

Statistics Anyone?



by Jürgen Pohl
<sept.sapins(at)verizon.net>

About the author:

Jürgen Pohl works as an R&D Engineer and technical translator on the Pacific Coast of the US.



Abstract:

Most statistics packages seem to be overwhelming in their scope, forcing the user on a very steep learning path, most of them are also very expensive. However, there are a few convenient alternatives for those who need them most: the beginner, who is being initiated to the secrets of the magic world of statistics, as well as the user for whom the elaborate commercial packages would be overkill. SalStat is one of those sought-after alternatives - thanks to its creator it is open source, meaning: free! Another advantage: the program is platform independent. Last but not least, it is very easy to use. The program is written in Python, however it can be used without any prior knowledge of that language.

SalStat - the Statistics Program

Introduction

Building on his own experience, the creator of SalStat was very well aware of the predicament in which many of those being introduced to statistics find themselves in: in order to expand their newly acquired knowledge (or just to do their homework or project...) an affordable, easy to use statistics program on their own machine is essential. With this in mind Alan James Salmoni developed SalStat and published it under the GNU license. It can be found at its homepage.

What can SalStat do?

Unfortunately this article cannot provide an introduction to statistics. In 'Resources' you will find some information on this topic. Here are the lists of the statistics and tests SalStat is able to generate:

Parametric and non-parametric tests are combined here:

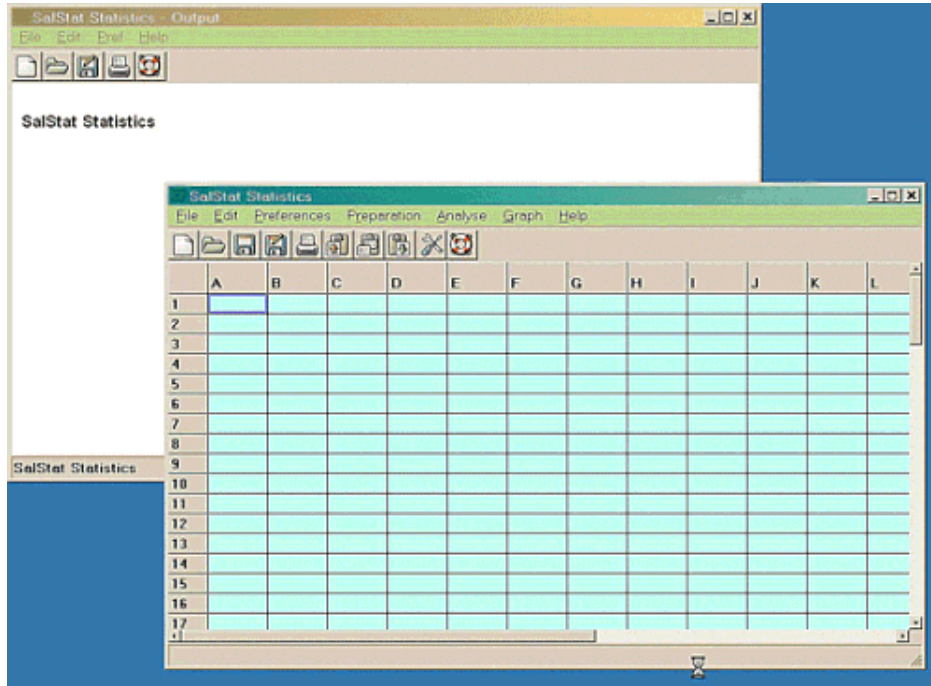
N (count)	range
sum	number of missing cells
mean	geometric mean
variance	harmonic mean
standard deviation	skewness
standard error	kurtosis
sum of squares	median
sum of squared deviations	median absolute deviation
coefficient of variation	mode
minimum	interquartile range
maximum	number of unique levels of data

Inference Statistics

t test (paired)	Pearsons correlation
t test (unpaired)	Spearman's rho correlation
1 sample sign test	Kendall's tau correlation
2 sample sign test	Point biserial r correlation
F test for variance ratio	linear regression
Wilcoxon Ranked Sums Test	Single factor analysis of variance (between subjects)
Mann-Whitney U Test	Single factor analysis of variance (within subjects)
Kolmogorov-Smirnov test	Kruskal-Wallis H test
Paired Permutation test	Friedman test

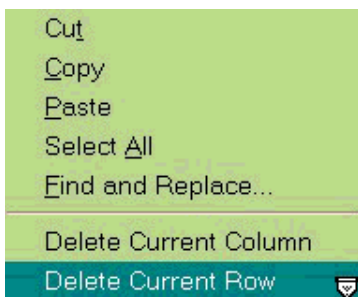
How Do We Work With SalStat?

Before you can take a look at SalStat you need to install it on your machine, but we will get to that later. First I would like to give you an idea of what you are getting. When you open the program these two windows appear on your screen:

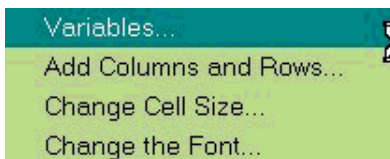


The first window (titled 'SalStat Statistics') in the foreground of the screenshot above shows a data entry grid like most spreadsheets: here we enter the data of our samples to be analyzed. At the top of the window the usual toolbar - clicking on one of the tools brings up a dropdown list of functions to select from.

- **File:** The usual.
- **Edit:** The usual too, but you can add columns or rows to your table:



- **Preferences:** we are getting a bit more specific in regard to statistics:



Here you find the tools to easily customize the cells of the table to match your input. Also important, you can name the specific variables of your test or for your statics. By clicking on 'Variables' the following window will open and you can type in your variable names for each

column:

	A	B	C	D	E	F	G
Variable Name	A	B	C	D	E	F	G
Decimal Places							
Missing Value							

Okay Cancel

- **Preparation:** we are invited to select which statistics we want - there is a long list to choose from, just mark the boxes of the statistics you wish or select all of them. Don't forget to mark the columns with the data you want to analyze (you cannot analyze data which don't exist...):

Descriptive Statistics

Select Column(s) to Analyse

- A
- B
- C

Select Descriptive Statistics

- Sum
- Mean
- Variance
- Standard Deviation
- Standard Error
- Sum of Squares
- Sum of Squared Devs
- Coefficient of Variation
- Minimum
- Maximum
- Range
- Number Missing
- Geometric Mean
- Harmonic Mean
- Skewness
- Kurtosis
- Median

Select All Select None

Okay Cancel

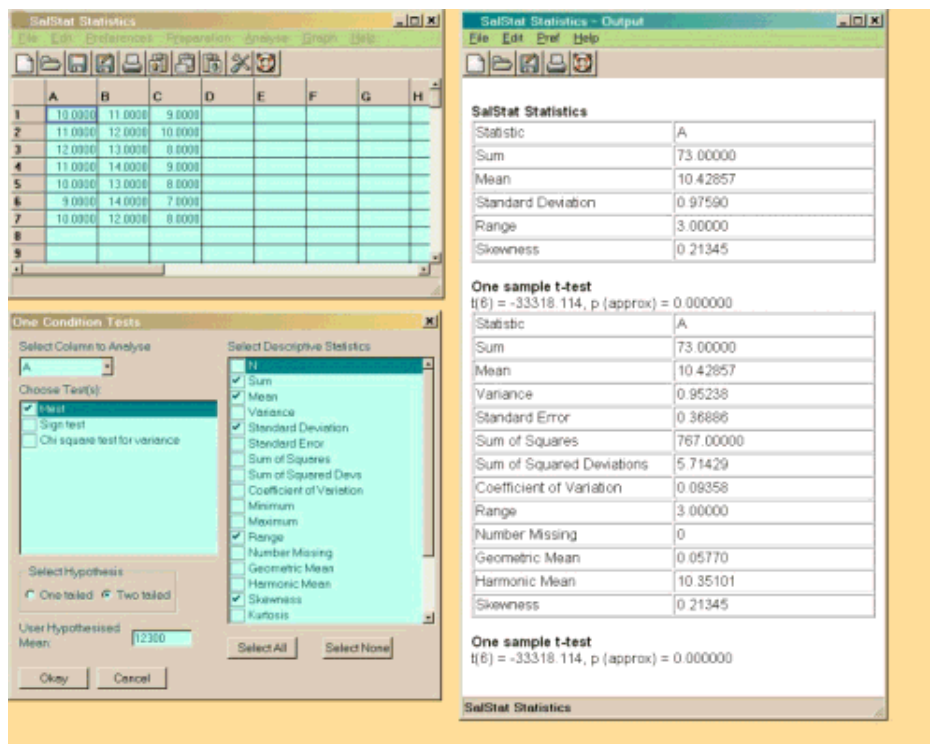
- **Analyze:** 'Analyze' is the next set of tools on our toolbar, where you choose the kind of test you want to run on your data
- **Graph:** is at present not yet functional.
- **Help:** here you find general and specific information to get you going.

The second window (titled 'SalStat Statistics - Output'), which is initially empty, will show the statistics results of the executed test.

Nothing is more frustrating than having a program like this installed without any data to play with - the author was so wise to include a test file (testreport1.txt) with known results. Simply enter the test data into your table and choose what kind of statistics or tests you would like to see and then hit 'Okay' and the result will appear in the output window. The windows below happened in following sequence :

1. We entered the data in the columns A, B, C of the data grid of our open SalStat Statistics window. The data are from the sample file testreport1.txt.
2. Next we went to the 'Analyse' tool on the tool bar. Here we chose the 'One Condition Test'
3. The 'One condition Test' window opened. We marked the specifics of our test:
 - 'Select Column to Analyse': we picked 'A'.
 - 'Choose Test(s)': we selected 't-test'.
 - 'Select Descriptive Statistics': we chose 'Sum', 'Mean', 'Standard Deviation', 'Range' and 'Skewness'.
 - 'Select Hypothesis': we kept the default 'Two tailed'
 - 'User Hypothesised Mean': we just guessed here for this demonstration and we entered the value of 12300
 - Click the 'Okay' button

The previously blank 'SalStat Statistics - Output' window (our screen shot below, left) will show the results of the test.



Getting and Installing SalStat

The program is waiting to be downloaded from its website. A number of alternatives are offered to accommodate your operating system(s). I have the program on two desktop machines with different operating systems in two locations. The source code is available for downloading as well - maybe you want to show off your (Python) programming skills...? Before trying any installation, please read the '*Basic Users Guide*', also available on the SalStat homepage under '*Documentation*'. The guide gives clear instructions on how to install the program, we need not to repeat them here - please have a look at the website.

Customizing SalStat

Another very useful part of SalStat is its built-in ability for users to write their own scripts - to automate tasks, build their own tests, etc. In the program's manual, which is the main part of the '*Help*' tool, you can find a detailed description on how to do that ('*Scripting and Making Your Own Tests*'). Have a look, it is a very helpful introduction to scripting. It should encourage the user to utilize the scripting feature. The last tool of the '*Analyse*'-dropdown list gives you access to the '*Scripting Window*' - here you can enter your scripts: try the samples given in the manual, they could convince you to use this feature.

One minor hang-up for some people: the program is written in Python. In order to do serious program customization you would need to learn the language. Fortunately Python is an open language, meaning free: you can download it with extensive instructions (addressing everyone from beginner to expert) from the Python webpage. If you do not want to deal with Python you can use SalStat as is - but with some Python knowledge you may enhance your benefits from this program.

Conclusion

SalStat was written with ease of use in mind. The user can click his or her way through a wish list of statistics and tests. The manual gives instructions for all the tests, including some hints on the value of their results. In general, however, it is assumed that you have at least basic knowledge of statistics or are in the process of acquiring it.

One word of caution: before you jump in and bank on your career as an up-and-coming scientist by using results generated with this program, listen to the recommendations of its creator and convince yourself of its merits by testing it! Those who are just starting with statistics will find many examples in text books - plug some of the available data into SalStat and see what you are getting. The test file (*testreport1.txt*), which comes with the download, gives you some result comparisons of tests run with other programs.

Resources

- Here we get a huge collection of statistics related links collected and published by Clay Helberg.
- An online Statsoft textbook and a glossary may be found here, Statsoft is known for its commercial statistics programs.
- The International Statistical Institute
- An excellent multilingual glossary is offered by the European Union
- *'Introduction to the Practice of Statistics'* by David S. Moore and George P. McCabe. Good book on the issue, seems to be widely used. The version I found in a public library had a CD with all the data for a large number of training exercises found in the book. The same sample data for the exercises and their results are also available for free on the authors website. I found used versions of the book available through online bookstores for very little money (\$ 5).

Webpages maintained by the LinuxFocus Editor team © Jürgen Pohl "some rights reserved" see linuxfocus.org/license/ http://www.LinuxFocus.org	Translation information: en --> -- : Jürgen Pohl <sept.sapins(at)verizon.net>
---	--