

Chapter 40

Introduction to the Popular Open Source Statistical Software (OSSS)

Zhijian Wu

New York University, USA

Zichen Zhao

Yale University, USA

Gao Niu

Bryant University, USA

ABSTRACT

This chapter first introduces the two most popular Open Source Statistical Software (OSSS), R and Python, along with their Integrated Development Environment (IDE) and Graphical User Interface (GUI). Secondly, additional OSSS, such as JASP, PSPP, GRETL, SOFA Statistics, Octave, KNIME, and Scilab, will also be introduced in this chapter with function descriptions and modeling examples. The chapter intends to create a reference for readers to make proper selection of the Open Source Software when a statistical analysis task is in demand. The chapter describes software explicitly in words. In addition, working platform and selective numerical, descriptive, and analysis examples are provided for each software. Readers could have a direct and in-depth understanding of each software and its functional highlights.

INTRODUCTION

In this chapter, the authors discuss the most popular Open Source Statistical Software (OSSS) with its creation history, target practitioners, and statistical usage examples. Although Programming languages such as Java, C++ can also perform statistical analysis with intensive coding, the authors limit discussion to the software specifically designed for statistical analysis.

DOI: 10.4018/978-1-7998-9158-1.ch040

The objective of this chapter is to create a reference for the readers and guide them to make proper selection of Open Source Software (OSS) when a statistical analysis task is in demand. The discussion includes the background information, research areas that the software designed for, and the overview of how to use the software. R and Python are the two most important and popular software, their applications are discussed in detail throughout chapter four to seven in this book. The authors focus on creating an overview of all open source statistical software in this chapter.

BACKGROUND

Open Source Software (OSS) is a type of computer software that had its code released to the public. St. Laurent (2008) indicated that users have the right to study, change and redistribute the software under the copyright granted by the software license holder. Closed source or proprietary software can only be modified and maintained by the people, teams and organizations who own the software. Microsoft Office and Adobe Photoshop are well-known proprietary software.

Open Source Software is popular to statistical analysis practitioners, not only because it is free, but also because it is more adaptive to the current rapidly developing academic research advancement environment.

This chapter first introduces the two most popular Open Source Statistical Software (OSSS) R and Python along with its Integrated Development Environment (IDE) and Graphical User Interface (GUI). Then, additional OSSS, like JASP, PSPP, GRETL, SOFA Statistics, Octave, KNIME and Scilab, are introduced with description of their functions and modeling examples. Figure 1 lists all of the popular open source statistical software and IDEs that are introduced in this Chapter.

Figure 2 and 3 demonstrate the popularity development within last five years of the Open Source Statistical Software discussed in this chapter. The value represents the Google search interest. A value of 100 is the peak popularity which happens on the third week of 2019 for Python, a value of 50 represents the software is half as popular. The data is extracted on 12/19/2019 from trends.google.com under the category of “Science” and “Web Search”. Since R and Python dominate the popularity charts, two figures are created in order to better presents the relationship between all of the software. Figure 2 demonstrates R and Python popularity. Figure 3 shows other Open Source Statistical Software (OSSS).

R

R is arguably the most popular Open Source Statistical Software. It has a strong statistical analysis capability and graphical visualization functionality. This section provides an overview of the software R by introducing most used packages, its popular IDEs and its functionalities. Following two chapters will introduce R with syntax examples and application. R was initially written by Ross Ihaka and Robert Gentleman from the University of Auckland, New Zealand (Contributors, 2019). R is licensed under GNU General Public License. As of June 2019, the most current version is 3.6.0, it was released on April, 26th, 2019. The official website of the software is <https://www.r-project.org/>.

As of July 18th, 2019, there are 14,580 CRAN packages (Contributed Packages, 2019). The authors select a few popular packages, categorize them by functionality, and list in table 1 based on Awesome R (2019).

33 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/introduction-to-the-popular-open-source-statistical-software-oss/286604

Related Content

The System for Population Kinetics: Open Source Software for Population Analysis

Paolo Vicini (2009). *International Journal of Open Source Software and Processes* (pp. 29-43).

www.irma-international.org/article/system-population-kinetics/38904

Studies of Project Tasks

(2018). *Free and Open Source Software in Modern Data Science and Business Intelligence: Emerging Research and Opportunities* (pp. 81-92).

www.irma-international.org/chapter/studies-of-project-tasks/193458

On the State of Free and Open Source E-Learning 2.0 Software

Utku Kose (2014). *International Journal of Open Source Software and Processes* (pp. 55-75).

www.irma-international.org/article/on-the-state-of-free-and-open-source-e-learning-20-software/124004

Design Approaches

Barbara Russo, Marco Scotto, Alberto Sillitti and Giancarlo Succi (2010). *Agile Technologies in Open Source Development* (pp. 133-143).

www.irma-international.org/chapter/design-approaches/36501

Transmission Line Routing Using Open Source Software Q-GIS

Shabbir Uddin, Karma Sonam Sherpa, Sandeep Chakravorty and Amitava Ray (2017). *International Journal of Open Source Software and Processes* (pp. 71-82).

www.irma-international.org/article/transmission-line-routing-using-open-source-software-q-gis/203649