ANTHROPOLOGIE * XXIV/2—3 * 1986

O. SOUDSKÝ, M. STLOUKAL

A STATISTICAL DESCRIPTION OF THE BEHAVIOUR OF THE MEASURES AND INDICES BASED ON THE HYPOTHESES ABOUT THE BURIAL-GROUNDS

ABSTRACT — The problem of hypothesis of the same importance of measures and indices in applications of multidimensional analyses is solved by accepting the hypotheses about dissimilarities of burial-grounds. The method of description of the behaviour of variables reflecting these dissimilarities is suggested. The step-wise discriminant analysis is used to solve the problem of rare occurrence of some variables and the problem of different size of burial-grounds. Results are applied to the group of old Slavonic burial-grounds from Czechoslovakia, namely from Mikulčice, Rajhrad, Josefov, Velké Bílovice, Lahovice and Ducové.

KEY WORDS: Discriminant analysis — Canonical variables — Early medieval burial-grounds — Old Slavonic populations.

One usually suppose a discovered material to be a random sample from some population (in the statistical sense) and some statistical tests are applied to prove dissimilarities between means and dispersions. By these tests a univariate description is usually complete.

The next step of the description usually concerns the multivariate analysis. Some heuristic or statistical methods, such as cluster analysis are applied to find whether there exist some structures among data sets (usually burial-grounds or groups of them).

To reduce the number of variables certain methods, e.g. analysis of principal components, factor analysis etc. are used. But these methods do not use any additional information, and they are usually only the first step in a further multidimensional analysis.

But wherever one describes some sets of burial-grounds he proceeds from a hypothesis about variables to a hypothesis (or conclusions) about burial-grounds. It seems that variables are presupposed to be for instance normally distributed or of the same covariance matrix or independent etc. When it is impossible to accept some of these presuppositions, one must transform the original data but always only as a means for subsequent analyses. One must not forget that in near all multidimensional analyses, when some special weights are not used (I do not mean for instance those in Mahalanobis' distance which must be used to reduce the influence of the linear dependence of variables), we positively presuppose all variables are of the same importance! However, the general validity of this hypothesis is extremely questionable.

Therefore, when we describe a group of Moravian burial-grounds, we also want to describe the behaviour of the variables connected with them. We tried this type of description on a group of Early Medieval Slavonic burial-grounds from the territory of Czechoslovakia.

In this study we wanted to accept some hypotheses about relations among burial-grounds (resulting from archaeological or geografical facts) and then to use them in describing the behaviour of variables (measures and indices).

First we should be aware of different "quality" (in the statistical sense) of our data sets. On the one hand we have large burial-grounds (in the sense of the number of well-preserved skulls), which can be either homogeneous or heterogeneous. As we shall see later, a very good example of a large homogeneous burial-ground is the set of male skulls from Rajhrad and an example of heterogeneous one is the data set from Ducové. On the other hand we have small data sets which can also be homogeneous or heterogeneous. Certainly the role of these types of burial-grounds will not be the same.

Nearly the same problems occur with variables. Some measurements are so specialized that there are only very few skulls preserved enough to comply with them. Therefore, these measures show very rare occurence and even in large burial-grounds it is difficult to make any statistical decision. Remember that simple statistics, e.g. arithmetic mean, can become an inaccurate estimate when derived from a little number of data. And the situation with indices is even worse not only because of their rare occurrence but also because their fit with normal distribution is worse what makes many statistical conclusions difficult, if not impossible, to draw.

For this study two types of behaviour of variables will be interesting. First they can be nearly constant throught all investigated burial-grounds, or they can vary through our data sets but with no differences among them and within them (nor in combinations with other variables).

The second type of variables vary sufficiently among data sets and, at the same time, they do not vary too much within data sets. These variables will be useful for comparing the studied burial-grounds but they will become a risky element as soon as we want to regard these burial-grounds as a single data set. Also here we can add "nor in combination with other variables".

When working with individual variables, the situation is not so difficult and is usually solved by tests of equality of means and variances.

In this paper we have focussed our attention on the combinations of variables of the second type of behaviour. We shall look for some combinations of variables which enable us to differentiate our burial-grounds.

We decided to define the hypothesis about burial-grounds in terms of their dissimilarity which can be presupposed owing to archaeological findings. Then the method of discriminant analysis can be applied to find a combination of variables corresponding to such a hypothesis.

For choosing the combination only large burial-grounds can be used because estimates derived from small burial-grounds are not reliable enough and we can not decide whether the failure results from lack of skulls or lack of dissimilarity. To be able to valorize our results and the reality they represent, we have applied them to all burial-grounds under study. These applications should be compared with archaeological deductions and also with other external information. We can apply our results also to individual skulls to identify some strange elements in data sets (certainly in terms of the presupposed hypothesis).

We have studied the group of Old Slavonic burial-grounds from Czechoslovakia from which individual metric data are available. We focussed our investigation on those localities from Moravia, which come from a relatively small area and are, except two cases about which we shall talk further, also accordingly dated into the 9th century. This group consists of 7 independent burial-grounds which were discovered in Mikulčice. The first 5 burial-grounds are dated into the 9th century, the 6th and 7th burial-grounds are a little younger, they reach from the 9th to the 10th century. From the 9th century is also the burial-ground in Rajhrad as well as the village burial-grounds in Josefov and Velké Bílovice.

For our intention there are 3 burial-grounds large enough, which can be assumed as distinguishable — the 2nd burial-ground in Mikulčice, which represents the population living in 9th century on Mikulčice castle, the set from 7th burial-ground, which is of some younger dating and at the same time represents the settlement outside the fortified sites of Mikulčice and, finally, the set from Rajhrad from 9th century, a significant economic centre with a magnate's court.

To make comparisons with other areas in Czechoslovakia, we have applied results also to one burial-ground from Bohemia — Lahovice, and to one large burial-ground from Slovakia — Ducové. The latter has its roots in the 9th century but the largest set of skulls comes from the high Middle Ages, from the 12th to 15th centuries and represents a further stage in the development of the European populations.

Important is the question of the initial set of variables (measures and indices). We could not leave out measures and indices belonging to those basic in anthropology. They are for exemple the length and breadth of cranium and the breadth and height of face. But it was necessary to exclude some scarce variables so that the set of skulls would not be too little for us to make any conclusion. As the aim of analysis was the discrimination of some burial-grounds, we could leave those variables and indices which had shown too rare and do not differ between the burial-grounds. We made several pair univariate comparisons of burial-grounds to get oriented in the significance and frequency of occurence of each quantity (measure and index). It is true that we resigned from the power, but even if it existed we should not be able to make any good estimation because of the scarcity of these combinations.

So we decided to choose the following set of eleven measures and eight indices out of the original set of 40 determinable measures and 13 indices. They are the most important from the anthropological point of view, and, moreover, they conform to a certain extent to the selection used by other authors.

The chosen variables are: maximum skull lenght (C 1), maximum skull breadth (C 8), minimum frontal breadth (C 9), basion-bregma height (C 17), bizygomatic breadth (C 45), nasion-prosthion height (C 48), orbital breadth and height (C 51 and C 52), nasal breadth and height (C 54 and C 55), angle breadth (C 66); indices: length-breadth (I 1), length-height (I 2), breadth-height (I 3), transversally frontoparietal (I 13), index of upper face (I 39), of orbit (I 42), of nose (I 48), and frontomandibular index according to Škerlj (IFM).

The omitted variables are: nasion-basion length (C 5), anterior forehead breadth (C 10), biauricular breadth (C 11), asterion breadth (C 12), auricular-bregma height (C 20), horizontal circumference (C 23), transverse arc (C 24), median sagittal arc (C 25), nasion-bregma arc (C 26), bregma-lambda arc (C 27), lambda-opisthion arc (C 28), nasion-bregma length (C 29), bregma-lambda length (C 30), lambda-opisthion length (C 31), cubic capacity with millet in cc (C 38), basion-prosthion length (C 40), upper facial breadth (C 43), inner orbital facial breadth (C 43/1), maxillary breadth (C 46), nasion-gnathion height (C 47) anterior interorbital breadth (C 50), upper nasal breadth (C 57), condylar breadth (C 65), chin height (C 69), height of ascending mandibular ramus (C 70), ramus breadth sin (C 71), facial index (I 38) and alveoral index (I 60).

Upon chosen variables we used the step-wise discriminat analysis as it is described in Dixon et al. 1977. We have tried many variants of parameters and we have seen that the results do not differ too much. It means that there is no danger that our results are rather caused by parameters of analysis than derived from data.

We present the variant where we wanted to find a linear combination of variables which are not too correlated. So we set the parameter of tolerance 0,4. It means that only that variable can in the next step enter the analysis, whose multiple correlation coefficient with variables yet entered does not exceed 0,6. This combination showes the best, in some way, the difference between (or among) the given burial-grounds. As we worked with three burial-grounds, we have got three pair discriminations and one discrimination of three burial-grounds. All these discriminations are represented by canonical variables.

We applied these linear combinations to original data sets (the numbers of skulls in the data sets arised because out of all initial variables only those chosen by analysis were used now). To see the behaviour of other variables we calculated correlation coefficients between all variables and canonical variables. So we obtained *Tab. 1* wich shows the relations of measures and indices of skulls found on three processed burial-grounds to "artificial" variables reflecting given dissimilarities. The analysis was performed, of course, for male and female skulls separately.

The canonical variables represent following discriminations: K 1 and K 2 (Discrim. I.) are derived from discrimination of all three burial-grounds (the 2nd burial-ground from Mikulčice, the 7th from Mikulčice and the burial-ground from Rajhrad). The first canonical variable (K 1) discriminates the best our three burial-grounds. As we can see on *Figure* 1, the second canonical variable (K 2) discriminates only Mikulčice the 7th, but not Mikulčice the 2nd and Rajhrad.

In discrimination II. we wanted to separate Mikulčice the 2nd and Mikulčice the 7th (K 3). The canonical variable K 4 is derived through discrimination of Mikulčice the 7th and Rajhrad (Dis. III) and, finally, Dis. IV designates the discrimination of Mikulčice the 2nd and Rajhrad (K 5). The canonical variables are following linear combinations of the initial variables:

a. male:

$$K\,1 = 0{,}2313 \cdot C\,48 + 0{,}12118 \cdot C\,51 + 0{,}16564 \cdot I\,2 + 0{,}04165 \cdot I\,13 + 0{,}07451 \cdot I\,48 - 0{,}01256 \cdot IFM - 38{,}74886.$$

$$K\,2 = 0{,}04798 \cdot C\,48 - 0{,}39345 \cdot C\,51 + 0{,}14573 \cdot I\,2 + 0{,}08591 \cdot I\,13 + 0{,}14676 \cdot I\,48 + 0{,}03919 \cdot IFM - 15{,}0266.$$

$$K\,3 = 0{,}52679 \cdot C\,51 - 0{,}06712 \cdot I\,2 + 0{,}0617 \cdot I\,42 - 0{,}10407 \cdot I\,48 - 0{,}03985 \cdot IFM - 12{,}60464.$$

$$K\,4 = 0{,}15901 \cdot C\,48 - 0{,}22525 \cdot C\,51 + 0{,}21685 \cdot I\,2 + 0{,}06123 \cdot I\,13 + 0{,}19085 \cdot I\,48 - 31{,}25657.$$

$$K\,5 = 0{,}2345 \cdot C\,48 + 0{,}08019 \cdot C\,52 + 0{,}18827 \cdot I\,2 + 0{,}05844 \cdot I\,13 + 0{,}09828 \cdot I\,48 - 41{,}9474.$$

b. female:

$$K\,1 = 0{,}01029 \cdot C\,8 + 0{,}100\,37 \cdot C\,45 + 0{,}22061 \cdot C\,48 - 0{,}178 \cdot C\,51 + 0{,}10809 \cdot I\,48 - 0{,}02693 \cdot IFM - 24{,}03644.$$

$$K\,2 = -0{,}14675 \cdot C\,8 + 0{,}1088 \cdot C\,45 + 0{,}05547 \cdot C\,48 - 0{,}48059 \cdot C\,51 + 0{,}04465 \cdot I\,48 - 0{,}06787 \cdot IFM + 26{,}76013.$$

$$K\,3 = 0{,}08228 \cdot C\,8 + 0{,}29289 \cdot C\,51 + 0{,}13173 \cdot C\,55 - 0{,}10307 \cdot I\,13 - 22{,}34647.$$

$$K\,4 = 0{,}13075 \cdot C\,45 + 0{,}1544 \cdot C\,48 - 0{,}54334 \cdot C\,51 - 0{,}10316 \cdot C\,66 + 0{,}18219 \cdot I\,13 + 0{,}10302 \cdot I\,48 - 12{,}52216.$$

$$K\,5 = 0{,}07239 \cdot C\,8 + 0{,}22789 \cdot C\,48 + 0{,}12067 \cdot I\,48 - 31{,}00394.$$

On Table 1. we can see: No variable has high correlation coefficient with all canonical variables. We can distinguish the following behaviour of variables

- the variable does not correlate with any canonical variable,
- the variable correlates only with the canonical variable which represents dissimilarity of one pair of burial-grounds — it corresponds only to dissimilarity between them, but not to any of the burial-grounds,
- the variable shows higher correlation whenever some burial-ground appears in discrimination and never when it is not compared. This variable can be considered specific for the given burial-ground.

Especially for male skulls: The variables which do not reflect differences among burial-grounds are C 1, C 8, C 25 to C 31, C 40, C 45, C 46, C 57, C 65, C 66, C 71, I 1, I 42, I 60, IFM.

The variables which reflect only dissimilarities are C 38 and C 70 (between Mikulčice the 2nd and Rajhrad), C 43 correlates only with discrimination of all three burial-grounds.

The variables which can be considered specific are: for Mikulčice the 2nd: C 10, C 11, C 12, C 23, C 47, C 48, C 43/1, C 52, C 55, I 38, I 39; for Mikulčice the 7th: C 51, I 2, I 3, C 52, and especially I 48; for Rajhrad: C 5, C 9, C 17, C 20, C 24, C 50, C 54, C 69, C 70, I 13.

Especially for female skulls: The variables which do not reflect differencies among burial-ground are C 1, C 5, C 17, C 25 to C 31, C 40, C 57, C 71, I 42, I 48, I 60, IFM. The variables reflecting only differencies of all three burial-grounds are C 9, C 54, C 66 and especially C 48; and differences between both burial-grounds from Mikulčice are I 1, and I 3.

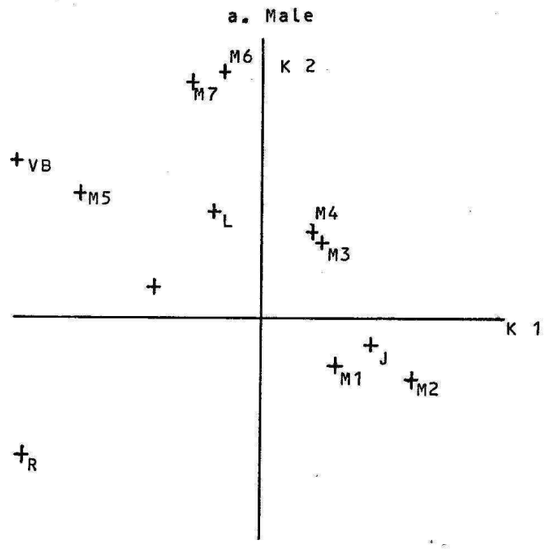
The variables considered as specific are: for Mikulčice the 2nd: C 8, C 10, C 11, C 12, C 20, C 23, C 24, C 38, C 43, C 43/1, C 45, C 46, C 47, C 52, C 55, C 65, C 70; for Mikulčice the 7th: C 51 and I 13; and for Rajhrad: C 50, C 69, I 39.

To describe the position of other burial-groups we apply our canonical variables to all data sets and computed arithmetic means, wich we drown on *Figure 1.* for the discriminations of all three burial-grounds and *Figure* 2. for pair discriminations. As we can see the burial-ground from Ducové do not show any similarity with one of three analysed burial-grounds and Lahovice shows similar behaviour with Mikulčice the 7th. The burial-grounds from Velké Bílovice and Mikulčice the 5th are too little and heterogeneous in canonical variables, that we can not accept the estimation of arithmetic mean to be representing the position of the burial-ground well. Another

TABLE 1. *Correlation coefficients between measures and indices and derived canonical variables*[1]

Measure or index	a. Male skulls					b. Female skulls				
	Discrim. I.		Dis. II.	Dis. III.	Dis. IV.	Discrim. I.		Dis. II.	Dis. III.	Dis. IV.
	K 1	K 2	K 3	K 4	K 5	K 1	K 2	K 3	K 4	K 5
C 1*	0.218	—0.160	0.229	—0.036	0.122	0.244	—0.069	0.152	0.098	0.265
C 5	0.427	0.032	0.118	0.279	0.348	0.233	—0.099	0.226	0.047	0.211
C 8*	0.114	—0.267	0.201	—0.078	0.094	0.515†	—0.569†	0.760†	—0.072	0.602†
C 9*	0.454	0.102	0.195	0.375	0.412	0.527	—0.125	0.252	0.309	0.474
C 10	0.279	—0.093	0.234	0.123	0.237	0.471	—0.282	0.449	0.128	0.496
C 11	0.452	—0.221	0.342	0.034	0.400	0.572	—0.401	0.664	0.040	0.577
C 12	0.300	—0.232	0.311	—0.019	0.26-	0.415	—0.227	0.436	0.099	0.412
C 17*	0.450	0.380	—0.148	0.594	0.468	0.197	—0.195	0.232	—0.027	0.223
C 20	0.369	0.308	—0.124	0.447	0.356	0.324	—0.142	0.216	0.118	0.360
C 23	0.285	—0.136	0.229	0.037	0.210	0.499	—0.300	0.463	0.098	0.541
C 24	0.344	0.229	—0.058	0.392	0.343	0.439	—0.340	0.441	0.047	0.508
C 25	0.236	0.127	—0.015	0.245	0.228	0.289	—0.149	0.217	0.095	0.357
C 26	0.224	0.170	—0.055	0.280	0.212	0.344	—0.080	0.177	0.168	0.364
C 27	0.081	0.125	—0.053	0.150	0.065	—0.001	—0.127	0.083	—0.080	0.054
C 28	0.124	—0.049	0.059	0.015	0.118	0.174	0.001	0.067	0.119	0.211
C 29	0.250	0.064	0.046	0.197	0.206	0.323	—0.124	0.206	0.103	0.368
C 30	0.137	0.083	—0.012	0.141	0.102	0.005	—0.109	0.078	—0.060	0.056
C 31	0.177	0.048	0.006	0.135	0.179	0.160	—0.113	0.128	0.024	0.190
C 38	0.378	—0.070	0.173	0.198	0.335	0.472	—0.446	0.604	—0.021	0.554
C 40	0.173	—0.016	—0.008	0.113	0.093	0.295	0.060	0.080	0.193	0.264
C 43	0.463	0.005	0.242	0.275	0.388	0.573	—0.240	0.418	0.198	0.556
C 43/1	0.338	—0.057	0.240	0.154	0.265	0.460	—0.269	0.390	0.104	0.465
C 45*	0.312	—0.070	0.183	0.130	0.276	0.690†	—0.285†	0.601	0.169†	0.599
C 46	0.160	0.212	—0.130	0.272	0.165	0.408	—0.208	0.305	0.083	0.418
C 47	0.707	—0.261	0.432	0.103	0.679	0.651	—0.213	0.495	0.152	0.683
C 48*	0.811†	—0.296†	0.518	0.135†	0.783†	0.805†	—0.140†	0.508	0.289†	0.821†
C 50	0.451	0.309	—0.063	0.475	0.455	0.546	0.126	0.096	0.438	0.463
C 51*	0.382†	—0.651†	0.759†	—0.281†	0.226	0.189†	—0.734†	0.694†	—0.493†	0.364
C 52*	0.321	—0.442	0.703	—0.206	0.364†	0.377	—0.189	0.389	0.045	0.364
C 54*	0.425	0.345	—0.126	0.549	0.391	0.425	—0.038	0.191	0.197	0.504
C 55*	0.594	—0.450	0.598	—0.122	0.555	0.460	—0.177	0.556†	0.067	0.475
C 57	0.122	0.280	—0.194	0.293	0.124	0.276	0.137	0.045	0.331	0.245
C 65	0.194	0.006	0.057	0.127	0.155	0.495	—0.170	0.347	0.140	0.436
C 66*	0.099	0.130	—0.061	0.064	0.109	0.330	—0.327	0.238	—0.148†	0.365
C 69	0.529	0.041	0.073	0.282	0.547	0.500	—0.002	0.197	0.252	0.514
C 70	0.426	—0.122	0.195	0.130	0.389	0.465	—0.107	0.309	0.133	0.446
C 71	0.084	—0.006	—0.009	0.034	0.036	0.266	0.075	0.018	0.196	0.178
I 1*	—0.070	—0.098	—0.009	—0.036	—0.015	0.268	—0.419	0.509	—0 117	0.312
I 2*	0.237†	0.518†	—0.344†	0.599†	0.351†	0.004	—0.163	0.102	—0.126	0.001
I 3*	0.269	0.487	—0.262	0.526	0.303	—0.260	0.335	—0.412	0.066	—0.284
I 13*	0.340†	0.296†	0.039	0.410†	0.320†	0.021	0.399	—0.433†	0.352†	—0.081
I 38	0.460	—0.186	0.272	0.028	0.467	0.160	—0.014	0.052	0.027	0.241
I 39*	0.599	—0.226	0.377	0.070	0.599	0.451	0.020	0.160	0.208	0.493
I 42*	0.076	—0.035	0.224†	—0.030	0.214	0.229	0.327	—0.145	0.383	0.098
I 48*	—0.070†	0.589†	—0.517†	0.530†	—0.067†	0.035†	0.097†	—0.234	0.123†	0.088†
I 60	—0.176	—0.047	—0.112	—0.128	—0.221	0.144	0.200	—0.143	0.219	0.119
IFM*	—0.244†	0.045†	—0.196†	—0.220	—0.219	—0.070†	—0.222†	0.032	—0.369	0.004
K 1	—	—0.002	0.379	0.544	0.976	—	—0.015	0.468	0.572	0.933
K 2	—0.002	—	—0.874	0.816	0.148	0.015	—	—0.798	0.800	—0.281
K 3	0.379	—0.874	—	—0.503	0.266	0.468	—0.798	—	—0.367	0.636
K 4	0.544	0.816	—0.503	—	0.643	0.572	0.800	—0.367	—	0.297
K 5	0.976	0.148	0.266	0.643	—	0.933	—0.281	0.636	0.297	—

[1]* *designates the members of initial set of variables* † *designates the variables used in the linear combinations.*



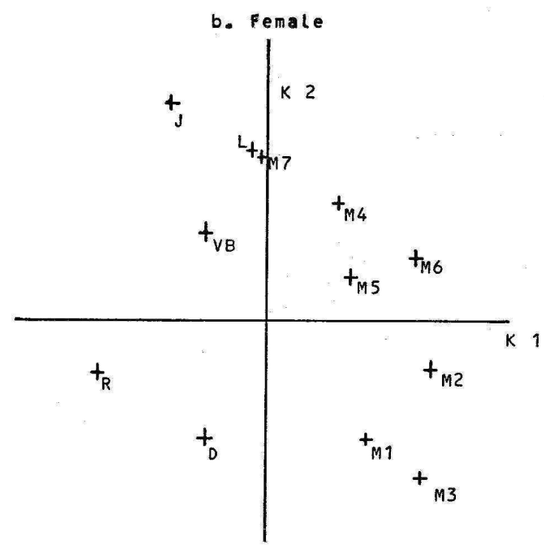


FIGURE 1. *Arithmetic means of canonical variables derived trough discrimination of M2, M7 and R. M1 to M7 ... burial-grounds Mikulčice the 1st to Mikulčice the 7th, R ... Rajhrad, D ... Ducové, J ... Josefov, L ... Lahovice, VB ... Velké Bílovice.*

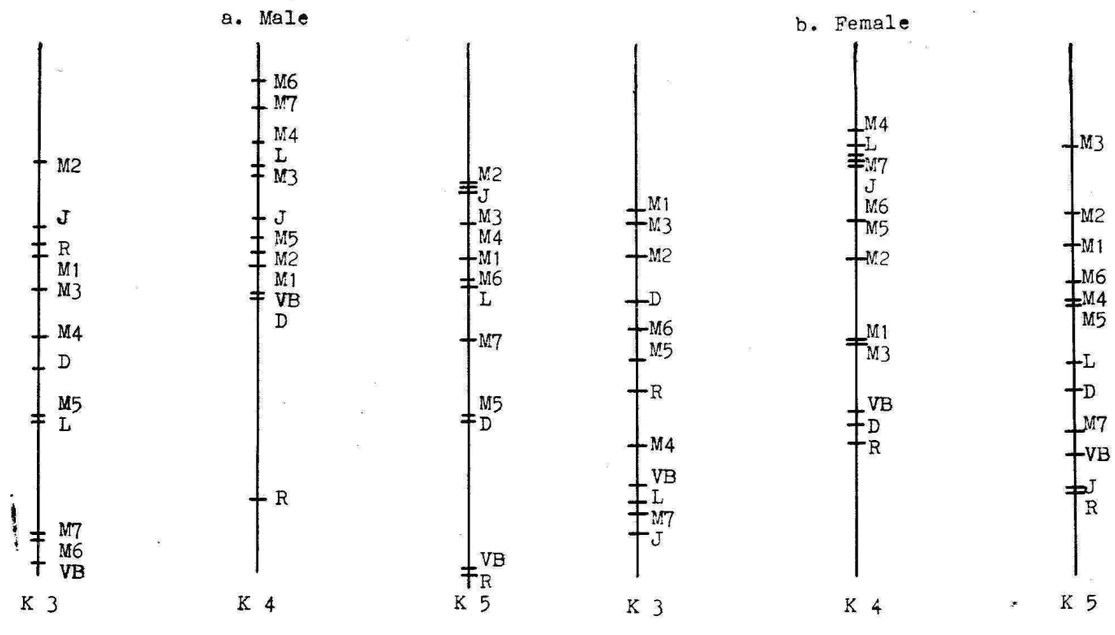


FIGURE 2. *Arithmetic means of canonical variables derived through pair discriminations.*

situation is with the burial-grounds from Mikulčice the 1st and the 3rd. They show similarity with Mikulčice the 2nd. The similarity with Mikulčice the 7th is perceptible for male skulls on Mikulčice the 6th. The position of burial-ground from Josefov is very different for male skulls, where it is similar to Mikulčice the 2nd, and for female skulls, where it is surely different from Mikulčice the 2nd. As we can see on Figure 1. the burial-ground from Rajhrad is much more isolated then burial-grounds from Mikulčice. The little burial-grounds from Mikulčice lays all in the space between Mikulčice the 2nd and the 7th.

The situation described by *Figures 1. and 2.* was also confirmed by the pair multivariate tests of equality of means applied to variables entered into the discriminant functions (and in canonical variables), which are derived from Mahalonobis' distance between burial-grounds. All burial-grounds, which are far from each other shows also a significant difference between means, with exception of burial-grounds from Velké Bílovice and Mikulčice the 5th.

The question of similarity or dissimilarity of the old slavonic populations on the territory of Czechoslovakia is a very important one, because it could make clear the problem of relationship and other respects between different settlements as well as between the streams of slavonic colonisation of this region. The soluting of this problem is not easy, being complicated by the fact that only skeletal series from the time some 4 centuries after Slavonic expansion are available. The results of their symbiosis with other groups as well as among themselves in this period manifest in the physical character of the inhabitants. There is a large number of sets at disposal for this research but in spite of all that it is not enough because of their different dating, their different geographical and social position and, from the statistical point of view, also because of the different size of the sets. The results of each attempt at such a comparison can be strongly influenced by the choice of samples as well as variables. We have presented here one possibility of this comparison, one of many possibles concerning the choice of samples, variables and methods. The further work and further experiments will show, whether our results in this paper are correct, whether they truely express the searched relations.

REFERENCES

DIXON W. J., BROWN M. B., 1977: BMDP — 77, Biomedical Computer Programs P-Series. University of California Press.

HOWELLS W. W., 1973: Cranial Variation in Man, Harvard University Cambridge, Massachusetts, Vol. 67.

PIETRUSEWSKY M., 1981: Cranial Variation in Early Metal age Thailand and Southeast Asia Studied by Multivariate Procedures. Homo, 32. Band, 1. Heft, pp. 1, 26.

Ing. O. Soudský
General Computing Centre
of ČSAV
Pod vodárenskou věží 2
182 07 Praha 8

Dr. Milan Stloukal, CSc.
National Museum in Prague
Vítězného února 74
116 79 Praha 1