

Les statistiques vous intéressent ?

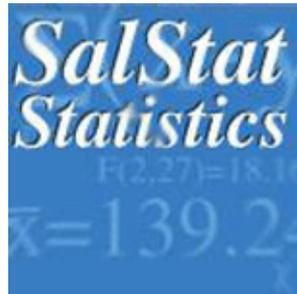


par Jürgen Pohl
<sept.sapins(at)verizon.net>

L'auteur:

Jürgen Pohl travaille comme ingénieur de recherche et développement et comme traducteur technique sur la côte pacifique des USA.

Traduit en Français par:
Jean-Etienne Poirrier
([homepage](#))



Résumé:

La plupart des suites logicielles de statistiques semblent avoir une portée accablante, forçant l'utilisateur à suivre une courbe d'apprentissage très raide et la plupart d'entre elles sont très onéreuses. Cependant, il existe quelques alternatives adaptées pour ceux qui en ont le plus besoin : le débutant, qui est initié aux secrets du monde magique des statistiques, tout comme l'utilisateur pour qui les suites commerciales seraient trop puissantes. SalStat est une de ces solutions de rechange recherchées et grâce à son créateur, il est Open Source, c'est-à-dire : libre ! Un autre avantage : le programme est indépendant de la plateforme. Dernier avantage, et non des moindres : il est très facile à utiliser. Le programme est écrit en Python mais peut, cependant, être utilisé sans connaissances préalables de ce langage.

SalStat – le programme de statistiques

Introduction

Comptant sur sa propre expérience, le créateur de SalStat était bien au courant de la situation fâcheuse dans laquelle de nombreuses personnes introduites aux statistiques se trouvent : pour étendre leurs connaissances nouvellement acquises (ou juste pour effectuer leurs devoirs ou projets...), un programme de statistiques accessible et facile à utiliser est essentiel sur leur propre machine. Avec cela en tête, Alan James Salmoni a développé SalStat et l'a publié sous la licence GNU. Il peut être trouvé sur sa [page web](#).

Que peut faire SalStat ?

Malheureusement, cet article ne peut pas fournir une introduction aux statistiques. Dans la rubrique « Ressources » (ci-dessous), vous trouverez quelques informations sur ce sujet. Voici les listes des statistiques et des tests que SalStat est capable de générer :

Les tests paramétriques et non-paramétriques sont combinés ici :

N (compte)	gamme
somme	nombre de cellules manquantes
moyenne	moyenne géométrique
variance	moyenne harmonique
écart type	skewness
erreur type	kurtosis
somme des carrés	médiane
somme des écarts types carrés	médiane de déviation absolue
coefficient de variation	mode
minimum	gamme interquartile
maximum	nombre de niveaux uniques de données

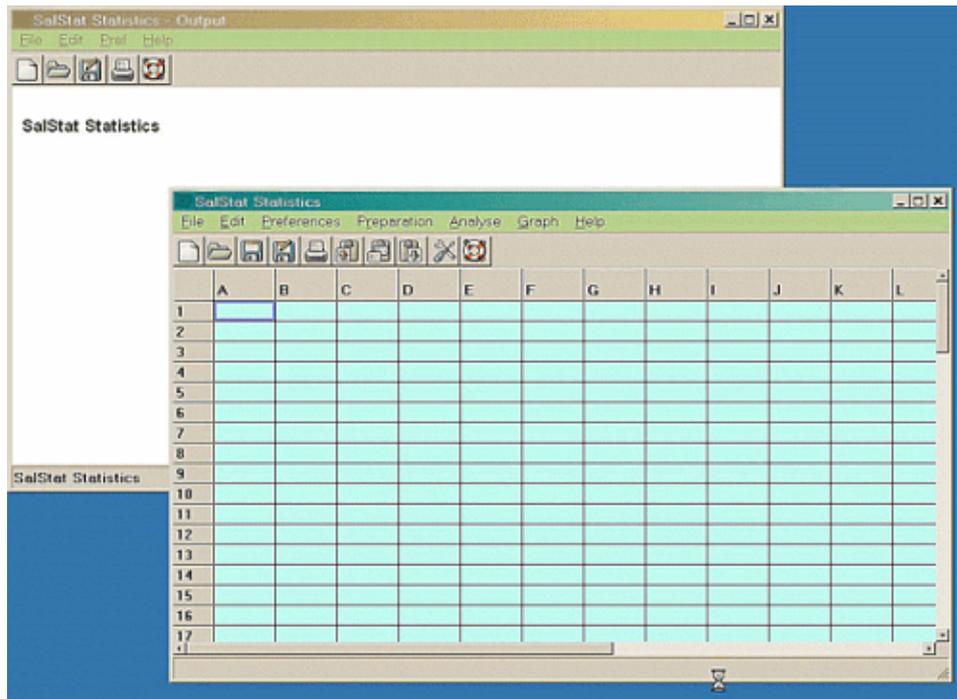
Statistique inférentielle

test t (pair)	corrélacion de Pearsons
test t (impair)	corrélacion rho de Spearmans
test du signe sur 1 échantillon	corrélacion thau de Kendalls tau
test du signe sur 2 échantillons	corrélacion r du point bisériel
test F pour le rapport de variance	régression linéaire
test des sommes rangées de Wilcoxon	analyse unifactorielle de la variance (entre sujets)
test U de Mann-Whitney	analyse unifactorielle de la variance (à l'intérieur des sujets)
test de Kolmogorov-Smirnov	test H de Kruskall-Wallis
test des permutations pairées	test de Friedman

Comment travaille-t-on avec SalStat ?

Avant de jeter un oeil à SalStat, vous devez l'installer sur votre machine, mais nous nous occuperons de cela plus tard. D'abord, je voudrais vous donner une idée de ce que vous aurez.

Lorsque vous lancez le programme, ces deux fenêtres apparaissent sur votre écran :

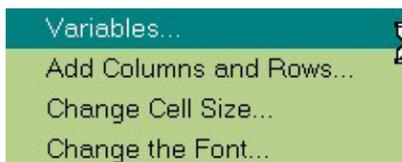


La première fenêtre (intitulée « *SalStat Statistics* »), à l'arrière plan de la capture d'écran ci-dessus, montre une grille d'entrée des données comme dans la plupart des tableurs : nous entrons ici les données de nos échantillons à analyser. Au sommet de la fenêtre se trouve la barre d'outils usuelle et cliquer sur un de ces outils amène une liste déroulante de fonctions à sélectionner.

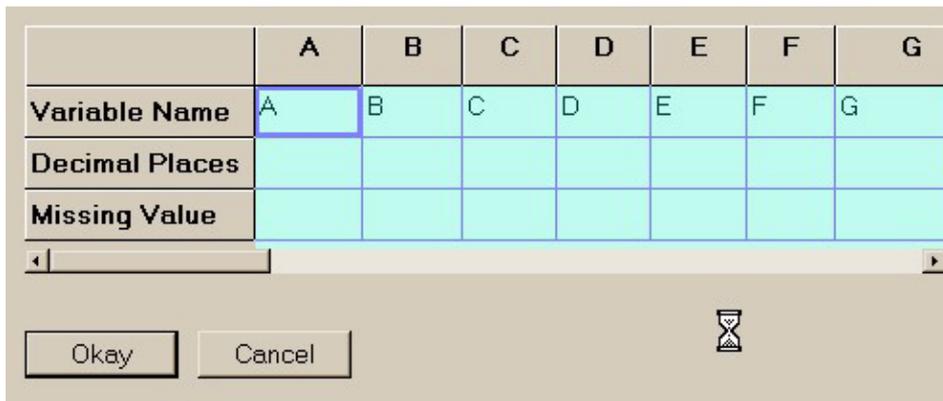
- **File:** Comme d'habitude.
- **Edit:** Comme d'habitude aussi, mais vous pouvez ajouter des colonnes ou des lignes à votre table :



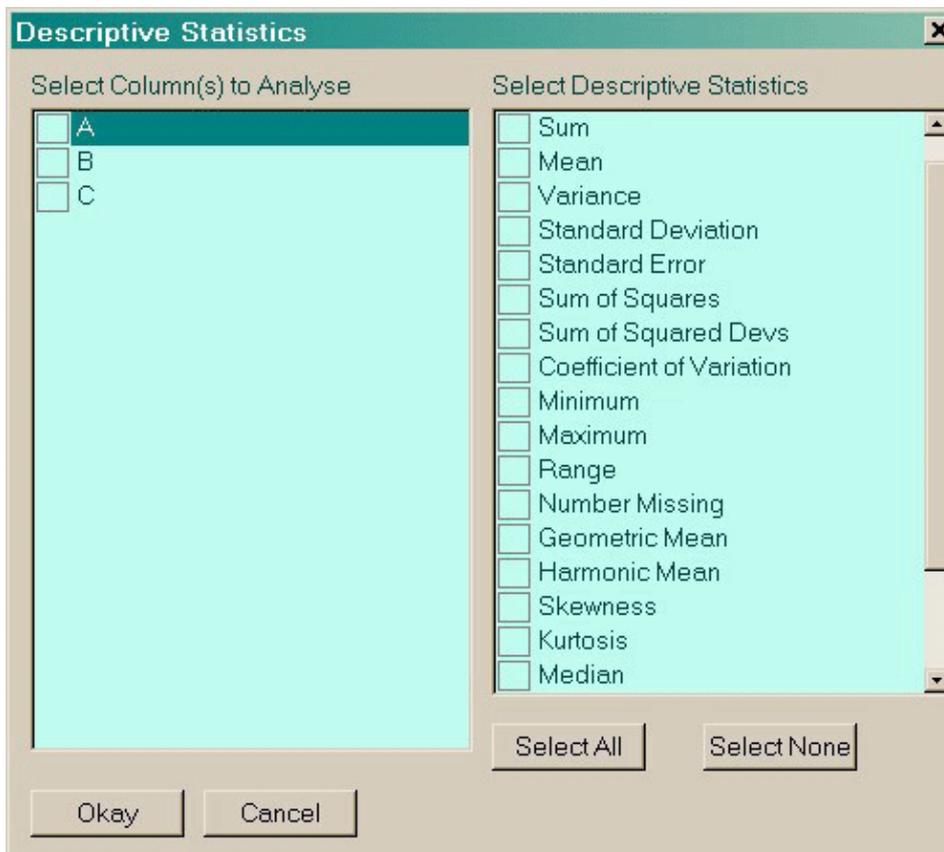
- **Preferences:** nous devenons un peu plus spécifiques quant aux statistiques :



Ici, vous trouvez les outils pour adapter les cellules de la table à vos entrées. Aussi important : vous pouvez nommer les variables spécifiques de votre test ou pour vos statistiques. En cliquant sur « *Variables* », la fenêtre suivante s'ouvre et vous pouvez taper les noms des variables pour chaque colonne :



- **Preparation:** nous sommes invités à sélectionner les statistiques que nous désirons – il y a une longue liste de choix possibles ; cochez juste les statistiques que vous souhaitez ou sélectionnez–les toutes. N'oubliez pas de cocher les colonnes de données que vous souhaitez analyser (vous ne pouvez pas analyser des données qui n'existent pas ...) :



- **Analyze:** « Analyze » est l'ensemble suivant d'outils dans notre barre d'outils, où vous choisissez le type de tests que vous voulez effectuer sur vos données
- **Graph:** n'est pas encore fonctionnel.
- **Help:** ici, vous trouverez des informations générales et spécifiques pour vous permettre de continuer.

La seconde fenêtre (intitulée « *SalStat Statistics –Output* »), initialement vide, va montrer les résultats statistiques des tests exécutés.

Rien n'est plus frustrant que d'avoir un programme comme celui-là sans aucune donnée pour jouer avec. L'auteur a été sage d'inclure un fichier de test (`testreport1.txt`) avec des résultats connus. Entrez simplement les données de test dans votre table et choisissez quel type de statistiques ou de tests vous voulez voir et cliquez sur « *Okay* » et les résultats vont apparaître dans la fenêtre de sortie (output). Les fenêtres ci-dessous sont apparues dans la séquence suivante :

1. Nous avons entré les données dans les colonnes A, B et C de la grille des données de notre fenêtre *SalStat Statistics* ouverte. Les données proviennent du fichier d'exemple `testreport1.txt`.
2. Ensuite, nous avons utilisé l'outil « *Analyse* » de la barre d'outils. Ici, nous choisissons le « *One Condition Test* »
3. La fenêtre « *One condition Test* » s'est ouverte. Nous avons marqué les spécificités de notre test :
 - ◆ « *Select Column to Analyse* » : nous avons choisi « *A* ».
 - ◆ « *Choose Test(s)* » : nous avons sélectionné le « *t- test* ».
 - ◆ « *Select Descriptive Statistics* » : nous avons choisi « *Sum* », « *Mean* », « *Standard Deviation* », « *Range* » et « *Skewness* ».
 - ◆ « *Select Hypothesis* » : nous avons gardé la valeur par défaut, « *Two tailed* ».
 - ◆ « *User Hypothesised Mean* » : nous avons juste deviné ici pour cette démonstration et nous avons entré la valeur de 12300.
 - ◆ Cliquez sur le bouton « *Okay* ».

La fenêtre « *SalStat Statistics – Output* » auparavant vide (voir notre copie d'écran ci-dessous, à gauche) va maintenant montrer les résultats du test.

The screenshot displays three windows from the SalStat Statistics application:

- SalStat Statistics (Main Window):** Shows a data table with 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 (Configuration Window):** Shows settings for the test:
 - Select Column to Analyse: A
 - Choose Test(s): t-test
 - Select Hypothesis: Two tailed
 - User Hypothesised Mean: 12300
 - Select Descriptive Statistics: Sum, Mean, Standard Deviation, Range, Skewness
- SalStat Statistics - Output (Results Window):** Displays the results of the t-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

Récupérer et installer SalStat

Le programme peut être téléchargé à partir de son [site web](#). Un certain nombre de solutions sont offertes pour s'accomoder de votre (vos) système(s) d'exploitation. J'ai ce programme sur deux ordinateurs avec des systèmes d'exploitation différents en deux endroits différents. Le code source est également disponible en téléchargement – peut-être voulez vous montrer vos talents de programmation (en Python) ... ? Avant d'essayer toute installation, lisez, s'il vous plaît, le « *Basic Users Guide* », également disponible sur le site web de SalStat dans la catégorie « *Documentation* ». Le guide donne des instructions claires sur la manière d'installer le programme. Ainsi, nous n'avons pas besoin de les répéter ici ; veuillez consulter le site web, s'il vous plaît.

Personnaliser SalStat

Une autre partie très utile de SalStat est sa capacité intrinsèque offerte aux utilisateurs d'écrire leurs propres scripts – pour automatiser des tâches, construire leurs propres tests, etc. Dans le manuel du programme, qui est la partie principale de l'outil « *Help* », vous pouvez trouver une description détaillée sur la manière de procéder (« *Scripting and Making Your Own Tests* »). Jetez-y un oeil, c'est une introduction très utile au langage à script. Il devrait encourager l'utilisateur à utiliser la fonctionnalité des scripts. Le dernier outil de la liste d'« *Analyse* » vous donne accès à la « *Scripting Window* » – ici, vous pouvez entrer vos propres scripts : essayez les exemples donnés dans le manuel, ils vous convaincront d'utiliser cette fonctionnalité.

Un inconvénient mineur pour certaines personnes : le programme est écrit en Python. Pour effectuer de sérieuses personnalisations du programme, vous devrez apprendre ce langage. Heureusement, Python est un langage ouvert, c'est-à-dire libre : vous pouvez le télécharger avec nombres d'instructions (s'adressant à tous, du débutant à l'expert) à partir du site web de [Python](#). Si vous ne voulez pas avoir affaire à Python, vous pouvez utiliser SalStat tel quel – mais, avec quelques connaissances en Python, vous pourriez tirer plus de bénéfices de ce programme.

Conclusion

SalStat a été écrit avec la facilité d'utilisation en tête. L'utilisateur peut cliquer à sa propre manière à travers une liste de statistiques et de tests. Le manuel donne des instructions pour tous les tests, incluant quelques conseils sur la valeur de leurs résultats. En général, cependant, il est supposé que vous avez au moins les connaissances de base sur les statistiques ou que vous êtes en train de les acquérir.

Un mot de mise en garde : avant de sauter dedans et de miser votre carrière en tant que scientifique d'avenir en utilisant les résultats générés par ce programme, écoutez les recommandations de son créateur et convaiquez-vous vous-même de ces mérites en le testant ! Ceux qui débutent juste avec les statistiques trouveront de nombreux exemples dans les livres de référence. Mettez quelques données disponibles dans SalStat et voyez ce que vous obtenez. Le fichier de test (*testreport1.txt*), qui est livré avec le programme, donne quelques comparaisons de résultats de tests obtenus avec d'autres programmes.

Ressources

- Ici, nous avons une importante collection de [liens](#) liés aux statistiques, collectée et publiée par Clay Helberg.
- Un livre en ligne de Statsoft et un glossaire peut être trouvé [ici](#), Statsoft est connu pour ces logiciels commerciaux de statistiques.
- Le [International Statistical Institute](#)
- Une excellent glossaire multi-langues est offert par l'[Union Européenne](#)
- « *Introduction to the Practice of Statistics* » par David S. Moore et George P. McCabe. Un bon livre sur le sujet, il semble être largement utilisé. La version que j'ai trouvé dans une bibliothèque publique possède un CD avec toutes les données pour un grand nombre d'exercices tirés du livre. Les données d'échantillons pour les exercices et leurs résultats sont aussi disponibles gratuitement sur le site web des auteurs. J'ai trouvé des versions usagées du livre disponibles dans les librairies en ligne pour très peu d'argent (5 US\$).

<p>Site Web maintenu par l'équipe d'édition LinuxFocus © 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 --> fr: Jean-Etienne Poirrier (homepage)</p>
--	--