JOURNAL OF APPLIED QUANTITATIVE METHODS

COMPARATIVE MULTIDIMENSIONAL ANALYSIS OF WHEAT PRODUCTION AT COUNTY LEVEL

Aura POPA
PhD Candidate, Department of Statistics and Econometrics,
University of Economics, Bucharest, Romania

Email: aura.popa@csie.ase.ro,
Web page: www.aurapopa.ase.ro

Abstract: *This article presents a series of multidimensional analysis methods on agricultural factors at county level, performing subsequently a comparison between the results obtained in 2004 and, respectively, in 2008. By applying factor analysis, principal components analysis and hierarchical classes analysis, this article will seek to highlight the main variables that influence the agricultural production of grain, by territorial grouping by county, and then forming some specific county typologies for Romanian wheat production.*

Key words: *agriculture; multidimensional data analysis; cluster analysis; factor analysis; county typologies for wheat production*

Introduction

Multivariate analysis used to study wheat cultivation in many articles, both in terms of its taxonomy, agronomy or agrometeorology, in the habitat assessment and the conduct of their parameters change scenarios at the limits considered normal. Such studies have been conducted in agriculture for decades, but in particular only inside experimental models, as well as those of **Waugh (1942)** which shows an example consisting of two sets of variables describing the type of spring wheat from Canada and the flour characteristics that resulted, or **Walton (1971)** who conducted studies on wheat genetics, or more recently Dixon (2006) on the same curriculum. In combination with engineering and cutting-edge technologies, it attempts to study production at the spatial level, by topographying and scanning the biomass, but with the latter it showed major disadvantages of data collection, yet unimproved. From a multidimensional perspective, **Moayedi et al. (2010)** investigated the effect of irrigation regimes during different phases of growth and development on yield and yield components, for some genotypes of wheat intended exclusively for bakery uses. Examples may continue, but it is to be remembered that the majority has as units of analysis the parcel, or small areas, homogeneous in terms of descriptors of natural and technical endowments utilised.

Several national and regional analysis, but also as derived from the literature, reveals that in the case of agriculture, matters must be considered more closely. The attempt to analyze at a certain level—even if it is at regional, macro-regional or national—makes many "players" lose their important role. Therefore, it is necessary to examine this issue at the county level, although the ideal, but nevertheless hypothetical, the most pertinent analysis is at the micro level, i.e. parcel level, of the area under cultivation. However, this would only be possible with the support of the implementation of a geographic information system software, a topographic database with detailed information about the soil type, meteorological situationtype of crop on each parcel of land, the amount and the type of fetilizers that are used, agro-technicalmethods implemented throughout the cultivation period, mapping the network of water etc. Because some of these data sources are not freely available, and others are not collected in an existing database, we must apply a multidimensional analysis at county level which is intended to outline for the whole group some areas for this specific analysis

The agricultural sector needs and uses more and more raw materials to improve its production and, therefore, it becomes even more dependent on agro-industrial complex than we might think. The farmer is subjected to a process of alienation with its means of production, which in turn affects later social and financial status **(Goodman et al., 1987)**.

Farmers, especially those with less land, can not afford to lose even the smallest piece of means of income and therefore will not bypass inputs that will ensure a great deal of success (equipment, mechanical equipment, chemical fertilizer, better quality seeds etc.) **(Friedman, 1991).** Therefore, it should not surprise us that a Romanian farmer, from his little land available, will buy fertilizer or pesticide with the risk to cover costs at the limit rather than to live a mere subsistence life.

According to analysis conducted by the National Meteorological Administration from Romania, for the years 2004 and 2008, agrometeorological characteristics showed no extreme events, vegetative rhythms on both field crops and fruit-growing species have evolved, in a whole, normally overall across the country, the processes of growth and development until harvest did not register impediments, although sometimes there were some extreme phenomena locally, on small areas, but without causing major changes at the county, regional or national levels. These are some reasons why the author of this article chose these years for the analysis.

In order to observe major changes at territorial level and concerning wheat production, requires a distance of five years between the current and the base year of study. Usually, it is taken as reference years that are multiples of five. In this case, the most recent data were those for 2000 and 2005, but according to ECA&D[1] many extreme events, were recorded, particularly in April 2005, when floods affected thousands of homes in over 140 localities and 30,000 hectares of cultivated farmland were covered in water, and in May of the same year, very large quantities of rain fell over more than 130,000 hectares of arable land. Several areas were also affected in the same year, in June and July—2,300 households, and in August about 1,400 homes were destroyed with all the agricultural land affected within their surrounding area. For this reason, a comparison analysis with this year could lead to a distortion of the overall situation, and because in terms of extreme weather phenomena such events have never been applicable in Romania for more than 50 years.

Multivariate models used to analyze wheat production in agriculture

Hierarchical methods, non-hierarchical and statistical classification try to recover, as far as possible, the real structure and form of groups of factors that underly the successful cultivation of wheat. Because these methods are already known, the following part of the article will focus on the study of results and their comparison. It will be presented in a series of multivariate data analysis methods used to identify some characteristics of wheat farming in Romania.

In the first part, in the descriptive analysis, conducted using SAS[2] software, data that were used in this study are summarized, both for the year 2004 and for the year 2008, and then will be presented the results of the methods used to identify features of clustering at the county level.

Descriptive data analysis

In a typical problem of numerical classification, expression of data will be in the form of a matrix, with *n* individuals—in our case represented by counties, on which were evaluated *p* characteristics, as follows:

- FOM—civilian employed population in agriculture at NACE Rev. 2 section, at counties level (per thousand persons);
- PROD—wheat production per counties (tons);

- SUP—surface cultivated with wheat, per counties (hectares);
- Park Numbers of agricultural tractors (TR), of plows (PLU), of mechanical cultivators (CUL), of mechanical drills (SEM), of self-propelled combine harvester, respectively, straw and hay balers (PRE) for each county (number of pieces);
- Insecticides (INS), fungicides (FUN), herbicides (ERB) and pesticides (PES) in whatever form or substance or mixture of substances, including their mixtures with ingredients for use in agriculture, forestry, the storage areas and other activities in order to prevent, reduce, remove or destroy pests, pathogens, weeds and other forms of animal or plant life, including viruses, harmful to plants and domestic animals, insects and rodents carrying diseases transmissible to humans, and adjusting products affecting plant growth, defoliation or splitting them; they are reported for each county substance (kilograms);
- Chemical fertilizers (CHIM) presented as industrial products as their contents may be nitrogen (AZ), phosphate (FOS), potassium (POS); they are expressed in active substance used on arable land in the county in question (tonnes);
- NAT—natural fertilizers including manure from all species of animals and birds (fresh or fermented) and liquid slurry, which is expressed in gross weight used at the county level (tonnes).

All data were collected for year 2004 and 2008, being extracted from regional statistical yearbooks published by the Romanian National Institute of Statistics, and the newly created base was first validated to eliminate any errors caused by any incorrect data entry, followed by application of proposed methods, interpretation of results for each year separately and so then reveal significant changes at temporal comparisons.

In the literature, some authors recommend standardisation of variables as a previous step to calculate the gap between values that may be due to different metrics, in order to eliminate the effect of scale on the results of the final hierarchization.

The most common types of standardisation are those using standard deviation or with the help of ranking. Thereby, for example, **Milligan and Cooper (1988)** recommended the second option because it has a linear standardisation, being more robust in the presence of aberrant values.

For the simple descriptive analysis of data, resulted in large overall coefficients of variation within several variables taken into account, therefore, because the data that we have present no aberrant values, we express information in a stadardized form applying as first proposed by standardized values of mean zero and dispersion equal to 1.

Even though in the 2008 data there have been increases in the case of some factors such as technical equipment—the fleet enriched in average with about 620 machines, the used quantities of fertilizers increased from 1,03 in the case of nitrates—the yield of this year was with 29,43% less than for the year 2004, this being due to the particular decrease of people employed in agriculture, an absolute number of 205 thousands persons, and the diminuation with 8,09 percentage points of agricultural area cultivated with wheat.

However, this change can be made on the behalf of the previous year, 2007, when there were recorded the highest annual average temperatures, causing drought in large areas, therefore the production was very low in almost all types of grains, and increasing wheat prices on international markets led farmers to be unable to buy enough seed for sowing in the autumn of that period.

As was expected, the top 10 are the counties of Timis, Calarasi, Teleorman, Olt and Dolj are found in both years chosen for analysis, and the bottom places are counties such as: Salaj, Hunedoara, Bistrita-Nasaud, Maramures, Harghita or Sibiu (we omitted from the analysis Bucharest which has the smallest agricultural area of all the adminsitrative areas considered).

One obvious reason is the profile of agricultural land these leading counties hold, the total area of arable land in each county in part plays an important role, which have the highest weights of the total arable land in the country (for example, in 2004, in Constanta county have been cultivated 5,03% of arable land from the total farmland with wheat across the county, while in Sibiu county it was only 1,01%), but also because of the high altitudes, in many counties the presence of hills and mountains is a major influence.

Principal components analysis and factor analysis.

Factor analysis is a multivariate statistical technique aimed at extracting a small number of hidden factors (latent) responsible for correlations between the original variables. If these correlations are relevant, it is believed to be caused by the existence of one or more common „hidden" factors for all variables. Factor analysis allows us to confirm the statistical results on the group of original variables. The variables are grouped together and, as such, they may be regarded as forming a homogeneous group and will be considered related to the same factor. Multivariate analysis techniques are increasingly used throughout many areas of research, and to characterize wheat production, in terms of quantity or quality, as certain aspects of the goal.

Based on descriptive information available about the total wheat production, the area under cultivation, the agricultural machinery, the amount of pesticides and the quantity of fertilizers used in every county in Romania, for 2004 and 2008, we can identify trends in certain areas not necessarily regionally, but links between factors that have been chosen.

Application of factorial analysis on data that we have, will result in a first correlation matrix and will conclude that:

- For 2004, there are strong direct links between the components of the available agricultural vehicles between the fleet and the production of wheat, between the fleet and fertilizer, but most importantly to note is that between the production of wheat and other factors considered in the analysis. As expected, the area under this kind of crop is a main „character" (r=0,875)[3], but also the sowers (r=0,897), the combines (r=0,788), and among the fertilizers, the nitrates are emphazised (r=0,75), between the machinery and fertilizers—the sowers and nitrates, with a total variation synthesing in 90,56% between the latter variable.
- For 2008, we can say that this time, the correlation matrix have strong direct links between the same pairs of variables as in the previous year considered, but it is most important to note again the strong linkage between the production of wheat and other factors taken into account in the analysis.
- This year, the area for this kind of agriculture has the greatest impact (r = 0,989), and other variables are slowing down in effect. Tractors and plows, with a correlation of 0,983 (also confirmed the fact that, logically, as the number of tractors increased, the more there will be bought plows for them). The sowers and the combines are in a direct linkage, with a correlation of 0,929, which indicates that the number of drills once increased, ads to more drilling, and so the harvest will be bigger;
- Both for 2004 and for 2008, it is noted that the employment variables in agriculture, insecticides, fungicides and natural and potassic fertilizers are poorly correlated, so in the further analysis they will be not considered. Therefore, from the 17 factors that we considered at first, only 12 remain.

Linking fertilizers with the rest of variables considered, only in the case of nitrates, it was expected as the agro-chemical substance used in such crops is the main one, and therefore, as known from the literature, there is a quatitatively role for the final harvest.

Therefore, the new group of values associated with previous correlation matrix, will be determined by its own values. A value greater than 1 for a component indicates the fact that it has a larger contribution than that of the original value, as indicated from the extract at (Figure 1).

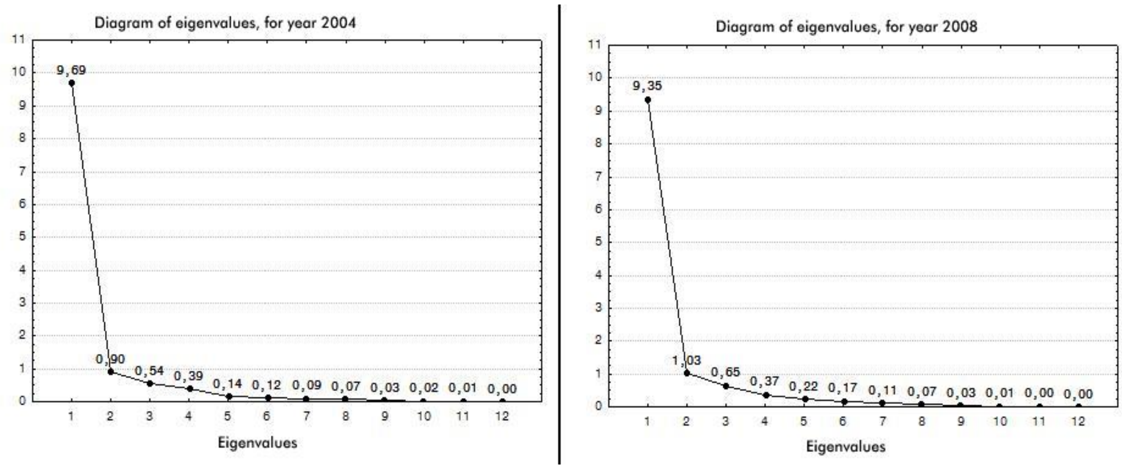

Figure 1.—Diagram of eigenvalues of correlation matrix (left for the year 2004, right for the year 2008).

From the obvious difficulty of viewing multi-dimensional space, having sizes larger than four, principal components analysis is used mostly to minimize the dimensionality of variables in at least two, but no more than three dimensions, as recommended in an example given by **Braun and Maindonald (2010).**

This method summarizes initial data variability on all the 12 factors with which we started, in number of twelve, in several unrelated parts together, called principal components. Each principal component is extracted in the linear combination of original variables, respectively with the minimum loss of information. The first principal component extracted from this combination is one that takes the maximum possible change from the original data. Thus, in both years, the biggest eigenvalue, of 9,69, it is of wheat yield, with a variance of 80,71% for the year 2004 and, respectively, of 77,88%, with an absolute value of 9,35, for year 2008. The second principal component, the area planted with wheat, takes less variance of only 7,46% (0,89) for the base year, and slightly higher in 2008, i.e. 8,54% (1,02 in absolute value), and so on for the rest of their eigenvalues that corresponds to the following considered values.

As the first principal components take more than 88,71% from the variance of initial data for the first year, and, respectively, of 86,43% for the other year of our study, then we can say that the scope to reduce the dimensionality was achieved.

The notion of score observations can be understood if we interpret the observations as vectors in an n-dimensional space of variables, as determined by **Stewart and Muller (2006).**

The last part of the principal components analysis involves the loadings's interpretation that are correlation coefficients between columns-score and original variables. Extremely important is the study of correlation coefficients between original variables and the first two principal components that emerged above. Strong coefficients show that corresponding variables can be considered responsible for the variance of the data.

Therefore, for the year 2004, we have strong correlations between the second factor and the wheat production (r=0,84) and the cultivated area (r=0,9), it also presents a powerful link between the second factor and the first variable (r=0,91) and the second one (r=0,90).

Normally, to obtain factors with small loads, that were not significant, and therefore to simplify the interpretation of factors, there is performed a rotation through the well-known varimax method.

For the remaining variables, it is not correlating with any other principal component, nor it is correlating with components that have small eigenvalues (for the year 2004, it is applicable to tractors (r=0,755) and plows (r=0,713) with the forth factor, the chemical fetilizers (r=0,767) and nitrates (r=0,712) with the third factor; for year 2008, in isolation, a correlation is revealed between the fourth main component and balers (r=0,845), and also,between the third principal component and the nitrates (r=07515))[4], this suggests that those variables have a minor contribution to the variance of the dataset, so that the “unimportant” variables will be removed in order to simplify the overall analysis.

Eigenvectors, associated eigenvalues, will be those weights in the calculation of linear combinations. Communality is part of the variance of initial variables, expressed as a percentage, which is due to all factors found.

When we assume that all communalities of the variables are equal to 100% — when all original variables are completely "explained" by factors— the actually result of the analysis coincides with the one of the principal components analysis, the results obtained in this case are:

Table 1. *Communality*

Variable	From factor 1 (in year 2004)	From factor 2 (in year 2008)
PROD	0,851	0,881
SUPR	0,822	0,905

Note: Extracton method: Principal Components Rotation: Varimax.

Factor loading coefficients form a matrix of size p x k, each element expressing the correlation between an original variable and a factor. We have p variables and latent factors k, and k <p. From the correlation matrix between original and new variables obtained, we can ascertain which variables are correlated with new factors.

Thus, for the both years considered, factor 1 is dirrectly and strong correlated with the wheat production, and the cultivated area, but weaker with the plows, seeders, sowers; the second factor is strongly correlated with the wheat yield and the area on which was cultivated, but weaker in intensity compared to other factors, and so on.

Classification methods.

Cluster analysis are designed to identify homogenous group of individuals, with similar variables that characterize them, it is necessary to define and to measure the similarity (or its complement, i.e. the distance) between two individuals or between groups of individuals, known as the proximity index. When all variables are continuous, the most commonly used method for calculating the distance between two individuals (observations) is the Euclidean or the Manhattan one. Hierarchical analysis was applied to the 42 cases, according to the two factors of the principal components analysis, production and area cultivated with wheat.

In the Figure 2, left, it can be easily observed that the biggest similarity will be found between counties that have the largest surfaces cultivated with wheat that resulted from the very simple descriptive analysis of the data from the beginning, and the high dissimilarity between these two variables proved to be the utmost importance from the previous analysis.

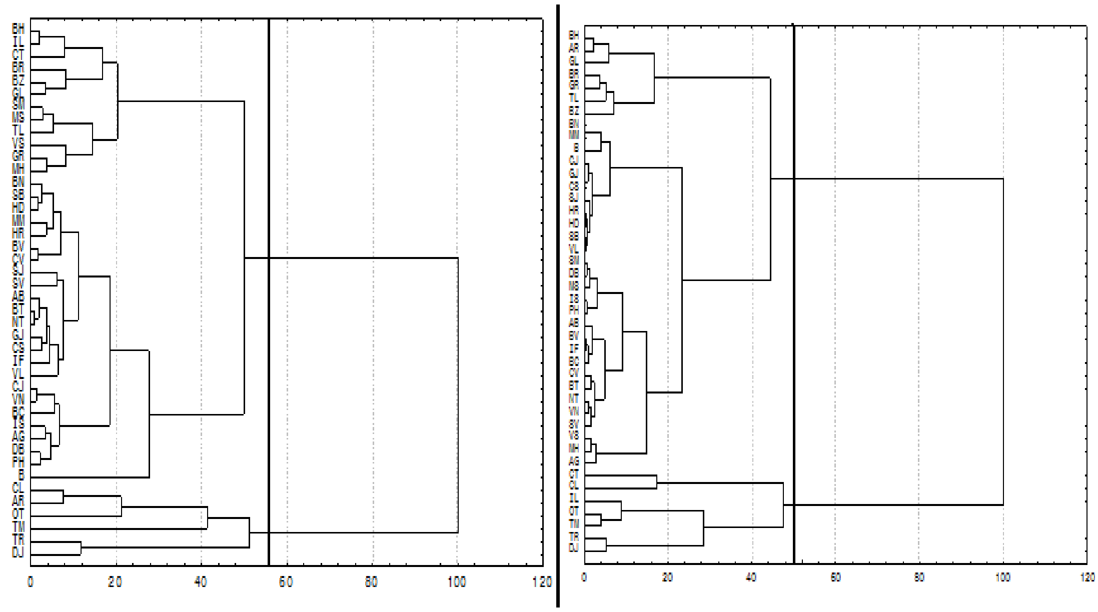

Figure 2.—Dendogram of wheat production and growing surfaces, at couties level in Romania (left for the year 2004, and right for the year 2008).

However the penultimate stage is highlighted at the end of the concatenation, that all counties by grouping according to these variables will be divided into two major groups. The second cluster consists of six counties—Calarasi, Arad, Olt, Timis, Teleorman and Dolj being grouped by similar characteristics. We re-create a dendogram for the year 2008 (Figure 2, right).

As analysists, we could cut the dendogram at the level of also two major groups, prob could be worded better but this time the steps of concatenation is achieved much earlier, and we can observe that the structure of groups, members that are forming the counties groups, changes in distance between the two years from the leading counties that places a greater emphasis this year than the year taken as a base of comparison.

Among the members of the two groups with few structural changes, then we can say that Calarasi, Teleorman, Olt or Dolj county are on top or both rankings in different years, at a distance of five years. These counties are known for their specific inclination towards agriculture and in these areas for the farmers tendency for wheat cultivation.

From the final dendograms which resulted, using cluster analysis, we can see that if we would have started analyzing at the macro-regional or at the regional level it would be lost in specificity and contribution for each county separately.

One reason for the loss of information, as much as is conducted through an overall analysis, it is due to the mismatch in territorial division that is purely administrative and not natural, within the specifics of agricultural land and crops found in each county separately.

Conclusions

The analysis in this study is based on data calculations for only two years, which have not been affected by any major influences from other factors, such as distribution of other types of crops on arable land available.

Instead, the great disadvantage of this comparison is the relatively small number of years and incomplete and unbalanced representation of all variables that are used in agricultural management, particularly the cultivation of wheat. If we could have access to a metadata base that holds all the necessary information at parcel level, this type of analysis would lead to more conclusive results. Such a database would be similar to that used by **Ziv and Goldman (1987)** for wheat cultures. In this case, the overall result indicates the potential of all multidimensional data considered in order to identify predictors that lead to increased production, and therefore profits.

References:

1. Braun, W. J. and Maindonald, J. **Data Analysis and Graphics Using R**, Cambridge University Press, UK, 2010, pp. 385
2. De Abdelmonem, A., Clark, V. and May, S. **Computer-aided multivariate analysis**, Volumul 6, USA CRC Press, Taylor & Francis, Inc., 2004
3. Dixon, J., Nalley, L., Kosina, P., La Rovere, R., Hellin, J. and Aquino, P.**Adoption and economic impact of improved wheat varieties in the developing world**,The Journal of Agricultural Science 144:6, Cambridge University Press, 2006, pp. 489-502
4. Duncan, C. J. and Scott, S. **Human Demography and Disease**, UK Cambridge University Press, 1998, pp. 103-110
5. Dunn, G. and Everitt, B. S. **An Introduction to Mathemathical Taxonomy**, UK Cambridge University Press, 1982, pp. 65-78
6. Fernandez, M. R., Ulrich, L., Brandt, S. A., Thomas A. G., Olfert, O., Yentner, R. P. and McConkez, B. G. **Impact of Crop Management Systems on Diseases of Spring Wheat on the Canadian Prairies**, Buck, H.T., Nisi, J.E., Salomón, N. (Eds.), Wheat Production in Stressed Environments, Proceedings of the 7th International Wheat Conference, 27 November - 2 December 2005, Mar del Plata, Argentina, Series: Developments in Plant Breeding, Vol. **12**, Berlin: Springer-Verlag, 2007, pp. 265-271
7. Friedmann, H. **Changes in the international division of labour: Agri-food complexes and export agriculture**, chapter in the book Friedland, W. F. (Ed.), „Towards a new political economy of agriculture", Boulder, Colorado, USA, Westview Press, 1991
8. Giri, C. Narazan **Multivariate Statistical Analysis**, USA Marcel Dekker, Inc., 2004, pp. 435-526
9. Goodmann, D., Sorj, B. and Wilkinson, J., **From farming to biotechnology: A theory of agroindustrial development**, Oxford, UK, B. Blackwell Ltd., 1987
10. Khattree, R. and Naik, D. N. **Applied Multivariate Statistics with SAS Software", Second Edition**, Cary, NC, SAS Institute Inc., 1999
11. Krzanowski, W. J. **Principles of Multivariate Analysis. A User's Perspective. Revised Edition**, USA, Oxford University Press, 2000, pp. 124-152
12. Martens, H. and Martens, M. **Multivariate Analysis of Quality: An Introduction**, USA, John Wiley & Sons, Inc., 2000, pp. 235-254
13. McLachlan, G.J., Basford, K.E., **Mixture Models, Inference and Applications to Clustering**, Marcel Dekker, New York, 1988
14. Milligan, G. W. and M. C. Cooper, M. C. **A study of standardization of variables in cluster analysis**, Journal of Classification, 5, 1988, pp. 181-204
15. Moayedi, A. A., Boyce, A. N. and Barakbah, S. S. **The Performance of Durum and Bread Wheat Genotypes Associated with Yield and Yield Component under Different Water Deficit Conditions**, Australian Journal of Basic and Applied Sciences, **4**(1), 2010, pp. 106-113
16. Rencher, A. C. **Methods of Multivariate Analysis**, USA, John Wiley & Sons, Inc., 2002
17. Spircu, L. **Analiza datelor: aplicatii economice**, Bucuresti, Editura ASE, 2005
18. Stewart, P. W. and Muller, K. E. **Linear model theory**, John Willez and Sons, 2006, pp. 35.
19. Walton, P. D. **Heterosis in spring wheat**, Crop. Sci., **11**, 1971, pp. 422-424
20. Waugh, F. V. **Regressions between sets of variables**, Econometrica, **10,** 1942, pp. 290–310
21. Ziv, O. and Goldman, A. **Comparisons of methods for statistical analysis of wheat growth factors in Israel**, Israel Agresearch **1**(2), 1987, pp. 53-66

Appendix1— *County Codification*

County	Code	County	Code	County	Code
Bihor	BH	Iasi	IS	Dambovita	DB
Bistrita-Nasaud	BN	Neamt	NT	Giurgiu	GR
Cluj	CJ	Suceava	SV	Ialomita	IL
Maramures	MM	Vaslui	VS	Prahova	PH
Satu-Mare	SM	Braila	BR	Teleorman	TR
Salaj	SJ	Buzău	BZ	Dolj	DJ
Alba	AB	Constanta	CT	Gorj	GJ
Brasov	BV	Galati	GL	Mehedinti	MH
Covasna	CV	Tulcea	TL	Olt	OT
Harghita	HR	Vrancea	VN	Valcea	VL
Mures	MS	Ilfov	IF	Arad	AR
Sibiu	SB	Bucuresti	B	Caras-Severin	CS
Bacau	BC	Arges	AG	Hunedoara	HD
Botosani	BT	Calarasi	CL	Timis	TM

Appendix 2—*Dataset of agriculture at the regional level in year 2004 (source: Romanian Statistical Yearbook, 2005) (I)*

Code	FOM04	PROD04	SUP04	TR04	PLU04	CUL04	SEM04	COM04
BH	92,3	881640	75141	9246	6918	1265	3089	1359
BN	45,9	169621	13302	2167	1665	143	552	269
CJ	78,1	426790	32051	5330	4452	593	1384	780
MM	77,4	133853	10796	3959	2724	81	433	235
SM	58,9	641429	46262	4535	3879	617	1735	829
SJ	37,2	241021	21598	4340	3654	498	1033	515
AB	51,6	377042	31524	3127	2373	720	1225	495
BV	33,7	151502	27518	4614	3283	839	994	443
CV	24,8	148469	24133	4002	2907	790	840	377
HR	43,9	88163	14553	3956	2400	1086	571	419
MS	74,8	601364	45007	5029	3950	654	1920	876
SB	29,5	204097	14478	2925	2135	444	813	309
BC	70,0	474330	22505	2424	1988	199	989	321
BT	79,0	367078	28383	2561	2289	361	1257	495
IS	99,7	461432	39341	3303	2922	696	1390	460
NT	87,1	355203	28541	2107	1840	239	987	402
SV	112,6	260088	33246	4497	3504	522	846	605
VS	73,0	689407	63888	2663	2413	249	1076	314
BR	41,7	960294	52962	3509	2827	830	1505	391
BZ	79,7	830255	45247	3549	2867	688	1247	323
CT	68,4	947655	86504	4938	3732	821	2150	660
GL	65,4	867146	49716	3243	2762	650	1464	346
TL	31,8	695579	43023	2521	1961	458	1102	448
VN	66,6	419478	29748	3191	2759	350	1119	316
Code	FOM04	PROD04	SUP04	TR04	PLU04	CUL04	SEM04	COM04
IF	36,1	307361	25628	1616	1243	404	829	177
B	8,8	2281	175	280	142	55	74	34
AG	74,1	504848	42400	3065	2426	406	1161	456
CL	51,2	1251109	113043	4504	3346	880	2278	851
DB	74,3	491899	34677	4771	3741	916	1685	378
GR	49,8	669364	81541	3756	2787	633	1704	478
IL	48,1	854255	75810	3147	2383	574	1678	514
PH	69,4	486930	30604	2162	1440	264	988	360
TR	93,0	1330201	216587	7058	6148	1556	3727	1243
DJ	115,9	1428540	234739	7422	6165	1731	4203	1952
GJ	41,3	344164	22232	2391	1925	214	837	304
MH	52,4	638239	75818	3592	3022	766	1730	622
OT	83,4	987795	145031	6041	5251	775	3509	1228
VL	60,1	323026	14770	2331	2569	96	651	229
AR	49,4	1148139	118651	7952	6132	1948	3464	1222
CS	43,0	379006	24208	5758	4099	566	869	396
HD	42,9	192911	17580	3969	2885	441	905	361
TM	76,0	1670001	142987	10260	8192	2578	5333	1861

Appendix 2—*Dataset of agriculture at the regional level in year 2004 (source: Romanian Statistical Yearbook, 2005) (II)*

PRE04	INS04	FUN04	ERB04	CHIM04	AZ04	FOS04	POS04	NAT04
279	31826	42391	226529	25373	15627	7402	2344	1127269
44	12808	27608	42966	5558	3293	1957	308	752073
134	30217	118985	144079	12631	7431	5155	45	1275402
50	7008	14547	19384	2992	1999	815	178	896211
255	16849	29140	174008	13288	8751	3753	784	503800
41	10272	36563	65767	4689	3486	1026	177	448545
63	13436	33859	77566	6816	4218	2252	346	740681
130	12656	46039	47477	4657	2862	1318	477	457430
170	18081	71243	42568	8469	4644	2566	1259	475351
137	7788	31508	25079	3621	2127	905	589	697423
182	24256	43526	149834	10151	6240	3425	486	893841
119	10005	22891	60493	3734	2786	741	207	752146
26	19039	31896	30991	8727	6399	2261	67	404995
112	13061	18522	23584	6190	3828	2189	173	1020449
77	86602	98484	47818	9678	6046	3042	590	416711
56	7675	14707	38309	6927	4354	2140	433	379700
154	20727	55296	51428	11407	7915	2594	898	837702
42	15157	22750	35280	6612	4211	2050	351	419890
130	54583	54693	166379	9026	6033	2966	27	6290
124	6821	33625	28530	3212	2514	563	135	207840
230	46511	82148	102492	5926	3606	1924	396	33956
147	42002	108035	98057	6431	4628	1580	223	124266
161	20250	34506	34355	3660	2837	777	46	10770
66	26270	389480	33533	4037	3181	811	45	179709
PRE04	INS04	FUN04	ERB04	CHIM04	AZ04	FOS04	POS04	NAT04
100	11721	10646	38320	3712	2416	1104	192	56970
15	140	241	40	32	22	7	3	361
68	41704	55166	94097	11864	8759	3059	46	315110
200	17578	20630	173666	9475	7133	2103	239	3936
110	37074	50534	70198	10926	7486	2561	879	524368
121	22371	24117	114163	10675	9076	1536	63	136160
110	14559	24435	94740	4651	3601	780	270	4805
79	14462	84185	30798	6526	5672	823	31	252470
131	15956	28115	157304	20796	15284	5456	56	302690
174	33100	127906	108298	18734	15736	2873	125	166856
17	4530	6611	12103	5093	4639	426	28	456690
22	1508	9855	55789	4950	4737	206	7	337281
90	31053	35421	77569	21327	17841	3303	183	128871
12	10267	29391	16838	7753	6855	862	36	459658
313	16777	33491	298335	16789	12718	3834	237	381568
43	1421	13323	84379	4640	3898	740	2	435901
50	2895	15559	47087	2841	1961	659	221	396645
537	45304	98807	530758	35408	23281	9529	2598	326036

[1] **Source:** European Climate Assement & Dataset (ECA&D) site: http://eca.knmi.nl/

[2] through access to the SAS Centre of Excellence, partnership between Acedemy of Economics Studies Bucharest and SAS Romania

[3] All data analyzed in this article were relevant statistically, with a significance level of 95%, and r is the correlation coefficient.

[4] α=0,05.