UNIWERSYTET SZCZECIŃSKI
WYDZIAŁ NAUK EKONOMICZNYCH I ZARZĄDZANIA

Studia i Prace WNEiZ US
nr 55 2019

EKONOMIA I FINANSE

DOI: 10.18276/sip.2019.55-06

Michał Nowakowski*
Uniwersytet Szczeciński

THE ANOVA METHOD AS A POPULAR RESEARCH TOOL

Abstract

The article presents theoretical and practical characteristics of the ANOVA analysis of variance, that is one of very popular statistical methods in experimental research. The study describes various types of analysis of variance, basic statistical assumptions, as well as the one research procedure of the selected types of the ANOVA method, that is one-way analysis of variance. The aim of the article was to show the theoretical assumptions and the research procedure of the ANOVA method, exemplary application of the method in research on the information usefulness on websites and literature review for selected applications of the method in various scientific fields. Presented in the article, the ANOVA method description can be useful in the practical selection of types of research experiments and expected statistical results, that can be implemented using analysis of variance.

Keywords: ANOVA analysis of variance, one-way analysis of variance, two-way analysis of variance, applications of the ANOVA method

Introduction

In research aimed at analyzing issues of cognitive nature and in situations of solving practical problems, some empirical methods of conducting research are usually used, which one of the examples is an experiment (experience). An **experiment** is based on

* ORCID: 0000-0003-1029-4594, E-mail: michal.nowakowski@usz.edu.pl

conscious induction of the phenomenon under strictly defined conditions in order to investigate its course and leads to a deliberate change by the researcher of the conditions and processes taking place in accordance with previously accepted premises.

1. The ANOVA analysis of variance

Analysis of variance, also called the **ANOVA method** (*ANalysis **Of** VAriance*), is a group of statistical data analysis methods, which are used to study and compare results of experiments or observations that depend on one or more factors operating simultaneously. Therefore, it is a technique of analyzing many populations by studying the influence of one or several factors, that are independent variables (explanatory variables), on one or several dependent variables (explained variables). The values of independent variables are also called the levels of the factor. The origin and development of the analysis of variance was initiated in 1923 by Ronald A. Fisher, english biologist and geneticist, who developed it for use in agriculture and biology, as well as data sorting. Over time, the analysis of variance, due to its versatility, has found wide applications in many different fields of experimental research (Simon, 2006; Wieczorkowska, Kochański, Eljaszuk, 2005).

The main goal of the analysis of variance is verification whether there is any relationship between the test data sets (e.g. direct effect of one factor on another), which is realized by testing the significance of differences between average values for the obtained results in many different populations. It can be also said, that the essence of the analysis of variance is the comparison of variance, that is the variability of the results obtained from the research, which are most often performed as comparisons between groups of independent variables. As practically, all the phenomena researched by scientists are characterized by some level of variability, the method of analysis of variance is used to determine the level and nature of this variability. This is done by dividing the total variability (i.e. the sum of squares of deviations of all measurements from the mean) into various sources related to the effects occurring in the study. Fisher distinguished three basic types of variability: total, between-group and within-group. From the mathematical side, the value of total variation is the sum of between-group variability and within-group variability (Bedyńska, Niewiarowski, Cypryańska, 2013; Ferguson, Takane, 1999).

As mentioned above, the main goal of the variance analysis is to determine whether the grouped data sets from research are related to each other in some way and there is a direct effect of one factor on the other. To prove this, the analysis of variance uses **two basic types of hypotheses**: **the H_0 hypothesis and the H_1 hypothesis**. The H_0 hypothesis assumes, that average values in populations are equal, which means that there is no influence of independent variables on dependent variables. The H_1 hypothesis assumes meanwhile, that the value of at least one average in populations differs significantly from at least one of the others, which means that there is an influence of independent variables on dependent variables (Wieczorkowska and others, 2005).

The analysis of variance, and in particular the *F test*, **requires meeting certain assumptions**, that are independent of the number of used factors, which means, that they are required for all types of the ANOVA method (i.e. one-way, two-way, etc.). The nature of these assumptions and the degree of non-compliance may lead to false conclusions from the analysis (Aczel, 2000; Ferguson, Takane, 1999; Snarska, 2005). These assumptions are as follows:

- the measurement values of the dependent variable within each group have a normal distribution,
- the variance of measurements between the compared groups is homogeneous,
- there is statistical independence of measurements within the group,
- the analyzed groups are as equally numerous as possible,
- individual elements in the experiment were assigned to experimental conditions in a random way.

2. Types of the ANOVA analysis of variance

The ANOVA analysis of variance is a set of statistical methods for comparing average values in three or more populations. There can be examined the influence of one classification factor within the one-dimensional model (so-called the one-way ANOVA analysis of variance), two classification factors within the one-dimensional model (so-called the two-way ANOVA analysis of variance with repeats and two-way ANOVA analysis of variance without repeats), many classification factors within the one-dimensional model (so-called the many-way ANOVA analysis of variance) or one/many classification factors within the many-dimensional model (so-called the many-way MANOVA analysis of variance) (Aczel, 2000; Ferguson, Takane, 1999; Finch, 2005; StatSoft.pl, 2018). If the assumptions of parametric tests are not met, other variants of the ANOVA method are applied for non-parametric tests, which additionally extend the possibility of analysis to the case of ordinal variables. Non-parametric equivalents of the variance analysis may include tests such as: the Kruskal-Wallis test, the Manna-Whitney test or the Friedmann test (Ferguson, Takane, 1999).

The types of the variance analysis most commonly used in scientific research include two types: the one-way analysis of variance and the two-way analysis of variance with repeats. Both types of analyzes are described below and, for the first of them, the research procedure is additionally described with an example of its application in the Microsoft Excel program.

The one-way analysis of variance is the simplest and most frequently used type of the ANOVA analysis of variance. It examines the impact of only one classification factor (for example: f_A – occurring on many levels) on the results of the conducted research. The one-way analysis assumes the presence of only one independent variable, which constitutes the studied groups or subgroups as part of the experiment, and only one dependent variable, which constitutes the results of the study. Using this type of the analysis of variance, the hypotheses concerning equality of means are tested (H_0(f_A) and H_1(f_A)) if there are more than two analyzed groups. Data differentiated by factor level are analyzed in the table in which they are placed in separate columns or rows (Snarska, 2005).

The research procedure using the one-way ANOVA analysis, assuming the study of the *k* kinds of experimental conditions, where in each of the *k* tests composed of *n* elements different experimental conditions are used, is presented below (Ferguson, Takane, 1999; Wieczorkowska and others, 2005).

1. **Identification of variables** (a dependent variable and an independent variable) and them **measuring scales** (types of variables: nominal, quantitative, etc.), **formulation of assumptions** (statistical requirements) and **main hypotheses** (H_0 and H_1).
2. **Calculation of the statistics distribution of the F_{otrz}** (calculation of the statistics distribution of the *F* Fisher value) based on the formulas below:

The *F* Fisher statistics: $$F_{otrz} = \frac{s_B^2}{s_W^2}$$

Independent estimators (s^2_B i s^2_W) of the variance in the population (σ^2):

$$s_B^2 = \frac{SS_B}{df_1} = \frac{SS_B}{df_B} = \frac{SS_B}{k-1} \qquad s_W^2 = \frac{SS_W}{df_2} = \frac{SS_W}{df_W} = \frac{SS_W}{k(n-1)} = \frac{SS_W}{N-k}$$

where:
N – the size of the whole sample (the total number of people participating in the study),
k – number of groups or number of factor levels (number of independent variable values),
n – number of people in the group.

Differentiations of between-groups (SS_B) and within-groups (SS_W):

$$SS_B = n\sum_{j=1}^{k}(M_j - M)^2 \qquad SS_W = \sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij} - M_j)^2$$

where:
i – index indicating the number of the individual in the group,
j – index indicating the number of the group to which the individual belongs,
M – general average,
M_j – average in the *j-th* group,
X_{ij} – value of the *i-th* individual result in the *j-th* group.

3. **Comparison of the statistics distribution of the F_{otrz} to its critical value of the F_{kryt}** (significance statistics of the F_{otrz} or the F test) for the assumed level of significance ($\alpha = 0{,}05$) and a certain number of degrees of freedom (for df_1 degrees of freedom in the numerator and df_2 degrees of freedom in the denominator) **based on values from the *F* distribution tables:**

   The F test: $F_{kryt} = F(k, N-k)$

4. **Establishing a decision-making rule and making a decision to reject or accept the H_0 hypothesis**. If $F_{otrz} \geq F_{kryt}$ the H_0 is rejected and it can be said, that there is at least one pair of means, which are different from each other and therefore the experimental factor affects statistically on feature, while if $F_{otrz} < F_{kryt}$ there are no grounds for the H_0 rejection.

As an example of the one-way analysis of variance application, the questionnaire data on the evaluation by the respondents of the selected aspects of the information usefulness used in various types of e-commerce online shopping websites were used. These data were obtained from a survey conducted in 2017 on a group of more than a dozen respondents representing various forms of education, age groups and experience in using the Internet (Nowakowski, Mazur, 2017).

The research procedure of the one-way analysis of variance was started with the determination of the independent variable and the dependent variable. The nominal variable was chosen as the independent variable, that is the type of e-commerce class shopping website (types: auction portal, price comparisoner portal, advertising portal, online shop). The dependent variable became the received quantitative results from the study on the level of perceived functional and visual transparency of the website as a whole in the assessment of individual respondents. In addition, the following main hypotheses were formulated: H_0 (the type of e-commerce shopping website does not affect the level of functional and visual transparency of the website as a whole) and H_1 (the type of e-commerce shopping website greatly affects the level of functional and visual transparency of the website as a whole). The next stages of the research procedure, that is calculating of the statistics distribution of the F_{otrz} and comparison of the statistics distribution of the F_{otrz} to its critical value of the F_{kryt}, were implemented using the data analysis tool (The analysis of variance: one-way) available in MS Excel. Automatically calculated results within the one-way analysis of variance tool are presented in the following Figure 1.

Figure 1. Screenshot of MS Excel presenting survey data and results obtained from data analysis for the one-way ANOVA analysis of variance

	The type of e-commerce class shopping site			
Respondents	**Auction portal**	**Price comparisoner portal**	**Advertising portal**	**Online shop**
Respondent-1	5	4	4	2
Respondent-2	5	4	5	3
Respondent-3	5	5	3	4
Respondent-4	3	4	4	3
Respondent-5	3	4	4	4
Respondent-6	4	4	4	4
Respondent-7	3	5	4	3
Respondent-8	3	5	4	5
Respondent-9	4	4	3	5
Respondent-10	5	5	4	5
Respondent-11	4	5	3	4
Respondent-12	4	4	4	3
Respondent-13	5	4	5	4
Respondent-14	4	5	5	3
Respondent-15	5	4	5	2

Analysis of variance: one-way ANOVA						
SUMMARY						
Groups	*Counter*	*Sum*	*Avarage*	*Variance*		
Auction portal	15	62	4,133333	0,6952381		
Price comparisoner	15	66	4,4	0,2571429		
Advertising portal	15	61	4,066667	0,4952381		
Online shop	15	54	3,6	0,9714286		
ANALYSIS OF VARIANCE						
Source of variance	*SS*	*df*	*MS*	*F*	*Value-p*	*Test F*
Between-group	4,983333	3	1,661111	2,7467192	0,0513595	2,7694309
Within-group	33,86667	56	0,604762			
Total	38,85	59				

Source: own study based on the MS Excel program.

On the basis of the results obtained from the data analysis, it was possible to move to the last stage of the research procedure that is determining the decision-making rule and making a decision on rejecting or adopting a H_0 hypothesis. As it results from the received data, the value of the statistics distribution F_{otrz} (F) = 2,7467, and critical value of the F_{kryt} (*Test F*) = 2,7694. Due to the fact, that F_{otrz} (F) < F_{kryt}

(*Test F*) and the *Value-p* (0,0513) is greater than the assumed level of significance (α = 0,05) there are no grounds for H_0 rejection. Accordingly, it can be concluded that the type of e-commerce shopping website does not affect significantly the level of functional and visual transparency of the website as a whole.

The two-way analysis of variance with repeats is more advanced and equally often used variation of the ANOVA analysis of variance, which can be helpful in many real situations. It examines the impact of two classification factors (for example: f_A and f_B – occurring on many levels) on the results of the conducted research and additionally, it takes into account the possibility of interactions between factors. **The interaction** between two factors, that is the combined effect of independent variables on a dependent variable, is expressed to the extent to which the impact of one factor depends on the level of the other factor.

The two-way analysis assumes the presence of two independent variables, which constitutes the studied groups or subgroups as part of the experiment, and only one dependent variable, which constitutes the results of the study. Using this type of the analysis of variance, the hypotheses concerning equality of means in subgroups for each factor are tested separately (H_0(f_A) and H_1(f_A) and H_0(f_B) and H_1(f_B)) and also equality of means in subgroups for both factors together (H_0(f_A-f_B) and H_1(f_A-f_B)). Verification of all hypotheses begins with a hypothesis that checks the existence of interactions between factors and in a situation where the interaction exists the remaining hypotheses are not verified by rejecting them. In the opposite situation it is assumed, that the model has no interaction and it goes into a test, that does not take into account the interaction. Data differentiated by classes due to both factors are analyzed in a table in which the number of rows and columns for both factors must be the same. The two-way analysis with repeats assumes the possibility of occurrence of the same data values within individual groups or subgroups with data (Snarska, 2005).

The research procedure using the two-way ANOVA analysis with repeats is very similar to the one-way ANOVA version and assumes the same stages of implementation. However, this procedure is more complex due to the existence of two independent variables and one dependent variable and also the need to test more hypotheses, including hypothesis that checks the existence of interactions between factors. The research procedure of the two-way ANOVA analysis of variance with repeats assumes the *a* kinds of experimental conditions in the scope of the first variable and the *b* kinds of experimental conditions in the scope of the second variable, in which there are *n* persons and *n* measurements for each of the *ab* combinations of experimental conditions (Ferguson, Takane, 1999; Wieczorkowska and others, 2005).

3. Selected applications of the ANOVA method

The ANOVA method has been for many years a very popular method of analyzing experimental data in research conducted by scientists around the world. The experimental research often analyzed with this method includes research in scientific fields such as: medicine, chemistry, agriculture, environmental protection, automation and robotics, economics and management, education or various applications of computer science.

In the area of widely understood **economics and management**, the ANOVA method has found many research applications, including in example studies on:

- analysis of the relationship between the level of savings in households and the various characteristics characterizing these households – application of the Kruskal-Wallis test, which is an equivalent to the non-parametric analysis of variance (Rozmus, Trzęsiok, 2017);
- using the ANOVA model to assess the variation of average indirect taxes on Polish households by decile groups – application of the one-way analysis of variance (Dobrowolska, 2016);
- the extent to which civilian spouses have used military financial education services provided by US military installations – application of the one-way analysis of variance (Plantier, Durband, 2007).

In the area of widely understood **behavioral economics**, the ANOVA method has found many research applications, including in example studies on:

- the impact of emotions on consumer behavior – application of the one-way analysis of variance (Łukasik, Witek, 2013);
- psychological mechanisms of effectiveness of managers' work and their success, understood as the result of creative management of people and information resources – application of the one-way ANOVA analysis of variance and the many-way MANOVA analysis of variance (Aranowska, Rytel, 2010);
- ways of behavior of Internet users during the search of information under the influence of experience – application of the two-way analysis of variance (Cothey, 2002).

In the area of widely understood **education**, the ANOVA method has found many research applications, including in example studies on:

- analysis of selected professional solutions in the field of personalization of e-learning for the benefit of its creators and users – application of the two-way analysis of variance (Szulc, 2018);

- experience in the use of the Moodle platform to support teaching at part-time under bachelor's program – application of the one-way analysis of variance (Klimczak, 2008).

In the area of widely understood **agriculture and environmental protection**, the ANOVA method has found many research applications, including in example studies on:

- assessment of the level of environmental genotoxicity and cytotoxicity in various places in the Baltic Sea affected by oil spills – application of the one-way analysis of variance (Barsiene, Rybakovas, Garnaga, Andreikenaite, 2012);
- understanding the impact about the influence of various quantitative and qualitative factors (e.g. January and February temperatures, genetic and geographical origin of winter wheat families) on their wintering in field experiments in 2009–2010 and carried out in 3 localities – application of the two-way analysis of variance (**Śmiałowski, Bogacka, Nita, Witkowski, 2011**).

Conclusions

The popularity of the ANOVA method for research applications in the world can be proved by the fact, that the number of indexed articles, documents and scientific papers in the Google Scholar search engine for scientists for the key password „anova", already has more than 2.37 million results (Google Scholar, 2018).

Another argument for the great popularity and wide applicability in the world of the ANOVA analysis of variance is the widely used implementation of this method in the form of computer software for many years. Appropriate use of IT tools to determine the analysis of variance significantly improves the work with different types of this method, and also allows the storage and proper organization of data. The ANOVA method is programmed and available for automatic use, among others in such popular and specialist software packages as: Microsoft Excel, Statistica or SPSS. The application of software for automatic performing the analysis of variance usually consists of: determination of the purpose of analysis and formulating hypotheses, preparation of data in terms of the number of factors and dimensions, appropriate organization and description of data in the table, activation of the appropriate calculation function and statistical analysis of the results obtained from the software (Simon, 2006; SPSS Tutorial, 2018; StatSoft Polska, 2018; Wątroba, 2002).

References

Aczel, A.D. (2000). *Statystyka w zarządzaniu*. Warszawa: Wydawnictwo Naukowe PWN.

Aranowska, E., Rytel, J. (2010). Wielowymiarowa analiza wariancji – MANOVA. *Psychologia Społeczna*, *5/2–3* (14), 117–141.

Barsiene, J., Rybakovas, A., Garnaga, G., Andreikenaite, L. (2012). Environmental genotoxicity and cytotoxicity studies in mussels before and after an oil spill at the marine oil terminal in the Baltic Sea. *Environmental Monitoring and Assessment*, *184*, 2067–2078.

Bedyńska, S., Niewiarowski, J., Cypryańska, M. (2013). *Statystyczny drogowskaz 2. Praktyczne wprowadzenie do analizy wariancji*. Warszawa: Wydawnictwo Szkoły Wyższej Psychologii Społecznej.

Cothey, V. (2002). A longitudinal study of World Wide Web users' information-searching behavior. *Journal of the American Society for Information Science and Technology*, *2* (53), 67–78.

Dobrowolska, B.M. (2016). Analiza wariancji – narzędzie do badań opodatkowania konsumpcji. *Annales Universitatis Mariae Curie-Skłodowska. Sectio H. Oeconomia*, *50/4*, 69–78.

Ferguson, G., Takane, Y. (1999). *Analiza statystyczna w psychologii i pedagogice*. Warszawa: Wydawnictwo Naukowe PWN.

Finch, H. (2005). Comparision of the performance of nonparametric and parametric MANOVA test statistics when assumptions are volated. *Methodology*, *5*, 27–38.

Google Scholar (2018). Wyniki wyszukiwania dla hasła „anova". Retrieved from https://scholar.google.pl/scholar?hl=pl&as_sdt=0%2C5&q=anova&btnG= (15.01.2019).

Klimczak, K.M. (2008). Narzędzia interaktywne – od quizów po warsztaty. Doświadczenia wykładowcy i opinie studentów. *e-Mentor*, *4* (26), 24–28.

Łukasik, I.M., Witek, A. (2013). Emocjonalne uwarunkowania zachowań konsumenckich młodych dorosłych. *Lubelski Rocznik Pedagogiczny*, *32*, 295–312.

Nowakowski, M., Mazur, M. (2017). *The impact of information usefulness of e-commerce services on users behaviors*. Cham: Springer Proceedings in Business and Economics. Neuroeconomic and Behavioral Aspects of Decision Making.

Plantier, J.H., Durband, D.B. (2007). Assessing the Use and Usefulness of Current Financial Resources for Civilian Military Spouses. *Financial Counseling and Planning*, *2* (18), 4–7.

Rozmus, D., Trzęsiok, J. (2017). Analiza czynników kształtujących oszczędności polskich gospodarstw domowych. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, *345*, 63–84.

Simon, J. (2006). *Excel. Profesjonalna analiza i prezentacja danych*. Gliwice: Wydawnictwo Helion.

Snarska, A. (2005). *Statystyka, ekonometria, prognozowanie. Ćwiczenia z Excelem*. Warszawa: Wydawnictwo Placet.

SPSS Tutorial (2018). SPSS ANOVA Tutorials – Overview. Retrieved from https://www.spss-tutorials.com/anova (15.01.2019).

StatSoft.pl (2018). Internetowy Podręcznik Statystyki StatSoft.pl. ANOVA/MANOVA. Retrieved from https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstanman.html (15.01.2019).
StatSoft Polska (2018). ANOVA. Retrieved from https://www.statsoft.pl/tag/anova (15.01.2019).
Szulc, J. (2018). Personalizacja w e-learningu. Przegląd rozwiązań. *Zeszyty Naukowe Wydziału Elektrotechniki i Automatyki Politechniki Gdańskiej*, *58*, 81–84.
Śmiałowski, T., Bogacka, M., Nita, Z., Witkowski, E. (2011). Wykorzystanie wieloczynnikowej analizy wariancji do oceny przezimowania wybranych rodów pszenicy ozimej. *Biuletyn Instytutu Hodowli i Aklimatyzacji Roślin*, *259*, 51–61.
Wątroba, J. (2002). *Wprowadzenie do analizy wariancji*. Retrieved from https://media.statsoft.pl/_old_dnn/downloads/wprowadzenie_do_analizy_wariancji.pdf (15.01.2019).
Wieczorkowska, G., Kochański, P., Eljaszuk, M. (2005). *Statystyka. Wprowadzenie do analizy danych sondażowych i eksperymentalnych*. Warszawa: Wydawnictwo Naukowe Scholar.

METODA ANOVA JAKO POPULARNE NARZĘDZIE BADAWCZE

Streszczenie

W artykule zaprezentowano charakterystykę teoretyczno-praktyczną jednej z bardzo popularnych w badaniach eksperymentalnych metod statystycznych, czyli analizy wariancji ANOVA. W pracy opisano różne rodzaje analizy wariancji, podstawowe założenia statystyczne, a także procedurę badawczą jednej z wybranych rodzajów metody ANOVA, czyli jednoczynnikowej analizy wariancji. Celem artykułu było pokazanie założeń teoretycznych i procedury badawczej metody ANOVA, przykładowego zastosowania metody w badaniach nad użytecznością informacji serwisów internetowych oraz przegląd literatury dla wybranych zastosowań metody w różnych dziedzinach naukowych. Zaprezentowany w artykule opis metody ANOVA może być przydatny w praktycznym wyborze rodzajów eksperymentów badawczych i spodziewanych wyników statystycznych, które mogą być zrealizowane z użyciem analizy wariancji.

Słowa kluczowe: analiza wariancji ANOVA, jednoczynnikowa analiza wariancji, dwuczynnikowa analiza wariancji, zastosowania metody ANOVA
Kody JEL: C1, M2

Cytowanie
Nowakowski, M. (2019). The ANOVA method as a popular research tool. *Studia i Prace WNEiZ US*, *55*, 67–77. DOI: 10.18276/sip.2019.55-06.