"HENRI COANDA"
AIR FORCE ACADEMY
ROMANIA

GERMANY

"GENERAL M.R. STEFANIK"
ARMED FORCES ACADEMY
SLOVAK REPUBLIC

INTERNATIONAL CONFERENCE of SCIENTIFIC PAPER
AFASES 2011
Brasov, 26-28 May 2011

VALUES AND PROPER VECTORS USED IN ECONOMIC-STATISTICS ANALYSIS

Cornelia GABER

PETROLEUM-GAS University of Ploieşti, România

Abstract: *The study of a statistics population through a poll aims at more topics of great interest. The answers to the used questionnaire contain some items of information (factors) among whom there are dependence connections. The methods of the multivariate analyses allow the reduction of the number of the initial factors, such a method is the factorial analysis.*

The paper presents the correspondence analysis as a factorial analysis method, the study being made on the basis of the values and proper vectors associated to a matrix resulted from the statistical units distribution according to two qualitative variables.

Mathematics Subject Classification 2010: *62H25.*
Keywords: *contingence*, *covariance matrix*, *principal component analysis*, *vectors*.

The socio-economical problems or the problems specific to the phenomena and processes which take place in nature can be solved using numerical or qualitative characteristics. The models methods used in the definition and analysis of the qualitative characteristics are varied and numerous, depending on the specific of the approached problems.

In any scientific research we resort to partial observations from which we can deduce general truths, the poll method proving to be fundamental.

As a result of a poll among the managers of the societies from the following field activities: working industry, commerce and buildings, one of questionnaire questions refers to the actions meant to counteract the impact of the present economical situation on the developed activities. The proposed variants for the actions would be: the increase of the prices/catered services, the reduction of the expenses, the reduction of production, the cut in personnel, loans, the postponing of investments.

The results of the answers to one of the given questions are presented in a contingence table, the data analysis being made on the basis of the correspondence analysis as factorial analysis method.

Contingence table

Nature of actions	Activity fields			Total
	working industry	commerce	buildings	
0	**1**	**2**	**3**	**4**
increase of the prices	5	7	7	19
reduction of the expenses	25	39	26	90
reduction of production	8	5	7	20
cut in	10	21	23	54

"HENRI COANDA"
AIR FORCE ACADEMY
ROMANIA

GERMANY

"GENERAL M.R. STEFANIK"
ARMED FORCES ACADEMY
SLOVAK REPUBLIC

INTERNATIONAL CONFERENCE of SCIENTIFIC PAPER
AFASES 2011
Brasov, 26-28 May 2011

personnel				
loans	6	10	9	25
postponing of investments	20	15	13	48
	74	97	85	256

The contingence table $(N)=(n_{ij})_{\substack{i=\overline{1,6}\\ j=\overline{1,3}}}$

where n_{ij} represents the number of managers from the sample who chose the answers to the two characteristics:

X = actions meant to counteract the impact on the present economical situation and

Y = activity field.

The model of factorial analysis defines the matrixes:

$$D_1 = X^T \cdot X = \begin{pmatrix} 19 & 0 & 0 & 0 & 0 & 0 \\ 0 & 90 & 0 & 0 & 0 & 0 \\ 0 & 0 & 20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 54 & 0 & 0 \\ 0 & 0 & 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 48 \end{pmatrix} \text{ and}$$

$$D_2 = Y^T \cdot Y = \begin{pmatrix} 74 & 0 & 0 \\ 0 & 97 & 0 \\ 0 & 0 & 85 \end{pmatrix}$$

The matrix for the lines profiles is:

$$D_1^{-1} \cdot N = \left(\frac{n_{ij}}{n_{i\bullet}} \right)_{\substack{i=\overline{1,6}\\ j=\overline{1,3}}} =$$

$$= \begin{pmatrix} 0{,}2632 & 0{,}3684 & 0{,}3684 \\ 0{,}2778 & 0{,}4333 & 0{,}2889 \\ 0{,}4 & 0{,}25 & 0{,}35 \\ 0{,}1852 & 0{,}3889 & 0{,}4259 \\ 0{,}24 & 0{,}4 & 0{,}36 \\ 0{,}4167 & 0{,}3125 & 0{,}2708 \end{pmatrix}, \quad \sum_{j=1}^{3} \frac{n_{ij}}{n_{i\bullet}} = 1$$

And the matrix for the columns profiles is:

$$D_2^{-1} \cdot N = \left(\frac{n_{ij}}{n_{\bullet j}} \right)_{\substack{i=\overline{1,6}\\ j=\overline{1,3}}} =$$

$$= \begin{pmatrix} 0{,}0676 & 0{,}0722 & 0{,}0824 \\ 0{,}3378 & 0{,}4021 & 0{,}3059 \\ 0{,}1081 & 0{,}0515 & 0{,}0824 \\ 0{,}1351 & 0{,}2165 & 0{,}2706 \\ 0{,}0811 & 0{,}1031 & 0{,}1058 \\ 0{,}2703 & 0{,}1546 & 0{,}1529 \end{pmatrix}, \quad \sum_{i=1}^{6} \frac{n_{ij}}{n_{\bullet j}} = 1$$

The statistical hypotheses to check the independence of the two characteristics X and Y are:

$$H_0 = \left\{ \frac{n_{ij}}{n} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}, i = \overline{1,6},\ j = \overline{1,3} \right\}$$

$$H_1 = \left\{ \frac{n_{ij}}{n} \neq \frac{n_{i\bullet} \cdot n_{\bullet j}}{n}, i = \overline{1,6},\ j = \overline{1,3} \right\}$$

And the test to be used is test χ^2 for which we determine the calculated and theoretical value.

$$\chi^2_{calculat} = \sum_{i=1}^{6} \left(\sum_{j=1}^{3} \frac{\left(n_{ij} - \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} \right)^2}{\frac{n_{i\bullet} \cdot n_{\bullet j}}{n}} \right) =$$

$$= \sum_{i=1}^{6} \left[\frac{\left(n_{i1} - \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} \right)}{\frac{n_{i\bullet} \cdot n_{\bullet j}}{n}} + \frac{\left(n_{i2} - \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} \right)}{\frac{n_{i\bullet} \cdot n_{\bullet j}}{n}} + \right.$$

$$+\frac{\left(n_{i3}-\frac{n_{i\bullet}\cdot n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet}\cdot n_{\bullet j}}{n}}\Bigg] = 10{,}65538$$

For a risk $\alpha = 5\%$ the theoretical value from the distribution table χ^2 having $(r-1)(q-1)=(6-1)(3-1)=10$ degrees of freedom is $\chi^2_{1-\alpha;(r-1)(q-1)} = \chi^2_{0{,}95;10} = 3{,}94$ and, as $\chi^2_{\text{calculat}} = 10{,}65528 > \chi^2_{0{,}95;10}$ means that the H_0 hypothesis is rejected, therefore the two qualitative variables are correlated.

As the variables are correlated it is necessary to do an analysis of the main components either for the lines profile, or the columns profile.

We define g_r weight center of the lines profiles by $g_r = \left(\frac{n_{\bullet j}}{n}\right)^T_{j=\overline{1,3}}$ that is $g_r = \left(\frac{74}{256}\ \frac{97}{256}\ \frac{85}{256}\right)^T$ and the weight center of the columns profiles is $g_q = \left(\frac{n_{i\bullet}}{n}\right)^T_{j=\overline{1,6}}$ that is

$$g_q = \left(\frac{19}{256}\ \frac{90}{256}\ \frac{20}{256}\ \frac{54}{256}\ \frac{25}{256}\ \frac{48}{256}\right)^T.$$

The total inertia calculated for the two weight centers has the expression:

$$I(g_r) = I(g_q) = \frac{1}{\sum_{i=1}^{6}\sum_{j=1}^{6} n_{ij}} \cdot \chi^2_{\text{calculat}} =$$

$$= \frac{10{,}65538}{256} = 0{,}04162.$$

As the obtained result is small we draw the conclusion that the two qualitative characteristics are independent.

For the correspondence of the lines profiles analysis we build the centered data matrix noted with B such as:

$$B = \left(\frac{n_{ij}}{n_{i\bullet}} - \frac{n_{\bullet j}}{n}\right)_{\substack{i=\overline{1,6}\\ j=\overline{1,3}}} =$$

$$= \begin{pmatrix} -0{,}0259 & -0{,}0105 & 0{,}0364 \\ -0{,}0113 & 0{,}0544 & -0{,}0431 \\ 0{,}1109 & -0{,}1289 & 0{,}018 \\ -0{,}1039 & 0{,}01 & 0{,}0939 \\ -0{,}0491 & 0{,}0211 & 0{,}028 \\ 0{,}1276 & -0{,}0664 & -0{,}0612 \end{pmatrix}.$$

The covariance matrix V, as compared to the weight center of the lines profiles g_r is:

$$V = \frac{1}{n}\cdot B^T D_1 B = \frac{1}{256}\cdot B^T D_1 B =$$

$$= \frac{1}{256}\begin{pmatrix} 1{,}69495 & -0{,}82475 & -0{,}52525 \\ -0{,}25295 & 0{,}82890 & -0{,}00412 \\ -0{,}8702 & -0{,}00415 & 0{,}87435 \end{pmatrix}$$

The reverse of the matrix D_2 is:

$$D_2^{-1} = \frac{1}{\det D_2}\cdot D_2^* =$$

$$= \frac{1}{74\cdot 97\cdot 85}\begin{pmatrix} 97\cdot 85 & 0 & 0 \\ 0 & 74\cdot 85 & 0 \\ 0 & 0 & 74\cdot 97 \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{1}{74} & 0 & 0 \\ 0 & \frac{1}{97} & 0 \\ 0 & 0 & \frac{1}{85} \end{pmatrix}.$$

The normalized proper vectors of the matrix $nV \cdot D_2^{-1}$ are the main axes $a_i, i = 1,2,3$.

The characteristic polynomial of the matrix $nV \cdot D_2^{-1}$ is:

$$\left| nV \cdot D_2^{-1} - \lambda I_3 \right| = P(\lambda)$$

And the solutions of the equation $P(\lambda) = 0$ are the proper values of the matrix $nV \cdot D_2^{-1}$:

$$nV \cdot D_2^{-1} =$$

$$= 256 \cdot \frac{1}{256} \begin{pmatrix} 1,69495 & -0,82475 & -0,52525 \\ -0,25295 & 0,82890 & -0,00412 \\ -0,8702 & -0,00415 & 0,87435 \end{pmatrix} \cdot$$

$$\cdot \begin{pmatrix} \frac{1}{74} & 0 & 0 \\ 0 & \frac{1}{97} & 0 \\ 0 & 0 & \frac{1}{85} \end{pmatrix} =$$

$$= \begin{pmatrix} 0,02290 & -0,01115 & -0,00710 \\ -0,00261 & 0,00854 & -0,00004 \\ -0,01024 & -0,00005 & 0,01029 \end{pmatrix}$$

$$P(\lambda) = \left| 256V \cdot D_2^{-1} - \lambda I_3 \right| =$$

$$= \begin{vmatrix} 0,02290 - \lambda & -0,01115 & -0,00710 \\ -0,00261 & 0,00854 - \lambda & -0,00004 \\ -0,01024 & -0,00005 & 0,01029 - \lambda \end{vmatrix} \Rightarrow$$

$P(\lambda) = -\lambda^3 + 0,04173\lambda^2 - 0,000333\lambda$ with the solutions $\lambda_1 = 0,03098$, $\lambda_2 = 0,01075$, $\lambda_3 = 0$.

For $\lambda_1 = 0,03098$ we obtain the main axis $a_1 = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$ solving the equations $nV \cdot D_2^{-1} \cdot a_1 = \lambda_1 a_1$ and for $\lambda_2 = 0,01075$ we get the main axis $a_2 = \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \end{pmatrix}$ solving the equation $nV \cdot D_2^{-1} \cdot a_2 = \lambda_2 a_2$.

The vectors $a_1 = \begin{pmatrix} -1,04378 \\ 0,11962 \\ 1,00000 \end{pmatrix}$ respectively $a_2 = \begin{pmatrix} 0,27246 \\ -0,33987 \\ 1,00000 \end{pmatrix}$ are not rectangular, that is $a_1 \cdot a_2 \neq 0$ and according to the Gram-Schmidt procedure the normalized proper vectors corresponding to the vectors a_1 and a_2 are:

$$V_1 = \frac{a_1}{\|a_1\|} = \begin{pmatrix} -0,71963 \\ 0,08247 \\ 0,68945 \end{pmatrix}, V_2 = \begin{pmatrix} 0,61565 \\ -0,38342 \\ 0,68846 \end{pmatrix}.$$

As $\lambda_3 = 0$ it means the corresponding proper vector is not interesting .

For the proper values we determine the "explained inertia" percentage such as:

– for

$$\lambda_1 = \frac{0,03098}{0,03098 + 0,01075} \cdot 100 = 74,24\%$$

– for

$$\lambda_2 = \frac{0,01075}{0,03098 + 0,01075} \cdot 100 = 25,76\%$$

– for $\lambda_3 : 0\%$,

therefore the factorial values associated to the proper values λ_1, λ_2 explain the whole inertia.

The main factors $u_i, i = \overline{1,3}$ are the proper vectors of the matrix $n \cdot D_2^{-1} V = D_2^{-1} N' D_1^{-1} N$ determined from the equations $nD_2^{-1} V u_i = \lambda_i \cdot u_i$, $i = \overline{1,3}$, where:

$$n \cdot D_2^{-1} V = D_2^{-1} N' D_1^{-1} N =$$

$$= \begin{pmatrix} 0{,}31199169 & 0{,}36774672 & 0{,}32026159 \\ 0{,}28056804 & 0{,}38745276 & 0{,}33197920 \\ 0{,}27884725 & 0{,}37884022 & 0{,}34231253 \end{pmatrix}.$$

We determine the main factors:

$$n \cdot D_2^{-1} V u_1 = 0{,}03098 u_1 \Rightarrow u_1 = \begin{pmatrix} -2{,}6298 \\ 1{,}1386 \\ 0{,}9699 \end{pmatrix}$$

$$n \cdot D_2^{-1} V u_2 = 0{,}01075 u_2 \Rightarrow u_2 = \begin{pmatrix} 0{,}1380 \\ -0{,}9842 \\ 0{,}9768 \end{pmatrix}$$

And the main components are:

$$C^1 = D_1^{-1} N\, n D_2^{-1} u_1 = \begin{pmatrix} -0{,}2114 \\ -0{,}3814 \\ -1{,}8655 \\ 0{,}7278 \\ 0{,}0701 \\ -2{,}0610 \end{pmatrix}$$

$$C^2 = D_1^{-1} N\, n D_2^{-1} u_2 = \begin{pmatrix} 0{,}2525 \\ -0{,}1430 \\ 0{,}5712 \\ 0{,}3312 \\ 0{,}1347 \\ 0{,}1839 \end{pmatrix}$$

Matrix $C = (C_1 \; C_2)$ contains the coordinates of the 6 weights line: the first column presents the corresponding components of the first factorial axis, the second column offers the components for the second axis.

The weight- columns are obtained from the weight-lines coordinates using the congruence:

$$Z = \Lambda \cdot C' \left(N D_2^{-1} \right)$$

where $\Lambda = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{pmatrix}$,

Where $Z = \begin{pmatrix} -4{,}5333 & -2{,}378 & -2{,}2646 \\ 0{,}1359 & 0{,}1042 & 0{,}1561 \end{pmatrix}$.

The components of the three weight-columns situated on the first line from Z represents the coordinate of the first factorial axis, the second line representing the coordinate of the second axis.

On the same factorial level we graphically represent the six types of actions meant to counteract the impact of the present economical situation and the three field activity.

The graphic representation on the factorial level is interpreted as follows:

- On the level of the action categories ensemble meant to counteract the impact of the present economical situation, the bond (association) indicates a similarity of frequencies form the point of view of the activity fields. It is the case of managers who have chosen as actions the expenses reduction (Y_2) respectively performing some loans (Y_5), affirmation confirmed by the corresponding percentages very close to the second and respectively fifth line from the matrix $D_1^{-1} N$.

- On the level of the ensemble of answer variants, the bond or association between the two weight-column representing two categories relative to the field activities, indicates a similarity of the frequencies regarding the mangers' distribution from the analyzed sample. This is achieved between the working industry (Z_1) and buildings (Z_3).

- Simultaneously for the two ensembles we assume taking into account the weight lines and the weight-columns to identify the groups (classes) which are responsible for certain associated. For example the managers who have chosen as action reducing the expenses (Y_2) respectively performing some loans (Y_5) come from the working industry (Z_1) and buildings (Z_3).

CONCLUSION

After the analysis performed, the managers' actions represent an important factor of influence on the present economical situation, regardless the activity field from where the companies whose managers they are.

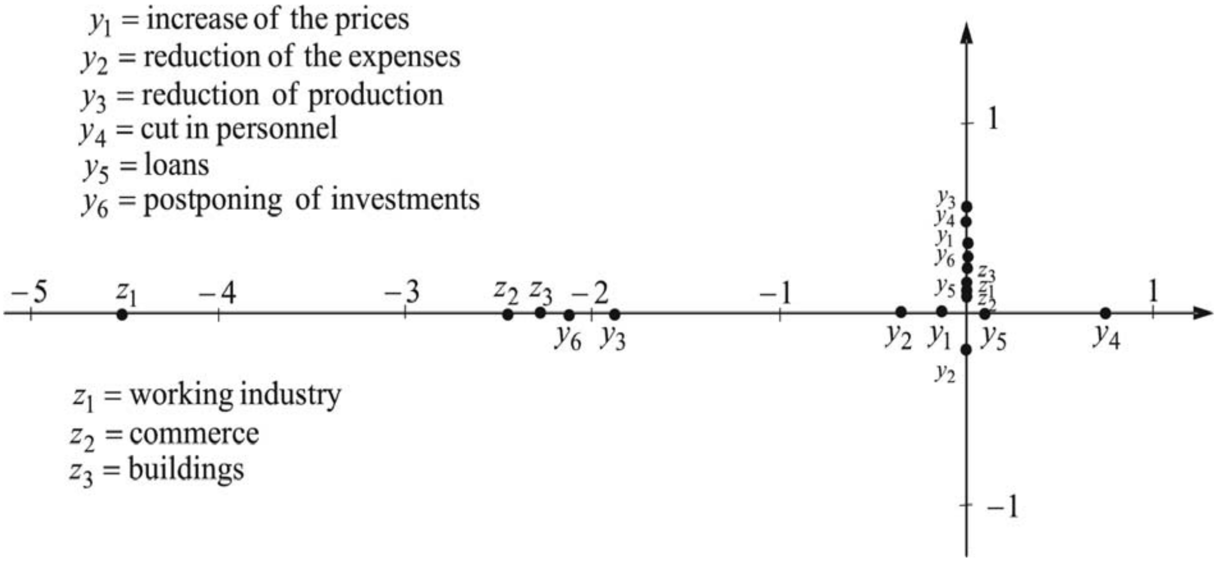

Fig.1 The graphic representation on the same factorial level of the categories of actions and activity fields.

REFERENCES

1. Dumitrescu Monica, „*Statistics polls and applications*", Technical Publishing House, Bucharest, 2000.

2. Gaber Cornelia, „*The poll – method of investigating the mass phenomena*", Ploieşti University Publishing House, 2005, pg.167-172, 228-234.

3. Isaic Maniu Al., Mitruţ C., Voineagu V., „*Statistics for business management*", Economical Publishing House, Bucharest, 1999, pg. 323-351.

4. Marinescu I., „*Analiza factorială*", Scientific and Encyclopedic Publishing House, Bucharest, 1984.

5. Maxim P.S., „*Quantitative research methods in the social sciences*", Oxford, 1999, p.233-250.

6. Voineagu V., Furtună F., Voineagu M.E., Ştefănescu C., "*Factorial analysis of the social and economic phenomena in the regional profile*", RAMIS SRL Publishing House, Bucharest, 2002.