Title: Choosing the Right Statistical Software for Radiology Research: A Critical Evaluation of Pros and Cons

Author: Arya H. Mirzaian, MD MPH

Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts; Harvard Medical School, Boston, Massachusetts. Ahaj-mirzaian@bwh.harvard.edu

Introduction

The field of radiology research has experienced a significant transformation in recent years with the growing availability of high-quality imaging modalities and advanced data collection techniques. This has enabled researchers to address a broad spectrum of clinical questions and also increased the complexity of data analysis. Different research objectives require different study designs, such as survival studies or meta-analyses, which in turn demand specific statistical software for accurate data analysis and interpretation (1). While numerous statistical software programs accessible to researchers, there is no widely accepted consensus on which one is superior for a particular analytical need. Hence, this review aims to provide an overview of the commonly used statistical software in radiology research and evaluate the advantages and disadvantages of each program to assist junior investigators and other researchers in selecting the most appropriate software for reliable and accurate data analysis.

Statistical Software Popularity: A Review of Published Studies in the Radiology Journal

The full texts of 100 original research articles published in the Radiology journal from November 2022 to April 2023 were consecutively assessed, and the utilized statistical software and methods in each study were extracted. Out of these papers, 13% used more than one statistical software. The most commonly used statistical software was R, which was mentioned in 43% of the articles, followed by SPSS (Statistical Package for the Social Sciences) (22%), SAS (Statistical Analysis System) (12%), STATA (10%), and GraphPad Prism (8%). Other tools, including

Python were used in less than 5% of studies. Besides the normality test to assess data distribution, the most commonly used statistical methods were non-parametric tests such as Mann-Whitney U and Wilcoxon Signed-Rank test (23%), regression analysis such as logistic and COX regression (18%), and correlation analysis such as Pearson or Spearman correlation and Intraclass Correlation Coefficient (18%).

Comparison of Statistical Software

Statistical programs have their own strengths and weaknesses, and differ significantly in terms of capabilities, ease-to-use, cost, and popularity. Choosing the right program is crucial for accurate and efficient data analysis and should be made based on the specific research needs, investigators' level of expertise, speed, etc. Here, the pros and cons of each statistical software was compiled for radiology research purposes.

Command-Line Interface Programs

R

R is a highly capable and open-source software that is designed to handle large datasets (2). With over 6000 available packages, it provides a broad range of functions that cover nearly all statistical methods. Additionally, it is an expectational tool for data visualizations. Despite numerous advantages, R does require users to become familiar with its syntax. Lack of user-friendly interface, comparing point-and-click programs, may present challenges for novice users. R is an ideal tool for tasks involving large datasets, meta-analysis, complex methods such as mixed models, and data visualization.

Broadly speaking, R is the most recommended statistical tool ignoring the obstacle of learning syntax. For new users, we recommend using the RStudio environment which simplifies the learning process (3). There are numerous online resources available to help users get started. One good example is "An Introduction to Statistical Learning," which covers key topics and is suitable for those beginning to explore data analysis or looking to advance their skills (4).

Python

Although Python was used in less than 5% of the radiology papers for the purpose of statistical analysis, it offers several unique features that make it a preferred tool for many researchers (5). Like R, Python is open-source software that covers almost all analytical methods required in radiology research and is suitable for large datasets. Python has superior capabilities for deep learning and machine learning, as it is a programming language with an extremely large community of support from all fields. Despite Python's advantages, R has more specialized libraries for sophisticated statistical methods and is generally considered a better tool for data visualization.

For researchers who seek to acquire coding skills and intend to conduct deep learning projects, Python is the ideal choice, despite being challenging to learn. Numerous resources are available to assist users in getting started with Python, including "Practical Statistics for Data Scientists" which is a valuable example of such resources for scientific research purposes (6).

SAS

SAS is another widely used software that offers an extensive range of statistical methods (7). Here, a comparison between SAS and R was made. SAS has an advantage over R when it comes to speed and efficiency when working on a very large dataset. Additionally, SAS offers both command-line and point-and-click interfaces, making it easier to learn for users without a programming background. Regarding disadvantages, SAS requires licensing which may be a limiting factor. R also has an advantage when it comes to graphics and novel methods.

Overall, SAS is a great but expensive alternative for beginners struggling with R coding or researchers who are seeking a faster tool to handle their large datasets. With many free resources available, including the "SAS University Edition" which is a great free resource offers by SAS, researchers can easily learn SAS programming (8).

STATA

STATA has features similar to SAS and offers both command-line and point-and-click interfaces (9). STATA is an easier program given a variety of built-in functions and is generally less expensive comparing SAS. However, STATA may not be as efficient as SAS when handling large

datasets and have limitations when dealing with complex statistical models. In radiology, STATA remains a popular choice for conducting meta-analysis (10).

Point-and-Click Programs

SPSS

SPSS is a user-friendly and relatively easy-to-learn software making it popular for researchers with basic statistical and coding skills (11). SPSS covers most, but not all, statistical functions required in radiology research and provides a limited range of graphical output options. SPSS is an appropriatetool for analysis tasks involving smaller and less complex datasets and routine methods such as regression and analysis of variance (ANOVA).

The cost of the software may be a limiting factor. Researchers can contact their institutions as most provide free or cheaper subscriptions. SPSS has a strong support community with numerous online resources. Among those, online tutorials and YouTube videos are preferred, comparing books/articles. Some advanced statistical methods not covered by the basic version, such as mediation analysis, can added as extensions by installing additional software. Despite advantages, researcher should consider other software for meta-analysis and novel statical methods such as Bayesian analysis.

Other Programs

While SPSS is frequently used in radiology research, it has several limitations, particularly in data visualization, meta-analysis, and sample size calculation. As such, we recommend considering complementary tools to overcome these shortcomings, as detailed below.

GraphPad Prism is user-friendly and well-regarded for its highly customizable graphics capabilities, making it an excellent choice for creating visualizations for simple analyses such as bar plots for ANOVA (12). However, GraphPad Prism is not open-source, has limited flexibility for advanced methods, and not be suitable for large datasets. Many researchers combine using SPSS for data analysis with GraphPad Prism for generating graphics.

Several software options are available to aid the process of power analysis and sample size calculation. Among those, G*Power is an open-source and user-friendly option widely used for its versatility (13).

For meta-analysis, besides R and STATA as the most popular tools, some alternatives are available for those seeking an easier option**.** Comprehensive Meta-Analysis (CMA) is one such software, provides a user-friendly interface and offers a range of analysis options, including meta-regression. However, it is not optimal for diagnostic accuracy meta-analysis (14). Meta-DiSc is another user-friendly software that specializes in diagnostic accuracy meta-analysis (15).

Lastly, for those who find even SPSS challenging to use, several online platforms are available to simplify the data analysis process. One such platform is https://www.socscistatistics.com/, where users can input or copy/paste their data and select the desired analytical methods to obtain the results.

Conclusions

This article provides an overview of the advantages and disadvantages of the commonly used statistical software in radiology research to assist researchers in selecting and learning the most suitable program. Selection of an appropriate statistical software should be based on various factors, such as the nature of the research, level of expertise, size and complexity of the data, as well as the software's speed and cost. By carefully weighing these factors, researchers can make informed decisions about the most appropriate statistical software to use for their research.

References:

1. Anvari A, Halpern EF, Samir AE. Statistics 101 for radiologists. Radiographics. 2015;35(6):1789–801.

2. R: The R Project for Statistical Computing [Internet]. Available from: https://www.r-project.org/

3. Posit [Internet]. Posit. Available from: https://www.posit.co/

4. An Introduction to Statistical Learning [Internet]. An Introduction to Statistical Learning. Available from: https://www.statlearning.com

5. Welcome to Python.org [Internet]. Python.org. Available from: https://www.python.org/

6. Practical Statistics for Data Scientists, 2nd Edition [Book] [Internet]. Available from: https://www.oreilly.com/library/view/practical-statistics-for/9781492072935/

7. Products & Solutions [Internet]. Available from: https://www.sas.com/en_us/software/all-products.html

8. SAS Analytics U | SAS [Internet]. Available from: https://www.sas.com/no_no/offers/sas-analytics-u/overview.html

9. Statistical software for data science | Stata [Internet]. Available from: https://www.stata.com/

10. Meta-analysis | Stata [Internet]. Available from: https://www.stata.com/features/meta-analysis/

11. SPSS Software | IBM [Internet]. Available from: https://www.ibm.com/spss

12. Home - GraphPad [Internet]. Available from: https://www.graphpad.com/

13. Universität Düsseldorf: G*Power [Internet]. Available from: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

14. Why Perform a Meta-Analysis - Overview 101 [Internet]. Available from: https://www.meta-analysis.com/pages/why_do.php?cart=BMZG8601859

15. Meta-DiSc. Get the software safely and easily. [Internet]. Software Informer. Available from: https://meta-disc.software.informer.com/

Table. Comparing pros and cons of commonly used statistical software in radiology research.

Program	Popularity in Radiology	Open Source	Cost * (USD)	Difficulty Level		Large Data Handling	Graphics Capability	Method Coverage
				Beginners	Experts			
R	High #1	Yes	Free #1	High	Mod	Yes	High #1	High
Python	Low	Yes	Free #1	High	Mod	Yes	High	High
SAS	Mod	No	~$7000/y	Mod	Low	Yes #1	High	High
STATA	Mod	No	~$400/y	Mod	Low	Yes	Mod to High	Mod to High
SPSS	Mod to High	No	~$1000/y	Low to Mod #1	Low #1	Yes	Low	Mod

* Prices may significantly vary per country and institution. Presented prices are for the first subscription of a single-user in the U.S.