



Paquete estadístico

SalStat



Descripción y uso elemental

Índice

Introducción	2
Interfaz	3
Hoja de datos	3
Ventana de salidas	5
Estadística descriptiva	6
Transformación de datos	7
Comparaciones para una muestra	8
Comparaciones paramétricas	9
Comparaciones no paramétricas	9
Comparaciones entre dos muestras	9
Comparaciones paramétricas	10
Comparaciones no paramétricas	10
Correlaciones entre dos muestras	10
Correlaciones paramétricas	11
Correlaciones no paramétricas	11
Comparaciones entre tres muestras o más	11
Comparaciones paramétricas	12
Comparaciones no paramétricas	12
Archivos de código	12
Ayuda	13
Futuras mejoras de <i>SalStat</i>	14
Referencias	14

Introducción

SalStat es un paquete estadístico gratuito para el análisis de datos. Su autor es Alan James Salmoni. Como él mismo indica, el programa fue diseñado originalmente para ser utilizado en ciencias como Psicología, y no pretende ser tan completo como para ser una alternativa a paquetes consolidados como *R* y los comerciales *SAS* y *S-plus*. Está pensado para gente que no muy familiarizada con los ordenadores o que no tiene acceso a paquetes comerciales.

Entre los objetivos que motivaron el que el autor se decidiese a crear este paquete están el de crear un paquete fácil de utilizar (con entorno gráfico), incluso para alguien con pocos conocimientos de Estadística, que no necesitase muchos recursos de la máquina en que se ejecutase, y disponible en el mayor número posible de plataformas y sistemas operativos. Su uso resultará bastante sencillo para quien haya manejado algunos de los paquetes estadísticos que hay en el mercado.

Aparte de los análisis que el paquete tiene implementados, una de las ventajas de *SalStat* es que el usuario puede ejecutar en él su propio código, con la condición de que esté escrito en *Python*.

SalStat está escrito en el lenguaje *Python* ([1]), que es a la vez fácil de aprender y potente; algunas de las características del paquete están determinadas precisamente por este lenguaje de programación. El entorno gráfico (*GUI*, de *graphic user interface*) ha sido creado con la herramienta *wxPython* ([4]). *Python* es un lenguaje orientado a objetos moderno y fácil de utilizar, su sintaxis es clara, existen intérpretes para la mayoría de plataformas, y dispone de algunas buenas herramientas para la creación de entornos gráficos.

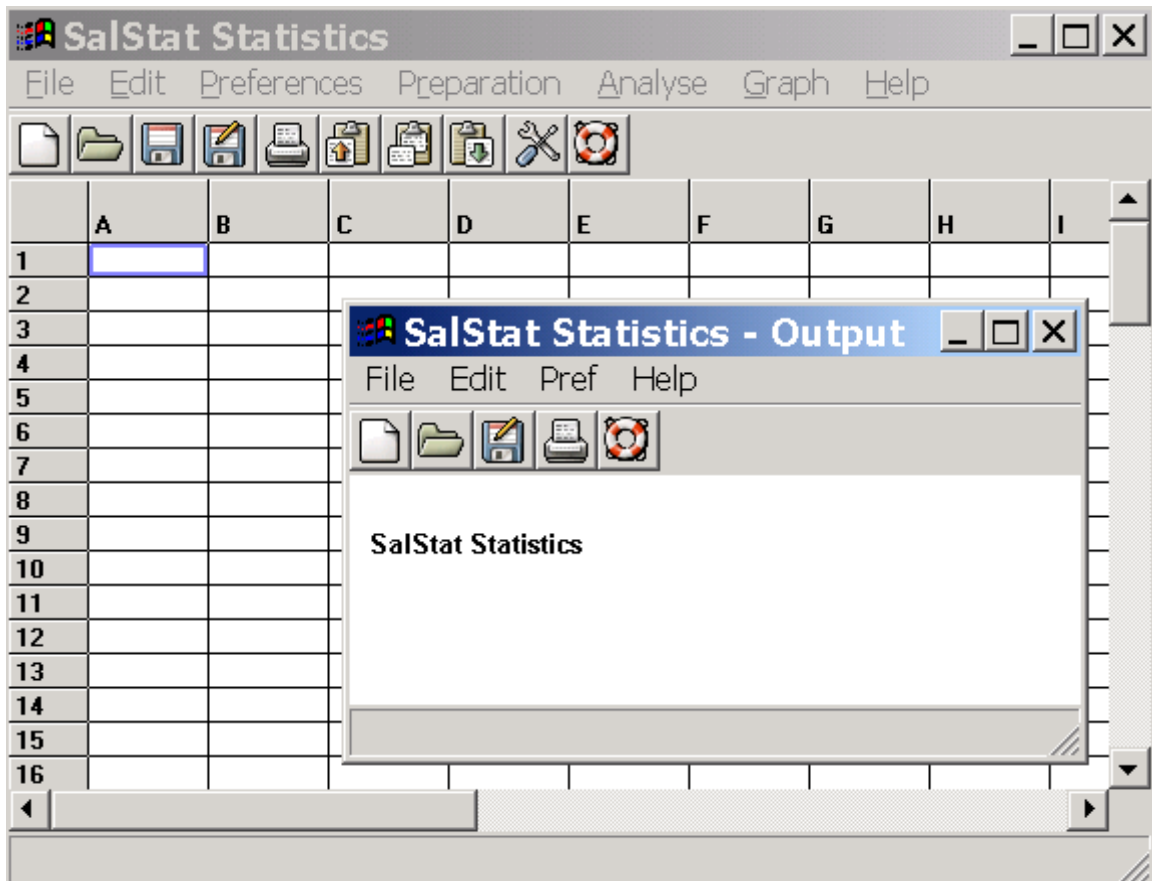
Una ventaja del paquete es que, por ser de código abierto, puedes modificar a tu antojo el programa para adaptarlo a tus necesidades, o puedes revisar el código para ver qué hace exactamente. Se puede acceder al código de las funciones estadísticas utilizando *Python*: basta poner una copia del módulo *salstat_stats.py* en el directorio de paquetes de *Python* y cargarlo (importarlo). Esto también significa por ejemplo que, frente a los programas que no son de código abierto, puedes asegurarte de que el código no envía información tuya o hace otras acciones que tú no quieres.

Por motivos prácticos el autor ha decidido distribuirlo bajo licencia pública: *GNU General Public License (GPL)*. Esto significa que cualquiera puede cambiarlo, venderlo, distribuirlo gratuitamente, etcétera, pero siempre poniendo a disposición el código fuente.

SalStat está actualmente disponible para las plataformas *Windows*, *Linux*, *MacIntosh*, *OSX* y *Unix*. Se puede obtener desde su sitio web [2]. Su código fuente está también disponible, para quien quiera compilarlo. Los usuarios que no utilicen *Windows* o *Linux*, tienen que instalar antes *Python* y *wxPython* para poder utilizarlo.

Interfaz

Al abrir el programa, aparecen dos ventanas diferentes. Una tiene la apariencia de una hoja de cálculo, será la que contendrá los datos; la otra, enteramente blanca, será la ventana en la que se mostrarán los resultados.

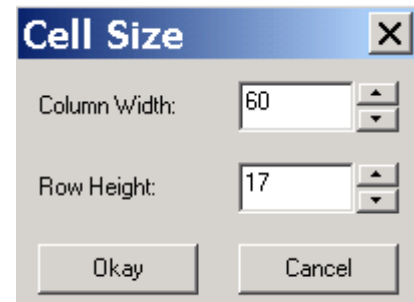


Hoja de datos

Mientras que la mayoría de los paquetes estadísticos suelen requerir que los datos se introduzcan de una forma concreta, *SalStat* está diseñado de forma intuitiva, y se diferencia en que **cada condición del modelo se introduce en una columna diferente**. Es decir, si una variable puede tomar valores en tres posibles niveles, se introducirán los datos utilizando tres columnas, una para cada nivel. Para **pasar de una celda a la que está debajo** hay que pulsar el botón *Intro*. Para **pasar a la celda de la derecha** hay que pulsar el tabulador (*Tab*). Para **volver a la celda de la izquierda** hay que pulsar el tabulador a la vez que se mantiene pulsado la tecla de las mayúsculas.

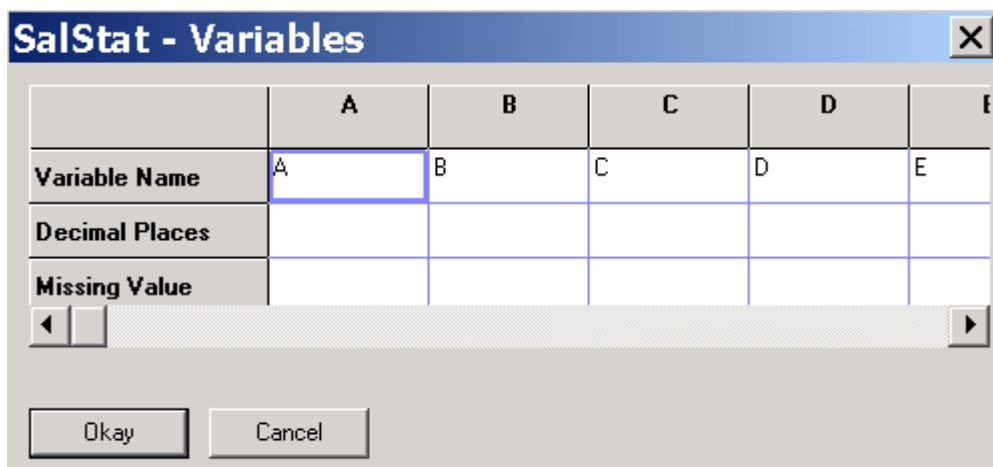
Se puede **personalizar la apariencia de la hoja de datos** en el menú:

Preferentes → *Change Cell Size...*



Si lo que se quiere es **cambiar el nombre de una columna**:

Preferentes → *Variables...*



Aunque el programa asigna el valor -99.999 a los datos omitidos, **el usuario puede elegir el valor que desee asignar a las celdas vacías**.

Dado que *SalStat* todavía no puede generar datos, se deben introducirse a mano o se pueden **cargar desde un fichero de texto plano ASCII** (que no contenga los nombres de las variables), que es el único formato desde el que actualmente *SalStat* puede importar datos:

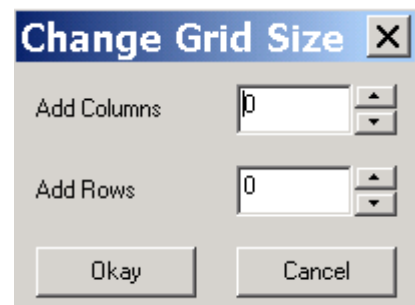
File → *Open...*

Para **guardar datos**:

File → *Save...*

Para **añadir alguna columna o fila** en la hoja de datos:

Preferences → *Add Columns and Rows...*



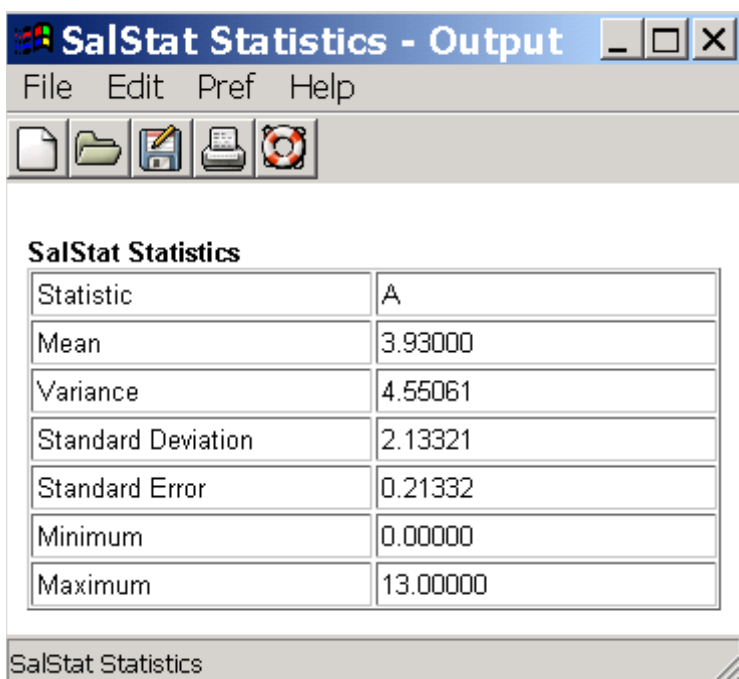
Para **eliminar alguna columna o fila** en la hoja de datos:

Edit → *Delete Current Column* y *Edit* → *Delete Current Row*

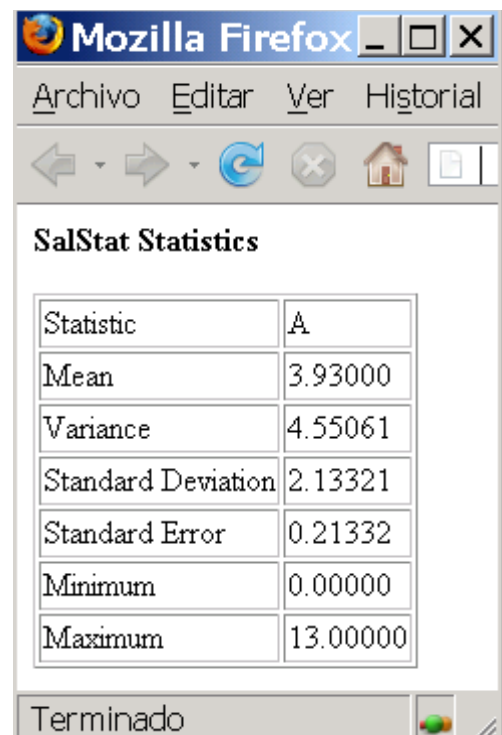
La columna o la fila añadida o eliminada será la que en ese momento contenga el cursor.

Ventana de salidas

El **formato en el que SalStat proporciona los resultados de sus análisis es el HTML**, es decir, el del código en que se escriben los documentos web. Esto los hace inmediatamente publicables por este protocolo, y hace que los resultados puedan ser abiertos con cualquier navegador de internet, pero dado que algunos procesadores de texto manejan el código HTML, pueden ser también vistos y editados con ellos.

A screenshot of the "SalStat Statistics - Output" window. It has a menu bar with "File", "Edit", "Pref", and "Help". Below the menu bar are icons for file operations. The main content area displays a table titled "SalStat Statistics" with the following data:

Statistic	A
Mean	3.93000
Variance	4.55061
Standard Deviation	2.13321
Standard Error	0.21332
Minimum	0.00000
Maximum	13.00000

The status bar at the bottom shows "SalStat Statistics".A screenshot of a Mozilla Firefox browser window displaying the HTML output from the SalStat window. The browser's address bar shows "SalStat Statistics". The content area displays the same table as in the SalStat window:

Statistic	A
Mean	3.93000
Variance	4.55061
Standard Deviation	2.13321
Standard Error	0.21332
Minimum	0.00000
Maximum	13.00000

The status bar at the bottom shows "Terminado".

Los resultados pueden también guardarse. Para **guardar los resultados**:

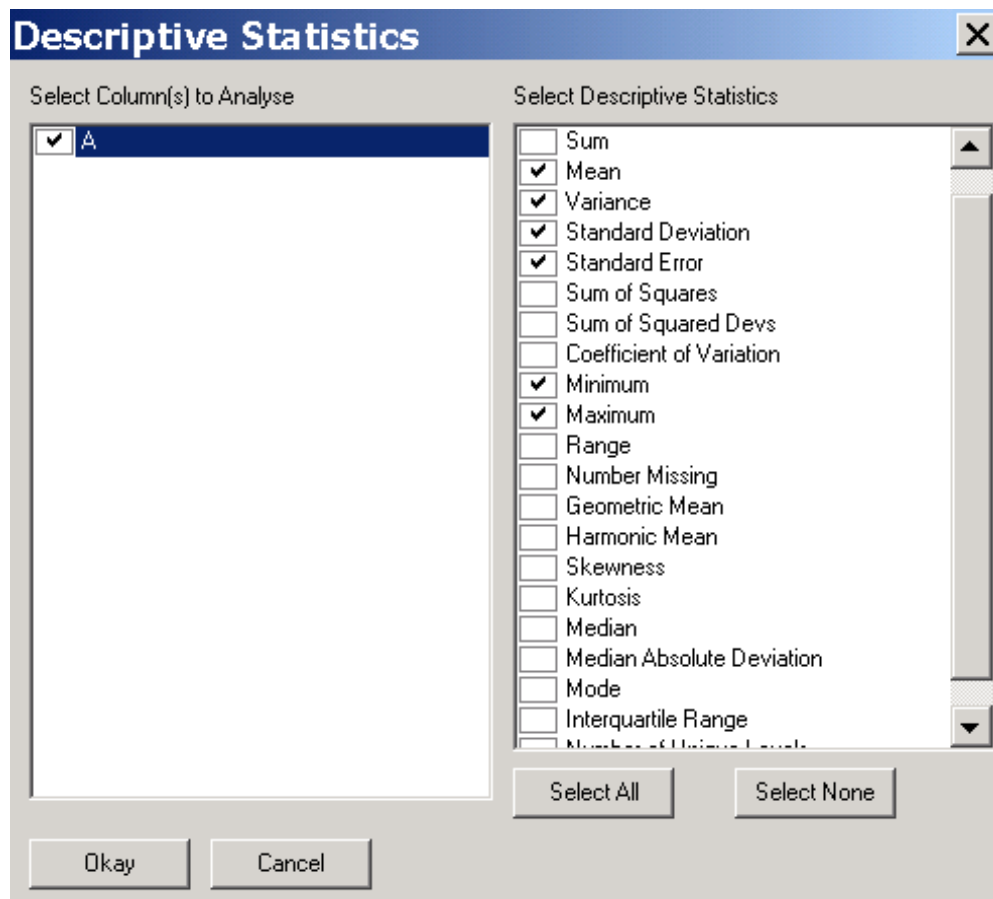
File → *Save as...*

Estadística descriptiva

Una vez que tenemos las variables en la hoja de entrada (una condición —o nivel— en cada columna), **para hacer un análisis descriptivo de los datos**:

Preparation → *Descriptive Statistics...*

Aparecerá un cuadro en el que elegiremos por un lado las columnas que queremos analizar y por otro los estadísticos descriptivos que queremos obtener.



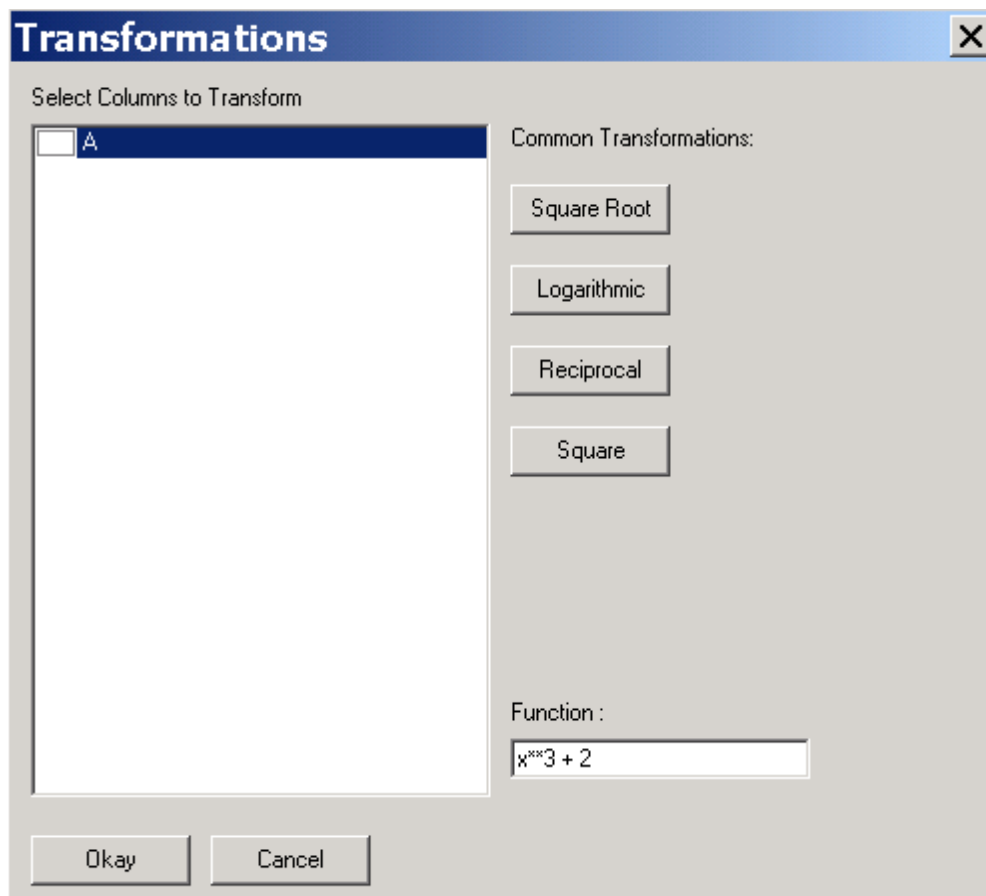
SalStat proporciona varias medidas de estadística descriptiva, algunas que son paramétricas y otras no: tamaño de la muestra, N , (*count*), suma (*sum*), media (*mean*), varianza (*variance*), desviación estándar (*standard deviation*), error estándar (*standard error*), suma de los cuadrados (*sum of squares*), suma del cuadrado de las desviaciones (*sum of squared deviations*), coeficiente de variación (*coefficient of variation*), mínimo (*minimum*), máximo (*maximum*), rango (*range*), número de datos ausentes —celdas vacías— (*number of missing cells*), media geométrica (*geometric mean*), media armónica

(*harmonic mean*), asimetría (*skewness*), curtosis (*kurtosis*), mediana (*median*), mediana de las desviaciones absolutas (*median absolute deviation*), moda (*mode*), rango intercuantílico (*interquartile range*), número de niveles únicos —que sólo aparecen una vez— de los datos (*number of unique levels of data*).

Transformación de datos

SalStat permite transformar fácilmente unos datos que ya están en una columna de la hoja de entrada para guardar los nuevos datos en otra nueva columna. Para **transformar datos**:

Preparation → *Transform...*



De nuevo aparece un cuadro en el que elegiremos tanto la columna que queremos transformar como la transformación que le queremos aplicar. Además de las funciones que aparecen por defecto, el hecho de que *SalStat* esté construido sobre *Python* hace que sea fácil aplicarles a los datos cualquier función que queramos definir nosotros, sólo tenemos que escribirla con la sintaxis de este lenguaje de programación, o utilizar las que ya existen en *Python*.

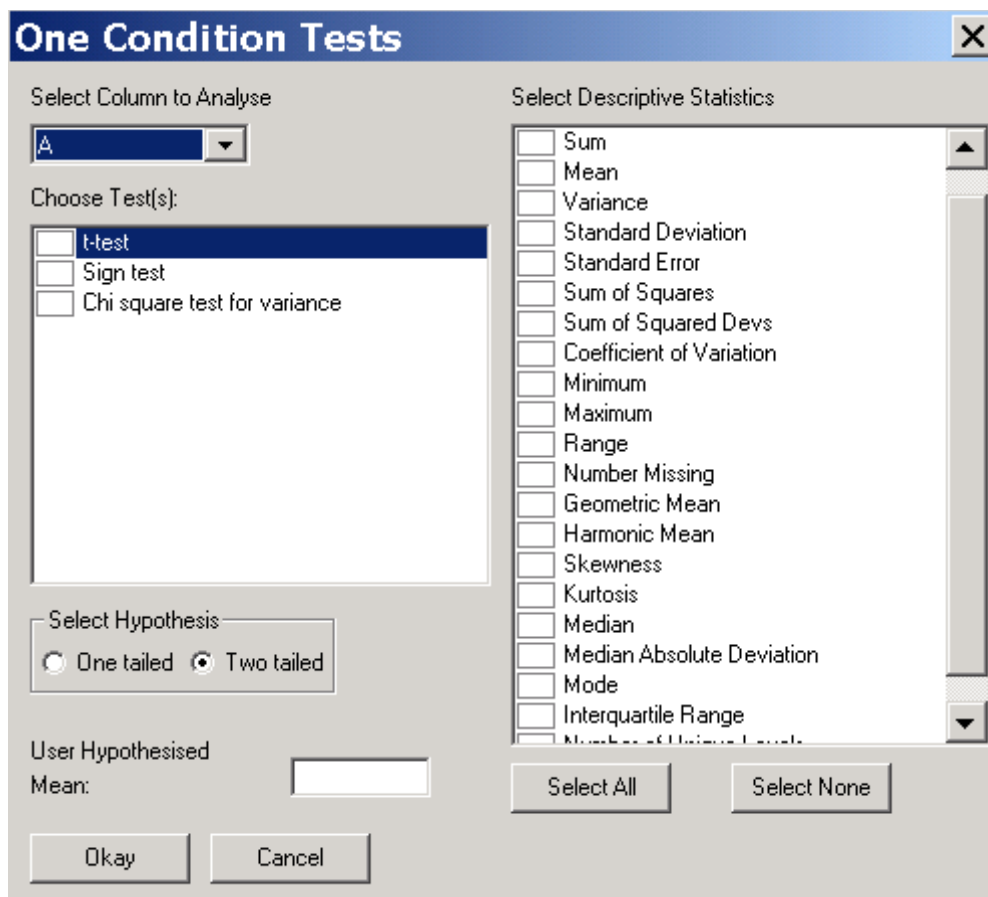
Ejemplos:

- 1) 'x+2' ó '2+x'
- 2) 'x*x*x' ó 'x**3'
- 3) Del módulo *math* de *Python*, añadiendo el prefijo «math.»: 'math.atan(x)', 'math.cos(x)', 'math.exp(x)', etcétera. Véase la documentación en <http://www.python.org>.

Comparaciones para una muestra

Para **hacer comparaciones de una muestra** respecto a un valor hipotético concreto:

Analyse → *1 Condition Tests...*



Aparecerá una ventana en la que se pueden seleccionar tanto algunos contrastes, a la izquierda, como algunos estadísticos descriptivos, a la derecha; unos y otros se calcularán a la vez. Es necesario indicar también la columna que se quiere analizar, si el contraste es unilateral o bilateral, y, a falta de otra muestra, el valor (*User Hypothesised Mean*) con el que queremos comparar nuestra muestra. Los contrastes que se pueden aplicar a una muestra son:

Comparaciones paramétricas:

- Contraste de la t (t -test)

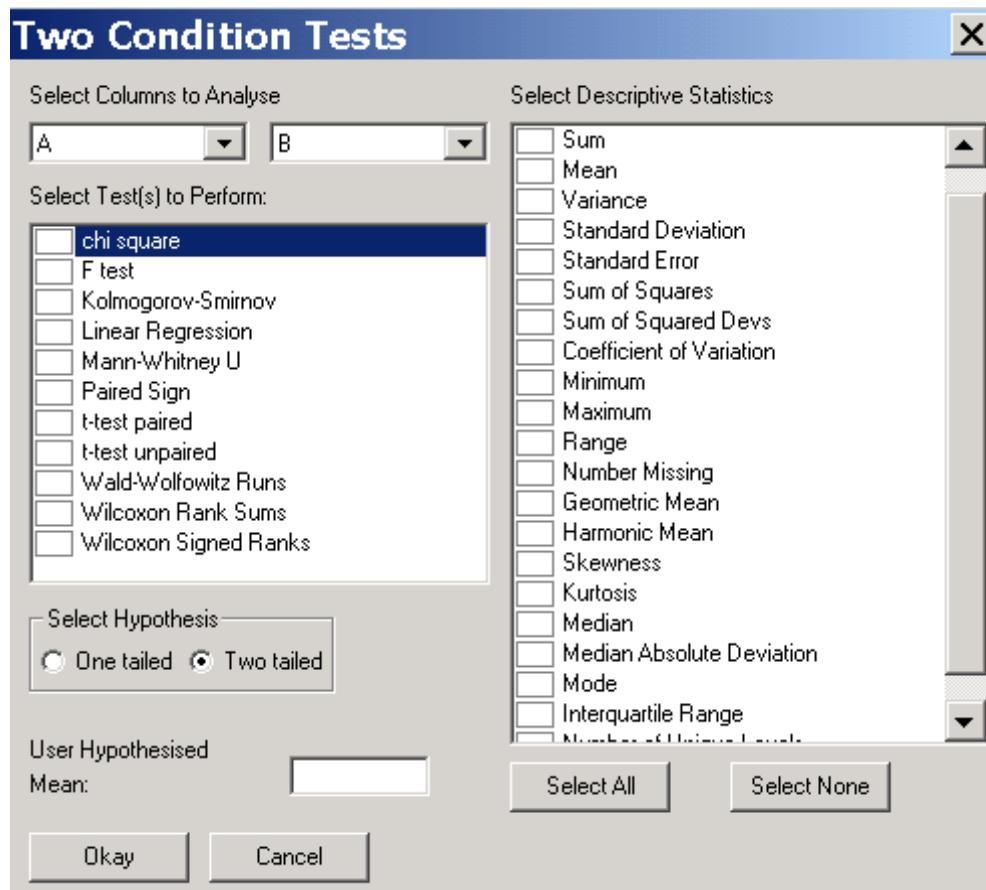
Comparaciones no paramétricas:

- Contraste de los signos ($sign$ test)
- Contraste Chi-cuadrado para la varianza (chi square test for variance)

Comparaciones entre dos muestras

Para hacer comparaciones entre dos muestras:

Analyse → 2 Condition Tests...



Aparecerá una ventana en la que se pueden seleccionar tanto algunos contrastes, a la izquierda, como algunos estadísticos descriptivos, a la derecha; unos y otros se calcularán a la vez. Los contrastes están ordenados por orden alfabético, los hay paramétricos y no paramétricos (y se pueden aplicar a la vez unos y otros). Es necesario indicar también las columnas que se quieren contrastar, si el contraste es unilateral o bilateral, y, a falta de otra

muestra, el valor (*User Hypothesised Mean*) con el que queremos comparar nuestra muestra. Los contrastes que se pueden aplicar a dos muestras son:

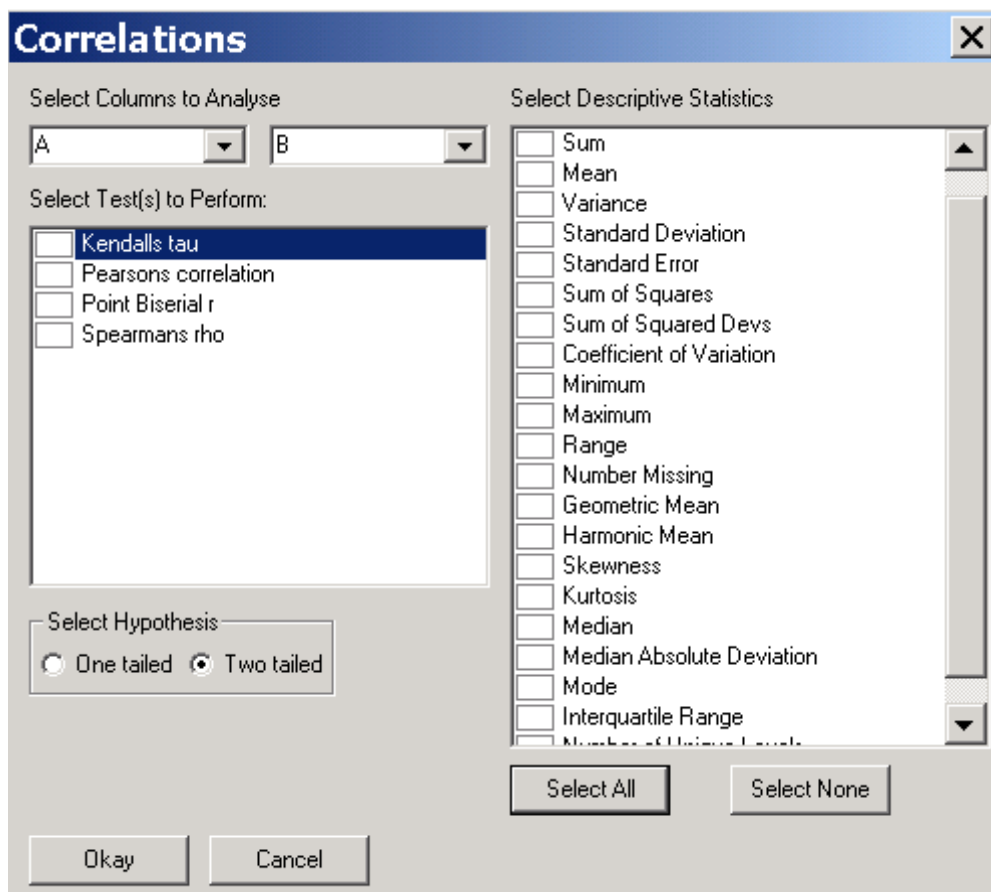
Comparaciones paramétricas:

- Contraste de la t no pareado (*unpaired t-test*)
- Contraste de la t pareado (*paired t-test*)
- Contraste de la F para el cociente de varianzas (*F test for variance ratio*). Este contraste necesita en valor hipotético con el que comparar (*User Hypothesised Ratio*)

Comparaciones no paramétricas:

- Contraste de los signos pareado (*paired sign test*)
- Contraste de los rangos signados de Wilcoxon (*Wilcoxon signed ranks test*)
- Contraste de la suma de rangos de Wilcoxon (*Wilcoxon rank-sums test*)
- Contraste U de Mann-Whitney (*Mann-Whitney U test*)
- Contraste de Kolmogorov-Smirnov (*Kolmogorov-Smirnov test*)
- Regresión lineal (*linear regression*)

Correlaciones entre dos muestras



Para estudiar la **correlación entre dos muestras**:

Analyse → Correlations...

Correlaciones paramétricas:

- Coeficiente de correlación de Pearson (*Pearson's correlation coefficient*)

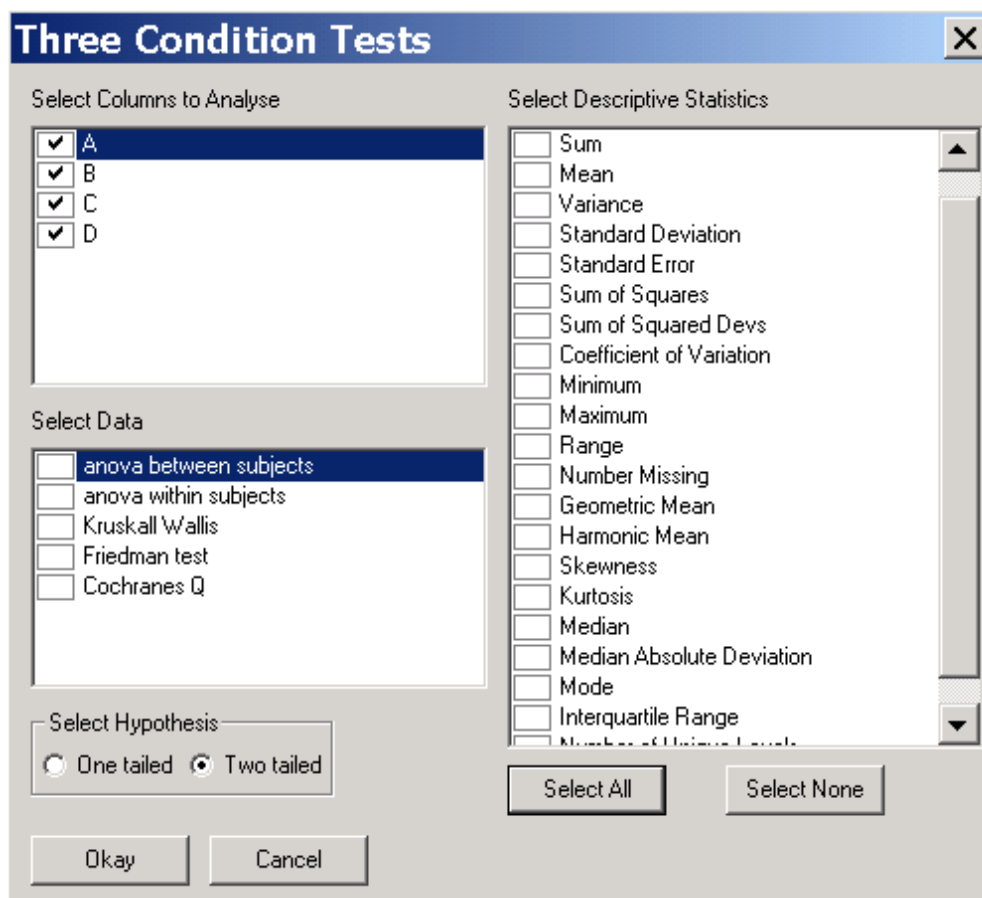
Correlaciones no paramétricas:

- Correlación de rango de Spearman (*Spearman's rank correlation*)
- Coeficiente de rango de Kendall (*Kendall's coefficient of rank*)
- Correlación biserial (*point biserial r correlation*)
- Regresión lineal (*linear regression*)

Comparaciones entre tres muestras o más

Para **hacer comparaciones de tres muestras**:

Analyse → 3+ Condition Tests...



Aparecerá una ventana de diálogo distinta a la de los casos anteriores, en la que se pueden seleccionar tanto algunos contrastes, a la izquierda, como algunos estadísticos descriptivos, a la derecha; unos y otros se calcularán a la vez. Es necesario indicar también las columnas que se quieren analizar, si el contraste es unilateral o bilateral y, a falta de otra muestra, el valor (*User Hypothesised Mean*) con el que queremos comparar nuestra muestra. Los contrastes que se pueden aplicar a una muestra son:

Comparaciones paramétricas:

- Análisis de la varianza (*analysis of variance* [anova]) interelementos e intraelementos.

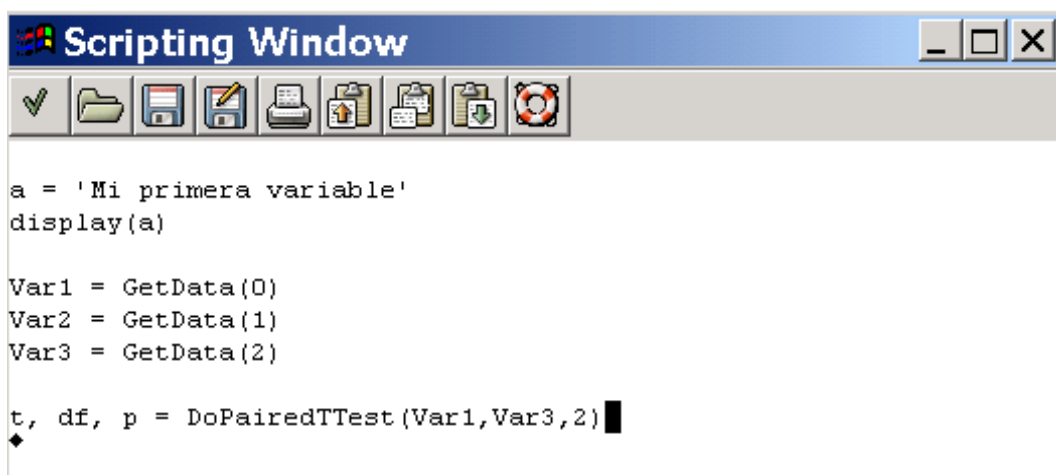
Comparaciones no paramétricas:

- Contraste *H* de Kruskal-Wallis (*Kruskal-Wallis H test*)
- Contraste Chi cuadrado de Friedman (*Friedman Chi Square*)
- Contraste *Q* de Cochranes (*Cochranes Q test*)

Archivos de código

Una de las principales vías para sacarle el máximo provecho a *SalStat* es la de **programar archivos de texto** (*scripts*). Conociendo la sintaxis básica del lenguaje *Python* es fácil programar código que haga prácticamente cualquier cosas.

Todos los análisis que hemos visto gráficamente, se pueden hacer con las correspondientes funciones o comandos. El manual oficial del paquete ([3]) dedica un capítulo y un apéndice a ver todos esos comandos y funciones, cómo hacer archivos de código (*scripts*), y ver cómo ejecutar código.



```
a = 'Mi primera variable'
display(a)

Var1 = GetData(0)
Var2 = GetData(1)
Var3 = GetData(2)

t, df, p = DoPairedTTest (Var1,Var3,2)
```

Para **abrir la ventana de código** hay que ir a:

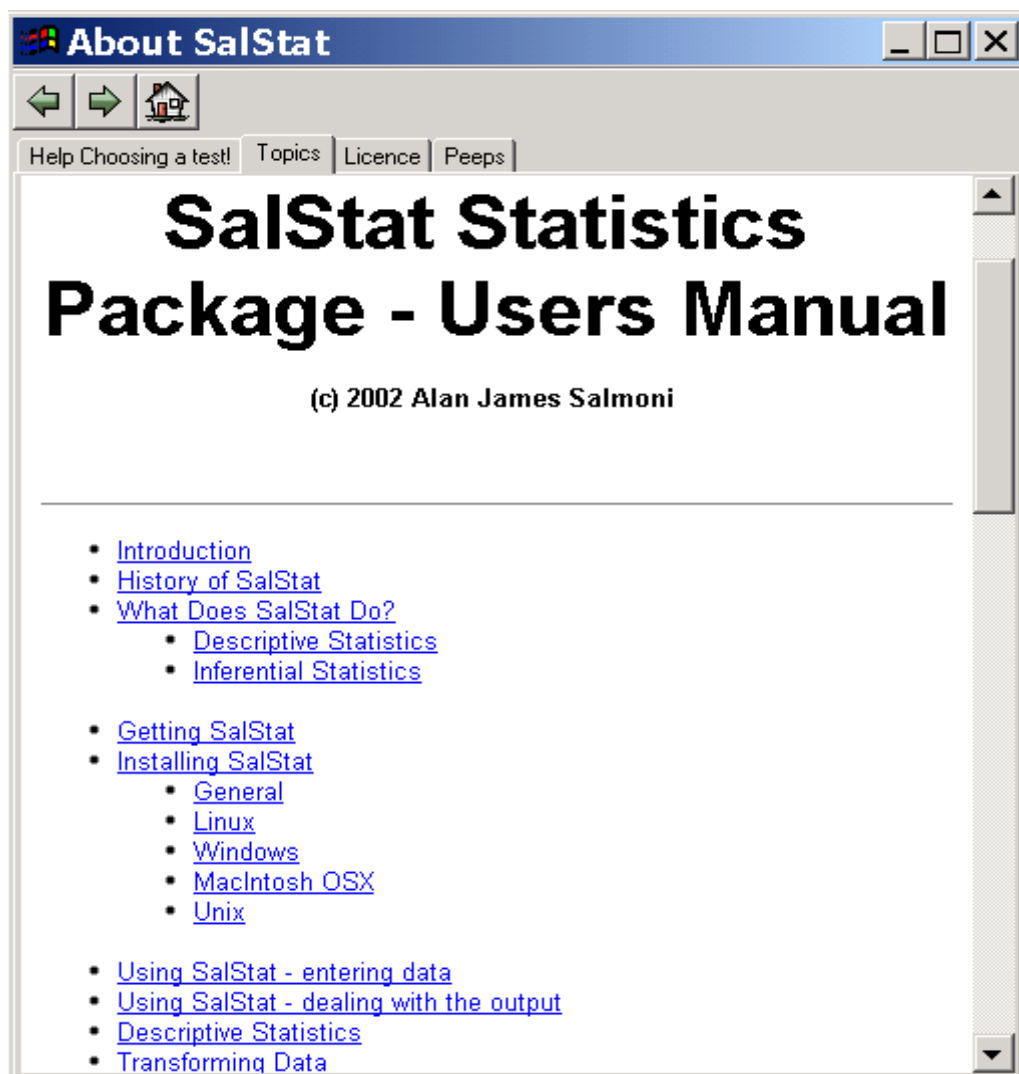
Analyse → *Scripting Window...*

Esta venta permite **editar** (copiar, cortar, pegar) **código**, **guardar y abrir los archivos** que se creen, y **ejecutar el código** pulsando el primer botón de la barra de herramientas.

Esta característica de *SalStat* lo convierte —salvando las diferencias— en una interfaz del lenguaje de programación *Python* sin necesidad de tenerlo instalado.

Ayuda

La ayuda del paquete se divide en cuatro bloques. Uno es la licencia bajo la que se distribuye *SalStat*, otro está dedicado a cómo hacer la referencias a él, la autoría y los agradecimientos. Los dos bloques más interesantes son: *What Test Shoud I Use...* y *Topics...* El primero está todavía en construcción, pero será sin duda de gran utilidad.



Futuras mejoras de *SalStat*

Según el apéndice que se incluye en el manual oficial ([3]), algunas de las próximas mejoras con que el autor prevé dotar al paquete son:

- Intérpretes para otros lenguajes: *Tcl* y *Perl*, principalmente, y quizá *Lisp* y *Scheme*
- Anova multifactorial con efectos simples
- MANOVA and MANCOVA
- Regresión múltiple, análisis factorial, análisis de la potencia y otros contrastes
- Otros formatos de datos desde los que importar datos: al menos *Excel* y *SPSS*
- Mejoras en las acciones de copiar, cortar y pegar de la hoja de datos; actualmente sólo se puede hacer celda a celda, no por bloques de ellas
- Incluir nuevos contrastes en la interfaz del programa
- Versiones para otras plataformas: principalmente *Mac OS7/8/9*, *BeOS* y *QNX*
- Mejores herramientas para hacer gráficos, quizá con enlaces externos a algunos programas profesionales, como *Gnuplot* o *DISLIN*
- Integración que haga posible el análisis de conjuntos de datos grandes. Se utilizará *Numeric* es un módulo de *Python* (compilado desde el lenguaje *C*) que permite trabajar con matrices de forma rápida y potente
- CD-ROM disponibles para vender el paquete desde el sitio web
- Capacidad de crear datos, para analizar los resultados con fines demostrativos o educativos

Referencias

[1] **Python**; <http://www.python.org>

[2] Salmoni, A. J. (2003); **SalStat Statistics Package** v. 20031022.;

<http://salstat.sourceforge.net/>; Available from <http://sourceforge.net/projects/salstat/>

[3] Salmoni, A. J. (2003); **SalStat Statistics Package - Users Manual**;

<http://salstat.sourceforge.net/manual.pdf>

[4] **wxPython**; <http://www.wxpython.org>

Nota: Este breve manual es el trabajo que hice para el curso «Recursos didácticos para usos docentes con aplicaciones informáticas de libre distribución», de la Universidad Nacional de Educación a Distancia (UNED), <http://www.uned.es>, dirigido por Antonio Colmenar Santos, e impartido en Ávila los días 17, 18 y 19 de noviembre del 2006. El contenido es una traducción, ordenación y actualización de materiales contenidos en el manual oficial de *SalStat*, [3], que está en inglés; pido perdón por si he cometido algún error en esa traducción e interpretación del manual.

David Casado de Lucas: <http://www.est.uc3m.es/dcasado/>, david.casado@uc3m.es

[Departamento de Estadística](#)

Universidad Carlos III de Madrid

C/ Madrid 126

28903 Getafe (Madrid)

