

Statistical Software Strategy

Dr Daisy Salifu

Biostatistician, Data Management, Modeling and Geo-Information Unit, *icipe*

Orientation Training for RSIF PhD Scholars (2022)

The **PASET** Regional Scholarship
and Innovation Fund

An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



Regional Coordination Unit



1

Statistics software champions...

- SAS
- SPSS
- JMP
- Stata
- SPSS
- GenStat
- R/RStudio

The **PASET** Regional Scholarship
and Innovation Fund

An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



Regional Coordination Unit



2

About SAS

- SAS is a statistical software suite developed by SAS Inst. for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation, and predictive analytics.
- SAS was developed at North Carolina State Univ. from 1966 until 1976 when SAS Inst took over
- Operating system - Windows, IBM mainframe, Unix/Linux, OpenVMS Alpha
- Most widely used statistical software packages in both industry and academia.
- Has lots of advanced statistics tools
- Highly customizable analysis options and output options.
- Publication-quality graphics with ODS (output delivery system).
- Large, active online community.
- *Access – Proprietary software (non-free). Pay for license*

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



3

About SPSS

- SPSS is a software program developed by IBM for data management, advanced analytics, multivariate analysis, business intelligence, criminal investigation
- Long produced by SPSS Inc., it was acquired by IBM in 2009.
- Current versions have the brand name: IBM SPSS Statistics
- Commands can be executed using the menu system or using command syntax - makes it easy to learn
- It is available for both Windows and Macintosh operating systems
- Syntax adds flexibility, customization, and automation options.
- Widely used in the social sciences.
- SPSS has also gone by the name PASW Statistics - "Predictive Analytics Software"

Access – trial version 7-30days then pay for license. License must be renewed yearly.

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



4

About JMP

- JMP is a software used for statistical analysis created by SAS Institute Inc.
- Unlike SAS (which is command-driven), JMP has a graphical user interface, and is compatible with Windows and Macintosh operating systems.
- The following techniques are only in JMP Pro and not JMP
 - Mixed models
 - Generalized linear models
 - Classification and regression trees
 - Neural network modeling
 - Text analysis
- *JMP & JMP Pro licenses must be renewed yearly*

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



5

About Stata

- Stata is a general-purpose statistical software developed by StataCorp for data manipulation, visualization, statistics, and automated reporting.
- It is used by researchers in many fields-biomedicine, epidemiology, sociology and science.
- Stata is a command- and menu-driven software package - available for Windows, Mac, and Linux operating systems.
- *Access – Proprietary software (non-free). Pay for license*

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



6

About GenStat

- Genstat is a statistical software package with data analysis capabilities, particularly in the field of agriculture.
- Was initially developed by Rothamsted Research (UK) in 1968, and now developed by VSN international, UK
- GenStat is designed to provide modular design, simple and advanced statistics and graphic functions.
- Has comprehensive menus to guide non-technical users and powerful programming language
- *Access – Proprietary software (non-free). Pay for license – Trial version*

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



7

About R and RStudio

- R is a language and environment for statistical computing and graphics
- R is free *software* developed by Ross Ihaka and Robert Gentleman in 1993
- R is supported by the R Core Team and the R Foundation for Statistical Computing
- R is used among data miners and statisticians for data analysis and developing statistical software
- Current version is R version 4.2.0 / 22 April 2022

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



8

R and Rstudio....

- RStudio is a free, open-source Integrated Development Environment (IDE) for R.
- Initial release: 28 February 2011; 11 years ago!!!
- You must install R before you can install RStudio.
- Its interface is organized so that the user can clearly view graphs, data tables, R commands, and output all at the same time.
- It also offers an Import-Wizard-like feature that allows users to import CSV, Excel, SAS (*.sas7bdat), SPSS (*.sav), and Stata (*.dta) files into R without having to write the commands.

Software strategy: R and Rstudio....

- Free and open source.
- Available for Windows, Macintosh, and Linux.
- Publication-quality graphs.
- Rivals (and in many cases, exceeds) SAS, Stata, SPSS, GenStat in terms of availability of advanced statistical methods and algorithms, through availability of user-created packages.
- Packages for *literate statistical programming* - weaving written reports and analysis code in one document.

How can I access R?

- R and RStudio are free and open-source.
- To install them on any computer, download the software from their respective websites.

<https://www.r-project.org/> for R

<https://www.rstudio.com/> for RStudio

The **PASET** Regional Scholarship and Innovation Fund



An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



11

R/RStudio

The **PASET** Regional Scholarship and Innovation Fund

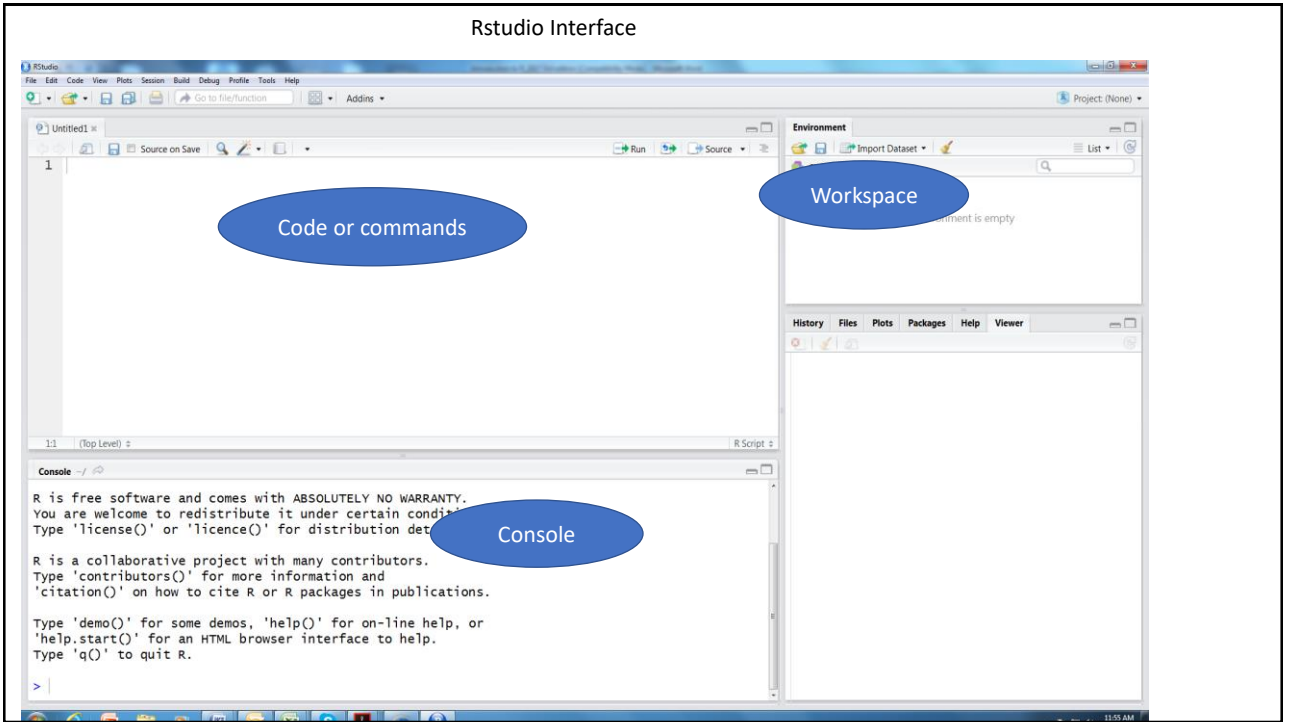


An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



12



13

Questions and Answers

The **PASET** Regional Scholarship and Innovation Fund

An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology



Regional Coordination Unit



14

Thank you



The **PASET** Regional Scholarship
and Innovation Fund



PASET
Partnership for Skills
in Applied Sciences,
Engineering & Technology

An Africa-led initiative to bridge the skills gap in Applied Sciences, Engineering, & Technology

Regional Coordination Unit



For more information contact

Regional Coordination Unit

Regional Scholarship and Innovation Fund

International Centre of Physiology and
Ecology (*icipe*)

P.O. Box 30772-00100, Nairobi, Kenya

Tel +254 (20) 8632000

 icipe@icipe.org rsif@icipe.org

 facebook.com/TheRSIF

 [@pasetsif](https://twitter.com/pasetsif)

 linkedin.com/in/PASET-RSIF/