

Statistik für Jedermann



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

About the author:
Jürgen Pohl arbeitet als
Entwicklungsingenieur und
Übersetzer an der
Pazifikküste der USA.



Abstract:

Die meisten Statistik-Pakete sind in ihrem Umfang etwas überwältigend, sie zwingen den Benutzer auf einen steilen Lernpfad, ausserdem sind sie teuer. Es existieren jedoch einige hilfreiche Alternativen für jene, die solche am dringendsten benötigen: Anfänger, die gerade in die magische Welt der Statistik einsteigen und jene, für welche die umfangreichen kommerziellen Pakete einfach zu umfangreich sind. SalStat ist eine der gewünschten Alternativen – dank seines Schöpfers ist es Open Source, d.h. es ist gratis! Ein weiterer Vorteil: das Programm läuft auf den meisten Plattformen und ist sehr benutzerfreundlich. Das Programm ist in Python geschrieben, es kann jedoch ohne Vorkenntnisse dieser Sprache benutzt werden. Auf der Homepage des Programms ist eine deutsche Version des Handbuchs zu finden.

SalStat – das Statistikprogramm

Einleitung

Mit seiner eigenen Erfahrung im Hintergrund, war sich der Entwickler von SalStat wohl bewusst, in welcher Klemme sich die neu in die Statistik Eingeführten befinden: um ihr neues Wissen aufzubauen (oder einfach ihre Hausaufgaben oder ein Projekt auszuführen), ist ein erschwingliches, benutzerfreundliches Statistikprogramm erforderlich. Mit diesen Gedanken begann Alan James Salmoni SalStat zu entwickeln. Es wurde unter der GNU-Lizenz veröffentlicht, es ist auf seiner [Homepage](#) zu finden.

Was kann man mit SalStat anfangen?

Leider kann dieser Artikel keine Einführung in die Statistik anbieten, in ('Ressourcen') finden sich einige Quellen zu dem Thema.

Nachfolgend die Listen mit den Statistiken und Tests die SalStat ausführen kann.

Parametrische and Nicht-parametrische Tests sind hier zusammengefasst:

N (Gesamtumfang)	Bereich
Summe	Anzahl der fehlenden Zellen
Mittelwert	geometrisches Mittel
Varianz	harmonisches Mittel
Standardabweichung	Schiefe
Standardfehler	Kurtosis
Quadratsumme	Medianwert
Summe der Abweichungsquadrate	durchschnittliche absolute Abweichung
Varianzkoeffizient	Modalwert, häufigster Wert
Minimum	Quartilabstand
Maximum	Anzahl der spezifischen Datenniveaus

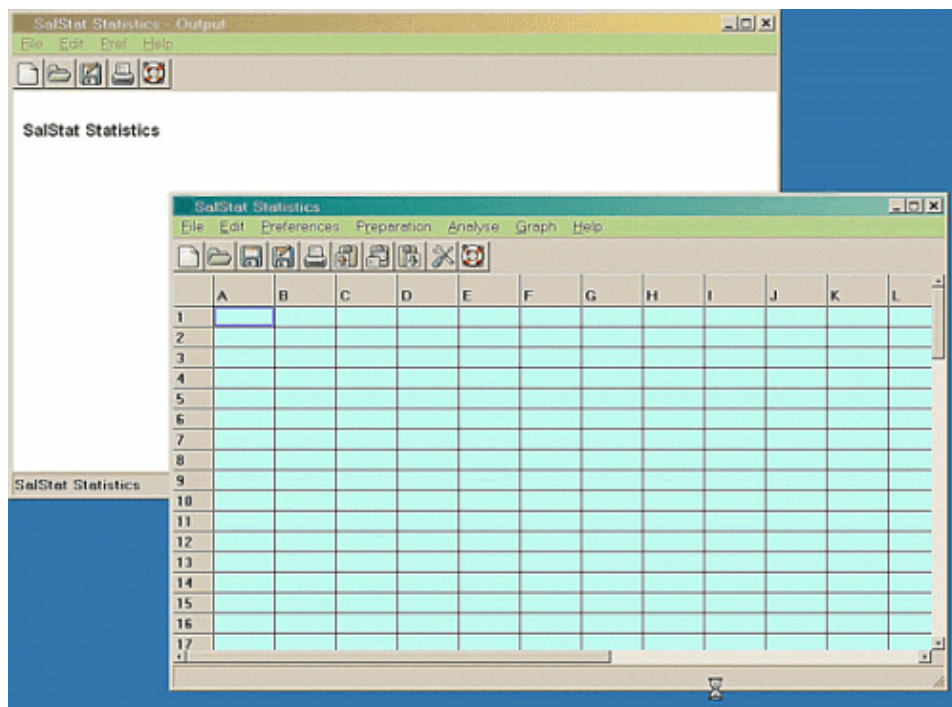
Inferenzstatistik

t-Test (paarweise)	Pearsons Korrelation
t-Test (nicht-paarweise)	Spearman's Rangkorrelation
Vorzeichenstest mit einer Stichprobe	Kendallsche Rangkorrelation
Vorzeichenstest mit zwei Stichproben	Punkt-biseriale Korrelation r
F-Test, Varianzverhältnistest	Lineare Regression
Wilcoxon Rangsummentest	Einfaktorvarianzanalyse (zwischen Datenreihen)
Mann-Whitney U-Test	Einfaktorvarianzanalyse (innerhalb einer Datenreihe)
Kolmogorov-Smirnov-Test	Kruskal-Wallis H-Test
Paarweiser Permutationstest	Friedmantest

Wie arbeitet man mit SalStat?

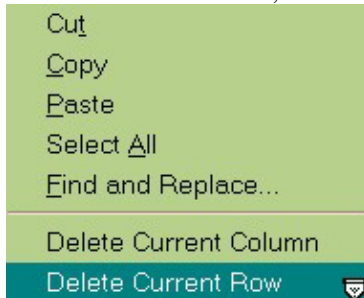
Bevor wir einen Blick auf SalStat werfen können, muss das Programm auf unserer Maschine installiert sein, darauf kommen wir jedoch später – einleitend möchte ich einen Eindruck von dem Programm geben. Beim Öffnen des Programms erscheinen zwei Fenster auf dem Bildschirm (ich muss hier einfügen, dass das Programm nicht lokalisiert ist, d.h. alle Werkzeuge, Auswahllisten, usw. sind in Englisch, z.Z. gibt es jedoch eine deutschsprachige Version des Benutzerhandbuchs – es kann von der Homepage heruntergeladen

werden):

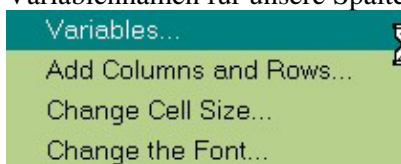


Das Fenster im Vordergrund oben (betitelt 'SalStat Statistics') ist im Tabellenformat: hier geben wir unsere zu analysierenden Daten ein. Am oberen Rand des Fensters das Menü mit den üblichen Werkzeugen – klicken wir auf ein Werkzeug, öffnet sich die zugehörige Dropdownliste zur Auswahl.


- **File:** Wie gewohnt.
- **Edit:** Die übliche Liste, wir können hier Spalten und Zeilen entfernen:



- **Preferences:** Unsere Tabelle kann hier an unsere Daten angepasst werden. Wir finden hier auch das erste statistikspezifische Werkzeug 'Variables', klicken wir darauf, wird ein Fenster geöffnet (siehe unten), es erlaubt uns anstelle der generischen A, B, C – Bezeichnungen die testspezifischen Variablenamen für unsere Spalten einzugeben.



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



- **Preparation:** hier suchen wir durch Markieren der Abhakkästchen die gewünschten (oder alle) Statistiken in der Wunschliste ('*Select Descriptive Statistics*') aus – natürlich dürfen wir nicht vergessen, die zu bearbeitenden Spalten auszuwählen ('*Select Column(s) to Analyse*');

Descriptive Statistics ✕

<p>Select Column(s) to Analyse</p> <div style="border: 1px solid black; padding: 5px;"> <input type="checkbox"/> A <input type="checkbox"/> B <input type="checkbox"/> C </div>	<p>Select Descriptive Statistics</p> <div style="border: 1px solid black; padding: 5px;"> <input type="checkbox"/> Sum <input type="checkbox"/> Mean <input type="checkbox"/> Variance <input type="checkbox"/> Standard Deviation <input type="checkbox"/> Standard Error <input type="checkbox"/> Sum of Squares <input type="checkbox"/> Sum of Squared Devs <input type="checkbox"/> Coefficient of Variation <input type="checkbox"/> Minimum <input type="checkbox"/> Maximum <input type="checkbox"/> Range <input type="checkbox"/> Number Missing <input type="checkbox"/> Geometric Mean <input type="checkbox"/> Harmonic Mean <input type="checkbox"/> Skewness <input type="checkbox"/> Kurtosis <input type="checkbox"/> Median </div>
---	--

- **Analyse:** hier wählen wir aus, welchen Test wir mit unseren Daten durchführen wollen. Das letzte Werkzeug in der Dropdownliste öffnet uns ein Skriptingfenster zur Ausführung von angepassten Tests, später mehr dazu.
- **Graph:** ist zur Zeit noch nicht funktional.
- **Help:** im Benutzerhandbuch finden wir ausführliche Informationen über das Programm. Eine deutschsprachige Version des Benutzerhandbuchs kann von der Homepage heruntergeladen werden.

Das zweite Fenster ('*SalStat Statistics – Output*') ist beim Öffnen leer, dort werden nach Durchführung des Tests die Ergebnisse ausgegeben.

Um dem Einstiegsfrust vorzubeugen, hat der Autor von SalStat in weiser Voraussicht eine Testdatei beigelegt (testreport1.txt). Wir geben also die Daten aus dieser Datei in unsere Tabelle ein, wählen die gewünschten Statistiken und Tests – nach Bestätigung mit 'Okay' wird kurz darauf das Ergebnis im Output-Fenster erscheinen.

Der Screenshot unten zeigt diesen Vorgang in dieser Reihenfolge:

1. Wir tragen die Daten der Testdatei in die Spalten A, B, C der Tabelle ein ('SalStat Statistics'-Fenster, unten links oben).
2. In der Werkzeugleiste wählen wir 'Analyse'. Im folgenden Dropdownmenü wählen wir den 'One Condition Test' (nur ein Beispiel...).
3. Das 'One Condition Test'-Fenster öffnet sich. Hier markieren wir in den Abhakkästchen unsere gewünschten Statistiken und Test(s):
 - ◆ Im 'Select Column to Analyse' wählen wir 'A'.
 - ◆ In 'Choose Test' haken wir 't-test' ab.
 - ◆ In der nächsten Liste 'Select Descriptive Statistics' wählen wir 'Sum', 'Mean', 'Standard Deviation', 'Range' und 'Skewness' aus (wieder ziemlich willkürlich...).
 - ◆ Bei 'Select Hypothesis' lassen wir die Grundeinstellung 'Two tailed'.
 - ◆ Als letztes müssen wir einen vom Benutzer hypothetisierten Mittelwert angeben, wir wählen den Wert 12300 (nur zur Demonstration, ziemlich willkürlich – nach einem kurzen Blick auf Testdaten...) .
4. Wir klicken auf 'Okay' und das vorher leere 'SalStat Statistics – Output'-Fenster zeigt uns die Ergebnisse.

The screenshot displays the SalStat Statistics application interface. It is divided into three main windows:

- SalStat Statistics (Main Window):** Shows a spreadsheet with data entered in columns A, B, and C. The data is as follows:

	A	B	C	D	E	F	G	H
1	10.0000	11.0000	9.0000					
2	11.0000	12.0000	10.0000					
3	12.0000	13.0000	8.0000					
4	11.0000	14.0000	9.0000					
5	10.0000	13.0000	8.0000					
6	9.0000	14.0000	7.0000					
7	10.0000	12.0000	8.0000					
8								
9								
- One Condition Tests (Dialog Box):** Shows the configuration for the test.
 - Select Column to Analyse:** A
 - Choose Test(s):** t-test, Sign test, Chi square test for variance
 - 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
 - Select Hypothesis:** One tailed, Two tailed
 - User Hypothesised Mean:** 12300
- SalStat Statistics - Output (Output Window):** Shows the results of the test.
 - SalStat Statistics:**

Statistic	A
Sum	73.00000
Mean	10.42857
Standard Deviation	0.97590
Range	3.00000
Skewness	0.21345
 - One sample t-test:**

t(6) = -33318.114, p (approx) = 0.000000

Statistic	A
Sum	73.00000
Mean	10.42857
Variance	0.95238
Standard Error	0.36886
Sum of Squares	767.00000
Sum of Squared Deviations	5.71429
Coefficient of Variation	0.09358
Range	3.00000
Number Missing	0
Geometric Mean	0.05770
Harmonic Mean	10.35101
Skewness	0.21345

SalStat herunterladen und installieren

Das Programm kann von seiner [Homepage](#) heruntergeladen werden. Dort finden wir einige Alternativen, um den entsprechenden Betriebssystemen gerecht zu werden. Ich habe zum Beispiel das Programm auf zwei Desktopmaschinen mit unterschiedlichen Betriebssystemen laufen. Der Sourcecode steht ebenfalls zur Verfügung – vielleicht möchte jemand seine Pythonerfahrung demonstrieren und zur Erweiterung des Programms beitragen... Vor der Installation bitte den *'Basic Users Guide'* lesen, der auf der SalStat-Homepage unter *'Documentation'* zu finden ist. Die Instruktionen dort sind sehr klar, so dass wir diese hier nicht wiederholen müssen.

SalStat Einstellungen

Eine sehr nützliche Eigenschaft von Salstat ist die Möglichkeit, durch Skripts persönliche Einstellungen vornehmen zu können, um Vorgänge zu automatisieren, eigene Tests zu bauen, usw. Im Handbuch (deutsche Version herunterladen, falls nötig) finden wir ausführliche Beschreibungen zum Skripting im *Anhang C – API-Details zum Skripting*, das ist auch eine gute Einführung zu diesem Thema, der Benutzer sollte sich ermutigt fühlen, Skripts zu schreiben.

Das letzte Werkzeug der Analyse-Dropdownliste führt direkt zum Skriptingfenster (*'Scripting Window'*), hier werden wir unsere Skripts eingeben: probieren wir doch einige von den Beispielen, die im Handbuch aufgeführt sind, gar nicht so schwierig, oder...?

Um weitgehende persönliche Anpassungen am Programm vornehmen zu können, sind Kenntnisse in Python Voraussetzung – in dieser Sprache ist das Programm geschrieben. Glücklicherweise ist Python eine freie Sprache, sie kann also gratis heruntergeladen werden. Ausführliche Einführungen und Dokumentation für Anfänger und Experten stehen auf der Python-Webseite zur Verfügung.

Falls sich Ihnen bei dem Gedanken die Haare sträuben – na ja, SalStat kann einfach benutzt werden wie es ist – mit Pythonkenntnissen könnten Sie eventuell mehr aus dem Programm herausholen.

Fazit

Die Grundidee für SalStat entsprang dem vielfachen Wunsch, ein einfaches Programm zur Verfügung zu stellen, das den Einsteiger ermutigt und den seriösen Benutzer ebenfalls zufriedenstellen kann. Der Benutzer kann sich in der gewohnten Weise einfach durch die Menüs der Statistiken und Tests klicken. Das Handbuch gibt Anweisungen für die Tests, ausserdem einige Hinweise über den Wert der Ergebnisse. Ganz allgemein wird jedoch vorausgesetzt, dass der Benutzer mindestens Grundkenntnisse der Statistik hat oder dabei ist, diese zu erwerben.

Eine Warnung: bevor Sie versuchen, Ihre Karriere als zukünftiger Wissenschaftler auf Daten aufzubauen, die Sie mit diesem Programm erzeugt haben, hören Sie auf die Empfehlung des Programmentwicklers und testen Sie das Programm. Anfänger können wahrscheinlich Beispiele mit Testergebnissen in ihren Fachbüchern finden. Tragen Sie einige Daten von Testbeispielen in Salstat ein und vergleichen Sie die Ergebnisse mit denen, die mittels anderer Programme gefunden wurden. Die Testdatei (*testreport1.txt*) ist ein Anfang, dort sind auch Vergleiche mit den Ergebnissen von Tests mit anderen Programmen aufgeführt.

Resources

Wer mit Hilfe seiner bevorzugten Suchmaschine im WWW wühlt, kann in Bezug auf Statistik viele Quellen finden, die besten und umfassendsten scheinen in Englisch zu sein.

- Hier ist eine sehr umfangreiche Sammlung von [Links](#), zusammengestellt und veröffentlicht von Clay Helberg.
- Ein online Textbuch und ein Glossar von Statsoft können [hier](#) gefunden werden, Statsoft ist auch für ein kommerzielles Statistikprogramm bekannt.
- Das [International Statistical Insitute](#)
- Ein vorzügliches vielsprachiges Glossar über die [Europäische Union](#) zu erhalten.
- *'Introduction to the Practice of Statistics'* von David S. Moore and George P. McCabe. Das scheint ein gutes, weitverbreitetes Buch über die Anwendung von Statistik zu sein – mit vielen Beispielen und Übungsaufgaben. Einige Versionen des Buches kommen mit einer CD, welche die Daten für die Übungen, als auch die Lösungen enthält – diese Daten sind auch von der Webseite des Autors herunterladbar. Das Textbuch ist auch für wenig Geld antiquarisch online zu finden, soweit ich weiss, existiert es leider nur in Englisch.
- Bei der [Uni Bielefeld](#) ist auch was zu finden...

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