Dominika Polko–Zając

Summary of doctoral thesis

Methods of comparing populations in economics research

Comparing the populations is significant among the methods of statistical inference used to identify differences between the studied populations. Classical parametric methods of comparing populations require meeting certain assumptions. It is often necessary that the observations for the sample be taken independently from the normal distribution population. Distribution of economic data is usually right skewed and the assumption about the normality of the distribution is unjustified. In social studies, on the other hand, variables measured on nominal and ordinal scales are most often analyzed, which also makes it impossible to make an assumption about the normality of the distribution. In such cases, it is usually not possible to use classical parametric tests. An alternative to parametric tests are nonparametric tests, which do not require fulfillment of many assumptions, but have lower power.

In recent years, the role of computer simulations in scientific research has increased significantly. Simulation analyzes allow answering many interesting and important questions. In this doctoral dissertation, particular attention was paid to permutation tests and the possibility of their use in economic statistical studies. These tests do not require meeting the assumption regarding the form of distribution of the examined variable in the population, and are characterized by similar power to parametric tests. The substance of permutation tests is to define test statistics in such a way that the areas of variation are different for cases when the verified hypothesis is true or false and then to determine the distribution of this statistic for all permutations of the variable. Usually, executing all permutations are practically impossible and for this reason the randomization method does not consider all possible permutations of a given data set, but some pre–determined, sufficiently large number of permutations. The test procedure using such permutation test is used to estimate the distribution of test statistics by approximating it using computer simulation.

The main goal of the dissertation was to present own proposals of simulation methods and modification of existing testing procedures as well as to assess the properties of these statistical tests against the parametric and non–parametric methods of populations comparison. The main objective was achieved through two groups of specific objectives. Theoretical and cognitive goals included reviewing and systematizing knowledge about

methods for comparing populations in economic research, and proposing and evaluating simulation methods that allow comparing multidimensional populations. Practical goals included indicating the possibility of using statistical inference methods and Monte Carlo methods to compare multidimensional populations in economic research, and proposing author's own methods using computer simulations to compare multidimensional populations.

The dissertation verified the main research hypothesis that the use of simulation, multidimensional methods of comparing populations in economic research leads to the effective identification of differences between populations also in cases where these populations do not meet certain assumptions and there is no possibility to use parametric methods. This hypothesis has been verified according to the above objectives through the following detailed hypotheses:

$H^{(1)}$: In the study of economic processes in the situation where the assumption regarding the form of distribution of the examined variable (variables) in the population is not met, simulation methods including resampling methods can be successfully used;

$H^{(2)}$: Statistical inference using permutation tests allows for effective analysis of economic data.

The author of the dissertation presents her own proposals and modifications of existing statistical tests using simulation methods enabling statistical inference also in cases where classical parametric tests are impossible to use or their use brings the risk of making wrong decisions. The subjects of the comparative analysis are objects that are units of two or more populations. Objects of economic research are characterized by multidimensional variables. Both real data and simulation data were used.

The structure of the dissertation, which consists of four chapters, was subordinated to the realization of the goals of thesis and testing of research hypotheses. The thesis contains a theoretical part and proposed methods using permutation tests with examples of their application. The first part of the thesis contains a theoretical description, which is an effect of detailed literature studies in the field of population comparison methods. The second part presents author's own proposals, modifications of existing statistical tests and results of simulation experiments regarding the properties of these testing procedures.

In the first chapter entitled *Parametric methods of comparing populations*, the basic concepts of the general theory of statistical hypotheses are discussed. The analysis began with a review of test procedures applicable to the verification of population comparison hypotheses. The chapter presents selected tests for verification of parametric hypotheses

including tests for equality of expected values, variances or proportions in two or more populations based on the independent or dependent, one– or multidimensional samples.

In the second chapter entitled *Nonparametric methods of comparing populations*, selected test procedures are presented, used to verify hypotheses about the identity of distributions of the examined variable in populations. As part of nonparametric methods, a number of tests have been distinguished enabling comparisons of distributions in two or more populations. The tests used for independent and dependent samples were considered. The issue of testing data homogeneity in contingency tables was also discussed.

In the third chapter *Characteristics of selected simulation methods of comparing populations*, selected statistical inference methods using resampling and computer simulation methods are presented. First of all, attention was paid to permutation tests, which due to computational intensity have only recently found wider practical application. The idea and types of these tests were discussed. The rest of the chapter focuses mainly on presenting author's own proposals for detecting the differences between the studied populations and proposals for permutation tests that are modifications of test procedures known from literature. Referring to current research directions related to the use of multivariate tests, it was proposed to use permutation tests for population analysis in which test objects are characterized by multidimensional variables. The characteristics of permutation tests used for data recorded on interval and ratio scales are presented. Among these tests, a modification of location–scale test, test for equality of mean vectors, test of identity of multidimensional population distributions based on the distance function and multidimensional location–scale test were proposed. The third chapter also includes the characteristics of permutation tests for verification of hypotheses formulated on the basis of data measured on nominal and ordinal scales. Tests of homogeneity of structures for data in contingency tables, both for data measured on nominal and ordinal scales, test of homogeneity of structures based on chi–square statistics and test of homogeneity of structures with directional alternative hypotheses. The third chapter also contains the characteristics of permutation tests used in correlation and regression analysis. The test of equality of correlation coefficients, the test of identity of two regression functions and the test enabling comparison of two sets of variables using the elements of canonical correlation analysis were presented. The chapter presents simulation analyzes of the properties of selected parametric, nonparametric tests and permutation tests. The comparative analysis concerned the size and power of these tests. The aim of the study was also to assess the properties of selected statistical tests where assumption

about the normality of the distribution in the population was fulfilled or not, also in case of small and large samples and samples of different and equal sizes.

In the fourth chapter *The use of simulation methods of comparing populations in economic research* selected examples of the use of permutation tests introduced in the third chapter were presented. The possibility of effective use of these tests as part of economic and social research to verify hypotheses formulated on the basis of data recorded on all types of scales was indicated. Examples of analyzes in the area of statistical quality control, issues of analysis of quality of life, analyzes of the labor market and other economic and social research were presented.

The new proposals presented in the dissertation as well as modifications of existing statistical tests are mainly aimed at weakening the assumptions that must be met when using classical parametric methods. The theoretical description of the proposed procedures and modification of existing solutions has been complemented by simulation analyzes. The research involved conducting computer simulations to verify the properties of the proposed non-classical methods. In the dissertation, these analyzes were carried out in the R statistical environment based on author's own calculation scripts. The analyzes were also carried out for selected, widely available, real data sets. For this purpose, a set of data from the Social Diagnosis study was used, as well as data shared by the Central Statistical Office of Poland. Using the proposed simulation methods, the hypothesis about the stability of the production process was verified. Comparative analyzes related to the economic activity of the population, the situation of the unemployed on the labor market and the subjective assessment of job satisfaction of respondents in selected Polish voivodships were carried out. The social effects of economic changes were examined by comparing the relationship between the assessment of satisfaction with various aspects of life of people, and selected economic and social variables in Polish voivodships in two periods. The last example in the dissertation concerned the comparison of the financial situation in households in Poland. In the empirical research carried out, the author's intention was to confirm similarity or to detect the differences between populations characterized by economic and social variables.

Knowledge accumulated in data on economic and social phenomena can be an effective tool for improving the quality of various decisions made. The advantage of using the presented permutation tests is the possibility of using them without verification of the assumption regarding the form of distribution of the examined variable in the population and in the case of having a sample with a small number of observations. For the test statistics

used, it is not necessary to know its distribution, because the distribution estimate is obtained by simulation. It was shown in the course of simulation studies that permutation tests are characterized by similar power to parametric tests, and in economic studies they lead to effective identification of differences between populations.

Dominika Pohko-Zajpc