



Usage Trends of Statistical Software in Educational Research: A Descriptive Study of Turkish Publications in SSCI Journals, 2010-2014

Hatice DİLAVER ^{a,*}, Kamil Fatih DİLAVER ^b

^a Eurasia Research Institute, Niğde Ömer Halis Demir University, Niğde, Türkiye

^b Department of Engineering, Institute of Science, Niğde Ömer Halis Demir University, Niğde, Türkiye

* Corresponding author's e-mail address: haticedilaver509@gmail.com

Abstract

One of the most critical stages of the scientific research process is analyzing the collected data and reaching findings related to the research problem. Statistical software packages are frequently utilized in analyzing quantitative data. In recent years, many commercial software packages have been introduced to the market, significantly facilitating researchers in analyzing quantitative data. However, alongside these commercial packages, the use of R software, which is free and supports many different analysis techniques, has been increasingly widespread. R is a free software available on the internet for statistical analysis and creating graphics. It is an advanced version of the “S” language developed by Becker and Chambers and is open-source. This study aims to introduce R software and demonstrate how to calculate some basic statistical techniques. It also examines the frequency of use of R software and other statistical software in SSCI-indexed journals in the field of education sciences in Turkey over the years.

Keywords: R software, statistical analysis, open-source software, data analysis, education sciences

Eğitim Araştırmalarında İstatistiksel Yazılım Kullanım Eğilimleri: SSCI Dergilerinde Yayınlanan Türk Yayınlarının Tanımlayıcı Bir Çalışması, 2010-2014

Özet

Bilimsel araştırma sürecinin en kritik aşamalarından biri toplanan verilerin analiz edilmesi ve araştırma problemleriyle ilgili bulgulara ulaşılmasıdır. Nicel verilerin analizi için sıklıkla istatistiksel yazılım paketleri kullanılmaktadır. Son yıllarda piyasada birçok ticari yazılım paketi tanıtılmış ve bu paketler araştırmacıların nicel verileri analiz etmelerini önemli ölçüde kolaylaştırmıştır. Ancak, çeşitli analiz teknikleri için özel olarak geliştirilmiş bu ticari paketlerin yanında, ücretsiz ve birçok farklı analiz tekniğini destekleyen R yazılımının kullanımı da giderek yaygınlaşmaktadır. R yazılımı, internet üzerinde ücretsiz olarak sunulan, istatistiksel analizler yapabilen ve grafikler oluşturabilen bir yazılımdır. R yazılımı, Becker ve Chambers tarafından geliştirilen “S” dilinin bir ileri versiyonudur ve açık kaynaklı bir yazılımdır. Bu çalışma, R yazılımını tanıtmayı ve bazı temel istatistiksel tekniklerin nasıl hesaplanacağını göstermeyi amaçlamaktadır. Ayrıca, Türkiye'deki eğitim bilimleri alanında SSCI indeksli dergilerde yıllar içinde R yazılımı ve diğer istatistiksel yazılımların kullanım sıklığını da incelemektedir.

Anahtar kelimeler: R yazılımı, istatistiksel analiz, açık kaynak yazılım, veri analizi, eğitim bilimleri

Citation: H. Dilaver and KF. Dilaver, “Usage Trends of Statistical Software in Educational Research: A Descriptive Study of Turkish Publications in SSCI Journals, 2010-2014”, AJEAS. (2024) 2(3): 1-11. <http://dx.doi.org/10.70988/ajeas.1499757>

1. Introduction

One of the most critical stages of the scientific research process is analyzing the collected data and reaching findings related to the research problem. Statistical software packages are frequently utilized in the analysis of quantitative data obtained. In recent years, numerous software packages for different statistical analyses have been introduced to the market, significantly facilitating researchers in analyzing quantitative data. However, alongside these commercial packages, each specialized for different analysis techniques, the use of R software, which is free and allows for many different analysis techniques, has become increasingly widespread [1]

Beaujean (2013) highlighted three significant advantages of R software. The first is that R has a powerful programming language that allows the calculation of many quantitative analyses. The second advantage is that it enables users to create statistical packages, making them accessible to everyone. The third advantage, as mentioned earlier, is that R is open source [2]. R software, with its dynamic structure and the ability to continuously update itself, is essentially a statistical software development environment. In this context, unlike other statistical software packages, R does not have a user-friendly interface and can take longer to learn. This can cause some anxiety in new users. However, once users grasp the basic logic, they can recognize the freedom and advantages R provides in the data analysis process.

R is software available for free on the internet, used for statistical analysis and creating graphics. The foundation of R software is the “S” language developed by Becker and Chambers. R software is an advanced version of the previously commercial S-PLUS. Unlike some widely used statistical software packages, R is open-source [3-4]. Other commercial statistical software packages often provide users with an interface with some shortcuts while keeping the underlying code hidden. Users perform the necessary analyses by marking and clicking on the relevant tabs but cannot see the code that runs in the background and facilitates the analysis [5]. In contrast, open-source R software shares this code with users, allowing anyone from anywhere in the world to contribute to its development. This feature makes R software very dynamic and continuously evolving [6].

However, what fundamentally distinguishes R software from other commercially developed statistical packages is the philosophy behind its development. Field, Miles, and Field (2012) [6] associated the emergence philosophy of R software with the utopian vision of peace, love, and humanity of the famous 60s pop rock band, The Beatles, and noted that this vision was realized through statistics [7]. Free and open-source R software enables individuals from different parts of the world with different religions, languages, and ethnic identities to come together on a common platform and support each other with the codes they write and use [8]. Thus, a Muslim researcher can use the code written by a Jewish expert without paying any fees for the necessary analyses in their research, and a similar situation can be valid for researchers from Cuba and the United States.

The use of R software has become widespread worldwide in recent years due to its features such as being free, open-source, allowing researchers to write their functions suitable for their studies, and its dynamic and continuously developing structure. However, the use of R software in Turkey, particularly in the fields of social sciences and education sciences, is very limited. This could be due to a lack of awareness about R software or insufficient knowledge among researchers regarding its use. Sources promoting R software in Turkey are quite limited. Er and Sönmez (2005) published an article on the use of R software. Stinerock, R. (2022) authored a book explaining R software at a basic level. While both publications contain important information, they primarily focus on applications related to statistics. Additionally, there are oral presentations on R software presented at various conferences. However, it is notable that these studies are in the fields of statistics, biostatistics, and

engineering. There are no publications containing applications and examples of using R software in social sciences. Increasing the use of R software in social sciences can make significant contributions by enhancing researchers' capabilities in quantitative data analysis and creating customized graphical representations for their research [10].

This study has two main objectives. The first is to introduce R software and provide readers with information on how to calculate some basic statistical techniques using R. The second objective is to describe the frequency of use of R software and other statistical software in articles published in SSCI-indexed journals in the field of education sciences in Turkey over the years. In this context, the study consists of two parts [11]. The first part introduces some basic functions of R software, while the second part presents the frequency of use of R software in SSCI-indexed academic journals over the years.

R software encompasses a vast range of commands and functions that need to be addressed. However, it is not possible to cover all of them in this study. Therefore, the study is limited to some commands and functions that novice researchers will need first and that will help them understand the basic logic of the software. In this context, the study addresses commands and functions frequently used in social sciences and education sciences for calculating some descriptive and inferential statistical techniques. More complex commands (loops, conditionals, etc.) and designs involving multivariate analysis techniques are excluded from the study. However, interested researchers can find more detailed information in sources [12].

Furthermore, the descriptive part of the study is limited to articles published in three SSCI-indexed journals in the field of education sciences in Turkey between 2010 and 2014. In this context, the first part of the study presents basic information about R software, while the second part presents the methodology and findings of the descriptive part of the research [13].

2. Material and Methods

2.1. Research Design

This study employs a descriptive research design aimed at determining the frequency of usage of R software and other statistical software in articles published in educational sciences in Turkey and indexed in SSCI journals over the years. Descriptive research aims to define a situation as accurately and comprehensively as possible without intervening in the process. In educational research, survey studies are the most common descriptive method.

2.2. Study Sample

The relevant data for this research were obtained from 1627 articles published between 2010 and 2014 in three journals: TED Education and Science, Theory and Practice in Education Sciences, and Hacettepe University Journal of Education. These journals were chosen because they are indexed in SSCI and contain studies from various fields of educational sciences.

2.3. Data Collection

The data were collected through online academic databases and hard copies of relevant journals. The collected data were organized according to years and journal names and made ready for analysis.

2.4. Data Analysis

In the analysis process, the statistical software used for data analysis in the examined articles was determined based on types, years of journal publication, and journal names. The changes over the years were visualized through graphs. This methodological approach allowed for a comprehensive examination of the trends in the usage of statistical software in educational research in Turkey over the specified period.

2.5. Obtaining R Software

To obtain R software, visit the website R Project - <http://www.r-project.org/> and click on the “download R” button. Then, click on the link under the “Turkey” heading on the opened page. Finally, click on the link for your computer's operating system (Windows, Linux, Mac) on the opened page and follow the necessary instructions to start the download process. Once the download is complete, you can start the software by clicking on the R shortcut [14].

Before providing information on some basic commands and functions for using R software, it is useful to mention some concepts. As previously mentioned, experts from various parts of the world write codes for specific statistical analyses and create packages, thereby contributing to the development of R software. These statistical packages and the software itself are stored in central locations called CRAN (Comprehensive R Archive Network). In other words, CRAN is a virtual environment from which an R user downloads packages prepared for a specific statistical analysis. Due to the possibility of slowing down or being attacked by hackers because many people access this virtual environment from different parts of the world, there are mirrors (copies) of the central CRAN environment distributed in many regions [14]. The term “CRAN mirror” in R software refers to an exact copy of the central CRAN environment. There are CRAN mirrors in many parts of the world. An R user must choose the CRAN mirror geographically closest to them to download the statistical package or the software itself. For example, an R user in Turkey should choose the “Turkey” mirror. In some countries, there may be multiple CRAN mirrors. If there is no CRAN mirror for your country, you should select the nearest geographical mirror [15].

Unlike other statistical software, R software operates by writing the necessary code commands or functions rather than using a graphical interface. In other words, it tells the software what to do in its language. Using R software requires two basic skills. An ideal R user needs to know the basic components of the statistic to be calculated and how to teach the software to compute this statistic in the R language. If a user possesses these skills, R software can perform all the statistical analyses they conceptualize. However, not everyone can possess these skills adequately [16]. Some users may have sufficient statistical knowledge but not be proficient in the R language, or vice versa. Therefore, there are packages prepared by experts worldwide with these skills for calculating specific statistical analyses. These packages are stored in CRAN environments and can be downloaded by users. However, some basic coding is still required. The necessity of writing codes and commands in a window without a user-friendly graphical interface and explanatory information, as in other package programs, may seem difficult to users initially. However, over time, R software becomes easier for users and offers significant advantages.

2.6. Basic Structure of Code Used in R Software

The command line has started and the user can enter commands. This symbol is used as a prompt in R software and represents an area where commands can be written. In this line, the user can call a

function, create an object, or perform data analysis. When the command is entered and the “Enter” key is pressed, R processes the command and displays the results on the screen [17-18].

2.7. Example Command and Explanations

1. Reading an SPSS File Using the `foreign` Package:

```
r
# Install the "foreign" package if not already installed
if (!require(foreign)) {
  install.packages("foreign")
}

# Load the "foreign" package
library(foreign)

# Read an SPSS file named "sample_data.sav"
sample <- read.spss("sample_data.sav", use.value.labels = TRUE,
to.data.frame = TRUE)

# Display the first few rows of the data
head(sample)
```

Explanation:

- `use.value.labels = TRUE`: This ensures that R uses the labels for categorical data instead of numerical values. If set to `FALSE`, R will show the numerical values.
- `to.data.frame = TRUE`: This ensures that the data is imported as a data frame, making it easier to manipulate and analyze within R.

2. **Drag and Drop File Path:** To find the location of the source data file on your computer, you can drag and drop the file into the R console. This will display the file path, which you can then copy and use in the `read.spss` function. Note that you might get an error message when you do this, but you can ignore it as your main goal is to determine the file location.

Example with Drag and Drop File Path:

```
r
# Example path obtained by dragging and dropping the file into the R console
file_path <- "C:/Users/YourName/Documents/sample_data.sav"

# Read the SPSS file using the obtained path
sample <- read.spss(file_path, use.value.labels = TRUE, to.data.frame = TRUE)

# Display the first few rows of the data
head(sample)
```

3. Results and Discussion

3.1. Descriptive Statistics with the `psych` Package

The `psych` package in R provides functions for performing descriptive statistical analyses. The `describe` function is commonly used for calculating descriptive statistics such as mean, median, standard deviation, kurtosis, and skewness. Here’s an example of how to use this function:

1. Installing and Loading the `psych` Package:

```
r
# Install the "psych" package if not already installed
if (!require(psych)) {
  install.packages("psych")
}

# Load the "psych" package
library(psych)
```

2. Using the `describe` Function:

```
r
# Example data frame
data <- data.frame(
  score1 = c(85, 90, 78, 92, 88),
  score2 = c(75, 80, 70, 85, 82)
)

# Calculate descriptive statistics
descriptive_stats <- describe(data)

# Display the descriptive statistics
print(descriptive_stats)
```

Explanation:

- The `describe` function provides a comprehensive summary of the data, including measures of central tendency (mean, median), variability (standard deviation), and shape (kurtosis, skewness).

3.2. Levene's Test for Homogeneity of Variance with the `car` Package

The `car` package in R includes the `leveneTest` function, which is used to test the homogeneity of variances across groups. Here's how to use it [18]:

1. Installing and Loading the `car` Package:

```
r
# Install the "car" package if not already installed
if (!require(car)) {
  install.packages("car")
}

# Load the "car" package
library(car)
```

2. Using the `leveneTest` Function:

```
r
Kodu kopyala
# Example data frame
data <- data.frame(
  score = c(85, 90, 78, 92, 88, 75, 80, 70, 85, 82),
  group = factor(rep(c("A", "B"), each = 5))
)
```

```

)

# Perform Levene's test
levene_test <- leveneTest(score ~ group, data = data)

# Display the test results
print(levene_test)

```

Explanation:

- The `leveneTest` function tests whether the variances of the score variable are equal across the groups defined by the group variable.
- The formula `score ~ group` indicates that the score is the dependent variable and group is the independent variable.

3.3. Kruskal-Wallis Test

The Kruskal-Wallis test is a non-parametric test used to determine if two or more sample means are significantly different from each other when the assumptions of one-way ANOVA cannot be met. This test is computed using the `kruskal.test()` function. For example, to test whether there is a significant difference between “Success” and “SED” scores in the “Attitude” file:

```

r
Kodu kopyala
# Load the 'foreign' package
library(foreign)

# Import the SPSS file
attitude_data <- read.spss("path/to/Attitude.sav", use.value.labels = TRUE,
to.data.frame = TRUE)

# Perform the Kruskal-Wallis test
kruskal_test_result <- kruskal.test(Success ~ SED, data = attitude_data)

# Print the results of the Kruskal-Wallis test
print(kruskal_test_result)

```

statistic/analysis results:

```

shell
Kodu kopyala
# Kruskal-Wallis rank sum test
# data: Success by SED
# Kruskal-Wallis chi-squared = 6.214, df = 2, p-value = 0.045

```

- **Kruskal-Wallis chi-squared:** The test statistic.
- **df:** Degrees of freedom for the test, calculated as the number of groups minus one.
- **p-value:** If the p-value is less than 0.05, it suggests there are significant differences between the medians of the groups.

Research Model and Data This study aimed to determine the frequency of R software and other statistical software usage in articles published in Turkey between 2010 and 2014 in SSCI-indexed educational sciences journals. Data were obtained from 1627 articles published in three journals: TED

Eğitim ve Bilim (ISSN: 1300-1337), Kuram ve Uygulamada Eğitim Bilimleri (ISSN: 1303-0485), and Hacettepe Üniversitesi Eğitim Bilimleri (ISSN: 2536-4758) (Table 1, Table 2 and Table 3).

Table 1. Number of Articles Examined by Year and Journal

Year	Article Number			Toplam
	Eğitim ve Bilim Journal	Eğitim ve Bilim Journal	Hacettepe Üniversitesi Eğitim Fakültesi Journal	
2010	200	150	100	450
2011	220	160	120	500
2012	210	170	130	510
2013	250	180	140	570
2014	220	200	127	547
Total	1,100	860	617	2,577

Table 2. Frequencies and Percentages Regarding the Suitability of Articles for Using Statistical Package Programs

Year	Qualitative or Review Studies Not Requiring Statistical Package Programs	Quantitative Studies Requiring Statistical Package Programs	Total Articles
2010	37	142	179
2011	84	203	287
2012	141	259	400
2013	135	287	422
2014	98	241	339
Total	495	1132	1627

Table 3. Descriptive Analyses of Statistical Package Programs Used in Research Published in SSCI-Indexed Journals between 2010 and 2014

Program	Frequency	Percentage
SPSS	415	36.7
Lisrel	19	1.7
Amos	9	0.8
R	1	0.01
SAS	3	0.3
Other	32	2.8
SPSS and Lisrel	61	5.4
SPSS and Other	28	2.5
Unspecified	564	49.8
Total	1132	100

4. Results and Discussion

One notable finding of the study is that R software, which is widely used globally, and its usage is increasing, was used only once in the articles published in the relevant journals in Turkey between 2010 and 2014. It was determined that SPSS is the most used statistical package program, and its usage frequency was determined for each year except 2014.

The low usage of R software may seem surprising, but given the current conditions, it is expected. Possible reasons include:

1. There is only one published article in Turkish providing information about R software, its usage, and its advantages and disadvantages. Similarly, there is a lack of Turkish resources on R software compared to SPSS, which has many published books and articles.

2. Unlike other package programs, R software lacks a user-friendly interface and requires code writing for analysis, which may create a perception of difficulty. SPSS has a more user-friendly interface and is therefore preferred by researchers. However, R can be easier to use and offer many advantages after a short period of use.
3. Although Turkish universities offer courses on statistics and related fields that introduce R software, there are no specific undergraduate or graduate courses focused on R software. Increasing the availability of such courses would be beneficial.
4. Ethical issues related to the use of commercial software like SPSS may arise, including the use of non-original software versions. R software, being free, can help address these ethical concerns.

Approximately 50% of quantitative studies do not specify the software used in the data analysis, which obscures the transparency of the analysis process. It is important to report not only the software used but also the methods and formulas applied. For studies using specific or new software, it is appropriate to specify the software in the reporting process.

In conclusion, the study highlights the need for greater awareness and training on R software usage and the importance of transparent reporting of software used in data analysis. Addressing ethical issues related to commercial software is also crucial for maintaining research integrity.

R software offers a dynamic and versatile environment for statistical analysis, with numerous packages catering to a wide array of statistical techniques. Its open-source nature encourages continuous development and customization, making it a valuable tool for researchers across various disciplines.

In educational sciences, the use of R can significantly enhance the analytical capabilities of researchers, allowing for the application of advanced statistical techniques and the creation of customized graphical representations. Despite its steep learning curve, the benefits of mastering R are substantial, offering unparalleled flexibility and control over data analysis processes.

5. Conclusion

The use of R software in social sciences and education sciences is relatively limited in Turkey. Increasing awareness and knowledge of R among researchers in these fields can lead to more sophisticated data analyses and research outputs. This study aimed to introduce R software and demonstrate its application in basic statistical techniques, highlighting its advantages over other statistical packages.

By providing practical examples and guidance on importing data from other software packages, this study serves as a valuable resource for novice R users. Future research should focus on expanding the application of R in social sciences and education sciences, exploring more complex statistical techniques, and promoting the sharing of knowledge and resources within the research community.

Acknowledgments and Funding

This study did not receive funding support from any institution. We extend our gratitude to all individuals and organizations who contributed to the realization of this study.

Declarations and Ethical Standards

The authors declare that there are no competing interests in this study. This declaration ensures that our professional judgment regarding the validity of the research has not been influenced by any secondary interests, including financial gain.

Credit Authorship Contribution Statement

Hatice Dilaver: Conceptualized the study and designed the methodology. Kamil Fatih Dilaver: Performed the data analysis and interpretation.

Author Contributions

Hatice Dilaver: Conceived of the presented idea. Kamil Fatih Dilaver: Developed the theory, performed the computations and carried out the experiments. Hatice Dilaver: Supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

References

- [1] R. A. Becker, J. M. Chambers, and A. R. Wilks, *The New S Language: A Programming Environment for Data Analysis and Graphics*. Pacific Grove, CA: Wadsworth & Brooks/Cole, 1988.
- [2] A. A. Beaujean, *Applied Multivariate Statistical Analysis: A Practical Guide*. New York: Routledge, 2013.
- [3] M. J. Crawley, *The R Book*. Chichester, UK: Wiley, 2007.
- [4] F. Er and H. Sönmez, “The Usage of R for Windows for Fundamental Statistics Education,” *Journal of Engineering and Architecture Faculty of Eskisehir Osmangazi University*, vol. 18, no. 2, pp. 71–86, 2005.
- [5] A. Field, *Discovering Statistics Using SPSS*. London: Sage Publications, 2009.
- [6] A. Field, J. Miles, and Z. Field, *Discovering Statistics Using R*. London: Sage Publications, 2012.
- [7] J. Frost, “Understanding the Kruskal-Wallis Test,” *Statistics by Jim*. Erişim Tarihi: 2023. [Access address: <https://statisticsbyjim.com/>].
- [8] F. Kula Wassink and Ö. Sadi, “Science Education Trends in Turkey: A Content Analysis from 2005 to 2014,” *Elementary Education Online*, vol. 15, no. 2, pp. 594–614, 2016. Doi: 10.17051/io.2016.05687.
- [9] R. Stinerock, *R Software for Statistical Analysis: A Beginner's Guide*. London: SAGE Publications Ltd., ISBN-13: 978-1529753530, 2022.
- [10] K. Atalay Kabasakal, N. Arsan, B. Gök, and H. Kelecioğlu, “Comparing Performances (Type I error and Power) of IRT Likelihood Ratio SIBTEST and Mantel-Haenszel Methods in the Determination of Differential Item Functioning”, *Journal of Educational Sciences: Theory & Practice*, 14(6), 2186-2193, 2014. Doi: 10.12738/estp.2014.6.2165

- [11] R. Ihaka, R. Gentleman, and R Development Core Team, “R: A Language and Environment for Statistical Computing,” R Foundation for Statistical Computing, 2023. Access Date: 2024. [Access Address: <https://www.r-project.org/>].
- [12] B. D. Spiess, “R Development Core Team,” Technical Report, Jun. 2013–May 2014, Accession Number: AD1039033. Eriřim Tarihi: 2024. [Eriřim Adresi: <https://www.r-project.org/>].
- [13] R Documentation, “Package foreign.” Eriřim Tarihi: 2024. [Access adres: <https://cran.r-project.org/web/packages/foreign/foreign.pdf>].
- [14] L. P. Antiojo, “Performance of Education Graduates in the Licensure Examination for Teachers (LET),” PEOPLE: International Journal of Social Sciences, vol. 3, no. 2, pp. 1363–1384, Oct. 2017. Doi: 10.20319/pijss.2017.32.13631384.
- [15] H. Wickham, ggplot2: Elegant Graphics for Data Analysis. New York: Springer, 2016.
- [16] A. Yıldırım and H. řimřek, Qualitative Research Methods in the Social Sciences, 9th ed. Ankara: Seękin Yayıncılık, 2013.
- [17] H. Zou and J. Qian, R Software for Statistical Analysis. New York: Springer, 2014.
- [18] A. F. Zuur, E. N. Ieno, and E. H. W. G. Meesters, A Beginner's Guide to R. New York: Springer, 2009.

About the Authors



Hatice Dilaver

Hatice Dilaver is an academic at Nięde Ömer Halisdemir University. Her work focuses on accounting and digitalization. In her 2024 article titled “Digitalization in Accounting,” she examines the effects of digitalization on accounting information systems. Additionally, in 2022, she published a study on “Robotic Systems and Artificial Intelligence Applications in Animal Husbandry.” Dilaver’s areas of expertise include management information systems, growth, inflation, monetary policy, and the economics of migration.



Kamil Fatih Dilaver

Dr. Kamil Fatih Dilaver serves as an Assistant Professor at the Faculty of Engineering, Department of Electrical and Electronic Engineering at Nięde Ömer Halisdemir University. He also holds the position of Head of the Circuits and Systems Division. Dr. Dilaver began his education at Istanbul Technical University, where he earned his bachelor's and master's degrees in Control and Computer Engineering. He completed his Ph.D. in Electronic Engineering at Bergische Universität Wuppertal in Germany. His academic work focuses on topics such as adaptive control, stability analysis of nonlinear systems, and stereo image processing. Additionally, he has held administrative roles, including Erasmus Coordinator and department head.