ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 3(322)2016

http://dx.doi.org/10.18778/0208-6018.322.11

Małgorzata Misztal[*]

ON THE USE OF SELECTED ORDINATION TECHNIQUES TO ANALYZE THE PARLIAMENTARY ELECTION RESULTS

Abstract. Ordination techniques such as CCA (Canonical Correspondence Analysis) or RDA (Redundancy Analysis) are very popular in ecological research but almost completely unknown in, for example, socio-economic research.

The goal of this paper is to concisely organize the concepts and terminology associated with ordination and to present the possibilities of its application in social research with an example of the analysis of the 2015 parliamentary elections results.

Keywords: ordination techniques, redundancy analysis, parliamentary elections.

JEL: C38, D72

1. INTRODUCTION

The simplest definition of the term "ordination" was given by Goodall (1954) and it means the arrangement of units in some order. According to Gower (1984) the term "ordination" was brought into general use by ecologists.

In ecology "ordination" refers to the representation of samples or sites as points along one or more gradients on the basis of their species composition or environmental attributes. The end result of ordination is (usually) two-dimensional ordination plot "showing relations among samples and/or species. Samples that are near to one another in the ordination diagram are inferred to resemble one another in species composition and environmental attributes. There is a tacit assumption that samples with similar species have similar environments" (Calow [ed.] 1999: 91).

In multivariate statistics, ordination is "the process of reducing the dimensionality (i.e. the number of variables) of multivariate data by deriving a small number of new variables that contain much of the information in the original data. The reduced data set is often more useful for investigating possible structure in the observations" (Everitt, Skrondal 2010: 312).

[*] Ph.D., Department of Statistical Methods, University of Łódź.

Ordination techniques (called also gradient methods) are very popular in ecological research but little known in, for example, socio-economic analyses.

2. TYPES OF ORDINATION TECHNIQUES

Data for ordination typically consist of two matrices Y and X stacked one beside the other:

$$\mathbf{D} = [\mathbf{Y} \,|\, \mathbf{X}] = \left[y_{ij} \,|\, x_{ik}\right] \quad i = 1, 2, ..., n; \quad j = 1, 2, ..., m; \quad k = 1, 2, ..., p \qquad (1)$$

In ecological research rows of the matrix represent sites, the first block of *m* columns represents species and the second block of *p* columns represents environmental variables. Speaking more generally, rows of the matrix represent objects (cases), the first m columns represent response (dependent) variables, and the next *p* columns represent explanatory (independent) variables (predictors).

There are two major approaches to ordination:

1) indirect (unconstrained) ordination – when only **Y** data are analyzed; if there is any information about the **X** data, it is used to interpret the results from indirect gradient analysis;

2) direct (constrained) ordination – when **Y** and **X** data are analyzed simultaneously.

Taking into account the relationship between response end explanatory variables the basic types of ordination techniques can be summarized as in Table 1.

Table 1. Basic types of ordination techniques

Ordination method:	Type of model:	
	Linear	unimodal (Gaussian)
unconstrained	Principal Components Analysis, PCA	Correspondence Analysis, CA (Reciprocal Averaging) Detrended Correspondence Analysis, DCA
Constrained	Redundancy Analysis, RDA	Canonical Correspondence Analysis, CCA Detrended Canonical Correspondence Analysis, DCCA

Source: own elaboration based on ter Braak&Prentice (1988).

Principal components analysis (PCA; Pearson 1901, Hotelling 1933) is based on rotation of the original system of axes defined by the response variables, such that the successive new axes (so-called principal components which are linear functions of the original variables) are orthogonal to one another and account for decreasing proportions of the variance in the data.

Redundancy analysis (RDA; Rao 1964, Wollenberg 1977) is canonical form of PCA and consists of two steps (Legendre, Legendre 1998). Step 1 is a multivariate regression of $\mathbf{Y}$ on $\mathbf{X}$ leading to a matrix of fitted values $\hat{Y}$ through the linear equation: $\hat{Y} = [X^T X]^{-1} X^T Y$. Step 2 is a principal component analysis of $\hat{Y}$. Both – the fitted values of the multivariate linear regression and the canonical axes are linear combination of the explanatory variables in *X*.

Correspondence analysis (CA) aims at visualizing a table of data in a low-dimensional subspace with optimal explanation of inertia (Greenacre 2007: 185). Additional information can be included in the map in the form of supplementary (passive) points with zero mass and zero inertia in order to interpret their positions relative to the active points.

By contrast, in canonical correspondence analysis (CCA; ter Braak 1986) the dimensions are assumed to be responses in a regression-like relationship with external variables i.e. dimensions are found with the same CA objective but with the restriction that the dimensions are linear combinations of a set of explanatory variables (Greenacre 2007: 192).

Detrended correspondence analysis (DCA; Hill, Gauch 1980) has been developed as a modification of CA designed to eliminate the so-called "arch effect". The arch effect appears when the positions of the objects on the second (vertical) ordination axis are strongly and nor linearly dependent on their positions on the first (horizontal) axis (Lepš, Šmilauer 2003: 53).

Detrended canonical correspondence analysis (DCCA) is a constrained version of DCA, although, according to Lepš & Šmilauer (2003: 53–54), the detrending procedure is rarely needed for a constrained unimodal ordination because the arch effect in CCA is usually a sign of some redundant explanatory variables being present. Removing one variable from such a group usually turns out to be the best solution.

All the techniques mentioned above are described in detail in e.g. ter Braak & Prentice (1988), Jongman et al. (1995) and Legendre & Legendre (1998).

To decide which ordination method – linear or unimodal is appropriate, the gradient lengths are important. Since the axes in DCA are scaled in standard deviations units (Hill, Gauch 1980) it is helpful to do DCA first and establish the gradient lengths. If the longest gradient is larger than 4 unimodal methods (CA, DCA or CCA) should be used. If the longest gradient is shorter than 3, the linear methods (PCA or RDA) are better. In the range between 3 and 4 every method can be used with good result (see: Lepš, Šmilauer 2003: 51).

Popular statistical packages such as SPSS or STATISTICA can be used for calculations needed for PCA and CA. Calculations of DCA, CCA and RDA models can be performed with the use of R-project (vegan and ade4 packages) or CANOCO for Windows.

As an illustration of the use of ordination methods the parliamentary election results will be analyzed.

3. ANALYSIS OF THE PARLIAMENTARY ELECTION RESULTS

3.1. Material and methods

Parliamentary elections to the Sejm were held on 25 October 2015. There were 8 nationwide committees:

1. PIS – Prawo i Sprawiedliwość / Law and Justice;
2. PO – Platforma Obywatelska / Civic Platform;
3. RAZEM – Partia Razem / Together;
4. KORWiN – Koalicja Odnowy Rzeczypospolitej Wolność i Nadzieja / Coalition for the Renewal of the Republic – Liberty and Hope;
5. PSL – Polskie Stronnictwo Ludowe / Polish People's Party;
6. ZLEW – Zjednoczona Lewica / United Left;
7. KUKIZ'15 – Kukiz'15;
8. .N – .Nowoczesna / .Modern.

The election results (i.e. the set of response variables) are presented in Table 2. Votes for 9 regional committees are totalled in the last column (OTHERS).

Table 2. Election results (in %)

Voivodships (sites):	Committees:								
	PIS	PO	RAZEM	KORWIN	PSL	ZLEW	KUKIZ15	.N	OTHERS
1	2	3	4	5	6	7	8	9	10
Dolnośląskie	32.63	29.26	3.86	4.74	3.14	8.05	9.03	8.69	0.59
Kujawsko-Pomorskie	31.86	27.74	3.70	4.23	6.40	10.39	8.04	6.91	0.72
Lubelskie	47.76	14.83	2.60	4.74	9.24	6.45	9.79	4.22	0.38
Lubuskie	28.27	28.21	3.99	4.99	5.12	10.02	8.75	9.99	0.65
Łódzkie	38.35	23.15	3.79	4.29	5.93	8.79	8.65	6.70	0.36

Table 2 (cont.)

1	2	3	4	5	6	7	8	9	10
Małopolskie	48.18	19.43	3.08	5.20	4.19	4.73	8.14	6.58	0.47
Mazowieckie	38.30	22.61	4.21	5.15	4.84	6.92	7.89	9.53	0.55
Opolskie	27.77	26.23	3.02	3.95	3.68	6.75	12.57	7.14	8.88
Podkarpackie	55.09	13.37	2.30	4.69	5.69	4.47	9.23	4.09	1.05
Podlaskie	45.38	16.74	2.59	4.66	8.07	7.35	9.07	5.37	0.76
Pomorskie	30.45	34.06	4.02	4.70	3.13	6.62	7.60	8.67	0.75
Śląskie	34.82	25.56	3.91	4.88	2.52	8.33	10.69	8.06	1.23
Świętokrzyskie	42.81	17.25	2.80	4.14	9.51	7.87	9.41	4.98	1.23
Warmińsko-Mazurskie	30.91	28.38	3.76	4.94	7.69	8.30	8.66	6.39	0.97
Wielkopolskie	29.61	28.45	3.94	4.32	6.62	9.28	7.77	9.32	0.70
Zachodniopomorskie	28.91	31.25	4.04	5.01	3.97	9.59	8.78	8.44	0.00

Source: own elaboration based on PKW data.

The set of explanatory variables consists of 16 subjectively selected characteristics of the voivodships that may affect the voters' decisions and is presented in Table 3.

To choose between linear and unimodal ordination, detrended correspondence analysis (DCA) was performed with the use of CANOCO 4.5 software. The longest gradient was 0.665. As it is shorter than 3, the linear methods (PCA or RDA) should be selected to analyze the data. Since we have a set of explanatory variables, redundancy analysis (RDA) seems to be the best choice.

Another important problem is the selection of explanatory variables. If the number of independent variables is greater than (the number of object – 1) the analysis is unconstrained. The fewer the explanatory variables, the stronger the constraints are. The forward selection procedure implemented in CANOCO software is based on Monte Carlo permutation tests (see details in Lepš, Šmilauer 2003: 60–72).

The assessment of the usefulness of each potential predictor variable for extending the subset of explanatory variables used in the ordination model starts with the first step when each variable is tested separately to estimate its independent, marginal effect i.e. the amount of variability in the response data that would be explained by a constrained ordination model using that variable as the only explanatory variable. The variable with the greatest marginal effect is selected into the model. In the next steps consecutive variables are entered into the model on the basis of their conditional effect i.e. the ability to increase the variance explained by the model.

Table 3. Selected characteristics of voivodships

Voivodships (sites):	Explanatory variables:															
	Number of eligible voters	Voter turnout (in %)	Population per 1 km^2	Population in urban areas in % of total population	Population in % of total population at working age	Population in % of total population at post-working age	Unemployment rate	Persons employed in agriculture. forestry and fishing in % of total	Gross Domestic Product per capita in zl	Average monthly gross wages and salaries in zl	Average monthly per capita expenditures of households in zl	Dwellings in thous.	Beneficiaries of social welfare benefits per 10 thous. population	At-risk of poverty rate	Median age	The percentage of people with higher education
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Dolnośląskie	2315022	49.42	146	69.3	63.5	19.7	9.1	4.6	47440	4048.81	1130.57	1123.4	417.5	5.1	39.9	10.24
Kujawsko-Pomorskie	1636579	46.36	116	59.8	63.3	18.4	10.6	14.1	34095	3440.10	987.60	726.6	747.4	9.6	38.8	8.50
Lubelskie	1730825	49.02	85	46.2	62.4	19.5	9.9	23.3	29479	3608.07	982.13	756.0	555.4	9.4	39.0	9.41
Lubuskie	800699	44.63	73	63.1	63.8	18.0	8.4	7.9	34862	3428.60	1051.49	362.7	636.5	6.4	38.6	9.09
Łódzkie	2014767	51.63	137	63.2	61.9	21.2	8.8	12.4	39080	3620.78	1118.97	997.8	521.7	6.1	41.0	9.51
Małopolskie	2647532	54.90	222	48.6	62.8	18.2	9.1	11.0	36961	3702.90	1000.51	1132.1	412.4	6.0	37.9	9.75
Mazowieckie	4402322	58.71	150	64.3	62.0	19.5	7.2	11.1	66755	4933.78	1355.41	2166.8	421.2	5.7	38.9	11.86
Opolskie	808931	43.12	106	52.0	64.3	19.6	7.8	11.4	33888	3638.14	1048.79	350.2	408.0	6.1	40.7	8.38
Podkarpackie	1699179	50.43	119	41.3	63.5	17.7	14.0	16.2	29333	3414.00	900.95	641.4	641.1	9.4	37.7	9.26
Podlaskie	947710	47.10	59	60.5	63.4	19.0	9.1	23.1	30055	3533.05	978.15	436.4	637.3	11.2	39.2	8.51

Table 3 (cont.)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Pomorskie	1767283	51.88	126	64.9	62.8	17.7	8.6	6.6	41045	4017.54	1094.90	824.5	538.4	9.2	37.8	9.40
Śląskie	3627755	52.25	372	77.3	63.2	20.0	8.6	2.8	44372	4104.89	1129.50	1731.0	390.5	4.9	40.6	11.63
Świętokrzyskie	1031221	46.82	108	44.6	62.6	20.4	11.3	22.6	31459	3438.02	969.83	436.1	663.6	8.5	40.2	10.31
Warmińsko-Mazurskie	1139200	42.32	60	59.2	64.2	17.0	9.8	13.0	30065	3390.14	922.62	501.1	878.0	13.2	38.0	7.66
Wielkopolskie	2717872	50.16	116	55.1	63.1	17.7	7.7	13.0	44567	3600.32	947.06	1163.0	460.8	8.9	37.8	10.47
Zachodniopomorskie	1342253	45.88	75	68.7	63.8	18.7	8.5	8.1	35334	3653.02	1066.19	633.9	599.2	7.1	39.4	9.23

Source: own elaboration based on PKW data and *Regiony Polski* (2015).

The results of the forward selection procedure with 499 random permutations are summarized in Table 4 and Table 5. Four explanatory variables should be entered into the RDA model.

Table 4. Ranking of the explanatory variables according to their marginal effect

Variable number	Variable	Variance explained [%]
8	Persons employed in agriculture	41.89
4	Population in urban areas	38.75
7	Unemployment rate	34.65
5	Population at working age	14.24
2	Voter turnout	10.71
11	Average monthly per capita expenditures	8.28
9	Gross Domestic Product per capita in zl	8.25
14	At-risk of poverty rate	4.45
6	Population at post-working age	3.31
10	Average monthly gross wages and salaries	3.16
15	Median age	2.72
13	Beneficiaries of social welfare benefits	2.22
1	Number of eligible voters	1.92
16	People with higher education	1.72
3	Population per 1 km^2	1.71
12	Dwellings	1.70

Source: own calculations using CANOCO software.

Table 5. The final results of the forward selection procedure

Steps	Variable	Additional fit to the explanation of variance [%]	Variance explained by the variables selected [%]	p-value
1	Persons employed in agriculture	41.89	41.89	0.006
2	Voter turnout	18.81	60.70	0.012
3	Unemployment rate	16.39	77.09	0.008
4	Beneficiaries of social welfare benefits	6.17	83.26	0.036

Source: own calculations using CANOCO software.

3.2. Redundancy analysis results

The basic results are presented in Table 6.

There are 4 canonical axes (because there are 4 explanatory variables) and 5 non-canonical ones. Four canonical axes explain 83.1% of the total variability. The first canonical axis (axis 1) explains 91% of variability in the canonical space and 75,6% of the total variability.

Table 6. The basic results of the RDA

Canonical axes		1	2	3	4
Eigenvalues:		0.756	0.044	0.027	0.004
Species-environment correlations:		0.631	0.875	0.758	0.612
Cumulative percentage variance:	of species data:	75.6	80.1	82.7	83.1
	of species-environment relation:	91.0	96.3	99.5	100.0

Source: own calculations using CANOCO software.

RDA ordinations may be presented as a biplot or triplot. An RDA biplot presents objects as points and either response or explanatory variables as vectors. In a triplot, objects are ordinated as points while both response and explanatory variables are presented as vectors (arrows). Levels of nominal variables are plotted as points.

The interpretation of these plots depends on what scaling has been chosen. In general, type I scaling (focus on objects) should be considered if the distances between objects are of particular value, or if most explanatory variables are binary or nominal. Type II scaling (focus on response variables) should be considered if the correlative relationships between variables are of more interest (for more details see: ter Braak 1994; Legendre, Legendre 1998; Lepš, Šmilauer 2003).

The RDA triplot (type II scaling) for parliamentary elections results is presented in Figure 1. Objects (voivodships) are ordinated as black points, response and explanatory variables as arrows (solid black and dashed grey respectively). Since type II scaling was applied, distances between object points should not be considered to approximate Euclidean distances.

A lot of information can be deduced from ordination diagrams.

Perpendicular projections of object points onto vectors representing response or explanatory variables approximate variable values for a given object. This gives the approximate ordering of the objects in order of increasing value of that response/explanatory variable.

Analyzing the results presented in Figure 1 it can be seen, among others, that PSL achieved the best election results in 3 voivodships: Lubelskie, Podlaskie and Świętokrzyskie. These are the voivodships with the highest percentages of persons employed in agriculture. The best election result was gained by PiS in Podkarpackie – the voivodship with the highest unemployment rate. It can be also observed that KORWiN achieved the biggest % of votes in Mazowieckie voivodship where the voter turnout was the highest.

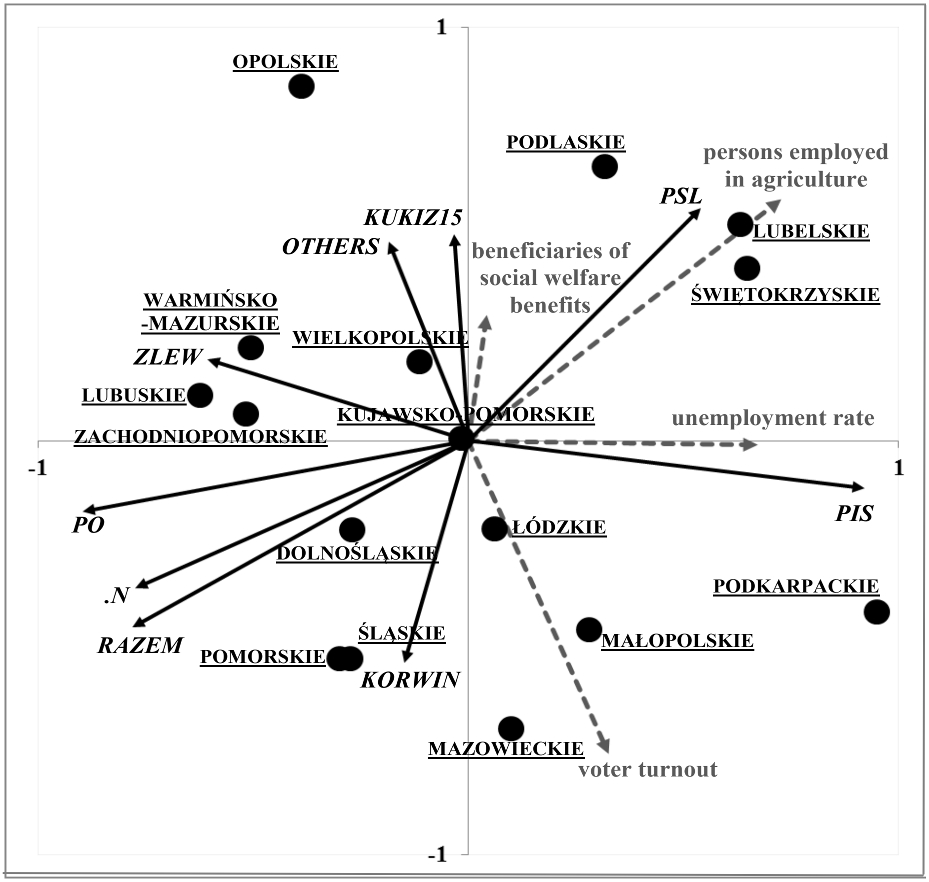

Figure 1. RDA ordination triplot of the parliamentary elections data

Source: own calculations using CANOCO software.

There is also one object point, representing Kujawsko-Pomorskie voivodship, projecting onto the beginning of the coordinate system – it means that this voivodship has an average value of the corresponding response or explanatory variables.

The angles between all vectors reflect their linear correlation. The approximated correlation between two variables is equal to the cosine of the angle between the corresponding vectors. Perpendicular vectors indicate the lack of correlation between the variables they represent. The angle less than 90° suggests positive correlation between variables and the angle approaching 180° – strong negative correlation between variables.

The following correlations can be observed, among others, on the RDA ordination triplot in Figure1:

1. Between response variables:
   - strong positive correlations between the election results of .Nowoczesna, Razem and PO and between Kukiz'15 and the Others;
   - strong negative correlation between the election results of PiS and Zjednoczona Lewica;
   - lack of correlation between the election results of KORWiN and PiS and KORWiN and Zjednoczona Lewica.
2. Between explanatory variables:
   - strong negative correlation between the voter turnout and the number of beneficiaries of social welfare benefits;
   - moderate positive correlations between the unemployment rate and the percentage of persons employed in agriculture.
3. Between response and explanatory variables:
   - strong positive correlation between the election results of PSL and the percentage of persons employed in agriculture;
   - strong negative correlations between the election results of Razem/ .Nowoczesna/PO and the percentage of persons employed in agriculture;
   - strong positive correlation between the election results of PiS and the unemployment rate;
   - strong positive correlations between the election results of Kukiz'15/Others and the number of beneficiaries of social welfare benefits;
   - strong positive correlation between the election results of KORWiN and the voter turnout.

The angles between vectors representing response or explanatory variables and the canonical axes can be also used to assess the linear correlation coefficients. The first axis is strongly and positively correlated with the rate of unemployment and the percentage of persons employed in agriculture. The second axis is negatively correlated with the voter turnout and positively with the number of beneficiaries of social welfare benefits.

Let us leave any political conclusions to political scientists.

4. CONCLUDING REMARKS

The goal of ordination is to represent object and response variables relationships as faithfully as possible in a low-dimensional space (Gauch 1982). But reduction of dimensionality is not the only reason to use ordination. Redundancy analysis as well as all other ordination methods is a technique of exploratory data analysis. Graphical presentation of the results of the ordination using the ordination biplots or triplots can facilitate the analysis of the relationship between the variation in the set of the response variables and the variation of the explanatory variables which can be measured on different measurement scales (interval, ordinal, nominal) with no need to satisfy any assumption (e.g. normality).

Ordination techniques should be popularized in socio-economic research.

REFERENCES

Calow P. (red.) (1999), *Blackwell's Concise Encyclopedia of Ecology*, Blackwell Science, London.

Everitt B.S., Skrondal A. (2010), *The Cambridge Dictionary of Statistics*, Fourth Edition, Cambridge University Press, Cambridge.

Gauch H.G., Jr. (1982), *Noise reduction by eigenvalue ordinations*, "Ecology", vol. 63, pp. 1643–1649.

Goodall D.W. (1954), *Objective methods for the classification of vegetation. III. An essay in the use of factor analysis*, "Australian Journal of Botany", vol. 2, pp. 304–324.

Gower J.C. (1984), *Ordination, multidimensional scaling and allied topics*, (in:) W. Lederman [ed.] "Handbook of Applicable Mathematics", vol. VI: E. Lloyd (ed.) "Statistics", Wiley, Chichester, pp. 727–781.

Greenacre M. (2007), *Correspondence Analysis in Practice*, 2ed., Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton.

Hill M.O., Gauch H.G. (1980),*Detrended correspondence analysis: an improved ordination technique*, "Vegetatio", vol. 42, pp. 47–58.

Hotelling H. (1933), *Analysis of a complex of statistical variables into principal components*, "Journal of Educational Psychology", Vol. 24, pp. 417–441, 498–520.

Jongman R.H.G., ter Braak C.J.F., van Tongeren O.F.R. (ed.) (1995), *Data Analysis in Community and Landscape Ecology*, Cambridge University Press, Cambridge.

Legendre P., Legendre L. (1998), *Numerical Ecology*, Elsevier Science B.V., Amsterdam.

Lepš J., Šmilauer P. (2003), *Multivariate Analysis of Ecological Data using CANOCO*, Cambridge University Press, Cambridge.

Państwowa Komisja Wyborcza: Wyniki wyborów do Sejmu RP z dnia 25 października 2015, Internet site:http://parlament2015.pkw.gov.pl, access: 15.11.2015.

Pearson K. (1901), *On lines and planes of closest fit to systems of points in space*, "Philosophical Magazine", Ser. 6, vol. 2, pp. 559–572.

Rao C.R. (1964), *The Use and Interpretation of Principal Component Analysis in Applied Research*, "Sankhyā: The Indian Journal of Statistics", Series A (1961–2002), vol. 26, no. 4 (Dec., 1964), pp. 329–358.

Regiony Polski(2015), GUS, Warszawa.

ter Braak C.J.F. (1986), *Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis*, "Ecology", vol. 67(5), pp. 1167–1179.

ter Braak C.J.F. (1994), *Canonical community ordination. Part I: Basic theory and linear methods*, "Ecoscience", vol. 2, pp. 127–140.

ter Braak C.J.F., Prentice I.C. (1988), *A theory of gradient analysis*, "Advances in Ecological Research", vol. 18, pp. 271 – 317.

van den Wollenberg A.L. (1977), *Redundancy analysis. An alternative for canonical correlation analysis*, "Psychometrika", vol. 42, no. 2, pp. 207–219.

Małgorzata Misztal

O ZASTOSOWANIU WYBRANYCH TECHNIK ORDYNACYJNYCH DO ANALIZY WYNIKÓW WYBORÓW PARLAMENTARNYCH

Streszczenie. Techniki ordynacyjne, takie jak kanoniczna analiza korespondencji (CCA) czy analiza redundancji (RDA), zyskały popularność przede wszystkim w badaniach ekologicznych, trudno natomiast znaleźć ich zastosowania np. w badaniach ekonomiczno-społecznych.

Celem pracy jest zwięzłe uporządkowanie pojęć i terminologii związanej z ordynacją oraz wskazanie możliwości aplikacyjnych metod ordynacyjnych w badaniach społecznych na przykładzie analizy wyników wyborów parlamentarnych w 2015 roku.

Słowa kluczowe: techniki ordynacyjne, analiza redundancji, wybory parlamentarne.