

AstroStat 1.0 Beta

Santosh Jagade
Virtual Observatory, India



Introducing AstroStat

- What is AstroStat?
 - VO-I package for statistical analysis of astronomical data.
- Why use AstroStat?
 - Uses **R**, a proven, open-source statistical computing environment, to implement the statistical tests.
 - Currently supports more than 30 statistical tests.
 - Also includes an up-to-date plotting system (*ggplot2*) with strong community support.
 - Provides interconnectivity with other VO tools using SAMP or PLASTIC.
 - Equipped to handle data in FITS, VOTABLE, and ASCII formats.



Essential features of AstroStat



Home window

- Divided into 4 distinct panels.
- Tests can be executed by selecting a test and loading a dataset, or vice versa.

The screenshot shows the AstroStat application window. The title bar reads "AstroStat" and the menu bar includes "File Interop". Below the menu bar are icons for "View File", "View Data", "VOPlot", and "Help".

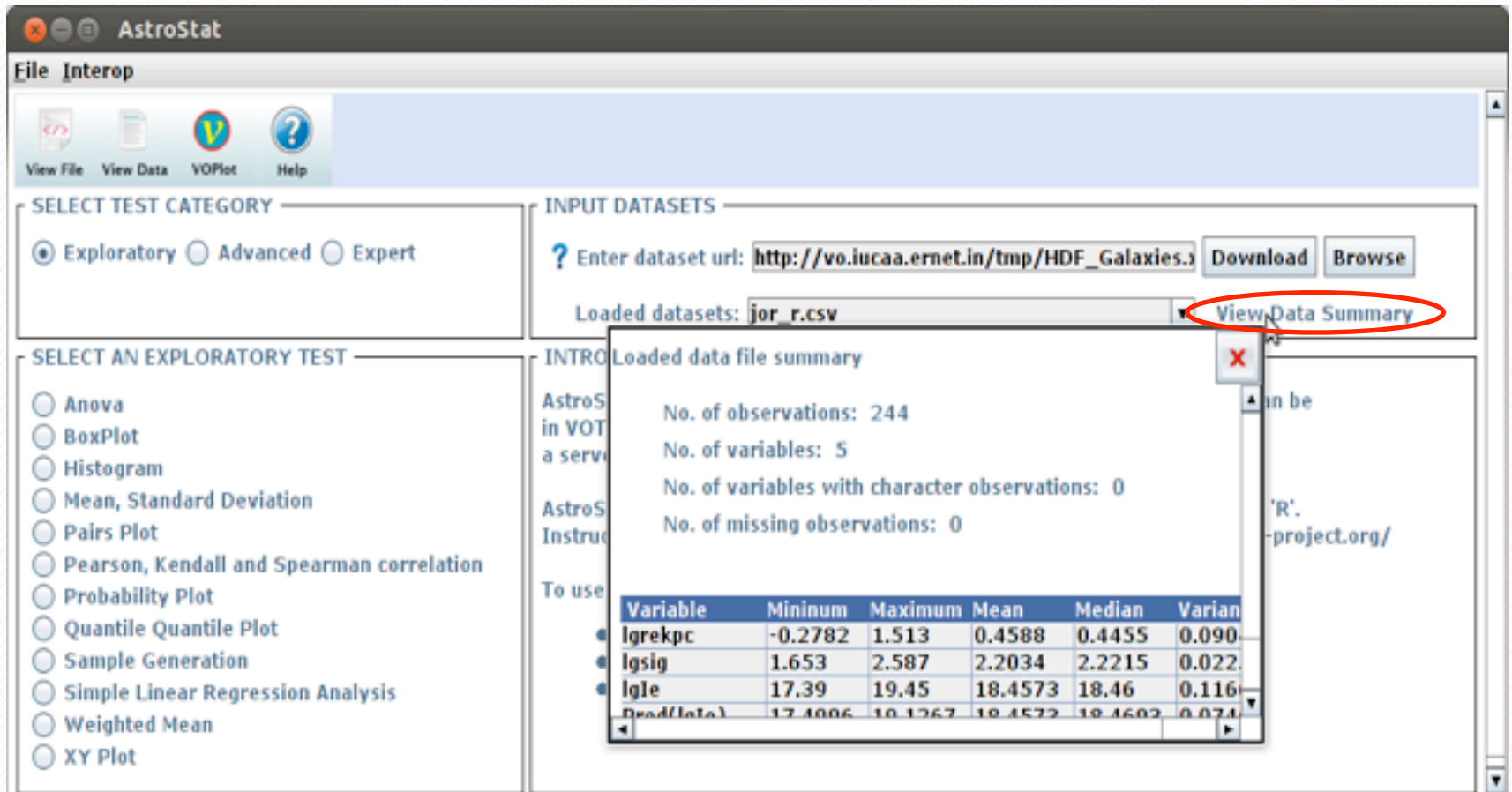
The main window is divided into four panels:

- SELECT TEST CATEGORY:** Contains radio buttons for "Exploratory" (selected), "Advanced", and "Expert".
- INPUT DATASETS:** Features a text input field with the URL "http://vo.iucaa.ernet.in/tmp/HDF_Galaxies.", "Download", and "Browse" buttons. Below it is a "Loaded datasets:" dropdown menu and a "View Data Summary" button.
- SELECT AN EXPLORATORY TEST:** Lists various statistical tests with radio buttons: Anova, BoxPlot, Histogram, Mean, Standard Deviation, Pairs Plot, Pearson, Kendall and Spearman correlation, Probability Plot, Quantile Quantile Plot, Sample Generation, Simple Linear Regression Analysis, Weighted Mean, and XY Plot.
- INTRODUCTION:** Contains text explaining that AstroStat provides statistical routines for VOTable, FITS, or ASCII data. It also mentions that AstroStat uses the R environment and provides a link to "http://cran.r-project.org/" for instructions. A section titled "To use AstroStat:" lists three steps: "Select test and data", "Provide required information", and "Run Test".



Data Summary

- Summary statistics of the active dataset are always available for a quick review.



The screenshot shows the AstroStat software interface. The main window has a menu bar with 'File' and 'Interop'. Below the menu bar are icons for 'View File', 'View Data', 'VOPlot', and 'Help'. The interface is divided into several panels:

- SELECT TEST CATEGORY:** Radio buttons for 'Exploratory' (selected), 'Advanced', and 'Expert'.
- INPUT DATASETS:** A text field for 'Enter dataset url:' containing 'http://vo.iucaa.ernet.in/tmp/HDF_Galaxies.', with 'Download' and 'Browse' buttons. Below it, 'Loaded datasets:' shows 'jor_r.csv' and a 'View Data Summary' button circled in red.
- SELECT AN EXPLORATORY TEST:** Radio buttons for various tests: Anova, BoxPlot, Histogram, Mean, Standard Deviation, Pairs Plot, Pearson, Kendall and Spearman correlation, Probability Plot, Quantile Quantile Plot, Sample Generation, Simple Linear Regression Analysis, Weighted Mean, and XY Plot.
- INTRO:** A section for 'Loaded data file summary' with the following statistics:
 - No. of observations: 244
 - No. of variables: 5
 - No. of variables with character observations: 0
 - No. of missing observations: 0
- Table:** A table showing summary statistics for variables. The table has columns: Variable, Minimum, Maximum, Mean, Median, and Variance.

Variable	Minimum	Maximum	Mean	Median	Variance
lgrekpc	-0.2782	1.513	0.4588	0.4455	0.090
lgsig	1.653	2.587	2.2034	2.2215	0.022
lg1e	17.39	19.45	18.4573	18.46	0.116
logd(log1e)	17.4006	18.1267	18.4572	18.4602	0.074

Input Panel

- Flexibility offered to the user in terms of variable transformations, choice of parameters, and format of output plots.

The screenshot shows the AstroStat software interface. The title bar indicates the file path: `/home/tejaskale/Ubuntu One/IUCAA/Datasets/sample_data/data/plane_data_19sep09/jor_r.csv`. The interface is divided into several sections:

- File Interop:** Contains icons for View File, View Data, VOPlot, and Help.
- SELECT TEST CATEGORY:** Radio buttons for Exploratory (selected), Advanced, and Expert.
- INPUT DATASETS:** A text field for the dataset URL (`http://vo.iucaa.ernet.in/tmp/HDF_Galaxies.1`), Download and Browse buttons, and a dropdown menu for loaded datasets (jor_r.csv).
- SELECT AN EXPLORATORY TEST:** Radio buttons for various tests, with Histogram selected.
- INFORMATION ON 'HISTOGRAM':** A text area explaining histograms and their construction.
- HISTOGRAM FOR JOR_R.CSV:** A section for configuring the histogram, including a table for columnwise transformations and a 'Run Test' button.

The 'Columnswise Transformations' table is highlighted with a red circle. It has the following structure:

Columns	x-val	log10(x)	loge(x)	exp(x)
Igrekpc	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Igsig	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Igle	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Pred(IgIe)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Delta(Ie)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Other parameters in the histogram configuration section are also highlighted with a red circle:

- Binwidth:
- On Y-axis: Frequency Probability
- Plots Format: PostScript JPEG PDF PNG
- Plots Per Page: 1 2 4 6
- Run Test button



Output

- Results computed by R are formatted to provide the user with the most relevant information.
- The following example shows the output of Pearson's Correlation analysis for Effective Radii and Mean Surface Brightness of galaxies. This output can be saved in either ASCII or PDF format.

PEARSON, KENDALL AND SPEARMAN CORRELATION Output for JOR_R.CSV

Pearson's Product Moment Correlation

var1	var2	size	f	r	t	P(>t)	CI-95%	data_file
lgrekpc	lg1e	244	242	-0.7997	-20.722	0	(-0.8409,-0.7494)	jor_r.csv

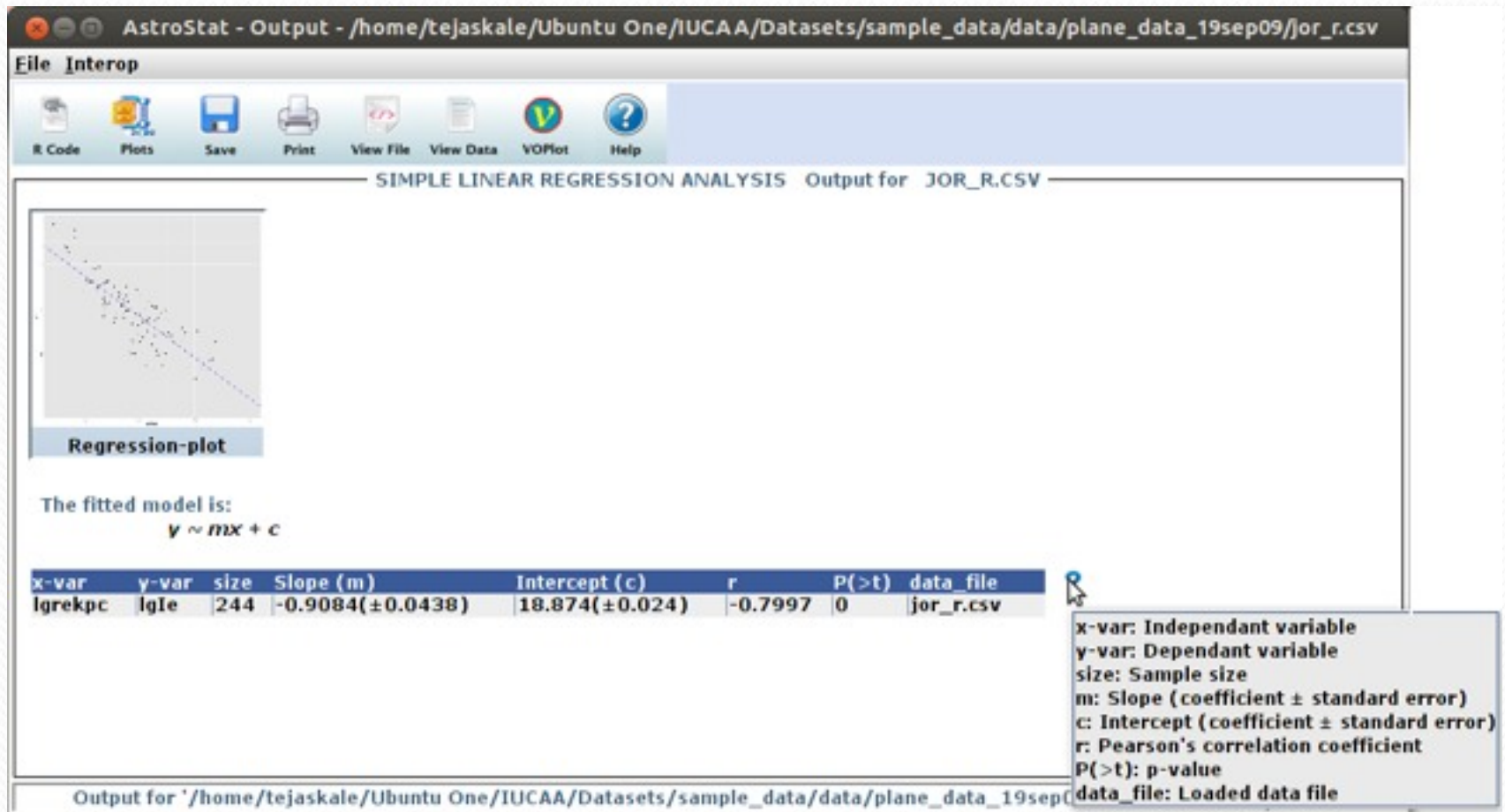
var1: Variable 1
var2: Variable 2
size: Sample Size
f: Degrees of Freedom
r: Pearson's correlation coefficient
t: t-Statistic
P(>t): p-value
C.I.: Confidence Interval
data_file: Loaded Data File

Output for '/home/tejaskale/Ubuntu One/IUCAA/Datasets/sample_data/data/plane_data_19sep09/jor_r.csv'. 244 rows



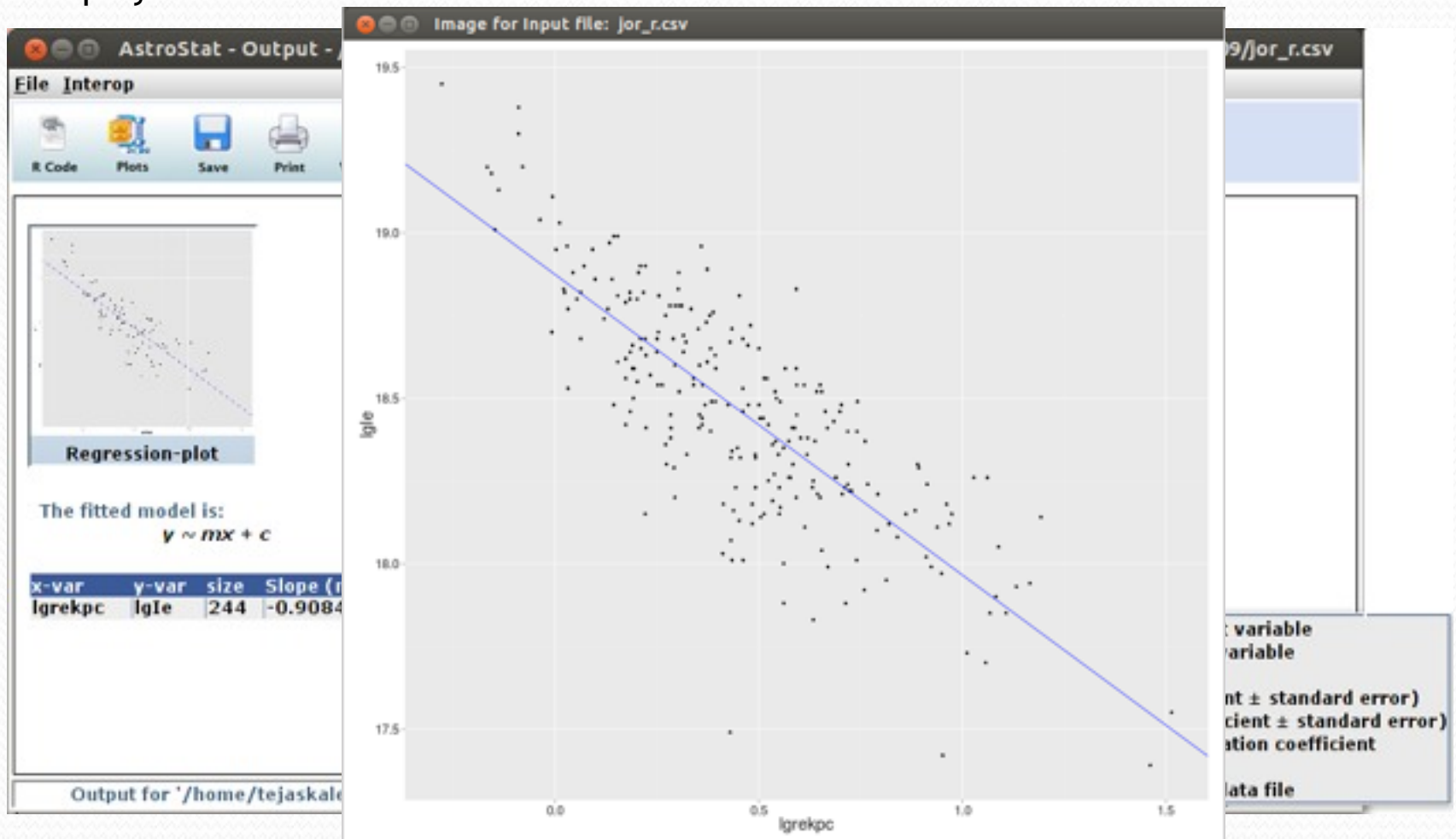
Output

- Output of Simple Linear Regression of Mean Surface Brightness on Effective Radii is displayed as



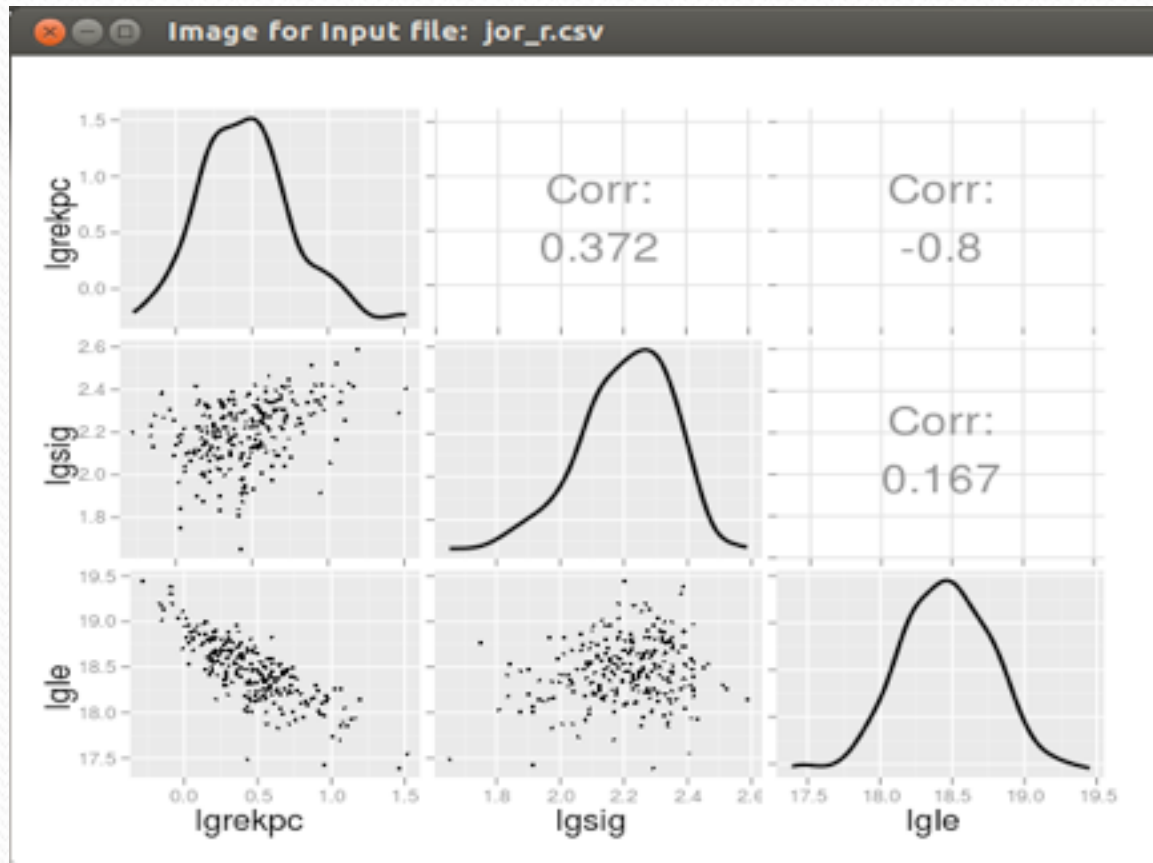
Output

- Output of Simple Linear Regression of Mean Surface Brightness on Effective Radii is displayed as



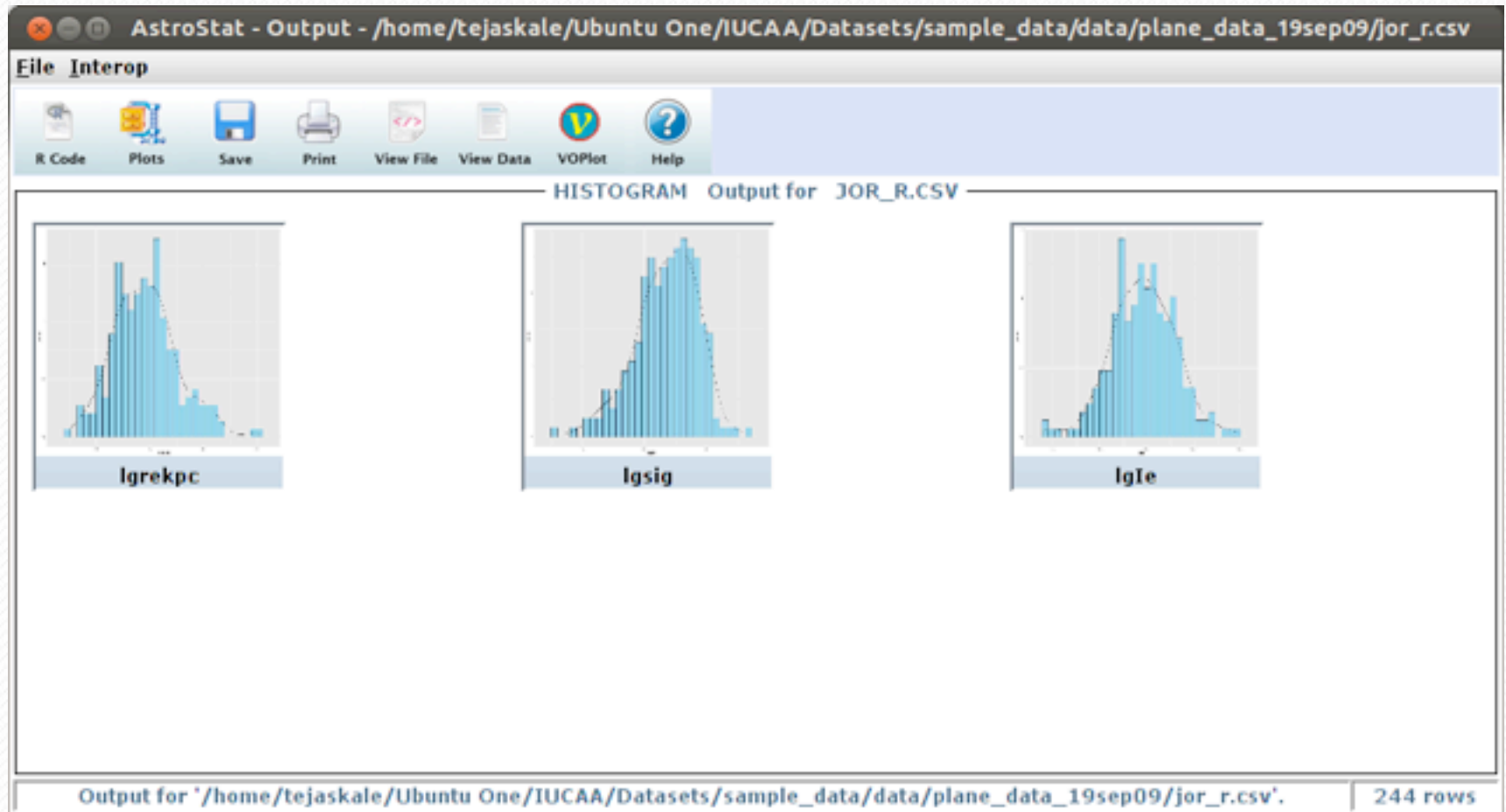
Plots

- New plotting system produces publication quality visualizations that can be saved in 4 different formats (*PS, JPEG, PDF, PNG*).



Plots

- Thumbnails allow a quick glance at all the generated plots.

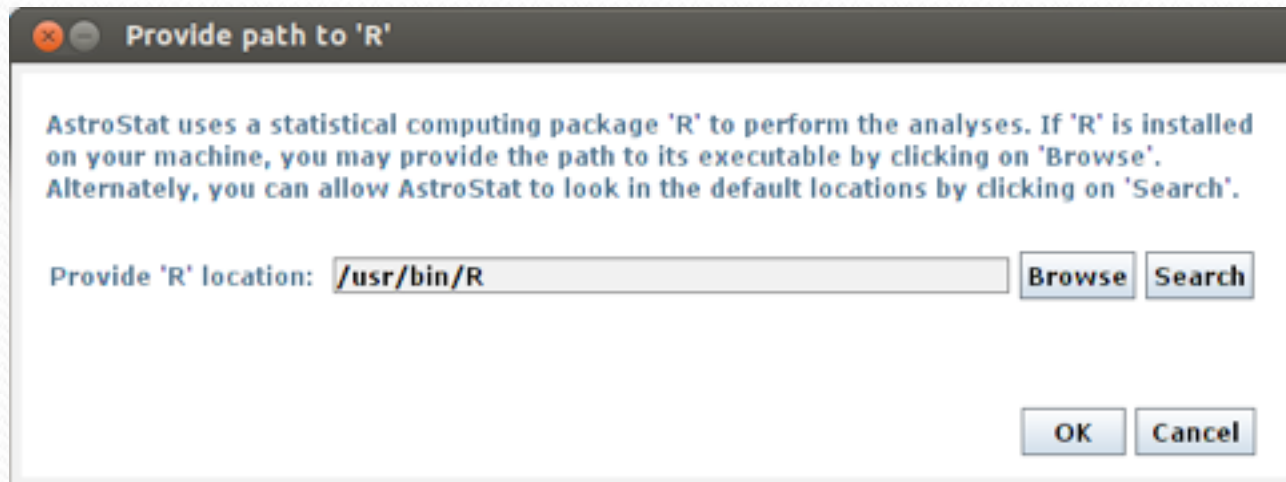


Other features of AstroStat



Path to R

- AstroStat requires the location of R to perform statistical computations.
- This path can now be automatically found using the tool's *Search* functionality.



Installation of R

- AstroStat uses numerous R libraries which have to be downloaded and installed by the user.
 - These libraries help produce
 - better visualizations
 - implement new tests
 - provide an interactive interface
- The download and installation of these libraries is now automated in AstroStat, if the user permits.



File Interop

SELECT AN EXPLORATORY TEST

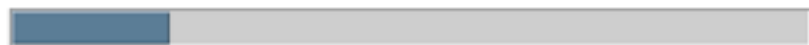
- Anova
- BoxPlot
- Histogram
- Mean, Standard Deviation
- Pairs Plot
- Pearson, Kendall and Spearman correlation
- Probability Plot
- Quantile Quantile Plot
- Sample Generation
- Simple Linear Regression Analysis
- Weighted Mean
- XY Plot

INFORMATION ON 'SIMPLE LINEAR REGRESSION ANALYSIS'

Simple Linear Regression Analysis is an extension of the correlation analysis between two variables. Considering one of the variables to be the dependent variable and the other to be independent, the aim of this test is to fit a straight line model between the variables that can be used for reasonable predictions of the dependent variable based on the interpolated values of the dependent variable.

Executing test

Loading package 'assertive' (20% complete)



Package missing

Would you like to install the package: 'assertive'?

OK Cancel

Columnwise Transformations:

Select Column	y-val	log10(y)		
lgrekpc	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Columnwise Transformations:

Select Column	x-val	log10(x)	loge(x)	exp(x)
lgsig	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Plots Format: PostScript JPEG PDF PNG

Run Test



Work in progress

- End-user documentation for summarizing various statistical tests and their usage.
 - Examples of analysis on astronomical datasets will also be provided.
- Saving textual output in ASCII or PDF format.
 - Currently, AstroStat can save output of individual tests.
 - Ability to append output of multiple tests will be added soon.
- Implementation of additional tests commonly used in the astronomical domain.



Thank You!

