Black Sea Journal of Statistics

Doi: 10.5281/zenodo.18234368

Open Access Journal
e-ISSN: 3108-4265

Research Article
Volume 1 - Issue 1: 1-6 / June 2025

USAGE OF R SOFTWARE IN EDUCATIONAL RESEARCH IN TÜRKİYE: FREQUENCY, FUNCTIONS, AND ADVANTAGES

Hatice DİLAVER[1]*, Kâmil Fatih DİLAVER[2]

[1]*Niğde Ömer Halisdemir University, Department of Eurasia Studies, 51200, Niğde, Türkiye*
[2]*Niğde Ömer Halisdemir University, Faculty of Engineering, Department of Electric and Electronics, 51200 Niğde, Türkiye*

Abstract: This study explores the use of R statistical software in educational sciences in Türkiye. Although R is a powerful, free, and open-source software widely used globally, its adoption in Türkiye, especially in social sciences, remains very limited. The study investigates the frequency of R and other statistical software in SSCI-indexed educational journals between 2010 and 2014. Results reveal that SPSS is overwhelmingly preferred, while R was used in only one instance. The advantages of R, its basic commands, assumptions testing, descriptive and inferential statistics, and non-parametric tests are also presented. The study concludes with recommendations for wider use of R in Turkish academia.

Keywords: Educational, Function, Statistical software, Frequency

***Corresponding author:** Niğde Ömer Halisdemir University, Department of Eurasia Studies, 51200, Niğde, Türkiye
E mail: haticedilaver509@gmail.com (H. DİLAVER)
Hatice DİLAVER https://orcid.org/0000-0002-4484-5297
Kâmil Fatih DİLAVER https://orcid.org/0000-0001-7557-9238

Received: April 22, 2025
Accepted: May 23, 2025
Published: June 15, 2025

Cite as: Dilaver, H., & Dilaver, K. F. (2025). Usage of R software in educational research in Türkiye: Frequency, functions, and advantages. *Black Sea Journal of Statistics, 1*(1), 1–6.

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 to analyze quantitative data. In recent years, many statistical packages have been developed for various types of analysis. These packages offer significant convenience to researchers during the data analysis process. In addition to commercial software specifically designed for different statistical techniques, the use of the free R software has been increasingly widespread.

R is a free software used for statistical analysis and graphical presentation and is accessible via the internet. The foundation of R is the "S" programming language developed by Becker and Chambers. R is an advanced version of the previously released commercial software known as S-PLUS (Er and Sönmez, 2005). Unlike many widely used commercial software, R is open source (R Development Core Team, 2002). While commercial software generally hides the underlying code and offers a graphical user interface for executing commands, R openly shares its source code with users. This transparency allows individuals from all over the world to contribute to the software's development, making R a dynamic and constantly evolving platform.

What sets R apart from other commercially driven software is the philosophy behind its development. Field at al. (2012) liken the philosophy of R's creation to the utopian vision of peace, love, and humanity popularized by The Beatles in the 1960s—realized through the medium of statistics. R enables individuals from diverse cultural and religious backgrounds around the world to contribute code to a shared platform, fostering global collaboration. For instance, a Muslim researcher can utilize code written by a Jewish expert to perform an analysis without any cost. Similarly, a Cuban and an American researcher can benefit from each other's contributions.

Beaujean (2013) emphasizes three main advantages of R. First, it is a powerful programming language capable of performing a wide range of quantitative analyses. Second, it allows users to develop and share statistical packages that others can access freely. Third, as previously noted, it is an open-source software.

Although R does not have a user-friendly graphical interface like some other statistical packages and may initially seem difficult to learn, it offers considerable flexibility and advantages once its basic logic is understood. Thanks to its free access, open-source structure, flexibility in custom function creation, and dynamic ecosystem, the use of R has grown globally in recent years. However, in Türkiye its usage remains limited. This could be due to unfamiliarity with the software or lack of training and resources on how to use it effectively.

In Türkiye, publications introducing R are quite limited. Er and Sönmez (2005) published an article on the use of R. Although both are informative, they mainly focus on statistical applications. Additionally, a few oral

presentations on R have been delivered at conferences (Özdemir et al., 2010; Baydoğan et al., 2014), primarily in statistics, bio-statistics, and engineering disciplines. There are no known publications demonstrating the use of R in social sciences. Yet promoting R in the social sciences can significantly enhance researchers' quantitative analysis capabilities and improve graphical representation of findings.

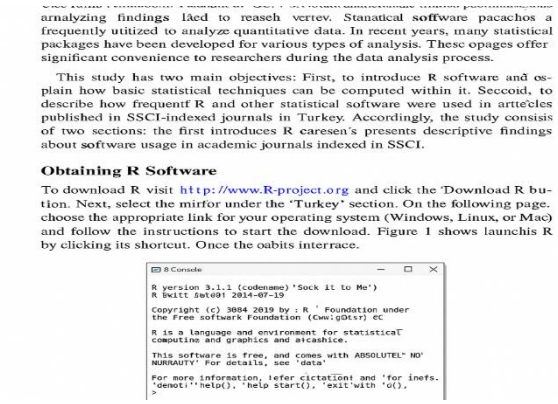
This study has two main objectives: First, to introduce R software and explain how basic statistical techniques can be computed within it. Second, to describe how frequently R and other statistical software were used in articles published in SSCI-indexed journals in the field of educational sciences in Türkiye. Accordingly, the study consists of two sections: the first introduces basic R functions; the second presented descriptive findings about software usage in academic journals indexed in SSCI.

R has a broad scope, with hundreds of commands and functions, which cannot all be covered here. Therefore, this study is limited to essential commands and basic statistical functions relevant to researchers new to R and frequently used in social sciences and education research. More advanced functions (e.g., loops, conditionals) and multivariate techniques are excluded but can be explored in more detail in sources such as Field et al. (2012), Zuur et al. (2009), and Crawley (2007).

The descriptive section is limited to articles published in three SSCI-indexed educational journals in Türkiye between 2010 and 2014. The structure of the study includes basic information about R in the first part and methodology/findings in the second.

2. R Software

To download R, visit http://www.R-project.org and click the "Download R" button. Next, select the mirror under the "Türkiye" section. On the following page, choose the appropriate link for your operating system (Windows, Linux, or Mac) and follow the instructions to start the download. Once the download is complete, you can launch R by clicking its shortcut. The basic interface is shown in Figure 1.


arnalyzing findings lâed to reaseh vertev. Stanatical software pacachos a frequently utilized to analyze quantitative data. In recent years, many statistical packages have been developed for various types of analysis. These opages offer significant convenience to researchers during the data analysis process.

This study has two main objectives: First, to introduce R software and explain how basic statistical techniques can be computed within it. Seccoid, to describe how frequentf R and other statistical software were used in artte'cles published in SSCI-indexed journals in Turkey. Accordingly, the study consisis of two sections: the first introduces R caresen's presents descriptive findings about software usage in academic journals indexed in SSCI.

Obtaining R Software

To download R visit http://www.R-project.org and click the 'Download R button. Next, select the mirfor under the 'Turkey' section. On the following page. choose the appropriate link for your operating system (Windows, Linux, or Mac) and follow the instructions to start the download. Figure 1 shows launchis R by clicking its shortcut. Once the oabits interrace.

Figure 1. The basic interface of R software.

3. Materials and Methods

3.1. Research Model

This study, which aims to determine the frequency of R and other statistical software used in articles published in SSCI-indexed educational journals in Türkiye between 2010 and 2014, is a descriptive research in the form of a survey. Descriptive research attempts to define a given situation as accurately and completely as possible, without interfering with the process. In educational research, survey studies are among the most commonly used descriptive methods (Büyüköztürk et al., 2009).

Sample

The data were obtained from 1,627 articles published between 2010 and 2014 in three SSCI-indexed educational journals in Türkiye:

- Education and Science
- Educational Sciences: Theory and Practice (Kuram ve Uygulamada Eğitim Bilimleri)
- Hacettepe University Journal of Education (Hacettepe Üniversitesi Eğitim Fakültesi Dergisi)

These journals were selected because they are indexed in SSCI and cover a variety of topics across different fields within educational sciences.

Table 1. Number of articles reviewed by year and journal

Year	Education and Science	ESTP (KUYEB)	Hacettepe Journal	Total
2010	57	63	59	179
2011	94	115	78	287
2012	96	174	130	400
2013	115	134	173	422
2014	162	100	77	339
Total	524	586	517	1627

3.2. Data Collection

Data were collected from both online academic databases and printed versions (hardcopies) of the selected journals. The articles were sorted by year and journal before analysis.

3.3. Data Analysis

During analysis, the statistical software used in each article was identified and categorized by:

- Type of statistical software
- Publication year
- Journal title

Frequencies and percentages were calculated, and trends across years were visualized using charts.

3.4. Basic Structure of R Commands

In R, the symbol ">" indicates the beginning of a new line for entering commands, and it appears at the start of every new command line. R commands typically consist of two key components: objects and functions. These two components are separated using the symbol <-. The part on the left side of this symbol represents the object, while the part on the right represents the function (i.e., object

<- function). The <- symbol instructs R to create the specified object using the function defined on the right-hand side.

In R, anything that is created is treated as an “object” (Knell, 2014). An object can be a variable or a statistical model, while functions are the instructions given to R to create the object.

For example, the command:

R

point<- c(45, 50, 55, 60, 65)

tells R to create a variable called point from the numerical values listed using the “concatenate” function, abbreviated as c( ). Similarly, when creating a variable with categorical data, the c( ) function is also used, but since the data are qualitative, each value must be enclosed in quotation marks:

R

gender <- c("female", "male", "female", "male")

R then identifies whether a variable is qualitative or quantitative based on the input format. When a user types puan or gender and presses enter, R will display the corresponding data.

An important point to note is that R is case-sensitive. A command typed in uppercase will be interpreted differently than the same command in lowercase.

If a function is incomplete or incorrectly formatted, R will display the symbol “+” on the next line instead of “>” as a prompt. This indicates that the previous command was incomplete and requires correction. For instance:

R

puan <- c(45, 60, 75

+

Here, the user forgot to close the parenthesis, so R awaits the completion of the command.

3.5. Data Transfer from Other Software to R

One of the important features of R is the ability to import data from other software platforms such as SPSS or Excel. Although data can be created directly in R, in many cases, researchers may want to import datasets already created in other software. In principle, data can be transferred from all major statistical programs into R. However, R needs to be informed about the structure and origin of the dataset through specific functions.

Once imported and saved, a dataset remains accessible within R unless manually deleted, offering practical advantages over some other software.

3.5.1. Importing Excel (CSV) files into R

One of the simplest and most common methods for importing data is using a CSV file (Comma-Separated Values) exported from Excel.

Example command:

R

data <- read.table("C:/Users/username/Desktop/data.csv", header=TRUE, sep=";")

data is the name of the object to be created.

read.table() is the function used to read the file.

header=TRUE tells R that the first row contains variable names.

sep=";" indicates that the values are separated by semicolons.

If you are unsure about the file path or prefer to choose the file manually, use:

R

data <- read.csv(file.choose())

This command opens a dialog box allowing you to select the file interactively.

3.5.2. Importing SPSS files into R

For SPSS data, you must install and load the foreign package (if not already installed). Then use the read.spss() function:

R

library(foreign)

data <- read.spss("C:/Users/username/Desktop/data.sav", use.value.labels=TRUE, to.data.frame=TRUE)

use.value.labels=TRUE ensures that R uses the value labels (e.g., "male", "female") instead of the numeric codes.

to.data.frame=TRUE tells R to convert the dataset into a data frame format, making it easier to work with.

You can also choose an SPSS file interactively:

R

data <- read.spss(file.choose(), use.value.labels=TRUE, to.data.frame=TRUE)

Note: For determining a file’s full path, you can drag and drop the file into the R console. R will show a file path error, but you can ignore the error and just copy the displayed file path.

Other useful import functions:

read.fwf() – for fixed-width text files.

read.dta() – for importing Stata .dta files.

3.6. Installing Statistical Packages in R

When R is initially installed, it includes a set of core packages. However, for conducting more specific or advanced statistical analyses, additional packages must be downloaded. For instance, a psychometrician estimating item parameters based on Item Response Theory (IRT) might use the “sirt” package, while someone conducting Confirmatory Factor Analysis might use the “sem” package.

To install a package in R, the following function is used:

R

install.packages("package_name")

Example:

R

install.packages("sirt")

Important: The name of the package must be enclosed in both parentheses and quotation marks.

After entering this command and pressing Enter, a dialog box will appear asking you to choose a CRAN mirror. It is recommended to select the mirror geographically closest to you. As of now, there are more than 5,000 statistical packages available, and this number continues to grow.

Once downloaded, the package will be stored in your local R library. To activate a package before analysis, use

the following function:
R
library(package_name)
Example:
R
library(sirt)
Note: This time the package name is written inside the parentheses without quotation marks.
R contains thousands of packages developed by experts worldwide. Sometimes, different packages may contain functions with the same name but different purposes. For example, both the "car" and "Hmisc" packages contain a function called recode. If both are installed, R may be unable to determine which function to use unless specified. Therefore, you must load the correct package using library() to ensure the right version is used.
To update a package that has already been installed:
R
update.packages()

3.7. R Software and Mobile Applications

Unlike many statistical software programs, R does not require powerful computer hardware and can be used easily on smartphones or tablets. For mobile devices, there is an application called "R Console", available on both iOS and Android platforms.
There are two versions:

- R Console Free: Available at no cost but with limited features.
- R Console Premium: Offers full desktop-level capabilities with a small one-time fee.

For data transfer to the mobile app, services like Dropbox or OneDrive are used. You can upload your data files to one of these cloud platforms and then access them directly from the mobile R Console app—allowing you to run statistical analyses on-the-go, such as while riding the subway or relaxing in a park.
If R Console is not found in your app store, you can search through a web browser and download it directly. In some cases, small adjustments in your device's security settings may be needed for installation.

3.8. Basic Functions in R – Descriptive Statistics

- summary(dataset_name)
  This function provides basic descriptive statistics (minimum, maximum, first and third quartiles, median, and mean) for all variables in the specified dataset.
- mean(dataset$variable)
  This calculates the arithmetic mean of a selected variable. The $ symbol is used to access a specific variable within a dataset.
- sd(dataset$variable)
  Computes the standard deviation of the selected variable.
- describe(dataset)
  This function gives detailed descriptive statistics including mean, median, standard deviation, kurtosis,andskewness.
  Requires the psych package to be installed.

3.8.1. Testing assumptions

Univariate Normality assumption

- shapiro.test(dataset$variable)
  Performs the Shapiro-Wilk test for normality. If the result is not statistically significant (p > .05), the distribution is considered normal.
- by(dataset$variable, dataset$group, shapiro.test)
  Tests normality within each group separately, often used before conducting independent-samples t-tests.
- ks.test(variable, pnorm)
  Performs the Kolmogorov–Smirnov test, comparing the distribution of the variable to a standard normal distribution.
  Used when sample size > 50, although ks.test is originally designed to compare two distributions.

Q-Q Plot (Graphical Normality check)

qqnorm(dataset$variable) qqline(dataset$variable, col="red")
yaml
These functions generate a Q-Q plot along with a reference line to visually assess normality.
Statistical Summary for Normality
- `stat.desc(dataset, basic=FALSE, norm=TRUE)`
Requires the **`pastecs`** package. This function provides skewness, kurtosis, and their standard error ratios for evaluating normality.
Homogeneity of Variance
- `leveneTest(dataset$dependent_variable ~ dataset$group_variable)`
Performs Levene's Test to check if variances are equal across groups.
Requires the **`car`** package.
If **p > .05**, the homogeneity assumption is met.
Inferential Statistics
Independent-Samples t-test
- `t.test(dependent ~ group, data=dataset)`
Used when comparing means of two independent groups. R automatically performs Welch's t-test, a robust alternative to Student's t-test.
Paired-Samples t-test
- `t.test(dataset$pretest, dataset$posttest, paired=TRUE)`
Tests whether the mean difference between two related groups (e.g., pre/post) is statistically significant.
Any of these tests can also be **stored as objects** for easier access:
```R
result1 <- t.test(dataset$pretest, dataset$posttest, paired=TRUE)
Later, typing result1 will display the output again.

**One-Way ANOVA**

anova_result <- aov(dependent ~ group, data=dataset) summary(anova_result)
bash
If the F-test is significant, **post hoc tests** can identify which groups differ:
- ```R
pairwise.t.test(dataset$dependent, dataset$group,
```

p.adjust.method="bonferroni")
Change "bonferroni" to "BH" for Benjamini-Hochberg or explore Tukey and Dunnet tests using the multcomp package.
Non-Parametric Tests
If normality or homogeneity assumptions are violated:
Wilcoxon Rank Sum Test (Mann-Whitney U)
wilcox.test(dependent ~ group, data=dataset)
Non-parametric alternative to independent-samples t-test.
Wilcoxon Signed-Rank Test
wilcox.test(dataset$pretest, dataset$posttest, paired=TRUE)
Used when assumptions for paired-samples t-test are violated.
Kruskal-Wallis Test
kruskal.test(dependent ~ group, data=dataset)
Non-parametric alternative to one-way ANOVA. However, it does not return group rank means.
To calculate rank means:
R
dataset$rank <- rank(dataset$dependent)
tapply(dataset$rank, dataset$group, mean)
For pairwise comparisons after a significant Kruskal-Wallis result, use wilcox.test() as explained earlier.

4. Results

4.1. Use of Statistical Software in SSCI-Indexed Turkish Journals (2010–2014)

Articles were first categorized based on whether they required statistical software (i.e., quantitative studies) or not (e.g., qualitative studies, literature reviews).

Table 2. Suitability of articles for statistical software use

Year	Qualitative/Review Studies	Quantitative Studies	Total
2010	37 (20.7%)	142 (79.3%)	179
2011	84 (29.3%)	203 (70.7%)	287
2012	141 (35.2%)	259 (64.8%)	400
2013	135 (31.9%)	287 (68.1%)	422
2014	98 (28.9%)	241 (71.1%)	339
Total	495 (30.4%)	1132 (69.6%)	1627

Out of 1,627 articles, 1,132 (69.6%) required the use of statistical software.
The statistical programs were categorized as follows:

- SPSS
- LISREL
- AMOS
- R
- SAS
- Others (e.g., Facets, Bilog, Multilog, Genova)
- SPSS + LISREL
- SPSS + Others
- Unspecified

Table 3. Statistical software used in quantitative studies (2010–2014)

Software	2010	2011	2012	2013	2014	Total (%)
SPSS	53	66	94	117	85	415 (36.7)
LISREL	–	5	2	4	8	19 (1.7)
AMOS	–	2	3	2	2	9 (0.8)
R	–	–	–	1	–	1 (0.01)
SAS	–	2	1	–	–	3 (0.3)
Others	3	8	5	3	13	32 (2.8)
SPSS + LISREL	8	10	14	12	17	61 (5.4)
SPSS + Others	5	1	5	3	14	28 (2.5)
Unspecified	73	109	135	145	102	564 (49.8)

Most notable findings:

- In 49.8% of the articles, the statistical software used was not reported.
- SPSS was the most used software (in about 45% of all quantitative studies when combined with SPSS+LISREL and SPSS+Others).
- R was used in only 1 article (0.01%).
- LISREL, AMOS, and SAS were rarely used.
- Use of more specialized software (e.g., Facets, Genova) was limited to 2.8%.

5. Discussion and Conclusion

One of the most striking findings is that R software, despite being free, open-source, and widely used internationally, was used in only 1 article among 1,132 quantitative studies. In contrast, SPSS was dominant across all years except 2014.

5.1. Possible Reasons for Underuse of R in Türkiye

- Lack of Turkish-language resources: Although some basic materials exist (Er and Sönmez, 2005), no R-based educational statistics book has been published in Türkiye. In contrast, SPSS-related books and guides are abundant.
- Misconceptions about R's difficulty: Since R lacks a graphical user interface and relies on scripting, it may be perceived as user-unfriendly. However, once familiar with R's logic, users find it efficient and flexible (Field, 2009).
- Limited inclusion in curricula: While R is taught in statistics, engineering, and biostatistics departments, it is not yet integrated into education faculties. Graduate-level courses in quantitative

analysis rarely incorporate R.

- Lack of in-service training and institutional promotion: Researchers may benefit from workshops and professional development that introduce R.

5.2. Ethical Concerns with Commercial Software

Some researchers may use unofficial or pirated versions of SPSS. R, being free and legal, offers an ethical alternative that avoids licensing issues.

5.3. On Reporting Software Use in Studies

Nearly 50% of the studies did not report the software used. Emphasizing statistical methods and rationale over software names is more academically meaningful—unless the software is novel or specialized (e.g., WinBUGS, Multilog, Facets).

When using open-source tools like R, sharing code improves transparency and reproducibility.

Author Contributions

The percentages of all authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	H.D.	K.F.D.
C	100	
D	100	
S		100
DCP	50	50
DAI	50	50
L	50	50
W	50	50
CR	50	50
SR	50	50
PM	50	50
FA	50	50

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declare that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans.

References

Baydoğan, M. G., Orbay, B., & Çetin, U. (2014). R ile programlamaya giriş ve uygulamalar. *XIX. Türkiye'de İnternet Konferansı Bildirileri Kitabı*. (ss. 57).

Beaujean, A. A. (2013). Factor analysis using R. *Practical Assessment, Research & Evaluation, 18*(4), n4.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2009). *Bilimsel araştırma yöntemleri* (5. baskı). Pegem Yayıncılık.

Crawley, M. J. (2007). *The R book*. Wiley.

Er, F., & Sönmez, H. (2005). The usage of R Windows for the fundamental statistics education. *Journal of Engineering and Architecture Faculty of Eskişehir Osmangazi University, 18*(2), 1–11.

Field, A. (2009). *Discovering statistics using SPSS* (2nd ed.). Sage Publications.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage Publications.

Knell, J. R. (2014). *Introductory R*. Knell.

Özdemir, A. F., Yıldıztepe, E., & Binar, M. (2010). İstatistiksel yazılım geliştirme ortamı: R. *XII. Akademik Bilişim Konferansı Bildirileri Kitabı*. (ss. 63).

R Development Core Team. (2002). *An introduction to R*. Network Theory.

Zuur, A. F., Ieno, E. N., & Meesters, E. H. W. G. (2009). *A beginner's guide to R*. Springer-Verlag.