An Examination of Free and Open-Source Software for Mathematical Statistics Data Analysis: Prospects for Use by Students of Social and Other Sciences

UDC: 316.774:004.42:519.8 DOI: https://doi.org/10.15421/172598

Udzhmadzhuridze Heorhii
Ph.D., Senior Lecturer, https://orcid.org/0000-0002-1401-8522, g.udzhmadzhuridze@gmail.com
Kharkiv University of Humanities «People's Ukrainian Academy» (Kharkiv, Ukraine)

Abstract.
The relevance of the article is connected with the necessity for students to acquire skills in mathematical statistics data analysis. It is established that these skills are key for the preparation of specialists in social, natural, medical, and technical sciences. It is revealed that the use of proprietary software is often limited by financial and organizational factors, which complicates the learning process in higher education institutions. The potential of Free and Open-source software as an alternative to commercial products is characterized.

The purpose of the article is defined as the identification, comparison, and evaluation of the capabilities of Free and Open-source software for mathematical statistics data analysis.

Results. An analysis of proprietary and free software for mathematical statistics data analysis was conducted. The evaluation was carried out through the execution of standard data analysis procedures, including descriptive statistics, correlation, factor, regression analyses, and clustering. It was established that the JASP and Jamovi programs demonstrate the best combination of functionality, usability, and result visualization.

Conclusions. Free and Open-source software is characterized as a relevant solution for developing countries and for the organization of distance learning. It was found that JASP supports a wide range of procedures and capabilities for data analysis and visualization, making it suitable for both student learning and scientific research. It was established that Jamovi, due to its modular structure and multilingual support, is more promising for students' educational needs. It was determined that further research may be directed towards the objectification of evaluations through surveys of students, teachers, and experts, or the experimental application of these programs in the student learning process.

Keywords: Data Analysis, Free Software, Mathematical Statistics, Open-Source Software, R, SPSS, Student Learning

Дослідження Вільного та Відкритого програмного забезпечення для математико-статистичного аналізу даних: перспективи використання студентами соціальних та інших наук

Уджмаджурідзе Георгій
Харківський гуманітарний університет «Народна українська академія» (Харків, Україна)

Анотація.
Актуальність статті пов'язана з необхідністю опанування студентами навичок математико-статистичного аналізу даних. Встановлено, що такі навички є ключовими для підготовки фахівців у соціальних, природничих, медичних і технічних науках. З'ясовано, що використання пропрієтарного програмного забезпечення часто обмежується фінансовими й організаційними чинниками, що ускладнює навчальний процес у закладах вищої освіти. Охарактеризовано потенціал Вільного та Відкритого програмного забезпечення як альтернативи комерційним продуктам.

Метою статті є виявлення, порівняння та оцінювання можливостей Вільного та Відкритого програмного забезпечення для математико-статистичного аналізу даних.

Результати. Проведено аналіз пропрієтарних і безкоштовних програм для математико-статистичного аналізу даних. Оцінювання здійснено шляхом виконання стандартних процедур аналізу даних (дескриптивна статистика, кореляційний, факторний, регресійний аналіз, кластеризація). Встановлено, що програми JASP і Jamovi демонструють найкраще співвідношення функціональності, зручності та візуалізації результатів.

Висновки. Охарактеризовано Вільне та Відкрите програмне забезпечення як актуальне рішення для країн, що розвиваються, та для організації дистанційного навчання. З'ясовано, що JASP підтримує широкий спектр процедур і можливостей для візуалізації та аналізу даних, що робить її придатною як для навчання студентів, так і для наукових досліджень. Встановлено, що Jamovi, завдяки модульній структурі та підтримці багатьох мов, є більш перспективною до навчальних потреб студентів. Встановлено, що подальші дослідження можуть бути спрямовані на об'єктивізацію оцінок шляхом опитування студентів, викладачів та експертів або експериментальне застосування цих програм у процесі навчання студентів.

Ключові слова: Відкрите програмне забезпечення, Вільне програмне забезпечення, математико-статистичний аналіз даних, студентське навчання, R, SPSS

Стаття надійшла / Article arrived: 11.09.2025 Схвалено до друку / Accepted:14.10.2025

Introduction.

Problem Statement.

The issue of students in various specialties mastering mathematical statistics analysis skills is a constant challenge for universities (Kandyba, Horban, Honcharova, & Honcharov, 2023). Such skills are necessary not only for students of social sciences (Khyzhniak, 2017) but also for those studying natural sciences (Luescher, Heckel, & Grass, 2025), medicine (Ghofrani & Taherdoost, 2025), humanities (Hyvönen, 2020), engineering and technical specialities (Porter, 2019). All these sciences can use mathematical statistics methods to test hypotheses, conduct tests, perform analyses, make predictions, and classify data.

Indeed, just as mathematics is the universal language of science, so the methods based on its use have an extremely wide range of applications and possibilities. Every day, as ordinary individuals, we encounter the results of the application of mathematical statistics methods in our daily practices (Shkoliar, 2023): from medical laboratory analyses to advertising algorithms displayed to us on social media; from smartphone colour options to determining the political course of a state.

All phenomena that denote probabilities contain an element of uncertainty and may be true – they can be calculated and determined using mathematical statistics methods. Therefore, their successful mastery by students of various specialities is a socially significant problem and, accordingly, may become the subject of attention of sociology as an originally positivist discipline.

In addition, this problem is particularly acute for Ukrainian society. The outbreak of coronavirus (Covid-19) in 2019 forced higher education institutions to switch to remote learning (Novikov, 2021). In this context, proprietary software, which is traditionally widely used, became difficult for students to access from their home computers.

On the other hand, the policy on the content of academic disciplines in mathematical statistics analysis de facto requires the mandatory use of proprietary software, which creates many problems for higher education institutions as economic entities. A licence for collective use of such software for higher education institutions costs from €230 per year (SPSS Faculty Packs for Academic Staff by IBM Software Editions, n.d.), and the equipment and maintenance of a computer lab where students can use this licensed software and perform laboratory work is also a significant financial burden. In the context of war, constant air threats, economic and migration crises, as well as distance learning, it is extremely difficult for students to acquire high-quality skills in working with proprietary software for mathematical statistics data analysis.

Therefore, this article is relevant for: developing societies; developed societies where the education system is forced to use distance learning methods; Ukrainian society, as a society of war and risks. In addition, this article may be of interest to representatives of the non-social sciences who are not familiar with specialised software for mathematical statistics data analysis.

Review of Research and Publications. The problems of using proprietary statistical software were addressed by D. Saparbaeva (2025), F. Strale (Strale, et al., 2025), F. Ramdani (Ramdani, Hilmiyah, & Indriyani, 2025); the problems of using Free and Open-source statistical software were addressed by L. Ashour (Ashour, 2024), J. Zhang (Zhang, et al., 2025), (Bleichrodt, Phan, Luo, & Kirpich, 2025). This has been given attention in Ukrainian scientific discourse by A. Boba (Boba, 2020), V. Zhukovska (Zhukovska, Mosiiuk, & Komarenko, 2018), M. Malezhyk (Malezhyk, 2016), D. Mazay (Mazay & Maltseva, 2019).

However, there is a lack of developments that propose replacing proprietary software for mathematical statistics data analysis by students with Free and Open-source software that would combine powerful computing capabilities with a high level of graphical user interface (GUI).

The purpose of this article is to identify, compare and evaluate the capabilities of Free and Open-source software for mathematical statistics data analysis.

Results.

Let us start with definitions of the concepts of software and some principles of its distribution and functioning, which may be unfamiliar to representatives of the social sciences.

The ISO/IEC/IEEE Standard 24765:2010 defines software as: «Computer programs, procedures, and possibly associated documentation and data pertaining to the operation of a computer system» (Glossary Term Computer Software, 2025); B. Matthews understands this concept as: «…a collection of computer programs, procedures and documentation that perform some task on a computer system» (Matthews, Giaretta, Mcilwrath, & Conway, 2008, p. 11); B. Galler provides the following definition: «To each user of a computer, the total computing facility provided for his use, other than the hardware, is the software» (Galler, 1962, p. 6).

The definition of Free software is provided by the founder of the social movement of the same name and author of the related licence (GNU-GPL) R. Stallman: «First, the freedom to copy a program and redistribute it to your neighbors, so that they can use it as well as you. Second, the freedom to change a program, so that you can control it instead of it controlling you; for this, the source code must be made available to you» (Stallman, 1986, p. 8). K. Fogel provides the following definition: «Free software is software that may be modified and redistributed freely by anyone, with no significant restrictions on how the code may be changed, the uses to which it may be put, or the parties with whom it may be shared» (Fogel, 2005, p. 1). The GNU-GPL licence requires that all further developments and products remain free and be distributed under the GNU licence.

The concept of Open-source software is defined as follows: «Open-source software refers to software that features freely available source code, which users may view, modify, adopt, and share for both commercial and non-commercial purposes» (What is Open Source Software (OSS)?, 2024). Today, the movement has evolved from a community of IT engineers into a social movement with its own ideology: «A pervasive example is open-source software, a global public good that plays a vital role in the economy and is foundational for most technology we use today» (Hoffmann, Nagle, & Zhou, 2024, p. 1). Open-source software uses a set of «open» licences (OSI Approved Licenses), which ensure legal compliance with the principles of this movement for various programmes and products. It should be noted that, unlike the principles of the Free Software movement, the principles of the Open-source software movement are not as imperative: products and developments created on the basis of open licences may become closed.

The rest of the software is proprietary or «closed». Its source code is secret, and it is forbidden to change or modify it; it belongs to certain developers not only as intellectual property, but also as material property. It should be noted that not all Free and Open-source software is free of charge, but a significant portion of it is. The following is also true: not all proprietary software is paid, but a significant portion of it is.

Within the scope of the issues we have outlined, the main criterion for choosing software is that it is free for students. Among proprietary software for mathematical statistics data analysis, the most well-known programmes are: IBM SPSS (Start your free 14-day trial), Stata (The complete statistical software for data science), SAS (Request demos, free trials, price quotes & more), Minitab Statistical Software (Free Trial of Minitab Solution Center), JMP (Download your free trial of JMP®) – all of them are paid and offer only a limited free trial period.

Therefore, the next stage will involve comparing and evaluating the functionality of free alternatives for mathematical statistics data analysis, namely the programmes PSPP, BlueSky Statistics, JASP, Jamovi, and R-Instat. The functionality of IBM SPSS will serve as a control group (reference). The following procedures will be performed in each programme: descriptive statistics, correlation analysis, principal component factor analysis, linear regression analysis, k-means clustering. The author will assess how similar and complete the results of applying these procedures in free programmes are in comparison with SPSS. The GUI component will be assessed separately, taking into account the following parameters: multilingual interface capabilities, the ability to transfer graphs and tables to Microsoft Excel, the clarity of additional parameters when performing statistical procedures, and the availability of hints or explanations. Each category will be evaluated on a four-point scale: 0 points – if the functionality is absent; 1 point – if the functionality is worse than in SPSS; 2 points – if the functionality is similar; 3 points – if the functionality is better. The evaluation results are presented in Table 1.

Table 1

Author's assessment of Free and Open-source software for mathematical statistics data analysis

	PSPP	BlueSky	JASP	Jamovi	R-Instat
Descriptive statistics	2	1	3	2	1
Correlation analysis	1	2	3	3	1
Factor analysis	1	1	2	2	1
Regression analysis	1	1	1	1	0
Clustering by k-means	1	1	3	2	0
GUI	1	1	3	3	1
Total	7	7	15	13	4

PSPP is a free copy of SPSS with a similar design structure. The program menu is not as colourful, but it is clear and easy to read, and all elements are easily distinguishable. The program only supports English; copying analysis results to Microsoft Excel does not work correctly; a separate module is required to build graphs. In correlation analysis, only Pearson's method of measurement is available. In factor analysis, it is not possible to output KMO values and save the factorisation results as a new variable. In regression analysis, it is possible to save residual values and predicted values (which is sufficient for student level), but it is not possible to select a method for including variables in the model. In k-means clustering, it is not possible to check the quality of the model by reviewing the analysis of variance, it is not possible to save the clustering results by respondents, and there is no information about the distance between the end centres of the clusters.

BlueSky greets users with a somewhat unclear display and unusual logic for structuring basic procedures. The programme only supports English; copying analysis results to Microsoft Excel works correctly; in some procedures, graphs can be constructed immediately. In descriptive statistics, the ability to obtain information about measures of central tendency is only available for numeric variables (a limitation of the R language). In factor analysis, KMO values are not available, nor is there any information about the proportion of explained

variance; factorisation results can be saved as a separate variable. In regression analysis, it is not possible to save residual values and predicted values; it is not possible to select a method for including variables in the model, but it is possible to create a formula for your own model. In k-means clustering, it is not possible to check the quality of the model by viewing the analysis of variance; it is possible to save the clustering results by respondents; information on the distance between the endpoints of the clusters is available.

JASP is characterised by clear display and understandable logic in structuring basic procedures. The programme supports many languages; copying analysis results to Microsoft Excel works correctly. Graphs can be constructed directly in the procedure menu, and the results of additional parameters or graph construction are displayed immediately. In descriptive statistics, the ability to obtain information on measures of central tendency is only available for numeric variables (a limitation of the R language); the programme provides extensive capabilities for constructing visual graphs. All necessary parameters are present in correlation analysis; it is possible to construct graphs; a convenient table designer is available. In regression analysis, it is not possible to use automatic methods for including variables; however, informative graphs (standardised residuals) are displayed, allowing the quality of the model to be assessed; only predicted values and residuals can be saved. In k-means clustering, the programme provides additional parameters for model quality control and informative graph construction; however, only the respondent's relationship to the cluster can be saved.

Jamovi has a clear design, understandable logic for structuring procedures, and a modular ideology. The programme supports a large number of languages (including Ukrainian); analysis results are correctly transferred to Microsoft Excel; graphs can be constructed directly in the procedure menu; the results of additional parameters or graphs are displayed immediately. In descriptive statistics, the ability to obtain information on measures of central tendency is available for ordinal and nominal scales, but there are limitations for weighted data; the programme provides extensive opportunities for building visual graphs. In correlation analysis, all major methods are available; graphing is possible; a specialised module for correlation analysis is available. In regression analysis, you can save residuals, predicted values, and distances; some graphs can be constructed; it is not possible to use automatic methods for including variables in the model. In k-means clustering, there are opportunities to construct informative graphs, but only the respondent's relationship to the cluster can be saved.

Despite its origins, R-Instat only supports Windows-based computers. The design and logic of the analysis procedures are not intuitive; the ability to build graphs simultaneously is often lacking; copying results to Microsoft Excel is only possible with additional modules. Data presentation is not visual; translation into other languages is of poor quality. Descriptive statistics do not allow for the simultaneous construction of visual graphs, and there are few descriptive parameters. The results of correlation analysis are displayed inconveniently; it is possible to construct graphs. Factor analysis does not allow rotation; KMO model quality parameters are not displayed; component values are not sorted; obtained factors can be saved as separate variables; it is possible to build visual graphs. The linear regression procedure was not found. The k-means clustering method is missing.

Conclusions.

Summarising the results of our research, we can highlight several key points. Firstly, teaching students' mathematical statistics data analysis is a truly pressing issue and challenge for higher education institutions. For the Ukrainian education system, this problem arises in two main areas: a lack of material and technical resources on the part of higher education institutions; and the forced distance learning, accompanied by problems with licensing specialised software and the possible lack of the minimum necessary computer equipment among students.

Secondly, all (popular) proprietary software for mathematical statistics data analysis is paid. The developers of these programmes do not provide the opportunity to use them for free, even for educational or scientific purposes. In addition, it should be noted that none of the verified companies offer regional prices or discounts for the Ukrainian market.

Therefore, Free and Open-source software is a relevant solution for developing countries – especially for low-income countries where it is impossible to organise face-to-face learning. Our comparative analysis and evaluation of software for mathematical statistics data analysis with a similar licence revealed that these programmes differ in ideology, structure, design, functionality and user-friendliness.

The JASP programme received the highest score in the author's evaluation. It is a powerful tool based on the R language; the programme is installed as a «large» package, offering a wide range of procedures for mathematical statistics data analysis and visualisation. The programme is distinguished by its modern intuitive interface, concise and informative output tables and graphs, and the ability to display (additional) results «on the fly». At the same time, JASP may be «too big» for students, as it is more of a full-fledged alternative to IBM SPSS for researchers and professional users.

The Jamovi programme is very similar to JASP in terms of properties, design and functionality. However, it has two distinctive features: firstly, it supports more languages (including Ukrainian); secondly, it has a modular component structure, which means the programme is smaller and has lower computer system requirements. Students receive a small but powerful tool that contains everything they need to learn mathematical

statistics methods. If their interests are broader, they can install additional modules through the program's internal menu. In addition, an additional advantage of Jamovi is the possibility of limited use via a web interface (web browser), which allows students without a personal computer to also work with the programme.

Of course, our conclusions are limited. It should be understood that the author evaluated the programmes based on his personal experience, preferences and tastes. Therefore, further research in this area could be aimed, on the one hand, at objectifying the assessments by surveying students, teachers and experts and, on the other hand, at experimentally applying these programmes in the process of teaching students. Undoubtedly, the experience of using these programmes in the educational process of students may reveal new aspects that need to be taken into account.

REFERENCES

Ashour, L. (2024). A review of user-friendly freely-available statistical analysis software for medical researchers and biostatisticians. *Research in statistics, 2*(1). doi:10.1080/27684520.2024.2322630

Bleichrodt, A., Phan, A., Luo, R., & Kirpich, A. (2025). StatModPredict: A user-friendly R-Shiny interface for fitting and forecasting with statistical models. *PLoS One, 20*(8). doi:10.1371/journal.pone.0329791

Boba, A. A. (2020). Prohrammnoe obespechenye dlia provedenyia y analyza rezultatov sotsyolohycheskykh oprosov [Software for conducting and analyzing the results of sociological surveys]. *Suchasni problemy rozvytku prava ta ekonomiky v innovatsiinomu suspilstvi: zb. nauk. prats za materialamy Mizhnarodnoi naukovo-praktychnoi konferentsii* (pp. 40-47). Velyko-Tyrnovo: Access Press.

Download your free trial of JMP®. (n.d.). Retrieved October 15, 2025, from JMP Statistical Discovery: https://www.jmp.com/en/download-jmp-free-trial

Fogel, K. (2005). *What Is Free Software.* Retrieved October 14, 2025, from Open Tech Strategies: https://opentechstrategies.com/archive/what-is-free-software.pdf

Free Trial of Minitab Solution Center. (n.d.). Retrieved October 15, 2025, from Minitab: https://www.minitab.com/en-us/products/minitab-solution-center/free-trial/

Galler, B. A. (1962). Definition of software. *Communications of the ACM, 5*(1), 6-7. doi:10.1145/366243.366276

Ghofrani, A., & Taherdoost, H. (2025). Biomedical data analytics for better patient outcomes. *Drug Discovery Today, 30*(2). doi:10.1016/j.drudis.2024.104280

Glossary Term Computer Software. (n.d.). Retrieved October 15, 2025, from Defense Acquisition University: https://www.dau.edu/glossary/computer-software?

Hoffmann, M., Nagle, F., & Zhou, Y. (2024). *The Value of Open Source Software.* Harvard Business School. doi:10.2139/ssrn.4693148

Hyvönen, E. (2020). Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery. *Semantic Web, 11*(1), 187-193. doi:10.3233/SW-190386

Kandyba, I. O., Horban, H. V., Honcharova, N. V., & Honcharov, D. S. (2023). Statystychnyĭ analiz danykh pro vykorystannia obchysliuvalnykh resursiv kompiuteriv zakladu vyshchoï osvity zasobamy Python [Statistical analysis of data on the use of computing resources of higher educational institutions using Python]. *Visnyk Khersonskoho natsionalnoho tekhnichnoho universytetu, 1*(84), 148-154. doi:10.35546/kntu2078-4481.2023.1.20

Khyzhniak, L. M. (2017). Mify pro sotsialnu statystyku, profesiina pidhotovka sotsiolohiv ta informatsiina bezpeka derzhavy [Myths about social statistics, professional training of sociologists, and information security of the state]. *Visnyk Kharkivskoho natsionalnoho universytetu imeni V.N. Karazina. Seriia «Sotsiolohichni doslidzhennia suchasnoho suspilstva: metodolohiia, teoriia, metody»*(38), 92-95.

Luescher, A., Heckel, R., & Grass, R. (2025). Exploring the intersection of natural sciences and information technology via entropy and randomness. *Nat Commun, 16*, 1-10. doi:10.1038/s41467-025-62353-1

Malezhyk, M. (2016). Mozhlyvosti analizu SPSS Statistics dlia opratsiuvannia rezultativ testuvannia [SPSS Statistics analysis capabilities for processing test results]. *Naukovyi visnyk Melitopolskoho derzhavnoho pedahohichnoho universytetu. Seriia: Pedahohika, 1*(16), 157-162.

Matthews, B., Giaretta, D., Mcilwrath, B., & Conway, E. (2008). *The Significant Properties of Software: A Study.* Chilton, UK: Rutherford Appleton Laboratory.

Mazay, D., & Maltseva, K. (2019). Vplyv sotsialnoi nerivnosti na zdorovia ukrainskoho naselennia (za danymy masyvu European Social Survey, 2012) [The impact of social inequality on the health of the Ukrainian population (according to the European Social Survey, 2012)]. *Naukovi zapysky NaUKMA. Sotsiolohiia, 2*, 23-34. doi:10.18523/2617-9067.2019.2.23-34

Novikov, V. M. (2021). Osvita v umovakh pandemii [Education in a pandemic]. *Demography and social economy, 46*(4), 80-97. doi:10.15407/dse2021.04.080

OSI Approved Licenses. (n.d.). Retrieved October 14, 2025, from Open Source Initiative: https://opensource.org/licenses

Porter, A. (2019). Data Analytics for Better Informed Technology & Engineering Management. *IEEE Engineering Management Review, 47*(3), 29-32. doi:10.1109/EMR.2019.2928265

Ramdani, F., Hilmiyah, F., & Indriyani, V. (2025). The Impact of SPSS on Research Completion. *TOFEDU: The Future of Education Journal, 4*(2), 419-429. doi:10.61445/tofedu.v4i2.338

Request demos, free trials, price quotes & more. (n.d.). Retrieved October 15, 2025, from SAS: https://www.sas.com/en_us/software/how-to-buy/request-price-quote.html

Saparbaeva, D. A. (2025). Implementation of Statistical Methods in SPSS Software. *International Journal of Pedagogics, 5*(6), 440-443. doi:10.37547/ijp/Volume05Issue06-117

Shkoliar, M. V. (2023). Povsiakdenni praktyky ta styl zhyttia ukraintsiv na tli viiny [Daily practices and lifestyle of ukrainians against the background of war]. *Habitus*(46), 47-53. doi:10.32782/2663-5208.2023.46.6

SPSS Faculty Packs for Academic Staff by IBM Software Editions. (n.d.). Retrieved October 13, 2025, from Hearne Software: https://www.hearne.software/Software/SPSS-Faculty-Packs-for-Academic-Staff-by-IBM/Editions

Stallman, R. (1986). What is the Free Software Foundation? *GNU'S Bulletin, 1*(1), pp. 8-10.

Start your free 14-day trial. (n.d.). Retrieved October 15, 2025, from IBM SPSS Statistics: https://www.ibm.com/products/spss-statistics#trial

Strale, F., Riddle, I., Geng, B., Oxford, B., Kah, M., & Sherwin, R. (2025). Confirming SPSS Results With ChatGPT-4 and o3-mini Models. *Cureus, 17*(4). doi:10.7759/cureus.82005

The complete statistical software for data science. (n.d.). Retrieved October 15, 2025, from STATA: https://www.stata.com/

What is Open Source Software (OSS)? (2024). Retrieved October 14, 2025, from GitHub: https://github.com/resources/articles/what-is-open-source-software

Zhang, J., Dai, S., Guo, C., Shen, M., Liu, Y., & Cenci, J. (2025). R Language for Environmental Design Academic Research Analysis. *Land, 14*(1084). doi: 10.3390/land14051084

Zhukovska, V. V., Mosiiuk, O. O., & Komarenko, V. V. (2018). Zastosuvannia prohramnoho paketu R u naukovykh doslidzhenniakh maibutnikh filolohiv [The development of the R software package from advanced scientific philologists]. *Informatsiini tekhnolohii i zasoby navchannia, 66*(4), 272-285. doi:10.33407/itlt.v66i4.2196