The clustering of agricultural products and determining important countries for these clusters by the factor analysis

Sabri ER[1], Ahmet ÖZÇELIK[2]

[1]*Ankara Development Agency, Ankara, Turkey*
[2]*Agricultural Economics Department, Ankara University, Ankara, Turkey*

Abstract: In the study, some important herbal agricultural products with respect to their production have been clustered, in addition to determining the most important or the best countries in terms of the production of certain herbal agricultural products by using the factor analysis. The FAO data set has been used in obtaining production of 30 agricultural products in 86 countries. 8 factors have been achieved by considering the Eigen values the numbers of which are greater than one. Each factor contains certain herbal agricultural products. First factor explains 40.51% of the total variation whilst the last factor explains only 3.89% of the total variability. 10 best countries for each factor have been revealed.

Key words: Eigen value, herbal product, total variability, Varimax

The first and foremost important priority of every living creature is to provide sufficient food in order to retain liveliness. Therefore, agriculture, a strategic as well as economic sector, is an inevitable activity for the human life. Today, the most significant goal of every government in the world is to meet sufficient and balanced food demand of its population. Herbal production such as wheat, paddy, maize, cassava has a particular importance due to the fact that these crops are used as stable food almost for the whole world population. The kind and variety of crop being produced depend heavily upon the climate and ecological conditions of the region. Consequently, we have a large variation in agricultural production especially in the herbal production in the terms of kind, quantity, quality, etc. produced in the world. Cassava or potatoes are the main product in some part of the globe while paddy or wheat is the major crop being produced in other part of the world to meet the basic nutrition of the population.

In some studies, there might be many variables to be used. However, using many variables in any study usually leads to the difficulty in interpreting the outcomes of the study. These variables in most cases are interrelated and it is possible to group these variables into one common variable called factor. When we group all interrelated variables into some factors, we obtain few variables (factors) than we have at the beginning. Factor analysis is a statistical technique being used in order to cluster the related several variables (medium or highly related) into few but independent variables called factors. Thus, large data set can be reduced and simplified in the study. For instance; gross profit, gross revenue, the total family income, the total agricultural revenue, etc. can be clustered in one common variable, named the income factor.

Rietveld and Van Hout (1993) state that the goal of factor analysis actually is to reduce the dimensionality of the original space and to give an interpretation to the new space, spanned by a reduced number of new dimensions which should underline the old ones. The obtained factor creates a new dimension that can be visualized as classification axes along which the measurement variables can be plotted. An important feature of factor analysis is that the axes of the factors can be rotated within the multidimensional variable space. Figure 1 shows what happens during a rotation when you only have two dimensions.

Factor analysis has been extensively used in many scientific areas, though it has been one of the leading statistical techniques in social sciences. Rummel (1970) suggested that there have been more methodology goods devoted to the topic of factor analysis than any other social science method or technique. When conducting a factor analytic study, a number

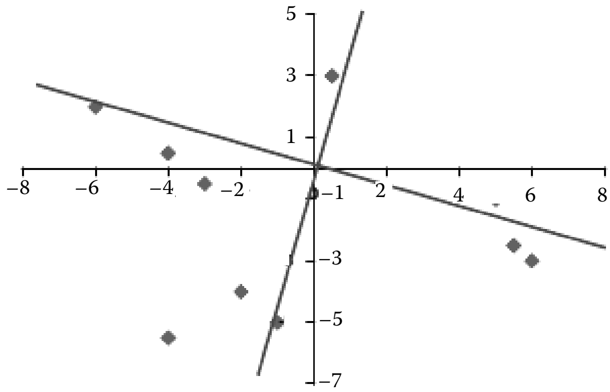

Figure 1. A factor rotation with only two factors

of issues must be considered. Ford et al. (1986) concentrated on four major steps; the choice of factor model to be used, the decision about the number of factors, the rotation method, and the interpretation of the factor solution.

There are also several studies conducted by using factor analysis in agriculture. Ghisay and Hosseini (2010) stated that the financial, technical, informational and regulatory challenges based on factor analysis have a greater impact on the emerging entrepreneurship in agricultural cooperatives. They also commented that those challenges have significant impacts on small farms in agricultural cooperatives. Akpınar and Yurdakul (2008) found that consumers' loyalty and brand preferences were two main factors for food products. Li (2012) suggested that agricultural policy and the number of technologic leaders have more influence on promoting the innovative capability of agriculture than the financial input of agricultural science and technology, the level of the local economic development, and the number of agricultural research institutions. Lashgarara (2004) proposed that 79% variance of the adoption of sustainable agriculture is determined based on five factors, namely the farming-economic factors, characteristics of innovation, the individuals' characteristics, the communication channels and educational participation. He also stated that social participations, access to market, rate of using mass media, the participation into extension classes, the knowledge and attitude about sustainable agriculture have a significant impact on the adoption of sustainable agriculture among wheat farmers of the Lorestan province. Dağistan et al. (2008) identified seven factors for the success of sheep production in Middle-South Anatolia (Turkey). These factors are firm size, rant ability, food input, unit costs, land, labour productivity, and pasturage time. Raven (1998) revealed that agricultural educators were reluctant to use factor analysis due to the fact that it requires large sample size. Çelik (2012) has introduced five factors in the examination of plant production in Turkey. The 69.315% of the total change can be explained by those factors.

Table 1. Results of the Kaiser-Mayer-Olkin (KMO) and the Bartlett tests

KMO and Bartlett's test		
Kaiser-Meyer-Olkin Measure of sampling adequacy		0.739
Bartlett's test of Sphericity	approx. Chi-Square	4.18E + 03
	df	378
	Sig.	0

MATERIAL AND METHODS

The purpose of this paper is to analyse and identify the factors that influence the herbal agricultural production in the world. In this study, we have used the production of 30 products in tonnes selected from 86 countries. The data set for 2011 has been obtained from the Food and Agriculture Organization of the United Nations (FAO). The selected products are; Apples, Apricots, Bananas, Barley, Beans (dry), Cassava, Cherries, Chick peas, Coffee (green), Cotton (lint), Cucumbers, Dates, Grapes, Groundnuts, Lentils, Maize, Oil (palm), Olives, Onions (dry), Oranges, Potatoes, Rice (paddy), Soybeans, Strawberries, Sugar beet, Sunflower seed, Tea, Tobacco, Tomatoes, and Wheat.

Factor analysis has been used to group agricultural products that have a medium or high correlation among or between them. Before the factor analysis is applied, the data set is supposed to be tested whether it is convenient for the study. Kaiser-Mayer-Olkin (KMO) and Bartlett tests are taken into account for this purpose. The KMO is an index (between 0 and 1) for comparing the magnitude of the observed correlation coefficients to the magnitude of the partial correlation coefficients. The closer KMO value to 1 indicates a sizeable adequacy. Less than 0.5 KMO value is unacceptable. The results of the KMO and Bartlett's test have been shown on Table 1.

The KMO value is 0.739 that shows the data set is good enough for the study. The Bartlett test checks whether the hypothesis correlation matrix is an identity matrix or not. The chi-square critical value below large enough not to reject the hypothesis that the population correlation matrix is an identity.

RESULTS AND DISCUSSION

Factor analysis is conducted through four steps. In the first step, the correlation matrix is generated so as to identify the variables that are related and most probably they will be in the same factor. Field (2000) states that variables in the study have to be inter correlated. However, this correlation should not be too high that may cause difficulties in determining the unique contribution of the variables to a factor. Correlation coefficients greater than 0.3 in the absolute value are indicative of the acceptable correlations. The primary objective of this stage is to determine factors. Factors are obtained by using the Principal Components Analysis, the most commonly used extraction method.

Eigen values are used to decide on how many factors we need to represent the data set we have in the study in addition to the scree the plot. As a general

Table 2. The total variance and % according to the individual factors

Total variance explained									
Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	total	% of variance	cumulative %	total	% of variance	cumulative %	total	% of variance	cumulative %
1	11.34	40.513	40.513	11.34	40.513	40.513	9.697	34.632	34.632
2	3.713	13.259	53.772	3.713	13.259	53.772	3.406	12.165	46.797
3	2.562	9.15	62.922	2.562	9.15	62.922	2.796	9.984	56.781
4	1.996	7.128	70.05	1.996	7.128	70.05	2.691	9.612	66.394
5	1.815	6.483	76.533	1.815	6.483	76.533	2.111	7.538	73.932
6	1.324	4.729	81.262	1.324	4.729	81.262	1.543	5.512	79.444
7	1.146	4.091	85.354	1.146	4.091	85.354	1.457	5.202	84.646
8	1.091	3.895	89.249	1.091	3.895	89.249	1.289	4.603	89.249
9	0.685	2.445	91.695						
10	0.598	2.137	93.832						
11	0.364	1.3	95.131						
12	0.328	1.173	96.304						
13	0.248	0.885	97.189						
14	0.208	0.742	97.931						
15	0.173	0.618	98.549						
16	0.106	0.378	98.927						
17	0.08	0.284	99.211						
18	0.07	0.251	99.462						
19	0.052	0.186	99.648						
20	0.029	0.102	99.751						
21	0.018	0.066	99.817						
22	0.015	0.055	99.871						
23	0.013	0.046	99.918						
24	0.01	0.035	99.952						
25	0.005	0.019	99.971						
26	0.004	0.015	99.987						
27	0.003	0.01	99.996						
28	0.001	0.004	100						

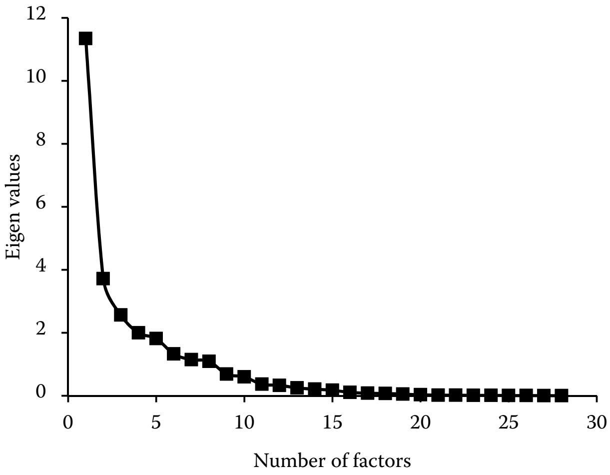

Figure 2. Scree plot

rule, factors whose Eigen values are greater than one are considered. Rietveld and Van Hout (1993) mention that the number of the positive Eigen values determines the number of dimensions needed to represent a set of scores without any loss of information. As seen in Table 2, there are 8 factors the Eigen values of which are greater than one. Thus, we have obtained 8 factors in our study. The first extracted factor has always the largest amount of variance in the sample. Then the second factor has the second largest amount of variance and so on. 40.51% of the total variance is explained by the first factor whilst only 3.89% of the total variance is explained by the eight factor. Almost 90% of the total variance is explained by 8 factors in this study.

Table 3. Rotated factor loads

Products	Factors							
	1	2	3	4	5	6	7	8
Apples	0.978	0.095	0.061	–0.111	0.077	0.056	0.004	–0.021
Cucumbers	0.977	–0.032	0.028	–0.184	0.053	0.022	–0.005	–0.005
Groundnuts	0.954	0.041	–0.007	0.212	0.080	–0.016	–0.056	0.067
Tomatoes	0.928	0.209	0.070	0.133	0.121	0.121	0.140	–0.002
Tobacco	0.921	0.057	–0.010	–0.170	0.304	0.004	–0.052	0.008
Tea	0.917	–0.038	–0.011	0.332	0.042	–0.004	–0.004	–0.026
Onions	0.911	0.071	0.076	0.371	0.085	0.018	0.056	0.003
Potatoes	0.874	0.107	0.375	0.225	0.051	–0.062	–0.033	–0.030
Rice	0.831	–0.020	–0.036	0.447	0.055	–0.020	–0.050	0.192
Cotton	0.764	0.304	–0.002	0.484	0.182	–0.056	–0.044	–0.029
Wheat	0.740	0.263	0.448	0.363	0.035	–0.007	0.024	–0.063
Strawberries	0.012	0.946	0.101	–0.007	–0.001	–0.022	0.066	–0.030
Maize	0.521	0.799	0.078	–0.042	0.189	–0.066	–0.104	0.023
Cherries	0.010	0.789	0.091	0.013	–0.039	0.319	0.280	–0.040
Soybeans	0.098	0.701	0.066	0.084	0.599	–0.081	–0.120	0.021
Barley	0.009	0.090	0.908	0.046	–0.048	0.211	0.025	–0.059
Sunflower seed	0.127	–0.101	0.877	–0.054	0.088	–0.055	–0.051	–0.027
Sugar beet	0.118	0.326	0.859	–0.054	–0.065	0.010	0.067	–0.019
Chickpeas	0.173	–0.001	0.000	0.971	–0.022	0.012	0.020	–0.024
Bananas	0.420	–0.037	–0.057	0.827	0.184	–0.031	–0.045	0.136
Beans	0.207	–0.076	0.002	–0.016	0.835	–0.021	–0.003	–0.009
Oranges	0.234	0.381	–0.017	0.172	0.789	0.096	0.029	0.101
Olives	–0.028	0.019	0.055	–0.005	0.016	0.921	–0.005	–0.034
Grapes	0.541	0.465	0.191	–0.061	0.017	0.547	0.120	–0.045
Dates	0.001	0.026	–0.077	–0.022	0.000	–0.172	0.873	–0.060
Apricots	0.002	0.153	0.142	0.010	–0.031	0.409	0.722	–0.006
Oil palm	–0.003	0.006	–0.029	–0.015	–0.129	–0.025	–0.014	0.881
Cassava	0.040	–0.053	–0.082	0.111	0.394	–0.028	–0.069	0.648

Extraction method: Principal Component Analysis; Rotation method: Varimax with Kaiser Normalization; Rotation converged in 6 iterations.

The scree plot is another method to decide the number of factors obtained in the study. Figure 2 shows the scree plot obtained in the study. As seen, the greater the Eigen value is the greater the slope of the curve is. As the Eigen value is getting smaller, the slope of the curve is getting flatter and after a certain point, it becomes flat. That is, the variance is not explained (or very little explained) by those factors. In other words, those factors are not significant at 5% level in this study.

The factors are rotated in order to make the factors more interpretable and more understandable. The rotation can best be explained by imagining factors as axes in a graph on which the original variables load. It is possible to make the factors load optimally by rotating these axes. The most popular rotational method is the Varimax rotation. The Varimax attempts to minimize the number of variables that have high loadings on a factor. This enhances the interpretability of the factors. As a general rule, the value of the common factor correlation ±0.3 or higher indicates a significant relation between variable and a factor. Table 3 displays the rotated factor loads obtained from the Varimax rotation.

We clearly see what products fall in what factor. Apples, Cucumbers, Groundnuts, Tomatoes, Tobacco, Tea, Onions, Potatoes, Rice, Cotton, and Wheat are in the first factor (F1). Strawberries, Maize, Cherries, and Soybeans are in the second factor (F2). Barley, Sunflower seed, and Sugar beet are in the third factor (F3). Chickpeas and Bananas are in the fourth factor (F4). Beans and Oranges are in the fifth factor (F5). Olives and Grapes are in the sixth factor (F6). Dates and Apricots are in the seventh factor (F7). Oil palm and Cassava are finally in the eighth factor (F8). It is possible to see some variable(s) that are not included in the factor but it (or they) has (have) a certain degree of relation with that factor. For instance; Strawberries, Maize, Cherries, and Soybeans are in the second factor. Grapes are not included in this factor due to having a smaller number of correlation compared to other products. However, there is a positive and powerful correlation with the same direction (0.465) between Grapes and factor 2. Similarly, the third factor in which

Table 4. List of countries based on the factor score of development

Factor 1		Factor 2		Factor 3		Factor 4	
China. mainland	8.958	United States of America	8.771	Russian Federation	6.13	India	8.881
India	1.608	Iran	0.902	Ukraine	3.962	Philippines	0.681
United States of America	0.180	France	0.569	Turkey	1.535	Ecuador	0.506
Pakistan	0.146	Italy	0.528	France	2.848	Australia	0.436
Egypt	0.105	Germany	0.525	Germany	1.814	Brazil	0.285
Bangladesh	0.104	Argentina	0.473	Argentina	1.039	Pakistan	0.261
Iran	0.097	Chile	0.452	Canada	0.717	Indonesia	0.220
Italy	0.093	Brazil	0.449	Australia	0.699	Turkey	0.141
Indonesia	0.033	Turkey	0.413	United States of America	0.265	Iran	0.117
Kenya	0.018	Romania	0.286	Hungary	0.250	United Republic of Tanzania	0.085
Factor 5		**Factor 6**		**Factor 7**		**Factor 8**	
Brazil	7.988	Spain	6.770	Iran	5.301	Indonesia	6.817
Hungary	3.364	Italy	4.087	Egypt	4.072	Malaysia	4.276
Argentina	1.149	Turkey	2.868	Turkey	2.991	Nigeria	3.247
Nigeria	0.887	Greece	1.196	Saudi Arabia	2.779	Thailand	1.706
Mexico	0.861	Morocco	0.953	Iraq	1.486	Brazil	0.779
Azerbaijan	0.717	France	0.874	Pakistan	1.265	Viet Nam	0.583
China. mainland	0.507	Iran	0.667	Italy	0.744	Mozambique	0.408
United Republic of Tanzania	0.486	Chile	0.472	Morocco	0.555	Turkey	0.217
Egypt	0.283	Australia	0.327	France	0.457	France	0.111
Spain	0.271	South Africa	0.131	Oman	0.404	Russian Federation	0.099

barley, sunflower seed, and sugar beet are grouped also contains wheat. There is a positive and pretty strong relationship (0.448) between wheat and F3.

The list of countries based on the factor score of development in terms of the herbal agricultural production has been given on Table 4.

The best 10 countries have been determined for each factor. Each value shows the degree to which each country's development contributes for agricultural production. The grater the value is, the more production the country has. The first factor in which Apples, Cucumbers, Groundnuts, Tomatoes, Tobacco, Tea, Onions, Potatoes, Rice, Cotton, and Wheat are included possesses 10 countries; China-mainland, India, the United States of America, Pakistan, Egypt, Bangladesh, Iran, Italy, Indonesia, and Kenya in order. These countries are the ones that produce the most of these stated products. In other words, these countries are the ones that are the best places in terms of climate and ecology to produce those products. China-mainland has the highest value by 8.958 among other countries shows that China is the best country in terms of growing and producing of those products. India, in the second order, has the value of 1.608 displaying that it is the second most important country with respect to production of those products included in the first factor, and so on. Similarly, United States of America is the most favourite country in the world in terms of growing and producing of the products included in the second factor; Strawberries, Maize, Cherries, and Soybeans. Russian Federation with the highest value of 6.631is the best place to produce Barley, Sunflower seed, and Sugar beet in the globe. In factor 6, olives are one of the two products included. Mediterranean is the best place in terms of ecology to produce olives. Thus, Spain is the number one country in the list by 6.770 followed by Italy with the value of 4.087 and Turkey with 2.868. Likewise, oil palm is one of the two products in factor 8. Indonesia and Malaysia are the biggest palm oil producers in the world. Indonesia has the highest value by 6.817 followed by Malaysia by 4.276.

REFERENCES

Akpınar M.G., Yurdakul O. (2008): Factors effecting brand preferences for food products. Akdeniz University Journal of Agricultural Science, *21*: 1–6.

Çelik Ş. (2012): Examination of plant production of provinces in Turkey by factor analysis. Yüzüncü Yıl University, Agriculture Faculty, Journal of Agricultural Science, *22*: 69–76.

Dağistan E., Koç B., Gül A., Gül M. (2008): Factor analysis of sheep production: A case study of Middle-South Anatolia. Yüzüncü Yıl University, Agriculture Faculty, Journal of Agricultural Science, *18*: 67–77.

Field A. (2000): Discovering Statistics Using SPSS for Windows. Sage publications, London-New Delhi.

Ford J.K., MacCallum R.C., Tait M. (1986): The application of exploratory factor analysis in technique: A critical review and analysis. Personnel Psychology, *39*: 291–314.

Ghisay F.G., Hosseini J.F. (2010): Challenges in developing in Iran's agricultural cooperatives: A factor analysis. World Applied Science Journal, *10*: 1032–1037.

Lashgarara F. (2004): Factor analysis of influencing factors on adoption of sustainable agriculture among wheat farmers of Lorestan province, Iran. Islamic Azad University, Science & Research Branch,Tehran.

Li J.Q. (2012): Factor analysis of agricultural innovative ability in Sichuan province. Environment and Natural Resources Research, *2*: 16–20.

Raven M.R. (1998): The Application of exploratory factor analysis in agricultural education research. Journal of Agricultural Education, *35*: 9–14.

Rietveld T., Van Hout R. (1993): Statistical Techniques for the Study of Language and Language Behavior. Mouton de Gruyter, Berlin.

Rummel R.J. (1970): Applied Factor Analysis. Northwestern University Press, Evanston.

Received: 5th February 2014
Accepted: 7th April 2014

Contact address:

Sabri Er, Ankara Development Agency, 1322 cad. No:11 Öveçler, Ankara, Turkey 06460
e-mail: sabrier06@gmail.com