

# **R Commander : Petit guide pratique**

## **1. Statistiques de base**

Jean-philippe Gaudron

Février 2016

Ce petit guide pratique a été écrit pour faciliter l'utilisation de R Commander. A partir de courts exercices et de questions de recherche simples, il permet d'apprendre à décrire des données univariées et bivariées et à les analyser avec ce logiciel simple, puissant et gratuit. La plupart des chapitres portent sur un test en particulier et peuvent donc être travaillés indépendamment les uns des autres. Mais tous nécessitent de connaître l'interface de R Commander (chapitre Prise en main) et de savoir comment construire ou importer un tableau de données (chapitre Construire un tableau de données...) avec R Commander.

Ce guide a été écrit en utilisant la version 2.2.3 de R Commander pour Windows. Mais compte tenu des faibles différences entre les versions pour Mac, LINUX (Ubuntu) et Windows (le chapitre Prise en main présente l'interface de R Commander sous les trois environnements), il peut être utilisé pour les trois. A ce titre, les annexes présentent les procédures de chargement et d'installation de R et de R Commander sous Windows, Mac et Ubuntu.

Enfin et surtout, ce petit guide pratique n'a pas vocation à se substituer à un manuel d'introduction à la statistique. Au contraire, il part de principe que les lecteurs et lectrices connaissent les statistiques descriptives ainsi que les tests d'inférence statistique usuels, leurs conditions d'utilisation et surtout leur pertinence par rapport aux questions de recherche posées.

Ce document est protégé par les droits d'auteur. Il peut être téléchargé et utilisé gratuitement. Toute modification, même partielle, n'est autorisée qu'avec l'accord écrit de l'auteur.

Pr Jean-Philippe Gaudron  
Département des Sciences de l'éducation et de la formation  
Université Toulouse Jean Jaurès  
F-31058 Toulouse  
France  
[gaudron@univ-tlse2.fr](mailto:gaudron@univ-tlse2.fr)

## Table des matières

1.	R Commander : découverte et prise en main .....	5
1.1	Démarrer R Commander .....	5
1.2	Faire des calculs avec R.....	8
1.3	Copier, enregistrer les résultats et fermer R Commander.....	9
2	Construire un tableau de données, spécifier les variables et modifier les valeurs .....	10
2.1	Construire un tableau de données avec R Commander .....	10
2.2	Enregistrer un tableau de données .....	13
2.3	Importer un jeu de données .....	14
2.4	Ouvrir un fichier de données déjà enregistré sous R.....	19
2.5	Coder les variables.....	20
2.6	Calculer de nouvelles variables .....	22
2.7	Découper une variable numérique en classes .....	23
2.8	Générer des jeux de données partiels .....	24
3	Statistiques descriptives .....	27
3.1	Statistiques descriptives numériques .....	28
3.1.1	Statistiques descriptives par variable .....	28
3.1.2	Statistiques descriptives par modalité de variable (sous-groupes).....	30
3.2	Construire des graphiques pour décrire des données.....	33
3.2.1	Dessiner un histogramme .....	33
3.2.2	Dessiner un graphe Tiges et feuilles .....	34
3.2.3	Dessiner un graphe quantile-quantile :.....	36
3.2.4	Dessiner une boîte à moustaches.....	37
3.2.5	Dessiner un graphe des moyennes .....	39
3.2.6	Dessiner un graphe en camembert .....	41
3.2.7	Dessiner un graphe en barres .....	42
4	Tableaux croisés et Chi-deux d'indépendance ( <i>Chi-square test of independance</i> ) .....	44
4.1	Calculer un Chi-deux à partir d'un tableau croisé.....	44
4.2	Calculer un Chi-deux sur un jeu de données .....	46
5	La corrélation avec le r de Bravais-Pearson ( <i>correlation with Pearson's r</i> ).....	48
5.1	Vérifier la nature linéaire de la relation entre les deux observations Avant et Après.....	49
5.2	Vérifier que les variables Avant et Après se distribuent normalement .....	50
5.3	Calculer la corrélation entre les deux variables.....	52
5.4	Calcul de la corrélation pour sous-groupes .....	52

5.4.1	Vérifier la nature linéaire de la relation entre Avant et Après pour le sous-groupe des filles .....	53
5.4.2	Vérifier la normalité des deux distributions des scores Avant et Après pour ce sous-groupe des filles .....	54
5.4.3	Calcul de la corrélation pour un sous-groupe.....	57
6	La corrélation avec le rho de Spearman ( <i>correlational analysis : Spearman's rho</i> ) .....	60
6.1	Vérifier la nature linéaire de la relation en observant le nuage de points .....	61
6.2	Vérifier que les variables Temps et Performance se distribuent normalement .....	62
6.3	Calculer la corrélation entre les deux variables Temps et Performance.....	63
7	Le test <i>t</i> de Student pour un échantillon ou test <i>t</i> univarié ( <i>one sample t-test</i> ) .....	64
8	Le test <i>t</i> de Student pour groupes indépendants ( <i>independent sample t-test</i> ) .....	66
8.1	Vérifier la normalité des groupes .....	67
8.2	Apprécier l'égalité des variances des 2 groupes.....	69
8.3	Calcul du <i>t</i> de Student pour groupes indépendants.....	70
9	Le test <i>U</i> de Mann-Withney (ou test de la somme des rangs de Wilcoxon) pour groupes indépendants ( <i>Mann-Withney U-test or Wilcoxon sum rank test</i> ).....	71
10	Le test <i>t</i> de Student pour groupes appariés ( <i>paired sample, ou related t-test</i> ).....	73
10.1	Vérifier la normalité des 2 groupes.....	73
10.2	Calcul du test <i>t</i> pour groupes appariés .....	75
10.3	Calcul du test <i>t</i> de Student pour sous-groupes appariés.....	76
10.3.1	Vérifier la normalité d'un sous-groupe .....	76
10.3.2	Calcul du test <i>t</i> pour deux sous-groupes appariés .....	79
11	Le test du signe de Wilcoxon pour groupes appariés ( <i>Wilcoxon signed ranks test</i> ) .....	81
12	ANOVA pour groupes indépendants ( <i>One-Way between-groups</i> ) .....	83
12.1	Tester la normalité des groupes .....	84
12.2	Tester l'égalité des variances des 3 groupes .....	86
12.3	Calcul de l'ANOVA pour groupes indépendants .....	87
12.4	Comparaisons <i>a posteriori</i> .....	88
13	La régression linéaire simple ( <i>Simple Regression</i> ) .....	89
13.1	Vérifier la nature linéaire de la relation entre <i>X</i> et <i>Y</i> en observant le nuage de points	90
13.2	Calculer la corrélation entre les deux variables Temps et Performance .....	91
13.3	La droite de regression.....	92
13.4	Les tests d'hypothèse : .....	93
13.4.1	Vérifier les conditions d'homogénéité des variances et de normalité dans les vecteurs en observant les graphiques .....	94

13.4.2	Pour aller plus loin .....	95
14	Références .....	98
15	Annexes .....	99
15.1	Installer R et R Commander sous Windows .....	100
15.2	Installer R et R Commander sous Mac .....	108
15.3	Installer R et R Commander sous Ubuntu (Linux) .....	124

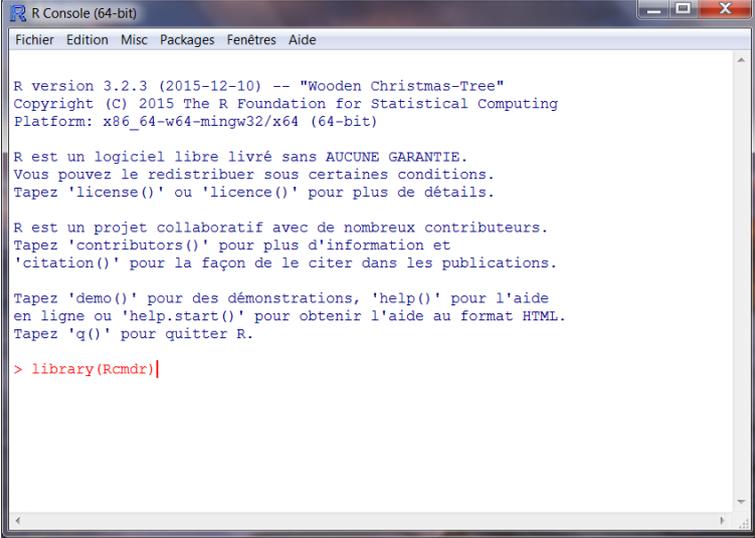
## 1. R Commander : découverte et prise en main

R Commander (Fox, 2005), ou encore Rcmdr est une interface graphique (*Graphical User Interface*, GUI) du logiciel libre R. R Commander permet de calculer des statistiques de base descriptives et inférentielles à partir de menus déroulants, sans avoir donc à écrire les commandes et maîtriser le langage R. R Commander repose sur la souplesse et la puissance de calcul de R et permet de construire des graphiques, autre point fort de R. Mais R Commander n'est que la partie visible de l'iceberg R, c'est à dire que c'est un paquet (*package*) parmi les centaines de paquets élaborés et offerts par la communauté des chercheurs pour enrichir les fonctionnalités de R, ou pour proposer des méthodes particulières. Pour autant, sa simplicité et son aspect intuitif permettent aux chercheur.e.s de répondre à nombre de questions quantitatives qu'elles et ils se posent ; et petit à petit, de découvrir les potentialités de R.

### 1.1 Démarrer R Commander

Pour démarrer R Commander, il faut tout d'abord démarrer R en double cliquant sur l'icône R se trouvant sur le bureau. La fenêtre R s'ouvre avec plusieurs paragraphes en bleu donnant quelques informations sur R (voir ci-dessous). Sous le texte, il y a le symbole rouge `>`, appelé prompt. C'est après ce symbole qu'il faut taper les lignes de commande de R qui s'affichent également en rouge.

Pour lancer R Commander, après le `>` il faut taper `library(Rcmdr)` et entrer.



```
R Console (64-bit)
Fichier Edition Misc Packages Fenêtres Aide

R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> library(Rcmdr)|
```

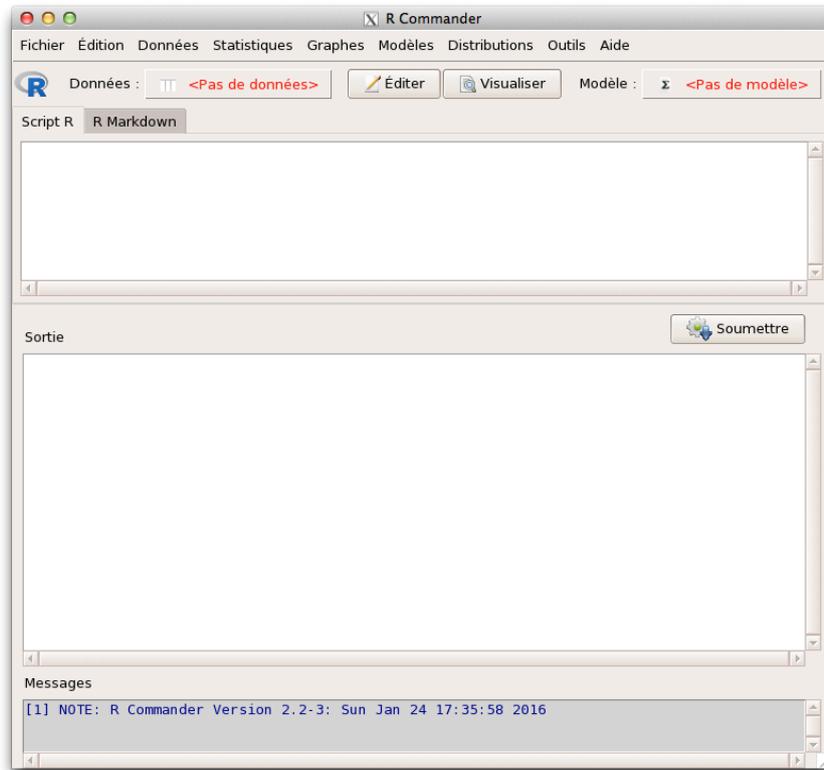
Si la commande est incomplète, le symbole `>` est remplacé par `+`

Si la commande est fautive (par exemple, la majuscule de Rcmdr a été oubliée), un message d'erreur s'affiche. R distingue donc majuscules et minuscules. Ajoutons que les `é`, `è`, `î`,... posent problème : il faut éviter tous les accents, même dans les noms des colonnes des fichiers importés. Dernière précision typographique : les décimales étant en anglais, il faut mettre des points à la place des virgules.

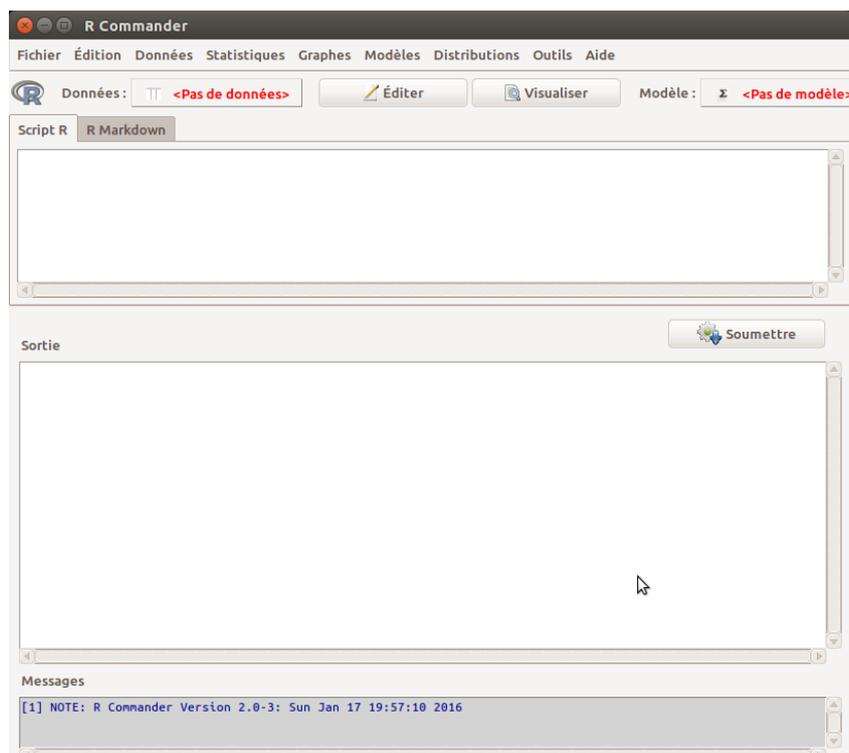
Lorsque R Commander est chargé, la fenêtre s'affiche à l'écran. Comme toute fenêtre, elle reproduit le graphique habituel de l'environnement de l'ordinateur, Windows, Mac ou Ubuntu (Linux). Pour autant, l'arborescence des menus est identique (il y a les mêmes options, les mêmes commandes) et les algorithmes de calculs sont les mêmes. Par la suite, nous utiliserons

le graphique de Windows, mais on aura compris que les procédures, les étapes sont identiques sous Mac ou sous Ubuntu. Ci-dessous sont présentées les fenêtres R Commander sous les différents environnements.

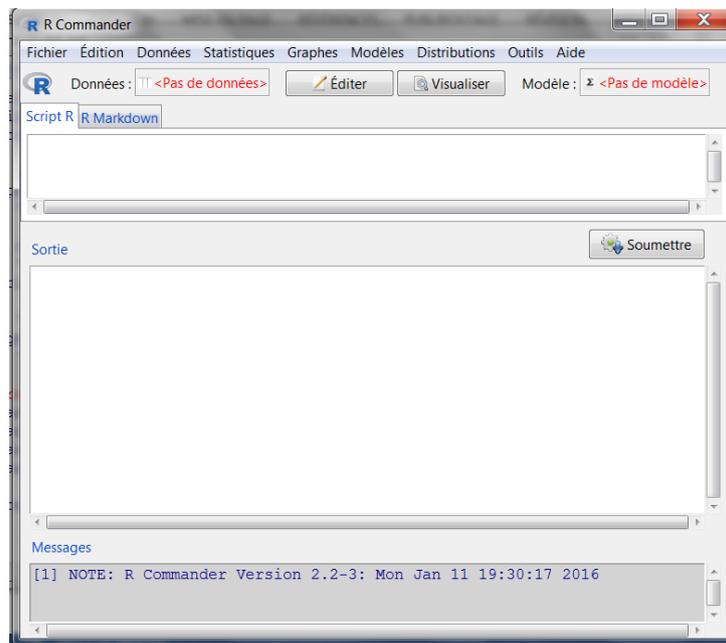
Pour Mac :



Pour Ubuntu :



Pour Windows :



Examinons la fenêtre R Commander

On peut l'agrandir en glissant la souris à partir des bords ou en cliquant le bouton d'agrandissement.

Tout en haut, le nom du programme R Commander et les 3 icônes pour masquer, agrandir/réduire, ou fermer.

En dessous, la barre « menu », avec Fichier, Édition, Données... En cliquant sur un mot, apparaît un menu déroulant à partir duquel on choisit les actions ou les analyses, on le verra plus loin.

Sous la barre des menus, quatre touches : Pas de données (lorsque l'on travaillera avec un jeu de données, son titre apparaîtra là ; Éditer ; Visualiser ; Pas de Modèle. Ce sont des touches raccourcis et que l'on retrouve dans les menus déroulants

Une première petite fenêtre blanche sous les deux onglets, script R et Markdown : c'est la fenêtre des commandes dans le langage de programmation de R, appelées scripts. Chaque clic de la souris sur une icône ou une commande d'un menu déroulant correspond à une programmation et y fera apparaître un script. On trouve d'ailleurs le même principe dans certains logiciels de statistiques tel SPSS avec la possibilité de visualiser le langage de programmation dans une fenêtre spécifique appelée Éditeur de syntaxe. Simplement, dans R Commander, 'l'éditeur de syntaxe' est sous les yeux en permanence, ce qui est une bonne façon de connaître le langage de R. D'ailleurs, lorsque nous souhaiterons programmer une analyse un peu spécifique, il suffira d'écrire le script dans la fenêtre, le sélectionner (le surligner avec la souris), puis de cliquer sur Soumettre. Et parfois, pour aller plus vite lorsqu'il s'agira de procéder à des traitements répétitifs, il suffira de reprendre le script, de le modifier à sa guise, de le sélectionner et de soumettre.

Notons que la fenêtre des scripts de R Commander est équivalente à la fenêtre de R : une commande lancée sous R Commander est reconnue par R et inversement.

La grande fenêtre en dessous est la fenêtre des sorties : c'est là que nous lirons les statistiques demandées (les commandes sont en rouge), c'est-à-dire les résultats (qui sont en bleu).

Tout en bas, une troisième petite fenêtre pour des messages, qui peuvent apparaître sous la forme d'**avis** (en vert), de **notes** (en bleu) ou d'**erreurs** (en rouge).

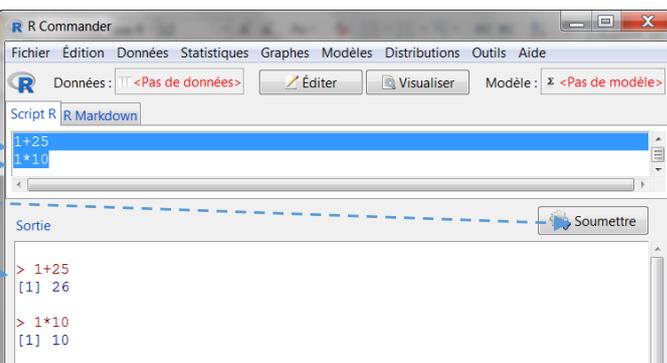
## 1.2 Faire des calculs avec R

R est avant tout un calculateur, donc R Commander aussi puisque les deux fenêtres sont équivalentes. Le tableau ci-dessous présente quelques opérations élémentaires possibles. Il suffit écrire l'opération dans la fenêtre commande, de surligner et de soumettre (dans R, il suffit d'écrire l'opération après le > et entrer) :

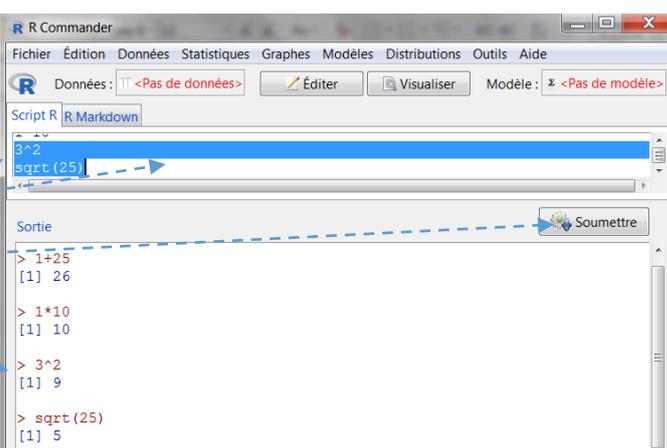
Opération	Signe	Exemple
Addition	$x + y$	1 + 25
Soustraction	$x - y$	5 - 1
Multiplication	$x * y$	1*10
Division	$x / y$	34/63
Elever à la puissance	$x ^ y$	3^2
Racine carrée	sqrt(x)	sqrt(25)
Transformation logarithmique	log10(x) ou log(x,10)	Log10(7)

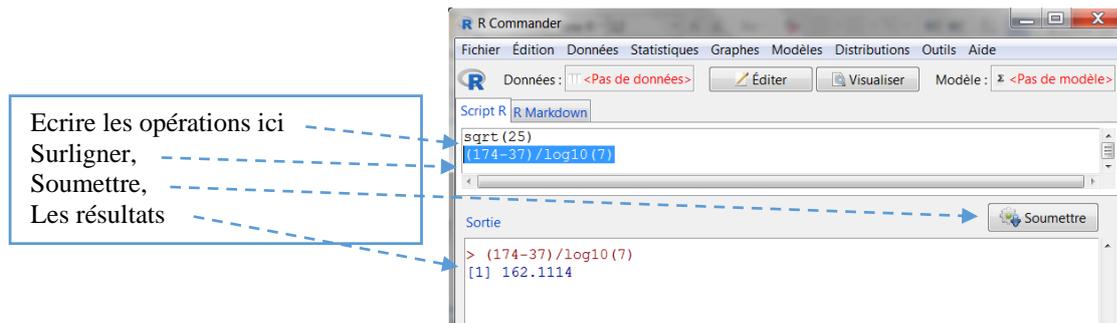
Faisons quelques opérations :

Ecrire les opérations ici  
Surligner,  
Soumettre,  
Les résultats  
[1] signifie que 26 est le premier élément de la ligne



Ecrire les opérations ici  
Surligner,  
Soumettre,  
Les résultats





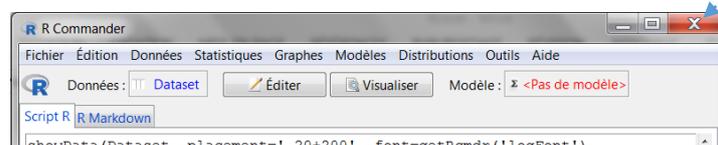
### 1.3 Copier, enregistrer les résultats et fermer R Commander

On peut rapidement copier ce qui apparaît dans la fenêtre des sorties, tout simplement en sélectionnant l'extrait que l'on souhaite copier avec la souris, puis Copier (bouton droit). Dans le traitement de texte, cliquer Coller : les résultats sont là.

Pour les graphiques, les images, l'option Copier les images est proposée dans le menu Editer ou en cliquant sur bouton droit de la souris. Il suffit ensuite de coller là où l'on souhaite, ou encore de l'enregistrer comme fichier

Avant de fermer R Commander, il ne faut pas oublier d'enregistrer le fichier actif au risque de perdre des modifications apportées.

Pour fermer R Commander, à moins de passer par le menu déroulant, Fichier → Sortir, il suffit de cliquer sur la croix rouge en haut à droite de la fenêtre de R Commander :

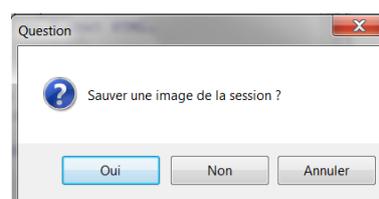


Quatre fenêtres se suivent, demandant la confirmation de la fermeture et l'enregistrement ou non le fichier script (pour réutiliser une autre fois les commandes utilisées), les scripts avec R Markdown, enfin les sorties (pour les conserver).



La fenêtre R Commander se ferme.

Il faut maintenant fermer R en cliquant également sur la croix rouge en haut à droite et répondre à une dernière fenêtre :



Notons que si l'on a fermé R Commander et que l'on souhaite le ré-ouvrir dans la même session de R, il faut écrire en face du prompt, la commande : `Commander()`

## 2 Construire un tableau de données, spécifier les variables et modifier les valeurs

Si l'on utilise des statistiques pour résoudre des problèmes, c'est que l'on dispose d'un grand nombre de données numériques. Pour effectuer les statistiques souhaitées, il est obligatoire de présenter ces données sous la forme d'un tableau dans lequel chaque colonne est une variable et chaque ligne un individu (individu au sens statistique du terme, c'est-à-dire une personne, une chose, une famille, un phénomène..., bref, tout ce à quoi s'intéresse la statistique). Chaque case, appelée cellule, est la valeur donnée à un individu (ou sa réponse s'il s'agit par exemple d'un questionnaire) à une variable (par exemple à une question).

Dans ce genre de tableau, la toute première ligne est réservée aux noms des colonnes, c'est à dire au nom des variables (ne pas mettre d'accent et d'espace). La première colonne est généralement réservée aux noms des lignes, c'est à dire les noms des individus qui sont souvent les numéros attribués aux individus, aux protocoles.

Un tel tableau des données peut-être très grand puisqu'il comporte autant de colonnes que de variables (dans un sondage, chaque question est a minima une variable) et autant de lignes que d'individus ayant participé. R Commander permet de construire des tableaux comportant moins de 10 000 valeurs. On peut aussi importer un tableau de données élaboré dans un tableur (Open Office, Excell...) ou dans d'autres logiciels de statistiques tels SPSS, SAS. Ces tableaux peuvent avoir plus de 10 000 valeurs.

Une pratique qui accélère la saisie consiste à remplacer les mots des modalités d'une variable par un chiffre (par exemple 1 pour les filles et 2 pour les garçons, ou encore les chiffres associés aux catégories socio-professionnelles, aux années d'études...). Ces chiffres ne font que nommer ces modalités. Nous verrons ci-dessous comment leur attribuer de nouveau leur nom (coder les variables).

Enfin, pour éviter les erreurs de saisie et pour aller plus vite, on entre les valeurs telles qu'elles apparaissent dans les protocoles. S'il faut faire la somme à plusieurs réponses, nous le ferons plus tard ; s'il faut inverser les réponses à une question négative, nous le ferons pareillement plus tard ; plus tard également s'il faut regrouper des scores en deux ou trois classes. En fait, ce sont des fonctions ou des calculs que permet de faire automatiquement R Commander lorsque que l'on souhaite modifier une variable, ce que nous verrons bientôt (modifier les variables).

### 2.1 Construire un tableau de données avec R Commander

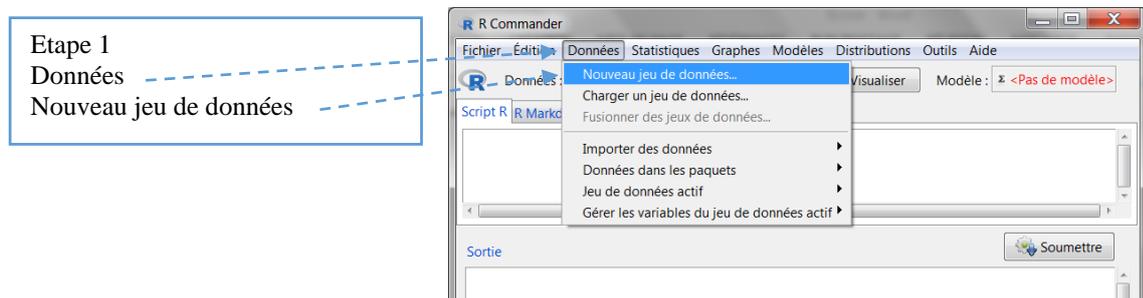
Construire un tableau de données avec R Commander se fait avec l'éditeur de données. Cet éditeur de données présente des fonctions semblables à celles des tableurs : il permet de créer un nouveau tableau de données, de modifier les données d'un tableau existant, d'ajouter ou de supprimer des lignes (*row*) et des colonnes (*column*) au tableau, de copier (*copy*), coller (*paste*) ou couper (*cut*) le contenu des cellules (*cell*). Cependant, il est moins ergonomique, moins rapide pour la saisie des données, il offre moins de possibilités qu'un vrai tableur et il est limité à 10 000 valeurs rappelons-le. Bref, il s'avère pratique pour de petits jeux de données et de petits exercices, sans plus.

Voici à titre d'exercice un petit jeu de données recueillies par une enseignante qui a relevé les notes à une évaluation passée par 6 élèves, trois filles et trois garçons. Comme il a été précisé plus haut, pour effectuer des statistiques, il est nécessaire de présenter les données sous la forme d'un tableau dans lequel chaque colonne est une variable et chaque ligne un individu et le nom des variables en haut des colonnes. Il y a ici trois variables, les élèves, leur sexe (codé ici 1 pour

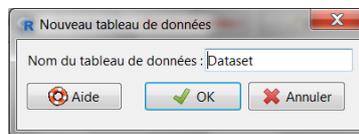
les filles et 2 pour les garçons) et leur note. Et il y a 6 individus, ici 6 élèves : 6 lignes donc, plus la première ligne pour les noms des variables. Voici le tableau que nous allons construire :

Elève	Sexe	Note
1	1	12,5
2	2	9
3	1	17
4	2	15,5
5	2	12
6	1	10

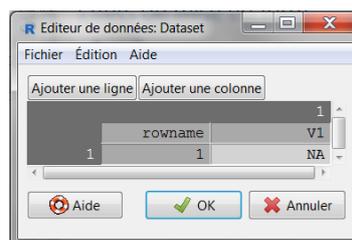
A partir de la fenêtre R Commander :



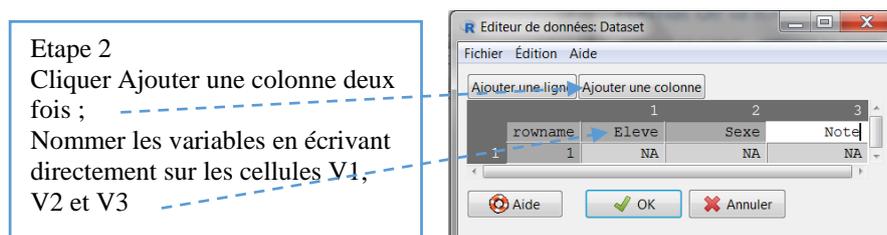
La fenêtre Nouveau tableau de données s'ouvre permettant de donner un nom. Laissons le nom par défaut (Dataset) et OK :



La fenêtre intitulée 'Editeur de données' s'ouvre avec le nom choisi (ici par défaut, Dataset).



Le plus pratique est de commencer par créer le tableau vierge, c'est-à-dire de créer le nombre de colonnes désirées. On ajoute une colonne en cliquant sur l'onglet 'Ajouter une colonne'. Puis on nomme la colonne.



Lorsque les colonnes sont prêtes, on peut passer aux lignes, c'est-à-dire que l'on remplit le tableau individu par individu, c'est-à-dire ligne par ligne. Pour remplir le tableau, il suffit de cliquer dans une cellule et de mettre la valeur correspondante qui s'affiche à la place de NA. NA signifie valeur manquante : s'il y en a, il suffira de sauter la cellule. Comme on le verra au cours des différentes parties, R Commander gère ces valeurs manquantes en proposant par exemple d'en tenir compte, ou pas, dans certains calculs ; ou encore en proposant une option dans laquelle, les lignes comportant des valeurs manquantes sont supprimées lors de la création de nouveaux fichiers avec une partie des données. On passe d'une cellule à l'autre à l'aide des flèches du clavier. Arrivé au bout de la première ligne, on ajoute une seconde ligne en cliquant sur l'onglet Ajouter une ligne etc.

Attention, rappelons que dans la numérotation anglaise, les décimales s'écrivent avec un point à la place de la virgule : par exemple ici, à la place de 12,5, il faut entrer 12.5.

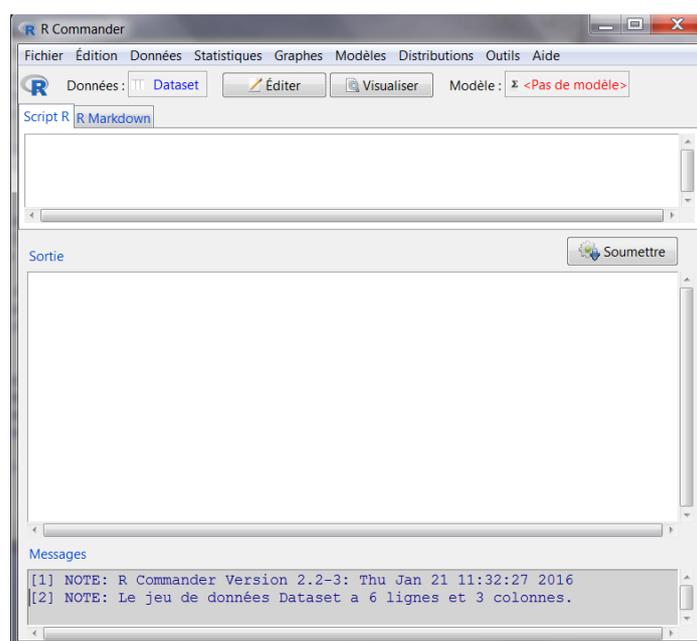
Voici le tableau construit :

Etape 2  
 Cliquer Ajouter une ligne pour créer l'individu suivant ;  
 Mettre les valeurs dans les cellules ;  
 Les virgules ont été remplacées par des points  
 OK

rowname	Elevé	Sexe	Note
1	1	1	12.5
2	2	2	9
3	3	3	17
4	4	2	15.5
5	5	2	12
6	6	1	10

Comme il n'y a pas de donnée manquante, le tableau final ne comporte aucune cellule avec NA.

Lorsque la saisie est terminée, appuyez sur OK. La fenêtre 'Editeur de données' disparaît et laisse place à la fenêtre R Commander. En haut, après 'Données', le nom du tableau 'Dataset' apparaît : c'est le jeu de données actif. Tout en bas, une note bleue signale qu'un jeu de données est actif, avec son nom (Dataset) et les nombres de lignes et de colonnes.



On peut visualiser le tableau sans modification possible, en cliquant sur le bouton Visualiser :

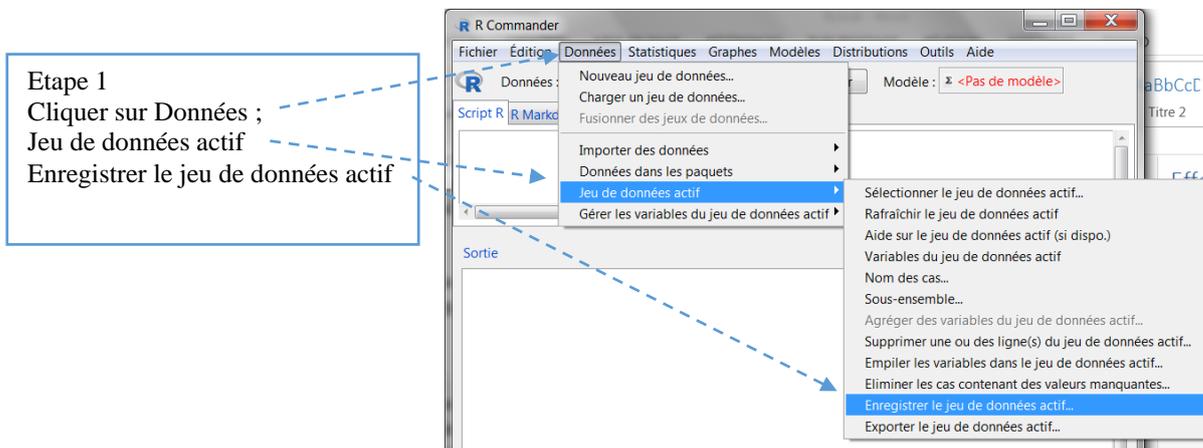
	Elève	Sexe	Note
1	1	1	12.5
2	2	2	9.0
3	3	1	17.0
4	4	2	15.5
5	5	2	12.0
6	6	1	10.0

Pour retirer la fenêtre, cliquer sur la croix rouge.

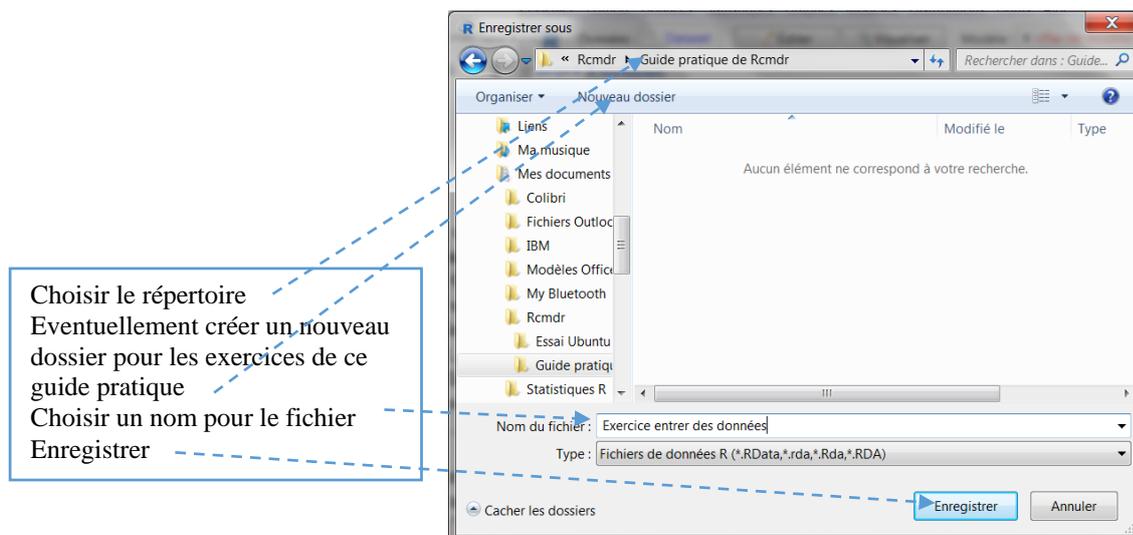
On peut l'éditer pour procéder à des modifications, à l'ajout de nouvelles données ou du nom des variables ou des modalités en cliquant sur le bouton Editer. Et pour retirer cette fenêtre, cliquer sur OK.

## 2.2 Enregistrer un tableau de données

Il est temps d'enregistrer ce fichier actif.



La fenêtre Enregistrer sous apparaît :



### 2.3 Importer un jeu de données

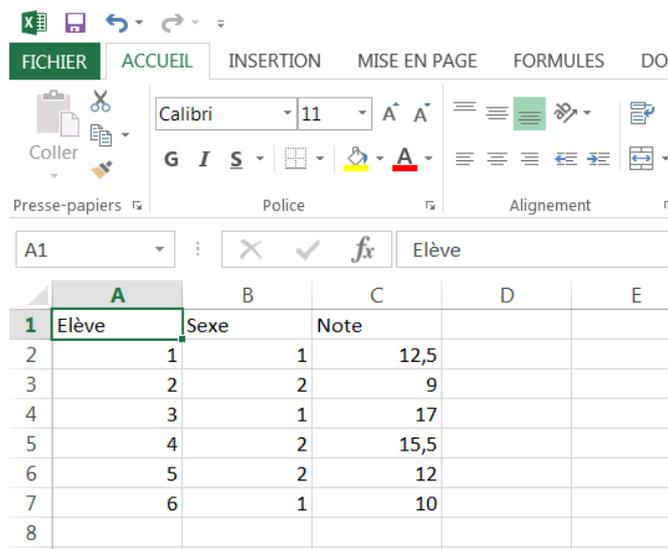
Comme beaucoup de logiciels de statistiques, R Commander peut travailler à partir de tableaux construits sur des tableurs (Excel, Open Office...) ou d'autres logiciels (SPSS, SAS, Minitab et STATA...). Il offre à ce titre une très grande souplesse et certains paquets (*packages*) sont dédiés à cela. On peut ainsi choisir le logiciel parmi une liste proposée et on se laisse ensuite guider et répondre aux questions posées quand il y en a. Ainsi, si l'on a un tableau construit dans l'un de ces logiciels, il suffit d'aller sur Données dans la barre de menu, puis Importer un jeu de données et de choisir le logiciel utilisé.

Si l'on a un jeu de données dans une autre logiciel que ceux proposés, ou si la démarche ci-dessus n'aboutit pas (ce qui arrive pour les nouvelles versions de ces logiciels qui ne sont pas encore intégrées dans la version présente de R Commander), la procédure reste néanmoins extrêmement simple, ce que nous allons faire ci-dessous.

Nous allons construire un tableau de données dans le tableur Excel, sachant que la procédure est identique pour le tableur gratuit de la suite Open Office. Nous utiliserons le même petit jeu de données que ci-dessus :

	Elève	Sexe	Note
	1	1	12,5
	2	2	9
	3	1	17
	4	2	15,5
	5	2	12
	6	1	10

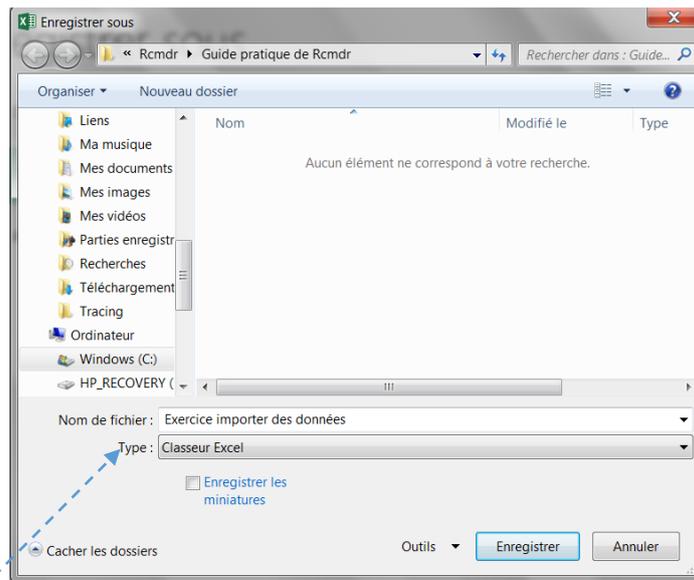
Voici dans Excel, le tableau créé avec le nom des variables sur la première ligne :



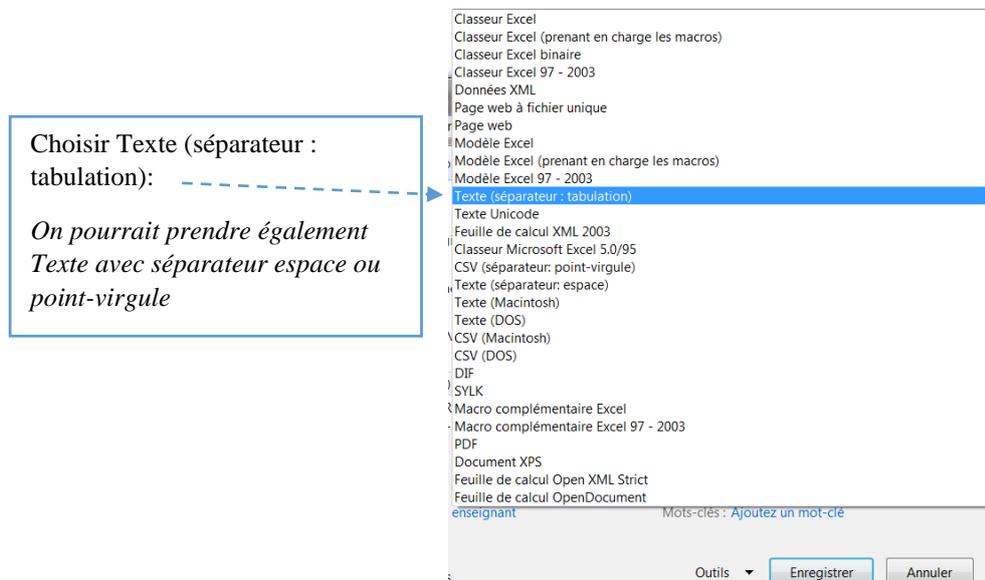
The screenshot shows the Microsoft Excel interface. The ribbon includes 'FICHIER', 'ACCUEIL', 'INSERTION', 'MISE EN PAGE', 'FORMULES', and 'DO'. The 'ACCUEIL' ribbon is active, showing options for 'Coller', 'Police', and 'Alignement'. The formula bar shows 'Elève' in cell A1. The data table is as follows:

	A	B	C	D	E
1	Elève	Sexe	Note		
2	1	1	12,5		
3	2	2	9		
4	3	1	17		
5	4	2	15,5		
6	5	2	12		
7	6	1	10		
8					

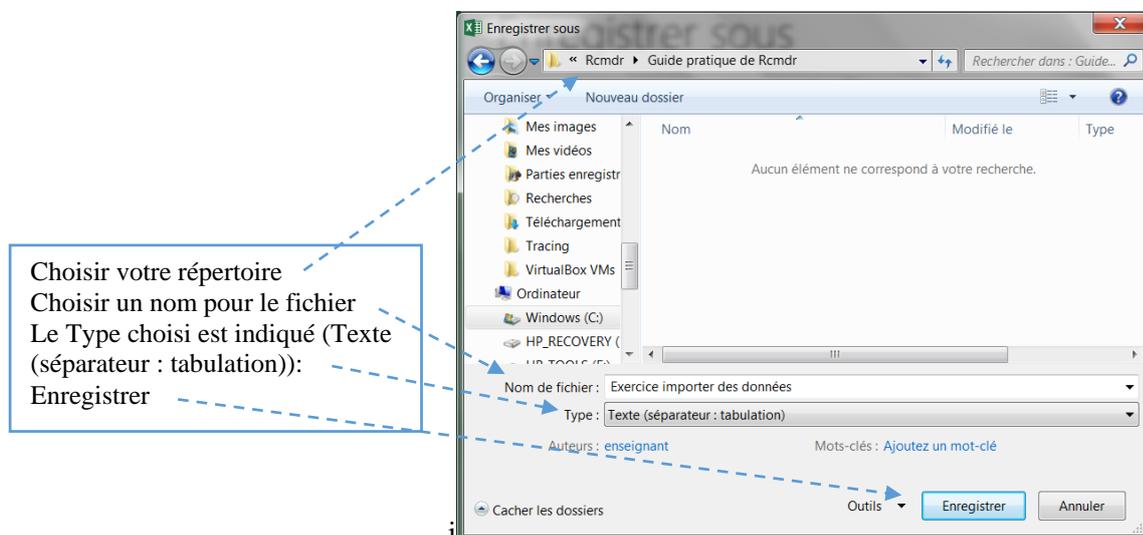
Aller dans Fichier → Enregistrer sous, pour ouvrir la fenêtre Enregistrer sous :



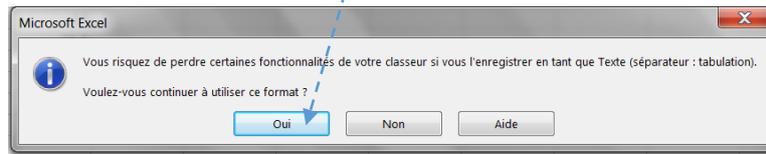
Cliquer sur Type et la fenêtre des formats d'enregistrement s'ouvre :



La fenêtre Enregistrer sous revient

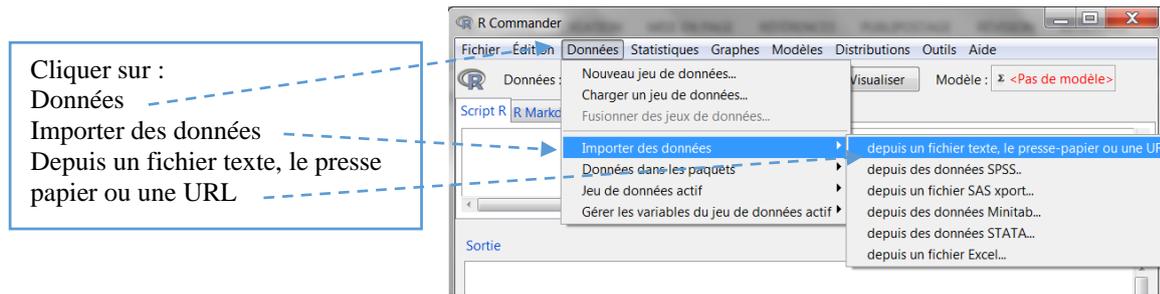


Si la fenêtre suivante apparaît, cliquer Oui

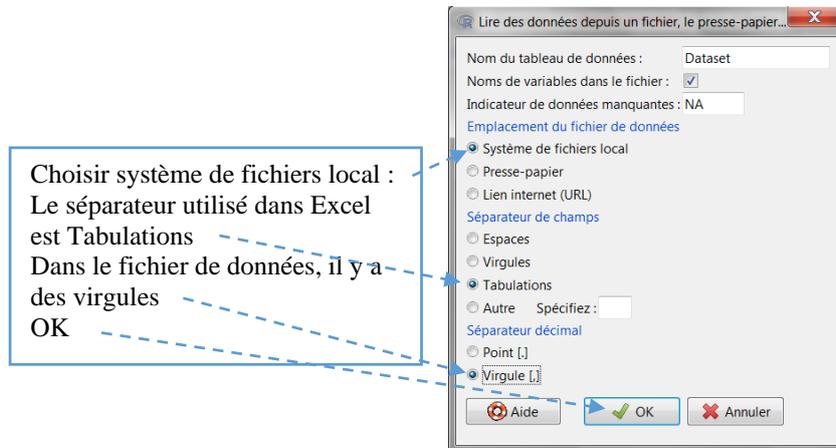


Fermer le fichier Excel.

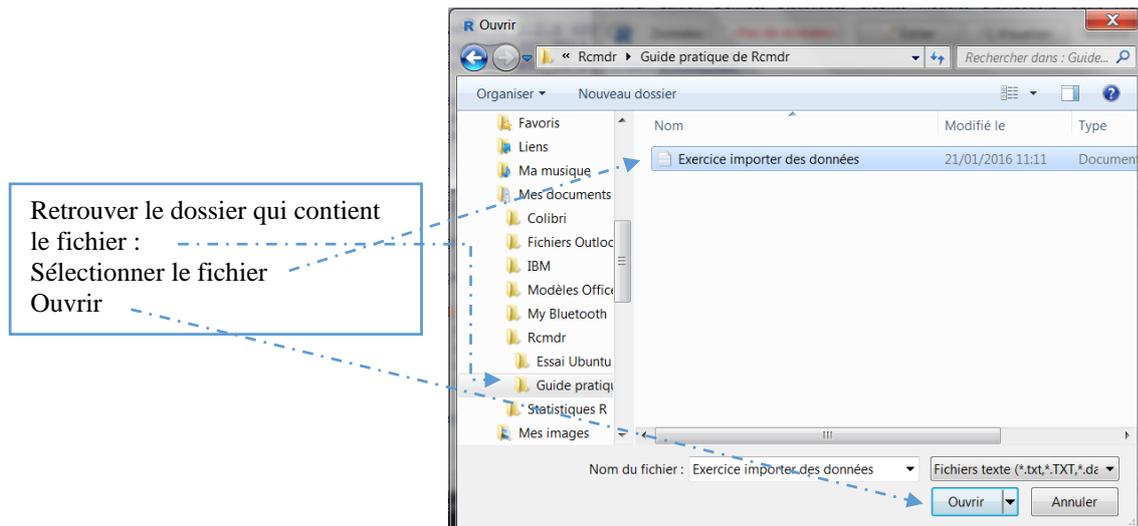
Retournons maintenant dans R Commander :



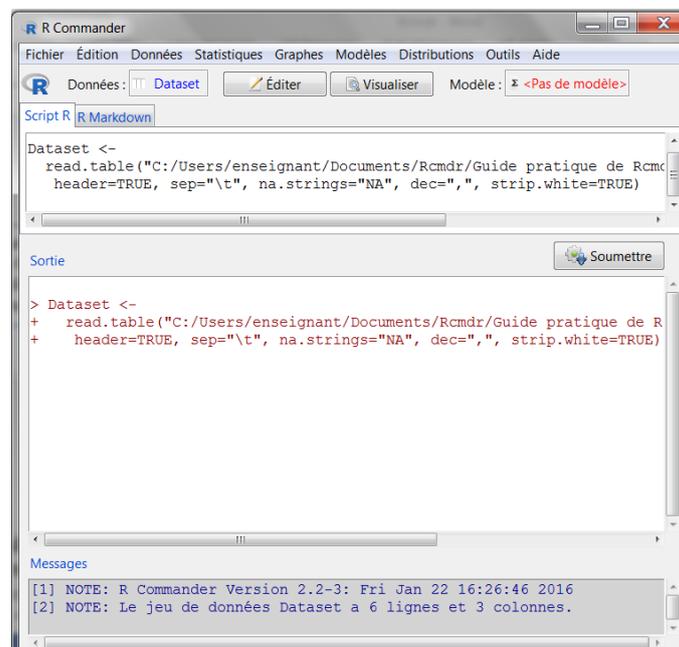
La fenêtre 'Lire depuis un fichier, le presse papier' s'ouvre. Elle offre propose plusieurs options qu'il faudra spécifier : l'emplacement du fichier de données ; le séparateur de champs, le séparateur décimal.



La fenêtre Ouvrir apparaît. Aller dans le dossier dans lequel le tableau Excel a été enregistré.



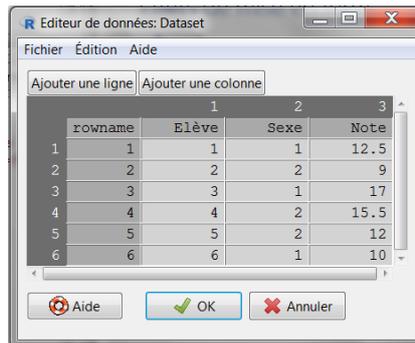
La fenêtre R Commander s’ouvre avec tout en bas une seconde note bleue indiquant, le fichier ouvert, son nom et les nombres de lignes et colonnes.



On peut visualiser le tableau sans modification possible, en cliquant sur le bouton Visualiser :

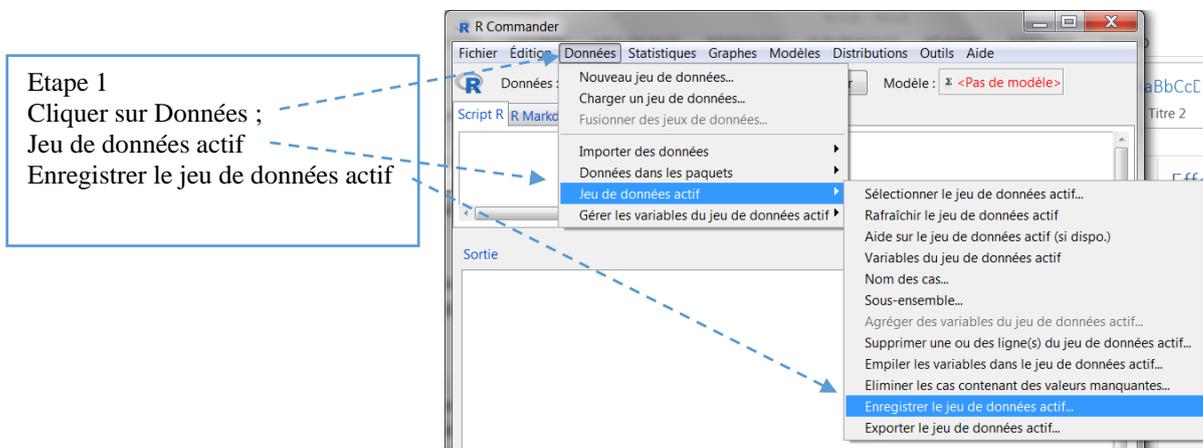
	Elève	Sexe	Note
1	1	1	12.5
2	2	2	9.0
3	3	1	17.0
4	4	2	15.5
5	5	2	12.0
6	6	1	10.0

Pour retirer la fenêtre ‘Visualiser’, cliquer sur la croix en haut à droite. On peut éditer le fichier pour procéder à des modifications, à l’ajout de nouvelles données ou du nom des variables ou des modalités. Il faut cliquer sur le bouton Editer. Il s’agit exactement du même tableau que si l’on passe par l’éditeur de données comme on a vu plus haut :

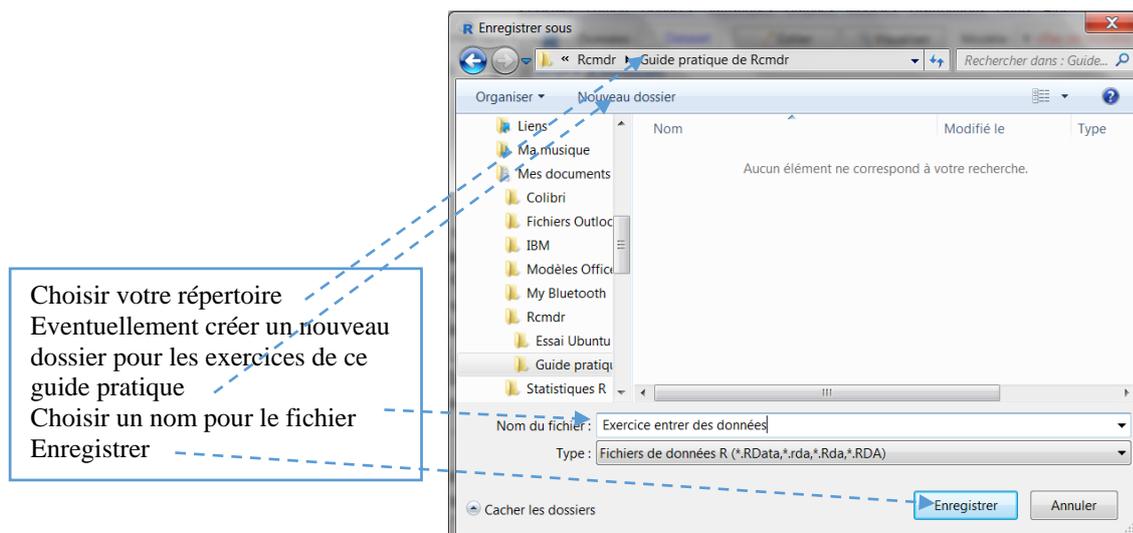


Enregistrer un tableau de données

Pour enregistrer ce fichier actif :

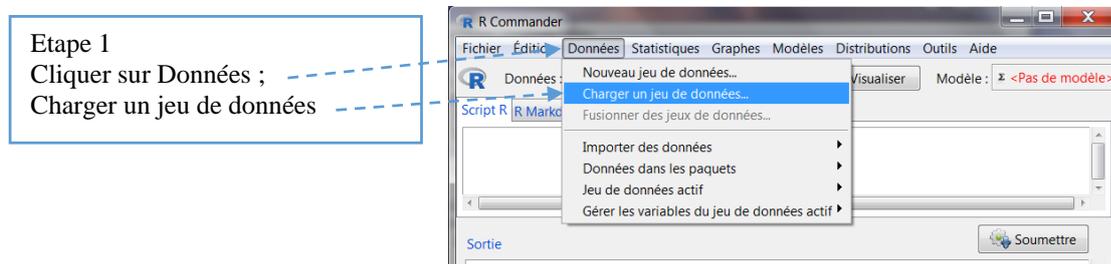


La fenêtre Enregistrer sous apparaît :

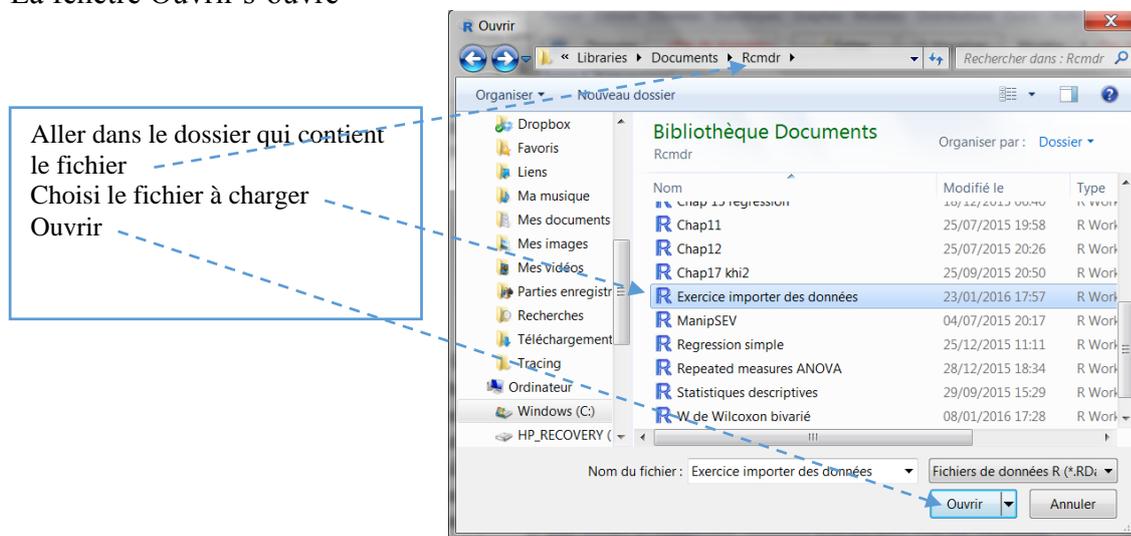


## 2.4 Ouvrir un fichier de données déjà enregistré sous R

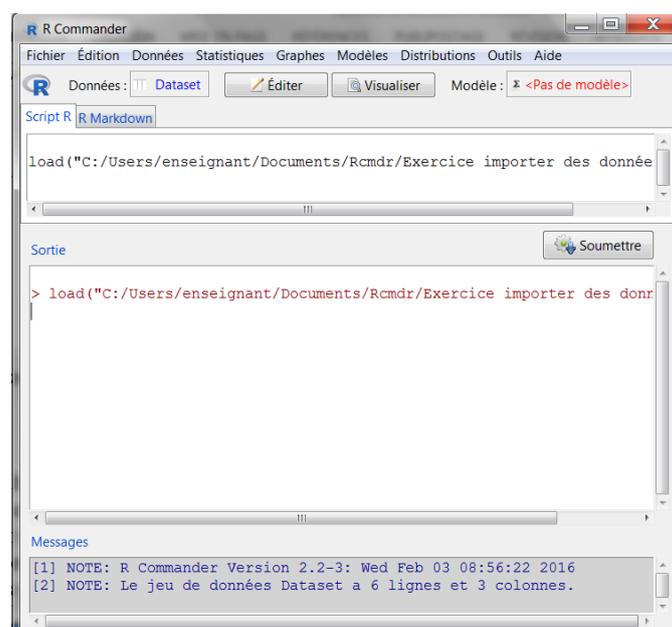
Au cours d'une nouvelle session avec R Commander, ou lorsque l'on souhaite travailler sur un jeu de données déjà enregistré sous R, il faut ouvrir ce jeu, c'est-à-dire « charger le jeu ».



La fenêtre Ouvrir s'ouvre



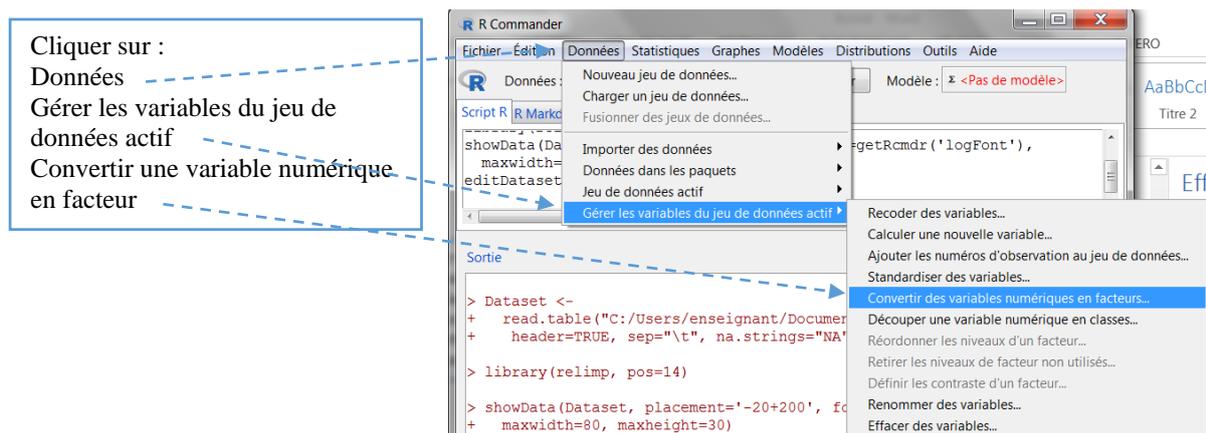
La fenêtre R Commander revient, avec tout en bas le message bleu indiquant le chargement du fichier comme fichier actif les nombres de lignes et de colonnes :



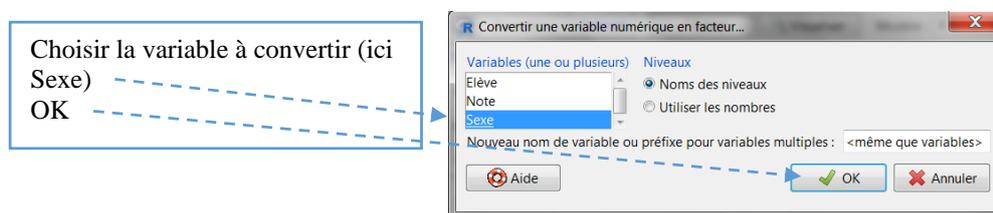
## 2.5 Coder les variables

Il existe trois sortes de variables dans R Commander : (1) les variables numériques, qui sont des nombres et qui correspondent en fait aux variables dépendantes ou variables à expliquer. Ce sont des valeurs numériques, des scores, des notes, des réponses chiffrées sur une échelle de type Likert, des grandeurs physiques, ... ; (2) les facteurs, qui sont des mots et qui correspondent aux variables indépendantes ou variables explicatives. Par souci de simplification, on l'a vu plus haut, on attribue généralement un chiffre aux différentes modalités ou niveaux de ce type de variables, par exemple, 1 pour fille, 2 pour garçon pour la variable sexe. (3) les variables logiques (*true*, *false*), dont il ne sera pas question dans ce guide.

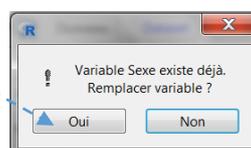
La spécification entre les deux types de variables utilisées dans ce qui suivra, variables numériques et facteurs, se fait automatiquement par R Commander. Autrement dit, en entrant un nombre dans une colonne, R Commander la considère automatiquement comme une variable numérique ; en entrant un mot, R Commander considère la colonne comme un facteur. Si l'on nomme une variable 'Sexe' et que l'on rentre 'fille' ou 'garçon' dans les cellules, R Commander considère la variable 'Sexe' comme un facteur. Mais si l'on rentre 1 pour les filles et 2 pour les garçons, comme dans l'exemple ci-dessus, R Commander considèrera cette même variable 'Sexe' comme une variable numérique. Il faut donc convertir la variable 'Sexe' pour en faire un facteur. La procédure est très simple. Utilisons le même fichier de données que ci-dessus.



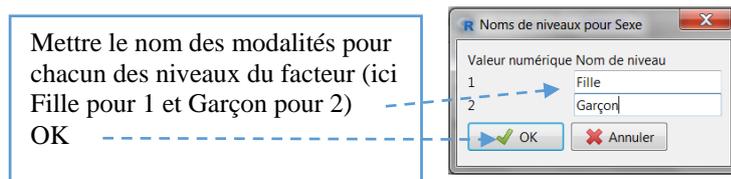
La fenêtre 'Convertir une variable numérique en facteur' s'ouvre :



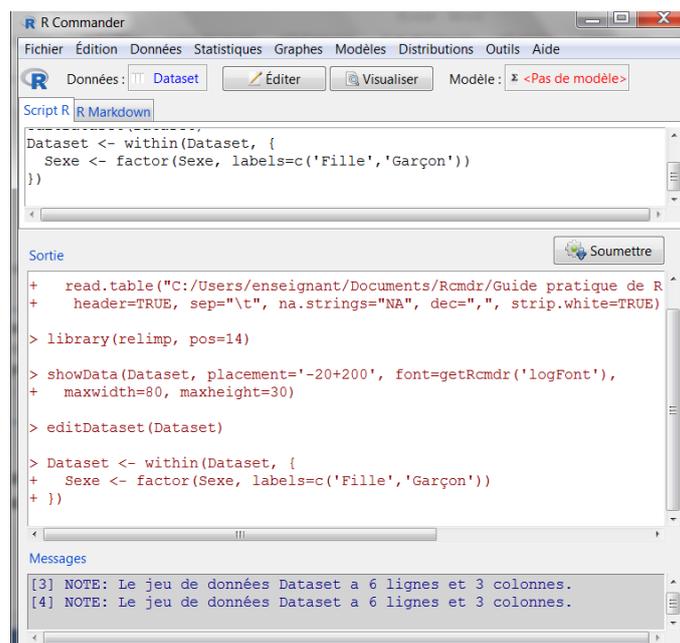
Une fenêtre d'avertissement prévient qu'il existe déjà un nom pour la variable. Cliquer Oui :



La fenêtre 'Noms de niveaux pour Sexe' s'ouvre :



La fenêtre R Commander apparaît de nouveau avec tout en bas un nouveau message bleu. On voit également dans la fenêtre des scripts le facteur Sexe et les noms (*labels*) de ses deux modalités, Fille et Garçon.



Cliquons sur Visualiser pour vérifier que les chiffres ont été remplacés par des noms :

Elève	Sexe	Note
1	Fille	12.5
2	Garçon	9.0
3	Fille	17.0
4	Garçon	15.5
5	Garçon	12.0
6	Fille	10.0

Enregistrer le fichier pour conserver les codages effectués.

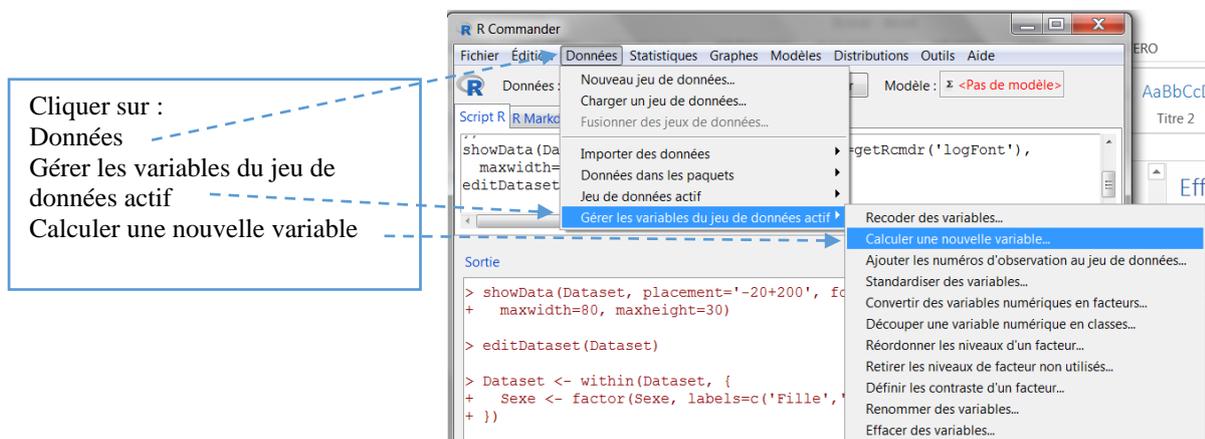
Ce codage des variables est primordial. En effet, R Commander est un logiciel qui intègre dans ses menus et les options offertes, les variables et leur statut. Par exemple, si le jeu de données ne comporte qu'une seule variable quantitative, la commande Corrélation ne peut pas être activée. Ou encore, si le jeu de données ne comporte qu'une variable indépendante (ou facteur) avec 2 modalités, la commande Test pour groupes appariés reste inactive, de même que l'ANOVA. Autrement dit, si cela évitera bien des erreurs, il faut en revanche ne pas se tromper dans le codage des variables ou oublier de coder un facteur.

## 2.6 Calculer de nouvelles variables

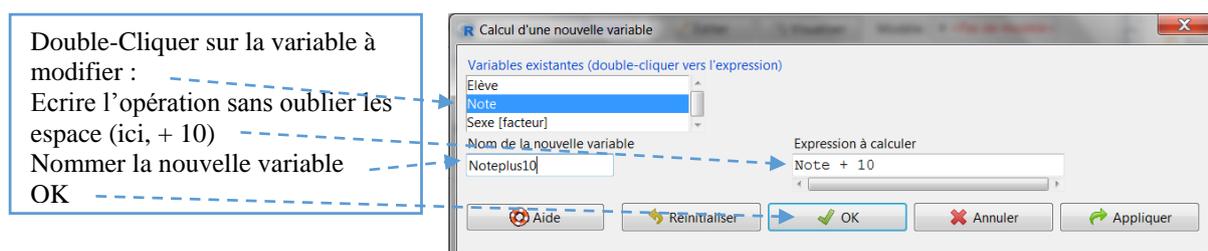
R Commander permet de calculer de nouvelles variables sur un jeu de données. Voici quelques opérations possibles :

Opération	Signe	Exemple
Addition	$x + y$	Variable 1 + 25
Soustraction	$x - y$	5 - Variable 1
Multiplication	$x * y$	Variable 1*10
Division	$x / y$	Variable 1/63
Elever à la puissance	$x ^ y$	Variable1^2
Racine carrée	sqrt(x)	sqrt(variable 1)
Transformation logarithmique	log10(x) ou log(x,10)	Log10(variable 1)

Faisons quelques exercices sur le jeu de données ci-dessus.



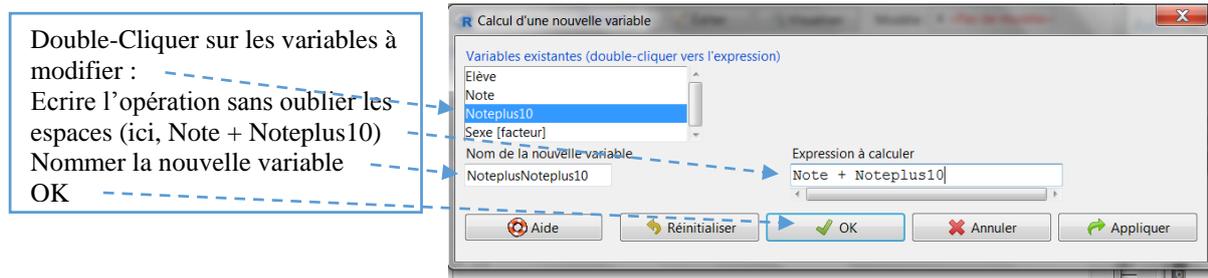
La fenêtre apparaît. Ajoutons 10 aux notes des élèves



La fenêtre R Commander revient. Visualisons le fichier :

Blève	Sexe	Note	Noteplus10
1	Fille	12.5	22.5
2	Garçon	9.0	19.0
3	Fille	17.0	27.0
4	Garçon	15.5	25.5
5	Garçon	12.0	22.0
6	Fille	10.0	20.0

On peut aussi effectuer les opérations avec deux variables. Après avoir Cliqué Données → Gérer les variables du jeu actif → Calculer de nouvelles variables :

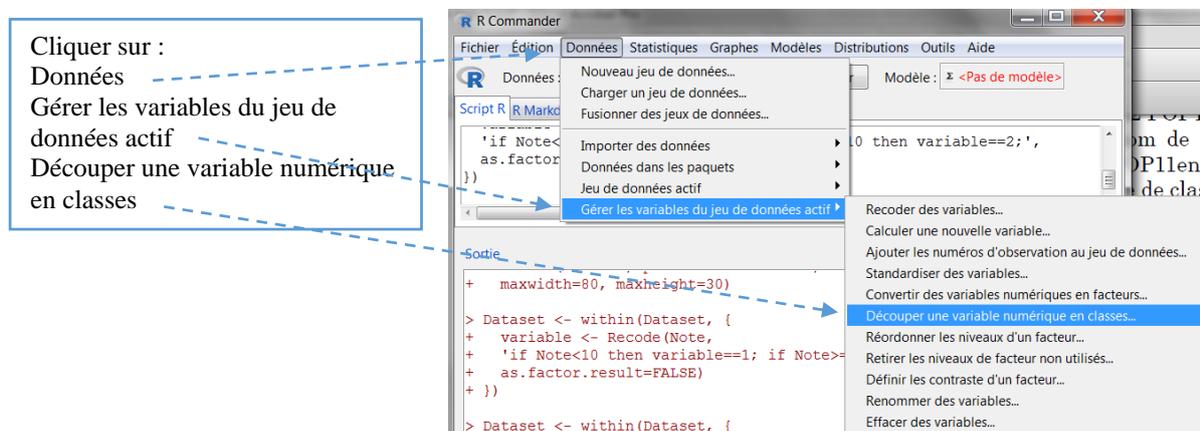


Une fois la fenêtre R Commander revenu, cliquons sur visualiser :

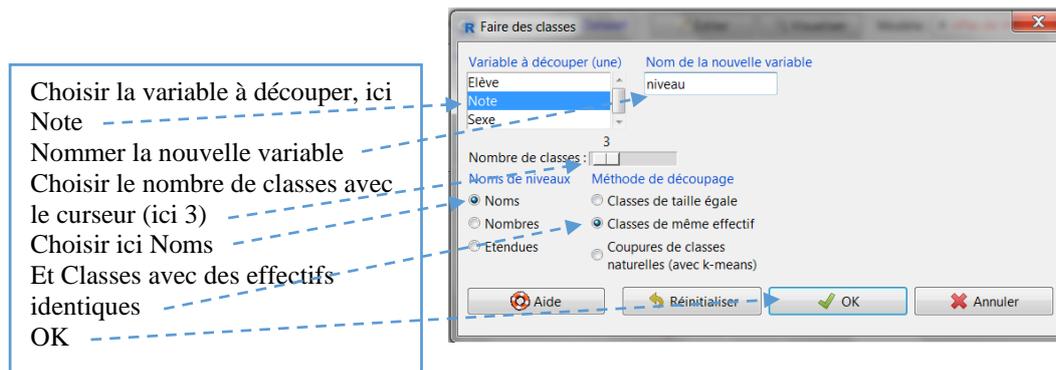
Elève	Sexe	Note	Noteplus10	NoteplusNoteplus10	NoteplusNoteplus10
1	Fille	12.5	22.5	35	
2	Garçon	9.0	19.0	28	
3	Fille	17.0	27.0	44	
4	Garçon	15.5	25.5	41	
5	Garçon	12.0	22.0	34	
6	Fille	10.0	20.0	30	

## 2.7 Découper une variable numérique en classes

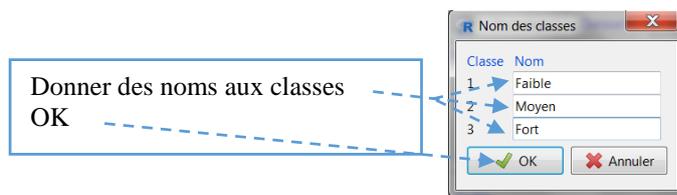
On peut souhaiter découper une variable numérique en classes, par exemple, regrouper les scores à une épreuve en trois classes, scores faibles, moyens et forts, ou encore par déciles... Plus précisément, cette nouvelle variable est un facteur et donc les classes sont autant de modalités. Voici comment procéder :



La fenêtre Faire des classes apparaît :



Comme on a souhaité ici donner des noms aux classes, la fenêtre ‘Nom des classes’ apparaît :



La fenêtre R Commander s’ouvre et cliquons sur Visualiser pour voir les modifications :

Elève	Sexe	Note	niveau
1	Fille	12.5	Moyen
2	Garçon	9.0	Faible
3	Fille	17.0	Fort
4	Garçon	15.5	Fort
5	Garçon	12.0	Moyen
6	Fille	10.0	Faible

A partir de la variable Note, une autre variable, a bien été créée avec 3 classes aux effectifs égaux. Plus précisément, il s’agit d’un facteur à trois modalités dont les effectifs sont égaux.

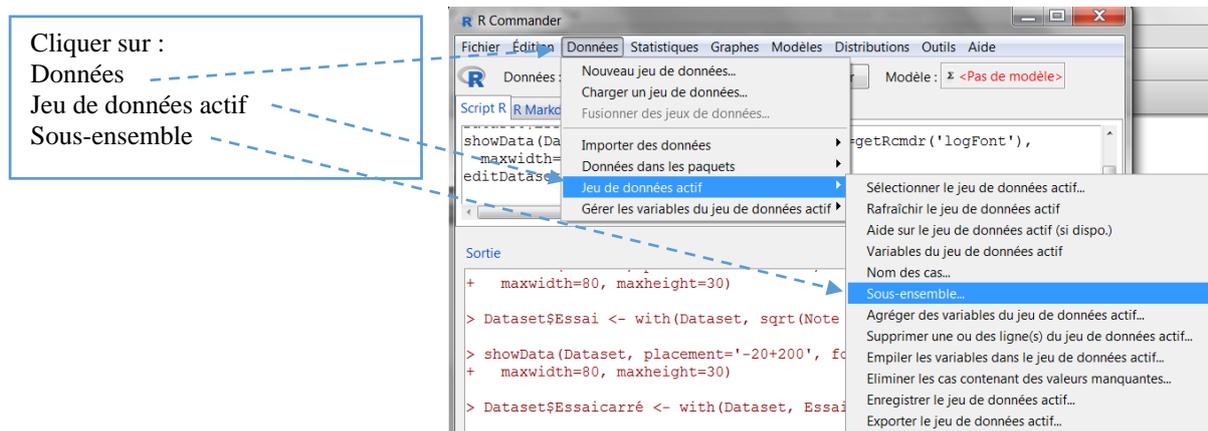
## 2.8 Générer des jeux de données partiels

R Commander permet de créer des jeux de données partiels, c’est-à-dire de nouveaux tableaux mais avec une partie des données. On peut ainsi avoir un nouveau jeu de données avec certaines variables seulement, ou encore avec certains individus qui présentent des valeurs particulières sur une variable donnée. Plusieurs symboles ou codes peuvent être utilisés :

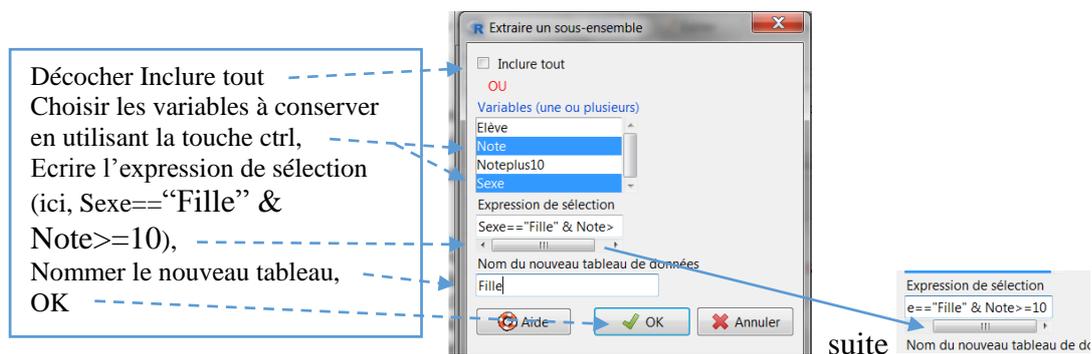
Symbole/code	Nom
==	égalité
!=	inégalité
&	et
	ou
>	Plus grand
<	Plus petit
>=	Plus grand ou égal
<=	Plus petit ou égal
is.na(variable)	Pour inclure les valeurs manquantes d’une variable
!is.na(variable)	Pour exclure les valeurs manquantes d’une variable

Si l'on utilise un nom, il faut mettre le nom entre guillemets (par exemple "Fille"). Attention, R Commander tient compte des majuscules et n'apprécie pas les lettres avec accents.

A titre d'exemple, nous allons créer un nouveau jeu de données dans lequel on souhaite n'avoir que la variable Note pour les filles seulement (expression de sélection : Sexe=="Fille") et (&) qui ont la moyenne, soit dont la note est supérieure ou égale à 10 (expression de sélection : Note>=10).



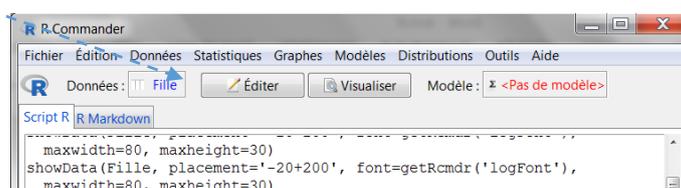
La fenêtre Extraire un sous-ensemble s'ouvre :



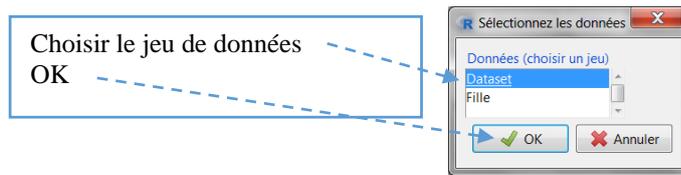
La fenêtre de R Commander revient et cliquons sur Visualiser :

Note	Sexe
12.5	Fille
17.0	Fille
10.0	Fille

On remarquera également que le fichier actif est devenu Fille ainsi que la note bleue tout en bas qui rappelle le nom et indique que le tableau 3 lignes et 2 colonnes. Si l'on souhaite revenir sur le fichier précédent, il suffit de cliquer sur Fille et de choisir l'autre fichier en mémoire :



La fenêtre Sélectionner les données s'ouvre :



Dans la fenêtre R Commander, c'est ce jeu Dataset qui est maintenant actif.

### 3 Statistiques descriptives

L'analyse statistique de données quantitatives commence par la description des données. Cette description se fait sous deux formes : à l'aide de tableaux et de diagrammes d'une part ; à l'aide de statistiques descriptives d'autre part.

Une enseignante souhaite tester l'efficacité de deux méthodes pédagogiques. Elle a divisé sa classe de 40 élèves en deux et de façon aléatoire : 20 élèves ont eu une pédagogie dite différenciée, cotée 1 ci-dessous ; 20 élèves ont eu une pédagogie dite traditionnelle, cotée 2 ci-dessous. Les 40 élèves ont ensuite passé la même épreuve notée sur 100. Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableaux de données) et convertir la variable Pédagogie en facteur, avec 1 = Différenciée et 2 = Traditionnelle.

Ces données sont celles utilisées dans le chapitre Comparaison de moyennes avec le  $t$  de Student. Elles ont été scindées en 2 ci-dessous pour gagner de la place, mais le tableau comporte bien 3 colonnes (Elève, Pédagogie et Note) et 40 lignes (hors la ligne titre).

Elève	Pédagogie	Note	<i>(Suite)</i>	Elève	Pédagogie	Note
1	1	87		21	2	82
2	1	95		22	2	72
3	1	84		23	2	95
4	1	74		24	2	60
5	1	76		25	2	90
6	1	92		26	2	87
7	1	77		27	2	89
8	1	90		28	2	86
9	1	94		29	2	76
10	1	84		30	2	74
11	1	91		31	2	85
12	1	90		32	2	75
13	1	75		33	2	90
14	1	93		34	2	91
15	1	87		35	2	88
16	1	85		36	2	63
17	1	90		37	2	78
18	1	89		38	2	72
19	1	87		39	2	84
20	1	85		40	2	60

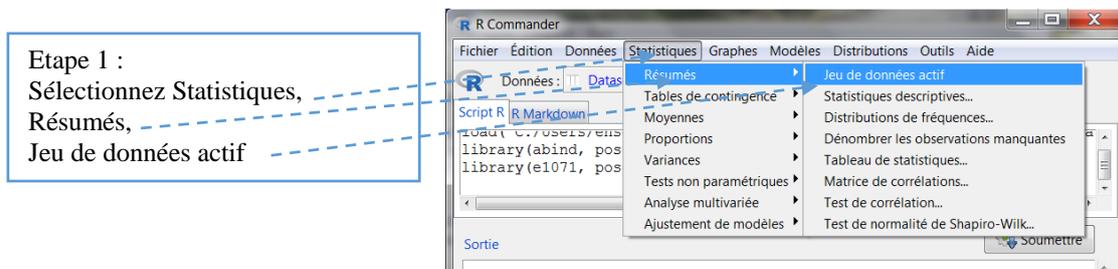
Nous verrons dans un premier temps les statistiques descriptives numériques et dans un second temps la description des données à l'aide de graphiques.

### 3.1 Statistiques descriptives numériques

R Commander offre plusieurs commandes pour résumer les données et les variables.

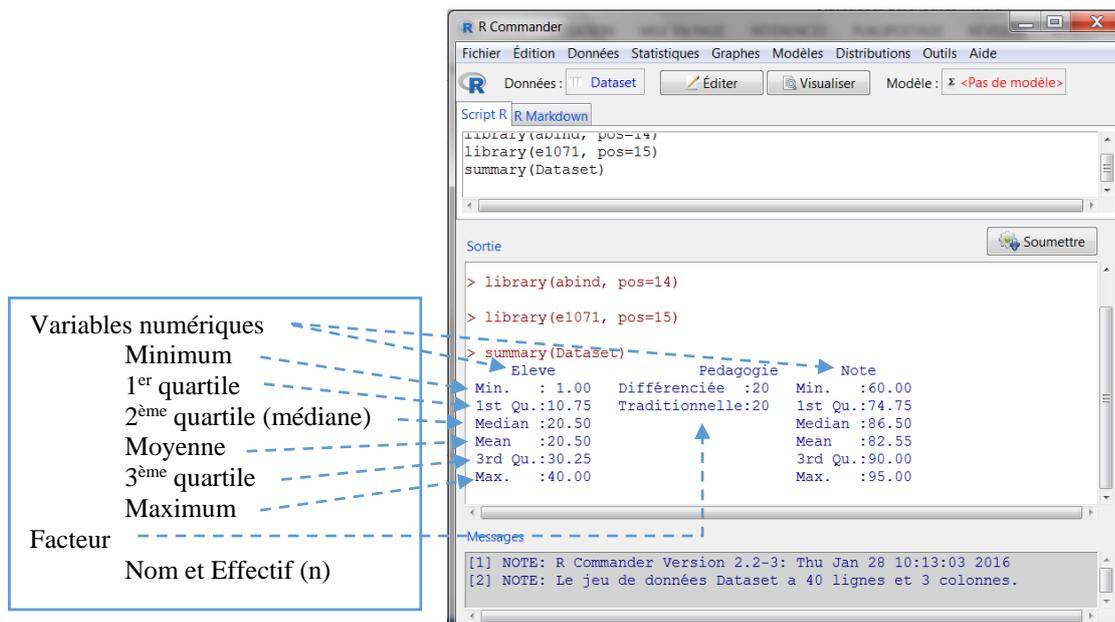
#### 3.1.1 Statistiques descriptives par variable

La commande ‘Jeu de données actif’ permet une première lecture très rapide de l’ensemble des variables du tableau, aussi bien les variables numériques que les facteurs, c’est là son intérêt:



La fenêtre R Commander revient avec les résultats suivants :

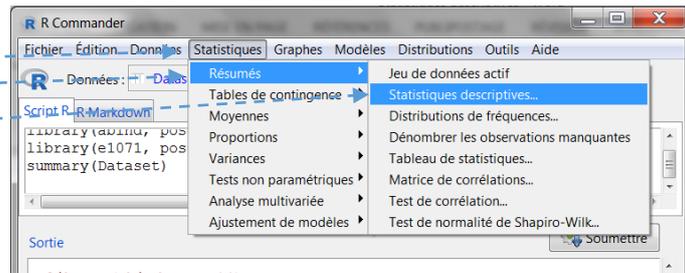
pour les variables numériques, les notes minimum et maximum, les 1<sup>er</sup>, 2<sup>ème</sup> (médiane) et 3<sup>ème</sup> quartiles, la moyenne ;  
pour les facteurs, le nom des modalités et leur effectif



Cette commande permet de vérifier le statut des variables, ici 2 variables numériques, (Eleve et Note), et un facteur (Pédagogie). On remarque aussi que la liste des élèves (première colonne) est considérée comme une variable numérique.

La commande Jeu de données actif permet également d’obtenir quelques statistiques de tendance centrale (moyenne et médiane) et de dispersion (minimum, maximum et quartiles). Cependant, pour avoir davantage de statistiques descriptives, il faut passer par d’autres commandes, en particulier la commande Statistiques descriptives.

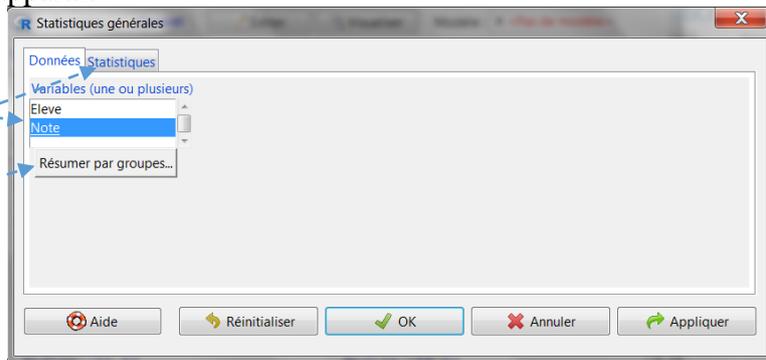
Etape 1 :  
 Statistiques  
 Résumés  
 Statistiques descriptives



La fenêtre Statistiques générales apparaît :

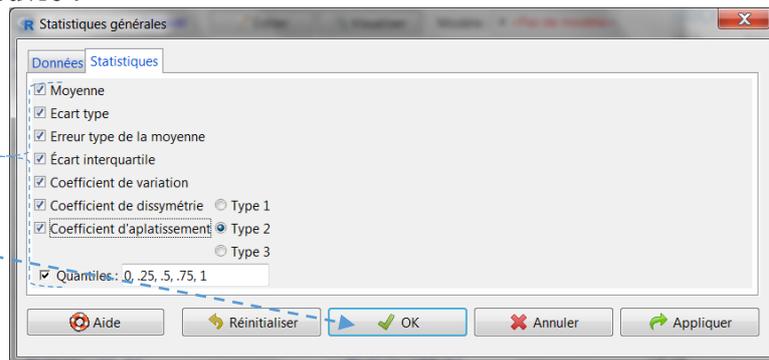
Etape 2 :  
 Choisir la variable (ici Note)  
 Cliquer sur l'onglet Statistiques

Cette touche permet de calculer les statistiques pour chaque modalité des facteurs (nous y reviendrons ci-dessous)



L'onglet Statistiques générales s'ouvre :

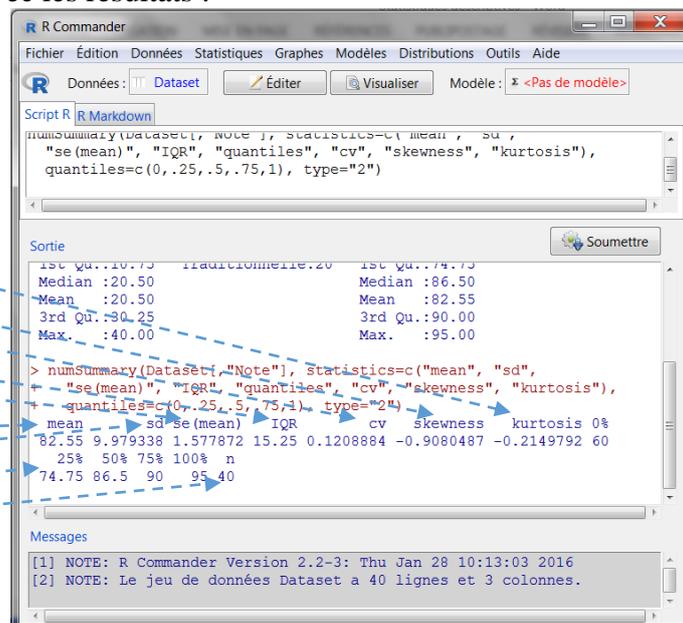
Etape 3 :  
 Choisir les statistiques en cliquant devant.  
 Ici, sélectionnons toutes les statistiques proposées :  
 OK



La fenêtre R Commander apparaît avec les résultats :

Résultats  
 Coefficient d'aplatissement  
 Coefficient de symétrie  
 Coefficient de variation  
 Espace inter-quartile  
 Erreur type de la moyenne

Moyenne  
 Ecart-type  
 Quartiles  
 Effectif (n)



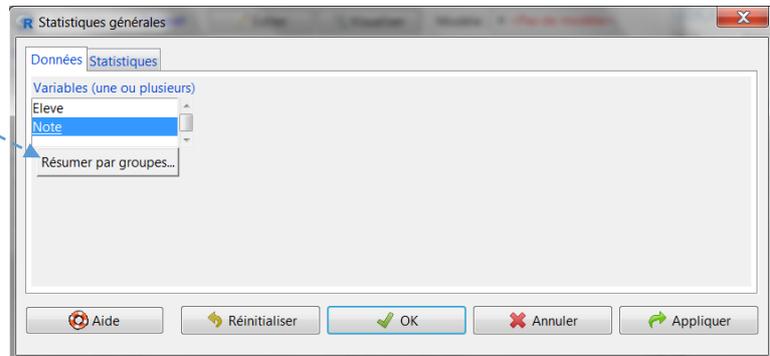
## 3.1.2 Statistiques descriptives par modalité de variable (sous-groupes)

Nous allons reprendre les mêmes étapes que ci-dessus mais en cliquant sur le bouton Résumer par groupes dans la fenêtre Statistiques descriptives. Retournons à cette fenêtre (Statistiques → Résumés → Statistiques descriptives) :

La fenêtre Statistiques générales apparaît comme nous l'avions laissée :

Étape 1 :

La variable est déjà choisie (ici Note)  
Cliquez sur Résumer par groupes

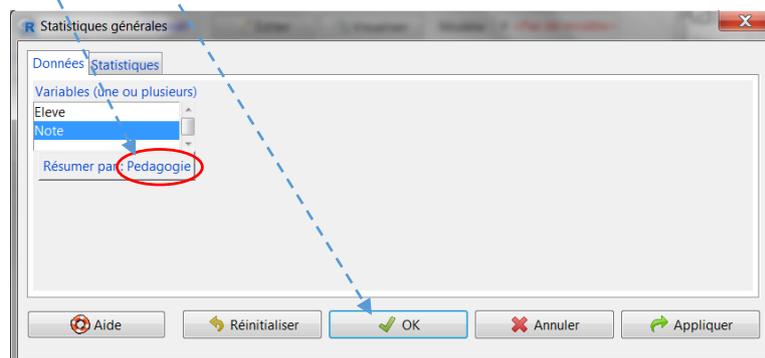


La petite fenêtre Groupes s'ouvre permettant de choisir la variable de regroupement (le facteur) souhaité. Ici, comme il n'y a qu'un seul facteur, R Commander le prend par défaut. Il suffit de cliquer sur OK

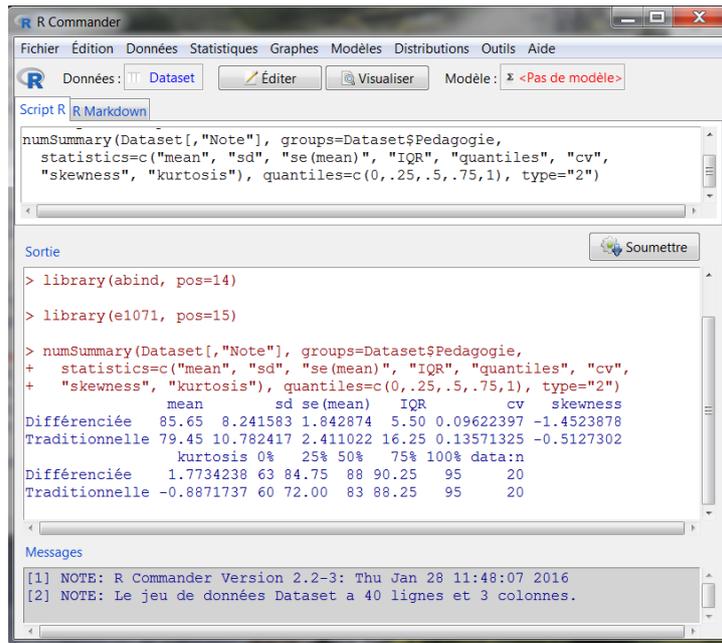


La fenêtre Statistiques générales revient avec le nom du facteur Pédagogie dans le bouton Résumer par.

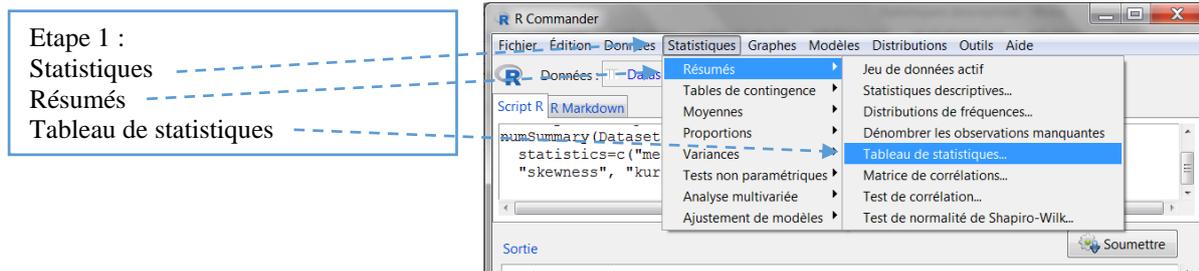
Enfin, comme nous avons déjà spécifié les Statistiques dans l'onglet statistiques, R Commander reprendra automatiquement celles-ci (nous avons demandé toutes les statistiques). Il suffit de cliquer sur OK



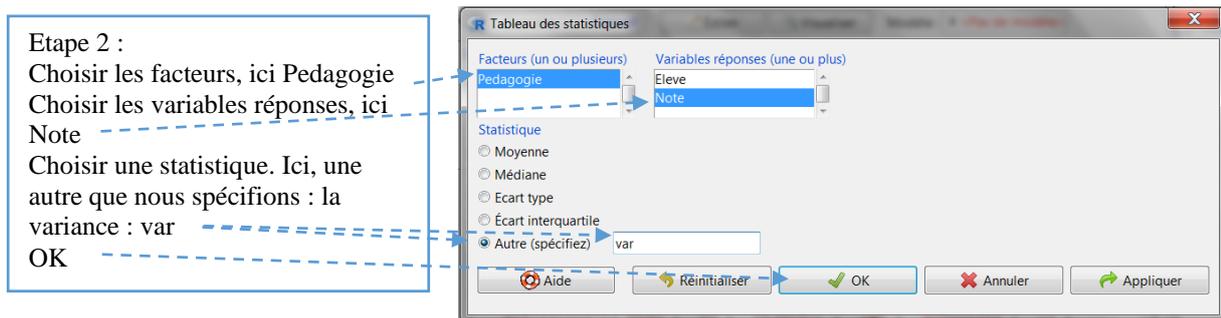
La fenêtre R Commander revient avec les mêmes résultats que ci-dessus mais par groupe de pédagogie, Différenciée vs Traditionnelle :



Une autre commande permet de présenter les statistiques descriptives de différents groupes sous la forme de tableau : la commande Tableau de statistiques

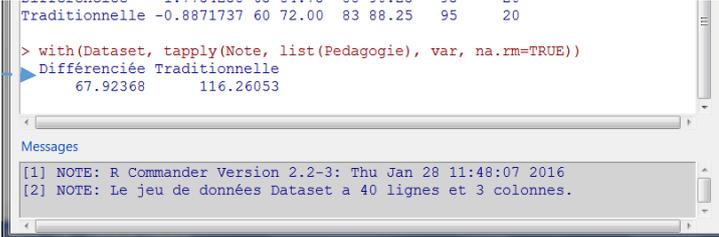


La fenêtre Tableau de statistiques s'ouvre :



La fenêtre R Commander apparait avec les résultats sous la forme d'un tableau avec les modalités du facteur (VI) en colonne les variances des notes en lignes (VD) :

Variance des 2 groupes



```

-----
Traditionnelle -0.8871737 60 72.00 83 88.25 95 20
-----
> with(Dataset, tapply(Note, list(Pedagogie), var, na.rm=TRUE))
Différenciée Traditionnelle
67.92368      116.26053

Messages
[1] NOTE: R Commander Version 2.2-3: Thu Jan 28 11:48:07 2016
[2] NOTE: Le jeu de données Dataset a 40 lignes et 3 colonnes.

```

La variance du sous-groupe à pédagogie différenciée est de 67,92 ; celle du sous-groupe à pédagogie traditionnelle, de 116,26.

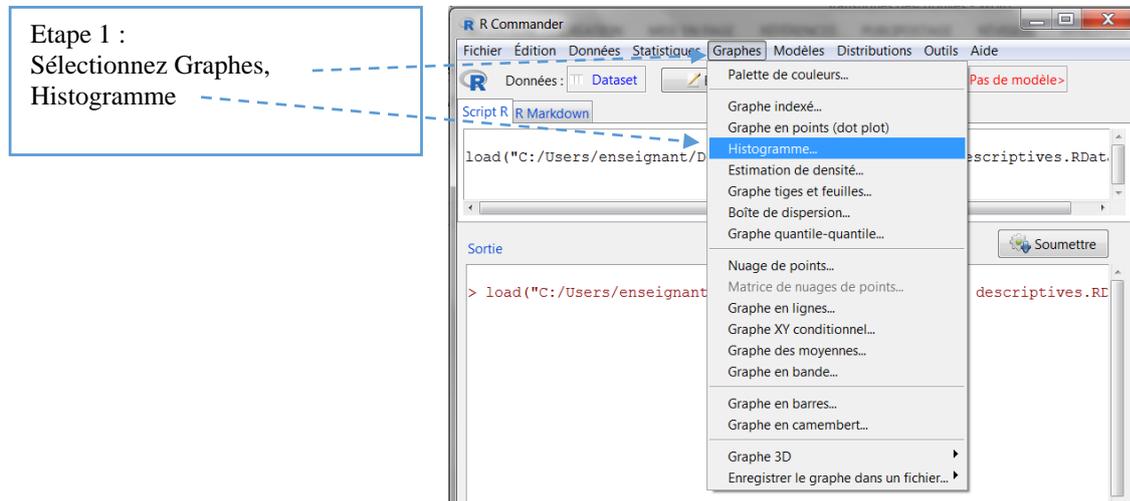
Cette commande ‘Tableau de statistiques’ permet donc de construire des tableaux avec les VI en colonnes, les VD en lignes. De plus, il permet de croiser les modalités de plusieurs facteurs (ce que nous ne pouvons pas faire dans cet exemple où il n’y a qu’un seul facteur). Enfin, l’option ‘Autre’ permet de calculer des statistiques qui n’apparaissent pas dans les choix proposés, ce que nous avons fait avec la variance et sa commande ‘var’. Le tableau suivant présente les commandes de R pour ces statistiques descriptives :

Statistique	Commande
Moyenne	mean
Médiane	median
Variance	var
Ecart-type	sd
Valeurs max et min	range
quartile	quantile
Espace interquartile	IQR
Somme totale	sum
Toutes les valeurs	c

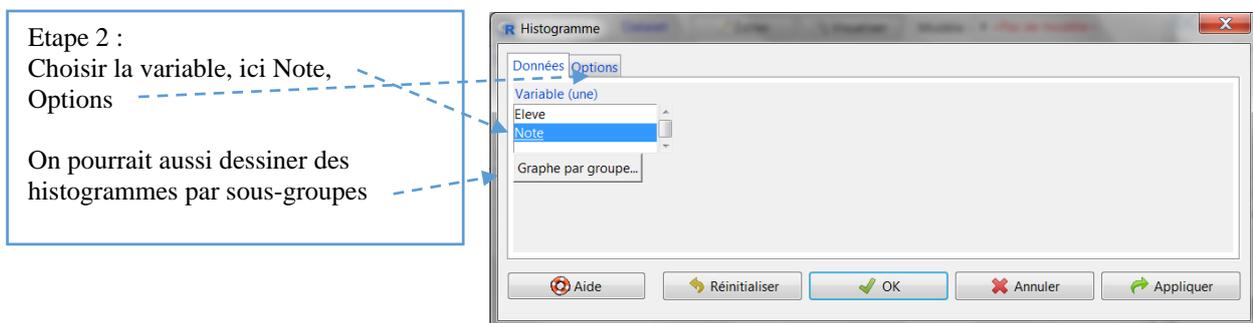
### 3.2 Construire des graphiques pour décrire des données

R est un logiciel extrêmement puissant pour construire des graphiques. Et il existe dans R Commander plusieurs commandes pour cela, regroupées sous le titre ‘Graphes’ du menu.

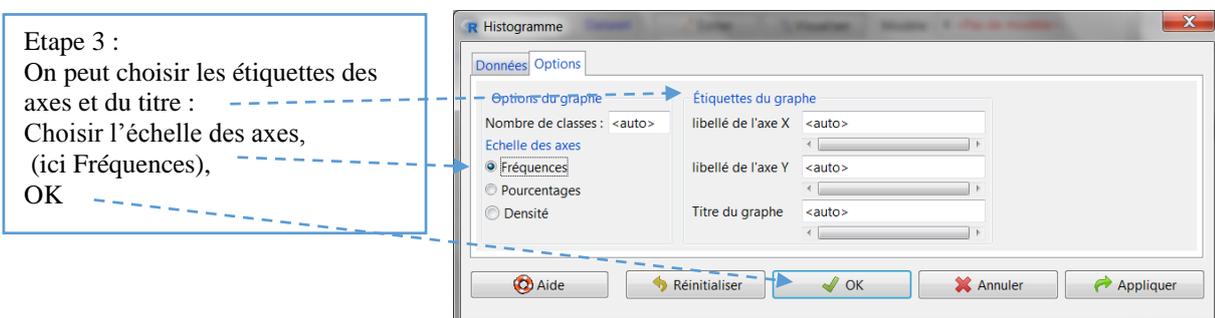
#### 3.2.1 Dessiner un histogramme



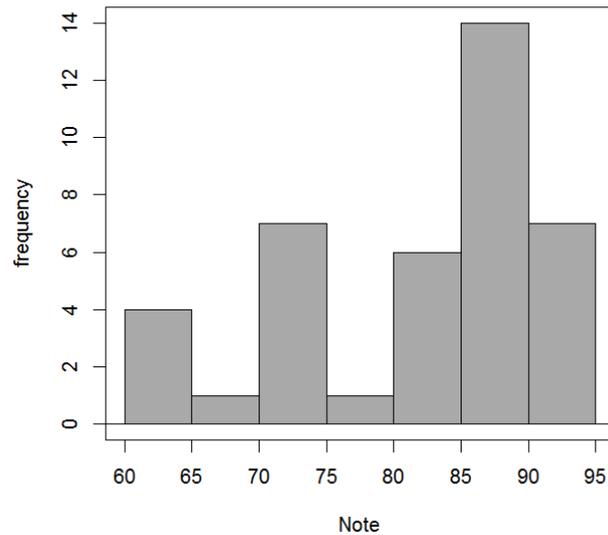
Dans la fenêtre Histogramme, il y a deux onglets : Données et Options. Renseignons tout d’abord la fenêtre Données :



La fenêtre Options s’affiche :



Le graphe apparaît dans une fenêtre spécifique :



Nous aurions pu changer les étiquettes, mettre un titre...

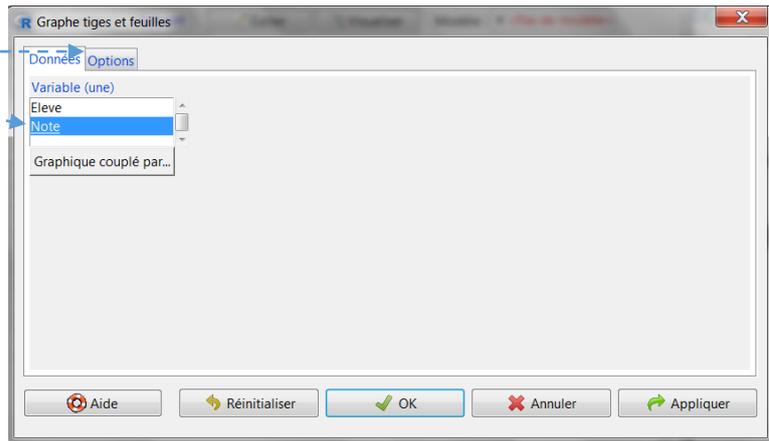
### 3.2.2 Dessiner un graphe Tiges et feuilles

Etape 1 :  
Sélectionnez Graphes,  
Graphe tiges et feuilles

Dans la fenêtre Graphe tiges et feuilles, il y a deux onglets : Données et Options. Renseignons tout d'abord la fenêtre Données :

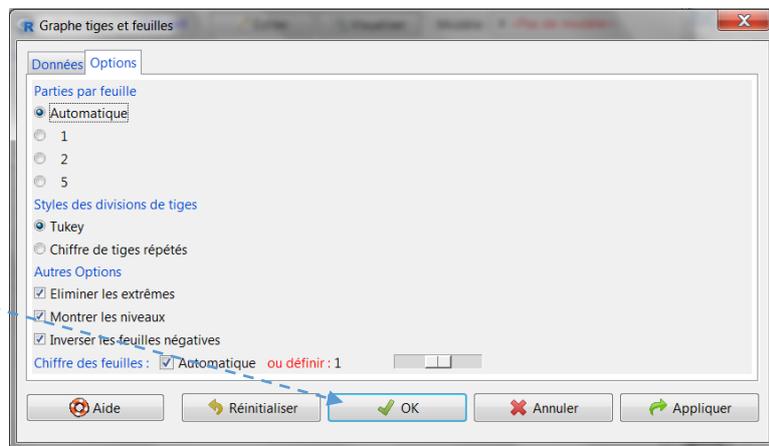
Etape 2 :  
Choisir la variable, ici Note,  
Options

On pourrait aussi dessiner des graphiques par sous-groupes

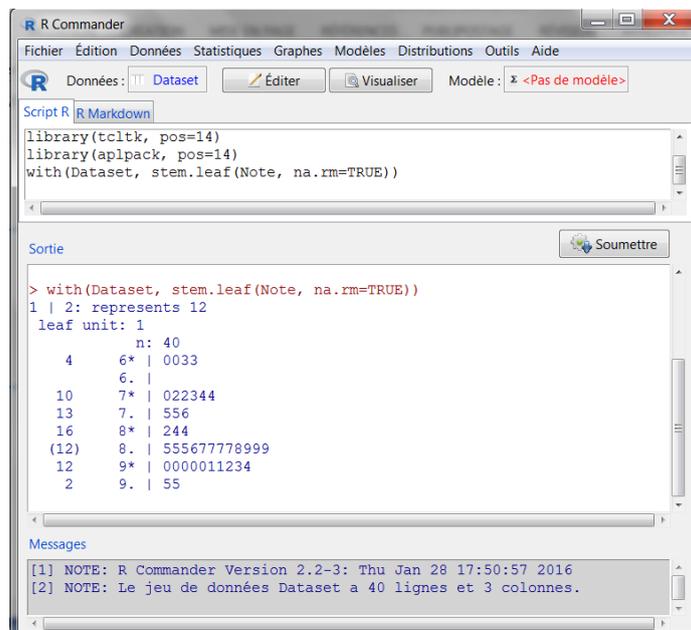


La fenêtre Options s'affiche :

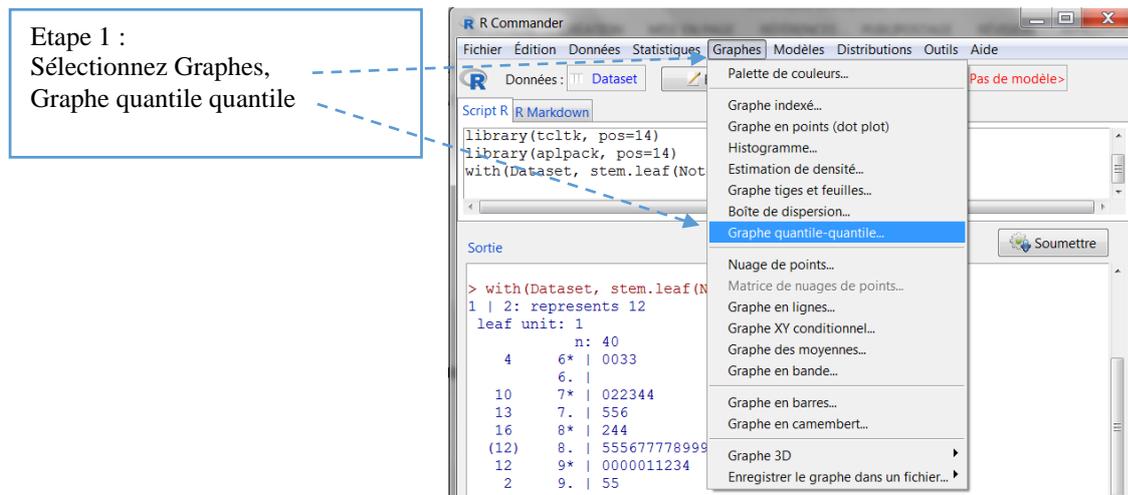
Etape 3 :  
Plusieurs options sont proposées.  
Laissons ici celles qui le sont par  
défaut  
OK



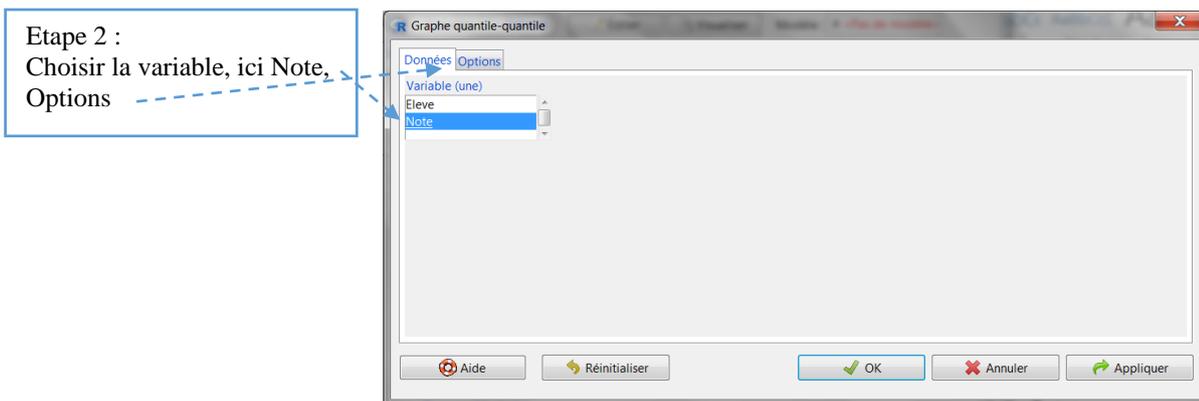
La fenêtre R Commander apparaît avec dans la sortie, le graphe tiges et feuilles :



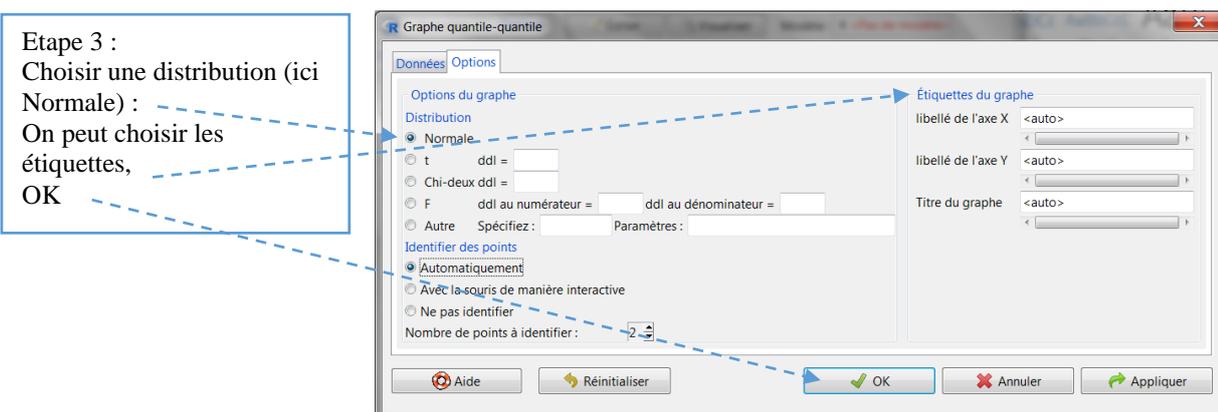
3.2.3 Dessiner un graphe quantile-quantile :



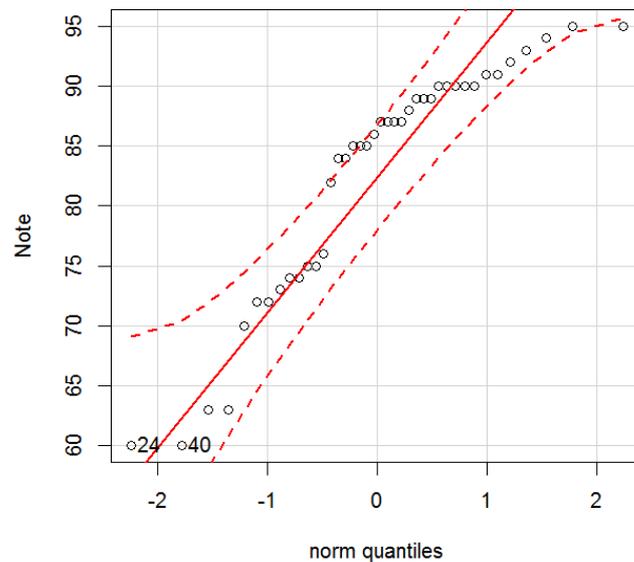
Dans la fenêtre Graphe quantile-quantile, il y a deux onglets : Données et Options. Renseignons tout d'abord la fenêtre Données :



La fenêtre Options s'affiche :



Le graphe apparaît dans une fenêtre spécifique :



Les points ne sont pas le long de la diagonale, la distribution ne semble pas normale.

### 3.2.4 Dessiner une boîte à moustaches.

Une boîte à moustaches permet de visualiser la dispersion des données d'une distribution (d'où son autre nom, boîte de dispersion) et sa relative symétrie. Le graphe a la forme d'une boîte, dont les charnières (les bords) sont les 1<sup>er</sup> et 3<sup>ème</sup> quartiles ; dans la boîte, un trait vertical au niveau de la médiane ; des moustaches de part et d'autre de la boîte, mesurent 1,5 fois la distance interquartile ; et au-delà des moustaches, la présence d'éventuelles valeurs extrêmes ou aberrantes est signalée. Ce type de graphe est d'autant intéressant à dessiner qu'il y a des groupes à comparer. Nous allons à titre d'exemple, dessiner les boîtes à moustaches, pour les deux groupes de pédagogie :

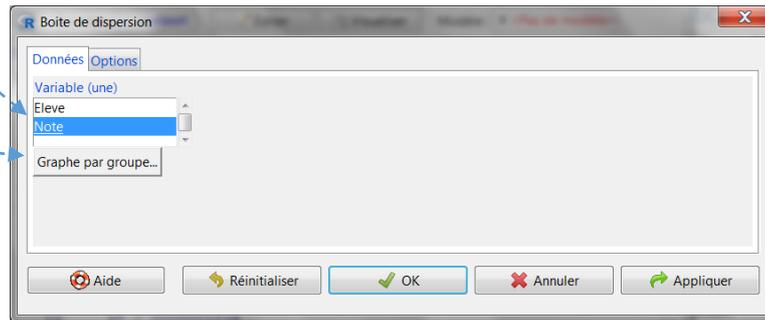
Etape 1 :  
Sélectionnez Graphes,  
Boîte de dispersion

```

> with(Dataset, qqPlot(Note
+   labels=rownames(Dataset
24 40
1 2
    
```

Dans la fenêtre boîte de dispersion, il y a deux onglets : Données et Options. Renseignons tout d'abord la fenêtre Données :

Etape 2 :  
Choisir la variable, ici Note,  
Cliquer sur Graphe par  
groupe

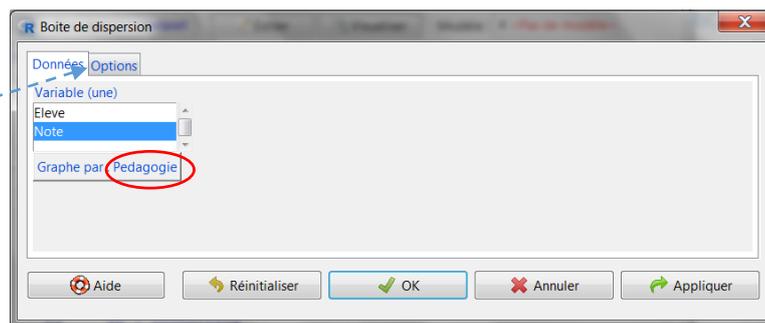


La petite fenêtre Groupes s'ouvre :  
Comme le jeu de données ne comporte qu'un seul facteur, R Commander le prend par défaut.  
Cliquons sur OK :



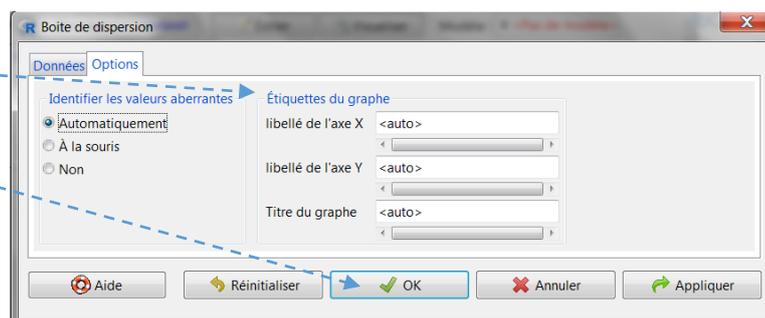
La fenêtre Boite de dispersion apparaît de nouveau

Etape 2 :  
Le facteur Pedagogie a été  
pris en compte  
Cliquer sur Options

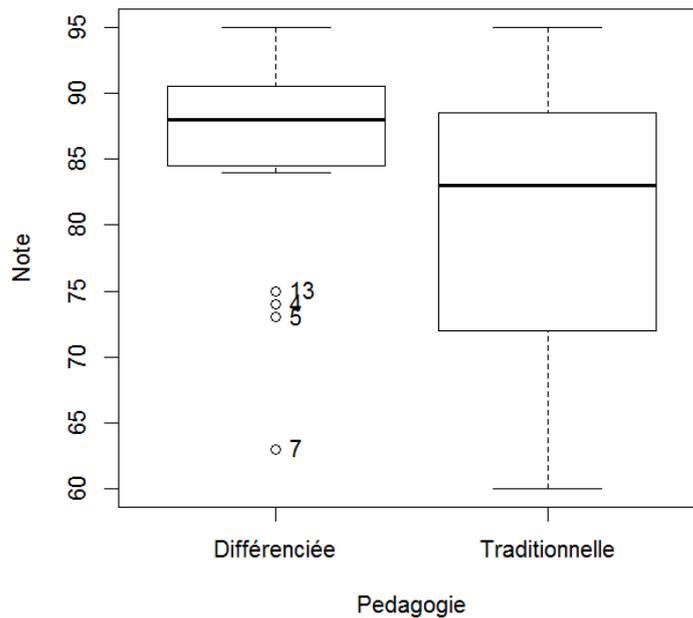


La fenêtre Options s'affiche :

Etape 3 :  
On peut choisir les étiquettes,  
OK



Les deux boîtes apparaissent côte à côte dans une fenêtre spécifique, permettant de comparer les dispersions des deux groupes de pédagogie :



On remarque des valeurs extrêmes dans le sous-groupe pédagogie différenciée. Ces valeurs apparaissent également dans la fenêtre Sortie de R Commander.

### 3.2.5 Dessiner un graphe des moyennes

Lorsque l'on souhaite comparer les moyennes de groupes, dessiner un graphe des moyennes permet de visualiser les différences entre les groupes. A titre d'exemple, prenons ici les deux groupes de pédagogie, différenciée et traditionnelle

Etape 1 :  
Sélectionnez Graphes,  
Boîte de dispersion

```

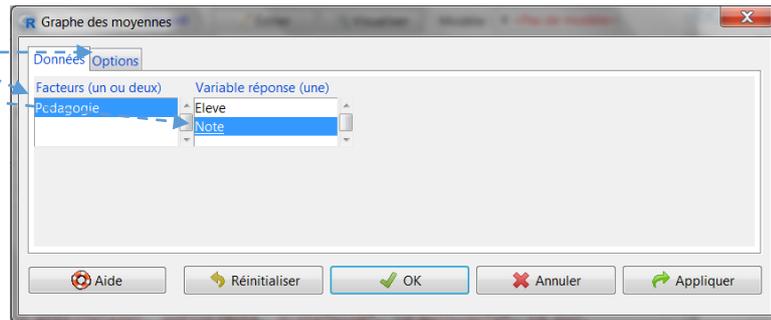
with(Dataset, qqPlot(Note,
  labels=rownames(Dataset))
Boxplot(Note~Pedagogie, dat
Sortie
13 7. | 556
16 8* | 244
(12) 8. | 55567778999
12 9* | 0000011234
2 9. | 55

> with(Dataset, qqPlot(Note
+ labels=rownames(Dataset)
24 40
1 2

> Boxplot(Note~Pedagogie, d
[1] "4" "5" "7" "13"
    
```

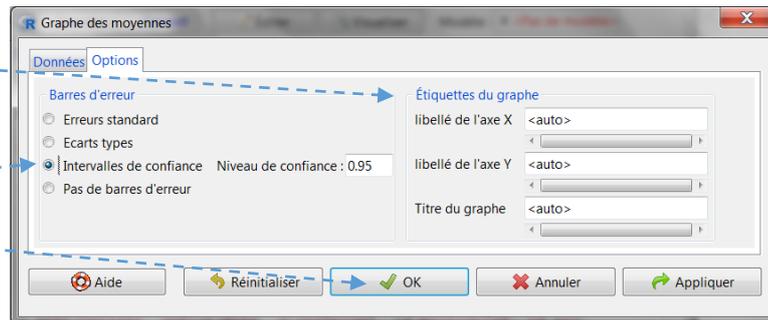
Dans la fenêtre Graphe des moyennes, il y a deux onglets : Données et Options. Renseignons tout d'abord la fenêtre Données :

Etape 2 :  
 Choisir le facteur (ici Pédagogie)  
 Choisir la variable, (ici Note)  
 Cliquer Option

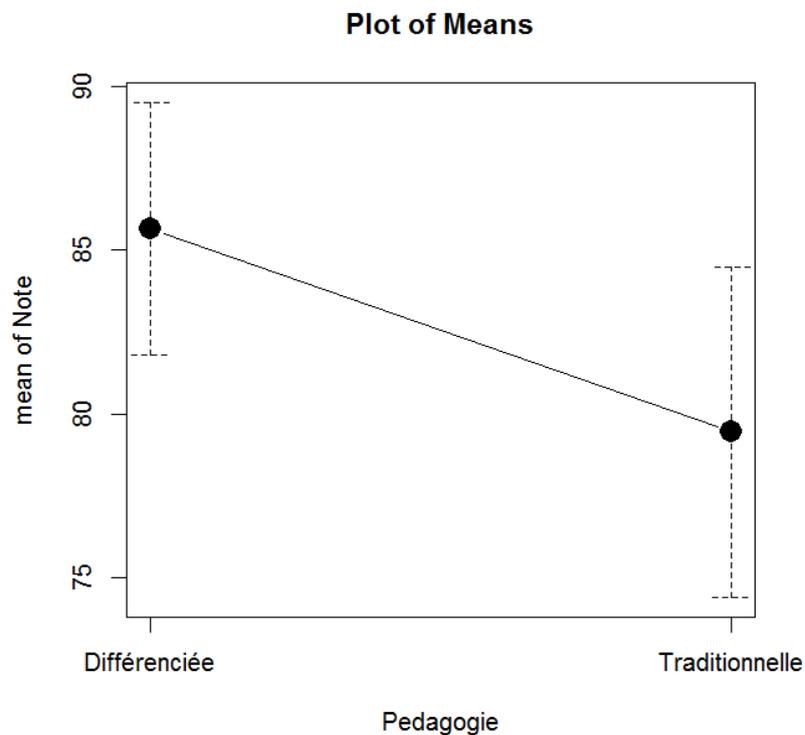


La fenêtre Options s'affiche :

Etape 3 :  
 On peut choisir les étiquettes,  
 On peut visualiser à l'aide  
 d'une barre d'erreur certaines  
 statistiques. Ici Intervalles de  
 confiance  
 OK



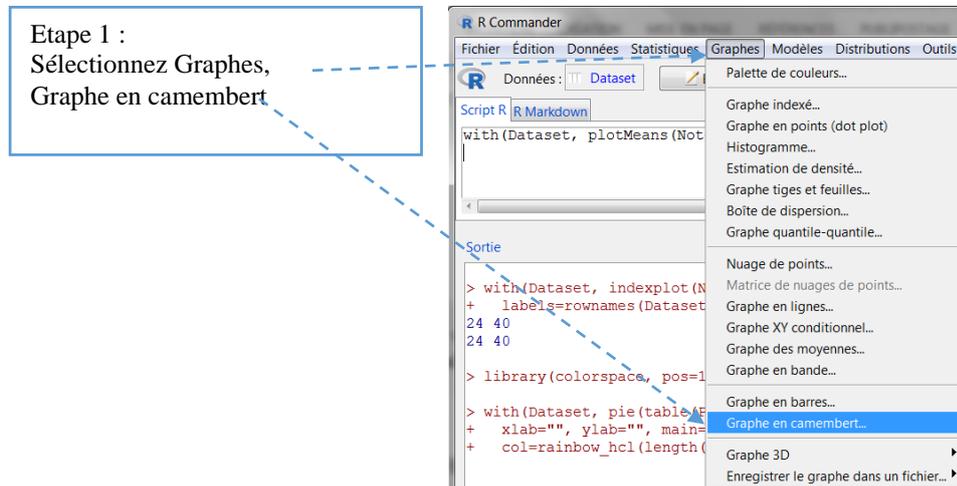
Le graphe des moyennes s'ouvre dans une fenêtre spécifique, permettant de comparer visuellement les deux moyennes et les intervalles de confiances :



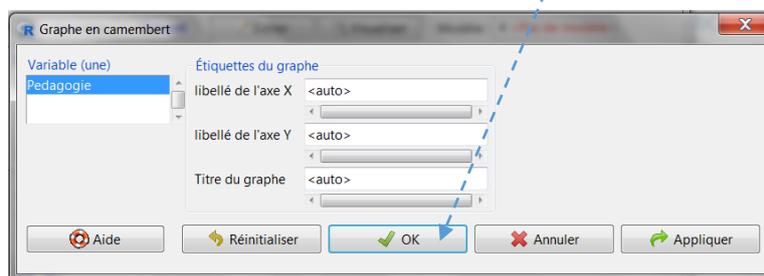
La moyenne des notes du groupe pédagogie différenciée est supérieure à celle du groupe pédagogie traditionnelle.

### 3.2.6 Dessiner un graphe en camembert

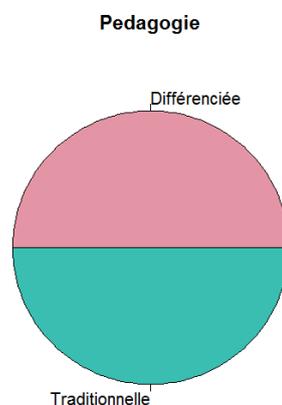
Lorsque l'on souhaite comparer les effectifs de groupes ou de catégories, on peut dessiner un graphe en camembert pour visualiser leur taille respective. Ici nos deux groupes (différenciée vs traditionnelles) sont équilibrés, donc le seul intérêt est de voir comment dessiner un camembert...



La fenêtre Graphe en camembert s'ouvre. Comme le jeu actif ne comporte qu'un seul facteur, R Commander le sélectionne par défaut. Cliquons sur OK.

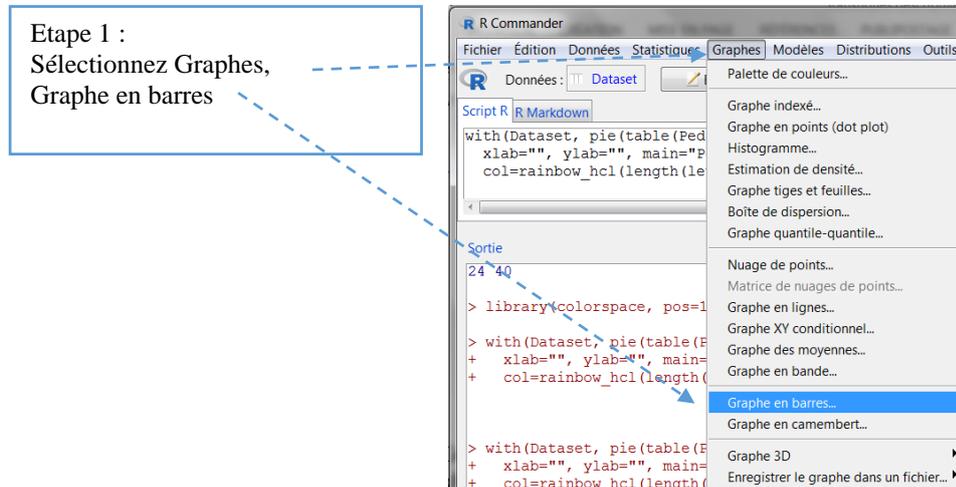


Le graphe en camembert apparaît dans une fenêtre spécifique :

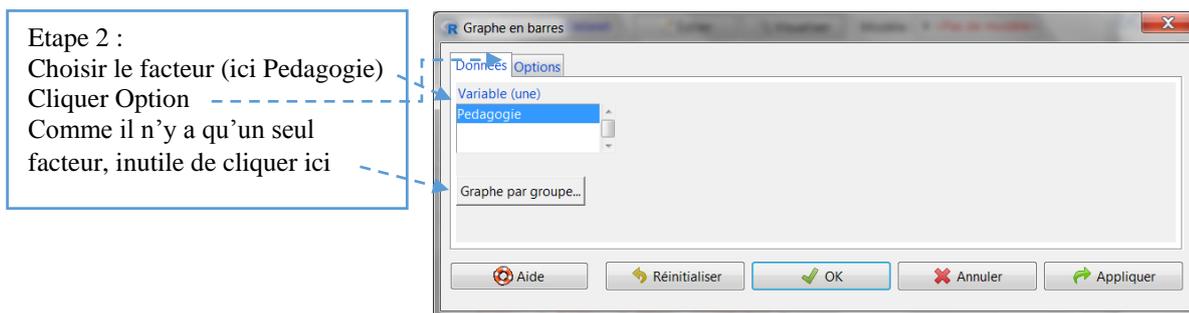


### 3.2.7 Dessiner un graphe en barres

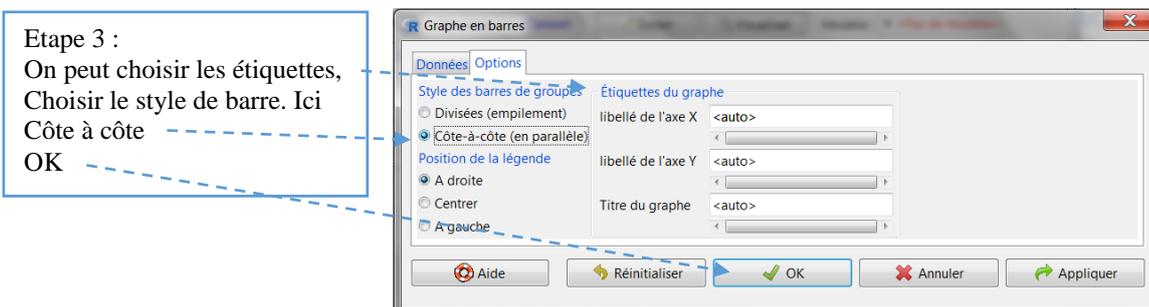
Pour comparer les effectifs de groupes ou de catégories, on peut aussi dessiner un graphe en barres. Comme nos deux groupes (différenciée vs traditionnelles) sont équilibrés, le seul intérêt est encore de voir comment dessiner ce graphe...



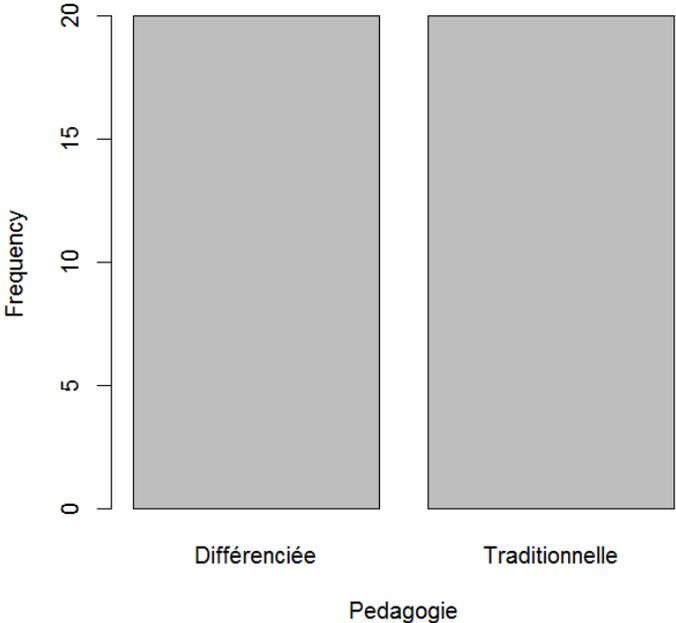
La fenêtre Graphe en barre s'ouvre.



La fenêtre Options s'affiche :



Le graphe barres s'ouvre dans une fenêtre spécifique, permettant de comparer visuellement la taille des deux groupes :



## 4 Tableaux croisés et Chi-deux d'indépendance (*Chi-square test of independence*)

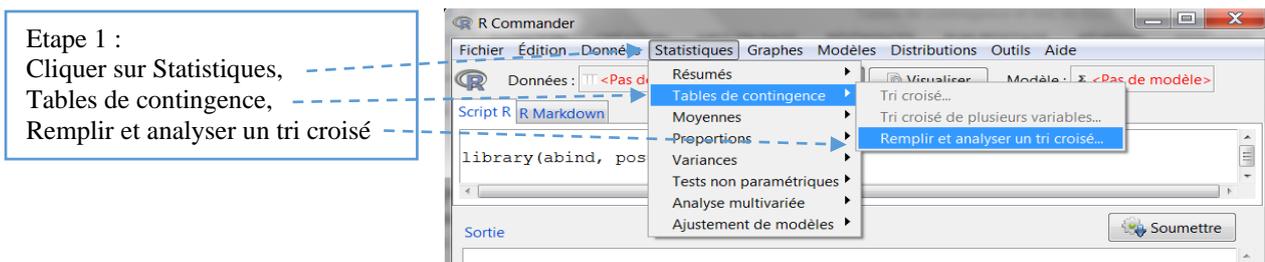
R Commander offre la possibilité de calculer le test du Chi-deux à partir de 2 sources de données différentes : un tableau croisé (ou table de contingence) ou un jeu de données.

### 4.1 Calculer un Chi-deux à partir d'un tableau croisé

Une conseillère d'orientation souhaite savoir s'il existe une relation entre l'orientation des élèves du lycée de son secteur et leur baccalauréat. Voici les données dont elle dispose :

	Bac général	Bac technologique
Université	180	35
Lycée pro-IUT	55	72
Autre	20	8

#### Calcul du Chi-deux

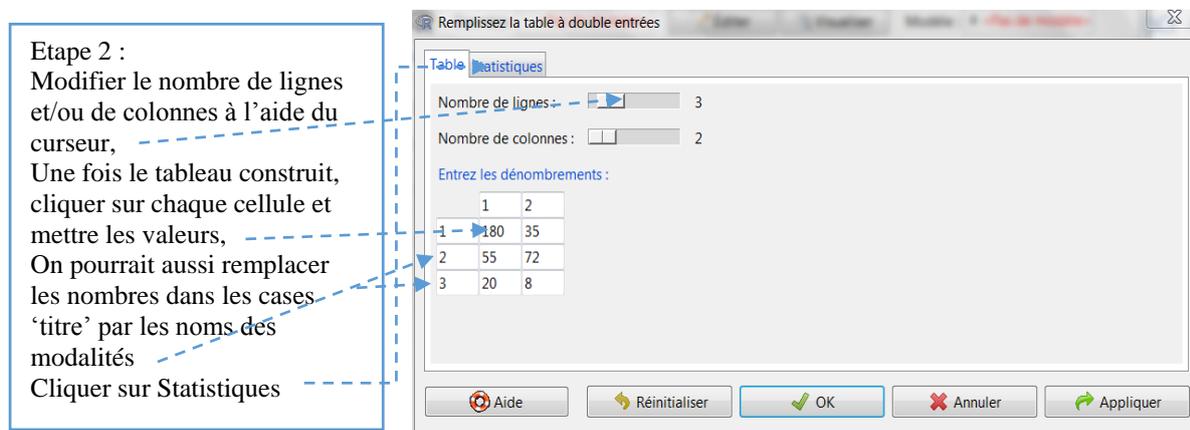


La fenêtre Remplissez la table à double entrées apparaît.

On spécifie le nombre de lignes (2 à 10) et le nombre de colonne (2 à 10) en faisant glisser le curseur à l'aide de la souris. Le tableau s'agrandit en tenant compte des modifications des nombres de lignes et de colonnes.

Il suffit ensuite de compléter chaque case du tableau ainsi créé avec les valeurs que l'on a. On pourrait aussi mettre des titres pour les colonnes et pour les lignes en écrivant dans les cellules Titre.

Dans notre exemple, on obtient alors :



L'onglet Statistiques apparaît. Il offre la possibilité de calculer les pourcentages ainsi que des éléments du Chi-2 et le test de Fischer.

Etape 3 :  
 Choisir le calcul des pourcentages (ici Pourcentages des colonnes)  
 Choisir le test d'hypothèse (ici, le Chi-deux)  
 OK

La fenêtre Résultats :

Résultats  
 Tableau croisé avec les fréquences  
 Tableau avec les pourcentages des colonnes  
 Glisser le curseur pour afficher la suite

```

> colPercents(.Table) # Column Percentages
      1 2
1 180 35
2  55 72
3  20  8

> colPercents(.Table) # Column Percentages
      1      2
1  70.6  30.4
2  21.6  62.6
3   7.8   7.0
Total 100.0 100.0
Count 255.0 115.0
    
```

Résultats (suite)  
 Valeur du Chi-deux  
 Degrés de liberté  
 Seuil de signification

```

> .Test
      Pearson's Chi-squared test

data:  .Table
X-squared = 60.9645, df = 2, p-value = 5.777e-14

> remove(.Test)
> remove(.Table)
    
```

Le test est significatif à .0005.

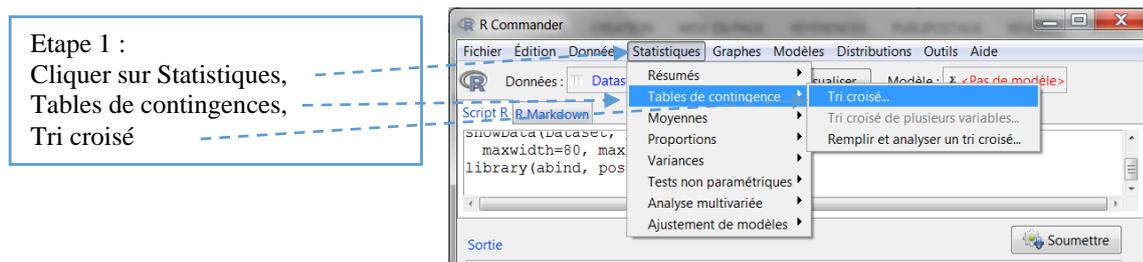
## 4.2 Calculer un Chi-deux sur un jeu de données

Un magazine de voyage et découverte effectue un sondage auprès de ses lecteurs afin de savoir si leur âge, jeunes (moins de 30 ans) ou séniors (plus de 50 ans), a une relation avec ce qui les intéresse le plus dans les reportages proposés, les textes ou les photos. Le tableau ci-dessous présente les données récoltées auprès de 40 lecteurs. Pour l'âge, les jeunes lecteurs sont codés 1, les séniors 2. A la question « Etes-vous plus intéressé par les photos ou par les textes ? », la réponse 'Plus intéressé par les photos' est codée 1, 'Plus intéressé par les textes' est codée 2. Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableaux de données) et convertir la variable Age en facteur, avec 1 = jeune et 2 = sénior ; la variable Intérêt avec 1 = photo et 2 = texte.

Pour gagner de la place, le tableau a été scindé en 2, mais il comporte bien 3 colonnes et 40 lignes (hors la ligne des titres).

Lecteur	Age	Intérêt	(suite)	Lecteur	Age	Intérêt
1	1	1		21	2	2
2	1	1		22	2	2
3	1	1		23	2	2
4	1	1		24	2	2
5	1	1		25	2	2
6	1	1		26	2	2
7	1	1		27	2	2
8	1	1		28	2	2
9	1	1		29	2	2
10	1	1		30	2	2
11	1	1		31	2	2
12	1	1		32	2	2
13	1	2		33	2	1
14	1	2		34	2	1
15	1	2		35	2	1
16	1	2		36	2	1
17	1	2		37	2	1
18	1	2		38	2	1
19	2	2		39	2	1
20	2	2		40	2	1

Pour effectuer le test du Chi-deux



La fenêtre Table de contingence à double entrée apparaît :

Etape 2 :  
Sélectionner la variable en colonne (VI),  
Sélectionner la variable en ligne (VD),  
Cliquer sur l'onglet Statistiques

La fenêtre de l'onglet Statistiques permet de calculer les pourcentages ainsi que des éléments du Chi-deux et du test de Fischer. Dans cet exemple, nous ne calculerons pas de pourcentage.

Etape 3 :  
Sélectionner Pas de pourcentages,  
Sélectionner Test Chi-deux,  
Cliquer OK

La fenêtre de résultats s'ouvre :

Résultats  
Tableau croisé avec les fréquences,  
Significatif au seuil :  
Degrés de liberté :  
Valeur du test du Chi-deux,

## 5 La corrélation avec le r de Bravais-Pearson (*correlation with Pearson's r*)

Ce test paramétrique est utilisé lorsque l'on souhaite apprécier le sens et la force d'une relation linéaire entre deux variables d'intervalles. Les deux distributions de scores se distribuent normalement.

Un enseignant a mesuré l'efficacité d'un renforcement en statistiques en comparant les moyennes des notes de ses 20 élèves, avant et après le renforcement. Son groupe étant composé de filles et de garçons, il a également relevé le sexe en codant 1 pour les garçons et 2 pour filles. Il souhaite apprécier la force de la relation entre les deux temps d'observation, Avant et Après. Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données...) et convertir la variable Sexe en facteur, avec 1 = Garçon et 2 = Fille.

Ces données sont celles utilisées dans le chapitre Comparaison de moyenne de groupes appariés avec le *t* de Student. Elles ont été scindées en 2 pour gagner de la place, le tableau comporte bien 4 colonnes et 20 lignes (hors ligne titre).

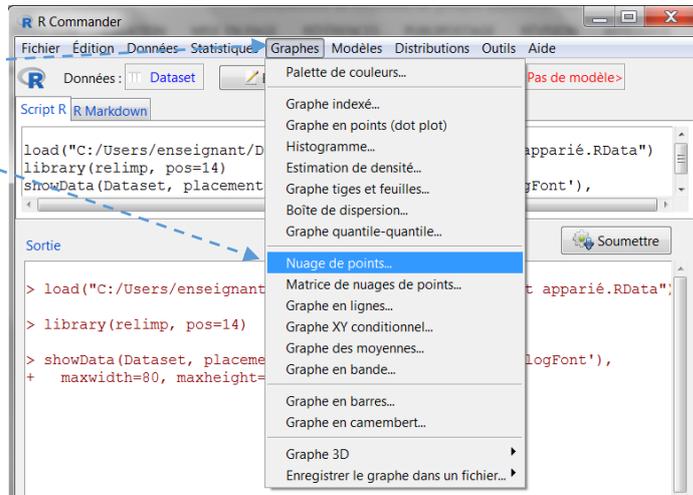
Participant	Avant	Après	Sexe	Participant	Avant	Après	Sexe
1	31	34	1	11	31	28	1
2	26	25	1	12	27	32	1
3	32	38	1	13	25	25	1
4	38	36	1	14	28	30	1
5	29	29	1	15	32	41	1
6	34	41	2	16	27	37	2
7	24	26	2	17	37	39	2
8	35	42	2	18	29	33	2
9	30	36	2	19	31	40	2
10	36	44	2	20	27	28	2

Nous allons procéder en 3 temps :

1. vérifier la nature linéaire de la relation entre les deux observations Avant et Après ;
2. vérifier que les observations Avant et Après se distribuent normalement ;
3. calculer la force et le sens de cette relation linéaire et effectuer le test d'hypothèse.

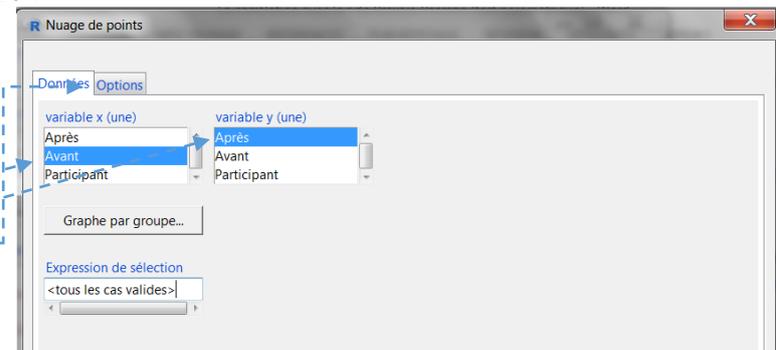
### 5.1 Vérifier la nature linéaire de la relation entre les deux observations Avant et Après

Etape 1 :  
Cliquez sur Graphes,  
Nuage de points

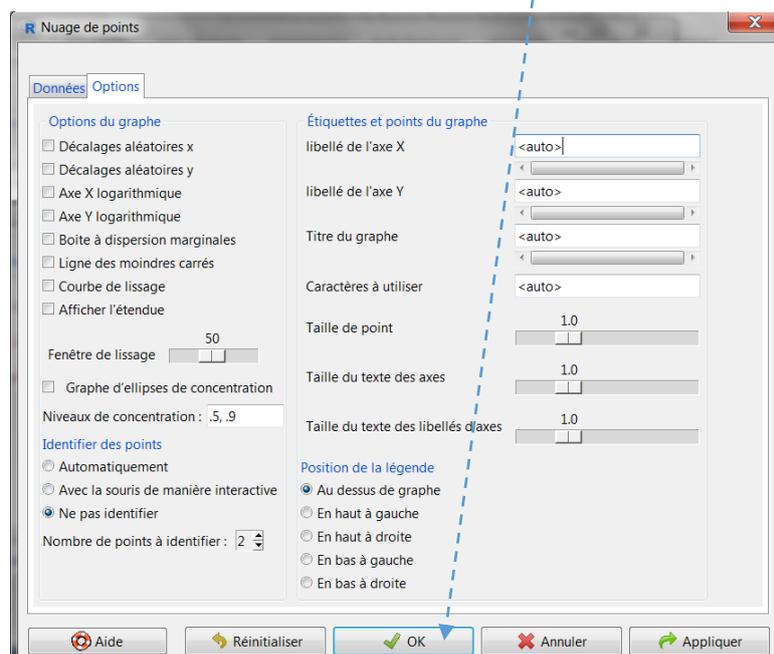


La fenêtre Nuage de points s'ouvre :

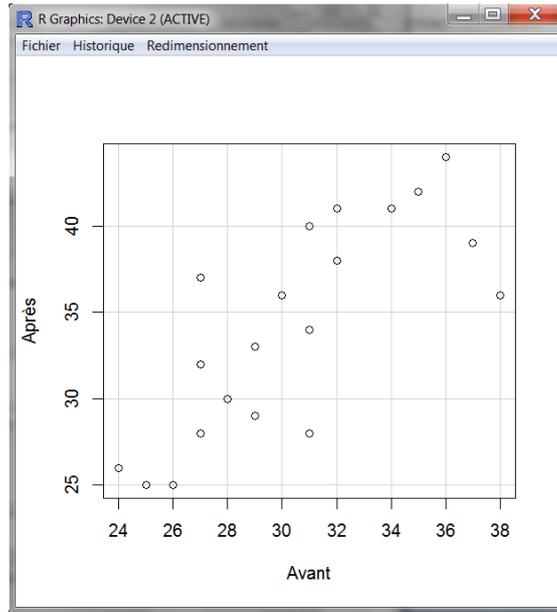
Etape 2 :  
Choisir la variable X, ici, Avant,  
Choisir la variable Y ici, Après,  
Cliquez sur Option



La fenêtre Option s'ouvre. Elle offre diverses options sur le graphe et sa présentation. Nous n'en prendrons aucune et cliquons directement sur OK :



Le nuage de points apparaît :



Le nuage de points ayant une forme allongée, on peut considérer qu'il y a une relation de nature linéaire entre les deux variables Avant et Après. La corrélation peut être calculée.

### 5.2 Vérifier que les variables Avant et Après se distribuent normalement

Commençons par vérifier la normalité de la distribution des scores pour l'observation Avant :

Etape 1 :  
Sélectionnez Statistiques,  
Résumés,  
Test de normalité de Shapiro-  
Wilk

La fenêtre Shapiro-Wilk Test for Normality apparaît:

Etape 2 :  
Sélectionnez Avant  
Cliquez sur OK

La fenêtre des résultats apparaît :

Résultats  
Variable Avant :  
Valeur du test de normalité,  
Significatif au seuil :

```
Sortie
> load("C:/Users/enseignant/Desktop/Rcmdr/Chap 9 t test apparié.RData")
> with(Dataset, shapiro.test(Avant))

      Shapiro-Wilk normality test

data:  Avant
W = 0.9646, p-value = 0.6381

Messages
[1] NOTE: R Commander Version 2.1-7: Sun Jan 03 16:42:55 2016
[2] NOTE: Le jeu de données Dataset a 20 lignes et 4 colonnes.
```

Les scores de la variable Avant se distribuent normalement.

Il faut recommencer la procédure ci-dessus pour la seconde variable Après. On peut cependant procéder encore plus rapidement : en effet, il suffit de modifier directement la dernière ligne de commande en remplaçant le nom de la première observation (Avant), par le nom de la seconde (Après), de surligner la commande et de soumettre.

Etape 1 :  
Remplacer 'Avant' par 'Après'  
directement dans le texte ;  
Surligner la commande avec la  
souris ;  
Cliquer sur Soumettre.

Voici les résultats :

Résultats  
Variable Après :  
Valeur du test de normalité,  
Significatif au seuil :

```
Sortie
data: Avant
W = 0.9646, p-value = 0.6381

> load("C:/Users/enseignant/Desktop/Rcmdr/Chap 9 t test apparié.RData")
> with(Dataset, shapiro.test(Après))

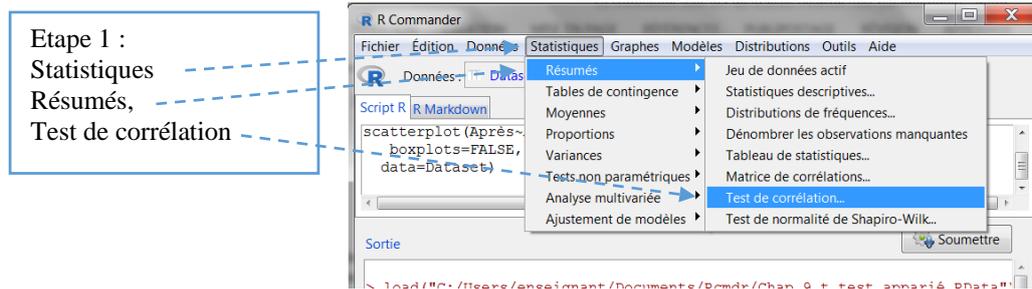
      Shapiro-Wilk normality test

data:  Après
W = 0.9438, p-value = 0.2832

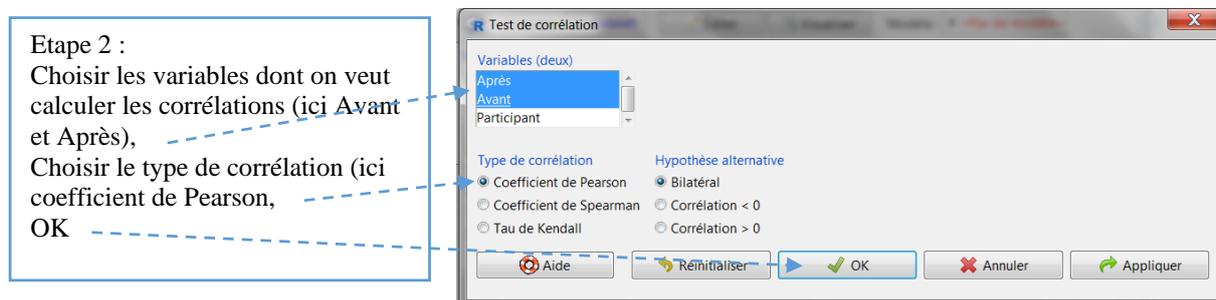
Messages
[1] NOTE: R Commander Version 2.1-7: Sun Jan 03 16:42:55 2016
[2] NOTE: Le jeu de données Dataset a 20 lignes et 4 colonnes.
```

L'hypothèse nulle (non différence avec la loi normale) est également acceptée pour la seconde variable, nous pouvons dès lors utiliser le test paramétrique  $r$  de Bravais Pearson (Rappel : si au moins l'une des deux distributions n'avait pas été normale (rejet de l'hypothèse nulle au seuil choisi), il aurait alors fallu utiliser un test non paramétrique, par exemple, le rho de Spearman).

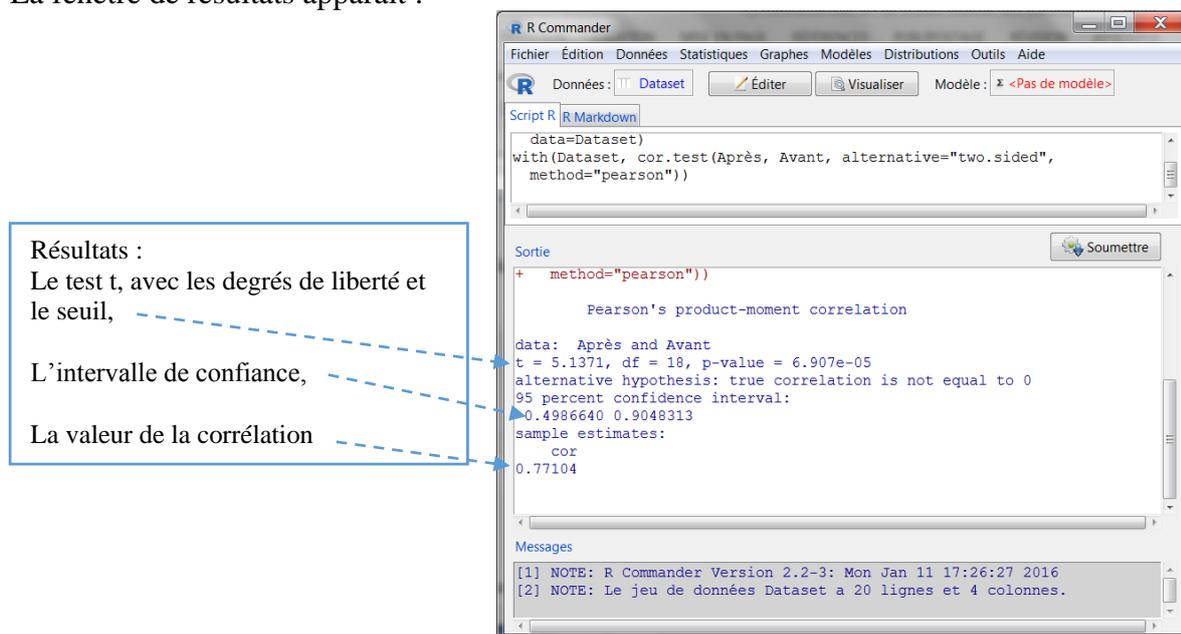
### 5.3 Calculer la corrélation entre les deux variables



La fenêtre Test de corrélation s'ouvre :



La fenêtre de résultats apparaît :



Avec  $r = 0,77$ , nous avons une relation linéaire très forte et positive, significative à  $.0005$ .

### 5.4 Calcul de la corrélation pour sous-groupes

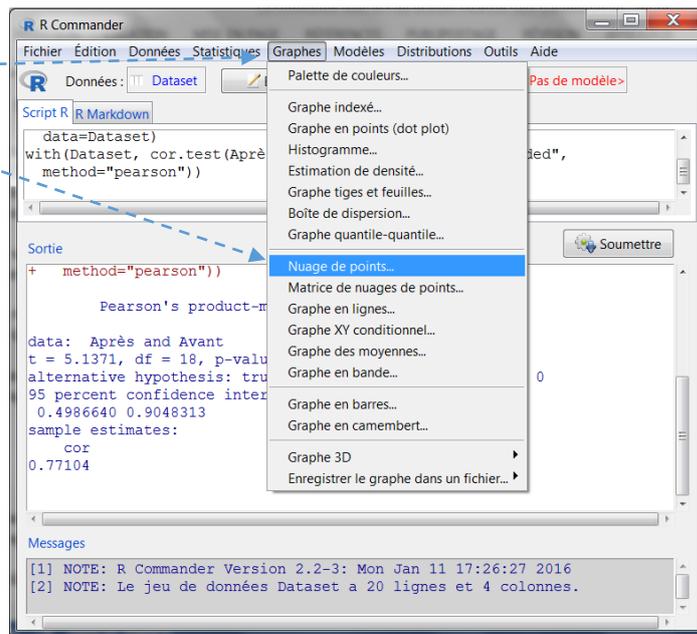
Il arrive souvent que l'on souhaite apprécier la relation entre deux variables non pas pour l'ensemble de l'échantillon, mais selon des sous-groupes qui nous intéressent, par exemple les garçons ou les filles. La démarche est très simple sur R Commander, mais elle nécessite d'ajouter quelques mots sur le script de la corrélation.

Cet enseignant voudrait ainsi savoir quels sont le sens et la force de la relation entre les scores avant et après le renforcement pour le sous-groupe des filles.

Les trois grandes étapes vues ci-dessus sont les mêmes ici : vérifier la nature linéaire de la relation, puis la normalité des distributions, enfin procéder au calcul de la corrélation.

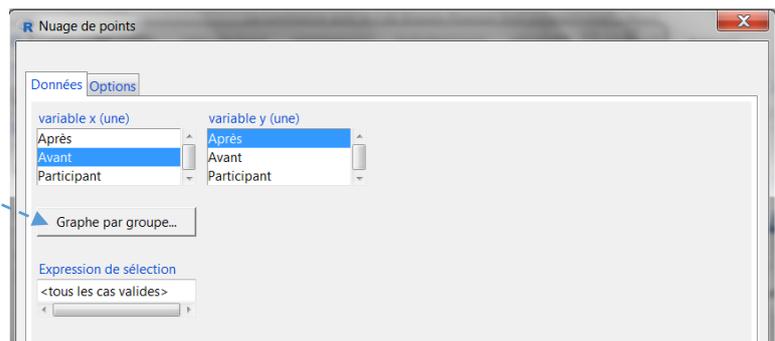
### 5.4.1 Vérifier la nature linéaire de la relation entre Avant et Après pour le sous-groupe des filles

Etape 1 :  
Sélectionnez Graphes,  
Nuage de points



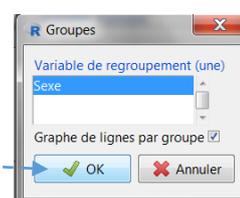
La fenêtre Nuage de points s'ouvre comme nous l'avons laissée :

Etape 2 :  
La variable X est déjà sélectionnée  
La variable Y est déjà sélectionnée  
Cliquer sur Graphe par groupe

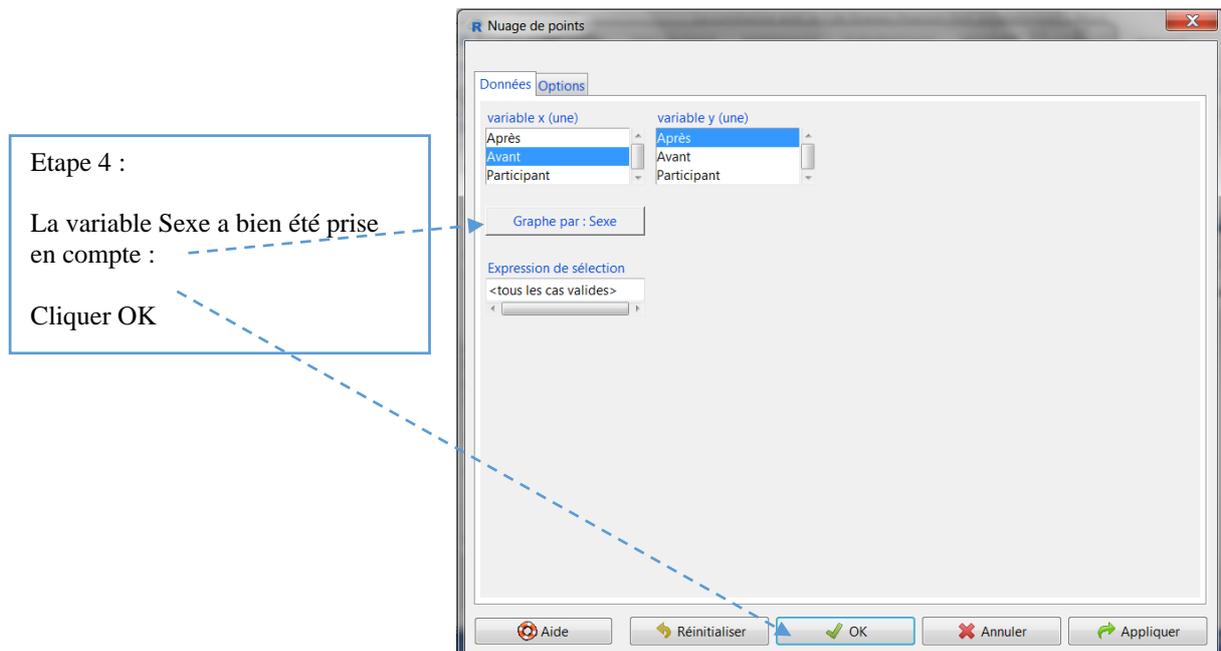


Une petite fenêtre s'ouvre qui permet de choisir la variable de regroupement. Dans cet exemple, il n'y a que la variable Sexe de disponible, elle est déjà sélectionnée.

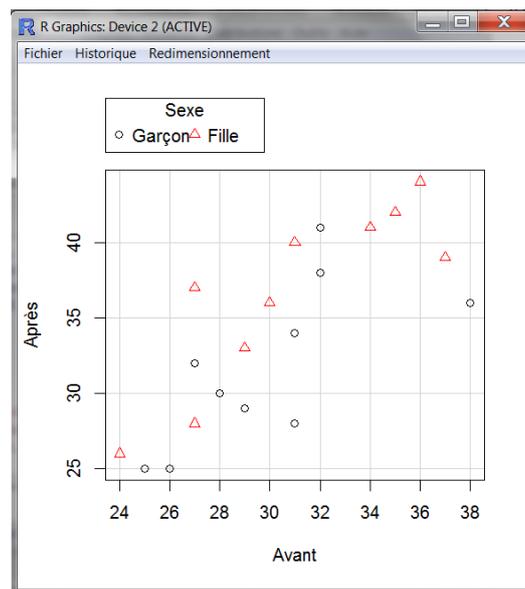
Etape 2 :  
La variable Sexe est déjà sélectionnée  
Cliquer sur OK



La fenêtre Nuage de point revient :



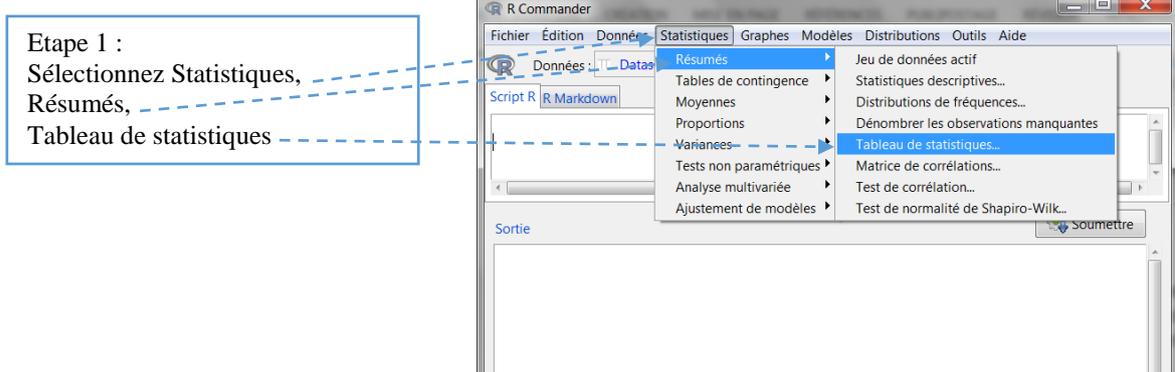
Le nuage de point apparaît. Les valeurs ont été différenciées selon le groupe d'appartenance, Fille vs Garçon :



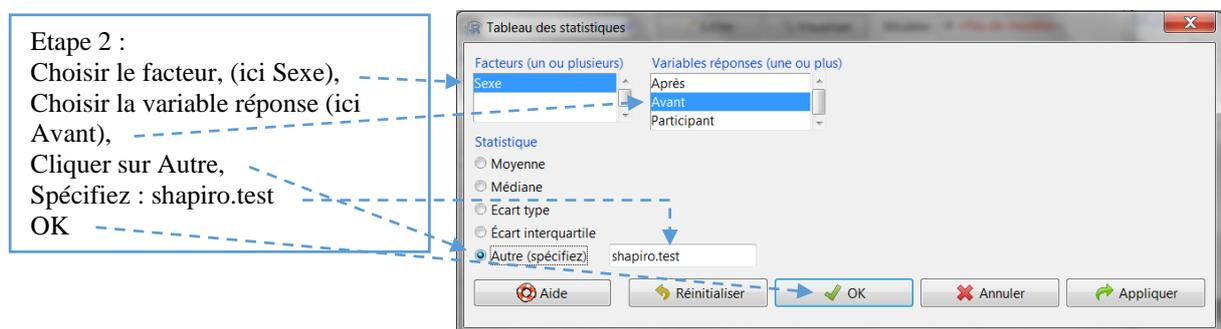
Le nuage de points du sous-groupe Fille ayant une forme allongée, on peut considérer qu'il y a une relation de nature linéaire entre les deux variables. La corrélation peut être calculée.

#### 5.4.2 Vérifier la normalité des deux distributions des scores Avant et Après pour ce sous-groupe des filles

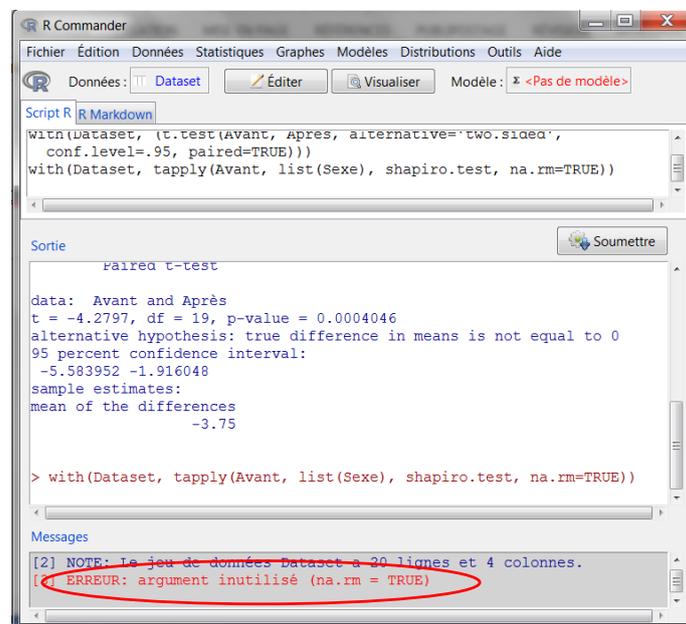
R Commander permet de tester la normalité d'une distribution avec le test de Shapiro-Wilk (Statistiques>Résumés>Test de normalité de Shapiro-Wilk), nous venons de le voir. Mais quand il s'agit, comme ici, de sous-groupes, le plus rapide est de passer par la commande Tableau de statistiques et de procéder comme suit :



La fenêtre Tableau de statistiques apparaît



La fenêtre R Commander apparaît avec tout en bas un **message d'erreur en rouge**, indiquant un argument inutilisé :

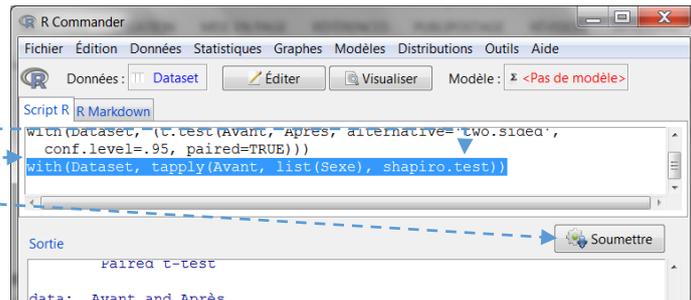


Nous allons modifier la dernière ligne de programmation dans la fenêtre des commandes en retirant cet argument non utilisé (`na.rm = TRUE`).

Nous effaçons l'argument inutile, c'est-à-dire ce qui se trouve après `shapiro.test` :

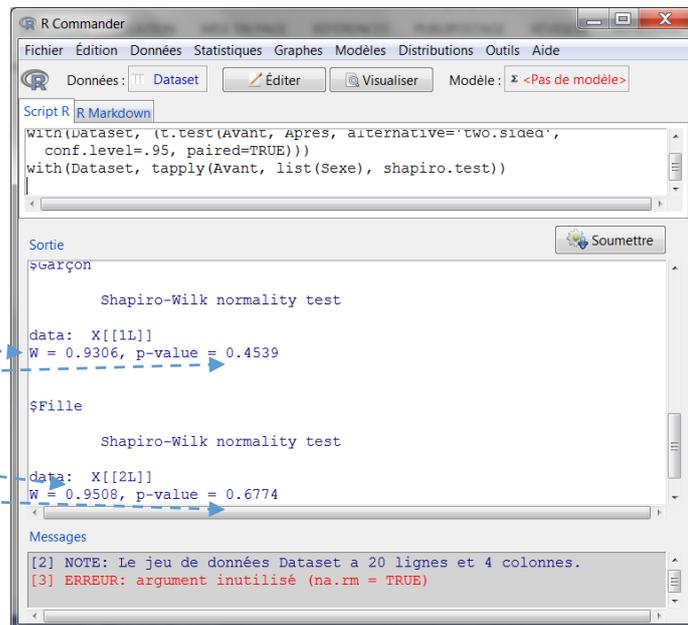
la ligne de commande devient : `with(Dataset, tapply(Avant, list(Sexe), shapiro.test))`

Etape 3 :  
 Effacer ‘, na.rm=TRUE’ directement dans le texte (ne pas oublier la virgule avant shapiro.test),  
 Surligner la ligne avec la souris,  
 Cliquer sur Soumettre.



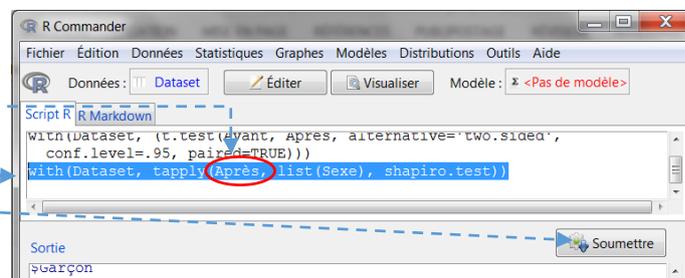
Dans la fenêtre sortie, les valeurs du test apparaissent, d’abord pour la modalité Garçon, puis pour la modalité Fille :

Résultats pour Avant  
 Garçon :  
 Valeur du test de normalité,  
 Significatif au seuil :  
  
 Fille :  
 Valeur du test de normalité,  
 Significatif au seuil :



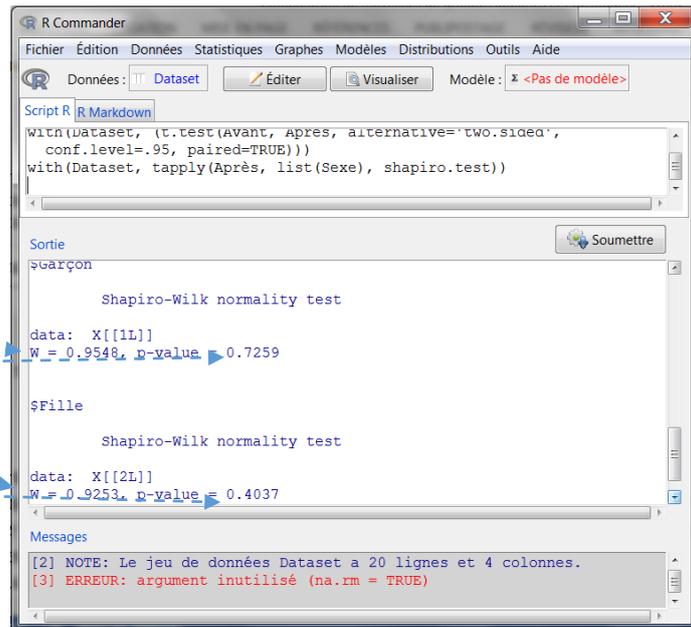
Pour les scores de la variable Avant, l’hypothèse nulle d’une non différence avec la loi normale n’est pas rejetée au seuil .05 pour chaque modalité de la variable Sexe, donc du sous-groupe Filles qui nous intéresse dans cet exemple. Il faut procéder de même pour les scores Après, soit en déroulant la même procédure que ci-dessus, soit, plus rapidement, en remplaçant dans la dernière ligne de la commande, Avant par Après.

Etape 1 :  
 Remplacer ‘Avant’ par ‘Après’ directement dans le texte ;  
 Surligner la commande avec la souris ;  
 Cliquer sur Soumettre



La fenêtre des résultats apparaît :

Résultats pour Après  
 Garçon :  
 Valeur du test de normalité,  
 Significatif au seuil :  
 Fille :  
 Valeur du test de normalité,  
 Significatif au seuil :

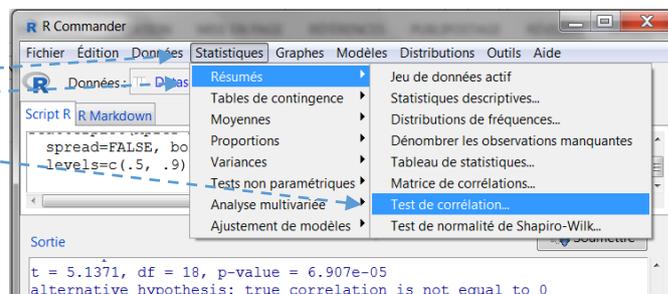


Pour les scores de la variable Après, l'hypothèse nulle d'une non différence avec la loi normale n'est pas rejetée au seuil .05 pour chaque modalité de la variable Sexe, donc du sous-groupe Fille qui nous intéresse dans cet exemple.

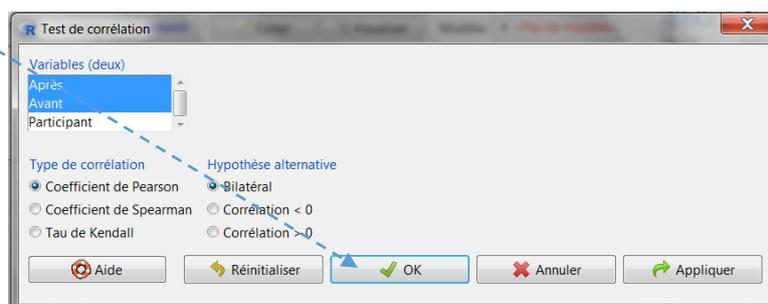
### 5.4.3 Calcul de la corrélation pour un sous-groupe

Nous allons devoir dérouler la même procédure qu'au début du chapitre pour obtenir la ligne de commande de la corrélation et ce, afin de la modifier comme nous le souhaitons.

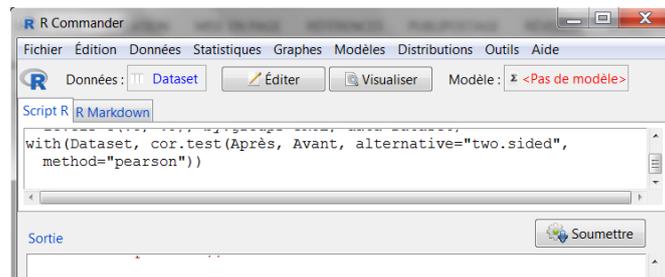
Etape 1 :  
 Cliquer sur Statistiques,  
 Résumés,  
 Test de corrélation



La fenêtre Test de corrélation s'ouvre comme nous l'avions laissée : les deux variables Avant et Après sont déjà sélectionnées, de même que le coefficient de Pearson. Il suffit de cliquer sur OK.



La même fenêtre résultats que ci-dessus apparaît. C'est dans la fenêtre Script R que nous allons modifier la commande :

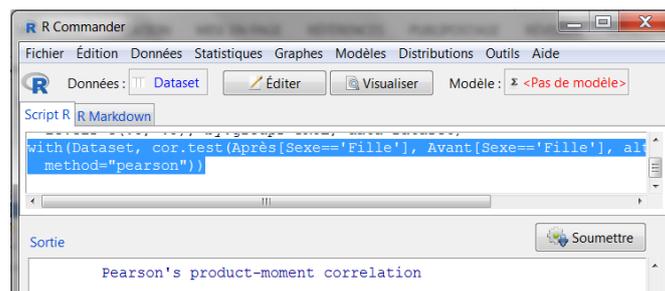


Le test a donc été calculé sur l'ensemble de l'échantillon, tous sous-groupes de la variable Sexe confondus. Nous allons maintenant spécifier le sous-groupe Fille, sur lequel va porter le test, en modifiant la dernière ligne de commande dans la fenêtre script qui est :

Ajoutons les noms de la variable Sexe et du sous-groupe Fille après chaque observation comme suit : [Sexe=='Fille'] (attention à l'absence d'espace, aux signes '=' et ==). Soit :

```
with(Dataset, cor.test(Après[Sexe=='Fille'], Avant[Sexe=='Fille'], alternative="two.sided", method="pearson")),
```

puis surlignons la commande et soumettre:



Voici la fenêtre des résultats :

Résultats pour le sous-groupe Fille

Valeur du test t, degrés de liberté et significatif au seuil

Intervalle de confiance de la corrélation

Valeur du r de Bravais-Pearson

La corrélation entre Avant et Après de 0,856 pour le sous-groupe Fille, est significative à .0016.

Si l'on souhaitait faire de même avec l'autre sous-groupe Garçon, il suffirait de remplacer Fille par Garçon dans la dernière ligne de commande, sachant que la nature linéaire de la relation ainsi que la normalité des distributions pour cette seconde modalité de la variable Sexe ont déjà été vérifiées plus haut.

## 6 La corrélation avec le rho de Spearman (*correlational analysis : Spearman's rho*)

Ce test non paramétrique est utilisé lorsque l'on souhaite apprécier le sens et la force d'une relation linéaire entre deux variables ordinales (deux classements), ou encore entre deux variables d'intervalles lorsque l'une des deux ne se distribue pas normalement.

Une formatrice en statistiques pense qu'il y a une relation entre la performance à l'examen et le temps consacré à s'entraîner. Pour tester cette hypothèse, elle met en place une aide en ligne pour les 25 étudiantes de sa classe et note en fin d'année, le temps total de connexions en heures. L'épreuve terminale est notée sur 40. Ainsi, chaque étudiante a 2 observations : le temps passé à s'entraîner (Temps) et la note à l'examen (Performance). Rentrer les données ci-dessous dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données).

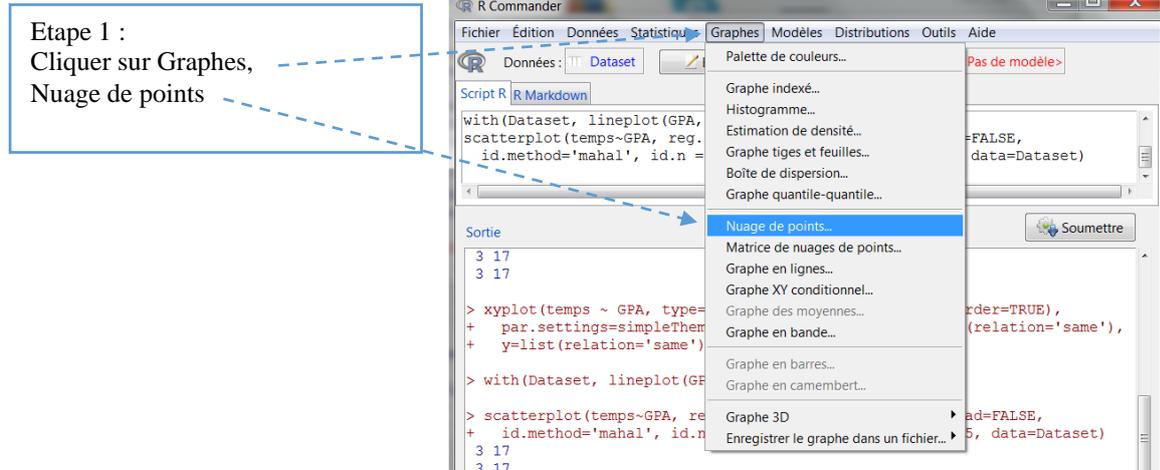
Les données ci-dessous sont les mêmes que pour la fiche Régression linéaire simple. Elles ont été scindées en 2 pour gagner de la place, le tableau comporte bien 3 colonnes et 25 lignes (hors ligne titre).

Etudiante	Temps	Performance	(suite)	Etudiante	Temps	Performance
1	2	21,2		14	2	22,2
2	5	30		15	7	35
3	15	40		16	0	16,8
4	1	20		17	14	40
5	3	25,6		18	7	38,6
6	0	17,3		19	9	40
7	7	39,1		20	1	18,4
8	12	37,7		21	5	35
9	2	21,2		22	7	37
10	6	35,5		23	4	33,3
11	10	38,5		24	10	38,8
12	8	31,2		25	3	29,8
13	5	36,8				

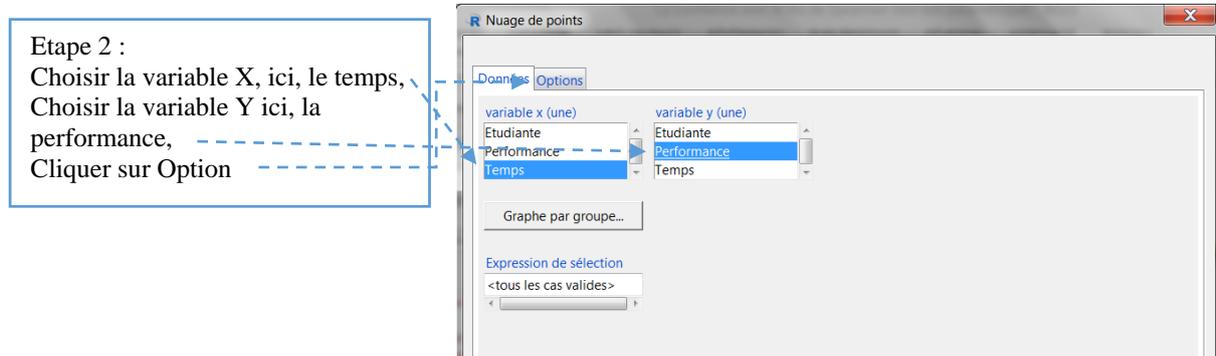
Nous allons procéder en 3 temps :

1. vérifier la nature linéaire de la relation entre les deux variables Temps et Performance ;
2. vérifier que les variables Temps et Performance se distribuent normalement
3. calculer la force et le sens de cette relation linéaire et effectuer le test d'hypothèse ;

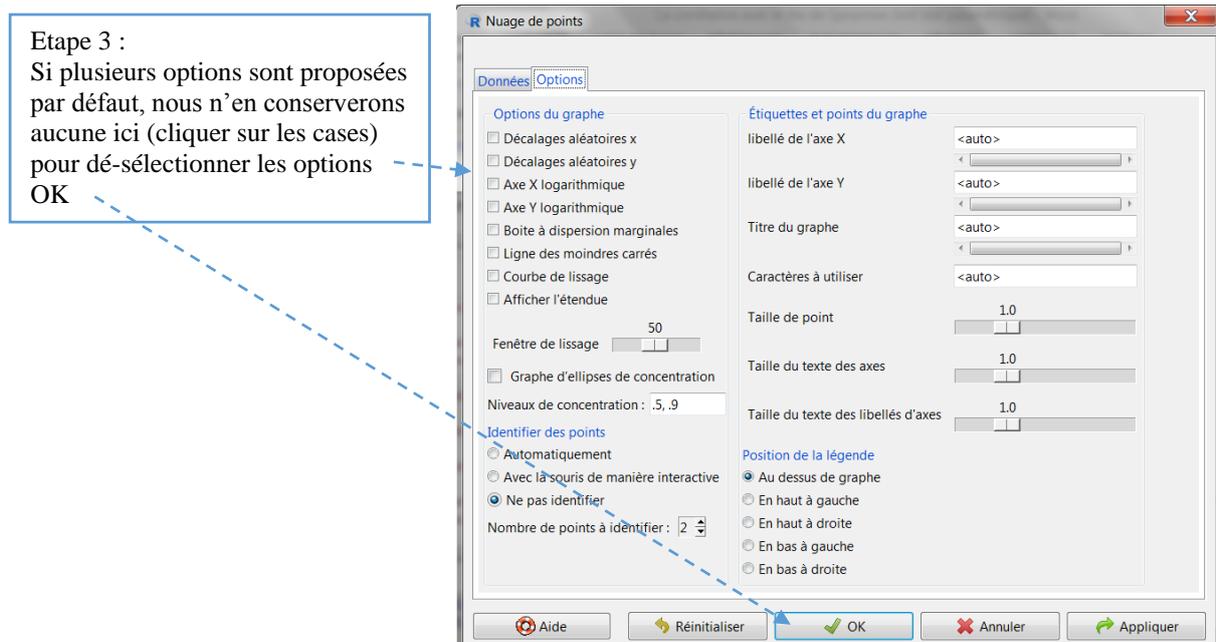
### 6.1 Vérifier la nature linéaire de la relation en observant le nuage de points



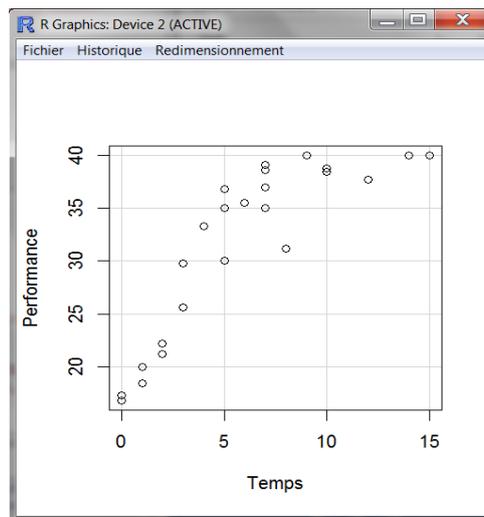
La fenêtre Nuage de points apparaît :



La fenêtre Options s'ouvre :



La fenêtre du nuage de points apparaît avec Temps en abscisse et Performance en ordonnée :



Le nuage de points ayant une forme allongée, on peut considérer qu'il y a une relation de nature linéaire entre les deux variables. La corrélation peut être calculée.

## 6.2 Vérifier que les variables Temps et Performance se distribuent normalement

Etape 1 :  
Cliquez sur Statistiques  
Résumés,  
Test de normalité de  
Shapiro-Wilk

La fenêtre Test Shapiro-Wilk de Normalité s'ouvre :

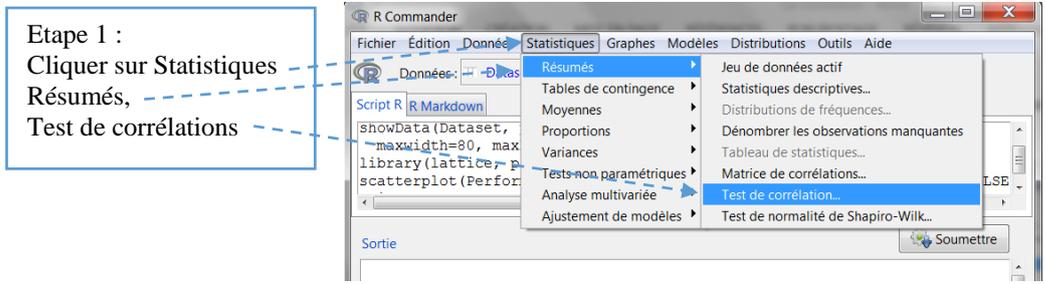
Etape 2 :  
Choisir une variable (ici  
Performance),  
OK

La fenêtre résultats apparaît :

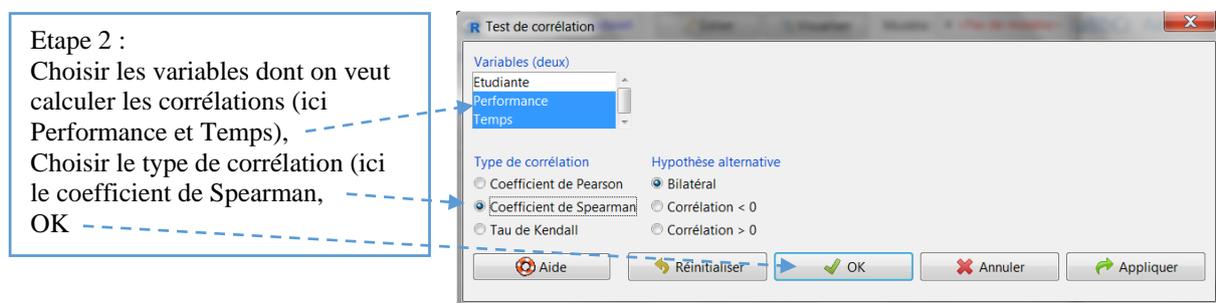
Résultats :  
Le test de normalité avec le seuil

Le test est significatif à .003, la variable Performance ne se distribue pas normalement. Il faut utiliser le test non paramétrique rho de Spearman pour apprécier la corrélation entre les variables Performance et Temps.

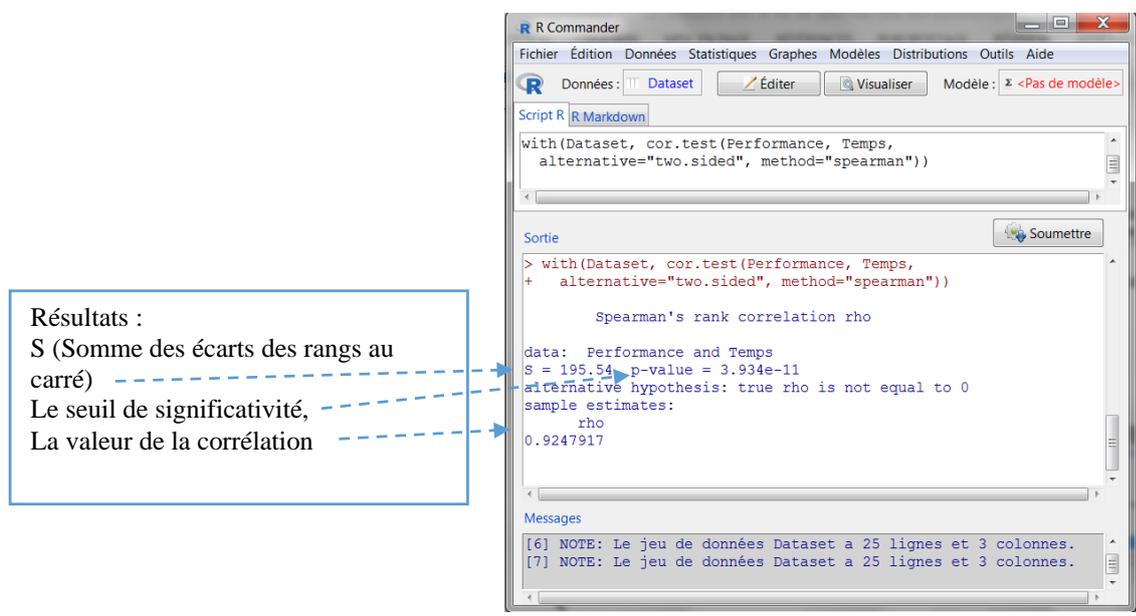
### 6.3 Calculer la corrélation entre les deux variables Temps et Performance



La fenêtre Test de corrélations apparaît :



La fenêtre des résultats apparaît :



Avec  $\rho = 0,92$ , nous avons une relation linéaire très forte et positive, significative à .0005 entre la performance à l'examen et le temps consacré à s'entraîner.

## 7 Le test t de Student pour un échantillon ou test t univarié (*one sample t-test*)

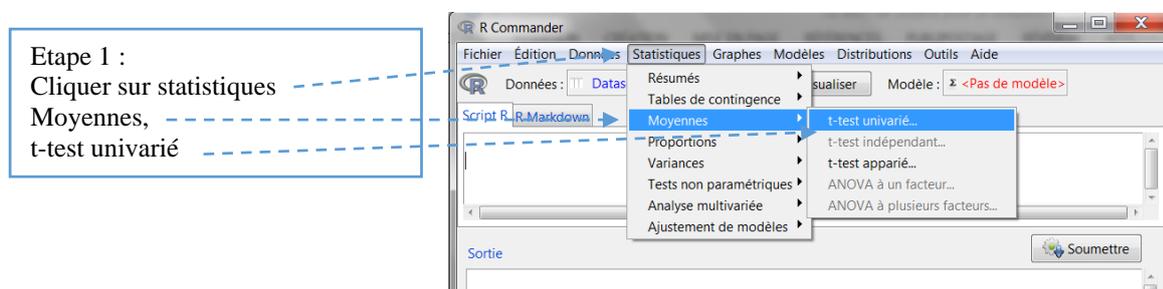
Ce test paramétrique est utilisé lorsque l'on souhaite comparer la moyenne d'un groupe à une norme. La distribution des scores de la norme se distribue normalement.

Une psychologue fait passer à une classe un test de QI. Elle constate des scores plutôt élevés et se demande si la moyenne des notes QI de la classe n'est pas différente de la norme (la moyenne des notes QI est de 100). Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données).

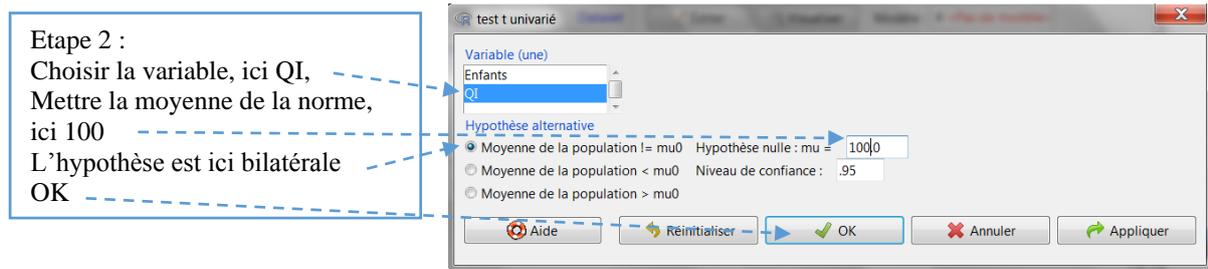
Pour gagner de la place, le tableau est scindé en deux. Mais il comporte bien 2 colonnes et 30 lignes (hors la ligne titre)

Enfants	QI	(suite)	Enfants	QI
1	110		16	110
2	106		17	100
3	115		18	120
4	125		19	125
5	115		20	120
6	103		21	110
7	109		22	105
8	100		23	104
9	99		24	122
10	107		25	112
11	120		26	107
12	112		27	124
13	102		28	98
14	105		29	117
15	110		30	110

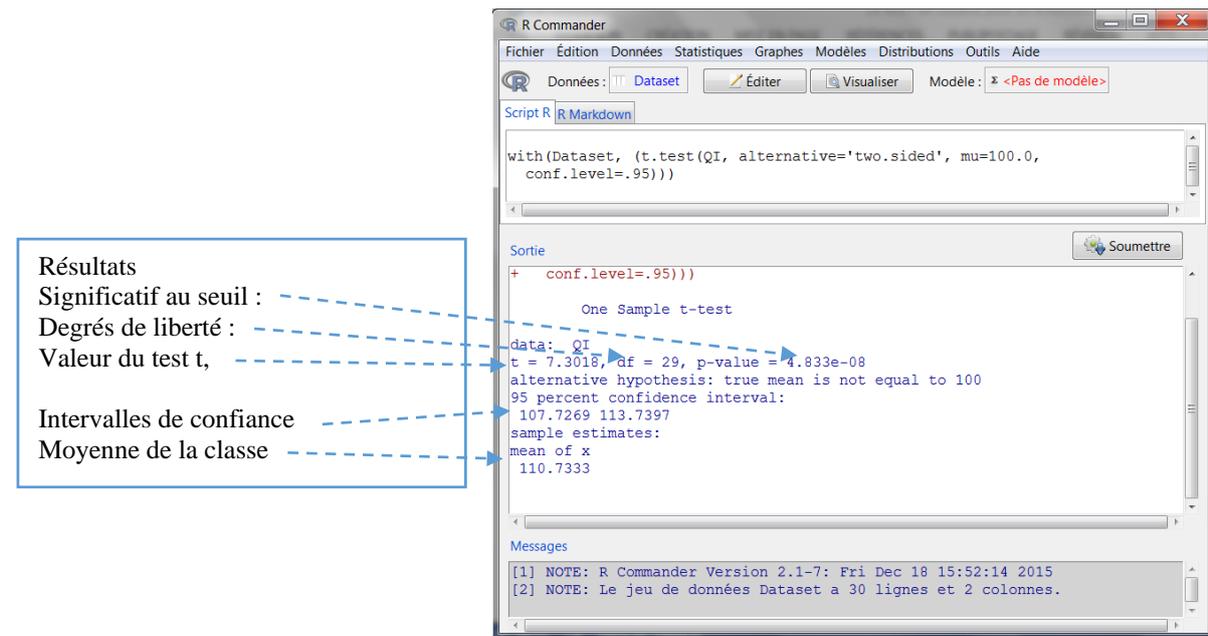
Les normes dont dispose cette psychologue se distribuent selon la loi normale (condition d'utilisation du t).



La fenêtre test t univarié s'ouvre :



La fenêtre de R Commander apparaît avec les résultats dans Sortie :



Avec  $t(29) = 7,30$ , il y a une différence significative à  $.0005$  entre la moyenne de cette classe et la norme

## 8 Le test *t* de Student pour groupes indépendants (*independent sample t-test*)

Ce test paramétrique est utilisé lorsque l'on souhaite comparer les moyennes de 2 groupes indépendants. Les deux distributions de scores dont sont issues les moyennes, se distribuent normalement.

Une enseignante souhaite tester l'efficacité de deux méthodes pédagogiques en comparant les moyennes des notes de deux groupes d'élèves : les 20 élèves qui ont eu une pédagogie dite différenciée, cotée 1 ci-dessous ; et les 20 élèves qui ont eu une pédagogie dite traditionnelle, cotée 2 ci-dessous. Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données...) et convertir la variable Pédagogie en facteur, avec 1 = Différenciée et 2 = Traditionnelle.

Ces données sont celles utilisées dans la fiche Statistiques descriptives. Elles ont été scindées en 2 pour gagner de la place, le tableau comporte bien 3 colonnes et 40 lignes (hors ligne titre).

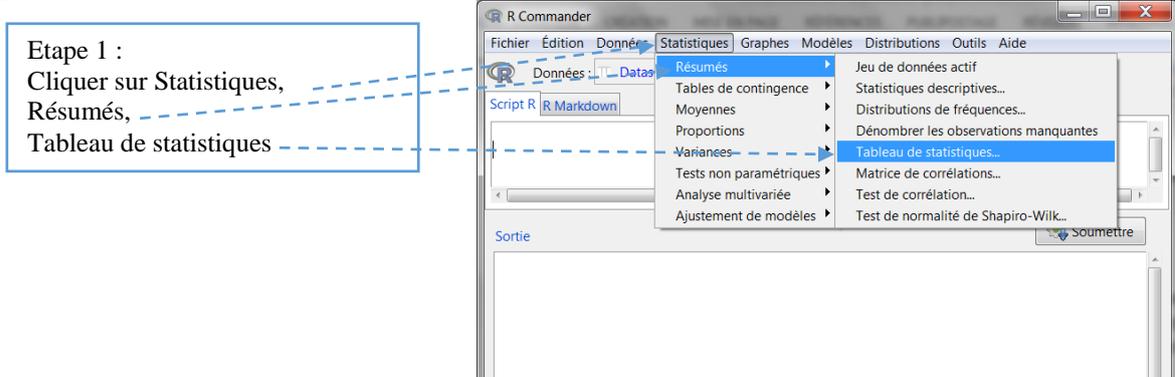
Elève	Pédagogie	Note	(Suite)	Elève	Pédagogie	Note
1	1	87		21	2	82
2	1	95		22	2	72
3	1	84		23	2	95
4	1	74		24	2	60
5	1	76		25	2	90
6	1	92		26	2	87
7	1	77		27	2	89
8	1	90		28	2	86
9	1	94		29	2	76
10	1	84		30	2	74
11	1	91		31	2	85
12	1	90		32	2	75
13	1	75		33	2	90
14	1	93		34	2	91
15	1	87		35	2	88
16	1	85		36	2	63
17	1	90		37	2	78
18	1	89		38	2	72
19	1	87		39	2	84
20	1	85		40	2	60

Nous allons procéder en 3 temps :

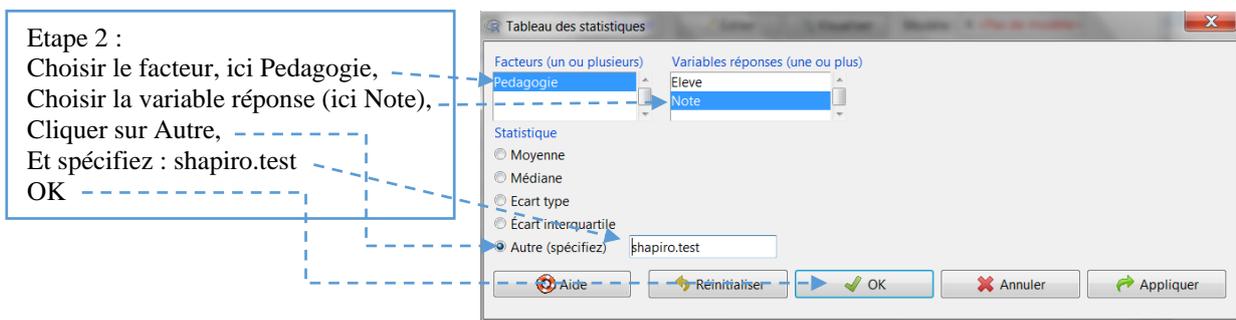
1. vérifier la normalité des distributions des 2 groupes ;
2. apprécier l'homogénéité des variances des 2 groupes ;
3. calculer le test d'hypothèse

## 8.1 Vérifier la normalité des groupes

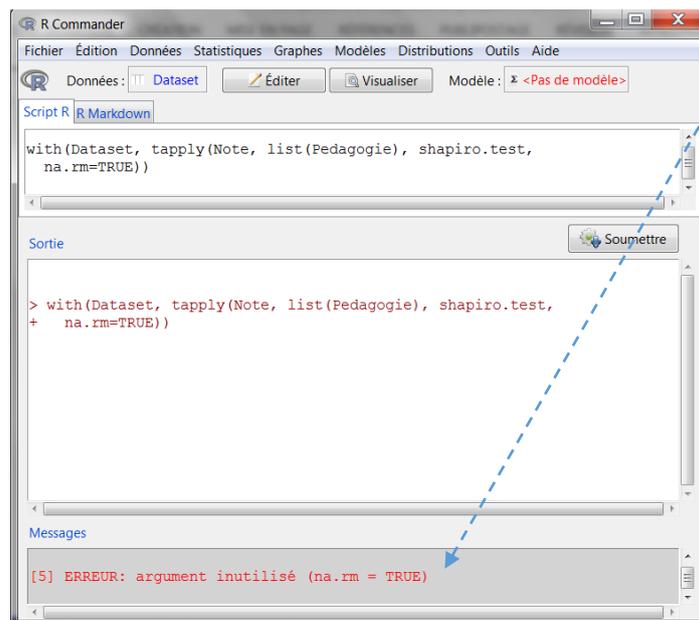
R Commander permet de tester la normalité d'une distribution avec le test de Shapiro-Wilk (Statistiques>Résumés>Test de normalité de Shapiro-Wilk). Mais quand il s'agit, comme ici, de deux sous-groupes, le plus rapide est de passer par la commande Tableau de statistiques et de procéder comme suit :



La fenêtre Tableau de statistiques apparaît :



La fenêtre R Commander apparaît avec tout en bas un message d'erreur en rouge :



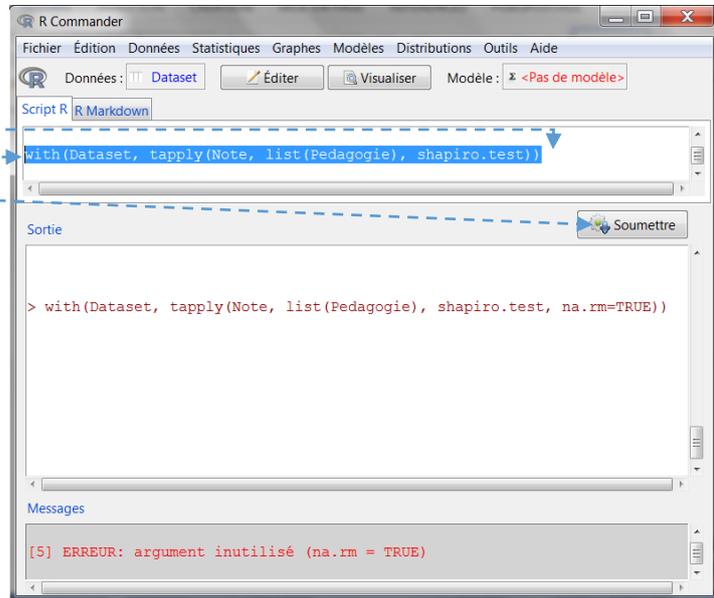
Nous allons modifier la dernière ligne de programmation dans la fenêtre des commandes en retirant cet argument non utilisé (`na.rm = TRUE`).

Nous effaçons l'argument inutile, c'est-à-dire ce qui se trouve après `shapiro.test`.

La ligne de commande devient : `with(Dataset, tapply(Note, list(Pedagogie), shapiro.test))`

#### Etape 3 :

Effacer '`na.rm=TRUE`' directement dans le texte (ne pas oublier la virgule après `shapiro.test`),  
Surligner la ligne avec la souris,  
Cliquer sur Soumettre.

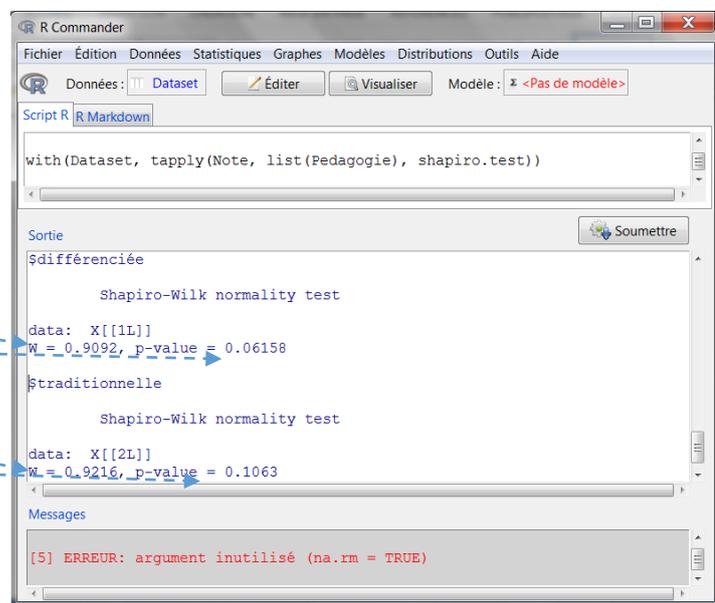


Dans la fenêtre sortie, les valeurs du test apparaissent, d'abord pour la modalité différenciée, puis pour la modalité traditionnelle :

#### Etape 4 : résultats

Pédagogie différenciée :  
Valeur du test de normalité,  
Significatif au seuil :

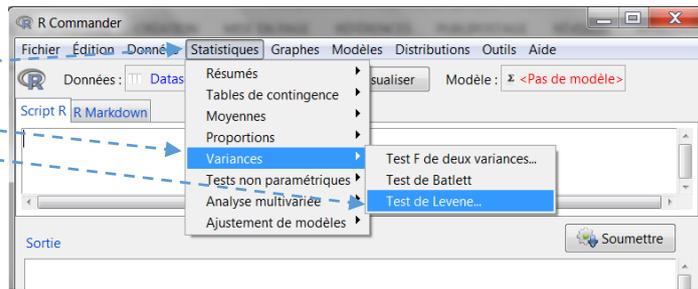
Pédagogie traditionnelle :  
Valeur du test de normalité,  
Significatif au seuil :



L'hypothèse nulle d'une non différence avec la loi normale n'est pas rejetée au seuil .05 pour chaque modalité. Une fois vérifiée la normalité des deux distributions, nous pouvons poursuivre le test statistique t. Une autre information est primordiale pour ce test : l'homogénéité (l'égalité) ou non des variances des deux distributions.

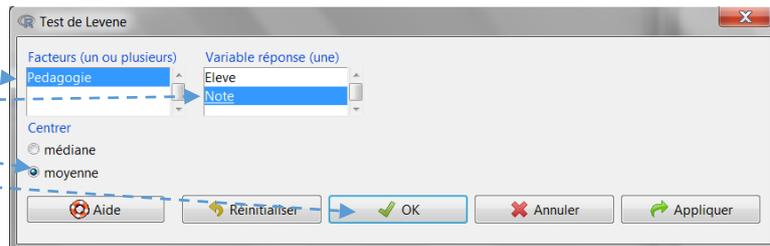
## 8.2 Apprécier l'égalité des variances des 2 groupes

Étape 1 :  
Cliquez sur Statistiques,  
Variances,  
Test de Levene



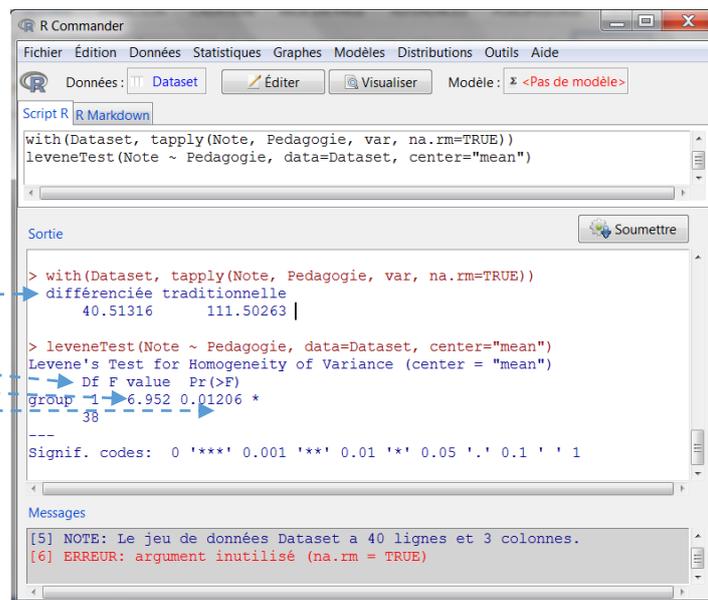
La fenêtre Test de Levene apparaît :

Étape 2 :  
Choisir le facteur, ici Pedagogie,  
Choisir la variable réponse (ici Note),  
Cliquez sur moyenne,  
OK



Le résultat du test de Levene apparaît dans la fenêtre Sortie :

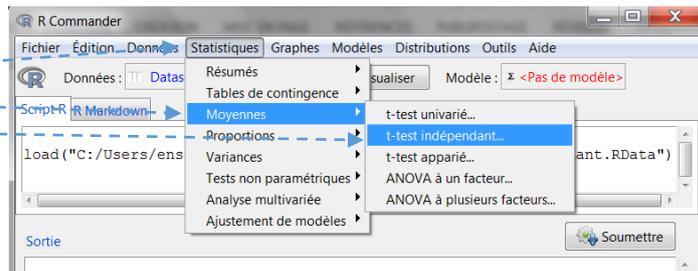
Résultats  
Valeur des variances des 2 groupes,  
  
Degrés de liberté :  
Valeur du test de Levene,  
Significatif au seuil :



Le test est significatif à .01 : on peut rejeter l'hypothèse nulle d'une homogénéité des variances. Nous pouvons à présent calculer le test de Student avec des variances non homogènes.

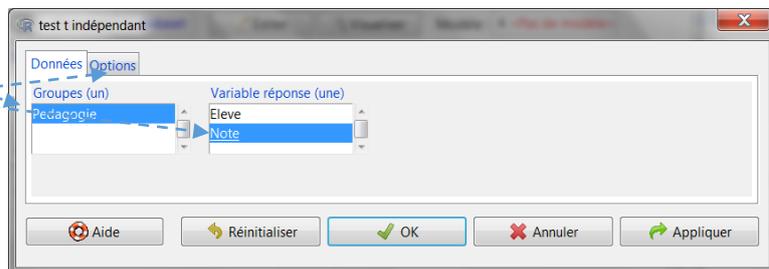
### 8.3 Calcul du t de Student pour groupes indépendants

Etape 1 :  
Cliquez sur Statistiques,  
Moyennes,  
t-test indépendant



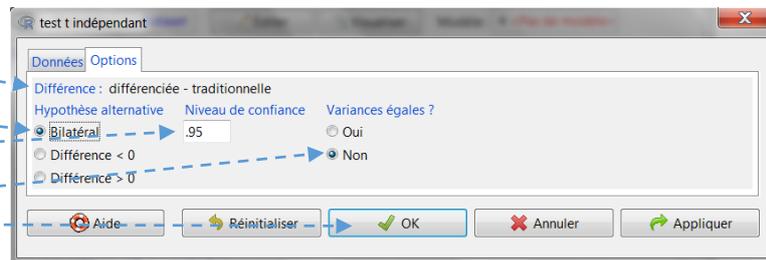
Dans la fenêtre test t indépendant, il y a deux onglets : Données et Options. Renseignons tout d'abord la fenêtre Données :

Etape 2 :  
Choisir le groupe, ici Pedagogie,  
Choisir la variable réponse (ici Note),  
Options



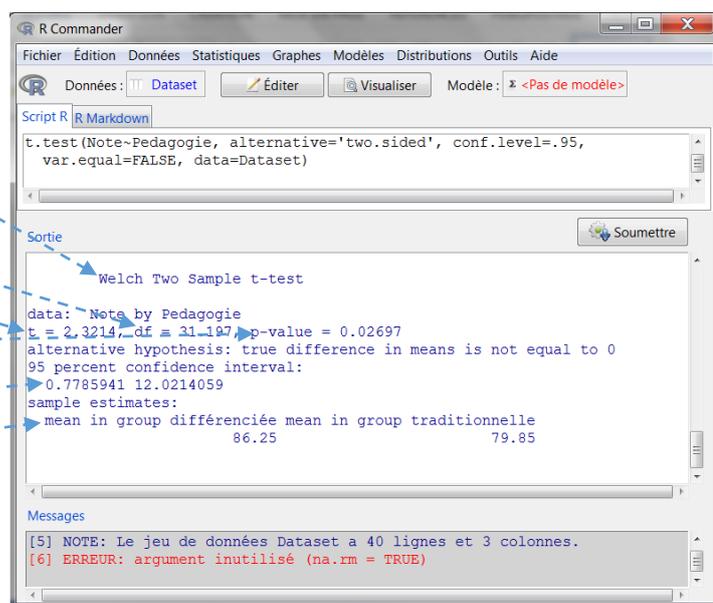
La fenêtre Options s'affiche :

Etape 3 :  
Les deux groupes sont affichés :  
Choisir test bilatéral,  
Choisir le niveau de confiance (ici .95),  
Variances égales (ici Non),  
OK



La fenêtre sortie apparaît :

Résultats  
Le t est calculé avec la correction de Welch (variances non homogènes)  
Degrés de liberté :  
Valeur du test t,  
  
Significatif au seuil :  
Intervalle de confiance de la différence des moyennes :  
Valeur des moyennes des 2 groupes,



## 9 Le test $U$ de Mann-Withney (ou test de la somme des rangs de Wilcoxon) pour groupes indépendants (*Mann-Withney U-test* or *Wilcoxon sum rank test*)

Ce test non paramétrique est utilisé lorsque l'on souhaite comparer les moyennes de 2 groupes indépendants, mais dont les deux distributions de scores ne se distribuent pas normalement, ou encore lorsque l'on souhaite comparer deux classements (variables ordinales).

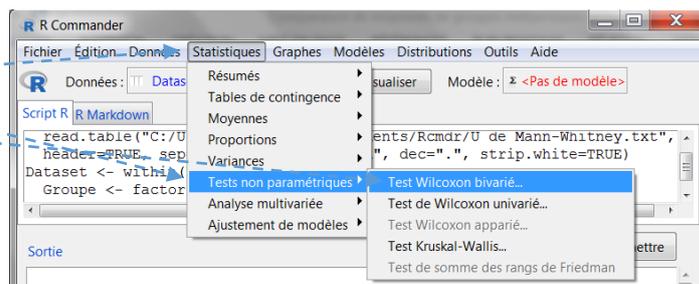
Une chercheuse souhaite tester l'effet de messages qui apparaissent sur un écran d'ordinateur, sur l'anxiété des participants. Les participants doivent remplir un tableau à partir d'une feuille, mais pour cinq d'entre eux (cotés 2 ci-dessous), des messages d'avertissement apparaissent sur l'écran (par exemple « Attention, vos données sont sur le point d'être effacées, cliquez vite sur Enregistrer » ; « Attention, la session semble vouloir s'arrêter, cliquer immédiatement sur « Ne pas arrêter »... Pour sept autres participants (cotée 1 ci-dessous), aucun message ne vient perturber la séance. Après l'exercice, la chercheuse propose aux participants de remplir une échelle d'anxiété dont elle relève les scores.

Rentrer les données dans R Commander, soit directement, soit en passant par un tableau (voir chapitre Tableau de données) et convertir la variable Groupe en facteur, avec 1 = Sans message et 2 = Avec messages.

Groupe	Anxiété
1	7
1	9
1	4
1	7
1	11
1	6
1	4
2	17
2	3
2	14
2	15
2	7

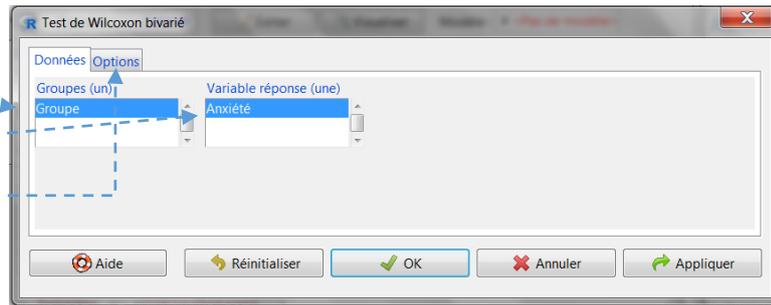
Calculons le test d'hypothèse  $U$  de Mann Withney ou test de la somme des rangs de Wilcoxon.

Etape 1 :  
Cliquez sur Statistiques,  
Tests non paramétriques,  
Test Wilcoxon bivarié



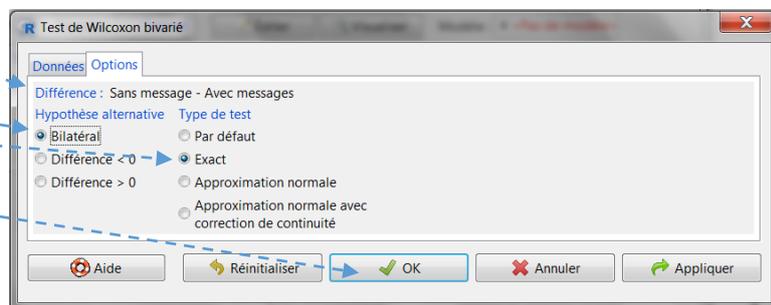
Dans la fenêtre Test de Wilcoxon bivarié, il y a deux onglets : Données et Options.  
Renseignons tout d'abord la fenêtre Données :

Etape 2 :  
Choisir le groupe, (ici Groupe),  
Choisir la variable réponse, (ici Anxiété),  
Options



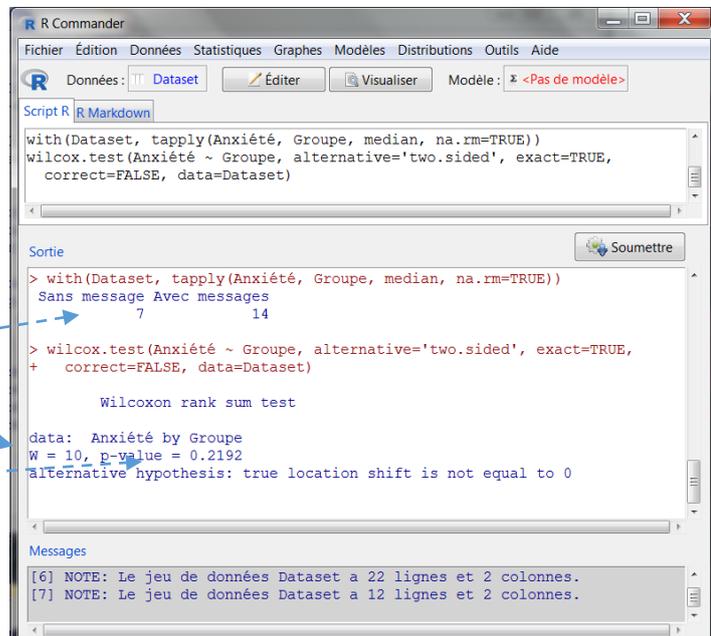
La fenêtre Options s'affiche :

Etape 3 :  
Les deux groupes sont affichés :  
Choisir test bilatéral,  
Choisir Exact  
OK



La fenêtre sortie apparaît :

Résultats  
Les médianes :  
Valeur du test W :  
Seuil de signification :



Il n'y a pas de différence significative entre les moyennes d'anxiété des deux groupes

La valeur du test de la somme des rangs de Wilcoxon calculée par R Commander est équivalente à celle du U de Mann-Withney.

## 10 Le test *t* de Student pour groupes appariés (*paired sample, ou related t-test*)

Ce test paramétrique est utilisé lorsque l'on souhaite comparer les moyennes de 2 groupes appariés, c'est-à-dire lorsque que ce sont les mêmes participants qui ont été observés 2 fois. Les deux distributions de scores dont sont issues les moyennes, se distribuent normalement.

Un enseignant souhaite apprécier l'efficacité d'un renforcement en statistiques en comparant les moyennes des notes de ses 20 élèves, avant et après le renforcement. Son groupe étant composé de filles et de garçons, il a également relevé le sexe en codant 1 pour les garçons et 2 pour filles. Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données) et convertir la variable Sexe en facteur, avec 1 = Garçon et 2 = Fille.

Ces données sont celles utilisées dans le chapitre Corrélation avec le *r* de Bravais Pearson. Ces données ont été scindées en 2 pour gagner de la place, le tableau comporte bien 4 colonnes et 20 lignes (hors ligne titre).

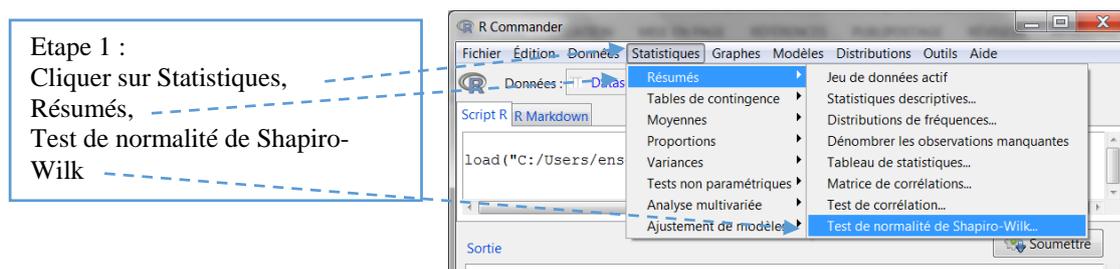
Participant	Avant	Après	Sexe	(suite)	Participant	Avant	Après	Sexe
1	31	34	1		11	31	28	1
2	26	25	1		12	27	32	1
3	32	38	1		13	25	25	1
4	38	36	1		14	28	30	1
5	29	29	1		15	32	41	1
6	34	41	2		16	27	37	2
7	24	26	2		17	37	39	2
8	35	42	2		18	29	33	2
9	30	36	2		19	31	40	2
10	36	44	2		20	27	28	2

Nous allons procéder en 2 temps :

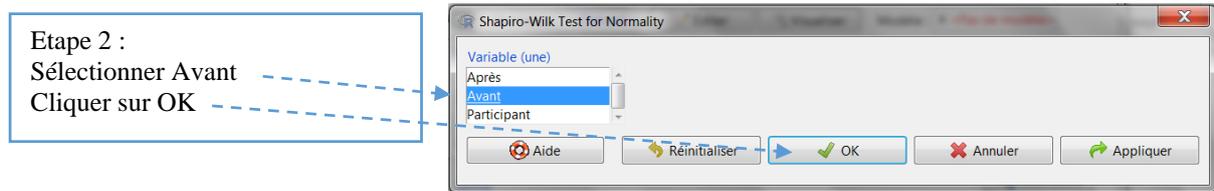
1. vérifier la normalité des distributions des 2 groupes, avant et après le renforcement ;
2. calculer le test d'hypothèse

### 10.1 Vérifier la normalité des 2 groupes

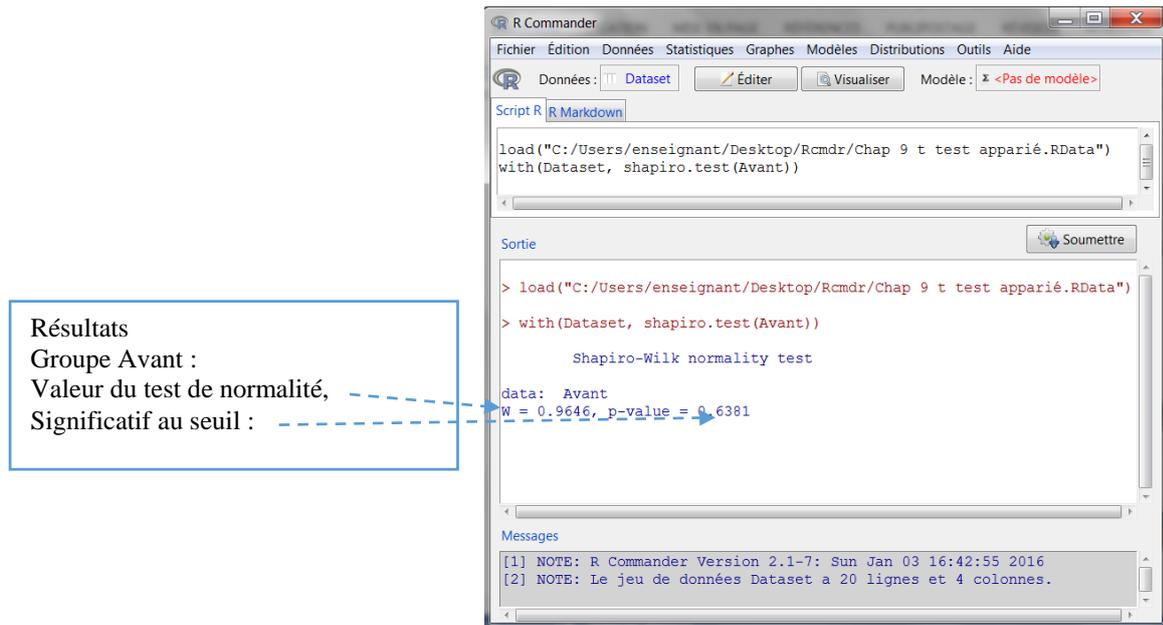
Commençons par vérifier la normalité de la distribution des scores pour le groupe Avant :



La fenêtre Shapiro-Wilk Test for Normality apparaît :

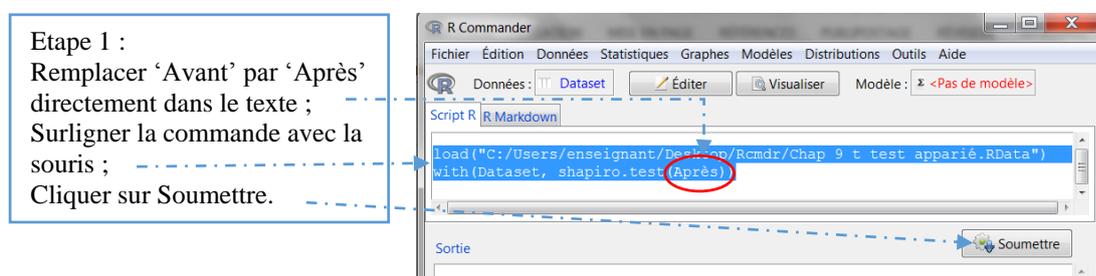


La fenêtre des résultats apparaît :

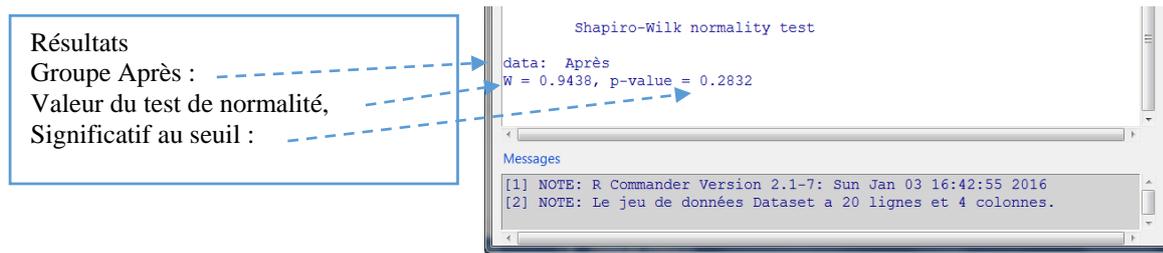


Les scores pour le groupe Avant se distribuent normalement.

Il faut recommencer la procédure ci-dessus pour le second groupe Après. On peut cependant procéder encore plus rapidement : en effet, il suffit de modifier directement la dernière ligne de commande en remplaçant le nom du premier groupe (Avant), par le nom du second groupe (Après), de surligner la commande et de soumettre.

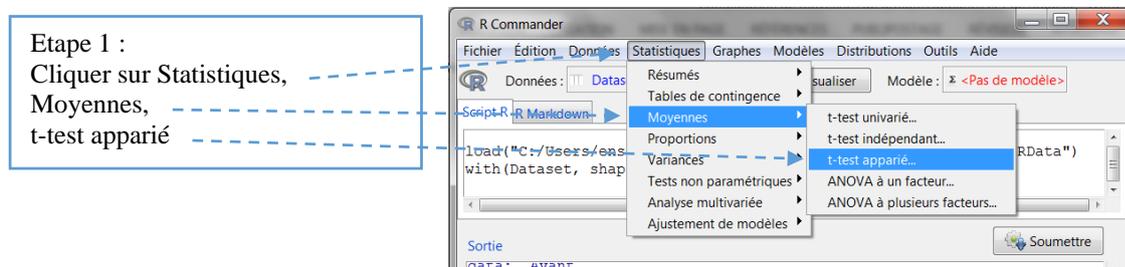


Voici les résultats :

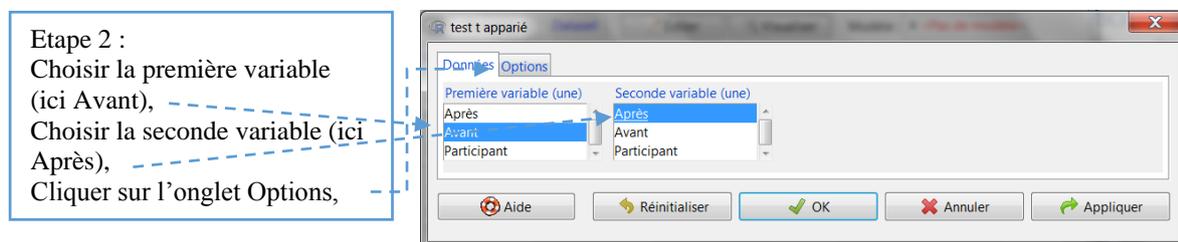


L'hypothèse nulle (non différence avec la loi normale) est également acceptée pour le second groupe, nous pouvons dès lors utiliser le test paramétrique *t* de Student. (Rappel : si au moins l'une des deux distributions n'avait pas été normale (rejet de l'hypothèse nulle au seuil choisi), il aurait alors fallu utiliser un test non paramétrique).

### 10.2 Calcul du test *t* pour groupes appariés

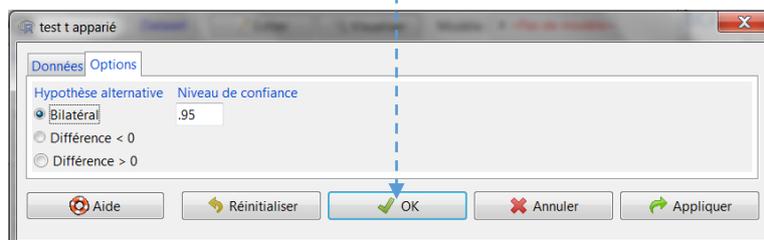


La fenêtre test *t* apparié apparaît :



La fenêtre de l'onglet Options permet de choisir d'une part la nature de l'hypothèse et d'autre part, le seuil alpha. Nous garderons ici les valeurs par défaut (test bilatéral et seuil à .95).

Cliquez sur OK



La fenêtre résultats apparaît :

Résultats  
Significatif au seuil  
Degrés de liberté  
Valeur du test *t*

Intervalle de confiance de la différence des moyennes  
Différence des moyennes des 2 groupes

```

Sortie
+ conf.level=.95, paired=TRUE))

Paired t-test

data: Avant and Après
t = -4.2797, df = 19, p-value = 0.0004046
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-5.583952 -1.916048
sample estimates:
mean of the differences
-3.75

Messages
[1] NOTE: R Commander Version 2.1-7: Sun Jan 03 16:42:55 2016
[2] NOTE: Le jeu de données Dataset a 20 lignes et 4 colonnes.
  
```

La différence entre les 2 groupes est significative à .0005.

### 10.3 Calcul du test *t* de Student pour sous-groupes appariés

Il arrive souvent que l'on souhaite comparer les moyennes, non pas de l'ensemble de l'échantillon, mais selon des sous-groupes qui nous intéressent, par exemple les garçons ou les filles, autrement dit, des sous-groupes d'une autre variable non confondus. La démarche est très simple sur R Commander, mais elle nécessite d'ajouter quelques mots sur le script du *t* de Student pour groupes appariés.

Par exemple, cet enseignant voudrait ainsi savoir si son renforcement a été bénéfique pour le sous-groupe des filles.

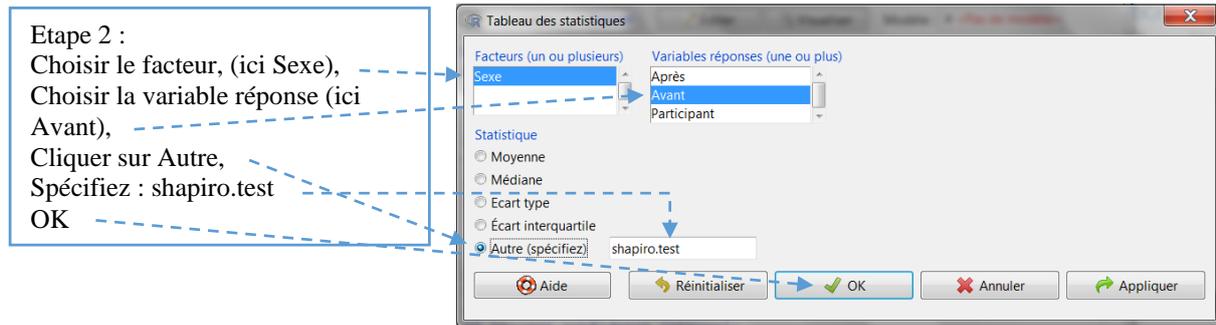
Il faut tout d'abord vérifier que les deux distributions de scores (ici Avant et Après) pour ce sous-groupe des filles se distribuent normalement.

#### 10.3.1 Vérifier la normalité d'un sous-groupe

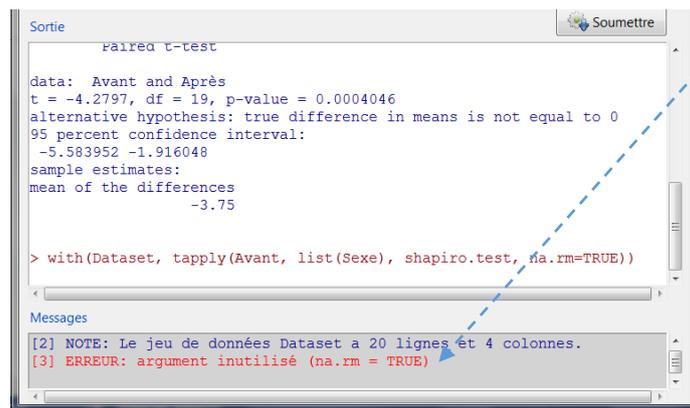
R Commander permet de tester la normalité d'une distribution avec le test de Shapiro-Wilk (Statistiques>Résumés>Test de normalité de Shapiro-Wilk), nous venons de le voir. Mais quand il s'agit, comme ici, de sous-groupes, le plus rapide est de passer par la commande Tableau de statistiques et de procéder comme suit :

Etape 1 :  
Cliquer sur Statistiques,  
Résumés,  
Tableau de statistiques

La fenêtre Tableau de statistiques apparaît :

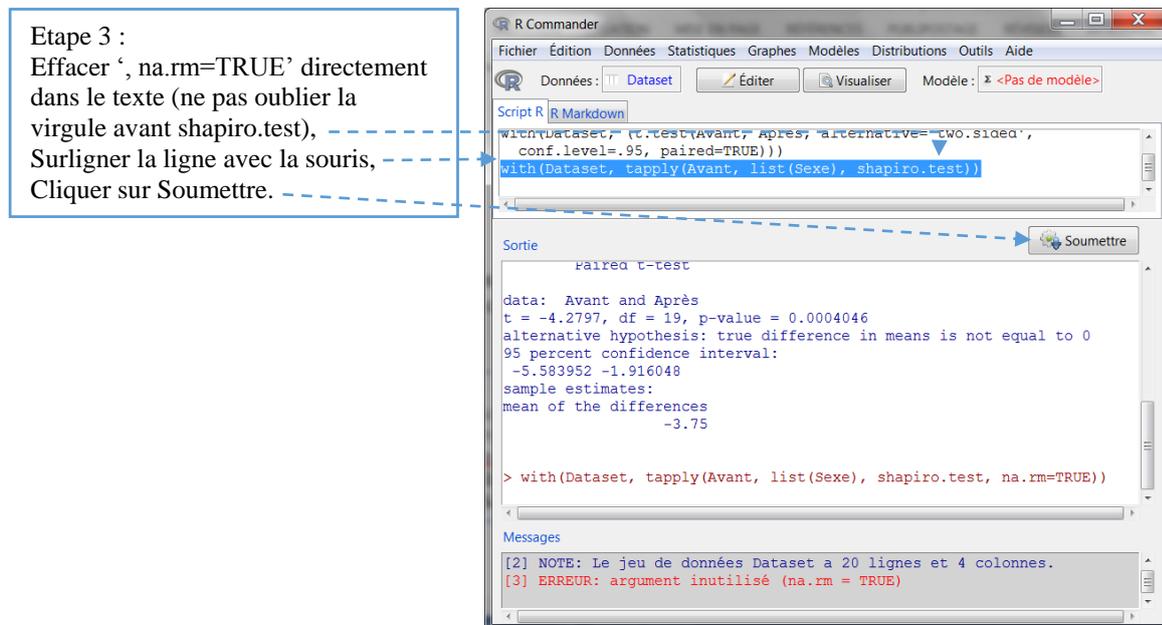


La fenêtre R Commander apparaît avec tout en bas un message d'erreur en rouge



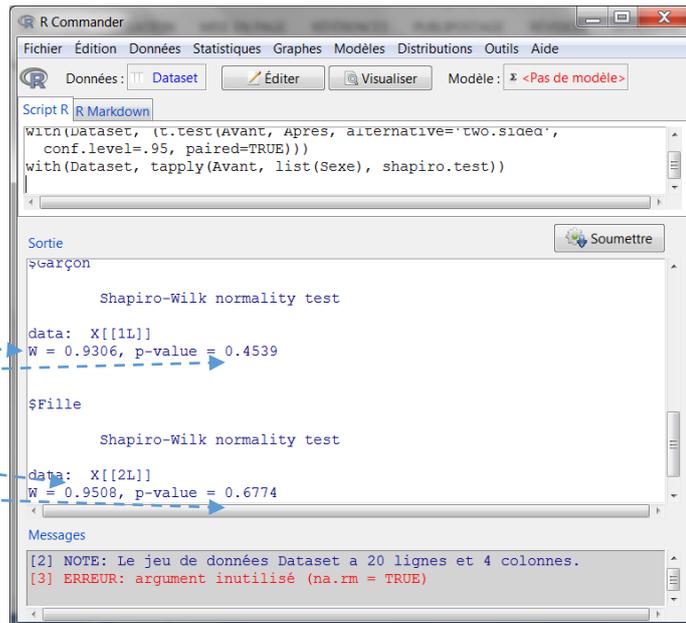
Nous allons modifier la dernière ligne de programmation dans la fenêtre des commandes en retirant cet argument non utilisé (na.rm = TRUE).

Nous effaçons l'argument inutile, c'est-à-dire ce qui se trouve après shapito.test : la ligne de commande devient : with(Dataset, tapply(Avant, list(Sexe), shapiro.test))



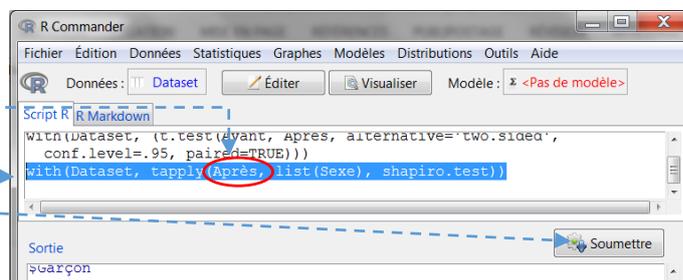
Dans la fenêtre sortie, les valeurs du test apparaissent, d'abord pour la modalité Garçon, puis pour la modalité Fille :

Résultats pour Avant  
 Garçon :  
 Valeur du test de normalité,  
 Significatif au seuil :  
  
 Fille :  
 Valeur du test de normalité,  
 Significatif au seuil :



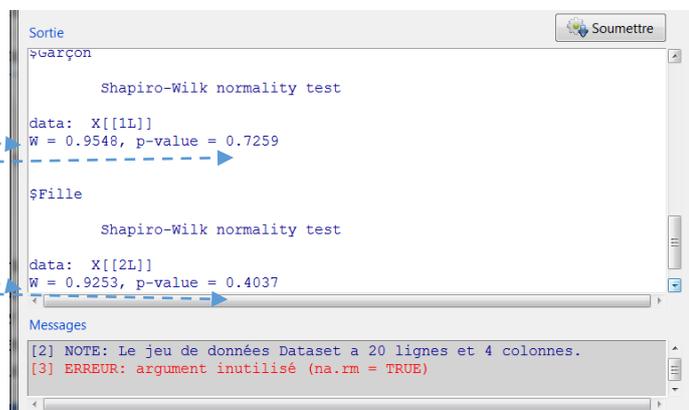
Pour les scores du groupe Avant, l'hypothèse nulle d'une non différence avec la loi normale n'est pas rejetée au seuil .05 pour chaque modalité de la variable Sexe, donc des filles qui nous intéresse dans cet exemple. Il faut procéder de même pour les scores Après, soit en déroulant la même procédure que ci-dessus, soit, plus rapidement, en remplaçant dans la dernière ligne de la commande, Avant par Après.

Etape 1 :  
 Remplacer 'Avant' par 'Après'  
 directement dans le texte ;  
 Surligner la commande avec la  
 souris ;  
 Cliquer sur Soumettre



La fenêtre des résultats apparaît :

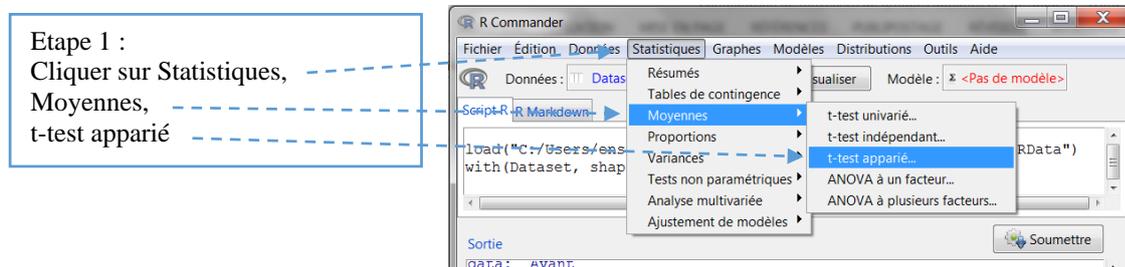
Résultats pour Après  
 Garçon :  
 Valeur du test de normalité,  
 Significatif au seuil :  
  
 Fille :  
 Valeur du test de normalité,  
 Significatif au seuil :



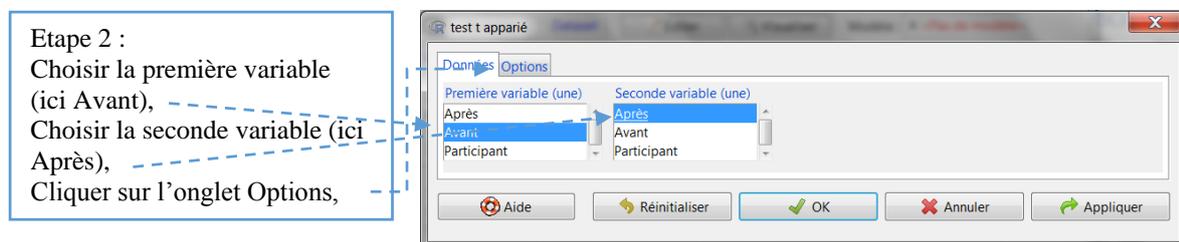
Pour les scores du groupe Après, l'hypothèse nulle d'une non différence avec la loi normale n'est pas rejetée au seuil .05 pour chaque modalité de la variable Sexe, donc des filles qui nous intéresse dans cet exemple.

10.3.2 Calcul du test  $t$  pour deux sous-groupes appariés

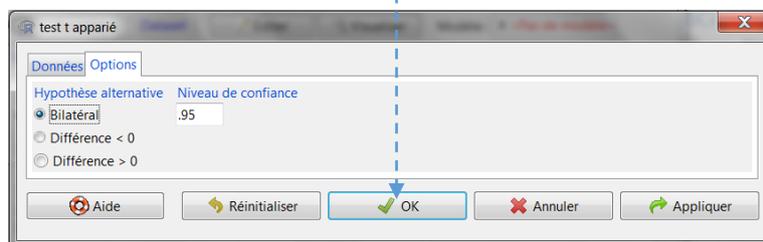
Nous allons dérouler la même procédure qu'au début du chapitre pour obtenir la ligne de commande du test  $t$  et ce, afin de la modifier comme nous le souhaitons.



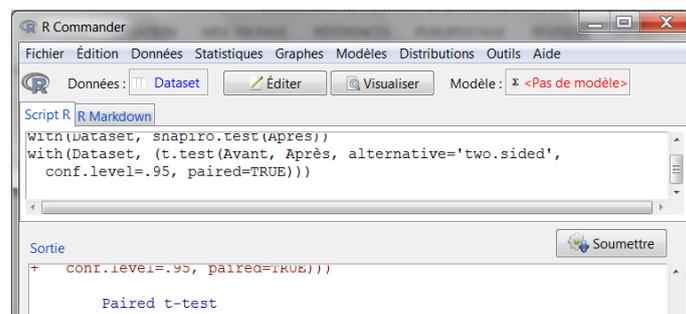
La fenêtre test  $t$  apparié apparaît :



La fenêtre de l'onglet Options permet de choisir d'une part la nature de l'hypothèse et d'autre part, le seuil alpha. Nous garderons ici les valeurs par défaut (test bilatéral et seuil à .95). Cliquez sur OK



La fenêtre résultats apparaît avec dans la fenêtre Script la commande du test  $t$  : c'est elle que nous allons maintenant modifier :



Comme au tout début, le test a donc été calculé sur l'ensemble de l'échantillon, tous sous-groupes de la variable Sexe confondus. Nous allons maintenant spécifier le sous-groupe Fille, sur lequel va porter le test, en modifiant la dernière ligne de commande dans la fenêtre script qui est :

```
with(ManipSEV, (t.test(Avant, Après, alternative='two.sided', conf.level=.95,
paired=TRUE)))
```

Ajoutons les noms de la variable Sexe et du sous-groupe Fille après chaque groupe comme suit : [Sexe=='Fille'] (attention aux espaces, aux signes ' et au == ). Soit :

```
with(ManipSEV, (t.test(Avant [Sexe=='Fille'], Après [Sexe=='Fille'], alternative='two.sided',
conf.level=.95, paired=TRUE))),
```

puis surlignons la commande et soumettre;

```
with(ManipSEV, (t.test(Avant [Sexe=='Fille'], Après [Sexe=='Fille'], alternative='two.sided',
conf.level=.95, paired=TRUE)))
```

Voici la fenêtre des résultats :

Résultats pour le sous-groupe Fille

Significatif au seuil

Degrés de liberté

Valeur du test t

Intervalle de confiance de la différence des moyennes

Différence des moyennes des 2 groupes

```
Paired t-test
data: _Avant[Sexe == "Fille"] and Après[Sexe == "Fille"]
t = -5.5876, df = 9, p-value = 0.0003396
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-7.867179 -3.332821
sample estimates:
mean of the differences
-5.6
```

La différence entre les scores Avant et Après pour le sous-groupe Fille est significative à .0005.

Si l'on souhaitait comparer l'autre sous-groupe Garçon, il suffirait de remplacer Fille par Garçon dans la dernière ligne de commande, sachant que la normalité des distributions pour cette seconde modalité de la variable Sexe a déjà été vérifiée plus haut.

## 11 Le test du signe de Wilcoxon pour groupes appariés (*Wilcoxon signed ranks test*)

Ce test non paramétrique est utilisé lorsque l'on souhaite comparer les moyennes de 2 groupes appariés, mais dont les deux distributions de scores ne se distribuent pas normalement, ou encore deux classements (variables ordinales).

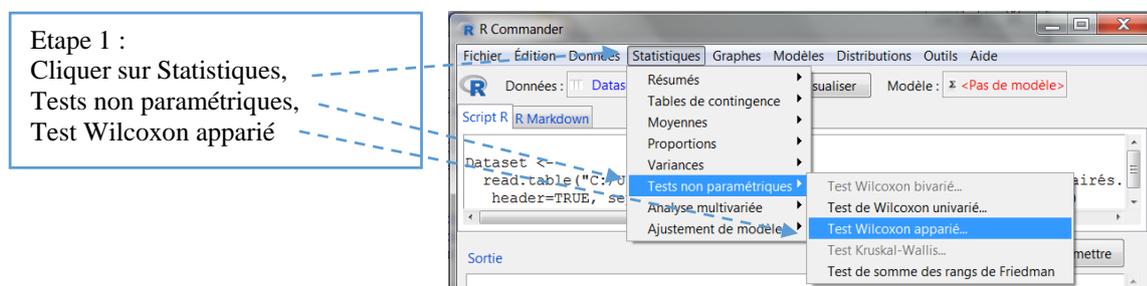
Un enseignant voudrait vérifier que son cours de français a eu un effet bénéfique sur les sentiments d'efficacité de ses 10 stagiaires au DAEU. A cet effet, il a fait passer une échelle de sentiments d'efficacité en français à son groupe avant ses interventions (Pré-test) et de nouveau la même échelle après ses interventions (Post-test).

Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données).

Elève	Pré-test	Post-test
1	5	7
2	6	6
3	2	3
4	4	8
5	6	7
6	7	6
7	3	7
8	5	8
9	5	5
10	5	8

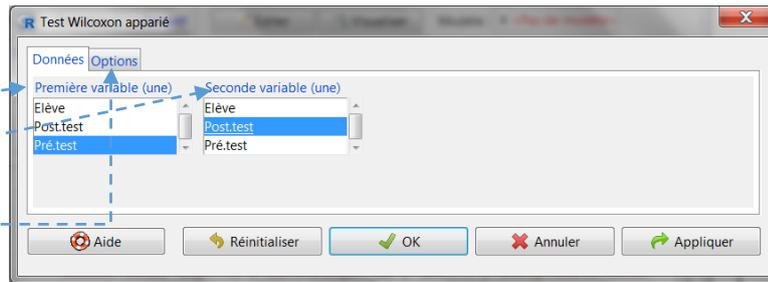
Tel que formulé (effet bénéfique), il s'agit de tester que la moyenne Post-test est plus grande que la moyenne Pré-test : c'est un test unilatéral. Ce qui signifie également que la différence Pré-test – Post-test sera négative.

Calculons le test d'hypothèse du signe de Wilcoxon pour groupes appariés



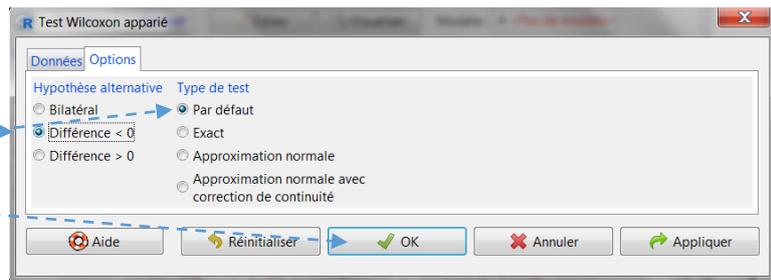
Dans la fenêtre Test de Wilcoxon apparié, il y a deux onglets : Données et Options. Renseignons tout d'abord la fenêtre Données :

Etape 2 :  
 Choisir la première variable (ici Pré-test),  
 Choisir la seconde variable (ici Post-test),  
 Options



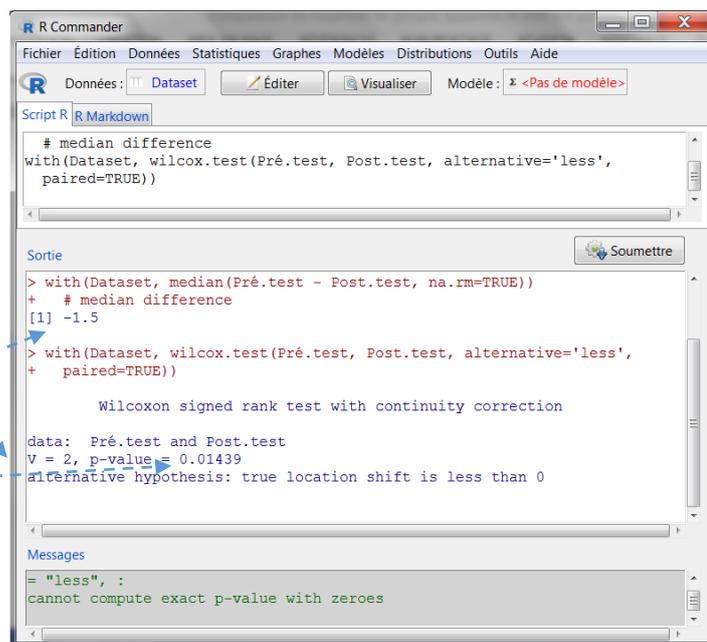
La fenêtre Options s'affiche :

Etape 3 :  
 Choisir Différence <0, c'est-à-dire test unilatéral avec Post-test plus grand que Pré-test  
 Choisir Par défaut  
 OK



La fenêtre sortie apparaît :

Résultats  
 Différence de médianes :  
 Valeur du test (Somme des rangs) :  
 Seuil de signification :



Il y a une différence significative entre les moyennes Pré-test et Post-test : le cours a bien eu un effet bénéfique sur les sentiments d'efficacité en français

## 12 ANOVA pour groupes indépendants (*One-Way between-groups*)

Le test paramétrique de l'analyse de la variance (ANOVA) permet de comparer les moyennes de 2 ou plus de 2 groupes indépendants. Plus précisément, il teste l'hypothèse nulle  $H_0$  que les moyennes sont identiques. Les distributions dont sont issues les moyennes doivent se distribuer normalement. Les variances doivent être homogènes.

Trente volontaires participent à une expérience visant à apprécier les effets de l'alcool sur la conduite automobile. Ils sont répartis au hasard dans trois groupes, trois conditions au début desquelles ils boivent une boisson très alcoolisée (codée 1 ci-dessous), ou une boisson peu alcoolisée (codée 3, ci-dessous), ou une boisson sans alcool (codée 2 ci-dessous). Une demi-heure plus tard, les participants conduisent dans un simulateur de conduite pendant 15 minutes et le nombre d'erreurs de conduite est enregistré automatiquement. La variable indépendante (ou facteur) est donc ici la boisson, avec 3 modalités (très alcoolisée, peu alcoolisée et sans alcool). La variable dépendante (ou variable réponse) est donc ici la conduite avec le nombre d'erreurs de conduite.

Rentrer les données dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données) et convertir la variable boisson en facteur, avec 1 = très alcoolisée, 2 = sans alcool et 3 = peu alcoolisée.

Les données ont été scindées en 2 pour gagner de la place, le tableau comporte bien 3 colonnes et 30 lignes (hors ligne titre).

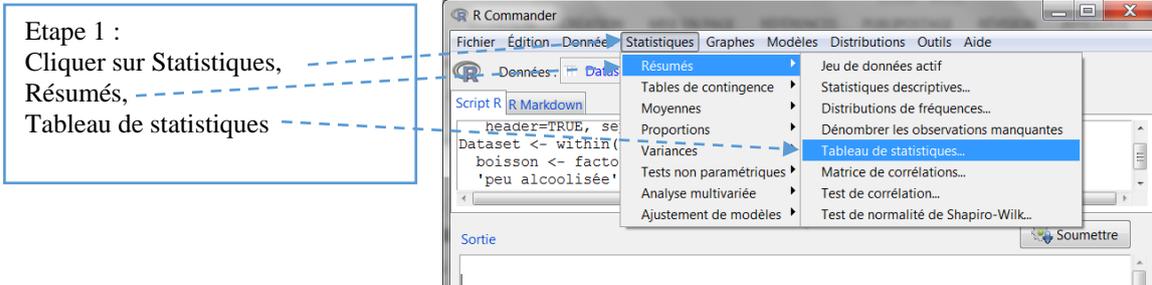
sujet	boisson	conduite	sujet	boisson	conduite
1	1	21	16	2	14
2	1	21	17	2	20
3	1	25	18	2	13
4	1	23	19	2	14
5	1	24	20	2	19
6	1	23	21	3	18
7	1	23	22	3	20
8	1	22	23	3	18
9	1	22	24	3	19
10	1	22	25	3	17
11	2	17	26	3	17
12	2	17	27	3	15
13	2	15	28	3	16
14	2	15	29	3	17
15	2	15	30	3	18

Nous allons procéder en 3 temps :

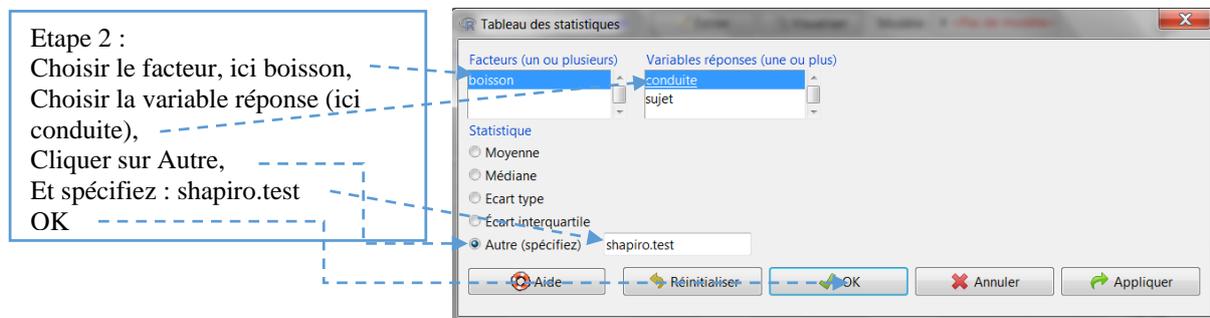
1. vérifier la normalité des distributions des 3 groupes ;
2. apprécier l'homogénéité des variances des 3 groupes ;
3. calculer le test d'hypothèse ANOVA

### 12.1 Tester la normalité des groupes

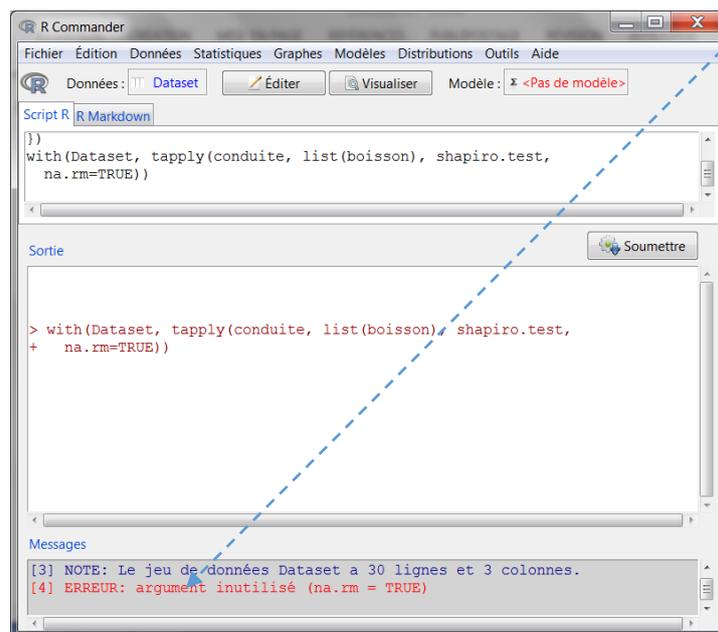
R Commander permet de tester la normalité d'une distribution avec le test de Shapiro-Wilk (Statistiques>Résumés>Test de normalité de Shapiro-Wilk). Mais quand il s'agit, comme ici, de trois sous-groupes, le plus rapide est de passer par la commande Tableau de statistiques et de procéder comme suit :



La fenêtre Tableau de statistiques apparaît :



La fenêtre R Commander apparaît avec tout en bas un **message d'erreur en rouge** :



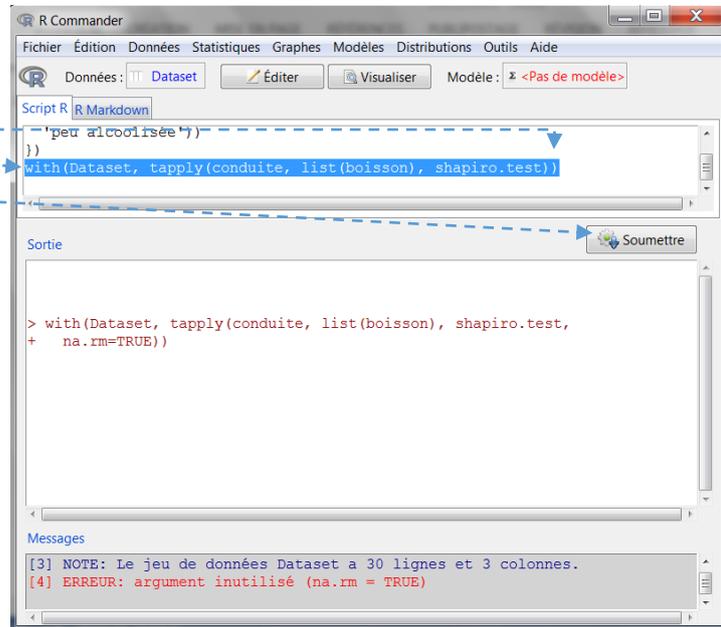
Nous allons modifier la dernière ligne de programmation dans la fenêtre des commandes en retirant cet argument non utilisé (`na.rm = TRUE`).

Nous effaçons l'argument inutile, c'est-à-dire ce qui se trouve après `shapiro.test`.

La ligne de commande devient : `with(Dataset, tapply(Note, list(Pedagogie), shapiro.test))`

### Etape 3 :

Effacer '`na.rm=TRUE`' directement dans le texte (ne pas oublier la virgule après `shapiro.test`),  
Surligner la ligne avec la souris,  
Cliquer sur Soumettre.



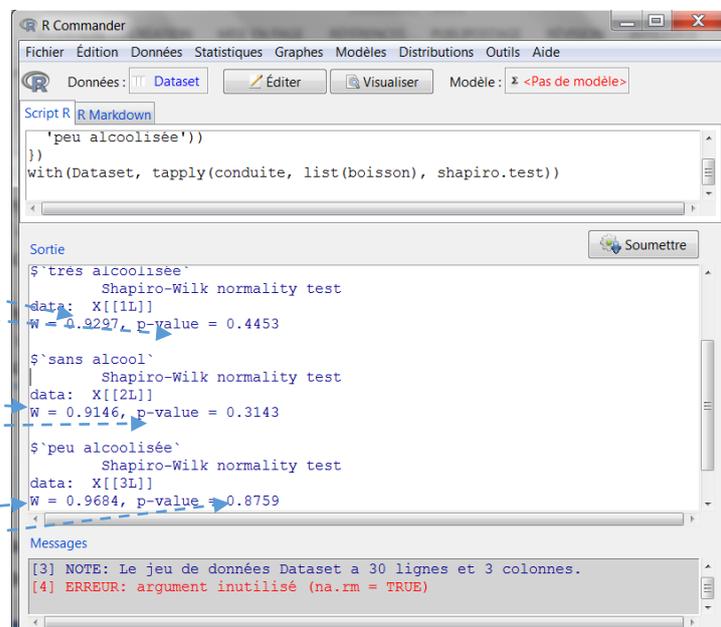
Dans la fenêtre sortie, les valeurs du test apparaissent, d'abord pour la modalité `boisson` très alcoolisée, puis pour la modalité `sans alcool`, enfin pour la modalité `peu alcoolisée` :

### Résultats

Boisson très alcoolisée :  
Valeur du test de normalité,  
Significatif au seuil :

Boisson sans alcool :  
Valeur du test de normalité,  
Significatif au seuil :

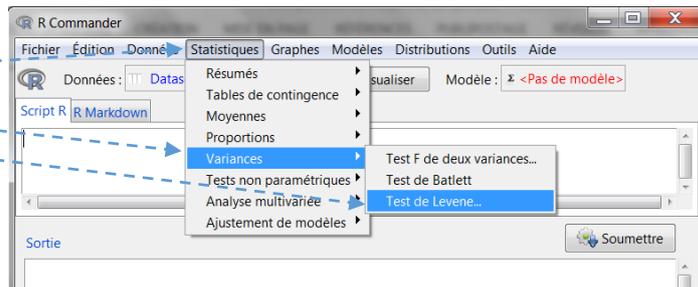
Boisson très alcoolisée :  
Valeur du test de normalité,  
Significatif au seuil :



L'hypothèse nulle d'une non différence avec la loi normale n'est pas rejetée au seuil `.05` pour chaque modalité. Une fois vérifiée la normalité des trois distributions, il faut vérifier l'homogénéité (l'égalité) ou non des variances des trois distributions.

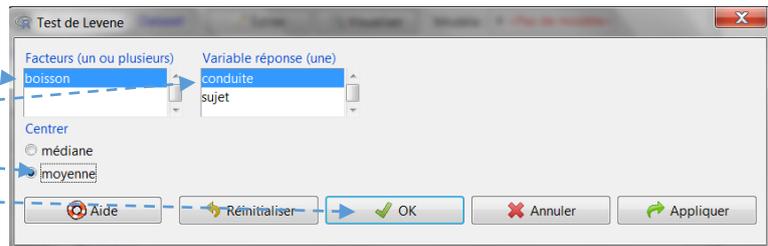
## 12.2 Tester l'égalité des variances des 3 groupes

Étape 1 :  
Cliquez sur Statistiques,  
Variances,  
Test de Levene



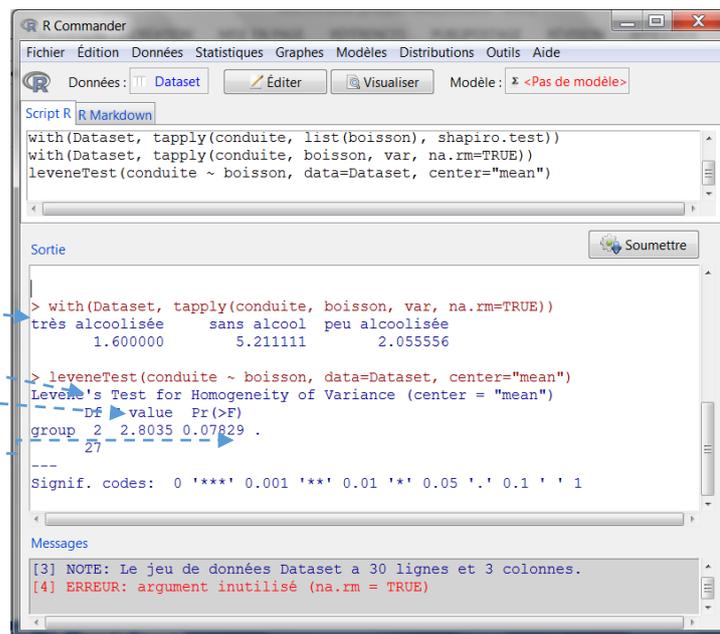
La fenêtre Test de Levene apparaît :

Étape 2 :  
Choisir le facteur, ici boisson,  
Choisir la variable réponse (ici  
conduite),  
Cliquez sur moyenne,  
OK



Le résultat du test de Levene apparaît dans la fenêtre Sortie :

Résultats  
Valeur des variances des 3 groupes,  
Degrés de liberté :  
  
Valeur du test de Levene  
  
Significatif au seuil :

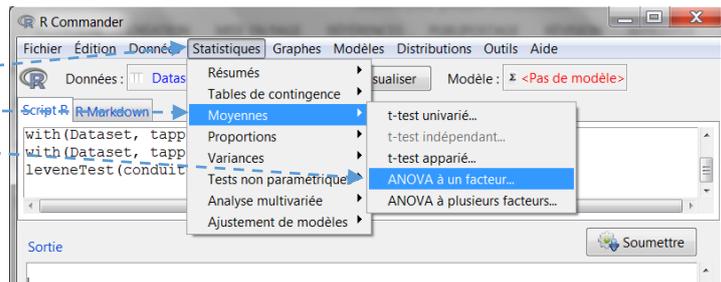


Le test n'est significatif pas à .05 : on peut accepter l'hypothèse nulle d'une homogénéité des variances.

Nous pouvons à présent calculer l'ANOVA.

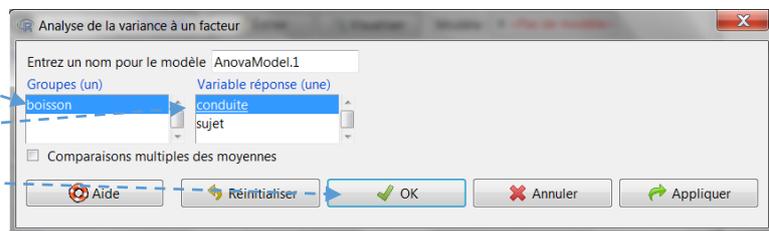
### 12.3 Calcul de l'ANOVA pour groupes indépendants

Etape 1 :  
Cliquez sur Statistiques,  
Moyennes,  
ANOVA à un facteur



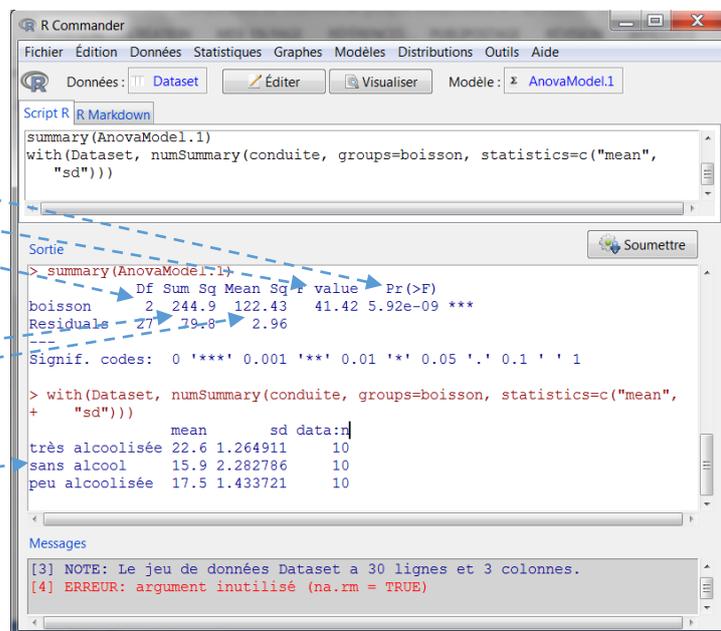
La fenêtre Analyse de la variance à un facteur apparaît. Elle permet de nommer l'analyse elle-même (le nom du modèle), de choisir le facteur (la VI) et la variable réponse (la VD). Il est possible également de procéder à des comparaisons *a posteriori* (post-hoc test).

Etape 2 :  
Choisir le groupe (VI), ici boisson,  
Choisir la variable réponse (ici conduite),  
OK



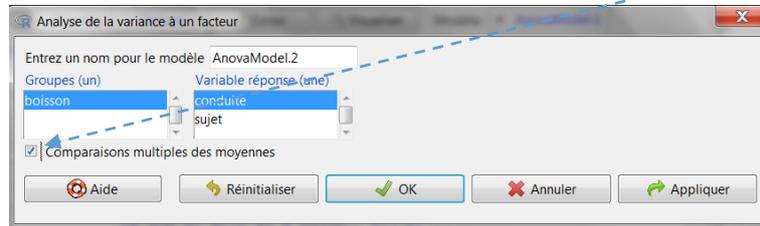
La fenêtre de sortie présente tout d'abord les différentes statistiques du F : pour les variations, la variation due aux boissons (variation entre les moyennes des groupes) et la variation due à l'erreur (ou résidus) (variation à l'intérieur des groupes), les degrés de liberté, les sommes des carrés et les carrés moyens ; la valeur du F et le seuil p. en dessous, les moyennes, écarts-types et effectifs des sous-groupes

Résultats  
Significatif au seuil :  
Valeur du test F,  
Degrés de liberté :  
  
Somme des carrés  
Carrés moyens  
  
Valeur des moyennes, écarts-types et effectifs des 3 groupes



### 12.4 Comparaisons *a posteriori*

R Commander permet de poursuivre l'ANOVA par des comparaisons multiples entre les moyennes, autrement dit des comparaisons *a posteriori* (post hoc). Il suffit pour cela de cocher Comparaison multiples de moyennes dans la fenêtre de l'Analyse de la variance à un facteur.



Les statistiques post hoc apparaissent après celles du F et des groupes, avec pour chaque comparaison, les différences entre les moyennes, le *t* de Student et le seuil *p* associé.

Résultats  
Significatif au seuil :  
Valeur des tests *t*,  
Différence entre les moyennes :

```

Simultaneous Tests for General Linear Hypotheses

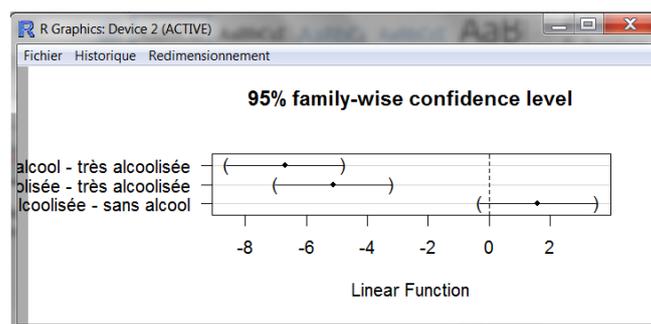
Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = conduite ~ boisson, data = Dataset)
Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
sans alcool - très alcoolisée == 0    -6.7000    0.7688  -8.714 <0.001
peu alcoolisée - très alcoolisée == 0  -5.1000    0.7688  -6.633 <0.001
peu alcoolisée - sans alcool == 0      1.6000    0.7688   2.081  0.113

sans alcool - très alcoolisée == 0    ***
peu alcoolisée - très alcoolisée == 0  ***
peu alcoolisée - sans alcool == 0
    
```

On constate des différences significatives à .001 entre la moyenne du groupe Très alcoolisée et celles des deux autres groupes ; pas entre celles de ces derniers. La fenêtre présente ensuite les intervalles de confiance des différences de moyennes calculés à .95 (non présenté ici).

Enfin, R Commander édite les graphiques de ces intervalles dans une fenêtre spécifique :



## 13 La régression linéaire simple (*Simple Regression*)

Une formatrice en statistiques pense que la performance à l'examen dépend en partie du temps consacré à s'entraîner. Pour tester son hypothèse, elle met en place une aide en ligne pour les 25 étudiantes de sa classe et note en fin d'année, le temps total de connexions en heures. L'épreuve terminale est notée sur 40. Ainsi, chaque étudiante a 2 observations : le temps passé à s'entraîner (Temps) et la note à l'examen (Performance). Dans le modèle de la régression linéaire, la performance est la variable dépendante  $Y$ , ou variable à expliquer, ou critère (*criterion variable*). Le temps est la variable indépendante  $X$ , ou variable explicative, ou prédicteur (*explanatory variable*). Rentrer les données ci-dessous dans R Commander, soit directement, soit en passant par un tableur (voir chapitre Tableau de données).

Les données ci-dessous sont les mêmes que pour la fiche Corrélation rho de Spearman. Elles ont été scindées en 2 pour gagner de la place, le tableau comporte bien 3 colonnes et 25 lignes (hors ligne titre).

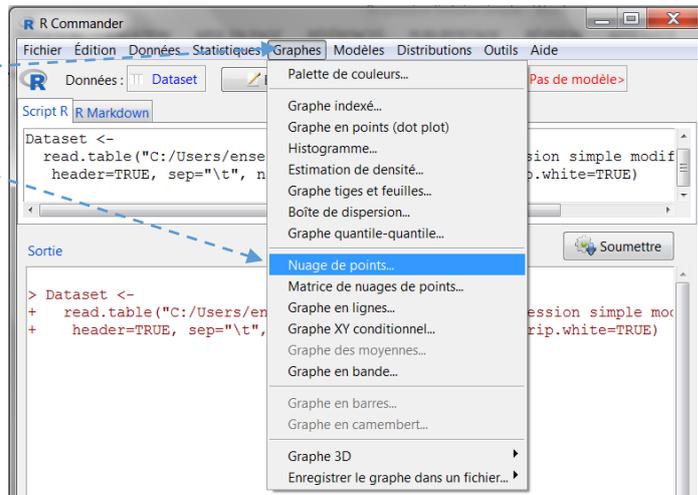
Etudiante	Temps	Performance	Etudiante	Temps	Performance
1	2	21,2	14	2	22,2
2	5	30	15	7	35
3	15	40	16	0	16,8
4	1	20	17	14	40
5	3	25,6	18	7	38,6
6	0	17,3	19	9	40
7	7	39,1	20	1	18,4
8	12	37,7	21	5	35
9	2	21,2	22	7	37
10	6	35,5	23	4	33,3
11	10	38,5	24	10	38,8
12	8	31,2	25	3	29,8
13	5	36,8			

Nous allons procéder en 4 temps :

1. vérifier la nature linéaire de la relation entre  $X$ , le temps et  $Y$ , la performance ;
2. calculer la force et le sens de la relation linéaire entre  $X$  et  $Y$  ;
3. estimer les paramètres de la droite de régression  $Y = bX + a$  et tester le modèle de prédiction
4. vérifier les conditions d'homogénéité des variances et de normalité dans les vecteurs pour interpréter correctement les tests d'hypothèse (la variance de  $Y$  pour toute valeur de  $X$  doit être constante dans la population et les valeurs de  $Y$  correspondant à toute valeur de  $X$  doivent être normalement distribuées autour de la valeur prédite de  $Y$ ).

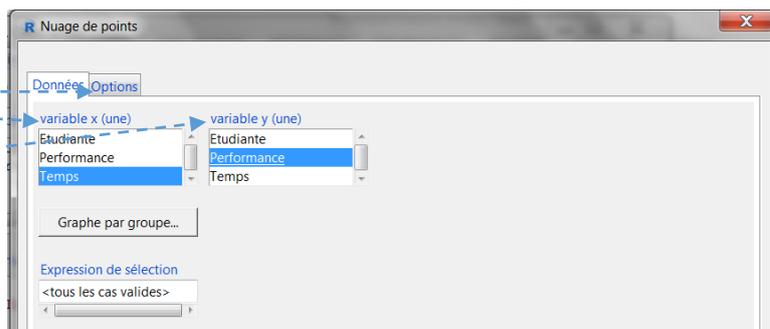
### 13.1 Vérifier la nature linéaire de la relation entre X et Y en observant le nuage de points

Etape 1 :  
Cliquez sur Graphes,  
Nuage de points



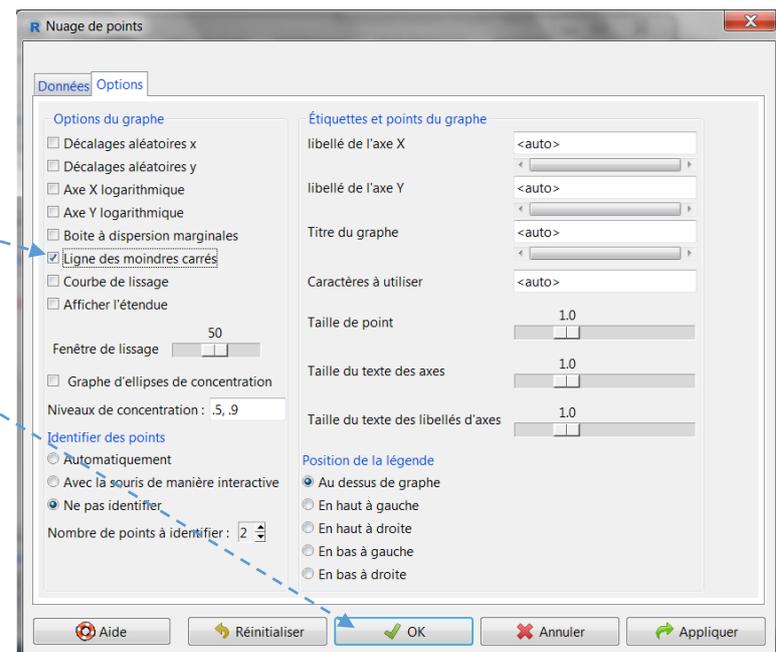
La fenêtre Nuage de point apparaît :

Etape 2 :  
Choisir la variable X (le  
prédicteur), ici, Temps,  
Choisir la variable Y (le critère)  
ici, Performance,  
Cliquez sur Options

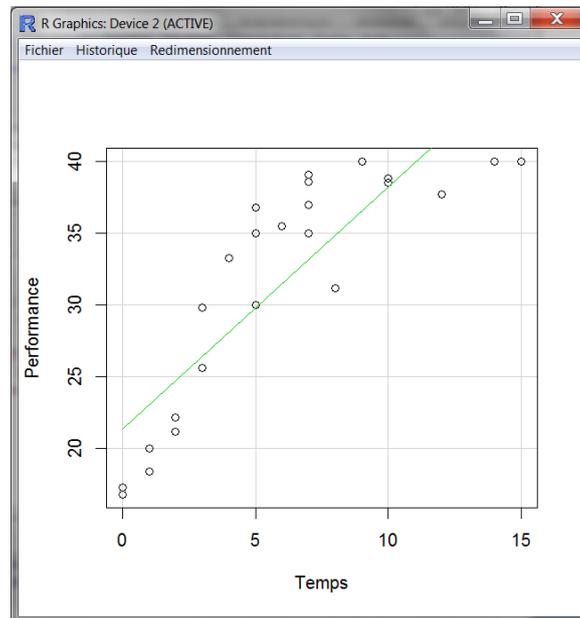


La boîte Options apparaît :

Etape 3 :  
Plusieurs options sont proposées.  
Nous choisissons ici la ligne des  
moindres carrés  
OK



La fenêtre du nuage de points apparaît avec le temps en abscisse (variable explicative ou prédicteur) et la performance en ordonnée (variable à expliquer ou critère ou réponse).



Le nuage de points ayant une forme allongée, on peut considérer qu'il y a une relation de nature linéaire entre les deux variables. La corrélation peut être calculée. La ligne des moindres carrés est en vert.

### 13.2 Calculer la corrélation entre les deux variables Temps et Performance

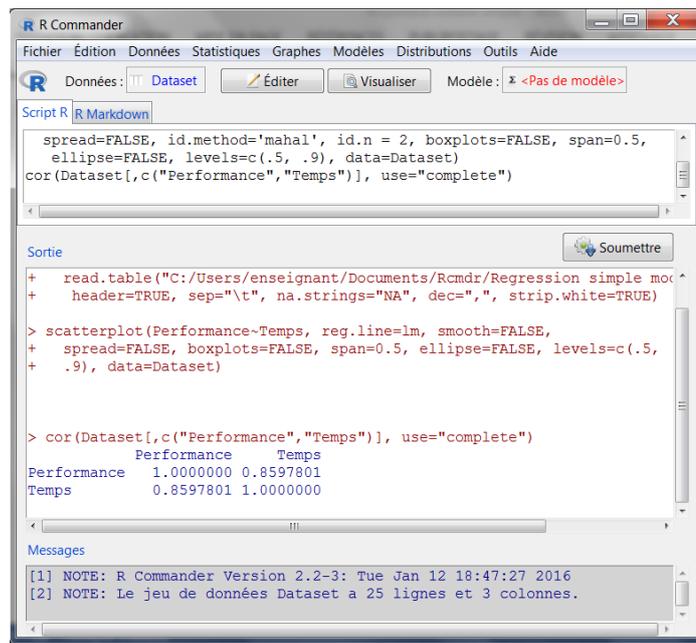
Etape 1 :  
Cliquez sur Statistiques  
Résumés,  
Matrice de corrélations

La fenêtre Matrice de corrélations apparaît :

Etape 2 :  
Choisir les variables dont on veut  
calculer les corrélations (ici  
Performance et Temps),  
Choisir le type de coefficient (ici  
le r de Bravais-Pearson,  
OK

La fenêtre des résultats apparaît :

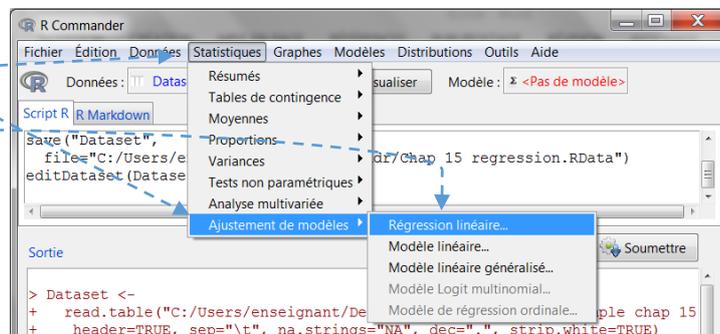
Etape 3 :  
La matrice de corrélation est calculée.  
Ici,  $r = 0,85978$



Avec  $r = 0,86$ , nous avons une relation linéaire très forte et positive entre  $X$  et  $Y$ .

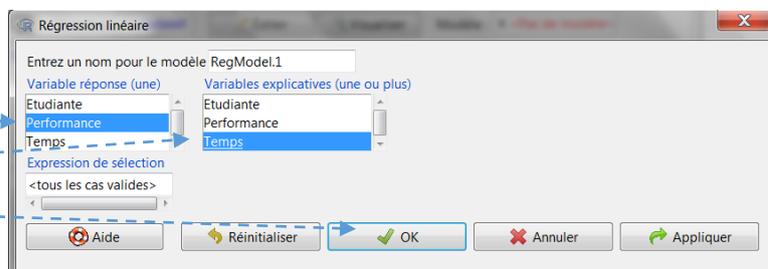
### 13.3 La droite de regression

Etape 1 :  
Cliquer sur Statistiques  
Ajustement de modèles,  
Régression linéaire



La fenêtre Régression linéaire apparaît:

Etape 2 :  
Choisir la variable réponse (ici Performance),  
Choisir la variable explicative (ici Temps),  
OK



La fenêtre Sortie affiche les résultats:

**Résultats**  
Des statistiques descriptives des résidus

La valeur de la constante (intercept)  
La valeur de la pente  
Les tests de signification des 2 coefficients (hypothèse nulle, la valeur = 0) avec t et seuil

Erreur type d'estimation  
Coefficient de détermination  $R^2$ ,  
ANOVA sur le modèle  
Degrés de liberté et seuil

```

R Commander
Fichier Édition Données Statistiques Graphes Modèles Distributions Outils Aide
Données: Dataset Éditer Visualiser Modèle: RegModel.1
Script R Markdown
cor(Dataset[,c("Performance", "Temps")], use="complete")
RegModel.1 <- lm(Performance~Temps, data=Dataset)
summary(RegModel.1)

Sortie
Residuals:
  Min       1Q   Median       3Q      Max
-6.6459 -3.6632  0.1866  3.8201  6.9866

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  21.3971    1.4829   14.429 5.14e-13 ***
Temps        1.6833    0.2085    8.075 3.66e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.293 on 23 degrees of freedom
Multiple R-squared:  0.7392, Adjusted R-squared:  0.7279
F-statistic: 65.2 on 1 and 23 DF, p-value: 3.657e-08

Messages
[1] NOTE: R Commander Version 2.2-3: Tue Jan 12 18:47:27 2016
[2] NOTE: Le jeu de données Dataset a 25 lignes et 3 colonnes.

```

L'équation de la droite de régression linéaire (équation des moindres carrés) est la suivante :

$$\text{Score prédit Performance} = (1,68 \times \text{Temps}) + 21,397$$

Le coefficient de détermination  $R^2$  est la proportion de variabilité de la variable dépendante (critère), ici la performance, qui est expliquée par la variable indépendante (prédictive), ici le temps. Autrement dit, environ 74% de la variabilité de la performance à l'examen est expliquée par le temps d'entraînement, ou encore la performance et le temps ont 74% de variabilité commune, ou encore que 74% de la variabilité totale est expliquée par la régression. Rappelons que dans le cas de la régression linéaire simple, c'est-à-dire entre 2 variables, le coefficient de détermination  $R^2$  est tout simplement le carré de la corrélation entre les 2 variables ( $0,86^2 = 0,74$ ).

Le coefficient standardisé  $\beta$  n'est pas calculé ici. Rappelons que dans le cas d'une régression simple, il est égal à la corrélation entre  $X$  et  $Y$ ., ici 0,86

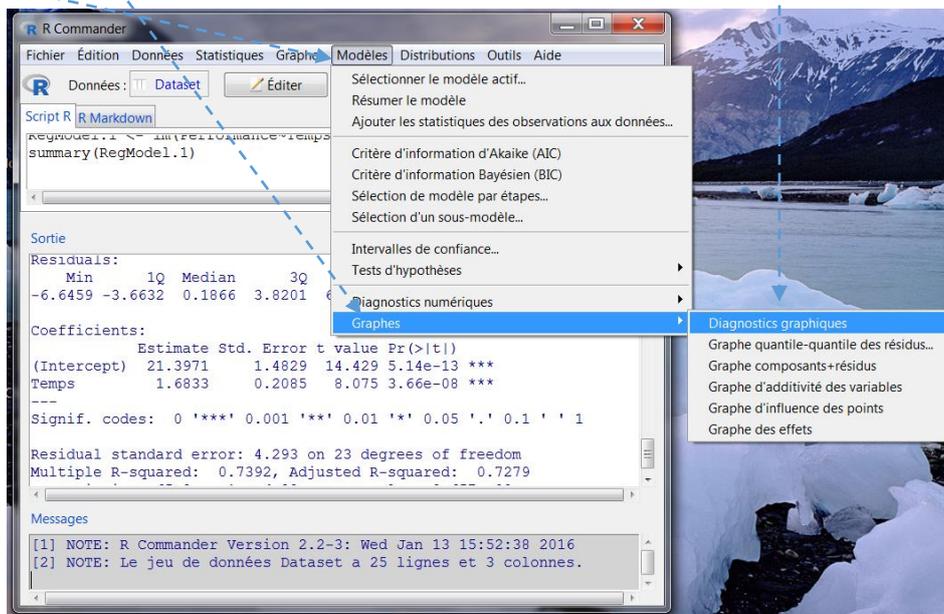
#### 13.4 Les tests d'hypothèse :

Les résultats de R Commander affichent également les tests d'hypothèse sur les coefficients de la régression  $b$  et  $a$  (test  $t$ ) et sur le modèle (ANOVA). On voit ainsi par exemple que le test  $F$  d'une valeur de 65,2 est significatif à .0005 (rappelons que cette ANOVA teste l'hypothèse nulle du  $R^2 = 0$ ).

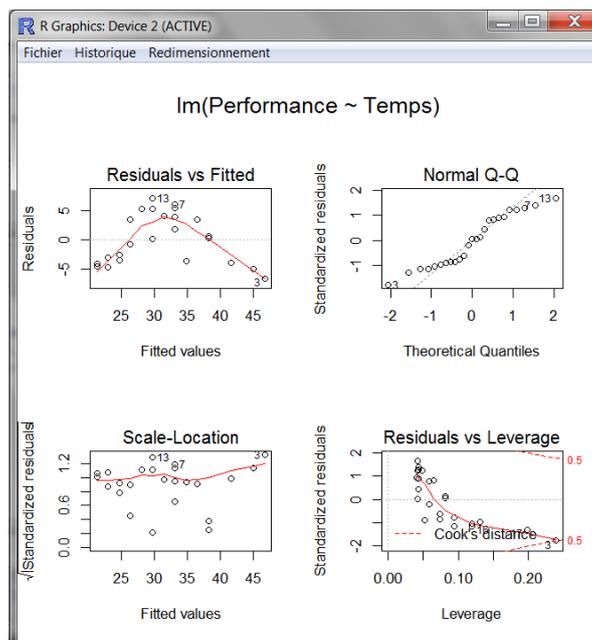
Mais rappelons ici que si l'on se limite aux seules données de l'échantillon, ces tests d'hypothèse sont inutiles, de même que les deux conditions ci-après. En revanche, si l'on souhaite faire des inférences sur la population dont sont issues ces données d'échantillon, il est nécessaire de vérifier au préalable les deux conditions de la régression simple avant d'interpréter ces tests : l'homogénéité des variances et de normalité dans les vecteurs.

### 13.4.1 Vérifier les conditions d'homogénéité des variances et de normalité dans les vecteurs en observant les graphiques

Pour vérifier l'homogénéité des variances dans les vecteur (encore appelée homoscedasticité ou égalité des variances dans les vecteurs) et de normalité dans les vecteurs, on peut représenter graphiquement respectivement les résidus par rapport aux valeurs prédites et faire un QQ plot entre les résidus normalisés et loi normale. Pour ce faire, il suffit d'aller dans Modèle, Graphes et Diagnostics graphiques.



La fenêtre suivante avec 4 graphiques apparaît :



Les deux graphiques du haut nous intéressent ici.

Dans le graphique en haut à gauche, les résidus (*residuals*) sont représentés par rapport aux valeurs calculées par la régression (*fitted values*). S'il y avait homogénéité des variances dans les vecteurs, le nuage des points ne devrait pas avoir de schéma particulier, la dispersion verticale ne devrait pas augmenter ou diminuer de façon systématique. Ce rassemblement aléatoire devrait plutôt avoir une forme d'un rectangle. La ligne rouge devrait se situer le long de celle en tirets bleus. Or, ce n'est pas le cas ici : on peut difficilement admettre l'homogénéité des variances dans les vecteurs. L'utilisation des tests d'hypothèse n'est pas conseillée. A titre d'exemple, poursuivons néanmoins avec la seconde condition.

Le second graphique en haut à droite est le QQ plot entre les résidus normalisés (moyenne 0, écart-type 1) et loi normale  $N(0, 1)$ . Il permet de vérifier graphiquement leur normalité. Ce qui n'est pas le cas ici, les valeurs élevées et les valeurs faibles s'éloignent fortement de l'axe. L'utilisation des tests d'hypothèse n'est pas conseillée.

Si l'on souhaite obtenir ces graphiques dans des fenêtres séparées :

Pour le graphique des résidus par rapport aux valeurs calculées :

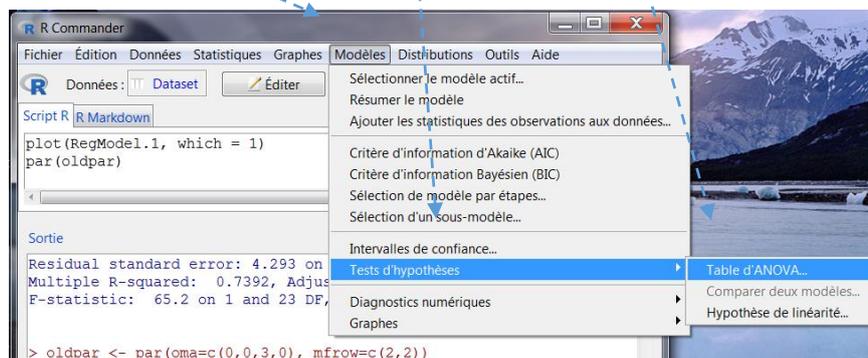
ajouter 'which = 1' dans la fenêtre Script R, la commande devient : `plot(RegModel.1, which = 1)`, puis surligner et enfin cliquer sur soumettre ;

Pour le QQ plot :

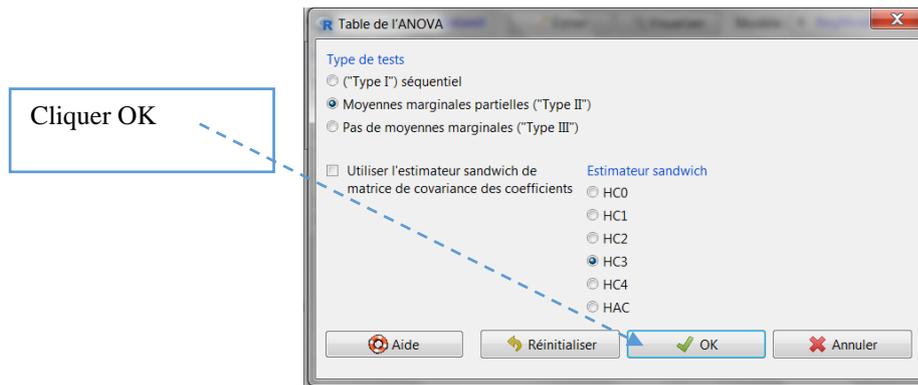
ajouter 'which = 2' dans la fenêtre Script R, la commande devient : `plot(RegModel.1, which = 2)`, puis surligner et enfin cliquer sur soumettre.

#### 13.4.2 Pour aller plus loin

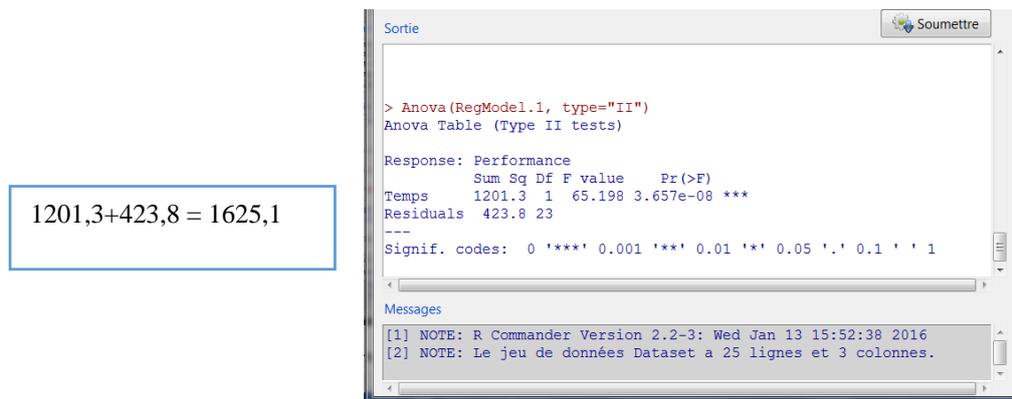
Si l'on souhaite connaître la décomposition de la variabilité totale de la variable réponse, (Somme des Carrés Totale, SCT) en somme des carrés de la régression (SCR) et des résidus ou Erreurs (SCE), (soit  $SCT = SCR + SCE$ ), il faudra effectuer une ANOVA de la réponse de  $Y$  sur  $X$ . Choisir Modèles, Tests d'hypothèses, Table d'ANOVA



La fenêtre Table d'ANOVA s'ouvre :

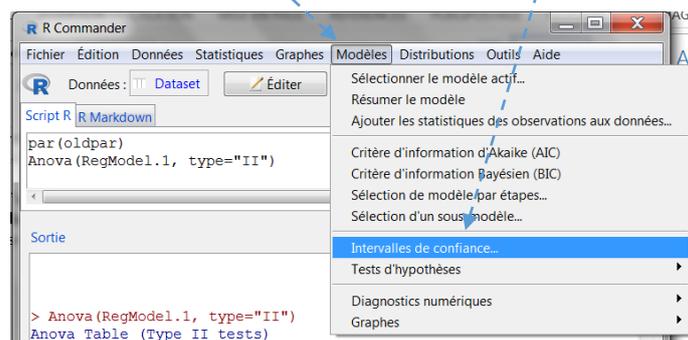


La fenêtre de sortie apparaît :

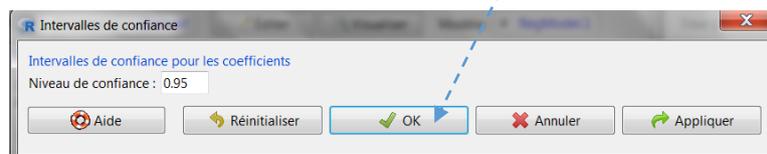


On remarquera que la valeur du F de cette ANOVA de la réponse de Y sur X est la même que celle calculée plus haut sur le modèle.

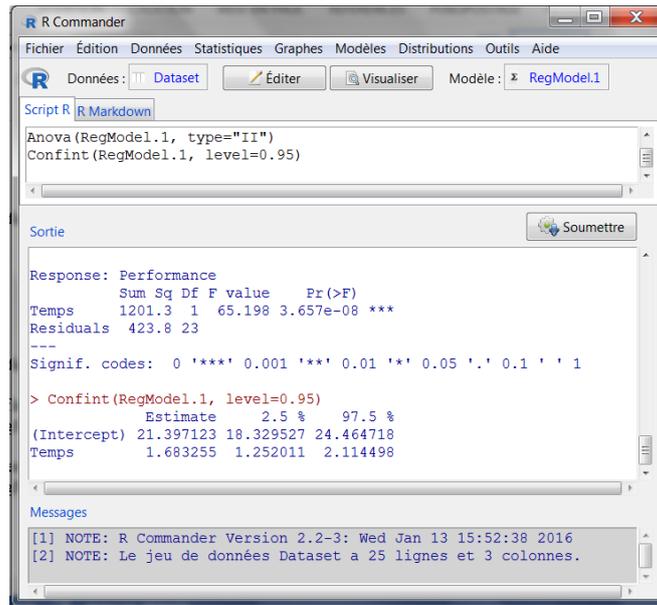
Si l'on souhaite calculer les intervalles de confiance sur les coefficients b et a, avec une probabilité alpha = 0.95, choisir Modèles et Intervalles de confiance



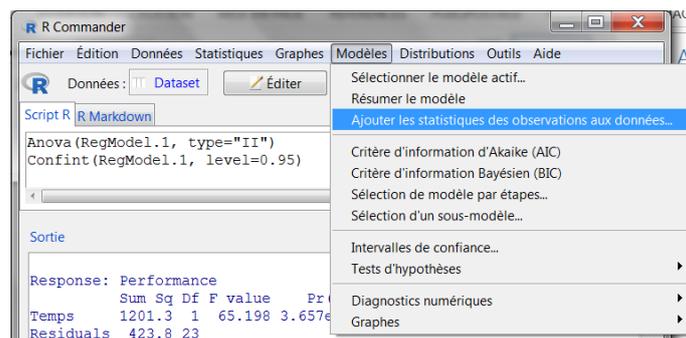
La fenêtre Intervalles de confiance s'ouvre et cliquer OK



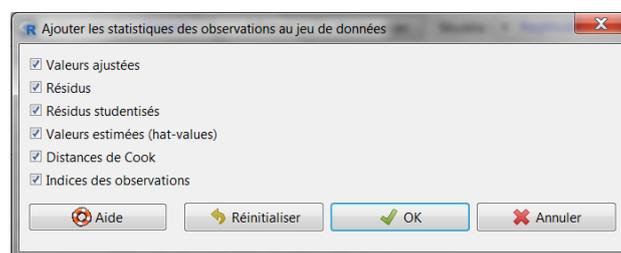
La fenêtre sortie apparaît avec tout en bas, les estimations de la constante et de la pente, suivies de leurs intervalles respectifs



On peut enfin obtenir les valeurs ajustées, les résidus et d'autres statistiques du modèle et les ajouter au tableau des données très simplement.



La fenêtre Ajouter les statistiques des observations au jeu de données s'ouvre et l'on choisit les statistiques souhaitées, puis OK (ce que nous ne ferons pas ici).



Enfin, pour visualiser dans la fenêtre Résultats les valeurs calculées par la régression d'après les variables explicatives : `RegModel.1$fitted.values` (ou `fitted(RegModel.1)`) ;

et les valeurs des résidus (c'est à dire, les valeurs effectives moins les valeurs calculées : `RegModel.1$residuals` (ou `residuals(RegModel.1)`).

## 14 Références

Cornillon, P.-A., Guyader, A., Husson, F., Jégou, N., Josse, J., Kloareg, M., Matzner-Lober, E., & Rouviere, L. (2010). *Statistiques avec R* (2<sup>nd</sup> ed.). Rennes, France : Presses Universitaires de Rennes.

Fox, J. (2005). The R Commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software*, 14, 1-42.

Howell, D.C. (2008). *Méthodes statistiques en sciences humaines* (2<sup>nd</sup> ed.). Bruxelles, Belgique : de Boek.

Howell, D.C. (2006). *Statistical method for psychology* (6<sup>th</sup>). Belmont, CA: Wadsworth Publishing Co Inc.

Karp, N.A. (2014). *R Commander: An introduction*. Retiré de <https://cran.r-project.org/doc/contrib/Karp-Rcommander-intro2.pdf>.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retiré de <http://www.R-project.org/>.

## 15 Annexes

Installation sous Windows

Installation sous Mac

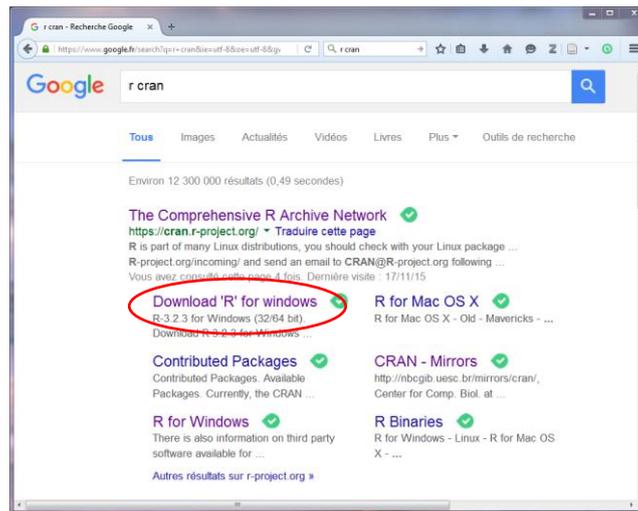
Installation sous Ubuntu

## 15.1 Installer R et R Commander sous Windows

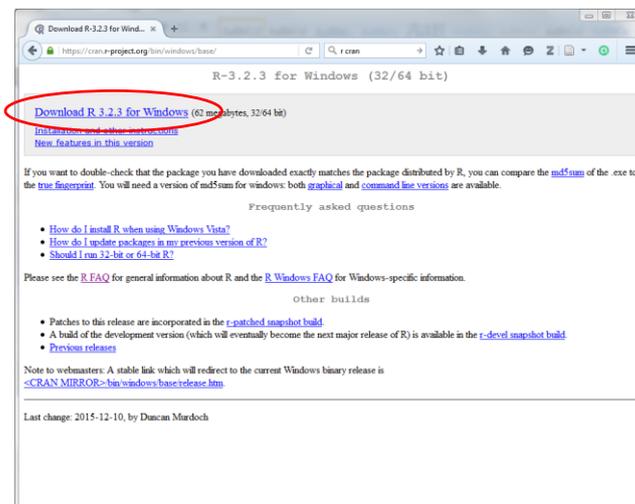
R Commander est une interface graphique de R. Il faut donc installer R, puis R Commander. La procédure est la même que pour la plupart des téléchargements sous Windows : aller sur le site choisi, cliquer sur télécharger et suivre la procédure en cliquant ‘OK’ ou ‘Accepter’ ou encore ‘Suivant’ lorsque les fenêtres s’ouvrent et le demandent.

A partir d’un moteur de recherche (ici Google) chercher ‘r cran’ ou encore taper l’adresse ‘cran.r-project.org’

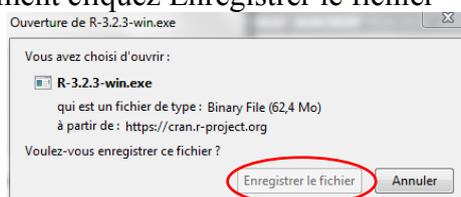
Parmi les choix proposés, choisir (1) le site The Comprehensive R Archive Network (cran) et (2) ‘Downlaod ‘R’ for windows’ (télécharger R sous windows (le français arrive bientôt...)).



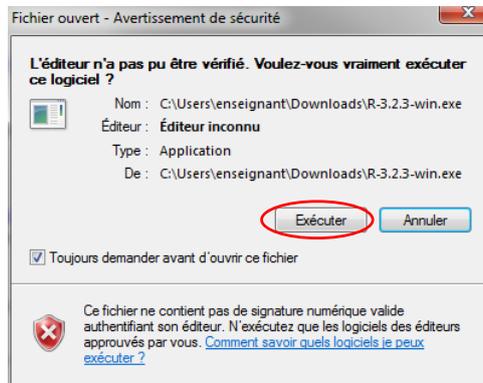
Dans la fenêtre qui s’ouvre cliquer sur ‘Download R.3.2.3. for Windows’ (ou la version plus récente depuis l’écriture de ce guide)



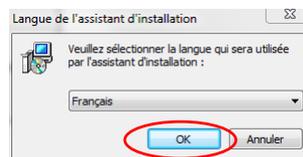
A l’invite pour le téléchargement cliquez Enregistrer le fichier



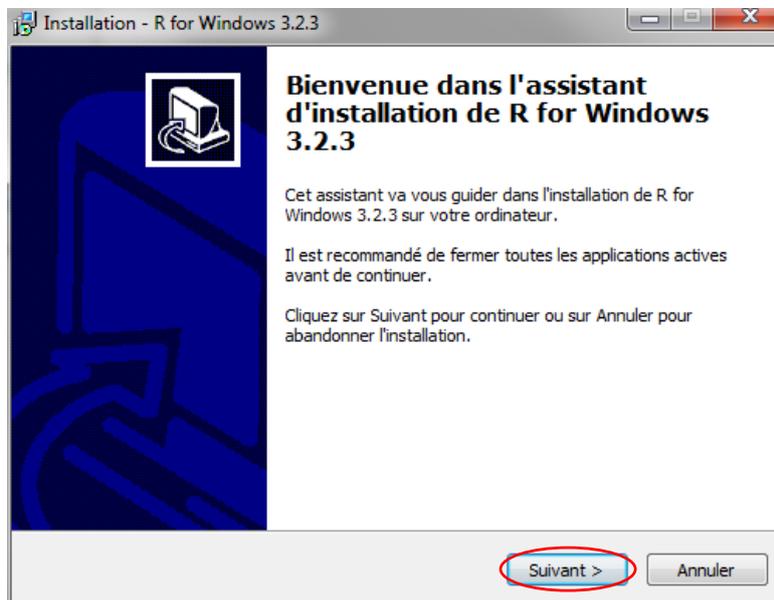
Lorsque le fichier est téléchargé, soit à partir de la fenêtre téléchargement, soit dans le dossier Download, double-cliquer sur R.3.2.3. win.exe pour l'exécuter.  
Si le logiciel Antivirus le demande, accepter l'exécution : Exécuter

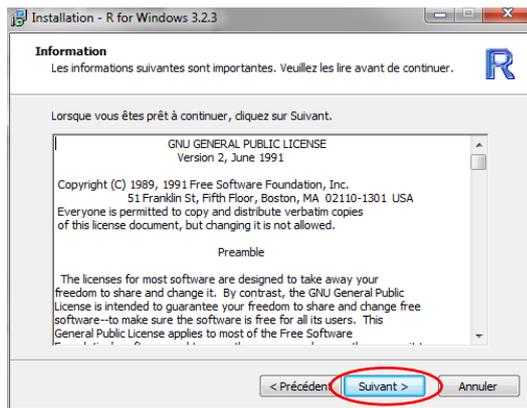


Choisir la langue d'installation et OK

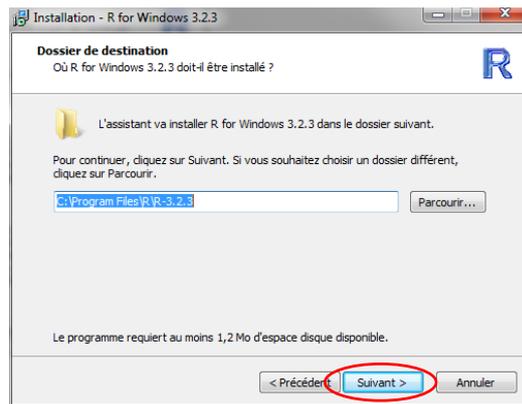


La fenêtre d'installation Windows s'ouvre : il suffit de suivre les propositions par défaut en cliquant suivant à chaque fenêtre

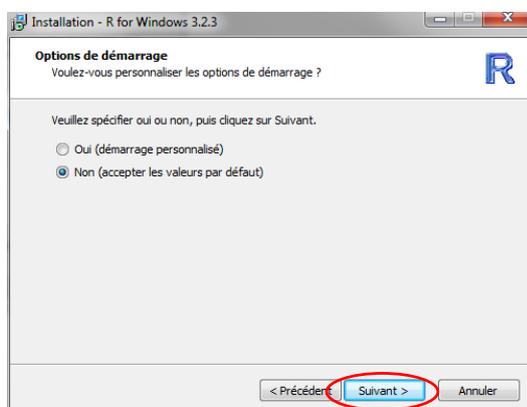
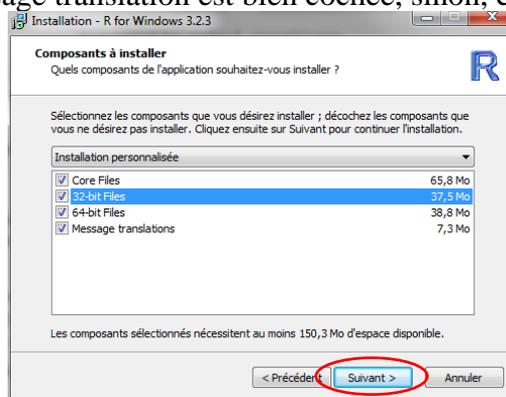




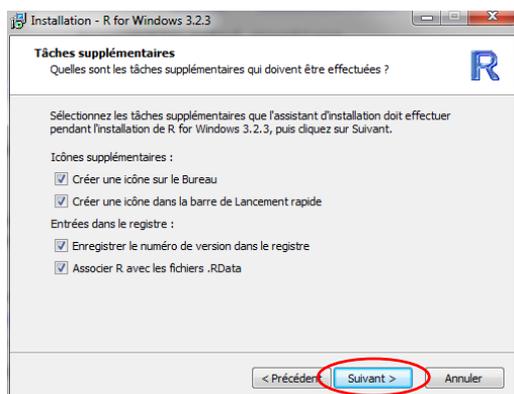
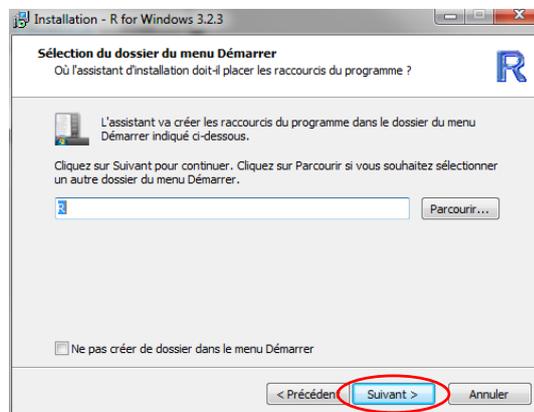
puis



Pour la fenêtre suivante, si le système d'exploitation est en 64 bits, la fenêtre comportera les 4 options cochées. S'il s'agit d'un système en 32 bits, l'option 64 bits files ne sera pas cochée. Vérifiez que l'option Message translation est bien cochée, sinon, cochez-la.



puis



puis



C'est fini. Si le système est en 32 bits, il y a une icône R i386 3.2.3 :



Si le système est en 64 bits, il y a 2 icônes, celle-ci-dessus, Ri386 3.2.3 et celle-là Rx64 3.2.3



Dans ce cas, on ne pourra garder que Rx64, l'autre étant mis dans la poubelle. En cliquant sur l'icône R du bureau, R s'ouvre avec la fenêtre suivante :

```

R Console (64-bit)
Fichier Edition Misc Packages Fenêtres Aide

R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |

```

### Installer R Commander

R Commander est installé ici avec une version de R 32 bits. Les deux fenêtres R se ressemblent (voir celle en 64 bits ci-dessus, et avec quelques icônes dans la fenêtre en 32 bits ci-dessous). Mais la procédure d'installation de R Commander est strictement identique pour R en 64 bits

```

RGui (32-bit)
Fichier Edition Voir Misc Packages Fenêtres Aide

R Console
R version 3.2.3 (2015-12-10) -- "Wooden Christmas-Tree"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

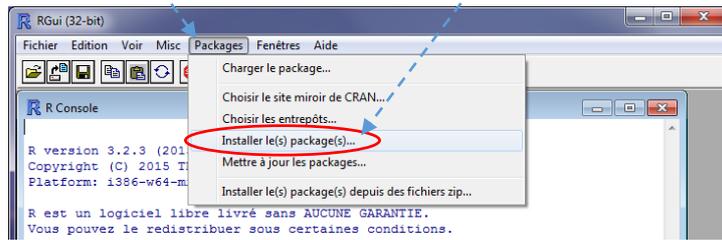
R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

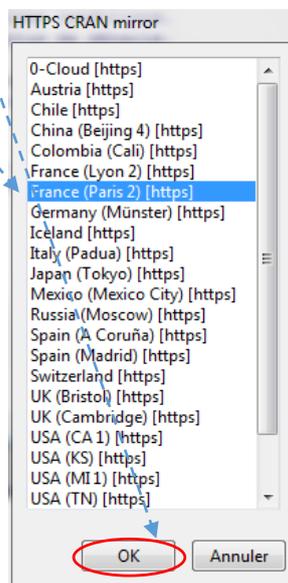
> |

```

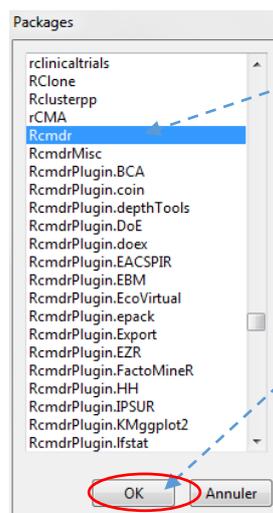
Aller dans le menu Packages et Installer le(s) package(s)



La fenêtre HTTPS CRAN mirror s'ouvre. En effet, l'installation des paquets de R se fait depuis des sites miroirs pour faciliter le téléchargement. Il faut choisir le site français le plus proche, ici France (Paris 2), puis OK :

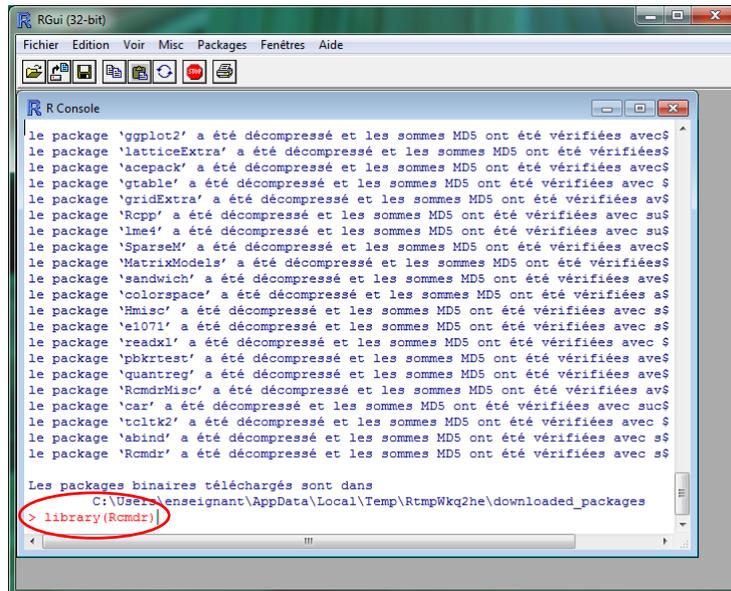


La fenêtre Packages s'ouvre. Chercher (par ordre alphabétique) Rcmdr, le sélectionner et OK

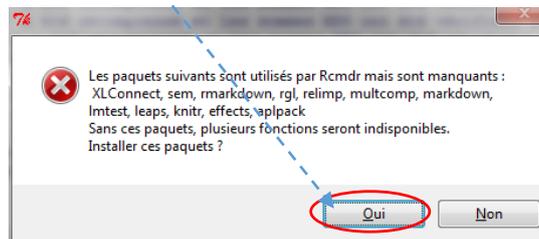


Rcmdr se télécharge et télécharge également des packages supplémentaires : le téléchargement peut prendre quelques minutes.

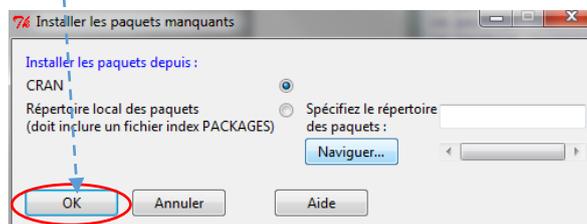
Lorsque le téléchargement est terminé, la fenêtre R ne défile plus et à la toute dernière ligne apparaît le prompt `>`. Il faut écrire la commande `library(Rcmdr)` et entrer



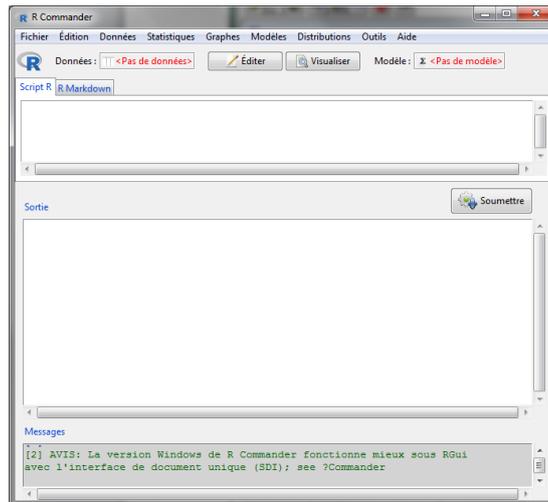
À la toute première ouverture de R Commander, une fenêtre peut s'ouvrir indiquant qu'il manque encore des paquets afin que toutes les fonctions soient disponibles. Il faut installer ces paquets manquants en cliquant Oui



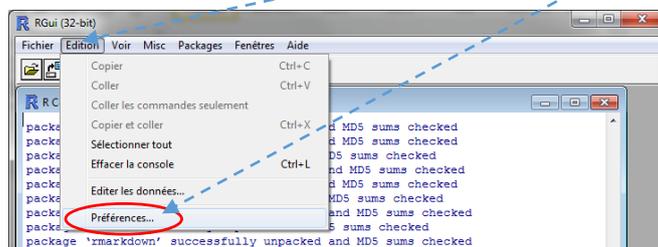
La fenêtre Installer les paquets manquants s'ouvre avec l'option Installer les paquets depuis CRAN cochée. Cliquer OK



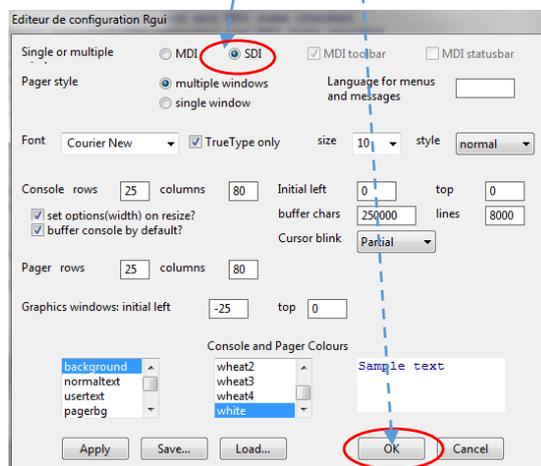
Voici la fenêtre R Commander. Tout en bas, en vert, un avis signale que R Commander fonctionne mieux avec SDI.



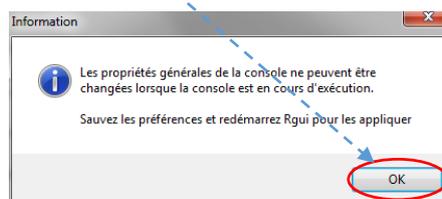
Il faut change l'interface. Dans le menu, ouvrir Edition et Préférences



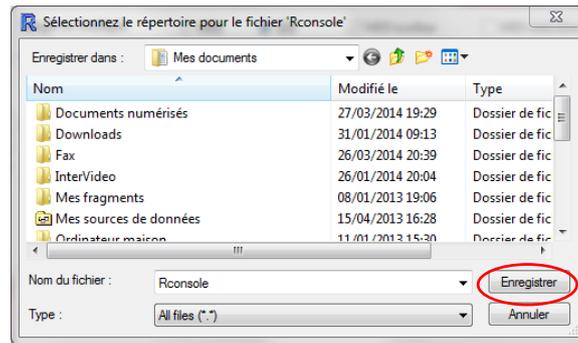
La fenêtre Préférences s'ouvre. Cliquer SDI, puis OK



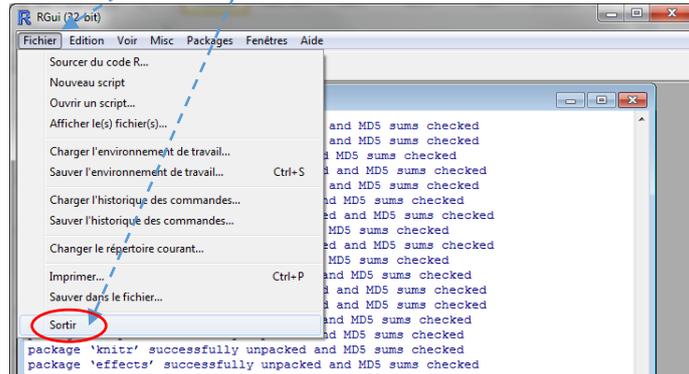
Dans la fenêtre Information, cliquer OK



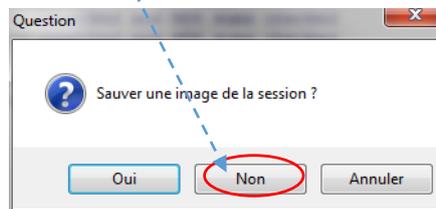
Dans la fenêtre suivante, cliquer sur Enregistrer :



Il faut fermer R avec Fichier et Sortir



Dans la dernière fenêtre, cliquer Non



Pour ouvrir R Commander, il faudra toujours ouvrir R, puis dans la fenêtre, écrire en face du prompt > la commande library(Rcmdr) et entrer.

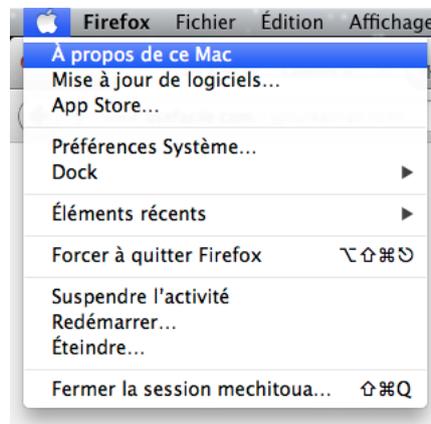
## 15.2 Installer R et R Commander sous Mac

R Commander est une interface graphique de R. Il faut donc installer R, puis R Commander. La procédure est la même que pour la plupart des téléchargements sous Mac : aller sur le site choisi, cliquer sur télécharger et suivre la procédure en cliquant 'OK' ou 'Accepter' ou encore 'Suivant' lorsque les fenêtres s'ouvrent et le demandent, en entrant son mot de passe...

La procédure est un peu longue parce que répétitive : il y a en effet trois fichiers à télécharger et à installer pour R, puis celui de R Commander, et enfin un logiciel spécifique X11 s'il n'est pas déjà installé sur l'ordinateur. Par ailleurs, la version de R dépendra de l'OS sur lequel tourne le Mac. C'est donc la première information qu'il faut rechercher et noter.

Rechercher la version de l'OS

Pour connaître la version de l'OS sur lequel tourne le Mac, il faut cliquer sur l'icône  tout en haut à gauche de la fenêtre d'accueil et dans le menu déroulant, cliquer sur A propos de ce Mac



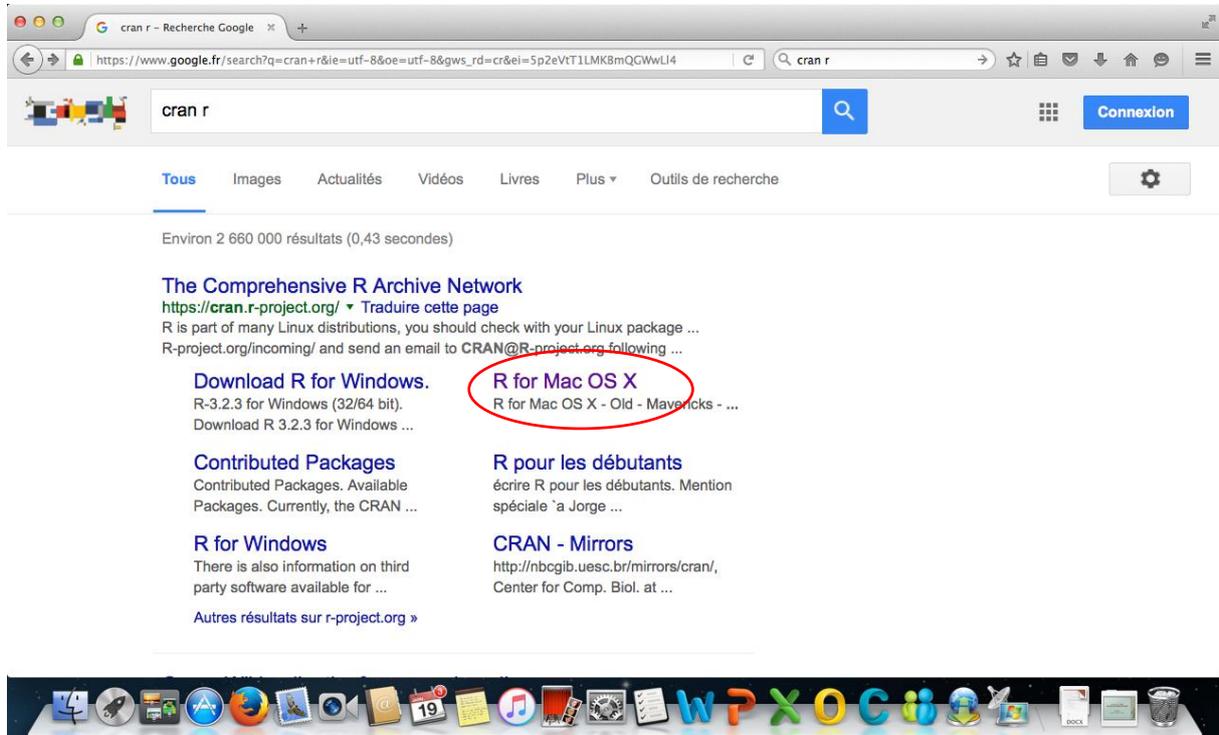
La fenêtre A propos de ce Mac s'ouvre indiquant la version utilisée, ici 10.8.5



## Télécharger R

A partir d'un moteur de recherche (ici Google) chercher 'r cran' ou encore taper l'adresse 'cran.r-project.org'

Parmi les choix proposés, choisir (1) le site The Comprehensive R Archive Network (CRAN) et (2) 'R for Mac OS X' (R pour Mac OS X (le français arrive bientôt...)).



Dans la fenêtre qui s'ouvre, il faudra télécharger 3 fichiers : le premier est sur la page, c'est un fichier R. Selon la version du Mac OS, cliquer sur la version R correspondante. Ici, nous téléchargeons R-3.2.1.Snowleopard.pkg., puisque la version MacOS utilisée est 10.8.

**R for Mac OS X**

This directory contains binaries for a base distribution and packages to run on Mac OS X (release 10.6 and above). Mac OS 8.6 to 9.2 (and Mac OS X 10.1) are no longer supported but you can find the last supported release of R for these systems (which is R 1.7.1) [here](#). Releases for old Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old](#) directory.

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

**R 3.2.3 "Wooden Christmas-Tree" released on 2015/12/10**

Please check the MD5 checksum of the downloaded image to ensure that it has not been tampered with or corrupted during the mirroring process. For example type `md5 R-3.2.3.pkg` in the *Terminal* application to print the MD5 checksum for the R-3.2.3.pkg image. On Mac OS X 10.7 and later you can also validate the signature using `pkgutil --check-signature R-3.2.3.pkg`

**Files:**

[R-3.2.3.pkg](#)  
MD5-hash: 6418ea2897709a230397d824231ch743  
SHA1-hash: e3ca68d6f414022e5c177ac0d0e01764cccd23  
(ca. 70MB)

**R 3.2.3** binary for Mac OS X 10.9 (Mavericks) and higher, signed package. Contains R 3.2.3 framework, R.app GUI 1.66 in 64-bit for Intel Macs, Tcl/Tk 8.6.0 X11 libraries and Texinfo 5.2. The latter two components are optional and can be omitted when choosing "custom install", it is only needed if you want to use the `tcltk` R package or build package documentation from sources.

Note: the use of X11 (including `tcltk`) requires [XQuartz](#) to be installed since it is no longer part of OS X. Always re-install XQuartz when upgrading your OS X to a new major version.

[R-3.2.1-snowleopard.pkg](#)  
MD5-hash: 586e9d01314d8cb75f0f0cc914f6a5  
SHA1-hash: 5ef6e91db12bac22a32d013a7e7ef6a9063cecd01  
(ca. 68MB)

**R 3.2.1** legacy binary for Mac OS X 10.6 (Snow Leopard) - 10.8 (Mountain Lion), signed package. Contains R 3.2.1 framework, R.app GUI 1.66 in 64-bit for Intel Macs. This package contains the R framework, 64-bit GUI (R.app), Tcl/Tk 8.6.0 X11 libraries and Texinfo 5.2. GNU Fortran is **NOT** included (needed if you want to compile packages from sources that contain FORTRAN code) please see [the tools directory](#).  
NOTE: the binary support for OS X before Mavericks is being phased out, we do not expect further releases!

Ensuite, glisser pour voir le bas de la fenêtre et cliquer sur tools :

[directory](#).  
NOTE: the binary support for OS X before Mavericks is being phased out, we do not expect further releases!

[Mac-GUI-1.66.tar.gz](#)  
MD5-hash: 450c99bab21178b2986db89e32771a

[NEWS](#) (for Mac GUI) News features and changes in the R.app Mac GUI

The new R.app Cocoa GUI has been written by Simon Urbanek and Stefano Iacus with contributions from many developers and translators world-wide, see "About R" in the GUI.

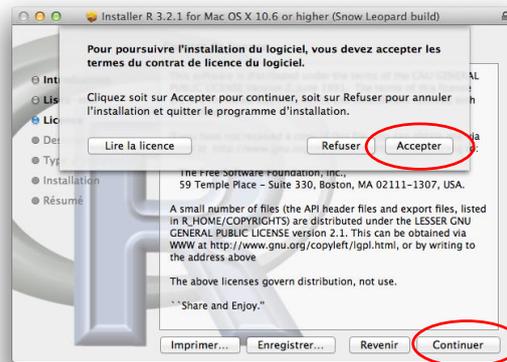
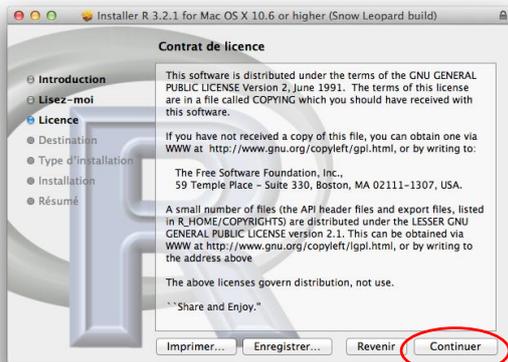
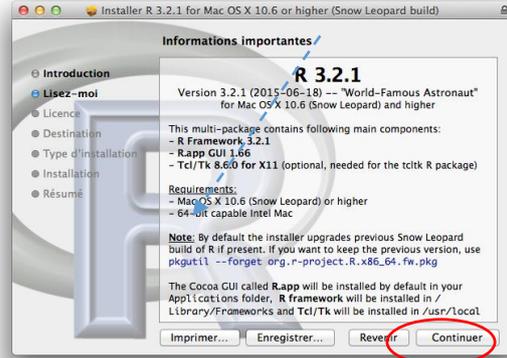
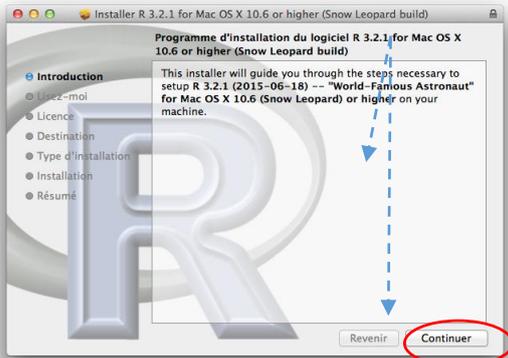
**Subdirectories:**

Additional tools necessary for building R for Mac OS X:  
Universal GNU Fortran compiler for Mac OS X (see [R for Mac tools page](#) for details).  
Binaries of package builds for Mac OS X 10.6 or higher (Snow Leopard build)  
Binaries of package builds for Mac OS X 10.9 or higher (Mavericks build)  
Legacy binaries of universal (32-bit and 64-bit) package builds for Mac OS X 10.5 or higher (Leopard build)  
Legacy binaries of universal (32-bit) package builds for Mac OS X 10.4 (Tiger build)  
Previously released R versions for Mac OS X

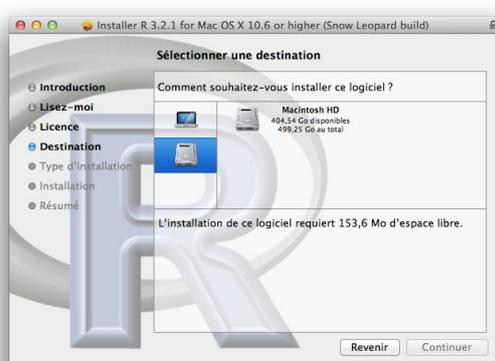
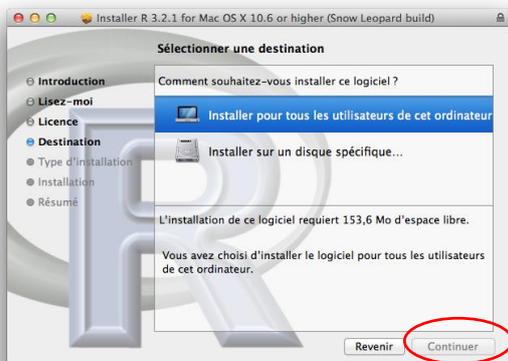
You may also want to read the [R FAQ](#) and [R for Mac OS X FAQ](#). For discussion of Mac-related topics and reporting Mac-specific bugs, please use the [R-SIG-Mac](#) mailing list.

La fenêtre tools s'ouvre et il y a deux fichiers à télécharger : un fichier gfortran et un fichier tcltk.

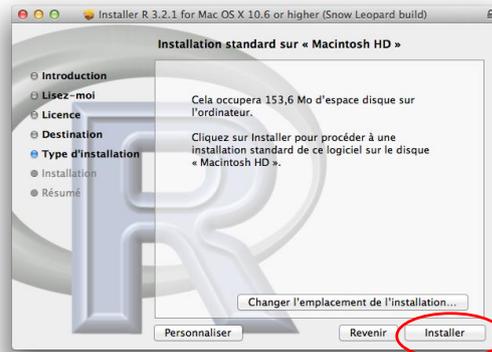
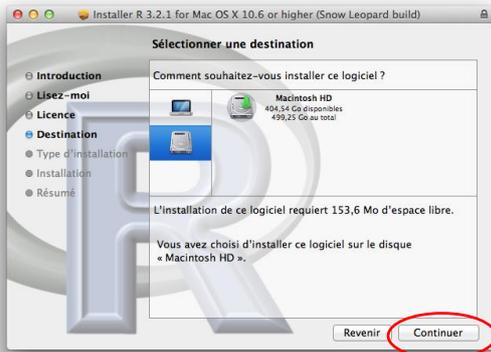




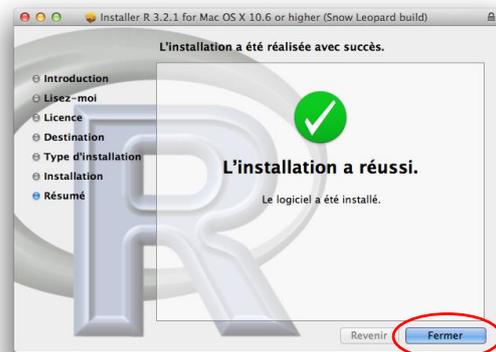
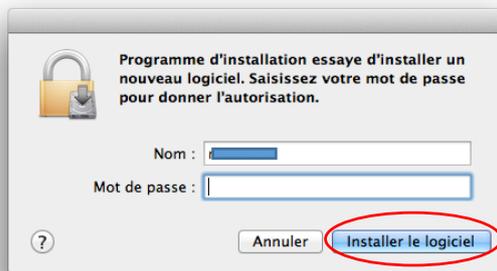
Dans les fenêtres suivantes, il faut choisir sur quel support installer R. Nous avons choisi ici Installer pour tous les utilisateurs et dans la fenêtre suivante Macintosh HD



Le disque choisi est confirmé



Lors de l'installation, une fenêtre demande le mot de passe :

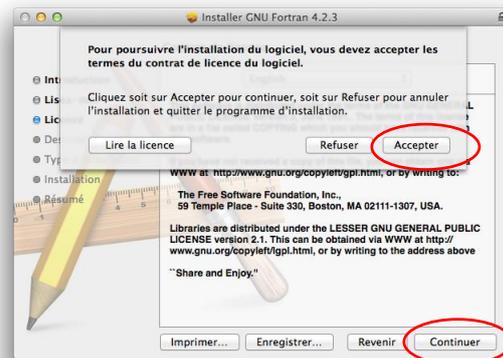
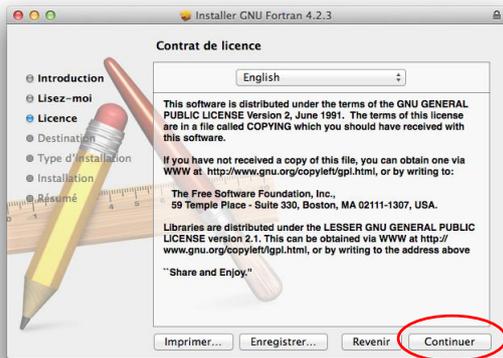


Lorsque l'installation de R est terminée, fermer la fenêtre. Il faut installer le second fichier gfortran

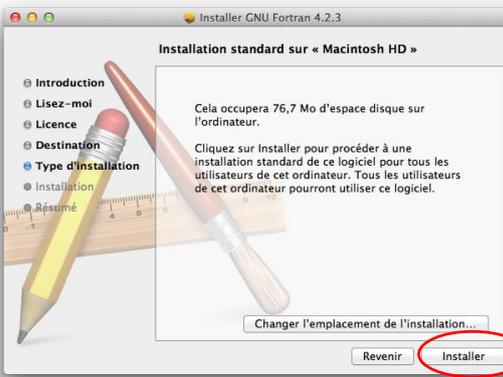
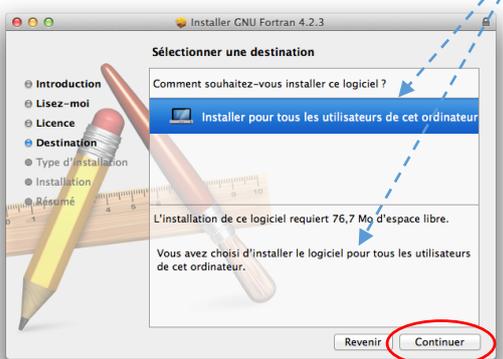
Installer le fichier gfortran

Revenir dans le dossier téléchargement et double cliquer sur le programme gfortran. La procédure d'installation reprend avec une succession de fenêtres assez similaires et pour lesquelles il faut continuer, accepter...

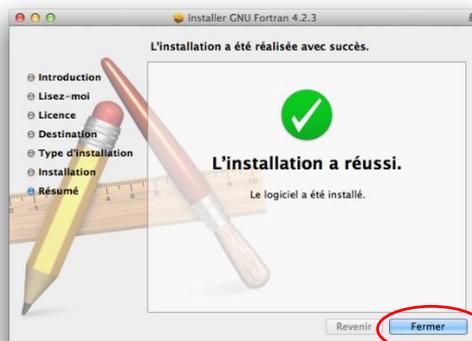
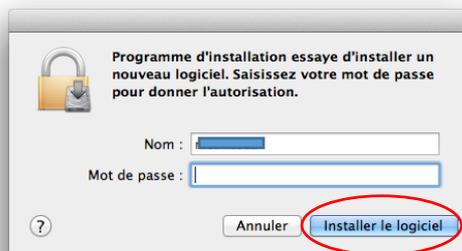




Choisir d'installer pour tous les utilisateurs



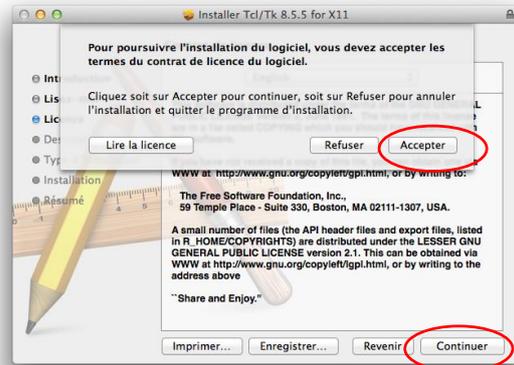
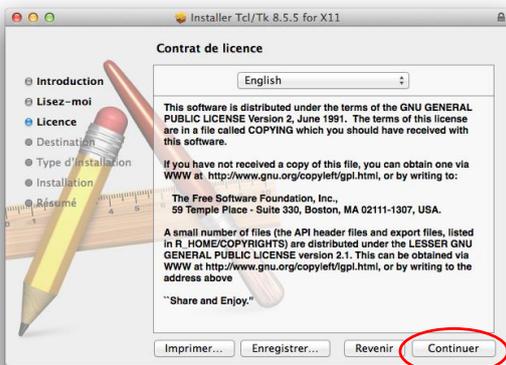
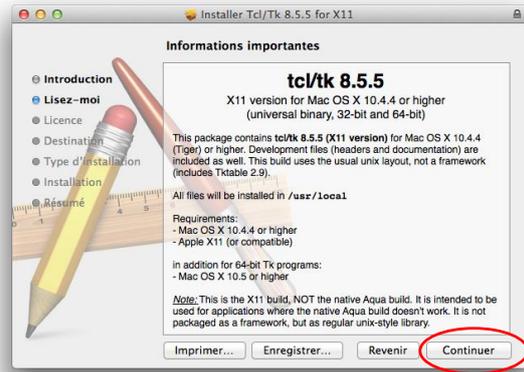
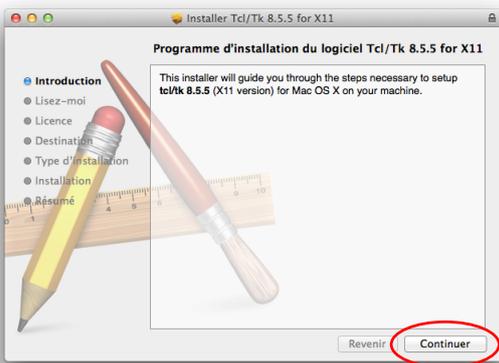
Au début de l'installation, le mot de passe est demandé



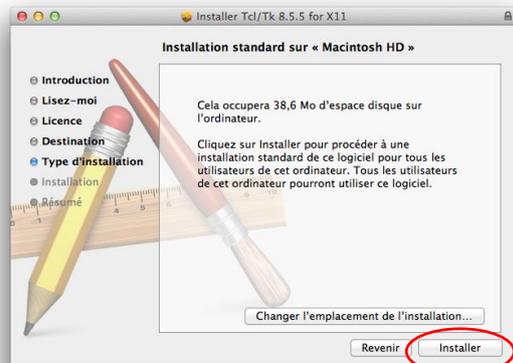
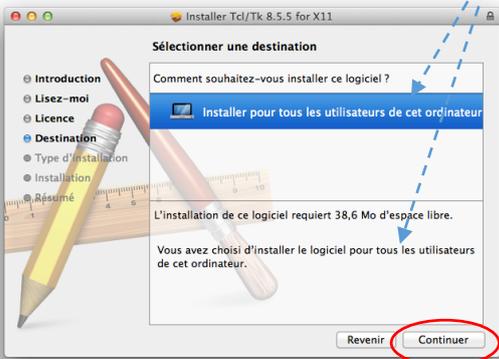
Le fichier gfortran est maintenant installé. Il reste le troisième fichier tcltk

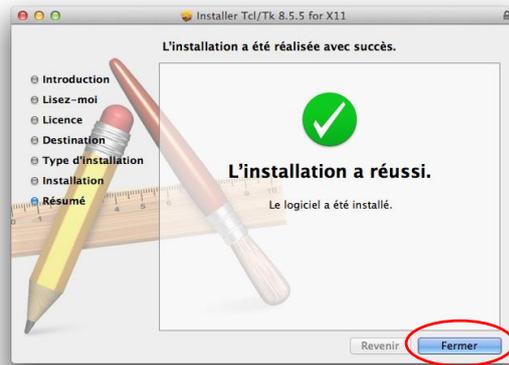
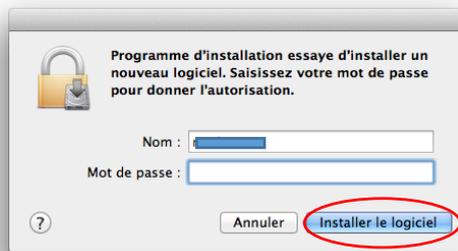
Installer le fichier tcltk

Revenir dans le dossier téléchargement et double cliquer sur le programme tcltk. La procédure d'installation reprend avec la même succession de fenêtres pour lesquelles il faut continuer, accepter...



Choisir Installer pour tous les utilisateurs.





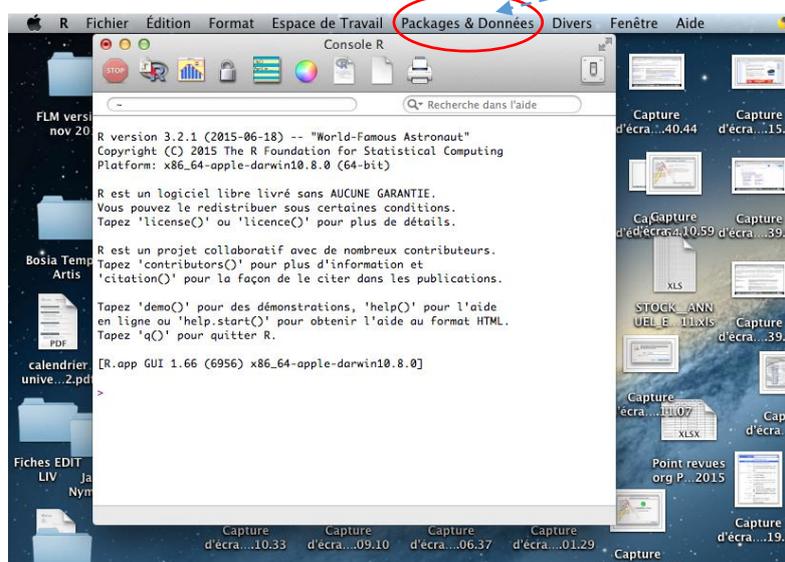
L'installation est terminée. La fenêtre se ferme.  
Une icône R a été créée



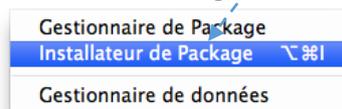
Pour ouvrir R, double cliquer sur l'icône R

## Installer R Commander

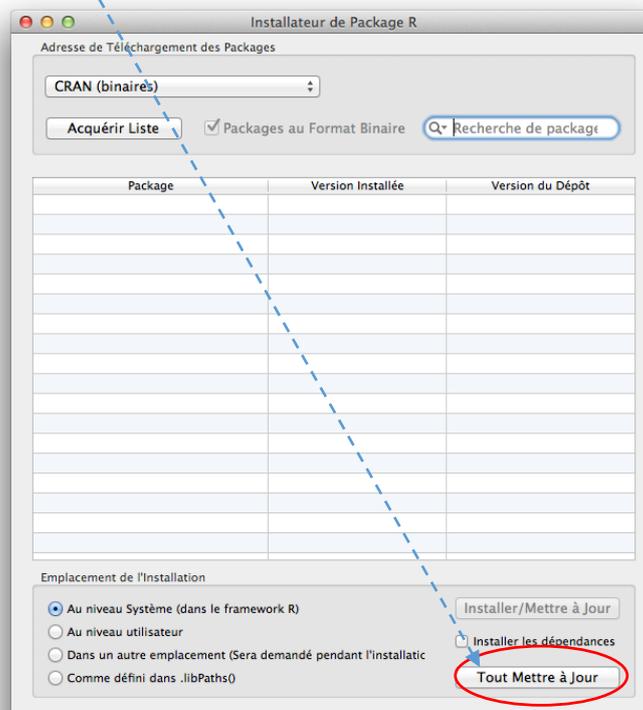
La fenêtre R est ouverte. Cliquer sur le menu Packages & Données



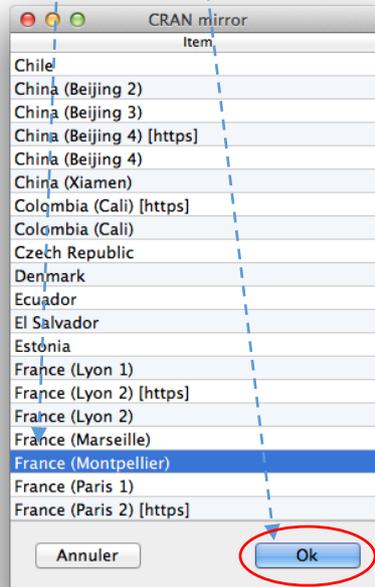
La fenêtre s'ouvre et choisir Installateur de Package,



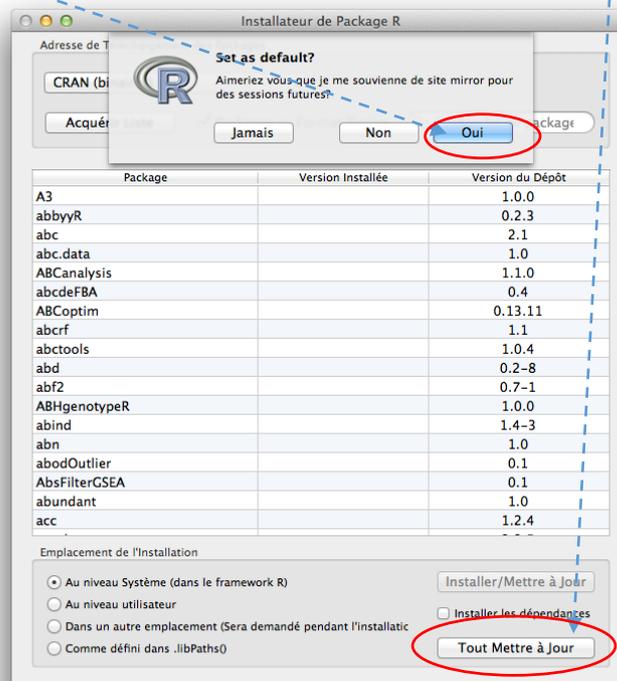
L'installateur de Package R apparaît. Il faut tout d'abord mettre à jour les packages de R en cliquant Tout mettre à jour



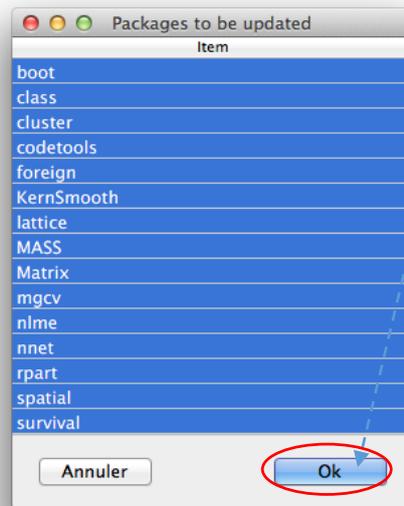
La fenêtre CRAN Mirror s'ouvre. Le téléchargement des packages se fait en effet à partir d'un site miroir pour aller plus vite. Choisir le site le plus proche dans la liste proposé. Nous choisissons ici France (Montpellier), puis OK



La liste des packages s'affichent, de même que la possibilité de garder en mémoire le site miroir. Choisir OUI puis tout en bas de la fenêtre, 'Tout mettre à jour'



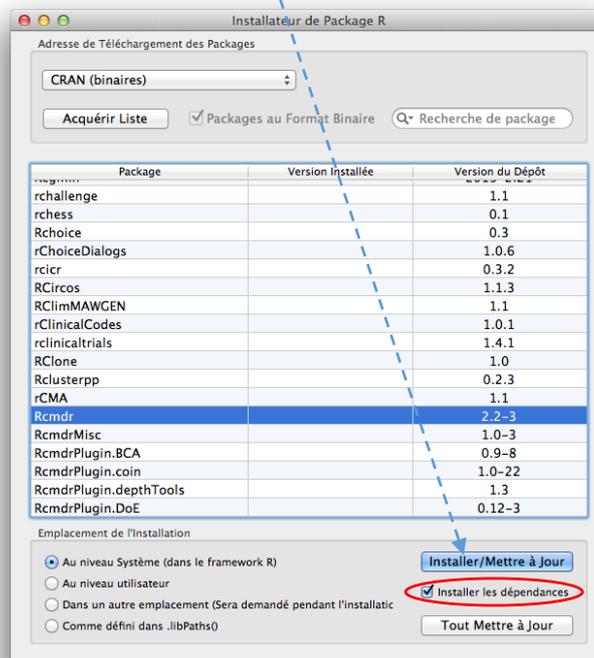
La fenêtre des packages devant être mis à jour s'ouvre. Cliquer OK.



Les packages sont téléchargés et installés dans R. On peut maintenant télécharger R Commander

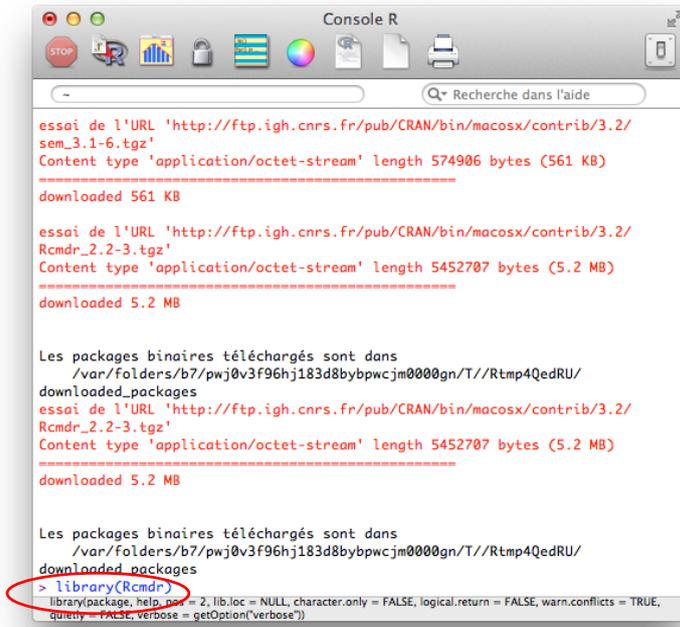
La fenêtre Installateur de package est ouverte. Dans l'ordre :

1. Dans la liste alphabétique des Package, chercher et sélectionner Rcmdr ;
2. cliquer sur Installer les dépendances ;
3. cliquer sur Installer/mettre à jour

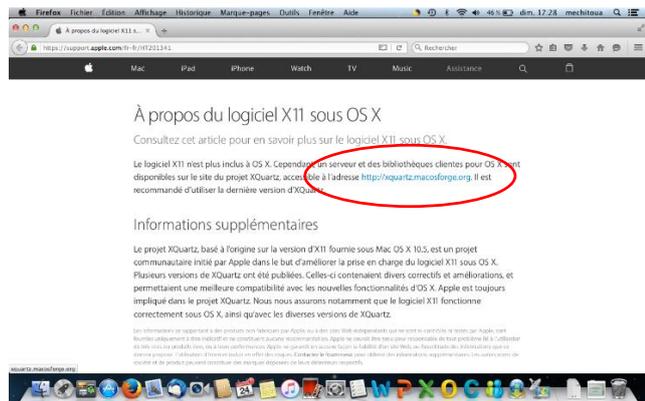


Le téléchargement commence, on peut voir sa progression dans la fenêtre R. Lorsqu'il est terminé, la fenêtre R arrête de défiler et le prompt > apparaît à la dernière ligne.

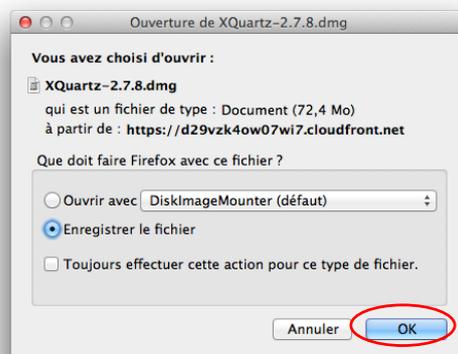
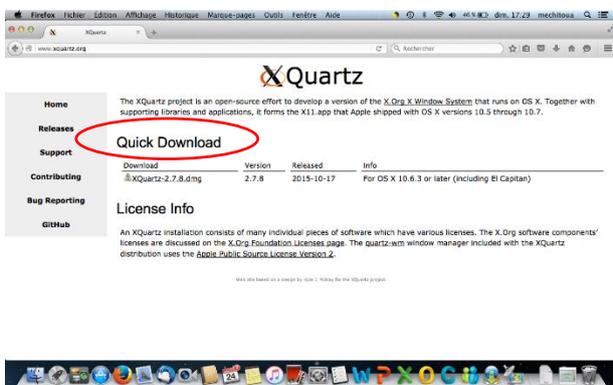
Pour ouvrir R Commander depuis la fenêtre R, écrire library(Rcmdr) devant > et entrer



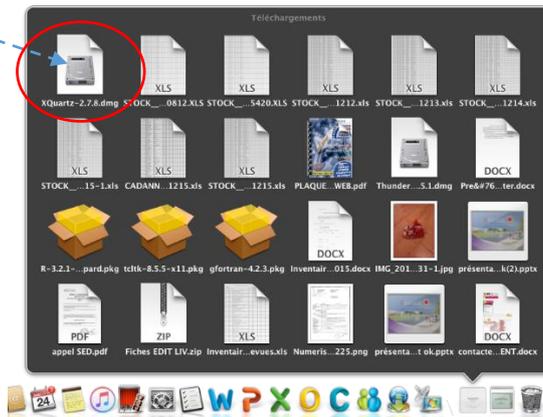
Une fenêtre demandant l'installation de X11 s'affiche s'il n'est pas déjà sur l'ordinateur. Ce sera le dernier téléchargement



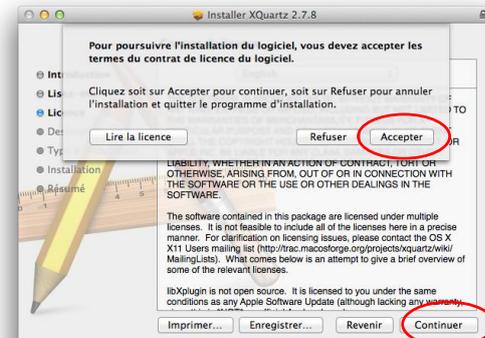
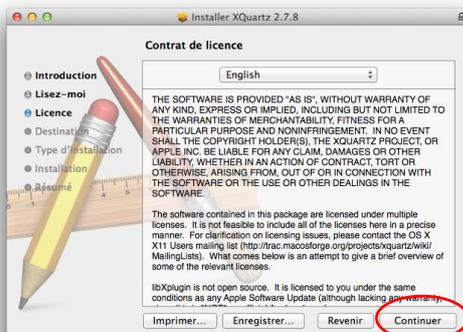
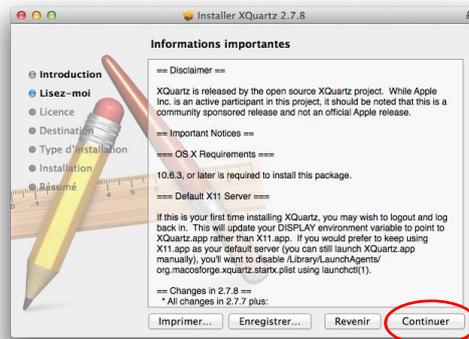
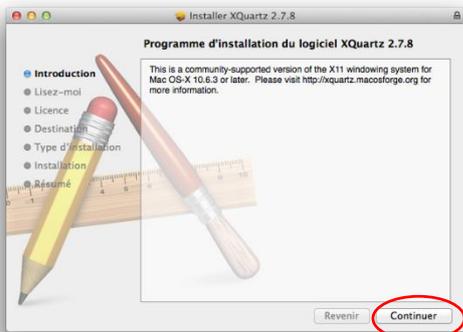
Cliquer sur Quick Download

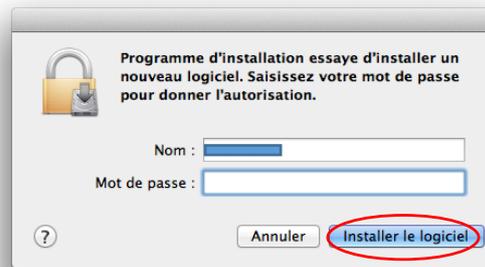
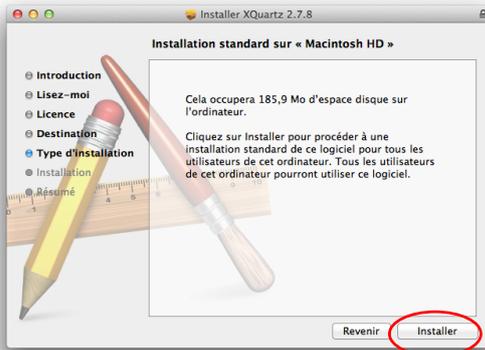


Lorsque XQuartz est téléchargé, aller dans le dossier de téléchargement et double cliquer sur XQuartz

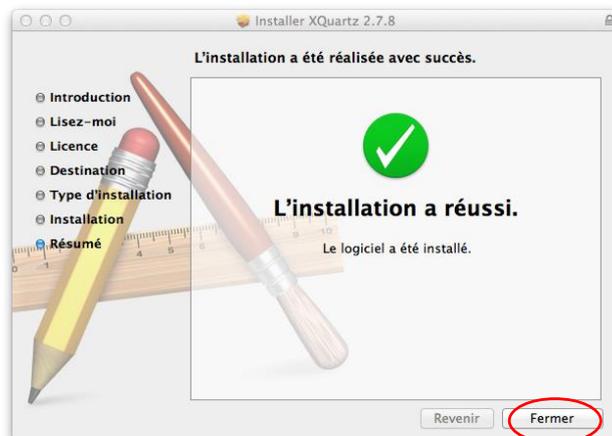


Suivre le programme d'installation, là également, cliquer sur continuer, accepter...



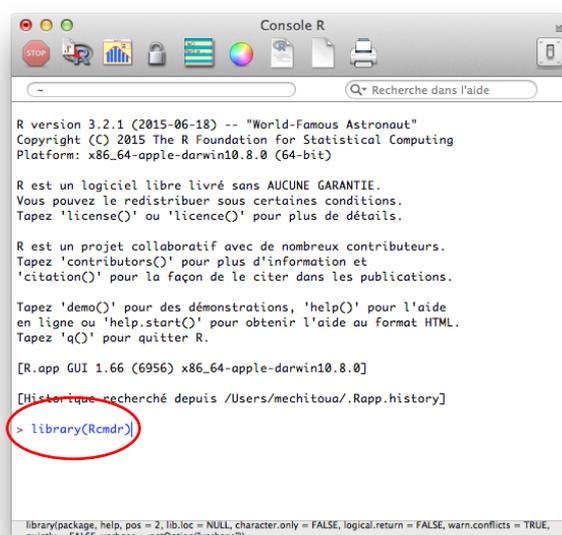


La petite fenêtre suivante indique qu'il faudra sortir puis relancer XQuartz pour en faire le serveur X11.



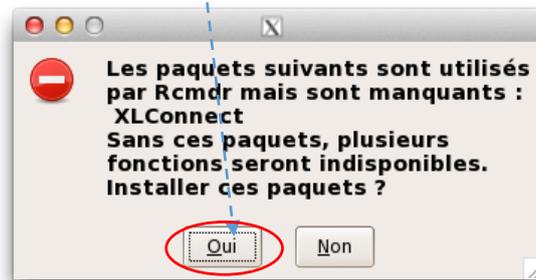
X Quartz est installé. Sur le bureau Mac, chercher et cliquer sur l'icône XQuartz pour relancer cette application. On peut ensuite la fermer.

Cliquer sur R pour ouvrir la fenêtre R puis écrire library(Rcmdr) après >

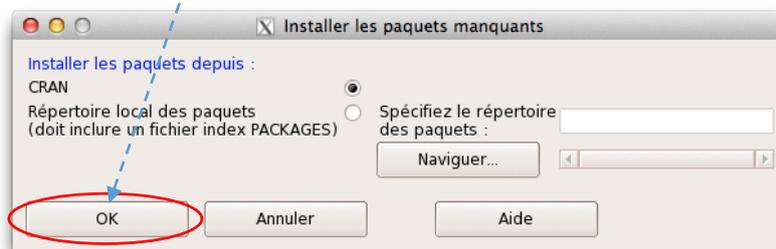


Notons qu'il faudra toujours ouvrir R, puis écrire la commande library(Rcmdr) pour ouvrir R Commander.

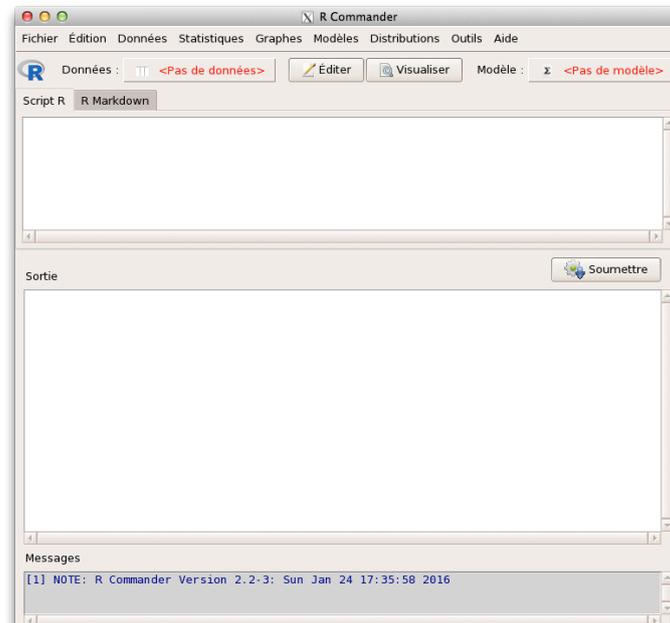
A la première ouverture de R Commander, une fenêtre peut signaler qu'il manque encore des packages qu'il faut télécharger si l'on souhaite avoir toutes les fonctions. Ici par exemple, il manque XLConnect. Cliquer Oui pour l'installer



La fenêtre Installer les package manquants s'ouvre. L'option installer les paquets depuis CRAN est cochée. Cliquer OK



Une fois les packages manquants installés, la fenêtre R Commander s'ouvre.

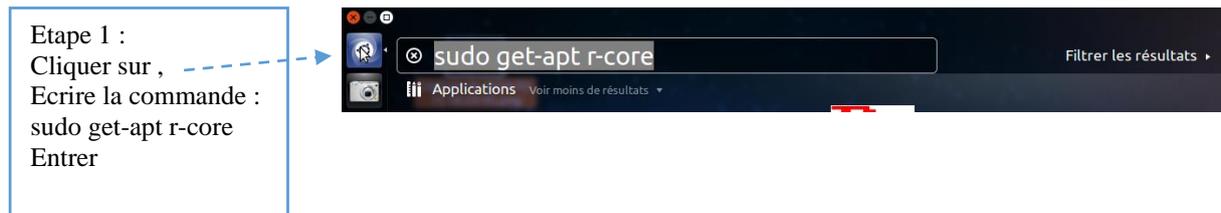


Pour fermer la fenêtre, cliquer sur la croix rouge tout en haut et fermer les différentes fenêtres Sortir sans enregistrer. Quand R Commander est fermer, fermer R de la même façon.

### 15.3 Installer R et R Commander sous Ubuntu (Linux)

R Commander est une interface graphique de R. Il faut donc installer R, puis R Commander. Le téléchargement de R et de R Commander s'effectue comme tout téléchargement, à partir de Ubuntu. Trois possibilités sont présentées ci-dessous : à partir de la ligne de commande, à l'aide d'Administration et enfin par le Gestionnaire de paquets Synaptic.

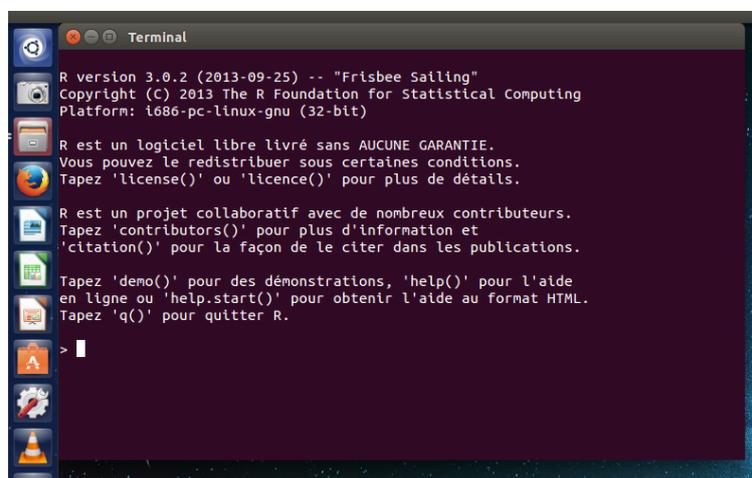
A partir de la ligne de commande.



La fenêtre demande du mot de passe s'ouvre : Ecrire le mot de passe et entrer R se télécharge pendant quelques minutes. Lorsque R est installé, l'icône R apparaît dans la colonne.

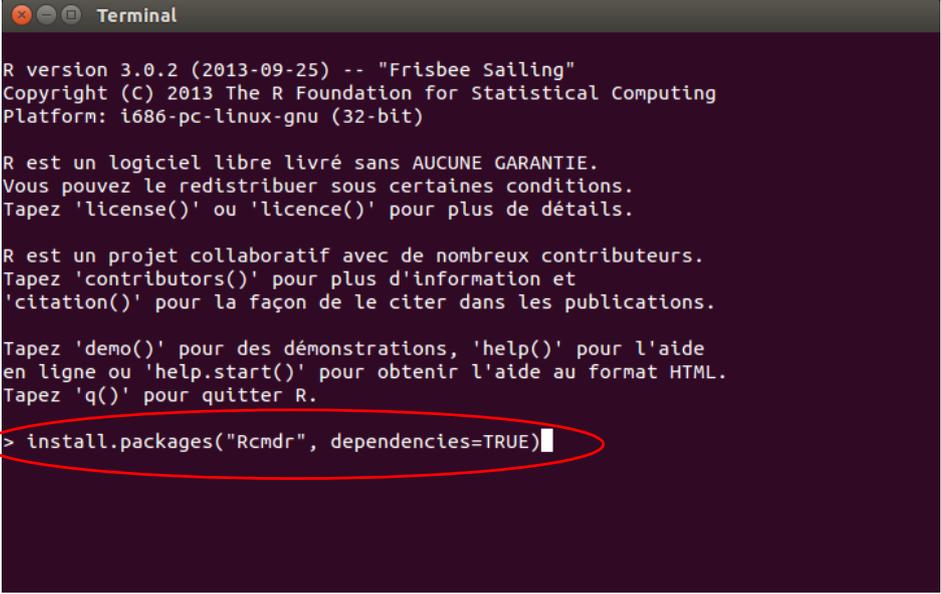


Cliquer sur R pour ouvrir le programme et la fenêtre R s'ouvre.



## Installer R Commander

En face du prompt `>`, entrer la commande : `install.packages(« Rcmdr », dependencies=TRUE)`



```

R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: i686-pc-linux-gnu (32-bit)

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

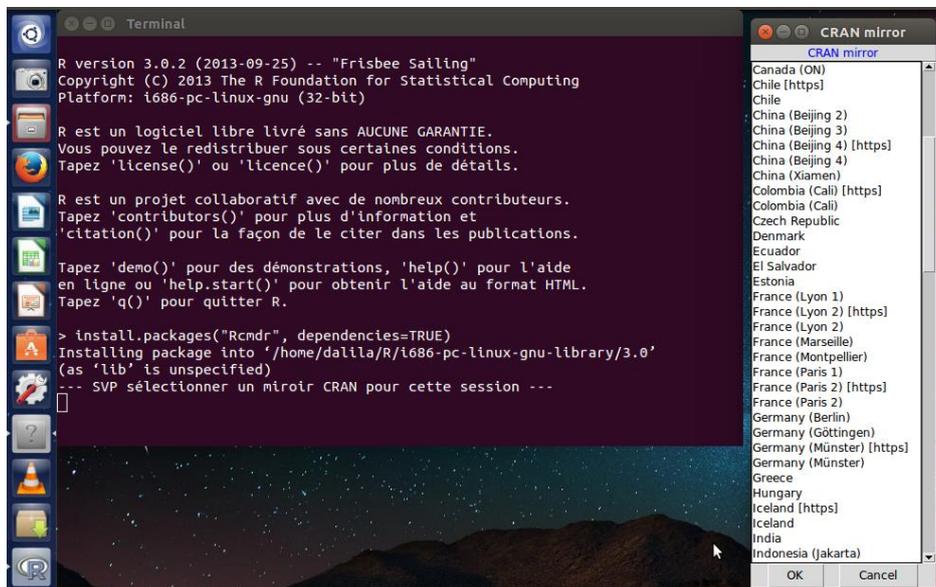
> install.packages("Rcmdr", dependencies=TRUE)

```

Puis entrer.

Le téléchargement commence et répondre ‘y’ (oui) quand les invitations le proposent.

Rapidement un tableau apparaît dans lequel il faut choisir le site miroir à partir duquel sera téléchargé R Commander : Choisir le site en France le plus proche et cliquer OK.

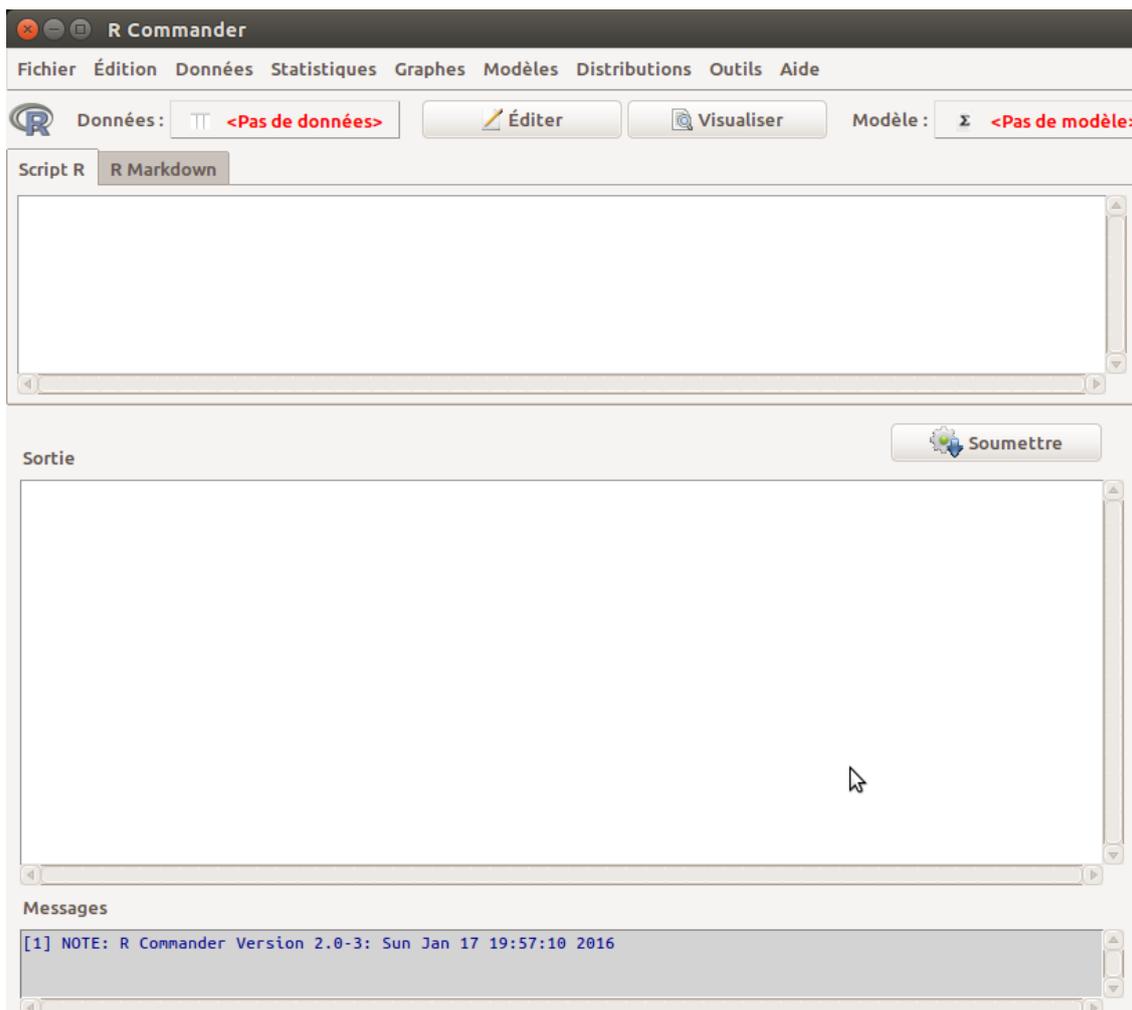


Le téléchargement continue pendant plusieurs minutes. Lorsqu’il est terminé, la fenêtre de R a le contenu suivant, avec le prompt tout en bas.

```
Terminal
* removing '/home/dalila/R/i686-pc-linux-gnu-library/3.0/Rcmdr'

Les packages source téléchargés sont dans
'/tmp/RtmpEyheTg/downloaded_packages'
Messages d'avis :
1: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'reshape2' a eu un statut de sortie non nul
2: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'scales' a eu un statut de sortie non nul
3: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'mi' a eu un statut de sortie non nul
4: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'rgl' a eu un statut de sortie non nul
5: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'ggplot2' a eu un statut de sortie non nul
6: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'Hmisc' a eu un statut de sortie non nul
7: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'RcmdrMisc' a eu un statut de sortie non nul
8: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'sem' a eu un statut de sortie non nul
9: In install.packages("Rcmdr", dependencies = TRUE) :
  l'installation du package 'Rcmdr' a eu un statut de sortie non nul
> library(Rcmdr)
```

Pour ouvrir R Commander, il faut écrire en face du prompt `> : library(Rcmdr)` puis entrer. La fenêtre de R Commander s'ouvre.



A l'aide d'Administration

- Dans l'ordre :
1. Cliquer sur ,
  2. Ouvrir l'Administration,
  3. Ecrire r-base dans la ligne commande
  4. L'icône R apparaît
  5. Cliquer sur l'icône R



La fenêtre R s'ouvre :

- Etape 2 :
- Cliquer sur Téléchargement,

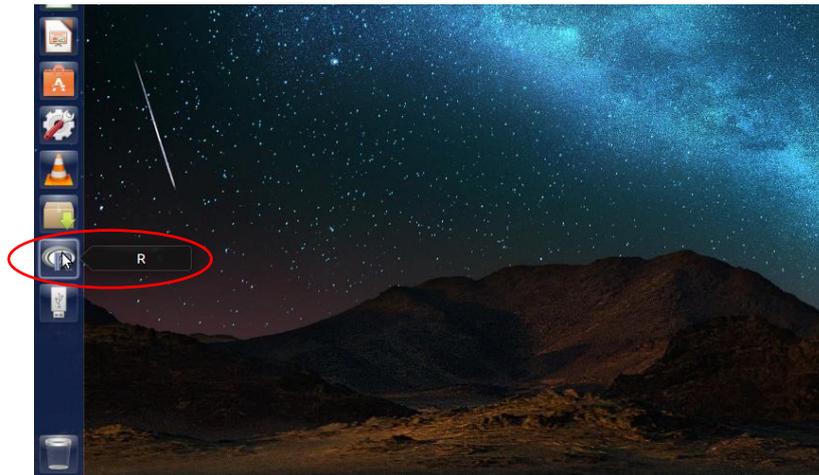


La fenêtre s'identifier apparaît :

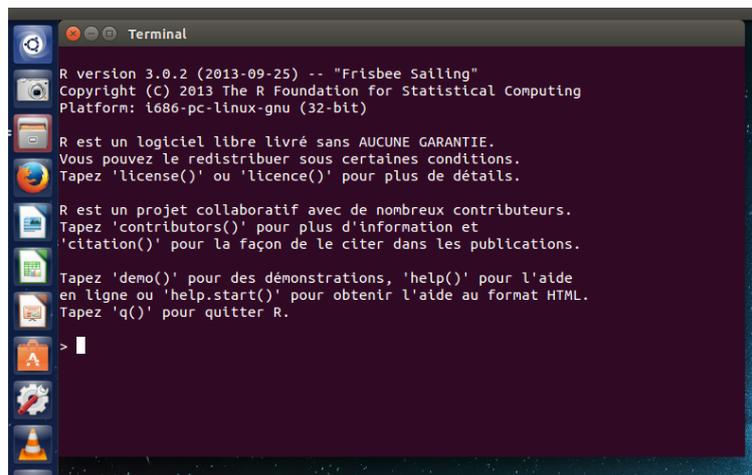


Rentrer le mot de passe et cliquez sur s'authentifier.  
R se télécharge pendant quelques minutes.

Lorsque R est installé, l'icône R apparaît dans la colonne.



Cliquer sur R pour ouvrir le programme et la fenêtre R s'ouvre.

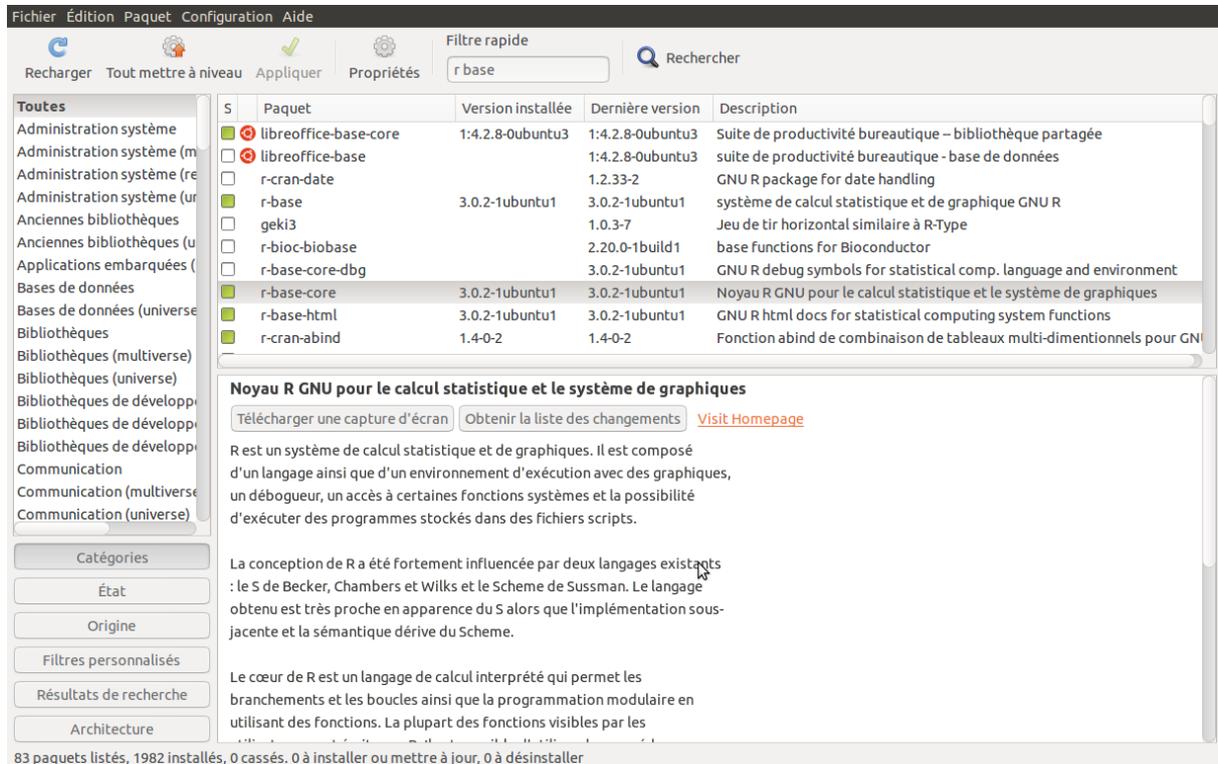


Installer R Commander

Pour installer R Commander, reprendre la même procédure que d'écrite plus haut.

## Téléchargement avec Synaptic

Pour les inconditionnel.le.s du Gestionnaire de paquets Synaptic, il suffit de taper r base et sélectionner r-base-core dans la fenêtre de Synaptic ; puis recommencer avec rcmdr, pour accéder aux 2 logiciels. Le téléchargement se fait avec toutes les dépendances.



Une fois R et R Commander installés, suivre les procédures décrites plus haut pour ouvrir R à partir de l'icône, puis R Commander en entrant la commande `library(Rcmdr)`.