

S-Plus

Prise en main rapide

Rachid BOUMAZA

INH
Département ETIC

rachid.boumaza@inh.fr

AVANT-PROPOS

Ce document n'est pas un manuel d'utilisation du logiciel S-Plus mais une invitation à aller découvrir ce logiciel. Merci d'avance à ceux qui prendront le temps de me signaler toute erreur (y compris d'orthographe).

Les manuels d'utilisation sont inclus dans le CD d'installation et sont disponibles à partir de l'aide en ligne (**Help**) après le lancement du logiciel.

Ce logiciel est disponible pour tous les enseignants et tous les étudiants de l'INH (postes réseau et postes individuels de l'INH).

Pour obtenir le CD d'installation, veuillez prendre contact avec le service informatique.

Son installation sur machine personnelle (à domicile) ou en entreprise (lors de stage) est possible et **soumise à conditions** (formulaire **obligatoire** à remplir et à envoyer à Insightful).

SOMMAIRE

1	LANCER / QUITTER S-PLUS	4
2	SAISIR / SAUVEGARDER / LIRE LES DONNEES	5
3	DESCRIPTION DE DONNEES UNIVARIEES	5
3.1	Cas qualitatif	5
3.2	Cas quantitatif	6
3.3	Ajustement à une loi de Laplace Gauss	6
4	TRAITEMENTS BIVARIES	6
4.1	Cas de deux variables quantitatives	7
4.2	Cas d'une variable qualitative et d'une variable quantitative	7
4.3	Cas de deux variables qualitatives	7
4.4	Comparaison de deux moyennes (test de Student)	7
4.5	Comparaison de plusieurs moyennes	8
4.6	Comparaison de deux ou plusieurs proportions ou test du χ^2 d'indépendance de deux variables qualitatives	8
4.7	ANOVA à un facteur	8
4.8	Test d'indépendance de deux variables quantitatives et régression linéaire simple	9
5	TRAITEMENTS MULTIVARIES : ANALYSE DES DONNEES	9
5.1	Analyse en composantes principales	9
5.2	Classification hiérarchique	9
6	AUTRES TRAITEMENTS MULTIVARIES	10
6.1	Régression linéaire multiple	10
6.2	Analyse de la variance à deux facteurs ou plus	10
7	SAUVEGARDER LES TRAITEMENTS REALISES	10
8	REDIGER / IMPRIMER UN RAPPORT	11
9	EXEMPLE D'UTILISATION DE <i>COMMANDS</i>	12

9.1	Test de Fisher – Snedecor de comparaison de deux variances-----	14
9.2	Test de Student-----	14
9.3	Test de Wilcoxon-----	15
10	EXEMPLE D'UTILISATION DE <i>SCRIPT</i>-----	15

1 Lancer / Quitter S-Plus

S-Plus se lance soit par un double clic sur l'une des icônes (Fig.1).



Fig.1

L'écran d'accueil se présente comme suit :

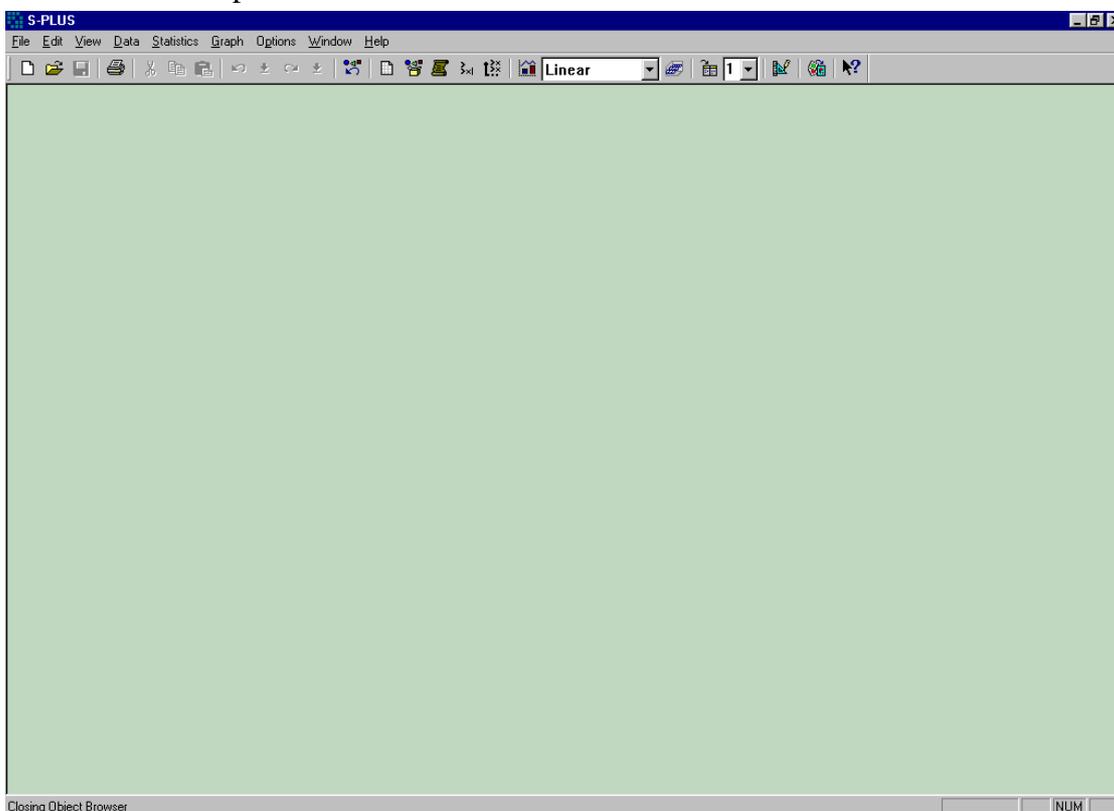


Fig.2

La première ligne est une *barre de titre* avec à droite les trois classiques boutons de Réduction - Agrandissement - Fermeture (ce dernier bouton permet de quitter S-Plus). A essayer si vous n'y êtes pas familier.

La deuxième ligne est la *barre de menus* (à découvrir au fur et à mesure), suivie par une ligne d'*icônes* qui permettent d'accéder un tout petit peu plus vite à certains des menus précédents. Suit un bureau où on peut trouver –suivant les paramètres d'initialisation du logiciel– une ou plusieurs fenêtres que l'on peut fermer.

La dernière ligne de la figure est la *barre d'état* de S-Plus et précède la *barre d'état* de Windows.

S-Plus se quitte en cliquant sur ► **File** ► **Exit**.

2 Saisir / Sauvegarder / Lire les données

En sélectionnant le menu **File**, puis par clics de souris successifs sur ► **New** ► **Data Set** ► **OK**, on obtient la fenêtre de saisie des données qui se présente comme une matrice (tableau rectangulaire de cellules) dont les lignes et colonnes sont automatiquement numérotées. Les colonnes correspondent aux variables statistiques, les lignes correspondent aux individus.

La première ligne porte les numéros des variables et est suivie d'une ligne grisée qui portera le nom de ces variables. Le nom des variables est obligatoire pour tout traitement statistique ; il est donné automatiquement par le logiciel sous la forme **Vn** (**n** est le numéro de la colonne), dès qu'on saisit une valeur dans une cellule quelconque d'une colonne.

De même, la première colonne numérote les individus et est suivie d'une colonne grisée qui contiendra le nom (optionnel) de ces individus.

On remarquera la nouvelle barre d'icônes : elle pourra servir à formater les données, voire les classer par ordre croissant ou décroissant.

On peut (re)nommer les colonnes (les variables), les formater, définir leur type (quantitative ou qualitative)... ; pour cela, on sélectionne la colonne par un clic gauche de souris –*bouton de gauche*– sur le nom de la colonne puis par un clic droit de souris –*bouton de droite*– on accède à un menu dans lequel :

- ► **Change Data Type** permet de définir le type de la variable... (type **factor** pour une variable qualitative) ;
- ► **Properties** permet de renommer la variable...

On peut commencer à saisir les données. L'enregistrement des données dans un fichier se fait au moyen de la commande ► **File** ► **Save**.

La lecture d'un fichier de données déjà existant pour consultation se fait par la commande ► **File** ► **Open**, lorsque ce fichier a été créé sous S-Plus ; lorsque c'est un fichier Excel, on peut le lire par la commande ► **File** ► **Import Data** ► **From File...**

Il existe d'autres possibilités pour importer des données saisies dans d'autres formats (fichier texte, fichier Access...).

3 Description de données univariées

Il s'agit de traitements qui ne font intervenir qu'une seule colonne (c'est-à-dire une variable) du fichier à la fois.

3.1 Cas qualitatif

La distribution de fréquences d'une variable qualitative (de type **Factor** dans S-Plus) s'obtient par la commande ► **Statistics** ► **Data Summaries** ► **Crosstabulations...** Le premier onglet permet de choisir la variable à traiter. Le deuxième onglet permet de choisir les calculs à opérer.

La représentation graphique de cette distribution de fréquences peut se faire sous forme de diagramme en barres par la commande ► **Graph** ► **2D Plot...** ► **Linear** ► **Bar with Base at Y Min(x,y1..yn)**, ou de camembert par la commande ► **Graph** ► **2D Plot...** ► **Pie** ► **Pie Chart (x)**.

Après avoir sélectionné le traitement, une fenêtre s'affiche pour sélectionner la variable dans la rubrique **x Columns**.

Une fois le résultat obtenu, on peut l'arranger (légendes, taille du graphique, couleurs...) en sélectionnant l'élément (axe...) à l'aide d'un clic gauche de souris, puis en accédant au menu correspondant par un clic droit. Ainsi, si par hasard il vous apparaît un diagramme en barres avec un nombre de barres inférieur au nombre de modalités, il faut "reconfigurer" l'axe vertical, en positionnant l'origine de l'axe en zéro.

Un titre peut être inséré par la commande ► **Insert** ► **Titles** ► **Main....**

3.2 Cas quantitatif

Pour une variable quantitative, on peut calculer et représenter sa distribution de fréquences, comme on peut calculer des paramètres de tendance (moyenne, médiane, quartiles...) et des paramètres de dispersion (variance, écart-type...).

Pour les paramètres, ils s'obtiennent par la commande ► **Statistics** ► **Data Summaries** ► **Summary Statistics....**

La distribution de fréquences d'une variable discrète se calcule comme celle d'une variable qualitative.

Pour une variable continue, sa distribution de fréquences peut être obtenue après avoir codé la variable ; ce que réalise la commande ► **Data** ► **Create Categories** en accédant au menu. Ce menu permet de sélectionner la variable à traiter, le nom de la variable qualitative résultante, les classes à créer, leur nombre, leur forme...

Une fois, l'opération de codage terminée, on obtient une variable qualitative, de type factor, dont la distribution de fréquences avec sa représentation graphique peuvent être obtenues comme au paragraphe précédent.

Cependant, l'histogramme d'une variable quantitative peut s'obtenir directement par ► **Graph** ► **2D Plot...** ► **Linear** ► **Histogram (x)**, ce qui est bien évidemment beaucoup plus rapide.

3.3 Ajustement à une loi de Laplace Gauss

Pour procéder à un ajustement par quantiles, ou QQ-Plot, on sélectionne la colonne puis par ► **Graph** ► **2D Plot...** ► **QQ Normal with Line (x)**, on visualise un nuage de points et une droite. On rappelle que si ces points sont "alignés", on peut considérer que les données s'ajustent à une loi de Laplace-Gauss.

On dispose aussi du test de Kolmogorov-Smirnov, qui est un test d'adéquation à une loi absolument continue, comme la loi normale. Après sélection de la variable, on procède au traitement ► **Statistics** ► **Compare Samples** ► **One Sample** ► **Kolmogorov-Smirnov GOF...** ► (GOF... pour Goodness of fit test) qui donne la P-valeur du test au vu de laquelle, on aboutit au rejet ou à la plausibilité de l'hypothèse de normalité.

4 Traitements bivariés

Ce sont des traitements qui font intervenir deux colonnes. Ce sera des traitements du type :

- comparaison de deux échantillons (test de Student par exemple) ;
- indépendance de deux variables (test du χ^2 – khi-2 – dans le cas qualitatif x qualitatif, ou le test sur le coefficient de corrélation linéaire dans le cas quantitatif x quantitatif) ;
- ANOVA (analysis of variance) à un facteur qui fait intervenir donc une "variable" qualitative et une variable quantitative ;
- régression linéaire simple qui fait intervenir deux variables quantitatives.

Avant de procéder à ces traitements, il est souhaitable de faire des représentations graphiques des données. Sans prétendre à l'exhaustivité, les paragraphes suivants donnent quelques exemples de traitements.

4.1 Cas de deux variables quantitatives

Chaque individu est représenté par un point dans un repère orthogonal. Ce graphique est obtenu par le traitement ► **Graph** ► **2D Plot...** ► **Linear** ► **Scatter Plot (x, y1, y2, ...)**. On porte en abscisse la première variable (dans la rubrique **x Columns**) et en ordonnée la deuxième variable (dans la rubrique **y Columns**). On peut à partir de ce graphique voir comment varie une variable en fonction de l'autre.

4.2 Cas d'une variable qualitative et d'une variable quantitative

La variable qualitative engendre une partition de la population en sous-populations, chacune constituée des individus prenant la même modalité. On peut représenter, par exemple par un histogramme, la distribution de fréquence de la variable quantitative pour chaque sous-population. Cela s'obtient par le traitement ► **Graph** ► **Multipanel Graph...** ► **Linear** ► **Histogram (x)** qui fait accéder à un menu où la rubrique **x Columns** sera renseignée par le nom de la variable quantitative. Après avoir validé par un clic sur **OK**, on accède à un second menu où la rubrique **Panel Type** sera mise à la valeur **Condition**, et la rubrique **Column List** sera renseignée par le nom de la variable qualitative.

Cela permet de visualiser les variations de la même variable quantitative dans chacune des sous-populations définies par la variable qualitative.

On peut aussi visualiser une boîte à moustaches (box plot) pour chaque sous-population, par le traitement ► **Graph** ► **2D Plot** ► **Linear** ► **Box Plot(x, grouping-optional)** où la rubrique **x Columns** est la variable qualitative et la rubrique **y Columns** est la variable quantitative.

4.3 Cas de deux variables qualitatives

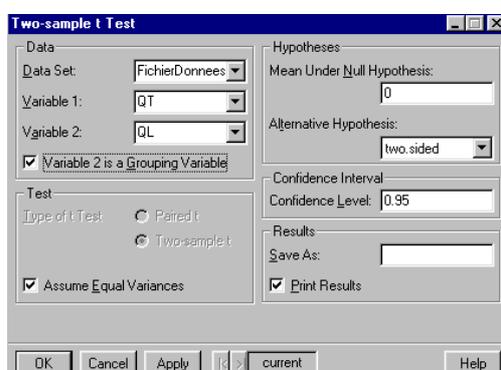
On calcule la distribution de fréquence conjointe de deux variables qualitatives X1 et X2 par le traitement ► **Statistics** ► **Data Summaries** ► **Crosstabulations...** La rubrique permet de sélectionner les variables X1 et X2 à croiser.

De même que précédemment, une des deux variables qualitatives, par exemple X1, engendre une partition de la population en sous-populations, chacune constituée des individus prenant la même modalité. On peut représenter, par exemple par un diagramme en barres, la distribution de fréquence de la deuxième variable qualitative, c'est-à-dire X2, pour chaque sous-population. Cela s'obtient par le traitement ► **Graph** ► **Multipanel Graph...** ► **Linear** ► **Bar with Base at Y Min(x,y1..yn)** qui fait accéder à un menu où la rubrique **y Columns** sera renseignée par le nom de la variable qualitative X2. Après avoir validé par un clic sur **OK**, on accède à un second menu où la rubrique **Panel Type** sera mise à la valeur **Condition**, et la rubrique **Column List** sera renseignée par le nom de la variable qualitative X1.

Cela permet donc de visualiser les variations de la même variable qualitative, X2, dans chacune des sous-populations définies par X1.

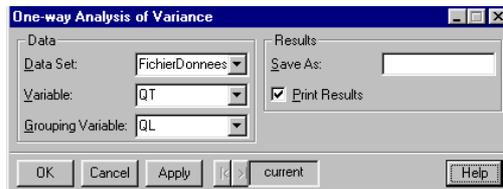
4.4 Comparaison de deux moyennes (test de Student)

Traitement : ► **Statistics** ► **Compare Samples** ► **Two Samples** ► **t Test...**



4.5 Comparaison de plusieurs moyennes

Traitement : ► **Statistics** ► **Compare Samples** ► **k Samples** ► **One Way Anova...**

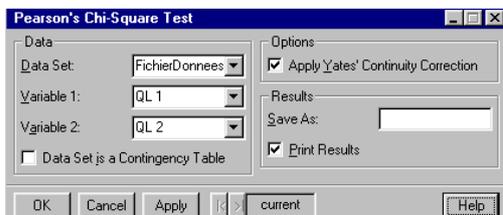


Comme l'indique le nom de ce traitement, c'est aussi l'ANOVA à un facteur, qui sera développée ci-après.

Si la variable qualitative est à deux modalités, ce test est équivalent au test de Student de comparaison de deux moyennes, du paragraphe précédent.

4.6 Comparaison de deux ou plusieurs proportions ou test du χ^2 d'indépendance de deux variables qualitatives

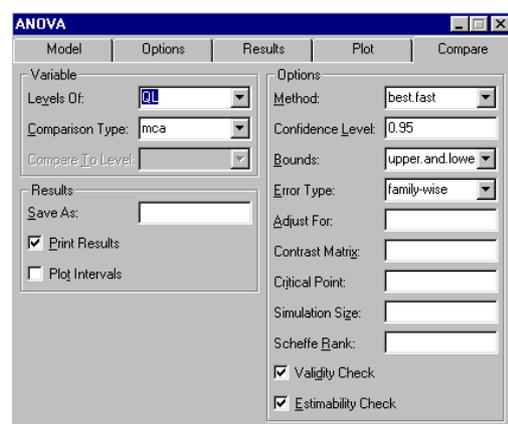
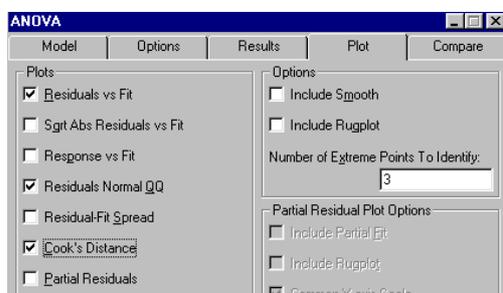
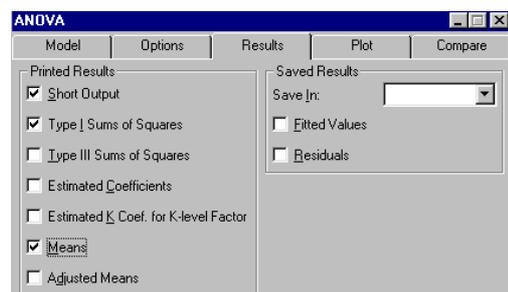
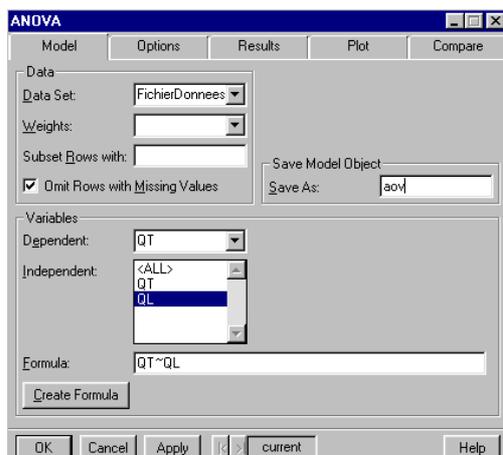
Traitement : ► **Statistics** ► **Compare Samples** ► **Counts and Proportions** ► **Chi square Test...**



Si les conditions d'application du test du χ^2 ne sont pas réunies, on peut mettre en œuvre le traitement : ► **Statistics** ► **Compare Samples** ► **Counts and Proportions** ► **Fisher's Exact Test...**

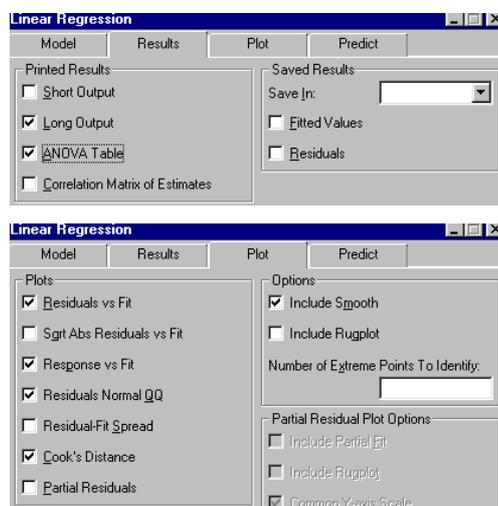
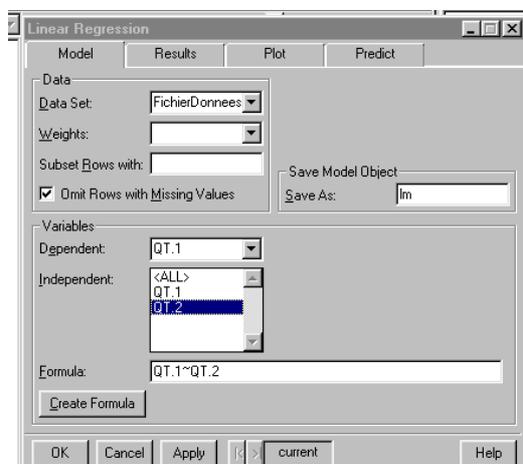
4.7 ANOVA à un facteur

Traitement : ► **Statistics** ► **ANOVA** ► **Fixed Effects...**



4. 8 Test d'indépendance de deux variables quantitatives et régression linéaire simple

Traitement : ► **Statistics** ► **Regression** ► **Linear...**



5 Traitements multivariés : analyse des données

Ces traitements font intervenir plus de deux colonnes (ou variables). On ne présente ici que quelques uns des traitements possibles sous S-plus. L'objectif est de présenter les méthodes étudiées dans le cadre des enseignements d'analyse des données de l'INH. Les traitements disponibles et les résultats fournis par S-plus peuvent être jugés insuffisants ; des méthodes (analyse des correspondances) et des résultats plus complets (outils d'aide à l'interprétation) sont disponibles en utilisant la librairie *multidim* qui sera développée dans les sections suivantes.

Les traitements multivariés tels que l'analyse de la variance à plusieurs facteurs ou la régression linéaire multiple, sont introduits en fin de document.

5. 1 Analyse en composantes principales

Le traitement ► **Statistics** ► **Multivariate** ► **Principal Components**, même s'il ne donne pas tous les éléments souhaitables, est néanmoins plus facile à mettre en œuvre que sa correspondante avec *multidim*, qui est décrite ci-dessous.

5. 2 Classification hiérarchique

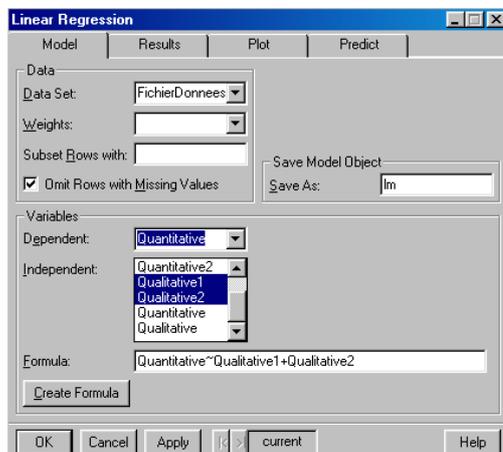
La classification ascendante hiérarchique s'obtient à l'aide du traitement ► **Statistics** ► **Cluster Analysis** ► **Agglomerative Hierarchical...** qui opère sur des variables quantitatives, éventuellement réduites. Il offre la possibilité de choisir entre plusieurs types de distances inter-groupes : distance min (**single**), distance max (**complete**), distance moyenne (**average**), distance de Ward (**ward**).

Ce traitement permet de visualiser la hiérarchie par un arbre et de sauvegarder la partition à nombre de classes voulu, sous forme de variable qualitative.

6 Autres traitements multivariés

6.1 Régression linéaire multiple

Traitement : ► **Statistics** ► **Regression** ► **Linear...**



C'est le même paramétrage que la régression linéaire simple. La page **Model** permet de préciser :

- la variable quantitative à expliquer dans la rubrique **Dependant** ;
- les variables quantitatives explicatives dans la rubrique **Independant**.

6.2 Analyse de la variance à deux facteurs ou plus

Traitement : ► **Statistics** ► **ANOVA** ► **Fixed Effects...**

C'est le même paramétrage que l'analyse de la variance à un facteur. La page **Model** permet de préciser la variable quantitative (dans la rubrique **Dependant**) et les facteurs dont on souhaite tester l'effet (dans la rubrique **Independant**).

La rubrique **Levels Of** de la page **Compare** permet de fixer le facteur pour lequel on veut obtenir les comparaisons des moyennes deux à deux.

7 Sauvegarder les traitements réalisés

On a vu (§1) que les données saisies peuvent être sauvegardées dans un fichier de données auquel S-Plus donne l'extension *.sdd*. De même, tous les résultats fournis par S-Plus peuvent être sauvegardés dans des fichiers.

Par défaut, les résultats numériques des traitements réalisés sont affichés à l'écran dans une fenêtre appelée **Report**, et les graphiques dans une fenêtre organisée en classeur, lequel classeur est constitué d'autant de pages que de graphiques demandés (Fig.7).

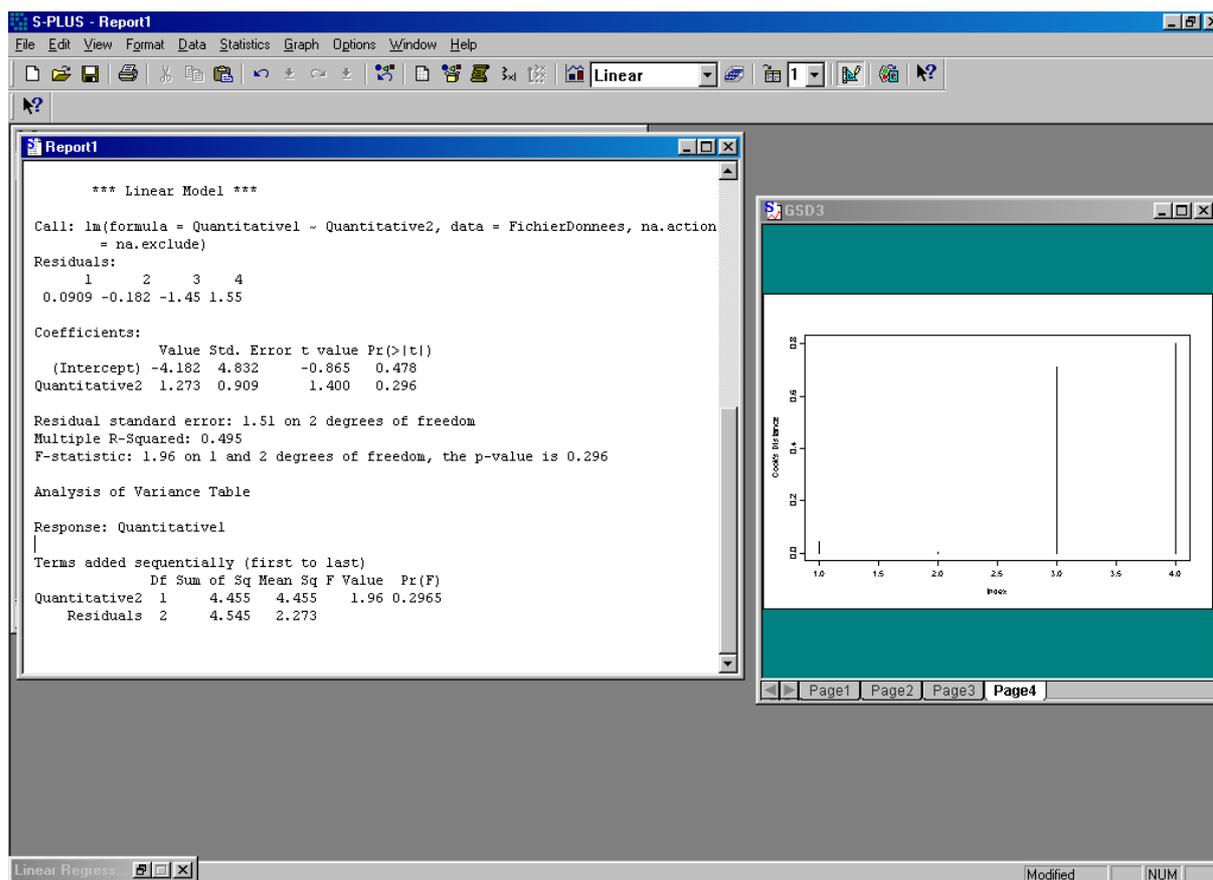


Fig.7

Les rapports et les graphiques peuvent être sauvegardés dans des fichiers auxquels S-Plus donne des extensions particulières (.srp pour un fichier rapport, .sgr pour un fichier graphique...). Ces fichiers ne sont accessibles qu'à partir de S-Plus.

On peut aussi sauvegarder ces résultats par un Copier-Coller dans un logiciel de traitement de texte comme Word (§ 11).

Certains traitements peuvent retourner des valeurs qu'on peut souhaiter récupérer sous forme de fichier de données ou éventuellement de variables (colonnes) ajoutées au tableau de données initial. Les rubriques du type **Save as** sont destinées à cet usage.

8 Rédiger / Imprimer un rapport

Pour rédiger un rapport, le plus simple est de procéder comme explicité au paragraphe précédent (Copier-Coller) avec le traitement de texte Word. Décrivons un exemple.

Exemple : Régression linéaire simple

Le fichier de données contient deux variables quantitatives dont les valeurs sont les suivantes.

Contenu de	Quantitative1 : 1 2 2 5
<i>FichierDonnees</i>	Quantitative2 : 4 5 6 6

Après application du traitement ► **Statistics** ► **Regression** ► **Linear...** on obtient dans la fenêtre **Report** les résultats qui par un Copier-Coller, donne ce qui suit sous Word. Ce texte peut ensuite être retravaillé pour en améliorer la présentation.

```
*** Linear Model ***
```

```
Call: lm(formula = Quantitatif1 ~ Quantitatif2, data = FichierDonnees, na.action
 = na.exclude)
```

```
Residuals:
    1     2     3     4
0.0909 -0.182 -1.45  1.55
```

```
Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -4.182   4.832    -0.865  0.478
Quantitatif2  1.273   0.909     1.400  0.296
```

Residual standard error: 1.51 on 2 degrees of freedom

Multiple R-Squared: 0.495

F-statistic: 1.96 on 1 and 2 degrees of freedom, the p-value is 0.296

3 observations deleted due to missing values

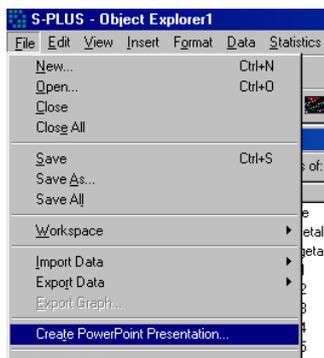
Analysis of Variance Table

Response: Quantitatif1

Terms added sequentially (first to last)

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
Quantitatif2	1	4.45	4.45	1.96	0.296
Residuals	2	4.55	2.27		

S-Plus permet aussi d'appeler directement Power Point pour la création de diapositives.



9 Exemple d'utilisation de *Commands*

On présente successivement les test suivants :

- Test de Fisher-Snedecor (comparaison de 2 variances),
- Test de Student (comparaison de deux moyennes),
- Test de Wilcoxon (comparaison des rangs moyens).

On suppose qu'on dispose d'un fichier de données appelé ici SDF1 qui comporte 2 colonnes V1 et V2 (les variables) et 13 lignes (les individus).

On accède à la fenêtre de commandes par clics de souris successifs sur ► **Window** ► **Commands Window** de la *barre des menus*.

The screenshot shows the S-PLUS interface with the following components:

- Object Explorer:** Shows the contents of 'D:\Splus6\users\RBoumaz\Essai0\SDF1'. It lists two objects: 'V1' (factor, position 1) and 'V2' (numeric, position 2).
- Data Window (SDF1):** A table with 15 rows and 4 columns. The first two columns are 'V1' and 'V2'. The data is as follows:

	1	2	3	4
	V1	V2		
1	a	1.00		
2	a	2.00		
3	a	3.00		
4	a	1.00		
5	a	2.00		
6	a	3.00		
7	a	4.00		
8	b	4.00		
9	b	5.00		
10	b	1.00		
11	b	2.00		
12	b	3.00		
13	b	2.00		
14				
15				
- Commands Window:** Shows the following commands and outputs:


```

> x <- SDF1$V2
> x1 <- x[SDF1$V1=="a"]
> x1
[1] 1 2 3 1 2 3 4
> x2 <- x[V1=="b"]
Problem: Object "V1" not found
Use traceback() to see the call stack
> x2 <- x[SDF1$V1=="b"]
> x2
[1] 4 5 1 2 3 2
> var.test(x1,x2)

      F test for variance equality

data: x1 and x2
F = 0.5714, num df = 6, denom df = 5, p-value = 0.5142
alternative hypothesis: true ratio of variances is not equal to
1
95 percent confidence interval:
 0.08189352 3.42146579
sample estimates:
variance of x variance of y
 1.238095      2.166667

> t.test(x1,x2)

      Standard Two-Sample t-Test

data: x1 and x2
t = -0.7639, df = 11, p-value = 0.461
alternative hypothesis: true difference in means is not equal t
o 0
95 percent confidence interval:
 -2.125379  1.030141
sample estimates:
mean of x mean of y
 2.285714  2.833333
      
```

La fenêtre de droite, *Commands*, permet d'entrer les commandes (ce qui est inscrit après le symbole supérieur > que présente S-Plus). Ces commandes sont exécutées dès qu'on les valide par la touche *Entrée* du clavier.

```
>x <- SDF1$V2
```

SDF1\$V2 est l'ensemble des valeurs de la colonne (la variable) V2 du fichier SDF1 ; ces valeurs sont affectées (par le symbole ->) au vecteur x.

Ainsi on désigne une colonne du fichier de données par le nom du fichier de données (ici SDF1), suivi du symbole dollar (\$), suivi du nom de la colonne (ici V2).

```
>x1 <- x[SDF1$V1=="a"]
```

Affecte à x1 les valeurs de x à la condition que SDF1\$V1 prenne la valeur a. Cela revient à isoler les valeurs correspondant au premier échantillon.

```
> x1
```

Demande d'affichage de x1.

```
[1] 1 2 3 1 2 3 4
```

Réponse de S-Plus à la demande d'affichage précédente

```
> x2 <- x[V1=="b"]
```

Censé affecter à x2 les valeurs de x à la condition que V1 prenne la valeur b.

Problem: Object "V1" not found

Use traceback() to see the call stack

S-Plus répond qu'il ne connaît pas l'objet V1, car on n'a pas précisé le nom du fichier de données.

```
> x2 <- x[SDF1$V1=="b"]
Correction de l'erreur précédente.
> x2
```

Demande d'affichage.

```
[1] 4 5 1 2 3 2
```

Réponse de S-Plus. Ce sont bien les valeurs correspondant au deuxième échantillon.

9.1 Test de Fisher – Snedecor de comparaison de deux variances

```
> var.test(x1,x2)
```

Demande de comparaison des variances de deux échantillons par le test de Fisher. La réponse de S-Plus est la suivante.

```
F test for variance equality
```

```
data: x1 and x2
```

```
F = 0.5714, num df = 6, denom df = 5, p-value = 0.5142
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
```

```
0.08189352 3.42146579
```

```
sample estimates:
```

```
variance of x variance of y
1.238095         2.166667
```

Par rapport à ce qui a été dit en cours, le f dit calculé n'est pas le maximum des deux rapports possibles. S-Plus prend au numérateur la première valeur de la commande `var.test(x1,x2)`, c'est-à-dire `x1`, et au dénominateur la deuxième valeur, c'est-à-dire `x2`.

En demandant de permuter ces valeurs, on obtient ce qui suit.

```
> var.test(x2,x1)
```

```
F test for variance equality
```

```
data: x2 and x1
```

```
F = 1.75, num df = 5, denom df = 6, p-value = 0.5142
```

```
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
```

```
0.2922724 12.2109783
```

```
sample estimates:
```

```
variance of x variance of y
2.166667         1.238095
```

On notera que la p-valeur est heureusement identique dans les deux cas.

9.2 Test de Student

```
> t.test(x1,x2)
```

```
Standard Two-Sample t-Test
```

```

data:  x1 and x2
t = -0.7639, df = 11, p-value = 0.461
alternative hypothesis: true difference in means is not equal
t
o 0
95 percent confidence interval:
 -2.125379  1.030141
sample estimates:
 mean of x mean of y
  2.285714  2.833333

```

9.3 Test de Wilcoxon

```
> wilcox.test(x1,x2)
```

Wilcoxon rank-sum test

```

data:  x1 and x2
rank-sum normal statistic with correction Z = -0.587, p-value
= 0.5572
alternative hypothesis: true mu is not equal to 0

```

Warning messages:

```
cannot compute exact p-value with ties in: wil.rank.sum(x,
y, alternative, exact, correct)
```

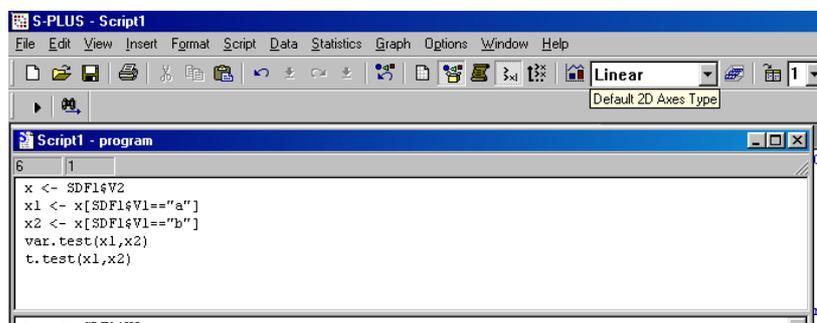
10 Exemple d'utilisation de *Script*

On présente les tests de Fisher-Snedecor et de Student, vus dans la section précédente traitant de l'utilisation de *Commands*). Ecrire un script c'est écrire une suite de commandes, un programme, puis on demande à S-Plus de l'exécuter.

On suppose donc qu'on dispose d'un fichier de données appelé ici SDF1 qui comporte 2 colonnes V1 et V2 (les variables) et 13 lignes (les individus).

On accède à la fenêtre de scripts, désignée par **Script1 – program**, par clics de souris successifs sur ► **File** ► **New** ► **Script File** de la *barre des menus*.

On introduit les commandes décrites au paragraphe précédent comme suit.



```

S-PLUS - Script1
File Edit View Insert Format Script Data Statistics Graph Options Window Help
Linear
Default 2D Axes Type
Script1 - program
6 1
x <- SDF1$V2
x1 <- x[SDF1$V1=="a"]
x2 <- x[SDF1$V1=="b"]
var.test(x1,x2)
t.test(x1,x2)

```

Puis on demande l'exécution de ces commandes par clics de souris sur ► **Script** ► **Run** de la *barre des menus*. S-Plus exécute les commandes l'une après l'autre. S-Plus affiche les résultats après chaque commande dans une fenêtre, désignée par **Script1 – output** (Fig.8).

```

S-PLUS - Script1
File Edit View Insert Format Script Data Statistics Graph Options Window Help
[Icons] Linear 1
Script1 - output
x1 <- x[SDF1$V1=="a"]
x2 <- x[SDF1$V1=="b"]
var.test(x1,x2)
t.test(x1,x2)

> x <- SDF1$V2
> x1 <- x[SDF1$V1 == "a"]
> x2 <- x[SDF1$V1 == "b"]
> var.test(x1, x2)

      F test for variance equality

data:  x1 and x2
F = 0.5714, num df = 6, denom df = 5, p-value = 0.5142
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.08189352 3.42146579
sample estimates:
variance of x variance of y
 1.238095    2.166667

> t.test(x1, x2)

      Standard Two-Sample t-Test

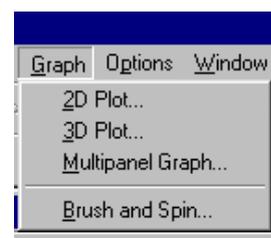
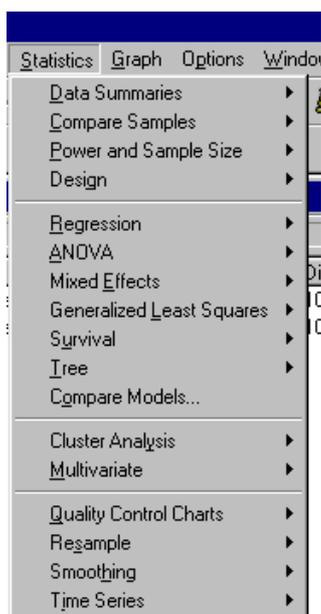
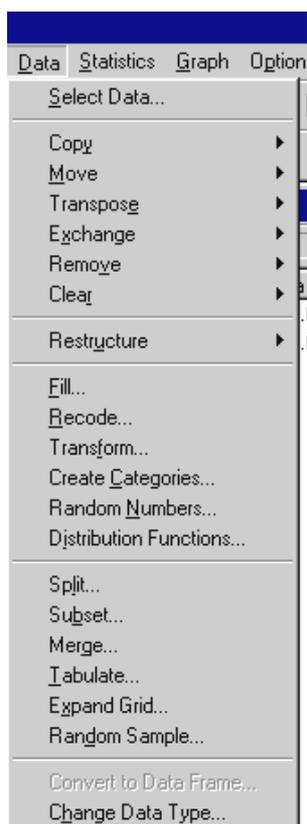
data:  x1 and x2
t = -0.7639, df = 11, p-value = 0.461
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-2.125379  1.030141
sample estimates:
mean of x mean of y
 2.285714  2.833333

```

Fig.9

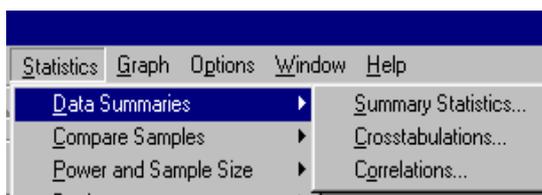
ANNEXE

Data	Gestion des données
Statistics	Calculs statistiques
Graph	Graphiques



Data Summaries

Moyenne, limites de confiance, variance, quartiles...
 Distributions conjointe, marginales, conditionnelles
 Coefficient de corrélation



Compare samples

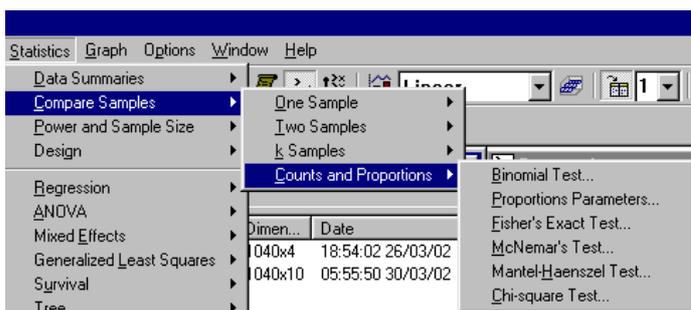
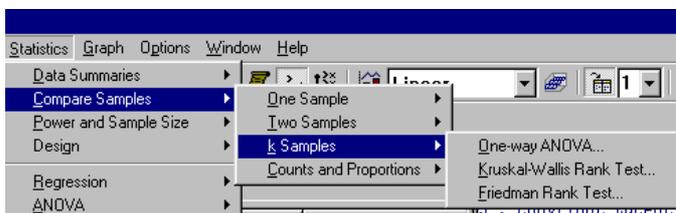
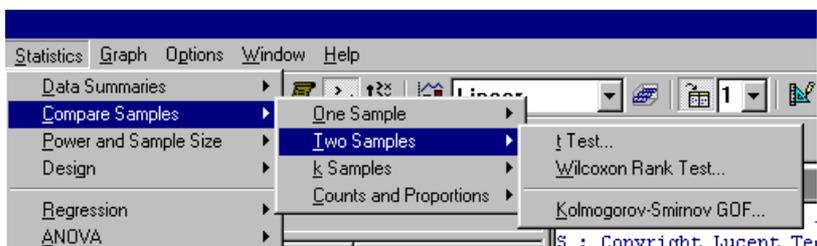
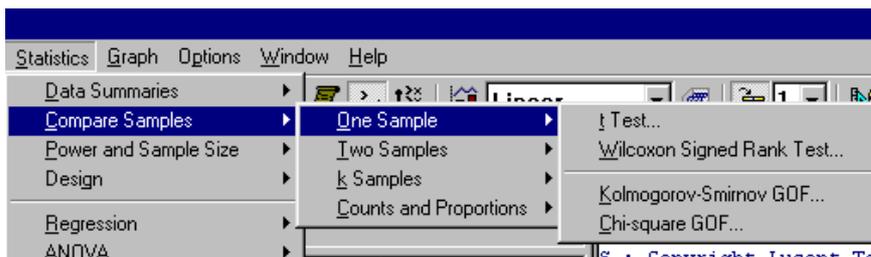
Tests de comparaison à une norme

Tests d'ajustement

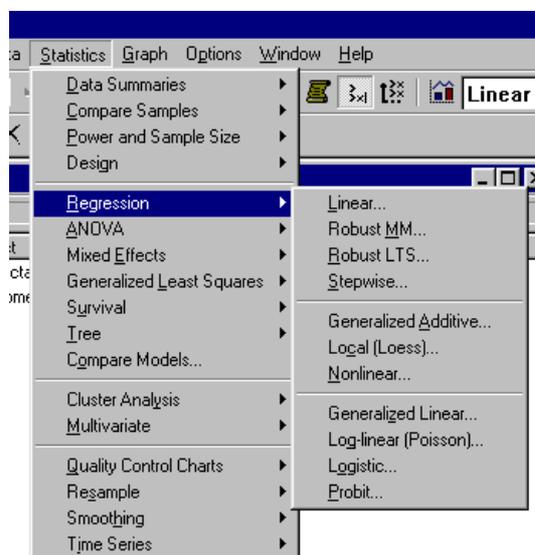
Test de comparaison de deux échantillons (Student, Wilcoxon)

Tests de comparaison de plusieurs échantillons (ANOVA à un facteur, Kruskal-Wallis, Friedman...)

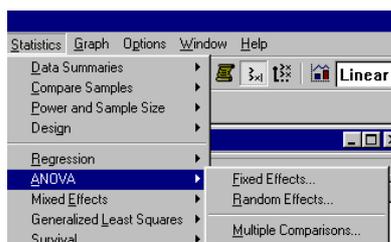
Tests portant sur des proportions (test du khi-deux d'indépendance, test exact de Fisher...)



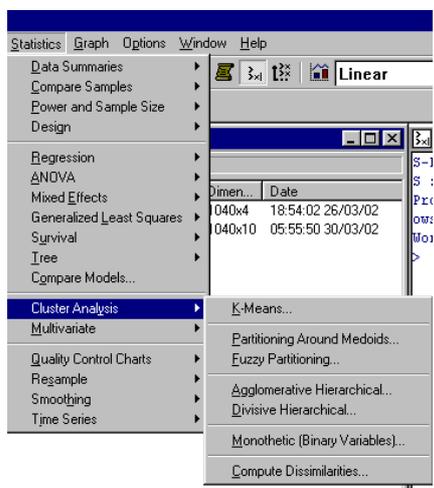
Régression linéaire simple
 Régression linéaire multiple, pas à pas...



Analyse de la variance à un facteur, à deux facteurs et plus.
 Analyse de la variance à effets aléatoires
 Comparaisons multiples (Scheffe, LSD...)



Classification automatique



Analyses factorielles (analyse en composantes principales, analyse discriminante...)
Analyse de la variance multiple

