The projection of* $\mathbf{T}$ *onto* $\mathcal{L}$ *exists, is unique, and has the expected (euclidean) properties.*

1. *For any array* $\mathbf{R} \in \mathcal{L}$,

$$|\mathbf{T} - \mathbf{R}|^2 \succcurlyeq |\mathbf{T} - \operatorname{proj}_{\mathcal{L}} \mathbf{T}|^2,$$

   with equality iff $\mathbf{R} = \operatorname{proj}_{\mathcal{L}} \mathbf{T}$;

2. $(\mathbf{T} - \operatorname{proj}_{\mathcal{L}} \mathbf{T}) \perp \mathcal{L}$;

3. $\operatorname{proj}_{\mathcal{L}}(K_1\mathbf{T}_1 + K_2\mathbf{T}_2) = K_1 \operatorname{proj}_{\mathcal{L}} \mathbf{T}_1 + K_2 \operatorname{proj}_{\mathcal{L}} \mathbf{T}_2$.

Proof. Let $\mathcal{X} \in \mathbb{R}^p_n$ be an arbitrary $(p \times n)$-matrix. As was shown, any submodule $\mathcal{L} \subset \mathbb{R}^p_n$ is equivalent to a linear subspace L in the space of n-rows, $L \subset \mathbb{R}^1_n$. Let Π be a projection matrix onto L in the space $\mathbb{R}^1_n$, that is, for any $x \in \mathbb{R}^1_n$

$$\operatorname{proj}_{\mathcal{L}} x = x\Pi.$$

To prove the theorem we need the following Lemma 1 and Theorem 4.

Lemma 1. *Let* $\mathcal{L} \subset \mathbb{R}^p_n$ *be a submodule in the space of* $(p \times n)$*-matrices, and let* $L \subset \mathbb{R}^1_n$ *be a linear subspace in the space of* n*-rows which generates* $\mathcal{L}$*. Then for any* $\lambda \in \mathbb{R}^p_1$

$$\lambda^T \mathcal{L} = L.$$

Proof of Lemma. Let $r = \dim L, r \leq n$. Let us choose within L the basis $e_1, \dots, e_r$. As we know, the subspace $\mathcal{L} \in \mathbb{R}^p_n$ can be represented as

$$\mathcal{L} = \{\mathcal{Y} \mid \mathcal{Y} = \sum_{k=1}^{r} \alpha_k e_k, \alpha_1, \dots, \alpha_k \in \mathbb{R}^p_1\}.$$

Let $\mathcal{Y} \in \mathcal{L}$, then for some $\alpha_1, \dots, \alpha_k \in \mathbb{R}^p_1$

$$\mathcal{Y} = \sum_{k=1}^{r} \alpha_k e_k.$$

Therefore, under any $\lambda \in \mathbb{R}^p_1$

$$\lambda^T \mathcal{Y} = \sum_{k=1}^{r} (\lambda^T \alpha_k) e_k \in L,$$

since $\lambda^T \alpha_1, \dots, \lambda^T \alpha_r$ are numerical coefficients. □

Theorem 4. *Let* $\mathcal{L}$ *be a submodule in the space of* $(p \times n)$*-matrices. Let* L *be a linear subspace in the space of* n*-rows which generates* $\mathcal{L} \subset \mathbb{R}^p_n$*. Let* Π *be a projection* $(n \times n)$*-matrix onto* L*, that is, for any* $x \in \mathbb{R}^1_n$

$$\operatorname{proj}_L x = x\Pi.$$

Then for any $\mathcal{X} \in \mathbb{R}^p_n$

$$\operatorname{proj}_{\mathcal{L}} \mathcal{X} = \mathcal{X}\Pi.$$

Proof. We must show that for any $\mathcal{Y} \in \mathcal{L}$

$$\langle \mathcal{X} - \mathcal{Y},\ \mathcal{X} - \mathcal{Y} \rangle \succcurlyeq \langle \mathcal{X} - \mathcal{X}\Pi,\ \mathcal{X} - \mathcal{X}\Pi \rangle \tag{1.20}$$

with equality if and only if $\mathcal{Y} = \mathcal{X}\Pi$. The inequality between two symmetrical $(p \times p)$-matrices in (1.20) means that for any $\lambda \in \mathbb{R}_1^p$

$$\lambda^T \langle \mathcal{X} - \mathcal{Y},\ \mathcal{X} - \mathcal{Y} \rangle \lambda \ \geq \lambda^T \langle \mathcal{X} - \mathcal{X}\Pi,\ \mathcal{X} - \mathcal{X}\Pi \rangle \lambda,$$

thus

$$|\lambda^T (\mathcal{X} - \mathcal{Y})|^2 \geq |\lambda^T (\mathcal{X} - \mathcal{X}\Pi)|^2,$$

thus

$$|\lambda^T \mathcal{X} - \lambda^T \mathcal{Y}|^2 \geq |\lambda^T \mathcal{X} - (\lambda^T \mathcal{X})\Pi|^2.$$

As was noted above, the n-row $y = \lambda^T \mathcal{Y}$ belongs to L and $\lambda^T \mathcal{X}\Pi = x\Pi$ is a projection of $\lambda^T \mathcal{X}$ onto L. Due to the properties of euclidean projection, we get for any $y \in L$

$$|x - y|^2 \geq |x - x\Pi|^2$$

with equality if and only if $y = x\Pi$. Thus, $\mathcal{X}\Pi$ is the nearest point to $\mathcal{X}$ in $\mathcal{L}$. ☐

Now we return to proving Theorem 3. From Theorem 4 we know $\mathcal{X}\Pi$ is the unique projection of $\mathcal{X}$ onto $\mathcal{L}$. So, statement 1 of Theorem 3 is proven.

The explicit expression $\text{proj}_{\mathcal{L}} \mathcal{X} = \mathcal{X}\Pi$ confirms that the operation of projection onto a submodule is a linear operation. So, statement 3 of Theorem 3 is proven as well.

To complete the proof of Theorem 3 we need to show statement 2. Let $e_1, \ldots, e_r$ be an orthogonal basis of L and $e_{r+1}, \ldots, e_n$ be an orthogonal basis of $L^\perp$. Then, $e_1, \ldots, e_n$ is the orthogonal basis of $\mathbb{R}_n^1$. In this orthogonal basis, if

$$\mathcal{X} = \sum_{i=1}^{n} \alpha_i e_i,$$

then

$$\mathcal{X} - \mathcal{X}\Pi = \sum_{i=r+1}^{n} \alpha_i e_i.$$

Since $\mathcal{Y} \in \mathcal{L}$,

$$\mathcal{Y} = \sum_{i=1}^{r} \beta_i e_i$$

for some $\beta_1, \ldots, \beta_r \in \mathbb{R}_1^p$. Therefore:

$$(\mathcal{X} - \mathcal{X}\Pi)\mathcal{Y}^T = \langle \sum_{i=r+1}^{n} \alpha_i e_i,\ \sum_{i=1}^{r} \beta_i e_i \rangle = \sum_{i=r+1}^{n} \sum_{j=1}^{r} \langle \alpha_i e_i,\ be_j e_j \rangle =$$

$$= \sum_{i=r+1}^{n} \sum_{j=1}^{r} \alpha_i e_i (\beta_j e_j)^T = \sum_{i=r+1}^{n} \sum_{j=1}^{r} \alpha_i e_i e_j^T \beta_j^T = 0,$$

since $e_i e_j^T = 0$ when $i \neq j$. □

1.6 Matrix Least Squares Method

Calculating projections onto a submodule $\mathcal{L} \subset \mathbb{R}_n^p$ become easier if the form of projection onto the linear subspace L which generates $\mathcal{L}$ is known. By the lemma from Section 1.5, for any $\lambda \in \mathbb{R}_1^p$

$$\lambda^T \operatorname{proj}_{\mathcal{L}} \mathcal{X} = \operatorname{proj}_L(\lambda^T \mathcal{X}). \tag{1.21}$$

Assume that for the right part of (1.21) we have an explicit formula $y = \operatorname{proj}_L x$. Then because of the linearity this gives us for $\operatorname{proj}_{\mathcal{L}}(\lambda^T \mathcal{X})$ an explicit formula $\lambda^T \mathcal{Y}$. Therefore

$$\lambda^T \operatorname{proj}_{\mathcal{L}} \mathcal{X} = \lambda^T \mathcal{Y}. \tag{1.22}$$

So we get an explicit expression for $\operatorname{proj}_{\mathcal{L}} \mathcal{X}$. One can say this is the calculation of $\operatorname{proj}_{\mathcal{L}} \mathcal{X}$ by Roy's method. [5]

Example: calculating the arithmetic mean. Let $X_1, X_2, \ldots, X_n \in \mathbb{R}_1^p$ be the set of p-columns. Let us consider the array $\mathbf{T} = \{X_i \mid i = \overline{1, n}\}$ and represent it in matrix form.

$$\mathcal{X} = \|X_1, X_2, \ldots, X_n\|. \tag{1.23}$$

Our task is to find the array $\mathcal{Y}$ with identical columns, i.e., an array of form

$$\mathcal{Y} = \|Y, Y, \ldots, Y\|, \ Y \in \mathbb{R}_1^p, \tag{1.24}$$

closest to (1.23). Arrays of form (1.24) produce a one-dimensional submodule. We shall denote it by $\mathcal{L}$, $\mathcal{L} \subset \mathbb{R}_n^p$. We have to find $\operatorname{proj}_{\mathcal{L}} \mathcal{X}$. The submodule $\mathcal{L}$ is generated by a one-dimensional linear subspace $L, L \subset \mathbb{R}_n^1$, spanned by n-row $e = (1, 1, \ldots, 1)$.

Let x be an arbitrary n-row, $x = (x_1, \ldots, x_n)$. The form of projection of x onto L is well known:

$$\operatorname{proj}_L x = (\overline{x}, \ldots, \overline{x}).$$

Applying Roy's method to matrix $\mathcal{X}$ (1.23) we get over to n-row $x = \lambda^T \mathcal{X}$, where $x_i = \lambda^T X_i, \ i = \overline{1, n}$. It is then clear that

$$\operatorname{proj}_L x = (\lambda^T \overline{X}, \ldots, \lambda^T \overline{X}).$$

Therefore,

$$\operatorname{proj}_{\mathcal{L}} \mathcal{X} = (\overline{X}, \ldots, \overline{X}). \tag{1.25}$$

Of course, this is not the only and not always the most efficient method. In this example, like in other cases, one can apply the matrix method of least squares and find

$$\hat{Y} = \arg\min_{Y \in \mathbb{R}_1^p} \sum_{i=1}^{n} (X_i - Y)(X_i - Y)^T. \tag{1.26}$$

Solution. Let us transform the function in (1.26): for any $Y \in \mathbb{R}_1^p$

$$\sum_{i=1}^{n}(X_i - Y)(X_i - Y)^T = \sum_{i=1}^{n}[(X_i - \overline{X}) + (\overline{X} - Y)][(X_i - \overline{X}) + (\overline{X} - Y)]^T =$$

$$= \sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T + n\sum_{i=1}^{n}(\overline{X} - Y)(\overline{X} - Y)^T = (1) + (2), \tag{1.27}$$

since “paired products” turn to zero:

$$\sum_{i=1}^{n}(X_i - \overline{X})(\overline{X} - Y)^T = 0, \qquad \sum_{i=1}^{n}(\overline{X} - Y)(X_i - \overline{X})^T = 0.$$

Now the function (1.27) is a sum of two nonnegatively defined matrices, and the first one does not depend on Y. The minimum attains at $Y = \overline{X}$: at that point the nonnegatively definite matrix (2) turns to zero.

The answer is an arithmetic mean, that is,

$$\hat{Y} = \overline{X}.$$

Of course, it is well known. It can be find by applying not the matrix but the ordinary method of least squares:

$$\hat{Y} = \arg\min_{Y \in \mathbb{R}_1^p} \sum_{i=1}^{n} (X_i - Y)^T(X_i - Y).$$

The results of the matrix method similarly relate to the traditional in the case of projection on other submodules $\mathcal{L} \subset \mathbb{R}_n^p$. The reason is simple: if an array $\mathcal{Y}$ is the solution of a matrix problem

$$\sum_{i=1}^{n}(X_i - Z_i)(X_i - Z_i)^T = (\mathcal{X} - \mathcal{Z})(\mathcal{X} - \mathcal{Z})^T \to \min_{\mathcal{Z} \in \mathcal{L}},$$

then Y is a solution of the scalar problem as well,

$$\mathrm{tr}\{\sum_{i=1}^{n}(X_i - Z_i)(X_i - Z_i)^T\} = \sum_{i=1}^{n}(X_i - Z_i)^T(X_i - Z_i) \to \min_{\mathcal{Z} \in \mathcal{L}}.$$

Thus, for instance, in calculating the projection on submodulqes one can use the traditional scalar method of least squares. Both least squares methods in linear models give us the same estimates of parameters. The necessity of matrix scalar products and the matrix form of orthogonality, projection, submodules, etc becomes apparent in testing linear hypothesis. We shall relate this in the next section.

2 Multivariate Linear Models

2.1 Arrays with Random Elements

Let us consider array (2.1), the elements of which are p-dimensional random variables presented in the form of p-columns.

$$\mathbf{T} = \{X_i \mid i = \overline{1,n}\}. \tag{2.1}$$

Remember that we treat such an array as a row composed of p-columns under algebraic operations. For arrays of form (2.1) with random elements, let us define mathematical expectation and covariance. The array

$$\mathrm{E}\,\mathbf{T}_X = \{\mathrm{E}\,X_i \mid i = \overline{1,n}\} \tag{2.2}$$

is called the *mathematical expectation* of $\mathbf{T}$. We define the covariance matrix of array (2.1) much like the covariance matrix of random vector. Let

$$t = (x_1, \dots, x_n)$$

be an n-row composed of random variables $x_1, x_2, \dots, x_n$. As we know, the covariance matrix $\mathrm{Var}\,t$ of the random vector t is an $(n \times n)$-matrix with elements

$$\sigma_{ij} = \mathrm{Cov}(x_i, x_j), \quad \text{where } i,j = \overline{1,n}.$$

Algebraically, with the help of matrix operations, the covariance matrix of the random vector t can be defined as

$$\mathrm{Var}\,t = \mathrm{E}\,(t - \mathrm{E}\,t)^T (t - \mathrm{E}\,t). \tag{2.3}$$

Following (2.3), we define the *covariance array* of random array (2.1) as

$$\mathrm{Var}\,\mathbf{T} := \mathrm{E}\,(\mathbf{T} - \mathrm{E}\,\mathbf{T})^T (\mathbf{T} - \mathrm{E}\,\mathbf{T}) = \{\mathrm{Cov}(X_i, X_j) \mid i,j = \overline{1,n}\}. \tag{2.4}$$

Here $\mathrm{Cov}(X_i, X_j)$ is a covariance matrix of random column-vectors X_i and X_j,

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}\,(X_i - \mathrm{E}\,X_i)(X_j - \mathrm{E}\,X_j)^T. \tag{2.5}$$

Note that we consider $\mathrm{Var}\,\mathbf{T}$ (2.4) as a square array of dimensions $(n \times n)$, the elements of which are $(p \times p)$- matrices (2.5).

Let us consider the array $\mathbf{R}$, obtained by the linear transformation of array $\mathbf{T}$ (2.1)

$$\mathbf{R} = \mathbf{T}Q, \tag{2.6}$$

where Q is a $(n \times n)$-matrix.

It is clear that

$$\mathrm{E}\,\mathbf{R} = (\mathrm{E}\,\mathbf{T})Q,$$
$$\mathrm{Var}\,\mathbf{R} = \mathrm{E}\,[(\mathbf{T}Q - \mathrm{E}\,\mathbf{T}Q)^T(\mathbf{T}Q - \mathrm{E}\,\mathbf{T}Q)] = Q^T(\mathrm{Var}\,\mathbf{T})Q. \tag{2.7}$$

In mathematical statistics, arrays with statistically independent random elements are of especial interest when the covariance matrices of these elements are the same. Let $\mathbf{T}$ (2.1) be an array such that

$$\mathrm{Cov}(X_i, X_j) = \delta_{ij}\Sigma, \quad i, j = \overline{1, n}. \tag{2.8}$$

Here Σ is a nonnegatively defined $(p \times p)$-matrix and δ_{ij} is the symbol of Kronecker. Let us consider an orthogonal transformation of array $\mathbf{T}$:

$$\mathbf{R} = \mathbf{T}C, \tag{2.9}$$

where C is an orthogonal $(n \times n)$-matrix. The following lemma is fairly simple but important.

Lemma 2.

$$\mathrm{Var}\,\mathbf{R} = \mathrm{Var}\,\mathbf{T} = \{\delta_{ij}\Sigma \mid i, j = \overline{1, n}\} \tag{2.10}$$

This lemma generalizes for the multivariate case the well-known property of spherical normal distributions.

Proof. The proof of the lemma is straightforward. To simplify the formulas, assume that $\mathrm{E}\,\mathbf{T} = 0$. Then, (2.7),

$$\mathrm{E}\,(\mathbf{T}C) = \mathrm{E}\,[(\mathbf{T}C)^T(\mathbf{T}C)] = C^T(\mathrm{Var}\,\mathbf{T})C =$$
$$= C^T\{\delta_{ij}\Sigma \mid i, j = \overline{1, n}\}C = \{\delta_{ij}\Sigma \mid i, j = \overline{1, n}\}.$$

□

Earlier, while discussing generating bases and coordinates (Section 1.3), we established that the transformation from the coordinates of array $\mathbf{T}$ in an orthogonal basis to coordinates of this array in another basis can be done through multiplication by an orthogonal matrix. Therefore if the coordinates of some array in one orthogonal basis are not correlated and have a common covariance matrix, then the coordinates of the given array hold these properties in any orthogonal basis. From the remark above and just established Lemma 2 follows

Lemma 3. *If the coordinates of a random array in an orthogonal basis are uncorrelated and have a common covariance matrix, then the coordinates of this array are uncorrelated and have the same common covariance in any orthogonal basis.*

This property is of great importance in studying linear statistical models.

Finally, let us note that in introducing and discussing covariance arrays of random arrays we have to work with the arrays themselves (1.1) and not with the matrices (1.7) representing them.

2.2 Linear Models and Linear Hypotheses

Definition 4. *One says that array* $\mathbf{T}$ (2.1) *with random elements obeys a linear model if*

a) for some given submodule $\mathcal{L}$

$$\mathrm{E}\,\mathbf{T} \in \mathcal{L}; \tag{2.11}$$

b) elements $X_1, \ldots, X_n$ *of array* $\mathbf{T}$ *are independent and identically distributed.*

If this is common for all X_i, with $i = \overline{1,n}$ a gaussian distribution, then we say that array $\mathbf{T}$ follows a *linear gaussian model*. We will now study linear gaussian models.

We shall denote with Σ the common covariance matrix for all p-columns. The array $\mathrm{E}\,\mathbf{T}$ and matrix Σ are *parameters of the model*. They are generally unknown; although, Σ is assumed to be nondegenerate.

For random arrays following the gaussian model, *linear hypotheses* are often discussed. Within the framework of the linear model (2.11) the linear hypothesis holds the form:

$$\mathrm{E}\,\mathbf{T} \in \mathcal{L}_1, \tag{2.12}$$

where $\mathcal{L}_1$ is a given submodule, and $\mathcal{L}_1 \subset \mathcal{L}$.

Let us show that the linear models and linear hypotheses discussed in multivariate statistical analysis have the structure of (2.11) and (2.12). The main linear models are factor and regression. For example, let us consider the one-way layout and regression models of multivariate statistical analysis.

The **One-way layout model** is the simplest of the "analysis of variance" models. It is a shift problem of several (say, m) normal samples with identical covariance matrices. The array of observations in this problem has to have double numeration:

$$\mathbf{T} = \{X_{ij} \mid j = \overline{1,m}, i = \overline{1,n_j}\}. \tag{2.13}$$

Here m is the number of different levels of the factor, which affects the expected values of the response. Here, n_j is the number of independently repeated observations of the response on the level j of the factor, $j = \overline{1, m}$. Finally, multivariate variables X_{ij} are independent realizations of a p-dimensional response, $X_{ij} \in \mathbb{R}_1^p$. Assume $N = n_1 + \cdots + n_m$. The main assumption of the model is: $X_{ij} \sim N_p(a_j, \Sigma)$.

We shall linearly order the observations which constitute the array (2.13) and then represent (2.13) as a $(p \times N)$-matrix.

$$\mathcal{X} = \|X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}, X_{m1}, \ldots X_{mn_m}\|. \tag{2.14}$$

Note that

$$\mathrm{E}\,\mathcal{X} = \| \underbrace{a_1, \ldots, a_1}_{n_1 \text{ times}}, \underbrace{a_2, \ldots, a_2}_{n_2 \text{ times}}, \ldots, \underbrace{a_m, \ldots, a_m}_{n_m \text{ times}} \|. \tag{2.15}$$

Let us introduce N-rows

$$\begin{aligned} e_1 &= (\underbrace{1, \ldots, 1}_{n_1 \text{ times}}, 0, \ldots, 0), \\ e_2 &= (\underbrace{0, \ldots, 0}_{n_1 \text{ times}}, \underbrace{1, \ldots, 1}_{n_2 \text{ times}}, 0, \ldots, 0), \\ &\ldots \\ e_m &= (\underbrace{0, \ldots, 0}_{n_1 \text{times}}, \underbrace{0, \ldots, 0}_{n_2 \text{ times}}, \ldots, \underbrace{1, \ldots, 1}_{n_m \text{ times}}). \end{aligned} \tag{2.16}$$

It is obvious that

$$\mathrm{E}\,\mathcal{X} = \sum_{i=1}^{m} a_i e_i.$$

Therefore $\mathrm{E}\,\mathcal{X}$ belongs to an m dimensional submodule of the space $\mathbb{R}_N^p$ spanned by n-rows (2.16).

The hypothesis $H_0 : a_1 = a_2 = \cdots = a_m$, with which one usually begins the statistical analysis of m samples, is obviously a linear hypothqesis in the sense of (2.12) $H_0 : \mathrm{E}\,\mathcal{X} \in \mathcal{L}_1$, where $\mathcal{L}_1$ is a one dimensional linear subspace spanned by the single N-row $e = e_1 + \cdots + e_m$.

Multivariate Multiple Regression in matrix form is

$$\mathcal{Y} = \mathcal{A}\mathcal{X} + \mathcal{E}, \tag{2.17}$$

where $\mathcal{Y} = \|Y_1, Y_2, \ldots, Y_n\|$. Here $\mathcal{Y}$ is theq observed $(p \times n)$-matrix of p-dimensional response; $\mathcal{X}$ is a given design $(m \times n)$-matrix; $\mathcal{A}$ is a $(p \times m)$-matrix of unknown regression coefficients; $\mathcal{E} = \|E_1, E_2, \ldots, E_n\|$ is a $(p \times n)$-matrix composed of independent p-variate random errors $E_1, E_2, \ldots, E_n$. In gaussian models

$$E_i \sim N_p(0, \Sigma),$$

where $(p \times p)$ matrix Σ is assumed to be non-degenerate. Generally Σ is believed to be unknown.

Let $A_1, A_2, \ldots, A_m$ be the p-columns forming matrix $\mathcal{A}$; let $x_1, x_2, \ldots, x_m$ be n-rows, forming matrix $\mathcal{X}$. Then

$$\mathcal{A}\mathcal{X} = \sum_{i=1}^{m} A_i x_i. \tag{2.18}$$

The resulting expression (2.18) shows that $\mathrm{E}\,\mathcal{Y} = \mathcal{A}\mathcal{X}$ belongs to an m dimensional submodule of the space $\mathbb{R}^p_n$, generated by the linear system of n-rows $x_1, x_2, \ldots, x_m$.

2.3 Sufficient Statistics and Best Unbiased Estimates

Let us consider a linear gaussian model (2.11) in matrix form

$$\mathcal{X} = \mathcal{M} + \mathcal{E}. \tag{2.19}$$

where $\mathcal{M} = \mathrm{E}\,\mathcal{X}$ is an unknown $(p \times n)$-matrix;

$$\mathcal{M} = \|M_1, M_2, \ldots, M_n\| \in \mathcal{L},$$

where $\mathcal{L}$ is a submodule of $\mathbb{R}^p_n$;

$$\mathcal{E} = \|E_1, E_2, \ldots, E_n\|$$

is a $(p \times n)$-matrix, the p-columns $E_1, E_2, \ldots, E_n$ of which are the independent $N_p(0, \Sigma)$ random variables.

The unknown parameter of this gaussian model is a pair $(\mathcal{M}, \Sigma)$. Let us find sufficient statistics for this parameter using the factorization criterion.

A likelihood of the pair $(\mathcal{M}, \Sigma)$ based on $\mathcal{X}$ is

$$\prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}}\right)^p \frac{1}{\sqrt{\det \Sigma}} \exp\{-\frac{1}{2}(X_i - M_i)^T \Sigma^{-1} (X_i - M_i)\} =$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{np} \left(\frac{1}{\sqrt{\det \Sigma}}\right)^n \exp\{-\frac{1}{2}\operatorname{tr} \Sigma^{-1} [\sum_{i=1}^{n} (X_i - M_i)(X_i - M_i)^T]\}. \tag{2.20}$$

The sum in square brackets is $\langle \mathcal{X} - \mathcal{M},\ \mathcal{X} - \mathcal{M} \rangle$. Let us represent $\mathcal{X} - \mathcal{M}$ as

$$\mathcal{X} - \mathcal{M} = (\mathcal{X} - \mathrm{proj}_{\mathcal{L}}\mathcal{X}) + (\mathrm{proj}_{\mathcal{L}}\mathcal{X} - \mathcal{M}) = (1) + (2)$$

and note that

$$(1) = \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X} \in \mathcal{L}^\perp, \quad (2) \in \mathcal{L}.$$

Thus, (Pythagorean Theorem)

$$\langle \mathcal{X} - \mathcal{M},\ \mathcal{X} - \mathcal{M} \rangle = \langle \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X} \rangle + \langle \mathrm{proj}_{\mathcal{L}}\mathcal{X} - \mathcal{M},\ \mathrm{proj}_{\mathcal{L}}\mathcal{X} - \mathcal{M} \rangle. \tag{2.21}$$

We conclude that the likelihood (2.20) is expressed through the statistics $\mathrm{proj}_{\mathcal{L}}\mathcal{X}$ and $\langle \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X} \rangle$, which are sufficient for $\mathcal{M}, \Sigma$.

The statistic $\mathrm{proj}_{\mathcal{L}}\mathcal{X}$ is obviously an unbiased estimate of $\mathcal{M}$. As a function of sufficient statistics it is the best unbiased estimate of $\mathcal{M}$. We can show that the best unbiased estimate of Σ is the statistic

$$\frac{1}{\dim \mathcal{L}^\perp} \langle \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X} \rangle \tag{2.22}$$

after proving the following theorem 5.

2.4 Theorem of Orthogonal Decomposition

Theorem 5. *Let $\mathcal{X} = \|X_1, X_2, \dots, X_n\|$ be a gaussian $(p \times n)$ matrix with independent p-columns $X_1, X_2, \dots, X_n \in \mathbb{R}^p_1$, and $\mathrm{Var}\, X_i = \Sigma$ for all $i = 1, \dots, n$. Let $\mathcal{L}_1, \mathcal{L}_2, \dots$ be pairwise orthogonal submodules $\mathbb{R}^p_n$, the direct sum of which forms $\mathbb{R}^p_n$:*

$$\mathbb{R}^p_n = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \dots$$

Let us consider the decomposition of $(p \times n)$-matrix $\mathcal{X}$ into the sum of orthogonal projections $\mathcal{X}$ on the submodules $\mathcal{L}_1, \mathcal{L}_2, \dots$:

$$\mathcal{X} = \mathrm{proj}_{\mathcal{L}_1}\mathcal{X} + \mathrm{proj}_{\mathcal{L}_2}\mathcal{X} + \dots$$

Then:

- *a) random $(p \times n)$-matrices $\mathrm{proj}_{\mathcal{L}_1}\mathcal{X}, \mathrm{proj}_{\mathcal{L}_2}\mathcal{X}, \dots$ are independent, normally distributed, and $\mathrm{E}\, \mathrm{proj}_{\mathcal{L}_i}\mathcal{X} = \mathrm{proj}_{\mathcal{L}_i}\mathrm{E}\, \mathcal{X}$;*
- *b) $\langle \mathrm{proj}_{\mathcal{L}_i}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_i}\mathcal{X} \rangle = W_p(\dim L_i, \Sigma, \Delta_i)$, where $W_p(\nu, \Sigma, \Delta)$ indicates a random matrix (of size $(p \times p)$), distributed under Wishart, with ν degrees of freedom and the parameter of non-centrality Δ. In this case*

$$\Delta_i = \langle \mathrm{proj}_{\mathcal{L}_i}\mathrm{E}\, \mathcal{X},\ \mathrm{proj}_{\mathcal{L}_i}\mathrm{E}\, \mathcal{X} \rangle.$$

Proof. Each submodule $\mathcal{L} \subset \mathbb{R}^p_n$ has a one-to-one correspondence to some linear subspace $L \subset \mathbb{R}^1_n$ which generates it, and $\dim \mathcal{L} = \dim L$. Let submodules $\mathcal{L}_1, \mathcal{L}_2, \dots \subset \mathbb{R}^p_n$ correspond to the subspaces $L_1, L_2, \dots \subset \mathbb{R}^1_n$. The subspaces $L_1, L_2, \dots \subset \mathbb{R}^1_n$ are pairwise orthogonal, and their direct sum forms the entire space $\mathbb{R}^1_n$. Let us denote the dimensions of submodules $\mathcal{L}_1, \mathcal{L}_2, \dots \subset \mathbb{R}^p_n$ (and subspaces $L_1, L_2, \dots \subset \mathbb{R}^1_n$) by $m_1, m_2, \dots$.

Let us choose in every subspace $L_1, L_2, \dots$ an orthogonal basis. For $\mathcal{L}_1$ let it be the n-rows $f_1, \dots, f_{m_1}$; for $\mathcal{L}_2$, the n-rows $f_{m_1+1}, \dots, f_{m_1+m_2}$ etc. With the help of these n-rows each of the submodules $\mathcal{L}_1, \mathcal{L}_2, \dots$ can be represented as the direct sum of one dimensional submodules from $\mathbb{R}^p_n$. For example, $\mathcal{L}_1 = \mathcal{F}_1 \oplus \mathcal{F}_2 \oplus \dots \oplus \mathcal{F}_{m_1}$, where

$$\begin{aligned}
&\mathcal{F}_1 = \{\mathcal{Y} \mid \mathcal{Y} = \alpha f_1,\ \alpha \in \mathbb{R}^p_1\},\\
&\mathcal{F}_2 = \{\mathcal{Y} \mid \mathcal{Y} = \alpha f_2,\ \alpha \in \mathbb{R}^p_1\},\\
&\dots\\
&\mathcal{F}_{m_1} = \{\mathcal{Y} \mid \mathcal{Y} = \alpha f_{m_1},\ \alpha \in \mathbb{R}^p_1\}.
\end{aligned}$$

The set of all n-rows $f_1, f_2, \dots, f_n$ forms an orthogonal basis in $\mathbb{R}^1_n$ and so does the generating basis in $\mathbb{R}^p_n$. Therefore any $(p \times n)$-matrix $\mathcal{X} \in \mathbb{R}^p_n$ can be represented in the form

$$\mathcal{X} = \sum_{i=1}^{n} Y_i f_i,$$

where $Y_1, \dots, Y_n$ are some p-columns, that is $Y_1, \dots, Y_n \in \mathbb{R}^p_1$, and

$$\begin{aligned}
\mathrm{proj}_{\mathcal{L}_1}\mathcal{X} &= \sum_{i=1}^{m_1} Y_i f_i,\\
\mathrm{proj}_{\mathcal{L}_2}\mathcal{X} &= \sum_{i=m_1+1}^{m_2} Y_i f_i \qquad \text{etc.}
\end{aligned}$$

Here p-columns $Y_1, Y_2, \dots, Y_n$ are coordinates of a $(p \times n)$-matrix $\mathcal{X}$ relative to the generating basis $f_1, \dots, f_n$, while the p-columns $X_1, X_2, \dots, X_n$ are coordinates of the same $(p \times n)$-matrix $\mathcal{X}$ relative to the orthogonal canonical basis $\mathbb{R}^1_n$: $e_1 = (1, 0, \dots)$, $e_2 = (0, 1, 0, \dots)$ etc. As was noted earlier (see Lemma 3), the transformation from some coordinates to others is performed through the right multiplication of an $(p \times n)$-matrix $\mathcal{X}$ by some orthogonal transformation $(n \times n)$-matrix, say by $(n \times n)$-matrix C:

$$\|Y_1, Y_2, \dots, Y_n\| = \|X_1, X_2, \dots, X_n\| C, \quad \text{or} \quad \mathcal{Y} = \mathcal{X} C.$$

Thus the p-columns $Y_1, \dots, Y_n$ are mutually normally distributed. Following Lemma 3,

$$\mathrm{Var}\,\mathcal{Y} = \mathrm{Var}\,\mathcal{X} = \{\delta_{ij}\Sigma \mid i, j = \overline{1, n}\}.$$

This means that $Y_1, \dots, Y_n$ are independent gaussian p-columns with common covariance matrix Σ, just like the p-columns $X_1, \dots, X_n$.

Let us consider random $(p \times p)$-matrices

$$\langle \mathrm{proj}_{\mathcal{L}_1}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_1}\mathcal{X}\rangle, \langle \mathrm{proj}_{\mathcal{L}_2}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2}\mathcal{X}\rangle, \dots$$

For example,

$$\langle \mathrm{proj}_{\mathcal{L}_1}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_1}\mathcal{X}\rangle = \sum_{i=1}^{m_1} Y_i Y_i^T.$$

The distribution of such random matrices is called a Wishart distribution. If $\mathrm{E}\,Y_1 = \mathrm{E}\,Y_2 = \cdots = \mathrm{E}\,Y_{m_1} = 0$, we get the so-called central Wishart distribution $W_p(m_1, \Sigma)$. Let us note that if one uses the notation $W_p(m, \Sigma)$ for a random matrix itself, not only for its distribution, then one can say that

$$W_p(m, \Sigma) = \Sigma^{\frac{1}{2}} W_p(m, I) \Sigma^{\frac{1}{2}},$$

if one represents as $\Sigma^{\frac{1}{2}}$ a symmetric matrix, the unique symmetric solution of the matrix equation: $Z^2 = \Sigma$.

One says that a random $(p \times p)$-matrix W has the noncentral Wishart distribution if

$$W = \sum_{i=1}^{m} (\xi_i + a_i)(\xi_i + a_i)^T,$$

where the p-columns $\xi_1, \xi_2, \ldots, \xi_m$ are iid $N_p(0, \Sigma)$, $a_1, a_2, \ldots, a_m$ are some nonrandom p-columns, generally distinct from zero. The distribution W somehow depends on the p-columns $a_1, a_2, \ldots, a_m$. Let us show that the distribution W depends on the noted p-columns through a so-called parameter of noncentrality: the $(p \times p)$-matrix

$$\Delta = \sum_{i=1}^{m} a_i a_i^T.$$

Let us introduce the $(p \times m)$-matrices

$$\xi = \|\xi_1, \xi_2, \ldots, \xi_m\|,$$
$$\mathcal{A} = \|a_1, a_2, \ldots, a_m\|.$$

In these notations

$$W = \langle \xi + \mathcal{A},\ \xi + \mathcal{A}\rangle.$$

Let C be an arbitrary orthogonal $(m \times m)$-matrix. Say $\eta = \xi C$. Note that $\eta \overset{\mathrm{d}}{=} \xi$, and

$$W \overset{\mathrm{d}}{=} \langle \eta + AC,\ \eta + AC\rangle.$$

We see that the noncentral Wishart distribution depends on $\mathcal{A} = \|a_1, \ldots, a_m\|$ not directly but through the maximal invariant $\mathcal{A}$ under orthogonal transformations, that is through $\langle \mathcal{A},\ \mathcal{A}\rangle = \sum_{i=1}^{m} a_i a_i^T$.

Therefore, in the general case

$$\langle \mathrm{proj}_{\mathcal{L}_i}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_i}\mathcal{X}\rangle = W_p(m_i, \Sigma, \Delta_i),$$

where $\Delta_i = \langle \mathrm{proj}_{\mathcal{L}_i}\mathrm{E}\,\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_i}\mathrm{E}\,\mathcal{X}\rangle$. □

Let us return to the unbiased estimate of parameter Σ of linear models. In linear model (2.19) $\mathrm{proj}_{\mathcal{L}^\perp}\mathrm{E}\,\mathcal{X} = 0$. Therefore the statistic (2.22) is

$$\frac{1}{\dim \mathcal{L}^\perp}\,\langle \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X}\rangle = \frac{1}{n-m}\Sigma^{\frac{1}{2}}W_p(n-m, I)\Sigma^{\frac{1}{2}}.$$

It is obvious that its expected value is Σ.

2.5 Testing Linear Hypotheses

Copying the univariate linear model, we shall define the hypothesis in the multivariate linear model (2.19) as

$$H : \mathrm{E}\,\mathcal{X} \in \mathcal{L}_1, \tag{2.23}$$

where $\mathcal{L}_1$ is a given submodule such that $\mathcal{L}_1 \subset \mathcal{L}$.

In this section we will propose statistics which may serve as the base for the construction of statistical criteria for testing H (2.23), free (under H) from the parameters $\mathcal{M}$, Σ.

Let us introduce the submodule $\mathcal{L}_2$ which is an orthogonal complement $\mathcal{L}_1$ with respect to $\mathcal{L}$:

$$\mathcal{L} = \mathcal{L}_1 \oplus \mathcal{L}_2. \tag{2.24}$$

Let us consider the decomposition of the space $\mathbb{R}^p_n$ into three pairwise orthogonal subspaces:

$$\mathbb{R}^p_n = \mathcal{L}_1 \oplus \mathcal{L}_2 \oplus \mathcal{L}^\perp.$$

Following theorem 5 the random matrices

$$S_1 := \langle \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X}\rangle \qquad \text{and} \qquad S_2 := \langle \mathrm{proj}_{\mathcal{L}_2}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2}\mathcal{X}\rangle$$

are independent and have Wishart distributions. Regardless of H

$$S_1 = \langle \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X}\rangle = W_p(n-m, \Sigma). \tag{2.25}$$

If the hypothesis H (2.23) is true, then

$$S_2 = \langle \mathrm{proj}_{\mathcal{L}_2}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2}\mathcal{X}\rangle = W_p(m_2, \Sigma). \tag{2.26}$$

(Here and further we denote $m = \dim \mathcal{L}$, $m_1 = \dim \mathcal{L}_1$, $m_2 = \dim \mathcal{L}_2$).

Under the alternative to H (2.23), the Wishart distribution of statistic (2.26) becomes noncentral with the parameter of noncentrality

$$\Delta = \langle \mathrm{proj}_{\mathcal{L}_2}\mathrm{E}\,\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2}\mathrm{E}\,\mathcal{X}\rangle.$$

The noncentrality parameter shows the degree of violation of the hypothesis H (2.23): $\mathrm{E}\,\mathcal{X} \in \mathcal{L}_1$.

In the one-dimensional case (when $p = 1$) the statistics (2.25) and (2.26) turn into random variables distributed as $\sigma^2\chi^2(n-m)$ and $\sigma^2\chi^2(m_2)$ respectively. Their ratio (under the hypothesis) is distributed free, and therefore it can be used as a statistical criterion for testing H. This is the well-known F-ratio of Fischer.

In the multivariate case the analogue of F-ratio should be the "ratio" of $(p \times p)$-matrices S_2 and S_1. Under $n - m \geq p$ the matrix S_1 (2.25) is non-degenerate, and therefore there exists a statistic ($(p \times p)$-matrix)

$$\langle \mathrm{proj}_{\mathcal{L}_2}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}_2}\mathcal{X}\rangle\ \langle \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X},\ \mathrm{proj}_{\mathcal{L}^\perp}\mathcal{X}\rangle^{-1} \tag{2.27}$$

Unlike the one-dimensional case ($p = 1$) the statistic (2.27) is not distributed free. By distribution, (2.27) is equal to

$$\Sigma^{\frac{1}{2}}\ W_p(m_2, I)\ W_p^{-1}(n-m, I)\ \Sigma^{-\frac{1}{2}}. \tag{2.28}$$

However the eigenvalues of matrix (2.27) under the hypothesis H (2.23) are distributed free (from $\mathcal{M}$, Σ). These eigenvalues coincide with the roots of the equation relative to λ

$$\det(W_p(m_2, I) - \lambda W_p(n-m, I)) = 0. \tag{2.29}$$

Therefore certain functions of the roots of equation (2.29) are traditionally used as critical statistics in testing linear hypotheses.

Here our investigation enters the traditional realm of multivariate statistical analysis, and therefore must be finished.

I thank E. Sukhanova, A. Sarantseva, and P. Panov for discussions and assistance.

The research is supported by RFBR, project 06-01-00454.

References

[1] **Kolmogorov, A.N.** (1946). Justification of the method of least squares, Uspekhi Matematicheskih Nauk 1, 57-70. (in Russian);

[2] **Durbin J., Kendall M.G.** The Geometry of Estimation. Biometrika, Vol.38, No.1/2. (Jun.,1951), 150-158.

[3] **Anderson, T.W.** An introduction to multivariate statistica analysis. New York: John Wiley and Sons, Inc. 1958, 374 p.

[4] **Scheffe, M.** The Analysis of Variance. New York: John Wiley and Sons, Inc. 1959, 334 p.

[5] **Roy S.N.** Some Aspects of Multivariate Analysis. Wiley, 1957.

[6] **Bilodeau M., Brenner D.** Theory of Multivariate Statistics. Springer-Verlag, 1999, 288 p

[7] **Horn R. A., Johnson C. R.** Topics in Matrix Analysis. Cambridge: Cambridge Univsity Press, 1994, 615 p.