APORTACION METODOLOGICA AL ESTUDIO DE LAS VARIACIONES DE POBLACION. EJEMPLO DE POLONIA

POR

ELZBIETA IWANICKA-LYRA

RESUMEN:

La experiencia de numerosos países prueba que la realización de modelos de procesos de población es un problema complejo, no sólo porque los planificadores tienen dificultades para determinar el óptimo poblacional, sino también porque es difícil determinar de antemano sus efectos propagadores. Por esta razón, las políticas demográficas tropiezan con serios obstáculos que parcialmente pueden ser superados conociendo la importancia de los diferentes factores demográficos y analizando cuál es la posible incidencia cuantitativa ante un cambio de los factores.

Para ello, la autora propone la utilización de un método de regresión múltiple que aplica a las diferentes voivodinas polacas, obteniendo resultados en el cálculo que son bastante parecidos a los que se producen en la realidad.

The lower than expected natural increase during the last years, observed on both the national and regional levels, has become the subject of inquiry of demographers in scholarly periodicals. It was also debated by the public in numerous articles in the daily press.

In statistical terms, the natural growth decrease results from a lower birth rate. In Poland, this phenomenon, though typical of societies undergoing rapid urbanization and industrialization, has been conditioned by numerous additional factors. One of these is, above all, the lack of an unequivocal and energetic pro-natalistic policy aimed at influencing family size on a level desired at a given period of development with an eye on the economic and social rise of the country in the future. The experience of numerous countries proves that the shaping of the population process is a complex problem not only because population planners have difficulties in determining the optimal population needs in the future, but also because it is difficult to anticipate propaganda effects. This is due to the fact that a properly conducted pro-natalistic policy resorts most often to economic incentives (development of a platform of benefits and help for large families, improved housing and living conditions, better health care), and to incentives aiming at influencing the consciousness of the society (the creation of a friendly atmosphere around the birth of a child and in favor of large families, the popularization of a hierachy of values in which having children would be a superior aim, the introduction of a system of rights, preferences and even privileges conducive to the organization of daily family life, and many others). Such incentives, however, function as stimulators for some time only, then their power of influence decreases. Consequently, population policy should be verified periodically, accompanied by renewed discussions of the family problem.

In the light of what has been said above, the steerage of the process of population growth appears as a multidimensional task which needs to be based on a program prepared on the basis of a precise cognizance of reality. At the diagnostic stage, population planners concentrate not only on the difficulties causing the decrease of the number of children in the family, but above all on demographic features which determine, so to say, in a natural way the reproductive abilities of the society. Studies in population geography and demography imply that the natural increase rate may be the result, among others, of the degree of population concentration calculated by the number of population per square kilometer, the migrational inflow and outflow per 1000 inhabitants, which indicates population movements, including —in accord with migrational regularities— mainly persons of procreative age and number of marriages per 1000 persons, fertility of women defined as the number of children born alive per 1000 women between the age of 15 and 49, and age structure expressed in population percentage in age groups of 0-14, 15-49 and above 50 years. Statistical indicators of these features were used as the basis of calculation in the examination of the power of influence of these features on the natural increase rate. Analyses were conducted for the years 1975 and 1976. The voivodships in their administrative boundaries of 1975 were the basic units for collecting data. They were analysed by the use of the multiple regression method by applying the following formula:

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n + \in$$

where $b_1 \dots x_n$ = variate independent coefficients (with *n* independent variates) and $\in$ = random components. In the analyses, the seven above mentioned demographic features were considered as independent variates, while natural increase as the dependent variate. The first two equations proved the need for a separate analysis of both components of natural increase, i.e., births and deaths.

While establishing the equation of birth, the pertinence of the selection of independent variates was ascertained, which found proof in the fact that the deviations of the dependent variate were 99 % explained by deviations of independent variates (Table 1).

Detailed figures of the estimated regression equation may be obtained by placing the coefficients of Table 1 in the general regression equation; thus the equation for 1976 appears as

Births 1976 = 15,32 + (—0,00) × density of population + 0,06 × migration inflow + (—0,07) × migration outflow + 0,62 marriages + + 0,16 × fertility of women + (—0,12) × age of population (15-49) + (—0,28) × age of population (≥ 50) = 15,32 — 0 +

$$+ 0{,}06 \times 28{,}30 - 0{,}07 \times 28{,}3 + 0{,}62 \times 9{,}50 + 0{,}16 \times$$
$$\times 73{,}80 - 0{,}12 \times 53{,}00 - 0{,}28 \times 23{,}2 = 19{,}879.$$

Average for Poland after the Main Statistical Office = 20,270.

Difference 0,391

The small diffrence between the birth index calculated from the equation and the index published in the year-book of the Main Statistical Office proves that the equation was well adjusted. It appears that a set of equations calculated for a longer period of time would make it possible to draw certain generalizations which might serve as a basis for population policy. While working on the project reported in this paper, an attempt was made to cover also the years 1973 and 1974, but the statistical figures available for those years are only tentative due to the change in the administrative division of the country at that time. As a result the analysis disclosed diverging values of the regression index, which prevented the establishment of a general equation.

Work on the independente variates was conducted by using the t-Student's test. The dejection region of the hypothesis of insignificance of particular features for 49 observations on the determined significance level 0,02 lay above the value $t = 2{,}42$. Statistical values t for the multiple regression coeficients in which the dependent variate were births, are shown in Table 2.

The obtained results confirm the expectations: the number of births is determined, above all, by the fertility of women and the number of marriages. The conclusion re-enforces the conviction of the correctness of population policy, which can influence the birth rate at a desired level by propagating an appropriate family model. On the other hand migrational population movements constitute an effective tool of controling procreation, especially at the regional level. At the same time migrational inflow demonstrates a positive dependence, while the outflow of population is connected with a decrease of the birth index. The conclusion drawn from Table 2 is confirmed by maps of distribution of types of population development; they are marked by a high share of depopulating areas (over 50 % of the country's territory in 1974), in which population decline caused by migration contributed to the decline of the natural increase. These areas tend to grow (over 60 % of the country's territory in 1977) covering mainly rural localities which in the past had an above the average birth rate.

A proper definition of an expedient scale of population outflow from communities having surpluses of man power, the channelling of population inflow into communities which are demographically less resilient, seem therefore to be means for an effective regulation of the procreating of a region.

The greatest problems in this type of work are usually connected with the definition of the measures of influence of particular independent variates on the value of dependent variates. One of the methods applied in mathematical statistics is the definition of the partial correlation coefficients which define the relative share of particular variates in the multiple correlation coefficient. It should be remembered that the sum of partial coefficients is not equal to the value of the multiple correlation coefficient. The values of partial correlation coefficients confirm the conclusions drawn from earlier analyses regarding the significance of particular demographic features in the process of shaping the birth rate.

The method which demonstrates the interdependence of features is the percentage change in birth rates caused by the 10 % change of the values of independent variates. The calculations are carried out by introducing average values of independent variates to the regression equation of a given year. Table 4 contains the results of analysis for 1976. It should be interpreted in the following way: the fertility increase of women by 10 % brings about a 6 % increase of the number of births, but a 10 % increase of migrational outflow results in a 1 % drop of the birth rate, etc.

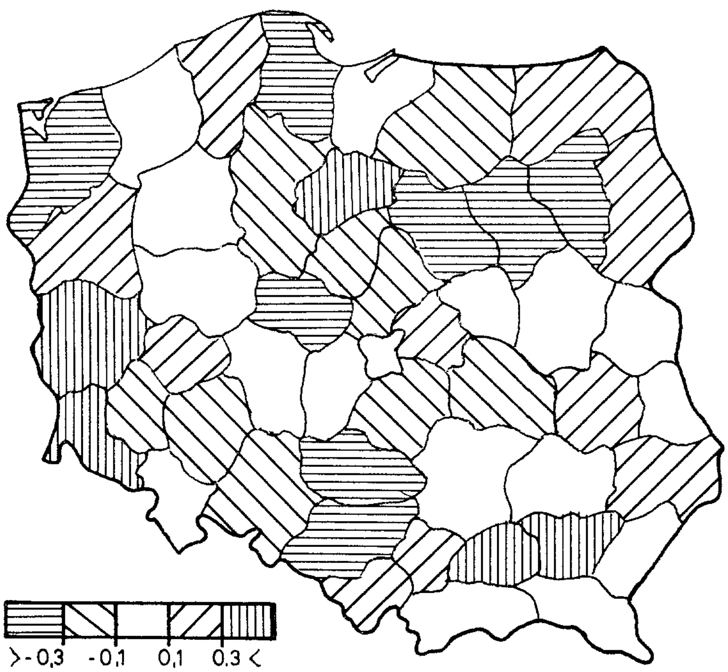
Table 4 presents information which may inspire population policy only during the years following directly the analyzed temporal profile, their usefulness in later years should be tested by repeating the work on the basis of up-to-date statistical data.

In all these analyses cause for reflection provides the negative correlation between the birth rate and the procreative age group of population, as well as the relatively weak indexes which are below the level of significance demonstrating the power of linkage between these two features. The reason for this is to be seen in the admission of the independent variates of the percentage share of population in the 15-49 age group in the total number of people. The results would be probably different if the variate would be determined in reference to the number of population above 15 years old.

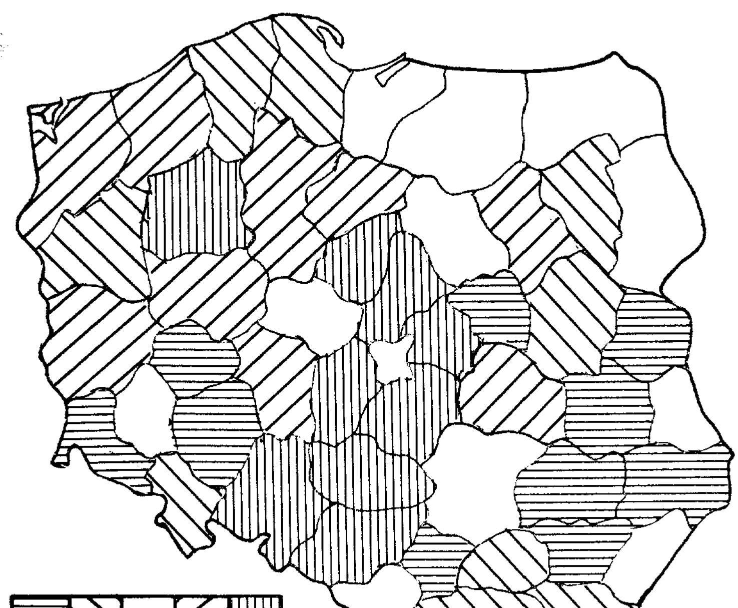
The second phase of the work consisted in determining the regression equation of the second component of the dependent variate, i. e., deaths. Since the equation of births and deaths is to serve the mathematical description of natural increase through one equation, the mathematical procedure of defining the dependence of the death rate was based on a set of variates applied earlier to the analysis of the birth rate. The high multiple correlation coefficients (Table 5) prove the pertinence of the selection of the independent variates and the high degree of adjustment of the regression equation.

Values of the regression coefficients inserted into the general regression equation permit a theoretical determination of the death index. The statistically irrelevant difference between the value of deaths elicited through the equation and the index value published in the year books of the Main Statistical Office confirm the proper adjustment of the equation. The work on the independent variates performed through the application of the t-Student's test (Table 6) point to a strong influence of age structure on the death rate —at the same time the percentage of inhabitants of older age correlates positively, while the percentage of people of procreative age— negatively. Such a conclusion is obvious, of course, but assumes diagnostic values when calculating the degree of connection between these two phenomena.

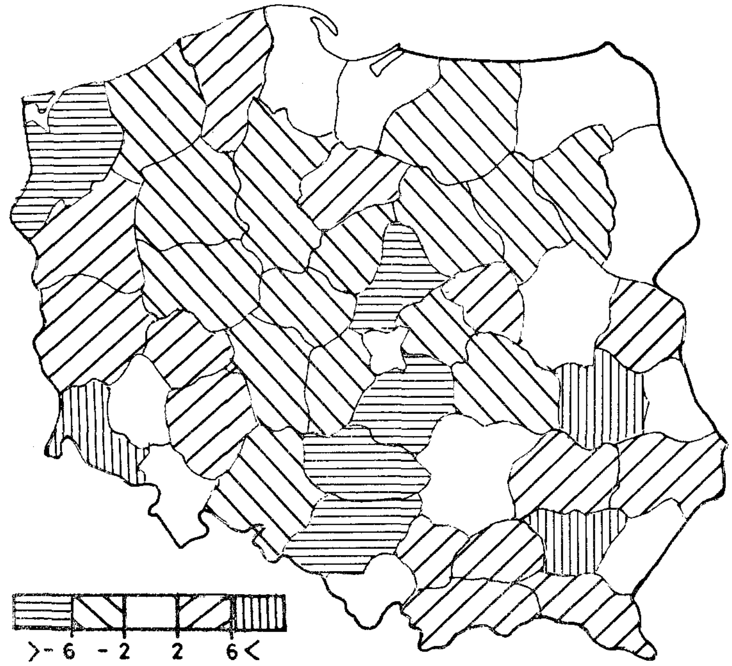
MAP. 1. MULTIPLE REGRESSION EQUATION RESIDUALS OF BIRTHS IN 1.976



MAP. 2. MULTIPLE REGRESSION EQUATION RESIDUALS OF DEATHS IN 1.976



MAP. 3. MULTIPLE REGRESSION EQUATION RESIDUALS OF NATURAL INCREASE IN 1.976



The analysis of the partial coefficients (Table 3) evinces that among the independent variates the features describing age structure are decisively influenced by the death rate, while the influence of other features is limited. Similar conclusions can be drawn from the analysis of the percentage of values of the death index at a 10 % of value of particular independent variates (Table 4).

The results of the work presented above aimed at defining the strength of connection between birth and death rates on the one hand, and selected demographic features on the other, on the national scale. The multiple regression method, however, may be also used to define the degree of adjustment of equations at the regional level. At this level conclusions can be drawn from the regression residuals. Assuming that a small value of the residual indicates analogic interdependencies in the country and a given region, and a greater value signals weaker connections between the examined features and the operation of other stimulators on births and deaths, there were introduced five classes differentiating deviations of estimated values from observed (regression residuals). Apart from a class of low regression residuals (oscilating from 0.1 to —0.1) there were established two classes of negative deviations and two with positive deviations. The spatial distribution of residuals from the regression equation of births and deaths have been demonstrated at the voivodship level, as illustrated on maps 1 and 2. The map of natural increase has been drawn on the basis of theoretical values of deaths and births, which helped to determine the estimated value of the natural increase compared with the index value of the year 1976 (Map 3).

The spatial distribution of residuals from the regression equation of natural increase indicates the diffusion of voivodships marked by an interdependence of features similar to that described by the country equation. Most of them are border areas marked by predominance of agriculture in the economic structure. There are, however, in this group also highly industrialized voivodships, such as Lódz, Walbrzych, Gdansk, Kielce. The spatial distances of these voivodships and the fact that they are economically and socially differentiated, renders a general interpretation difficult, which must be based on the analysis of particular enclaves.

Natural increase values are higher in the southwest voivodships than in those calculated theoretically. These areas were settled in the post Second World War period to a large degree by a young population, mostly rural. As a result, a large family model predominated there. Such a model was also propagated by the government of the Polish People's Republic during the last thirty years has had en undulatory character, the western territories of the country are demographically young today, which is revealed in the low average age of the inhabitants as well as in the above average family size. The demographic young age of this region is also illustrated on maps 1 and 2.

The natural increase of the southeastern voivodships is also higher than the theoretically expected. This is most likely the result of the lower urbanization of this and the traditionally large families in the Sub-Carpathian region. The positive deviation of the birth rate of some of the eastern voivodships has to be connected with the considerable improvement of living conditions in this area which until recent times was referred to as Poland B; at the same time, the construction of industrial plants in these areas and the intensive population inflow contributed to their demographic rejuvination.

The voivodships of Central Poland had a lower natural increase than that calculated theoretically; the highest regression residuals appeared in areas marked by high urbanization and industrial concentration, i. e., Upper Silesia. This is probably the effect of the intensive operation of the urbanization processes which manifest themselves, among others, in the form of a small family model, typical of urban milieus.

The reasons for the different degree of adjustment of the equation calculated for the country to particular regions are highly complex and call for more studies to provide a basis for interpretation. It should be remembered that in accord with the requirements of the applied method, reasons for the differentiation should not be looked for in the set of features acknowledged as independent variates, because their influence was eliminated in the computational procedure.

In the light of the study it seems to be clear that the multiple linear regression equation constitutes an adequate research tool in the work on the determination of the relation between birth and death rates on the one hand, and demographic features on the other. The establishment of the equation of natural increase is possible through the subtraction of estimated equations of births and deaths.

The most effective method of measuring the power of influence of demographic features on the natural increase rate seems to be the analysis of the partial correlation coefficient. Due to the elimination of the indirect influence of other variates, this method determines the interdependence among variates more accurately than other measuring tools. The most useful tool for the purpose of population policy, however, seems to be the method of percentage deviations of independent variates. The divergence in the estimation of regression coefficients for both years results probably from the degree of aggregation of initial data; these are available at the voivodship level, and consequently they could not be used for the differentiation of demographic features of the rural and urban population. The divergence also results from the narrowing of the group independent variates to demographic features without considering other characteristics which in various years might have had a bearing on the value of the dependent variate. In future work on this problem, several variates should be included in the group of independent features, such as percentage of urban population, medium age of population and the accumulated population inflow during the period preceding the given point in time.

TABLE 1

Multiple regression coefficients of birth equation

	Value	*Density of population*	*Migrational*		*Marriages*	*Fertility of women*	*Age*		*Correlatior coefficient*
			inflow	*outflo*			*15-49*	*50 and abore*	
	(a)	(b_1)	(b_2)	(b_3)	(b_4)	(b_5)	(b_6)	(b_7)	
1975	10,8945	—0,0019	0,0516	—0,0704	0,5620	0,1690	—0,0628	—0,2318	0,9901
1976	15.3203	—0,0012	0,0591	—0,0705	0,6196	0,1595	—0,1220	—0,2787	0,9892

SOURIE: The author's calculations based on data of the Department of Demographic and Social Studies of the Main Statistical Office.

TABLE 2

Values of statistics t for multiple regression of births

Years	*Densitrof population*	*Migrational*		*Marriages*	*Fertility of women*	*Age*	
		inflow	*ouflow*			*15-49*	*50 and abore*
	(b_1)	(b_2)	(b_3)	$(b)_4$	(b_5)	(b_6)	(b_7)
1975	—3,4691	4,4991	—5,6950	6,9609	10,7070	—0,9730	—4,7980
1796	—2,2470	3,1278	—4,3428	5,6123	7,6600	—1,2488	—4,2902

SOURCE: As in table 1.

TABLE 3

Partial coefficient correlation in 1976

	Density of population (b_1)	Migrational inflow (b_2)	Migrational outflow (b_3)	Marriages (b_4)	Fertility of women (b_5)	Age 15-49 (b_6)	Age 50 and abore (b_7)
Births	—0,33	0,47	—0,56	0,66	0,77	—0,19	0,56
Deaths	0,22	0,13	—0,22	0,68	x	—0,36	0,69

SOURCE: As in table 1.

TABLE 4

Birth a de ath deviations from averages caused by 10 % deviation from country average of particular independent variates

	Inflow (b_2)	Outflow (b_3)	Marriages (b_4)	Fertility tof women (b_5)	Age 15-49 (b_6)	Age 50 and abore (b_7)
Births	0,85	—1,00	2,95	6,00	—3,20	—3,25
Deaths	0,68	—0,91	0,68	x	—7,84	6,36

SOURCE: As in table 1.

TABLE 5

Multiple regression coefficients of cleaths

Years	Value	Density of population	Migrational		Marriage	Fertility of women	Age		Correlation coefficient
			inflow	outflow			15-49	50 and abore	
	(a)	(b_1)	(b_2)	(b_3)	(b_4)	(b_5)	(b_6)	(b_7)	
1975	5,2300	0,0014	—0,0704	0,0443	—0,2115	x	—0,0041	0,2659	0,9276
1976	9,7412	0,0016	0,0173	—0,0299	0,0591	x	—0,1308	0,2406	0,9556

SOURCE: As in table 1.

TABLE 6

Values of statistics t for multiple regression coefficient of death equation

Years	Density of population	Migrational		Marriages	Fertility of	Age	
		inflow	outflow			15-49	50 and abore
	(b_1)	(b_2)	(b_3)	(b_4)	(b_5)	(b_6)	(b_7)
1975	1,4100	—3,3831	1,9801	—1,5561	x	—0,0692	5,2922
1976	1.4398	0,8212	—1,4668	0,4922	x	—2,5104	6,1025

Notice: Significance limit of statistics t = 2,42.