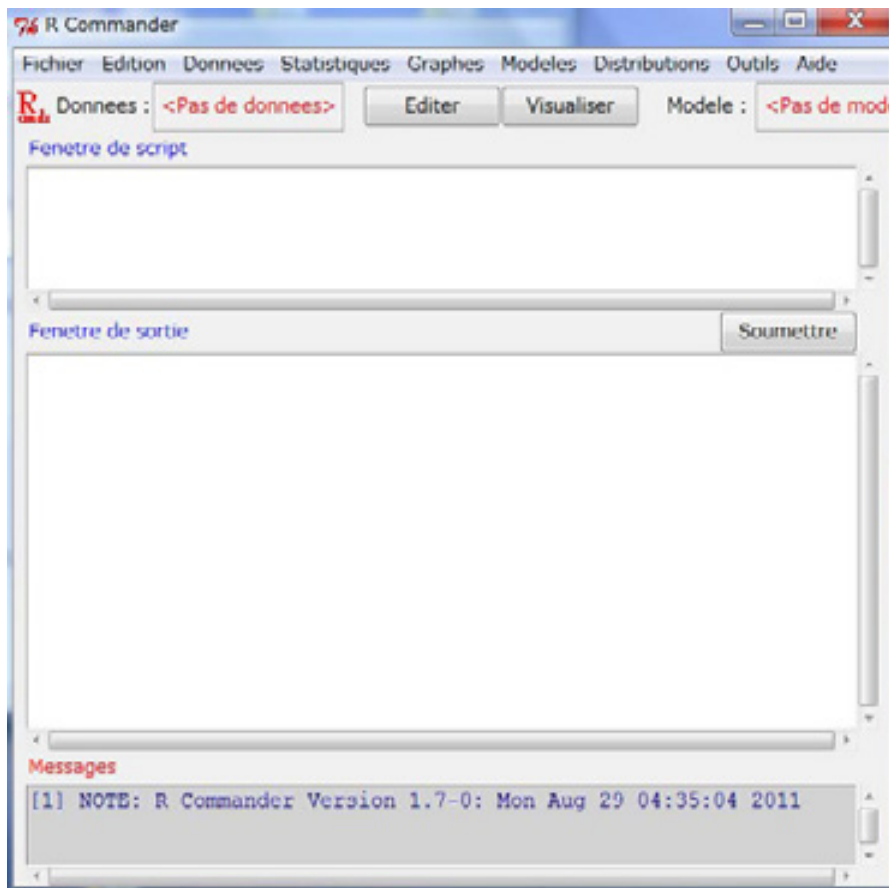


## Introduction du logiciel libre de statistiques : R et R commander



(La version 05/09/2011)

Toshiharu OKAYASU  
Conseiller Principal

Projet de Renforcement du Programme de Prévention du VIH /JICA



## Préambule

Le but du présent document de fournir un point de départ pour des débutants intéressés R et surtout le package R commander.

R est un système d'analyse statistique et graphique crée par Ross Ihaka et Robert Gentleman. R est distribué librement sous les termes de la GNU *General Public Licence*; son développement et sa distribution assurés par plusieurs statisticiens rassemblés dans le R *Development Core Team*. R est disponible pour WINDOWS, Macintosh et UNIX à partir du site internet du *Comprehensive R Archive Network* (CRAN). Cependant R est assez compliqué pour des débutants d'utilisateurs du logiciel statistiques. Pour surmonter ce problèmes, le package R commander (Rcmdr) a été développé par Pr. John FOX de l'université de McMaster au Canada. Ce manuel présente des principales utilisations de R commander en utilisant des données de l'enquête de l'évaluation du Projet de renforcement du Programme de Prévention du VIH /JICA et Programme National du sida(PNLS) du ministère de la Santé Publique à Madagascar. Si vous intéressez ce manuel et des données d'exercice et des suggestions, vous adressez au bureau du Projet.

Je vous souhaite un renforcement de capacité de l'analyse en statistiques de PNLS et des intéressés de ce manuel.

Toshiharu OKAYASU

## Tables des matières

Menu de statistiques	R Commander (version 1.7.0)	4
Installation de R		5
Installation du package R commander (Rmdr) et démarrage		8
Données dans les packages		9
Importation des données		11
Convertir des variables numériques en facteurs		14
Recoder des variables		15
Découper une variable numérique en classe		17
Les tableaux croisés à la main		18
Les tableaux croisés en double entrée		19
Tables de contingences avec tableau à plusieurs entrées		20
Fusionner des jeux de données		21
Nom des cas		23
Exporter le jeu de données actif		26
Distribution de fréquence		27
Utilisation de Fenêtre de Script		28
Eliminer les cas contenant des valeurs manquants		29
Graphe en camembert		31
Graphe en Nuage de points		32
Graphe en 3D		33
Test de corrélation		34
Test de normalité de Shapiro Wilk		36
Test F de deux variances		38
Test de Welch		39
Test de Wilcoxon bivarié (Test non paramétrique pour comparer deux moyennes (Médians))		40
R commander Personnalisé		43
ANNEXE : Test statistique de base et R commander		45
Type de données		46
Schéma de test statistique 1		47
Schéma de test statistique 2		48
Distribution T		49
Distribution Chi-carrée		50

## Menu de statistiques

R Commander (version 1.7.0)

### Statistiques –Résumé -Jeu de donnée actif

- | -distribution de fréquence
- | -Dénombrer les observations manquantes
- | -Tableau de statistiques
- | -Matrice de corrélation
- | -Test de corrélation
- | -Test de normalité de Shapiro-Wilk

### Table de contingences- Tableau de double entrée

- | -Tableau de plusieurs entrée
- | -Remplir et analyser un tableau à double entrée

### Moyennes -T test uni varié

- | -T test indépendant
- | -T test apparié
- | -ANOVA à un facteur
- | -ANOVA à plusieurs facteurs

### Proportions –Test de proportion uni-varié

- | -Test de proportions bi varié

### Variances-Test F de deux variences

- | -Test de Bartlett
- | -Test de Levene

### Tests non paramétriques- Test Wilcoxon bi varié

- | -Test Wilcoxon apparié
- | -Test de Kruskal-wallis
- | -Test de somme de rangs de Fridman

### Analyse multi varié –Fiabilité d'échelle

- | -Analyse en composantes principales
- | -Analyse factorielle
- | -Classification-Classification pour K-means
- | -Classification hiérarchique
- | -Résumé Classification hiérarchique
- | -Ajouter les groupes de la classification en jeu de données

### Ajustement de modèle-Régression linéaire

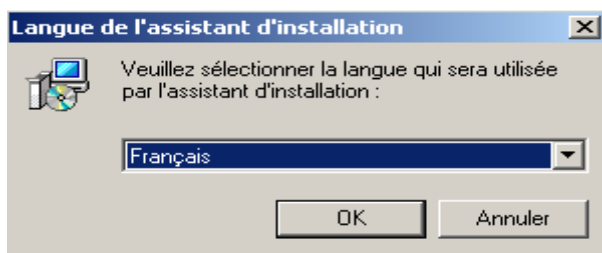
- | -Modèle linéaire
- | -Modèle linéaire généralisé
- | -Modèle logit multinominal
- | -Modèle de régression ordinaire

## Installation de R

Téléchargez R2-13.1-win.exe

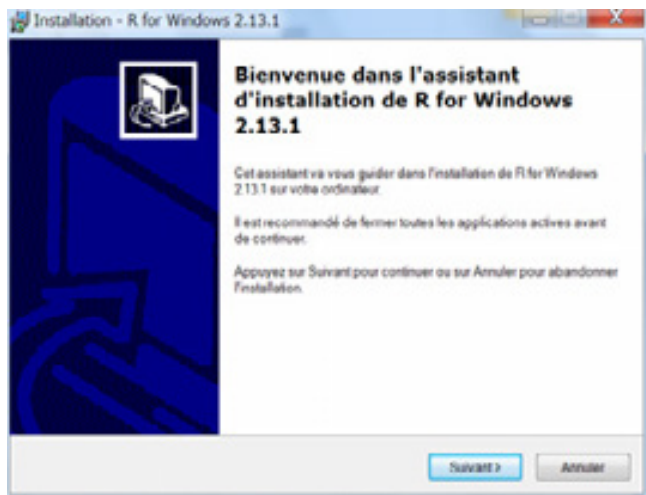
([www.r-project.org/](http://www.r-project.org/))

Double-cliquez sur R-2.13.1-win.exe

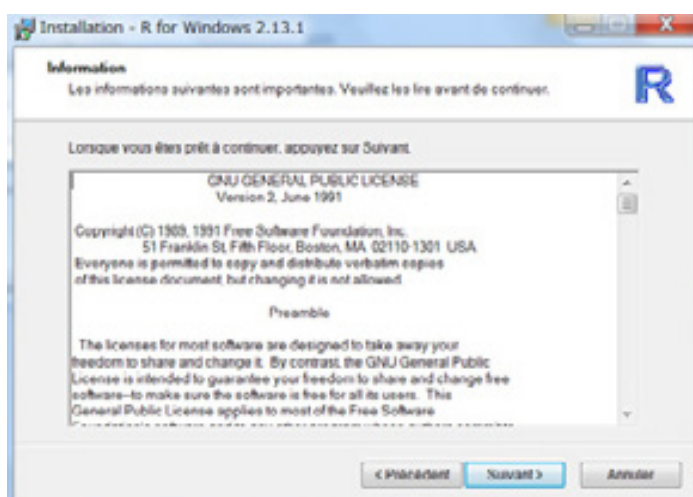


← Sélectionnez la langue d'installation.

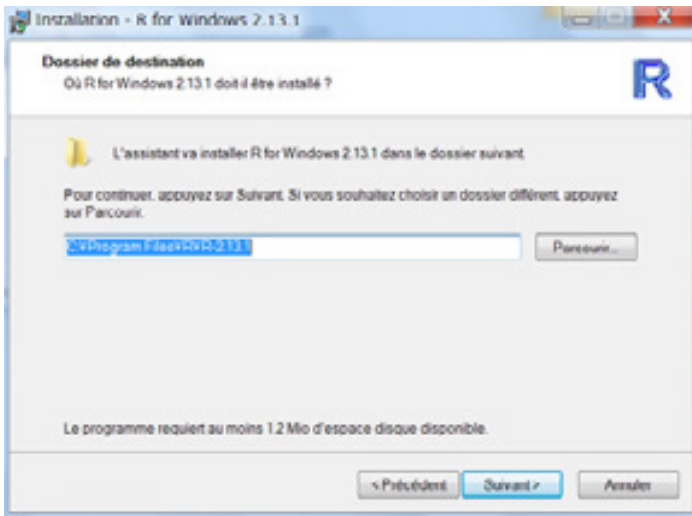
Cliquez sur OK.



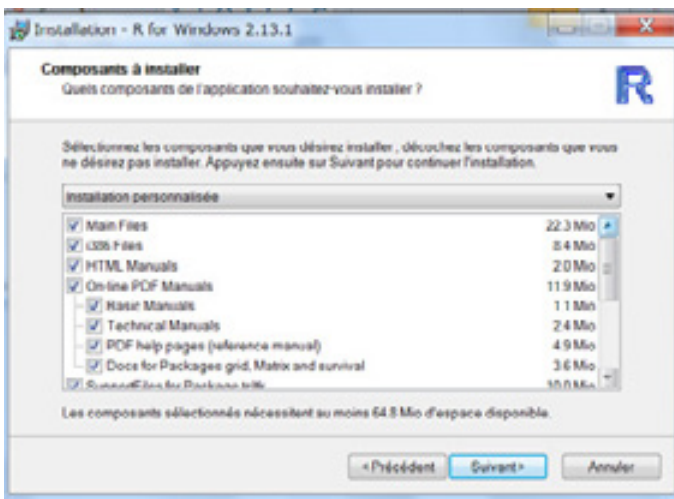
Cliquez sur Suivant.



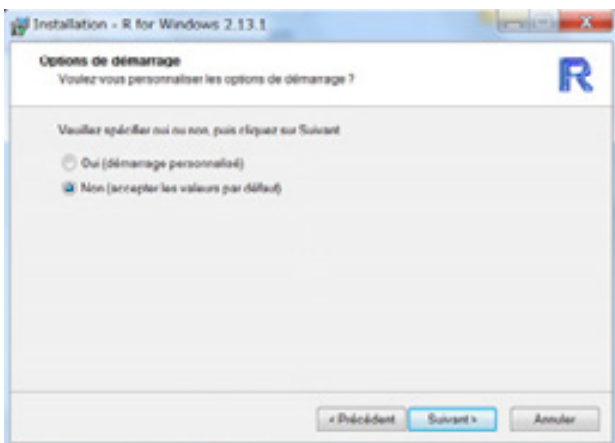
Cliquez sur Suivant.



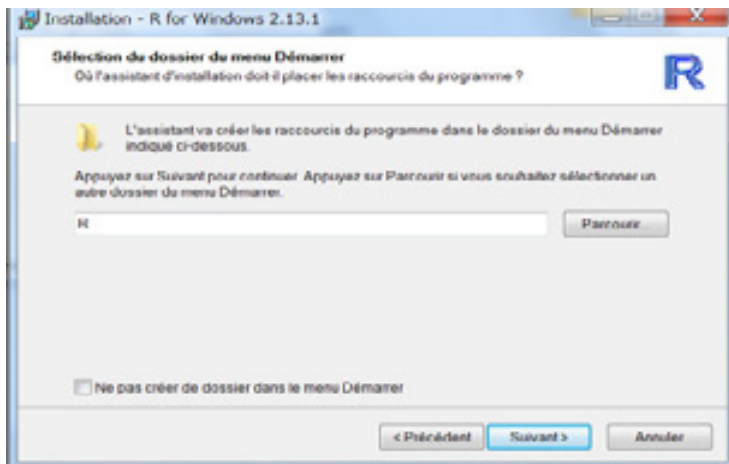
Cliquez sur Suivant.



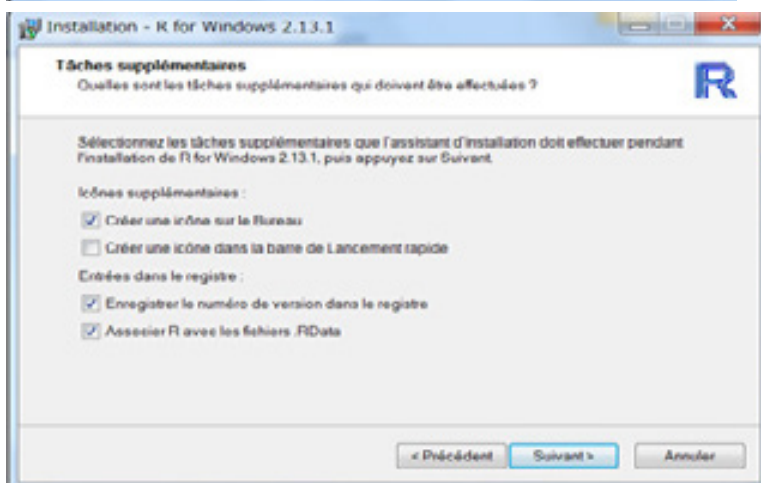
Sélectionnez toutes les options et Cliquez sur Suivant.



Sélectionnez sur Non et cliquez sur Suivant.



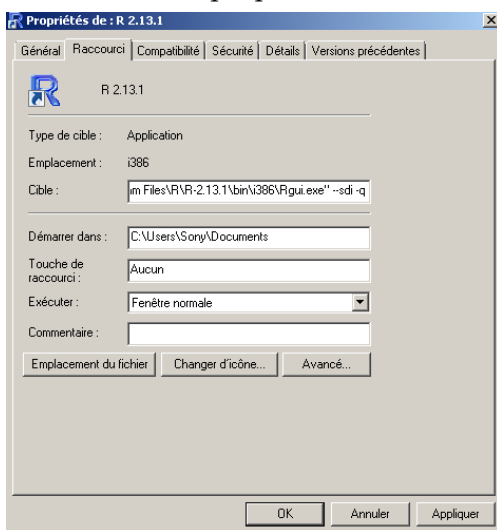
Cliquez sur Suivant.



Cliquez sur Suivant et attendez la fin de l'installation.

Note : Quand R se démarre, il peut vous demander les packages manquants. Vous allez télécharger ou vous allez copier et coller des packages manquants dans le fichier de library.

Quand l'icône de R est sur votre bureau de l'ordinateur, cliquez droite et sélectionnez la propriété.



Ajoutez --sdi -q après Rgui.exe" dans le cible

N.B --sdi signifie l'utilisateur simple. -q signifie l'enlèvement de message de démarrage.

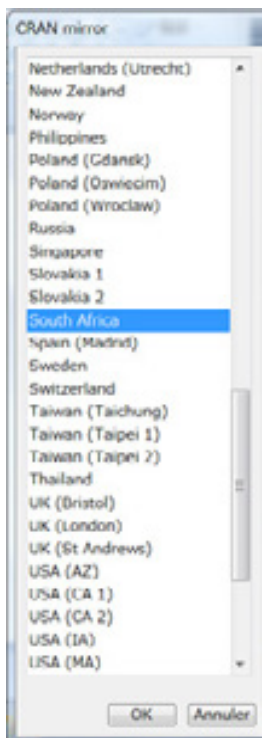
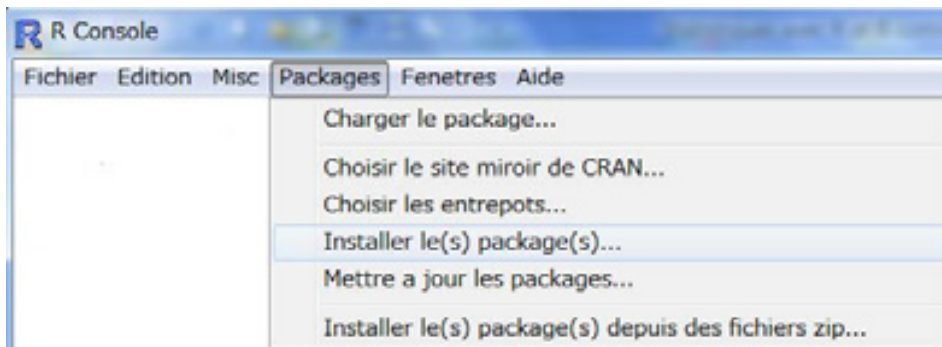
## Installation du package R commander (Rcmdr) et démarrage

Double cliquez sur l'icône de R sur votre bureau de l'ordinateur.

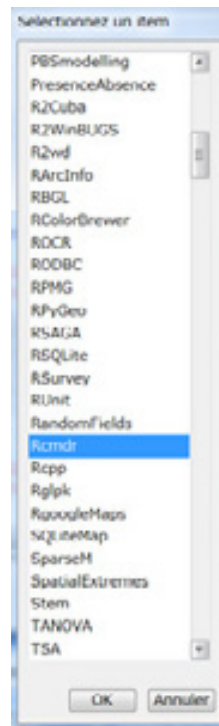


Si la connexion internet est disponible, vous pouvez télécharger un package de Rcommander.

**Packages > Installer le(s) package(s)**

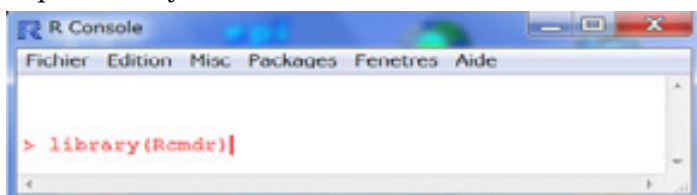


Sélectionnez le pays plus proche, en cas de Madagascar, Afrique de Sud est recommandé.



Sélectionnez **Rcmdr** et Cliquez sur OK.

Tapez library (Rcmdr) dans R Console.





R Commander sera apparu.



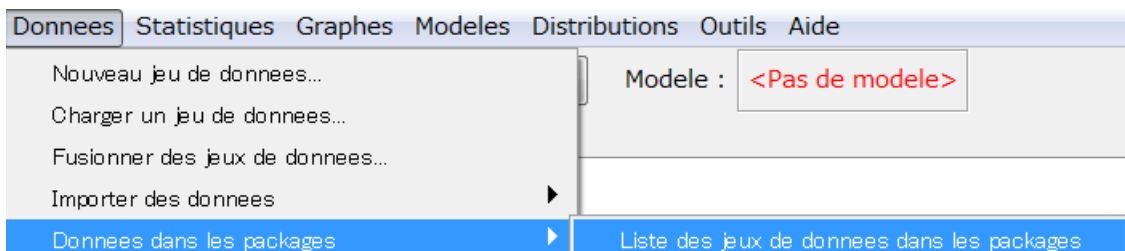
Quand vous mettez le curseur au coin de la fenetre R Commander, vous changez la taille de fenetre.

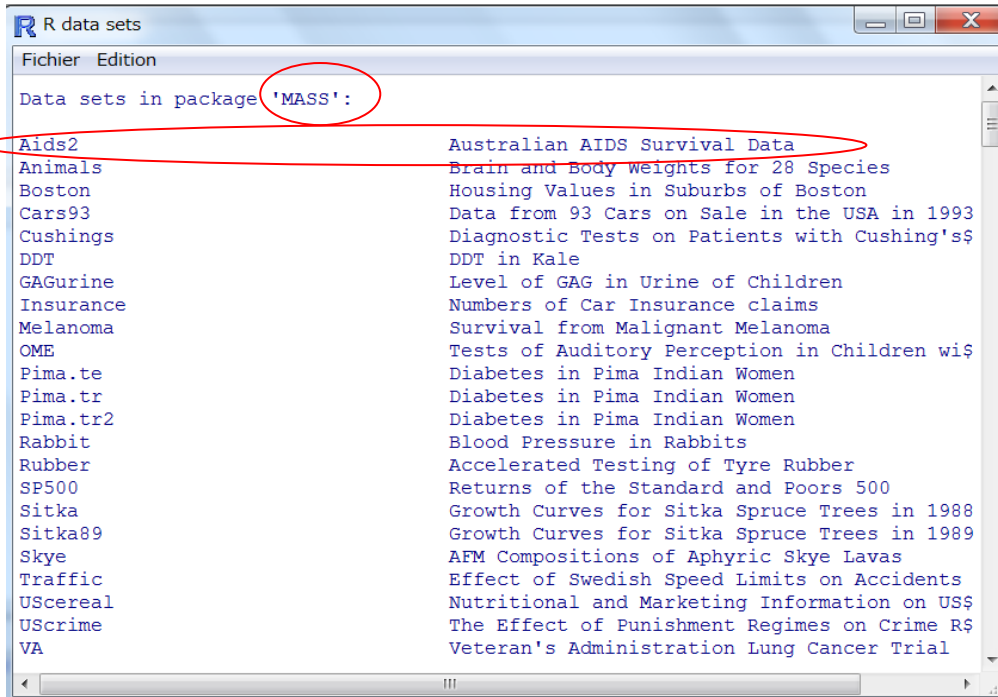
Messages seront affichés avec chaque l'opération.

### Données dans les packages

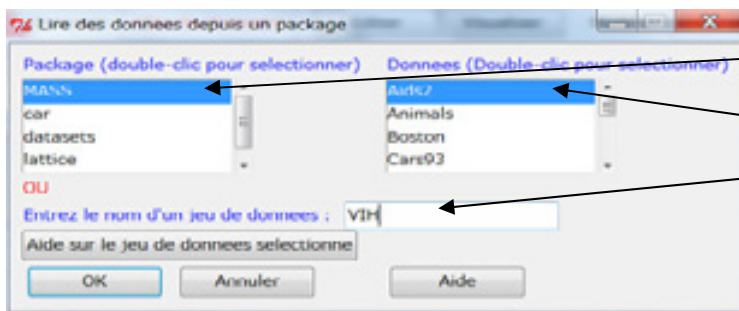
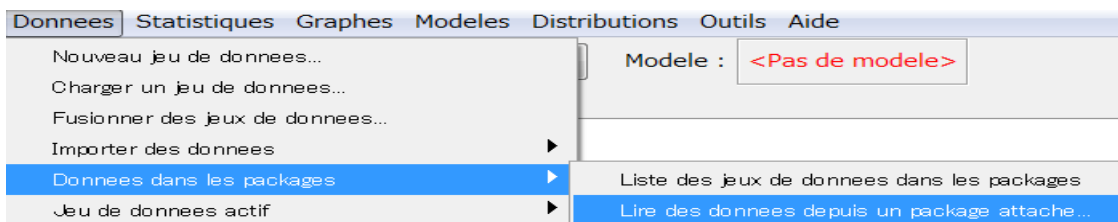
Quand vous installez R et Rcmdr, certaines données de packages aussi ont été installées en même temps. Pour voir et utiliser ces données, cliquez sur ;

**Données > Données dans les packages > Liste des jeux des données dans les packages**





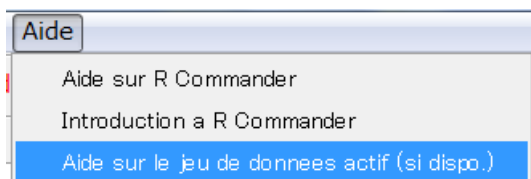
Données > Données dans les packages > Lire des jeux des données dans les packages



Sélectionnez le package et données.  
Mettez le nom de jeu de données

Pour voir le contenu de données dans les packages, cliquez sur ;

Aide > Aide sur le jeu de données actif (si dispo.)



Aids2 [MASS]

## Australian AIDS Survival Data

### Description

Data on patients diagnosed with AIDS in Australia before 1 July 1991.

### Usage

Aids2

### Format

This data frame contains 2843 rows and the following columns:

state

Grouped state of origin: "NSW" includes ACT and "other" is WA, SA, NT and TAS.

sex

Sex of patient.

diag

(Julian) date of diagnosis.

death

(Julian) date of death or end of observation.

status

"A" (alive) or "D" (dead) at end of observation.

T.categ

Reported transmission category.

age

Age (years) at diagnosis.

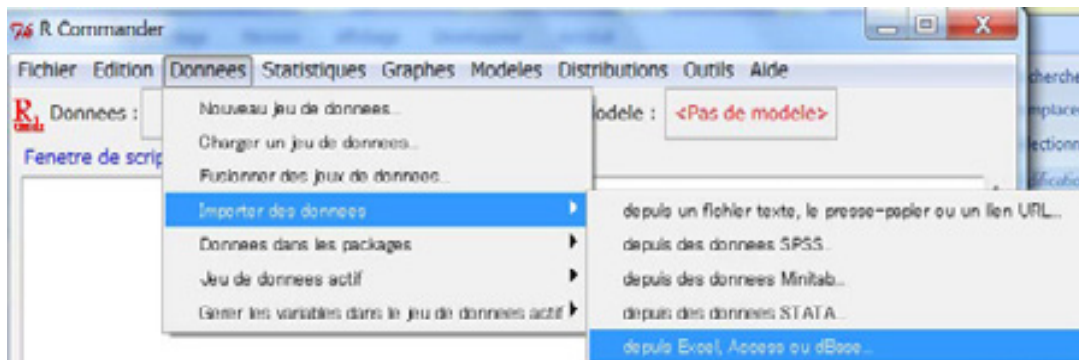
### Note

This data set has been slightly jittered as a condition of its release, to ensure patient confidentiality.

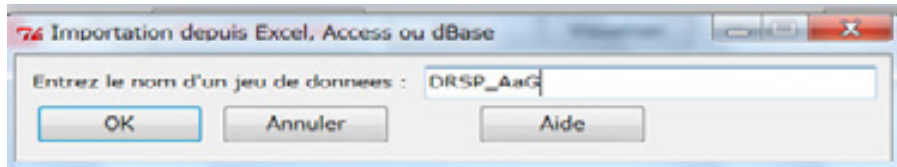
## Importation des données

R Commander peut lire des données un fichier text(.csv), le presse-papier, SPSS(.sav), données minitab, Excel , Access, dBase.

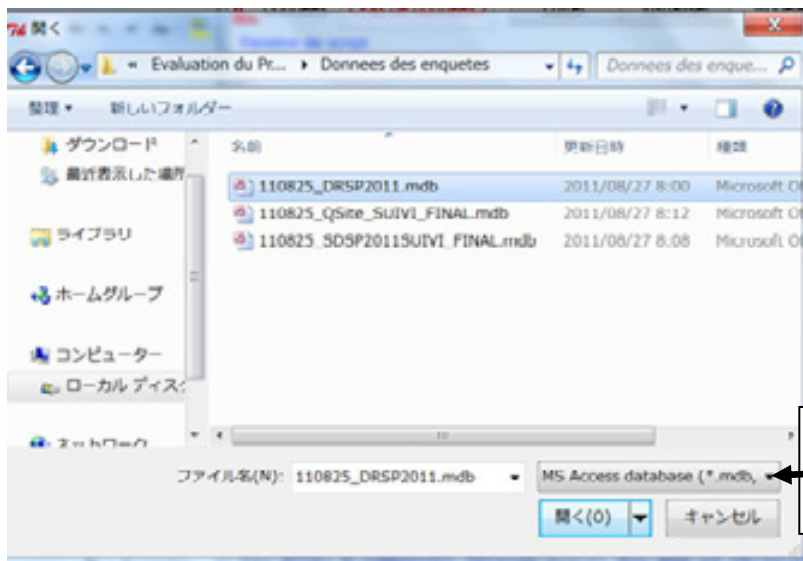
**Donnez >Importation des données >depuis Excel, Access ou dBase**



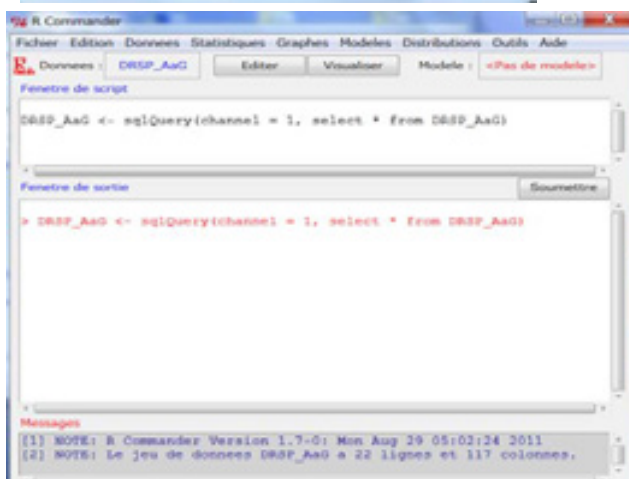
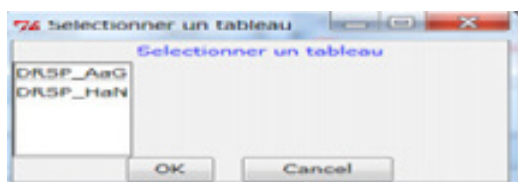
Mettez le nom d'un jeu de données : Par exemple < DRSP\_AaG > et cliquez sur OK.



Cherchez vos données dans votre ordinateur.



Si vos données d'Access qui a plusieurs tableaux, vous devez sélectionner un tableau et cliquez sur OK.



Dans la case de Messages, NOTE va expliquer nombre de lignes (nombre des échantillons) et nombre de colonnes (nombre des variables)

Si vous voulez voir le jeu des données, cliquez sur **Visualiser**.

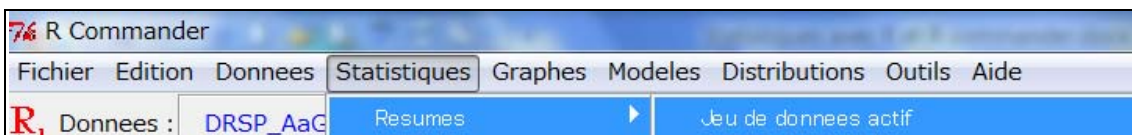
ID	CODE	TYPEETUDE	DATE	DSRF
1	TNR_DRS_1	2008	2011-07-29	ANALAMANGA
2	TMM_DRS_2	2008	2011-08-10	ANALANJIROFO
3	TLE_DRS_2	2008	2011-08-10	ANOSY
4	TLE_DRS_3	2008	2011-07-27	ATIMO ANDREFANA
5	WFI_DRS_3	2008	2011-07-26	HAUTE MATSIATRA
6	DIE_DRS_1	2008	2011-07-19	DIANA
7	TLE_DRS_4	2008	2011-07-15	MENABE
8	MJN_DRS_3	2011	2011-08-02	MELAKY
9	MJN_DRS_2	2008	2011-07-25	BOENY
10	TNR_DRS_3	2008	2011-07-18	ITASY
11	MJN_DRS_4	2008	2011-07-29	SOFIA
12	WFI_DRS_2	2008	2011-07-28	ATIMO ATSIANANA
13	MJN_DRS_1	2008	2011-07-29	BETSIBOFA
14	TMM_DRS_3	2008	2011-07-28	ATSIANANA
15	WFI_DRS_1	2008	2011-07-13	AMORON'I MANIA
16	TMM_DRS_1	2008	2011-07-25	ALACTRA MANGORO
17	TLE_DRS_1	2008	2011-08-10	ANDROY
18	TNR_DRS_2	2008	2011-07-21	BONGOLAVA
19	WFI_DRS_4	2008	2011-07-28	IHOROMBE
20	DIE_DRS_2	2008	2011-07-26	SAVA

Quand vous cliquez le triangle en bas droite, vous allez voir d'autres variables.  
Note : R va lire des données non entrée automatiquement comme NA ou case vide.

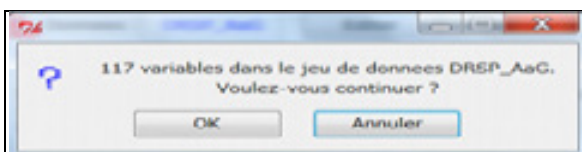
### Résumés des données du jeu des données

Pour voir des données entiers, cliquez sur ;

**Statistiques >Résumés >Jeu de données actif**



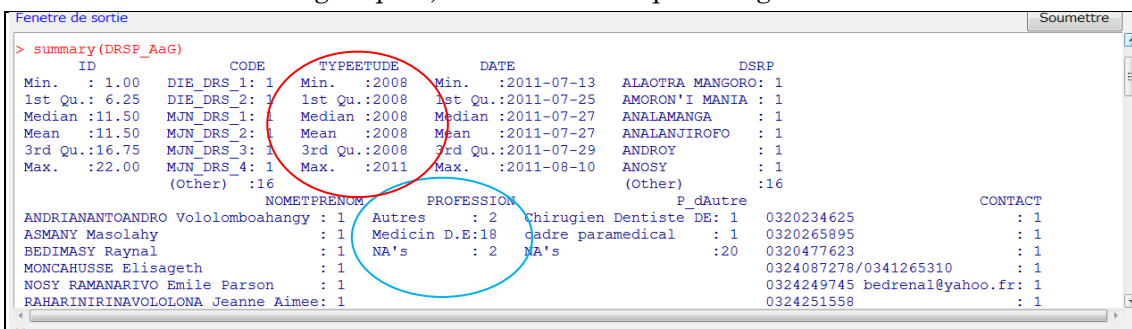
R vous demande la continuité, cliquez sur OK.



Dans la fenêtre de sortie, un résumé sera affiché.

-Pour des données numériques/quantitatives, Min (Minimum), 1st Qu. (Premier Quantile), Median (Médian), Mean (Moyenne), 3rd Qu. (Troisième Quantile), Max (Maximum) et NA's (Non Applicable) seront apparus.

-Pour des données catégoriques, nombre de chaque catégorie seront affichés.



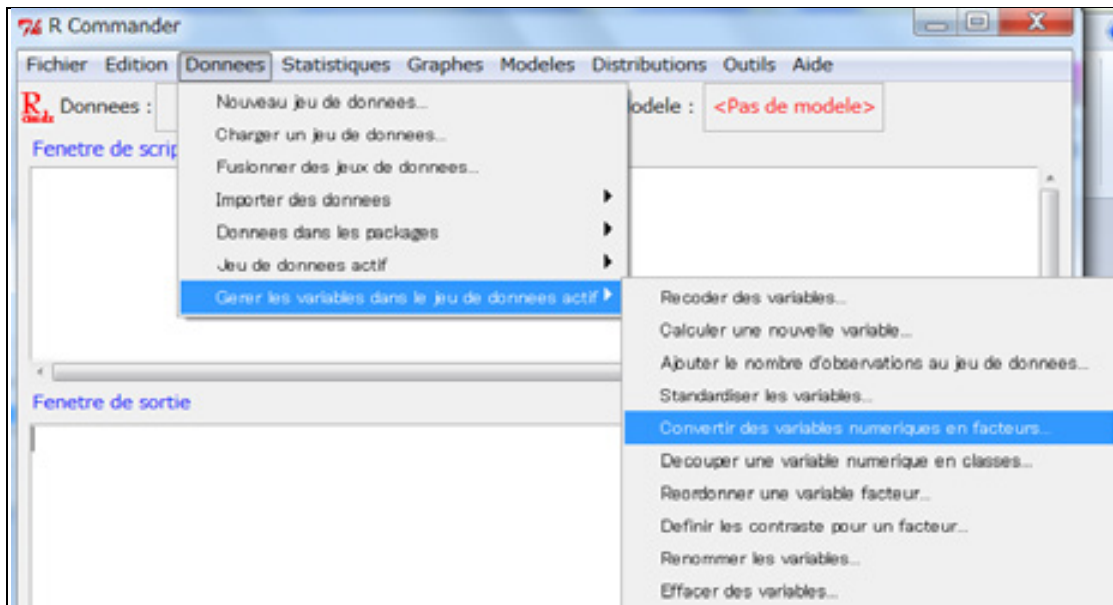
Note : Certaines données sont traitées comme numériques au lieu de données catégoriques, il faut changer le type de données comme Facteurs (données catégoriques). Par exemple, le type de l'étude, nous voudrions savoir combien pour 2008 et combien pour 2011 mais pas Médian, ni Moyenne. Il faut les changer pour les données en facteurs.

### Convertir des variables numériques en facteurs

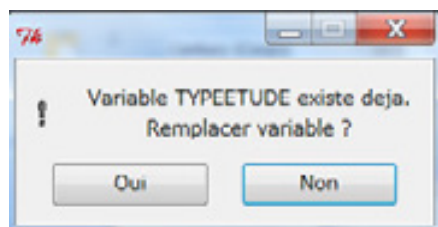
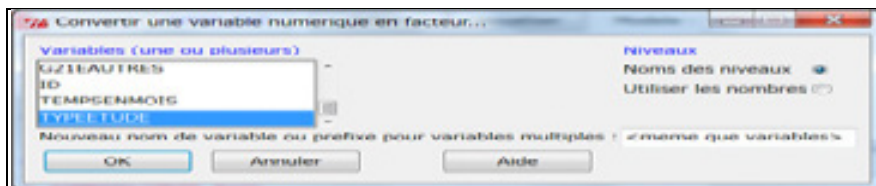
R a deux types de variables ; variables numériques et facteurs.

Note : pour faire le tableau croisé, il faut des variables en facteurs pour des données traitées.

Données > Gérer les variables dans le jeu de données actif > Convertir des variables numériques en facteurs

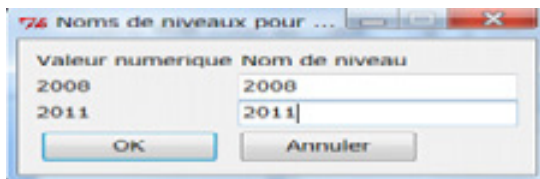


Pour la case de variables, sélectionnez sur TYPEETUDE,



Cliquez sur OK

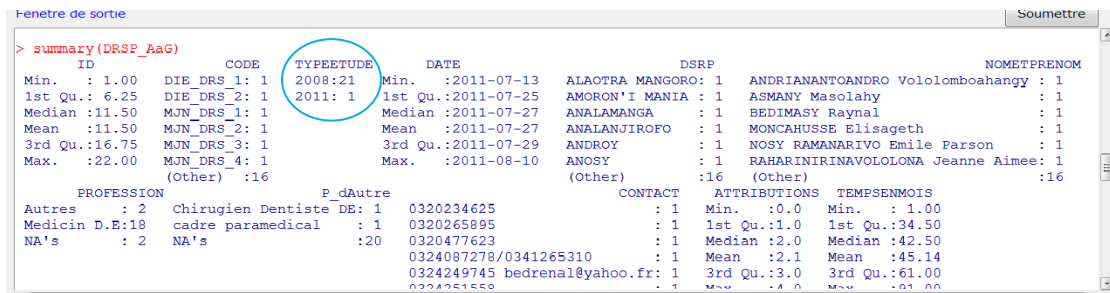




Ecrivez les noms de niveau. Et cliquez sur OK.

Note : ici, nous avons écrit les nombres, mais ils vont les traiter comme facteur (catégorie)

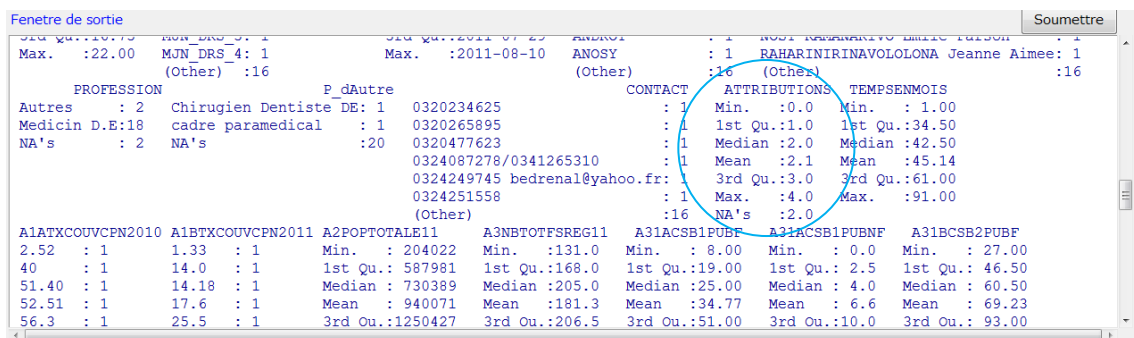
Vous vérifiez le changement avec la fonction de **Résumés des données du jeu des données**



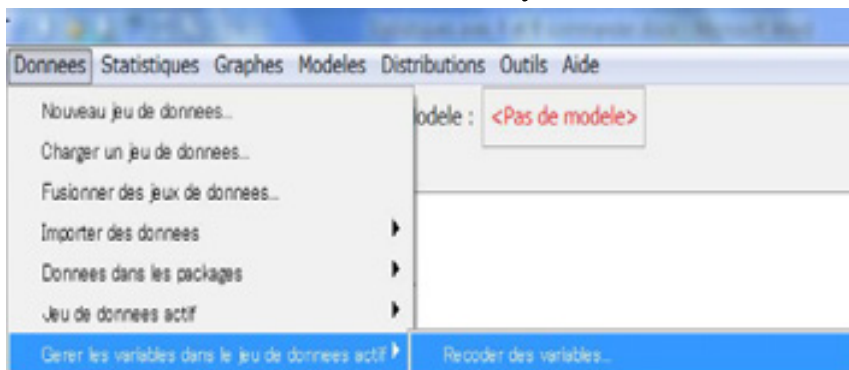
Exercices : vous allez changer le type de données pour des d'autre données.

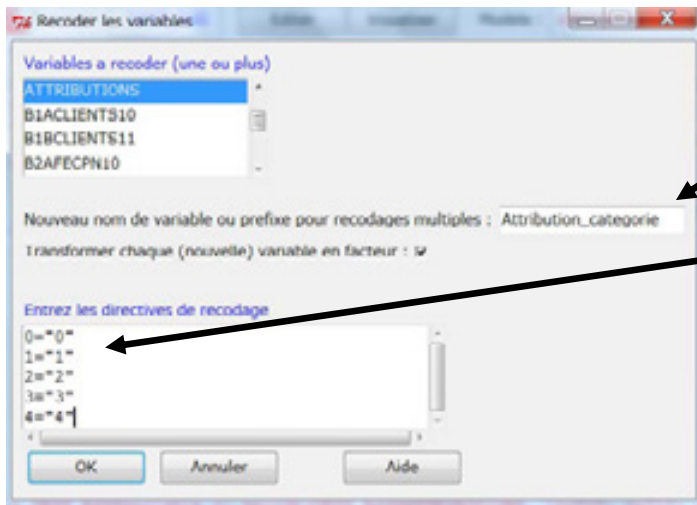
### Recoder des variables

Quand nous voudrions compter le nombre par catégorie avec des données numérique, il faut recoder des variables. Par exemple, on a besoin le nombre des attributions par catégorie au lieu de Moyenne et Médian.



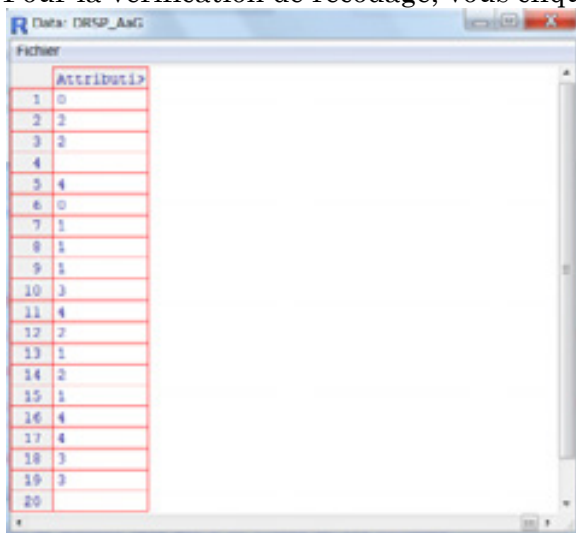
Données > Gérer les variables dans le jeu de données actif > Recoder des variables





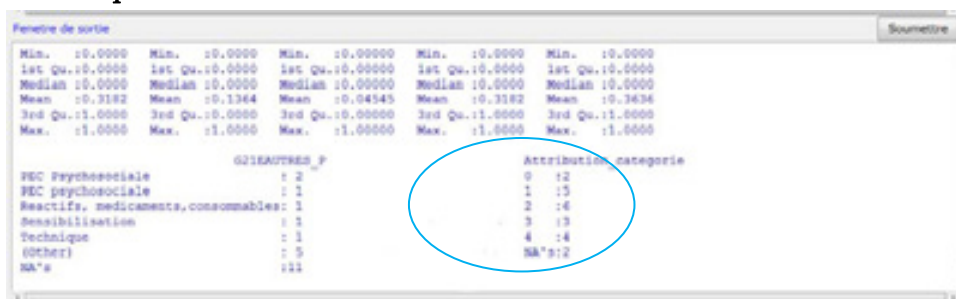
Sélectionnez une ou des variables. Ex.ATTRIBUTIONS  
 Et mettez le nouveau nom. Ex.Attribution\_categorie.  
 Entrez les directives de recodage. Pour des données en facteurs, il faut ajouter " ".  
 Et cliquez sur OK.

Pour la vérification de recodage, vous cliquez sur Visualiser.



Données recodées seront apparues à la fin de votre tableau actif.

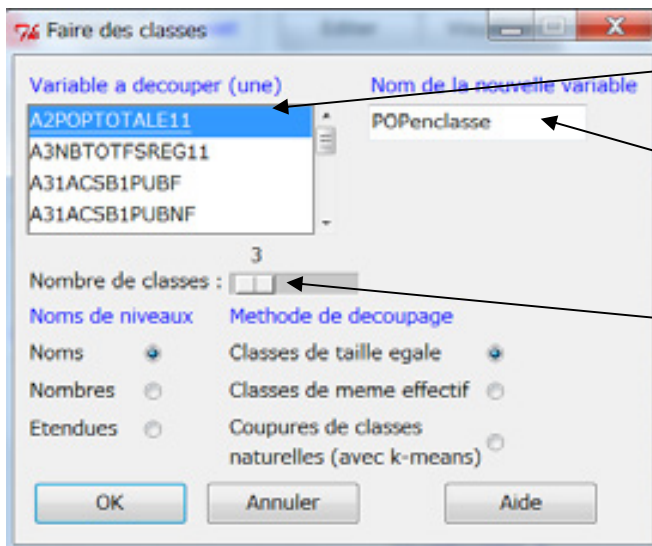
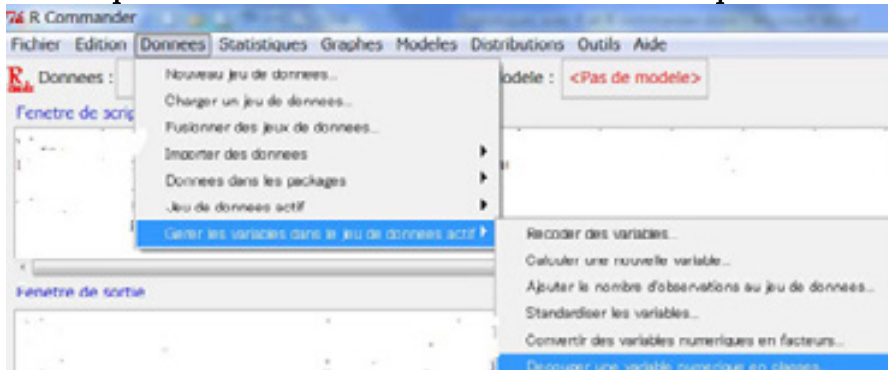
Pour la vérification de changement, cliquez sur ;  
**Statistiques >Résumés >Jeu de données actif**





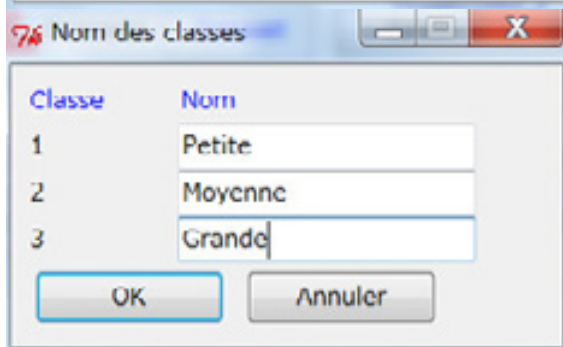
## Découper une variable numérique en classe

Statistiques > Donnée > Décoder une variable numérique en classe



Sélectionnez Variables à découper. Ex. A2 POPTOTALE. Et mettez le nom de la nouvelle variable. Ex. POP11enClasse. Fixez le nombre de classe. Ex.3. Cliquez sur OK.

Si vous glissez ce bouton, vous pouvez changer le nombre de classes.



Mettez des noms des classes, Cliquez sur OK.

Pour la vérification, cliquez sur Visualiser.

R Data: Dataset

Fichier

	POPenclas>
1	Grande
2	Moyenne
3	Petite
4	Moyenne
5	
6	Petite
7	Petite
8	Petite
9	Petite
10	
11	Petite
12	Petite
13	Petite
14	Moyenne
15	Petite
16	Petite
17	Petite
18	Petite
19	Petite
20	Petite

Données en classes seront apparues à la fin de tableau.

### Les tableaux croisés à la main

Statistiques > Tables de contingence > Remplir et analyser un tableau à double entrée

Remplissez la table à double entrees

Nombre de lignes :

Nombre de colonnes :

Entrez les denombrements :

	1	2
1	<input type="text"/>	<input type="text"/>
2	<input type="text"/>	<input type="text"/>

Calculer les pourcentages

Pourcentages des lignes

Pourcentage des colonnes

Pourcentages du total

Pas de pourcentages

Test d'hypotheses

Test Chi-deux d'indépendance

Composants de la statistique Chi-deux

Imprimer les frequences attendues

Test exact de Fisher

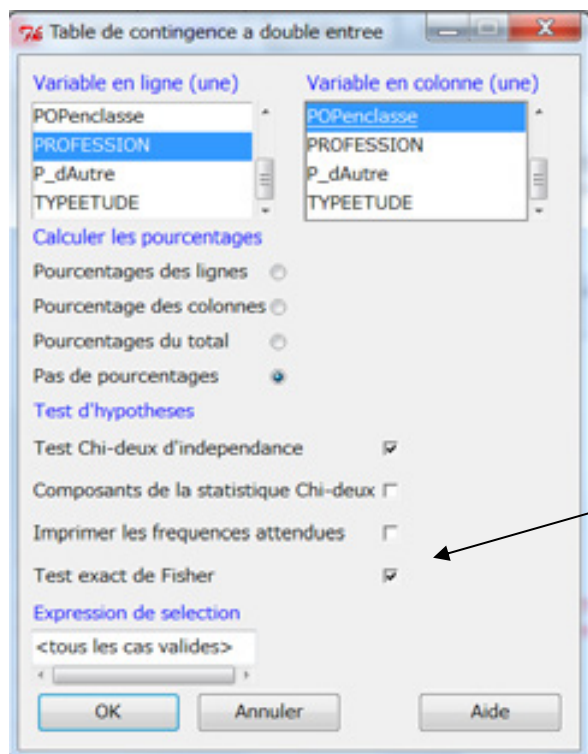
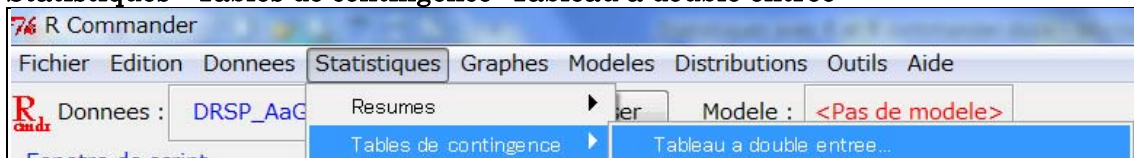
OK Annuler Aide

Vous pouvez changer nombre de lignes et de colonnes en glissant les buttons.

Mettez des noms de variables dans les case de 1, 2 et des chiffres dans les cases vides.  
Note : cette fonction est comme STATCALC d'Epi info.

## Les tableaux croisés en double entrée

### Statistiques > Tables de contingence > Tableau à double entrée



Sélectionnez une variable en ligne. Ex. PROFESSION.  
Sélectionnez une variable en colonne. Ex. POP11en classe.  
Cliquez sur OK.

Une case de Fréquence Escompté sera moins de 5, vous couchez sur Test exact de Fisher.

```
> .table
      POPenclasse
PROFESSION  Petite Moyenne Grande
Autres      2         0         0
Medicin D.E 11         5         1

> .Test <- chisq.test(.Table, correct=FALSE)
> .Test

      Pearson's Chi-squared test

data:  .Table
X-squared = 1.0317, df = 2, p-value = 0.597
```

```
> fisher.test(.Table)

      Fisher's Exact Test for Count Data

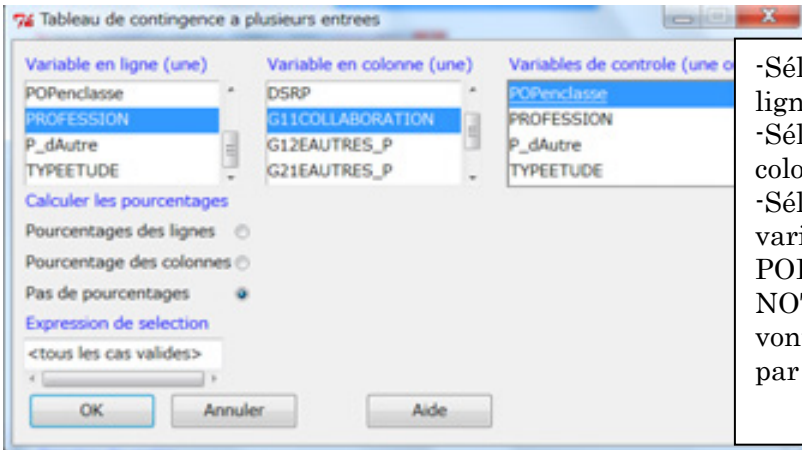
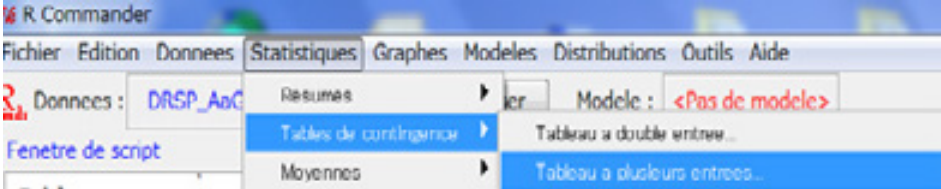
data:  .Table
p-value = 1
alternative hypothesis: two.sided

> remove(.Table)
```

Messages

```
[6] AVIS: 3 les frequences attendues sont inferieures a 1
5 les frequences attendues sont inferieures a 5
```

**Tables de contingences avec tableau à plusieurs entrées**  
**Statistiques > Tables de contingence > Tableau à plusieurs entrées**



-Sélectionnez une variable en ligne. Ex. PROFFESION.  
 -Sélectionnez une variable en colonne Ex. G11Collaboration.  
 -Sélectionnez une ou plusieurs variables de contrôles. Ex. POPenclasse.  
 NOTE : variables de contrôle vont stratifier le tableau 2x2 par chaque variable en facteur.

```
Fenetre de sortie

> .Table
, , POPenclasse = Petite

      G11COLLABORATION
PROFESSION  NON OUI
Autres      0   2
Medicin D.E 2   9

, , POPenclasse = Moyenne

      G11COLLABORATION
PROFESSION  NON OUI
Autres      0   0
Medicin D.E 0   4

, , POPenclasse = Grande
```

```

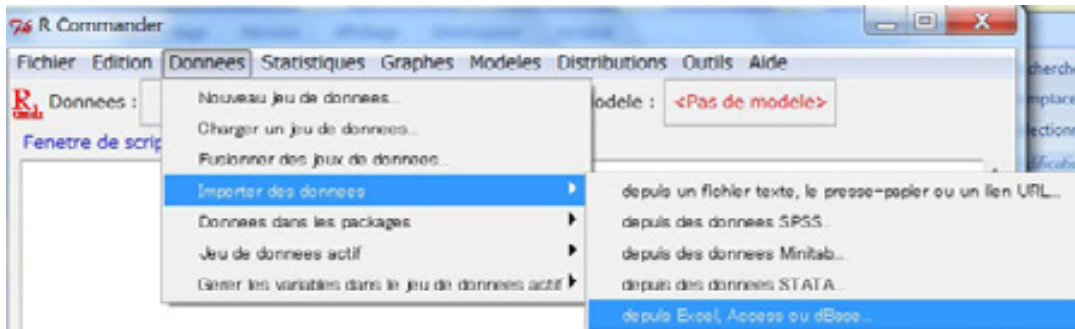
G11COLLABORATION
PROFESSION      NON OUI
Autres          0  0
Medicin D.E    0  1

```

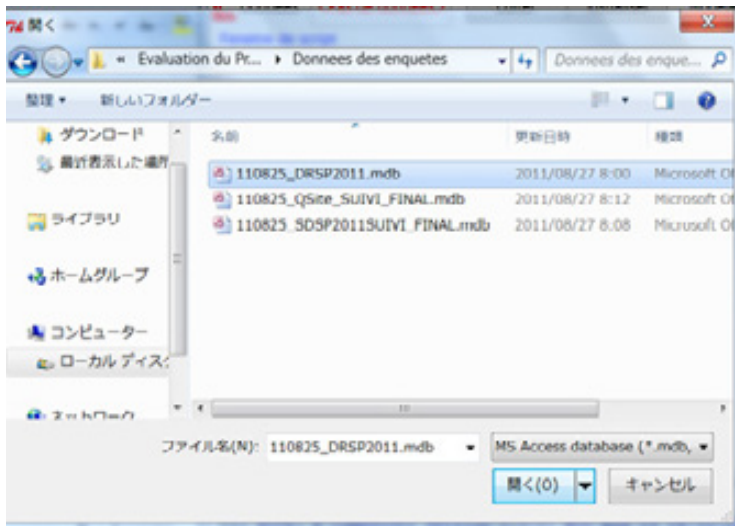
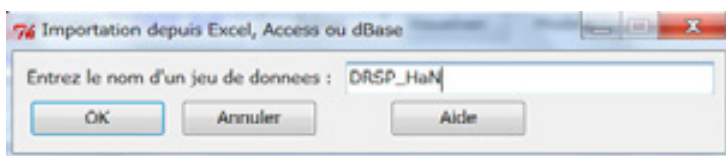
### Fusionner des jeux de données

Tous d'abord, il faut insérer le jeu de données que vous voulez ajouter.

**Donnez >Importation des données >depuis Excel, Access ou dBase**

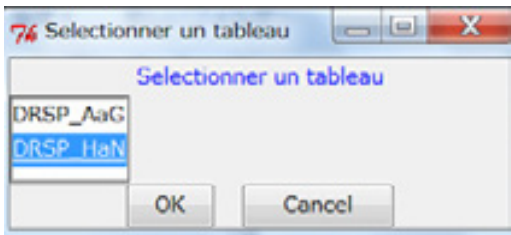


Mettez le nom d'un jeu de données : Par exemple < DRSP\_HaN > et cliquez sur OK.

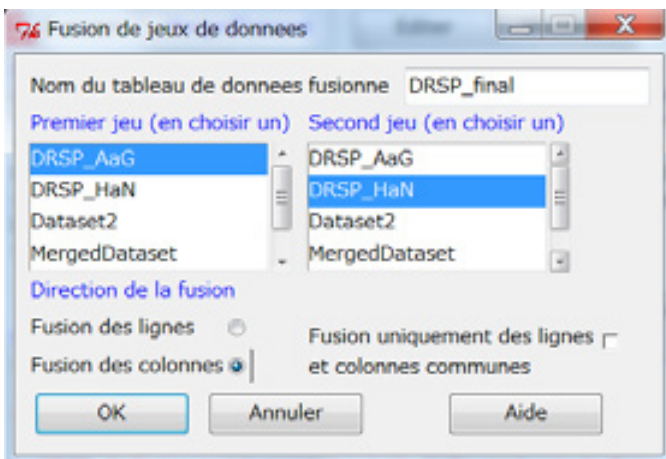
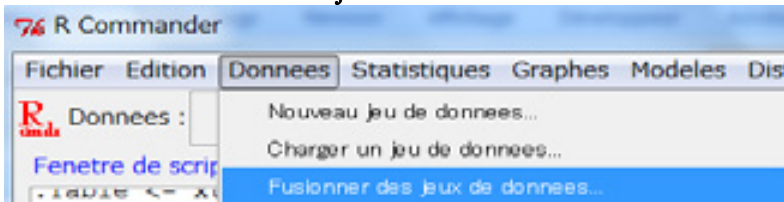


Si vous cliquez la case à cote de Données, deux tableaux seront apparus. Cliquez sur Cancel si la sélection est bonne.





**Données > Fusionner des jeux de données**



- Mettez le nom du tableau de données fusionnées.
- Sélectionnez le premier jeu et le second jeu
- Sélectionnez la Fusion des colonnes.
- Cliquez sur OK.

Pour la vérification de la fusion des données, cliquez sur Visualiser.

	N33ATAUXC>	N33BTAUXC>	N34ONGRAP>	N35LIEUEN>
1	96	93.4	OUI	Dans le R>
2	94	80	OUI	Rapports >
3	100	50	NON	
4	NA	NA	NON	
5	100	100	NON	SE/SNLS
6	100	60	OUI	CNLS
7	88.89	95.7	NON	
8	87.9	93.4	OUI	Siege.
9	NA	NA	NON	
10	91.9	76.6	NON	
11	85.6	88.53	OUI	Activites>
12	100	100	OUI	Mes archi>
13	NA	NA	OUI	Ces rappo>
14	93.7	88	NON	
15	96.5	93.3	OUI	Monsieur >
16	70.4	100	OUI	Base de d>
17	100	83.3		
18	81.1	98.2		
19	90.8	67.4	NON	
20	80	NA	OUI	PSI (M3M)
21	84.9	78.1	OUI	RMA CSB
22	86	16	NON	

## Nom des cas

### Données > Jeu de données actif > Nom des cas

Note : Quand vous avez fait la de saisie des données avec plusieurs tableaux (Access etc), vous devez utiliser le même code (ID) pour chaque données de tableaux différents. Cependant R va importer des ordres des données différents que vous avez fixés avec Access, car la fusion des jeux de données fait avec le nom de cas (**row.names**), pas le champ commun.

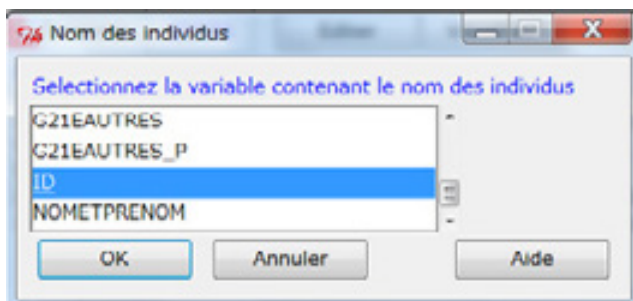
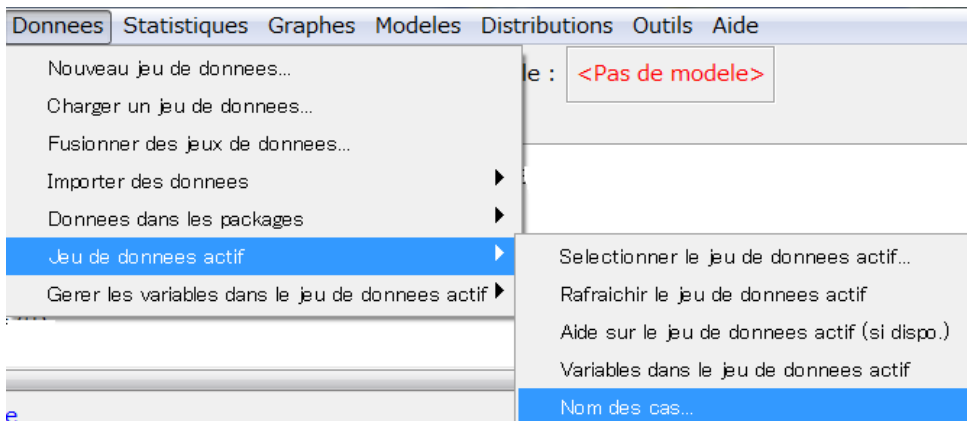
Pour que deux ou plusieurs tableaux se fusionner correctement, il faut utiliser la fonction de <Nom de cas>.

The image shows two screenshots of the R Data Viewer interface. The top window, titled 'Data: SDSP\_AaG', displays a table with 10 rows and 6 columns: ID, CODE, TYPEETUDE, DATE, and DSRP. The bottom window, titled 'Data: SDSP\_HaM', displays a table with 10 rows and 6 columns: ID, CODE, TYPEETUDE, H1DISPONIBILITEDOCUMENT, and H12POLITIQUEINTERNATIONALCTV. In both screenshots, the 'ID' column is circled in green, indicating its importance for row identification during data merging.

	ID	CODE	TYPEETUDE	DATE	DSRP
1	3	TNR_SSD_4	2008	2011-07-25	ANALAMANGA
2	4	TNR_SSD_6	2011	2011-07-28	ANALAMANGA
3	5	TNR_SSD_5	2008	2011-07-27	ANALAMANGA
4	82	TLE_SSD_3	2008	2011-07-22	ANDROY
5	83	WFI_SSD_4	2011	2011-08-16	ATSIMO AT SINANANA
6	84	TLE_SSD_6	2008	2011-07-27	ANOSY
7	22	TNR_SSD_11	2011	2011-07-18	ITASY
8	23	TNR_SSD_12	2008		ITASY
9	24	MJN_SSD_11	2008	2011-08-01	SOFIA
10	32	WFI_SSD_1	2011	2011-07-19	AMORON' I MANIA

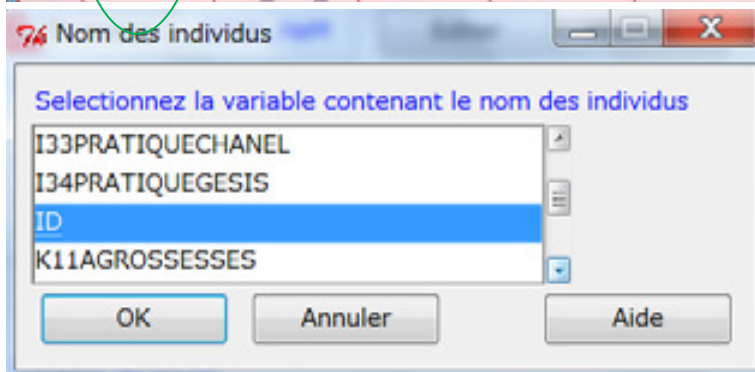
	ID	CODE	TYPEETUDE	H1DISPONIBILITEDOCUMENT	H12POLITIQUEINTERNATIONALCTV
1	14	TLE SSD 7	2008	OUI	NON
2	15	TLE SSD 15	2008	OUI	NON
3	16	TLE SSD 16	2008	OUI	NON
4	17	TLE SSD 12	2011	OUI	OUI
5	21	WFI SSD 12	2008	OUI	
6	22	TNR SSD 11	2011	OUI	NON
7	33	WFI SSD 2	2008	OUI	
8	34	WFI SSD 3	2011		NON
9	35	DIE SSD 2	2011	NON	NON
10	1	TNR SSD 2	2011	OUI	



R Data: SDSP\_AaG

Fichier

	row.names	CODE	TYPEETUDE	DATE	DSRP
1	3	TNR_SSD_4	2008	2011-07-25	ANALAMANGA
2	4	TNR_SSD_6	2011	2011-07-28	ANALAMANGA
3	5	TNR_SSD_5	2008	2011-07-27	ANALAMANGA
4	82	TLE_SSD_3	2008	2011-07-22	ANDROY
5	83	WFI_SSD_4	2011	2011-08-16	ATSIMO ATSIANANA
6	84	TLE_SSD_6	2008	2011-07-27	ANOSY
7	22	TNR_SSD_11	2011	2011-07-18	ITASY
8	23	TNR_SSD_12	2008		ITASY
9	24	MJN_SSD_11	2008	2011-08-01	SOFIA
10	32	WFI_SSD_1	2011	2011-07-19	AMORON' I MANIA





	CODE	TYPEETUDE	H11DISPONIBILITE	DOCUMENT	H12POLITIQUE	NATIONA
14	TLE_SSD_7		2008		OUI	
15	TLE_SSD_15		2008		OUI	
16	TLE_SSD_16		2008		OUI	
17	TLE_SSD_12		2011		OUI	
21	WFI_SSD_12		2008		OUI	
22	TNR_SSD_11		2011		OUI	
33	WFI_SSD_2		2008		OUI	
34	WFI_SSD_3		2011		<NA>	
35	DIE_SSD_2		2011		NON	
1	TNR_SSD_2		2011		OUI	

Note pour 1<sup>er</sup> tableau, ID sera affiché comme row.names.

Donnees Statistiques Graphes Modeles Dist

Nouveau jeu de donnees...

Charger un jeu de donnees...

Fusionner des jeux de donnees...

Fusion de jeux de donnees

Nom du tableau de donnees fusionne SDSP\_AaM

Premier jeu (en choisir un) Second jeu (en choisir un)

SDSP\_AaG SDSP\_AaG

SDSP\_HaM SDSP\_HaM

Direction de la fusion

Fusion des lignes  Fusion uniquement des lignes et colonnes communes

Fusion des colonnes

OK Annuler Aide

R Data: SDSP\_AaM

Fichier

	row.names	CODE.x	TYPEETUDE.x
1	1	TNR_SSD_2	2011
2	10	TMM_SSD_10	2011
3	11	TMM_SSD_9	2008
4	12	TLE_SSD_13	2011
5	13	TLE_SSD_14	2008
6	14	TLE_SSD_7	2008
7	15	TLE_SSD_15	2008
8	16	TLE_SSD_16	2008
9	17	TLE_SSD_12	2011
10	18	WFI_SSD_10	2008

R Data: SDSP\_AaM

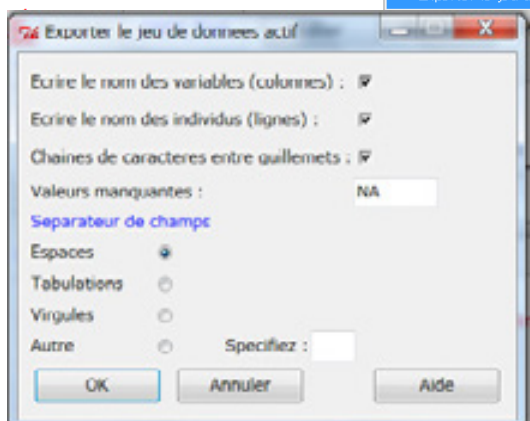
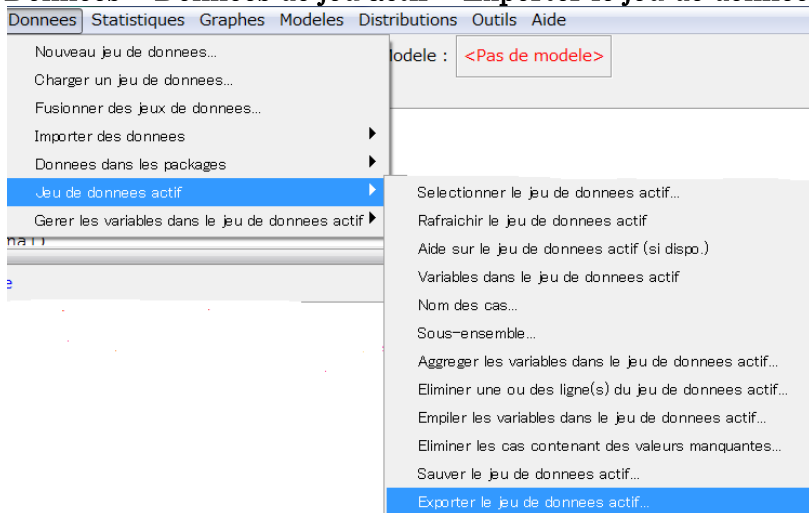
Fichier

	CODE.y	TYPEETUDE>	H11DISPON>	H12POLIT>
1	TNR_SSD_2	2011	OUI	
2	TMM_SSD_10	2011	OUI	NON
3	TMM_SSD_9	2008	OUI	NON
4	TLE_SSD_13	2011	OUI	NON
5	TLE_SSD_14	2008	OUI	NON
6	TLE_SSD_7	2008	OUI	NON
7	TLE_SSD_15	2008	OUI	NON
8	TLE_SSD_16	2008	OUI	NON
9	TLE_SSD_12	2011	OUI	OUI
10	WFI_SSD_10	2008	OUI	NON

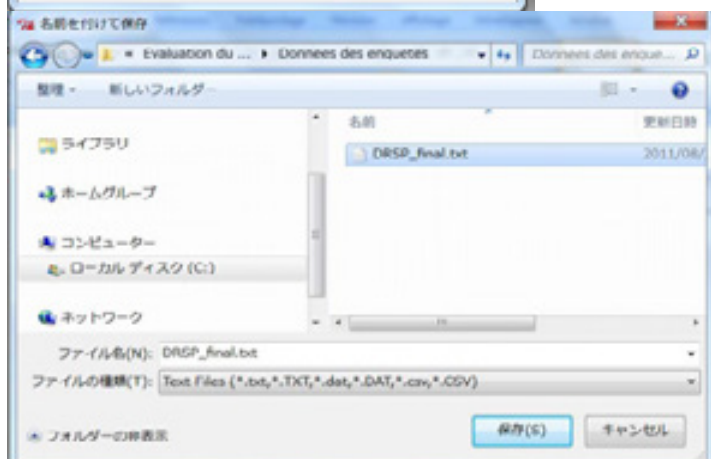
## Exporter le jeu de données actif

R peut exporter le jeu de données actif en tant que le fichier texte (.txt)

**Données > Données de jeu actif > Exporter le jeu de données actif**



Cliquez sur OK s'il y a pas de changement.

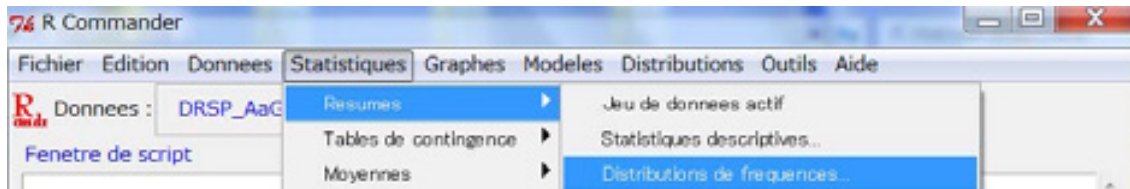


Vous allez indiquer un endroit de sauvegarder des données.

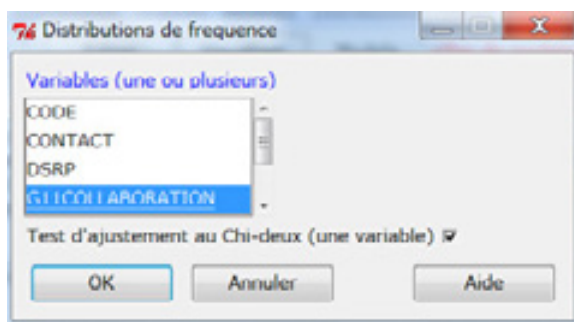
## Distribution de fréquence

### Statistiques > Résumé > Distribution de fréquence

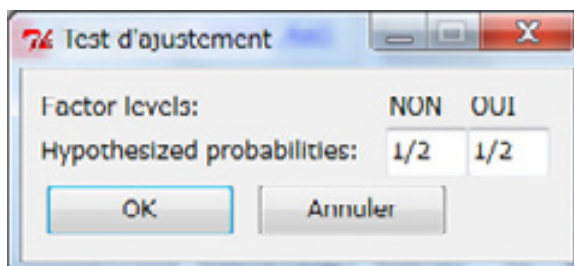
Vous pouvez compter des nombres et pourcentage de données en facteur et test d'ajustement de Chi-Carrée.



Sélectionnez une ou plusieurs variables.



Si vous voulez faire un Test d'ajustement au Chi-deux, sélectionnez une variable et cochez la case de Test d'ajustement.



Vous allez définir les probabilités d'hypothèse. Par défaut R peut calculer des probabilités en fonction de type de données.

```
Fenetre de sortie
> .Table # counts for G11COLLABORATION
NON OUI
  2 19

> round(100*.Table/sum(.Table), 2) # percentages for G11COLLABORATION
NON OUI
9.52 90.48

> .Probs <- c(0.5,0.5)
> chisq.test(.Table, p=.Probs)
Chi-squared test for given probabilities
```

```

> .Probs <- c(0.5,0.5)
> chisq.test(.Table, p=.Probs)

      Chi-squared test for given probabilities
data:  .Table
X-squared = 13.7619, df = 1, p-value = 0.0002075

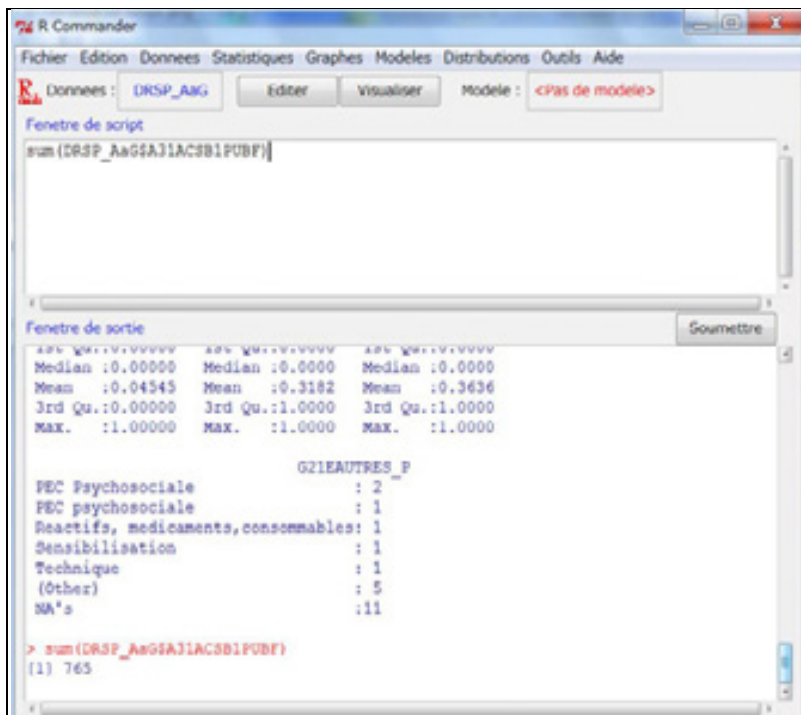
> remove(.Probs)
> remove(.Table)

```

df=Degrés de  
Liberté

### Utilisation de Fenêtre de Script

Si vous connaissez certains script de R comme sum() qui signe le total ou la somme, vous allez directement écrire des script directement dans la Utilisation de Fenêtre de Script.



sum(nom de jeu de données  
actif et ajouter \$ et ensuite  
nom de variable)

Ex.

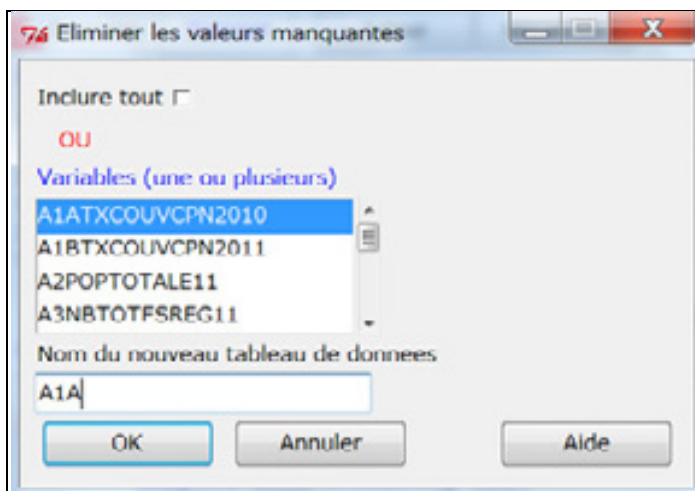
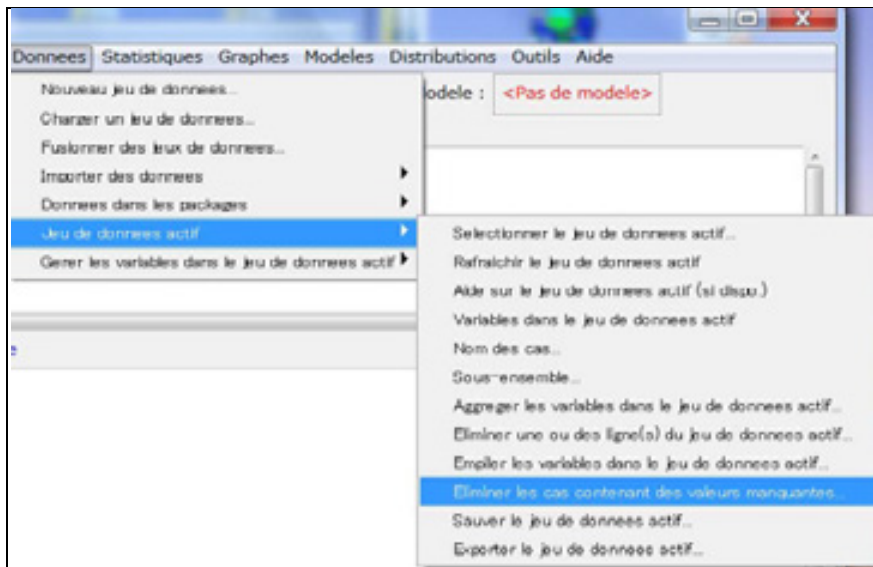
sum(DRSP\_AaG\$A31ACSB1  
PUBF)

Cliquez Soumettre.

Note : si vos variables  
contient NA, R ne peut pas  
calculer le la somme.

## Eliminer les cas contenant des valeurs manquants

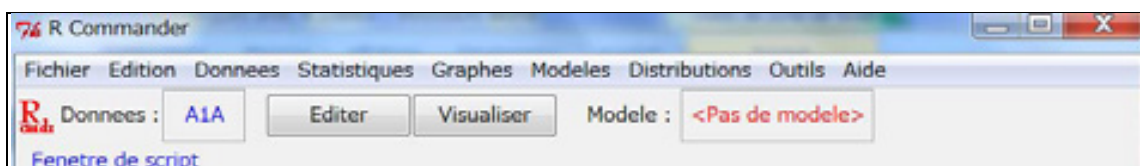
Données > Jeu de données actif > Eliminer les cas contenant des valeurs manquants



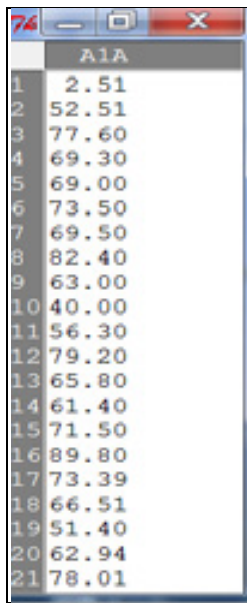
Décochez la case de Inclure tout si vos variables.

Sélectionnez une variable et mettez le nom du nouveau tableau.

Quand vous cliquez sur OK, le nouveau tableau sera apparu en tant que le jeu de données actif.



Pour la vérification des données, cliquez sur Visualiser.

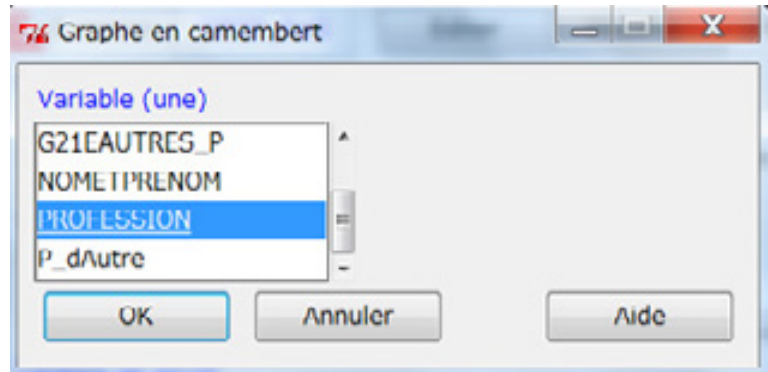
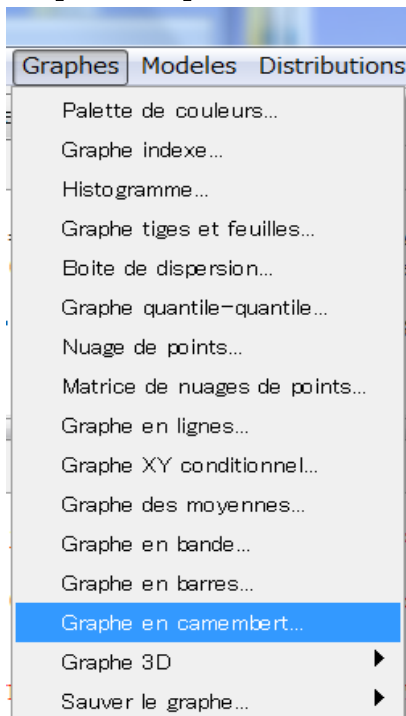


A screenshot of a spreadsheet application window titled "A1A". The window displays a list of 21 rows, each with a numerical value. The values are: 2.51, 52.51, 77.60, 69.30, 69.00, 73.50, 69.50, 82.40, 63.00, 40.00, 56.30, 79.20, 65.80, 61.40, 71.50, 89.80, 73.39, 66.51, 51.40, 62.94, and 78.01.

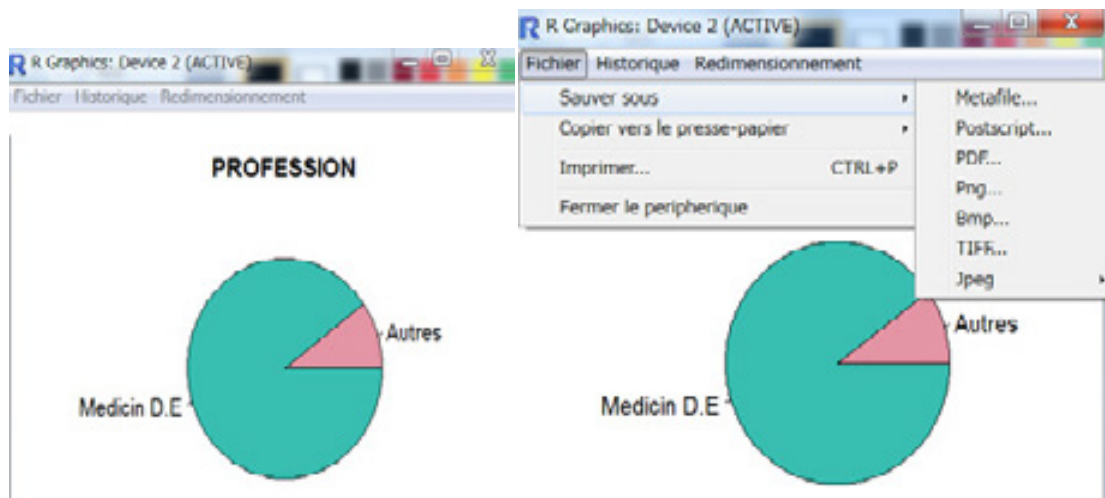
	A1A
1	2.51
2	52.51
3	77.60
4	69.30
5	69.00
6	73.50
7	69.50
8	82.40
9	63.00
10	40.00
11	56.30
12	79.20
13	65.80
14	61.40
15	71.50
16	89.80
17	73.39
18	66.51
19	51.40
20	62.94
21	78.01

## Graphe en camembert

Graphe > Graphe en camembert



Sélectionnez une variable et cliquez sur OK.



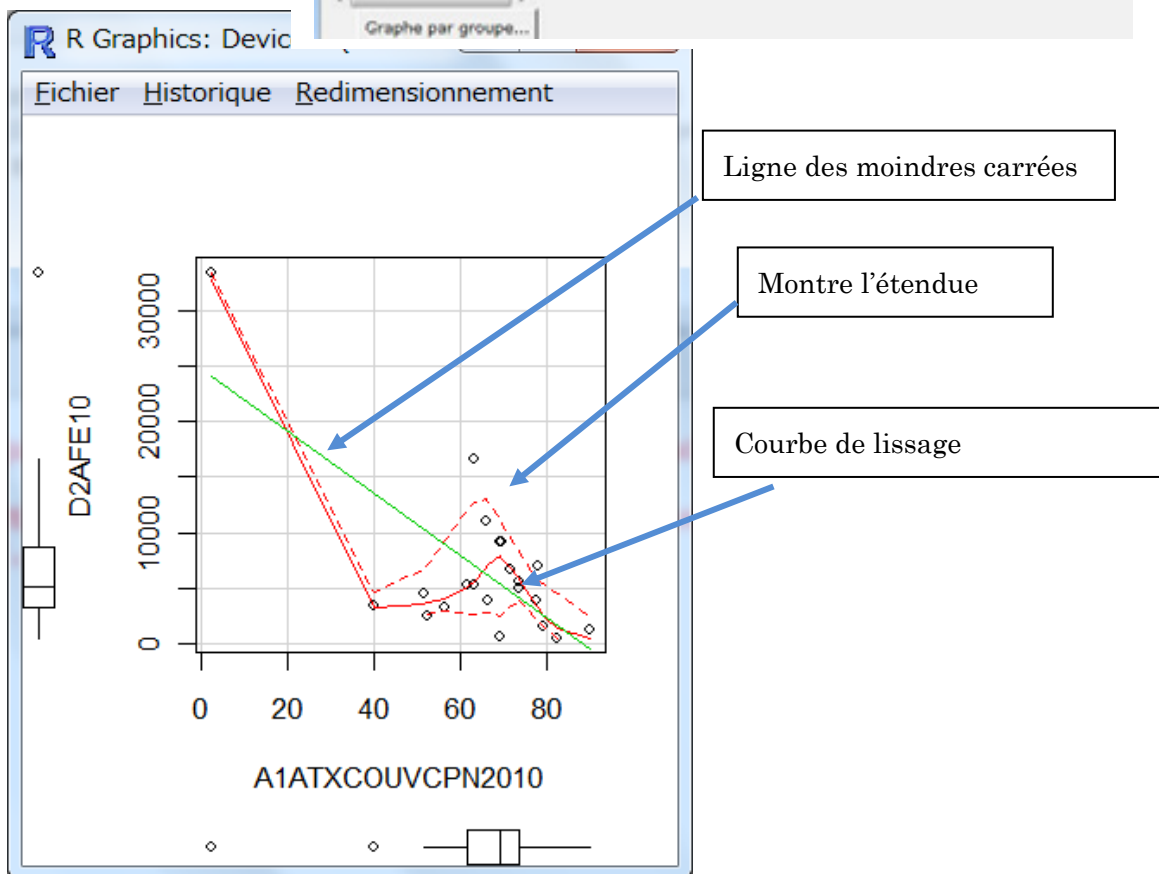
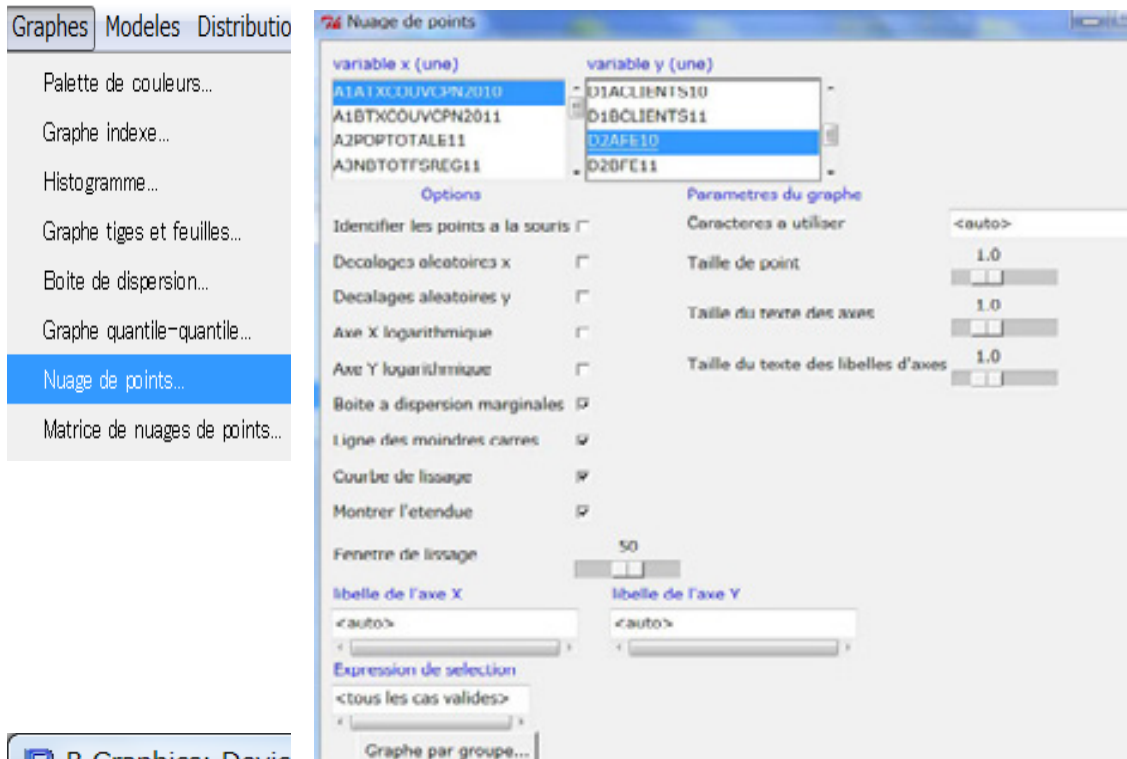
Pour sauver le graphe,

Fichier > Sauver sous > Png ou Jpeg etc



## Graphe en Nuage de points

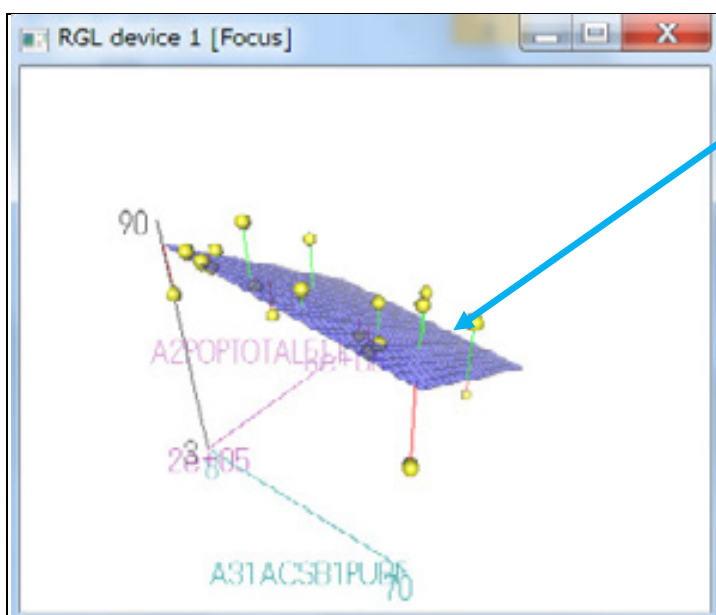
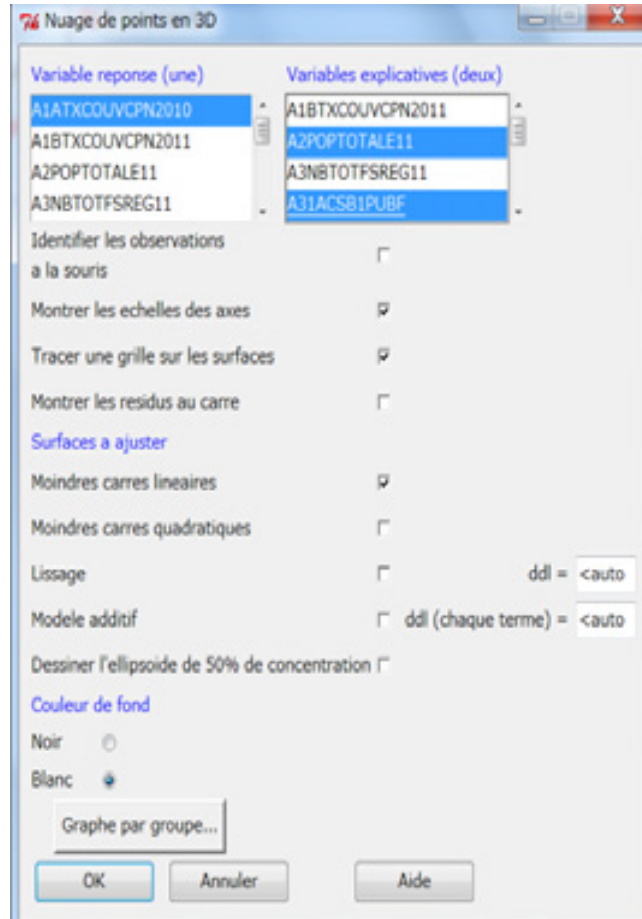
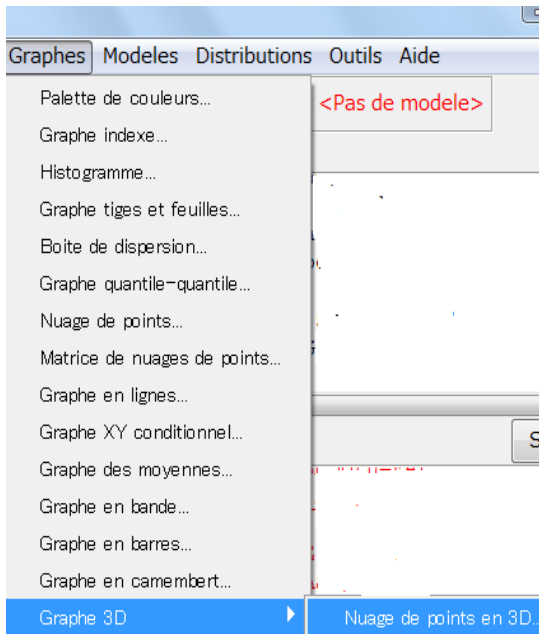
Graphe > Nuage de points





## Graphe en 3D

### Graphe > Graphe en 3D > Nuage de points en 3D

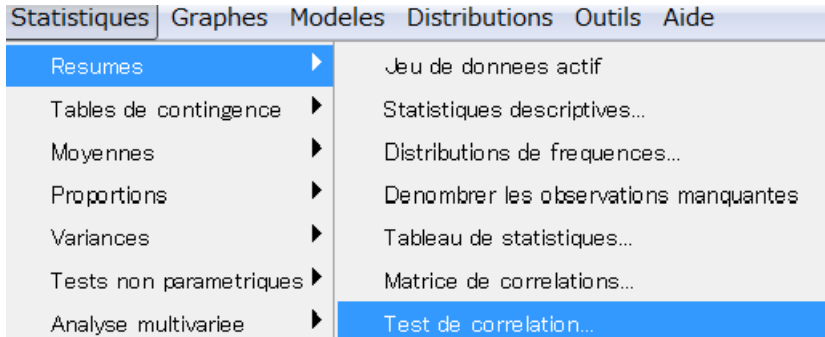


Moindre carrée linéaire

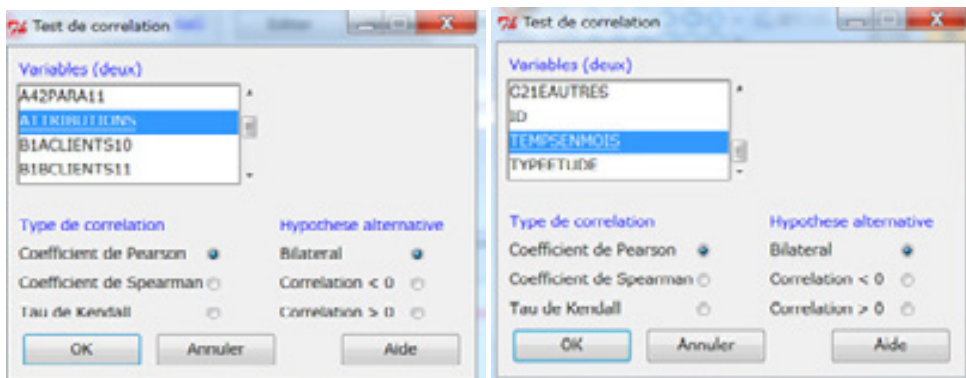
Note : Si vous glissez le curseur avec le clic gauche, le graphe va se tourner.

## Test de corrélation

Statistique >Résumé > Test de corrélation



Note. Cette exercice est avec les données de table SDSP\_AaG de SDSP final.(Access)



En cliquant Ctrl, vous devez sélectionner deux variables.

Pour le Type de corrélation, si votre variables sont comme la distribution normale, vous utilisez Coefficient de Person, si non coefficient de Sepearman. Taux de Kendall est moins utilisé par rapport de deux méthodes.

Pour Hypothèse alternative, vous sélectionnez Bilatéral.

```
> cor.test(SDSP_AaG$ATTRIBUTIONS, SDSP_AaG$TEMPSENMOIS,
+ alternative="two.sided", method="pearson")

Pearson's product-moment correlation

data: SDSP_AaG$ATTRIBUTIONS and SDSP_AaG$TEMPSENMOIS
t = 1.3975, df = 82, p-value = 0.166
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.06396956  0.35529477
sample estimates:
      cor
0.1525174
```

Cor signifie le coefficient de corrélations. Cor =r=0.1525174 est très faible corrélation.

Pour une vérification de la corrélation, vous allez créer le graphe de nuage des

points.

- Graphes
- Modeles
- Distributions
- Palette de couleurs...
- Graphe indexe...
- Histogramme...
- Graphe tiges et feuilles...
- Boite de dispersion...
- Graphe quantile-quantile...
- Nuage de points...

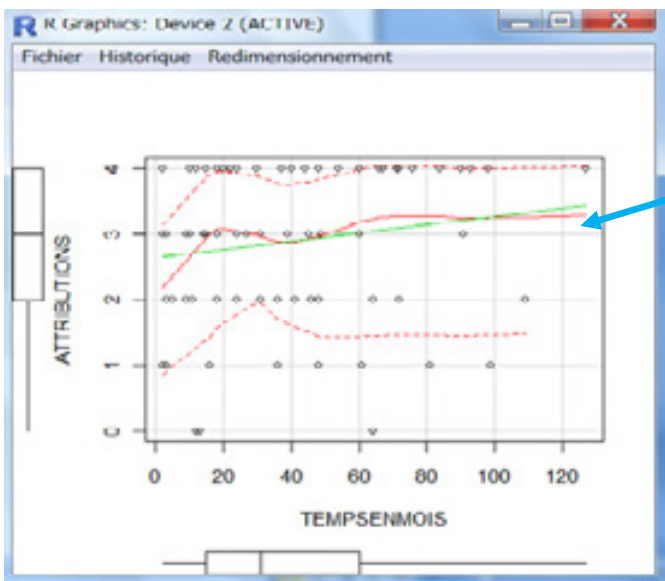
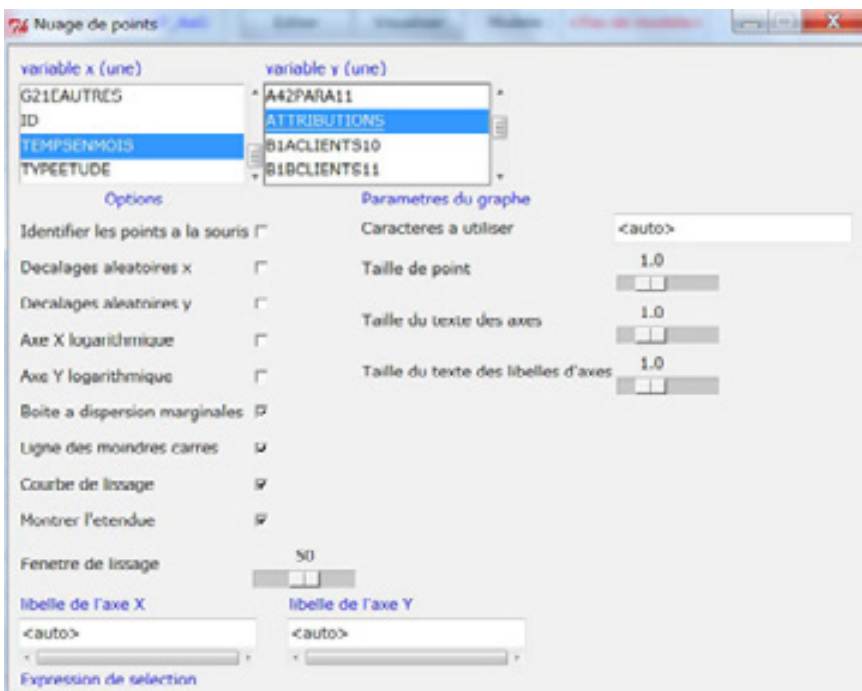
$r =$  Coefficient de *corrél*ation

$-0,2 \leq r \leq 0,2$                       corrélation très faible

$-0,4 \leq r < -0,2$     $0,2 < r \leq 0,4$    corrélation faible

$-0,7 \leq r < 0,4$     $0,4 < r \leq 0,7$    corrélation moyennement forte

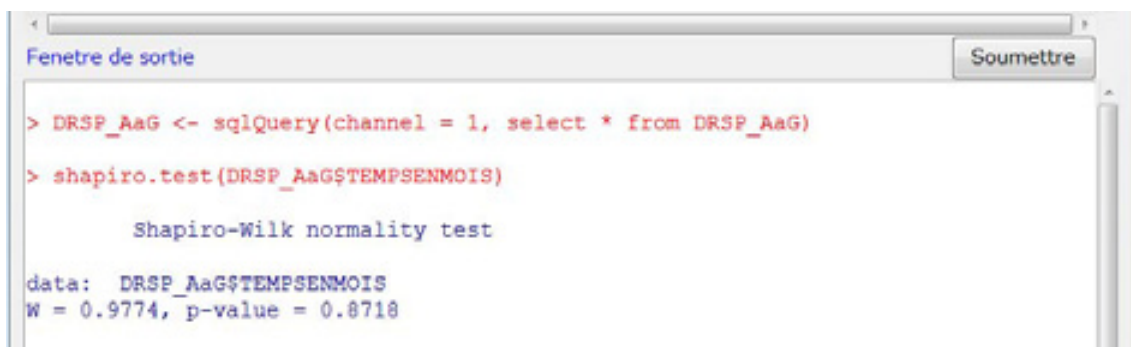
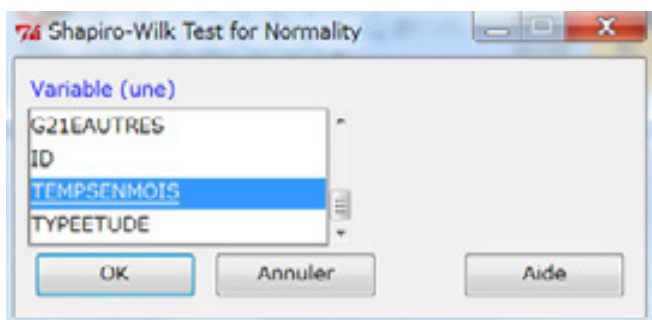
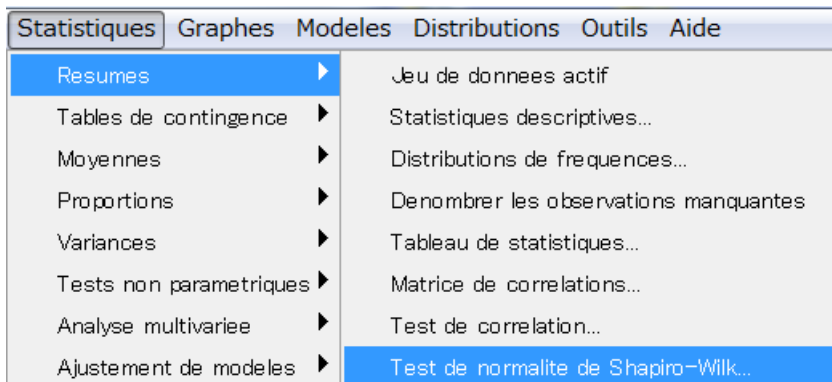
$-1,0 \leq r < -0,7$     $0,7 < r \leq 1,0$    corrélation forte



Vérifiez sur la ligne des moindres carrés

## Test de normalité de Shapiro Wilk

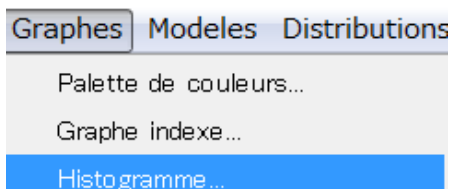
Statistiques > Résumé > Test de Normalité de Sapiro Wilk

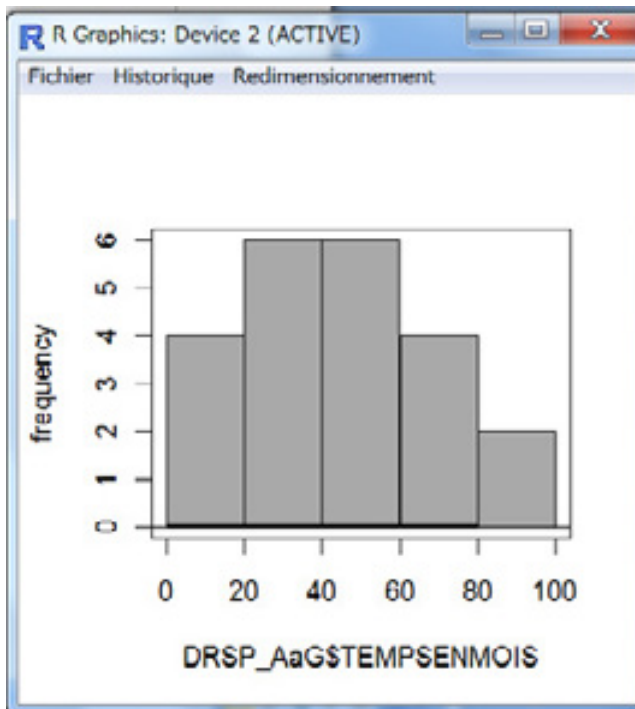
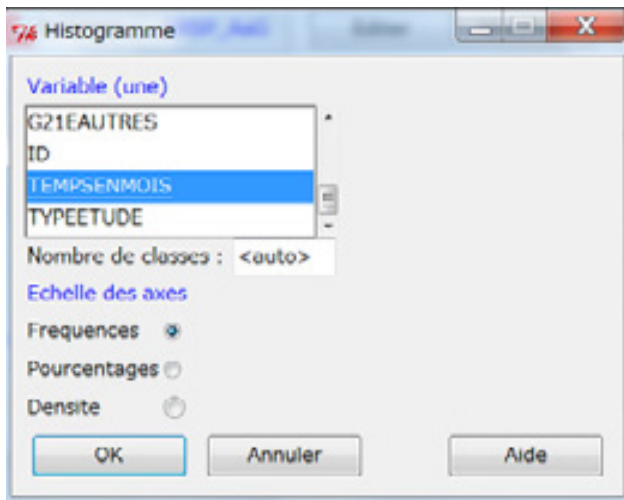


Comme la valeur de  $P > 0,05$  ( $P=0,8718$ ), on ne peut pas dire que cette variable ne suit pas la distribution normale Si bien que vous utiliser le test paramétrique.

On va vérifier cette distribution avec une graphe.

## Graphe>Histogramme

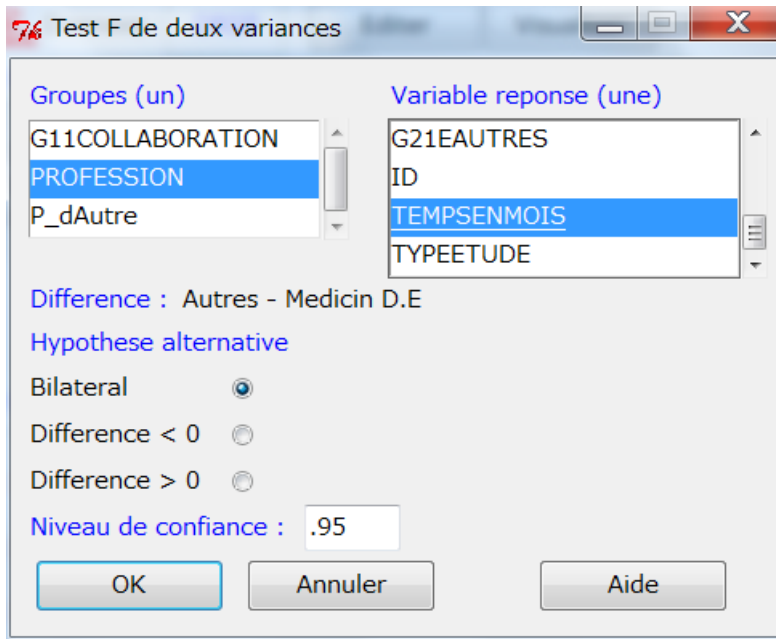
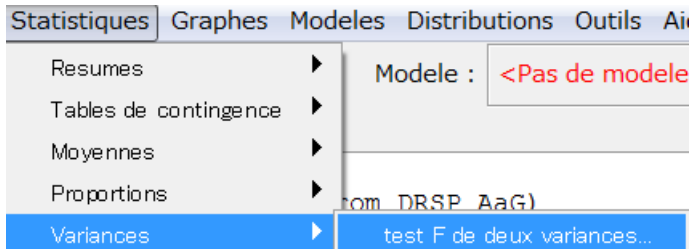




Le graphe est comme la distribution normale.

## Test F de deux variances

Avant que deux comparer deux moyennes, vous devez vérifier que variables cibles suivent la distribution normale (Test de normalité de Sapio Wilk )et les variances sont égales ou pas (Test F de deux variances) . Nous voudrions comparer les moyens de temps en tant que le responsable VIH par profession au niveaux de DRSP.



```
> tapply(DRSP$TEMPSENMOIS, DRSP$PROFESSION, var, na.rm=TRUE)
      Autres Medicin D.E
128.0000    634.9281
```

```
> var.test(TEMPSENMOIS ~ PROFESSION, alternative='two.sided', conf.level=.95,
+ data=DRSP)
```

F test to compare two variances

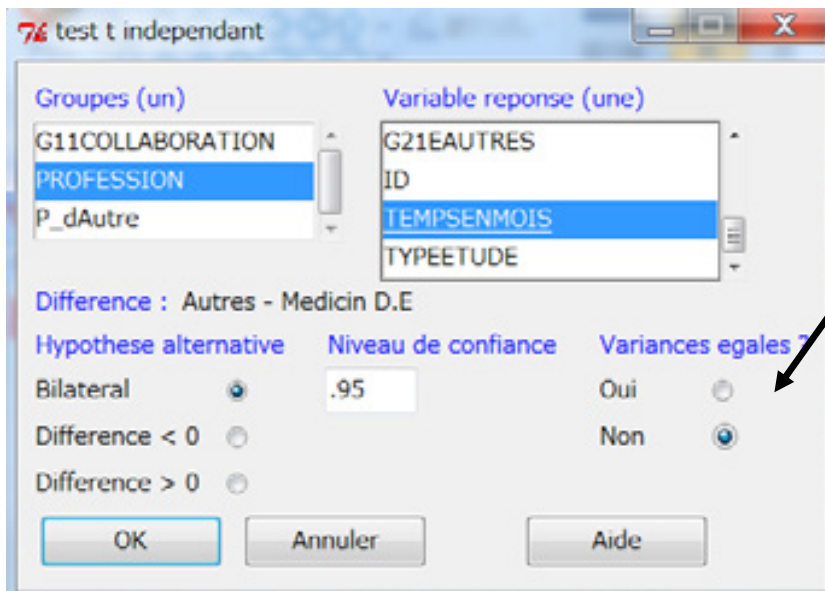
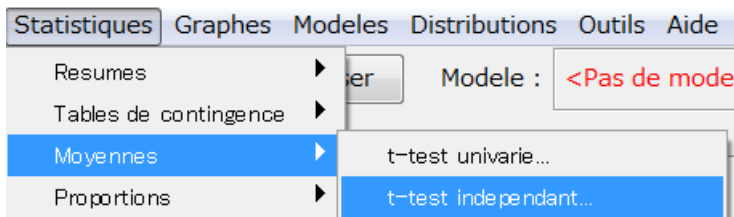
```
data:  TEMPSSENMOIS by PROFESSION
F = 0.2016, num df = 1, denom df = 17, p-value = 0.6818
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.03336597 199.32624246
sample estimates:
ratio of variances
 0.2015976
```

Comme le résultat de F test de Variance, on prend une hypothèse alternative <variance n'est pas égal>, il faut faire le test de Welch.

## Test de Welch

Test de Welch est inclus dans t-test indépendant.

Statistiques > Moyennes > T Test indépendant



Nous avons fait le test de variance (test F de deux variance) et résultat est que variances ne sont pas égales.

Vous cliquez sur NON.

Attention !

Par défaut, variance égales : NON.

Quand vous cliquez sur OUI, R va calculer T Test indépendant :

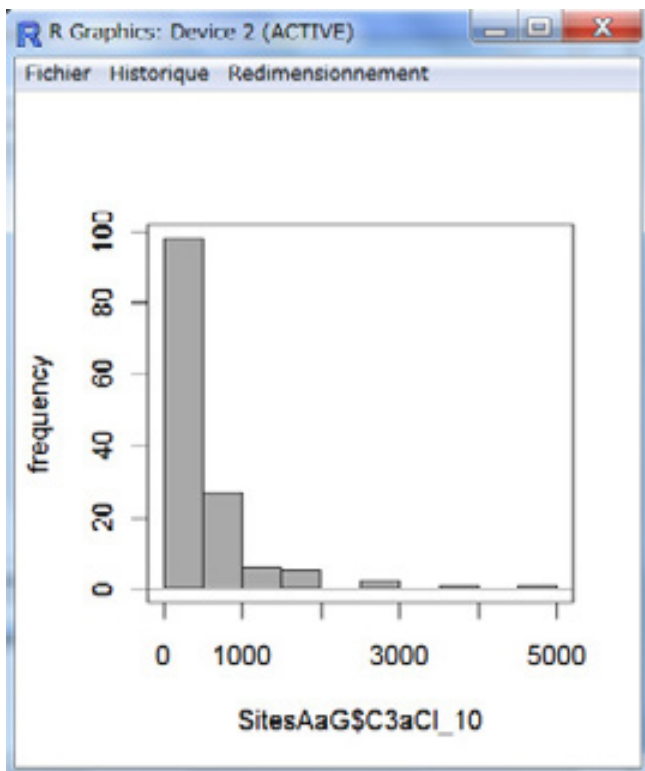
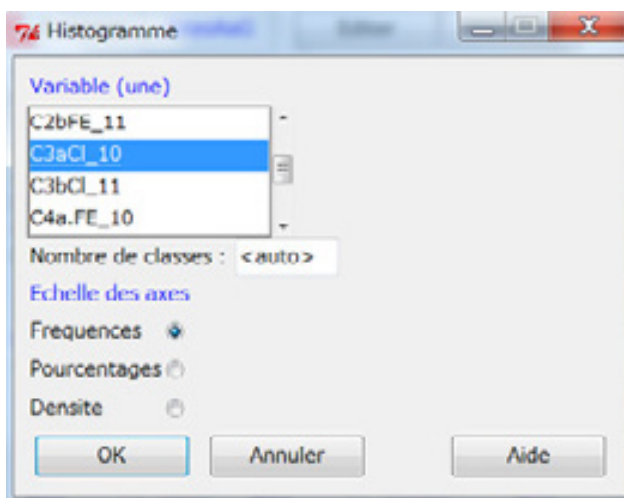
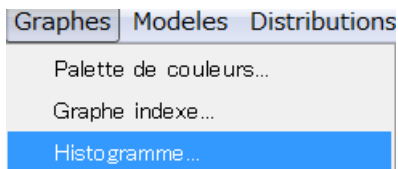
```
Welch Two Sample t-test

data:  TEMPSENMOIS by PROFESSION
t = 1.0929, df = 2.364, p-value = 0.3732
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-26.23924  48.01702
sample estimates:
mean in group Autres mean in group Medicin D.E
          55.00000          44.11111
```



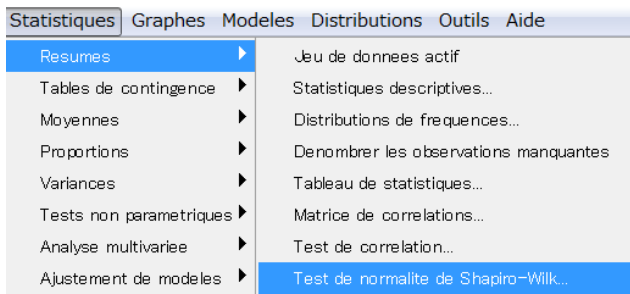
## Test de Wilcoxon bivarié (Test non paramétrique pour comparer deux moyennes (Médians))

Nous voudrions comparer la moyenne de deux groupes de FS (2008 et 2011) pour la réalisation de dépistage du VIH en 2011.



Cette variable ne se ressemble avec la distribution normale.



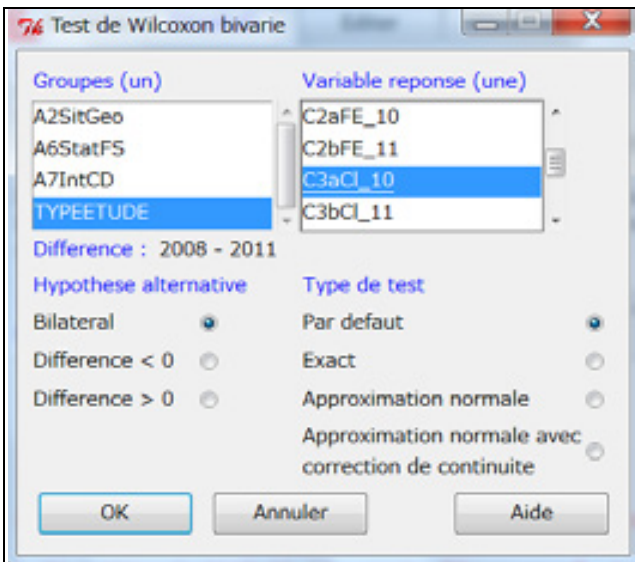
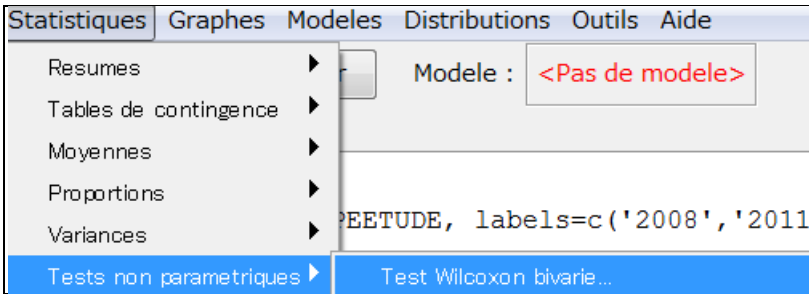


```
> shapiro.test(SitesAaG$C3aCl_10)

      Shapiro-Wilk normality test

data:  SitesAaG$C3aCl_10
W = 0.6484, p-value < 2.2e-16
```

Le résultat de test de Sapiro Wilk est  $P < 0,05$  si bien que cette variable ne suit pas la distribution normale.



```
> tapply(SitesAaG$C3aCl_10, SitesAaG$TYPEETUDE, median, na.rm=TRUE)
2008 2011
 470  160

> wilcox.test(C3aCl_10 ~ TYPEETUDE, alternative='two.sided', exact=TRUE,
+   correct=FALSE, data= SitesAaG)

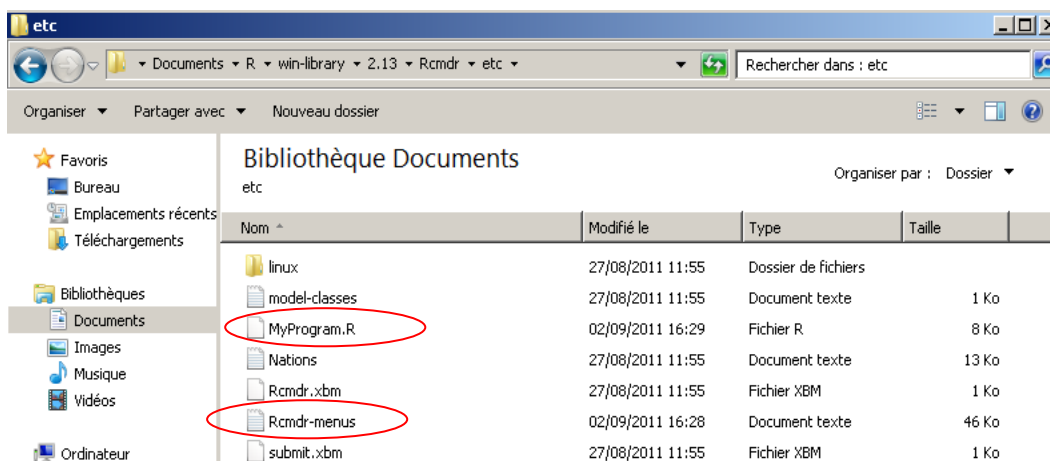
      Wilcoxon rank sum test

data:  C3aCl_10 by TYPEETUDE
W = 3571.5, p-value = 8.09e-07
alternative hypothesis: true location shift is not equal to 0
```

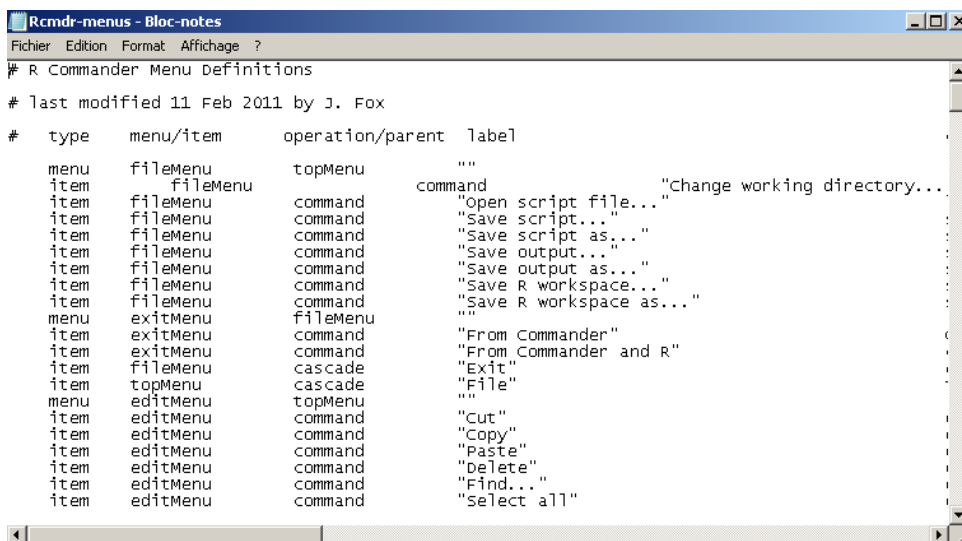
Note : Le test non paramétrique ne montre pas la moyenne mais le médian car les variables ne suit pas la distribution normale.

## R commander Personnalisé

Vous pouvez personnaliser R commander en modifiant le fichier MyProgramme.R et le fichier Rcmdr-menus. Si vous ne connaissez pas des programmations ou Rcmdr, il vaut mieux consulter cette partie à quelqu'un qui connaisse R commander et/ou programmation. Si vous avez des fichiers de MyProgramme.R et Rcmdr-menus qui ont déjà personnalisé, vous les copiez et collez dans le dossier avec Répertoire : MyDocement>R>win-library>2.13>Rcmdr>etc ou Programme files>R>library>2.13>Rcmdr>etc.



Si vous ouvrez Rcmdr-menus, le contenu va s'apparaître comme ci-dessous ; Si vous modifiez Rcmdr-menus, menus peuvent être changé.



Si vous ouvrez MyProgramme.R, le contenu va s'apparaître comme ci-dessous ; Si vous modifiez Myprogramme R, des programmes peuvent être changé

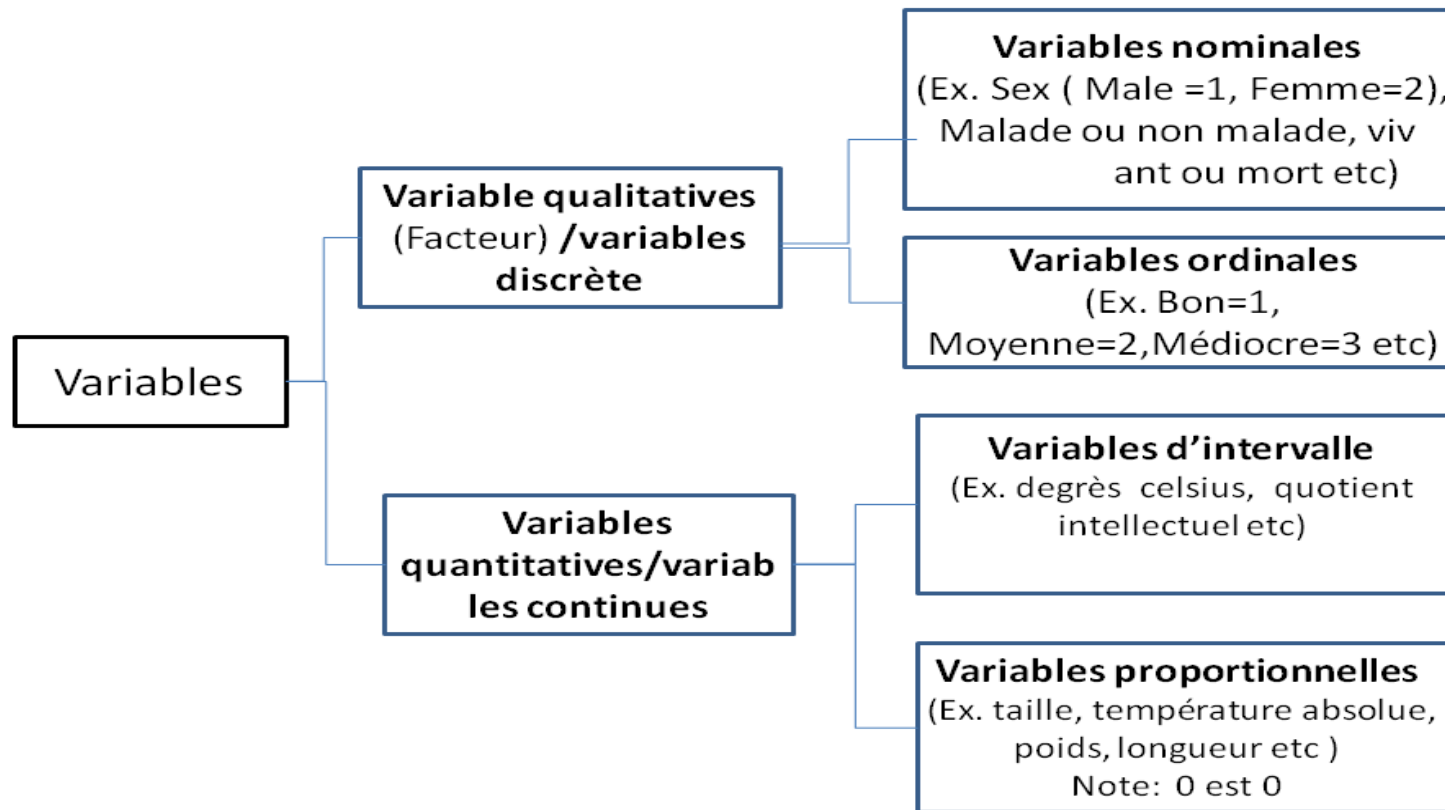


**ANNEXE : Test statistique de base et R commander**

**Test statistique de base et Rcmdr(R commander)**

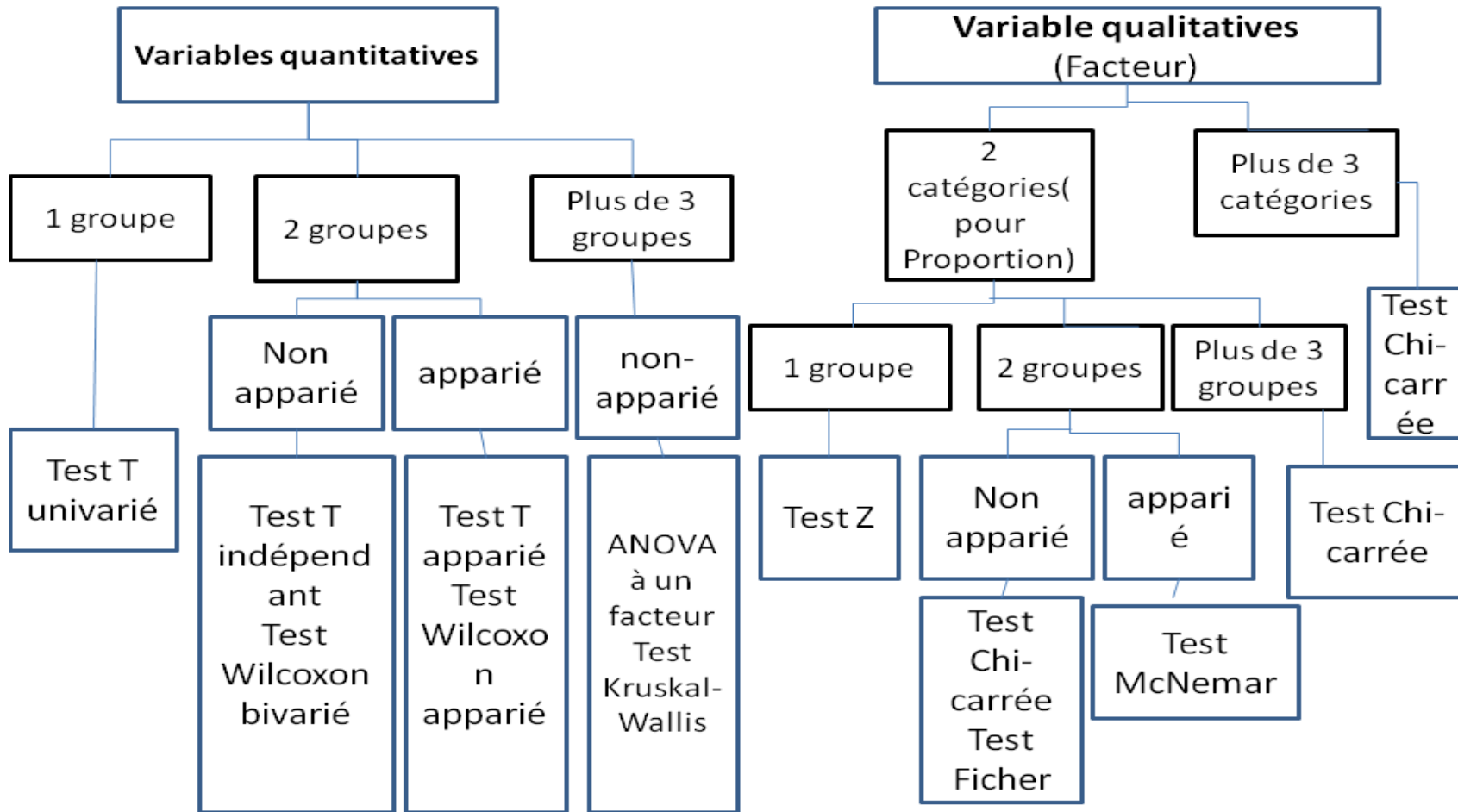
	<b>Test statistique</b>	<b>Hypothèse nulle</b>	<b>Rcmdr</b>
Test de la Différence de moyenne entre deux groupes	-Test T (Paramétrique) -Test Welch( Si la variance n'est pas égale) -Test Wilcoxon (Non paramétrique)	Moyenne de deux group est égale	-Statistiques>Moyenne >T test indépendant ou apparié (si vous sélectionnez -Statistiques>Test non paramétrique >Test Wilcoxon indépendant ou apparié
Test de la Différence de moyenne entre trois groupes	-ANOVA(Paramétrique) -Test Kruskal-Wallis(Non-Paramétrique)	Les moyennes de tous les groupes sont égal	-Statistiques>Moyenne >ANOVA à un facteur - Statistiques>Test non paramétrique >Test Kruskal-Wallis
Test de la Différence de proportion	-Chi-carrée -Test exact de Fisher (une celle de fréquence escompté sera mois de 5) -Test McNemar	La proportion de tous les groups de cases est égal	-Statistiques>Table de contingence >Tableau à double entrée ou plusieurs entrée
Test de corrélation	-Coefficient de Person -Coefficient de Spearman	Coefficient de correlation =0	-Statistiques>Résumés >Test de corrélation

## Type de données



Aoyama, Epidémiologie d'aujourd'hui

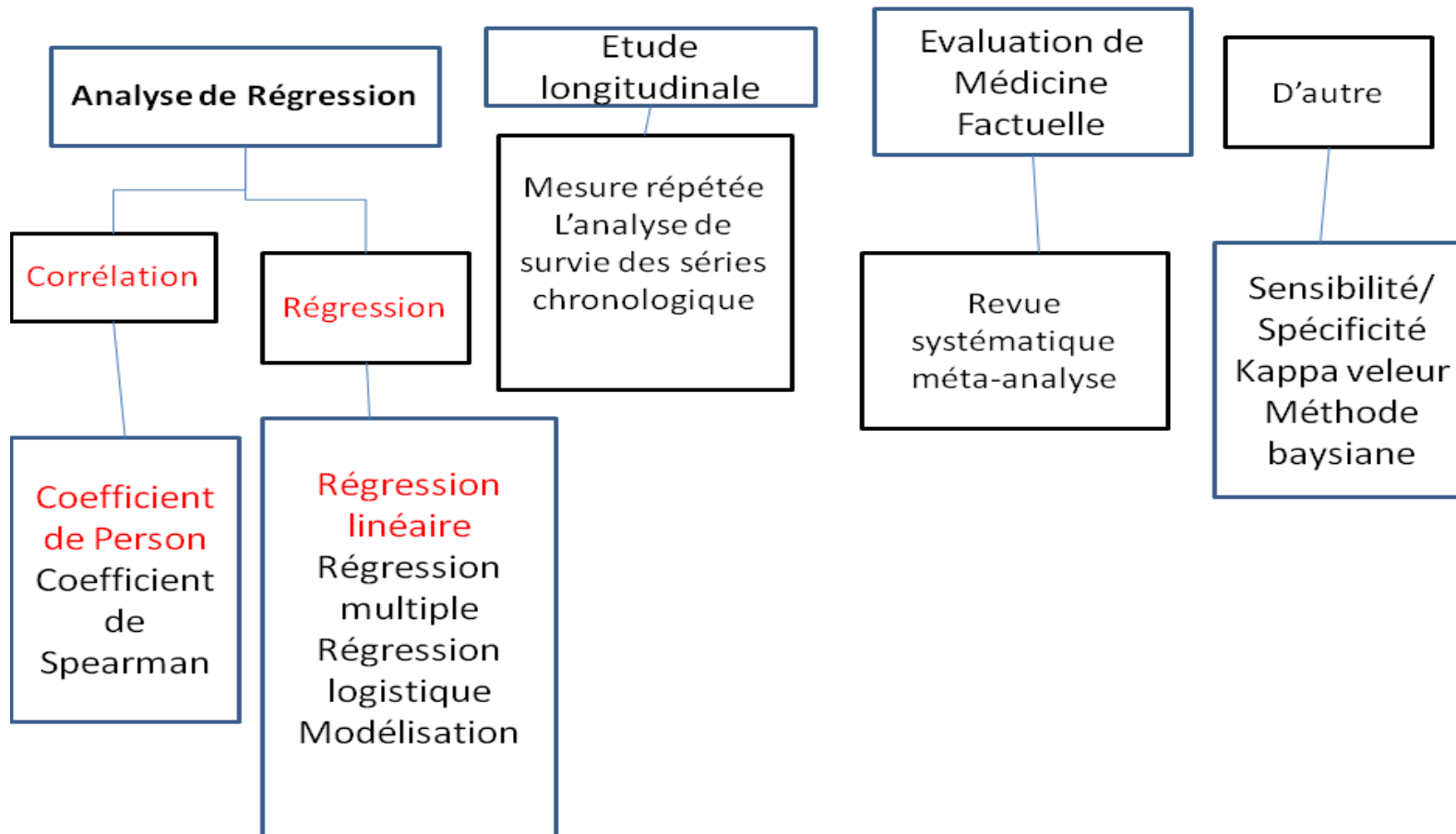
Schéma de test statistique 1



Aoyama, Epidémiologie d'aujourd'hui



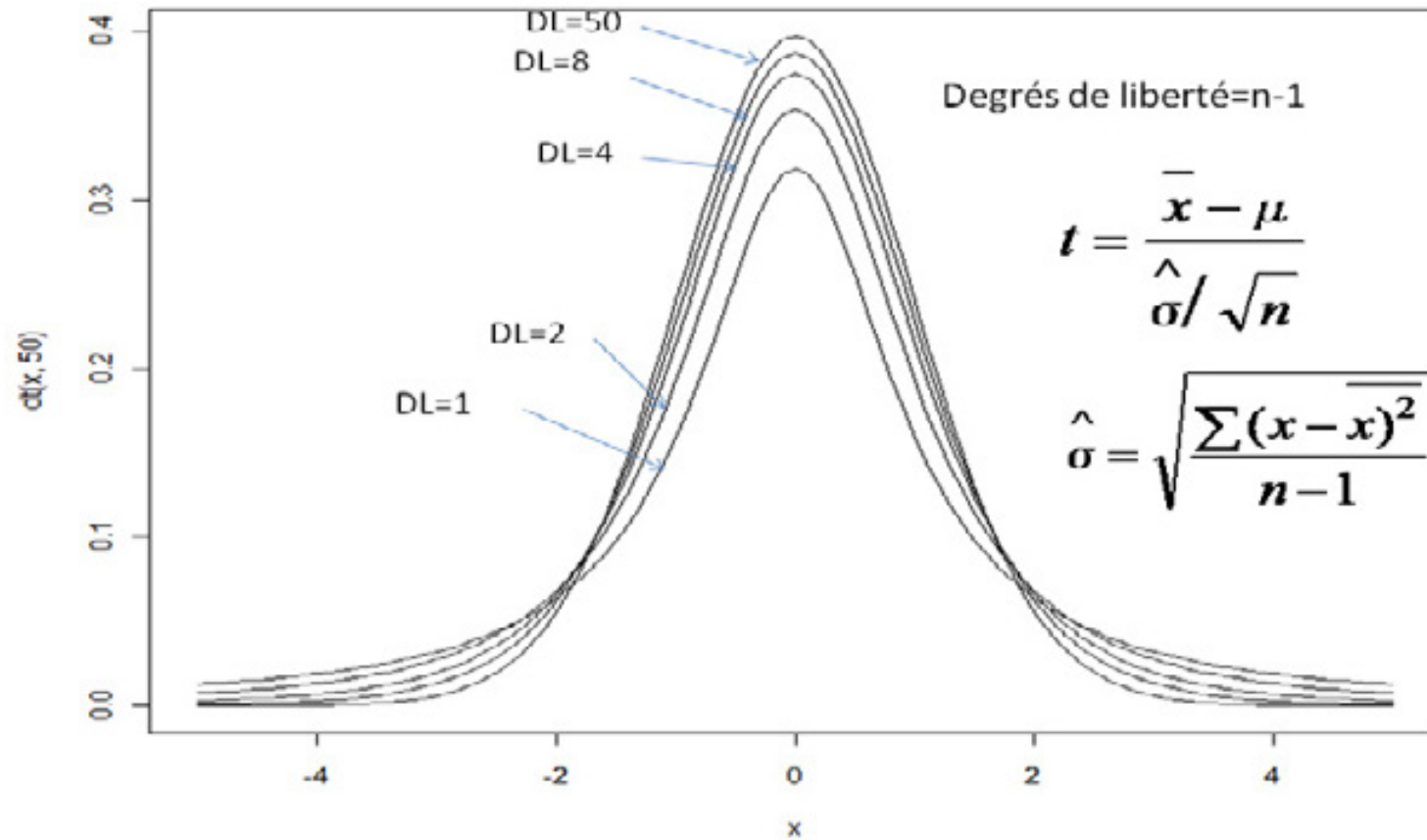
## Schéma de test statistique 2



Aoyama, Epidémiologie d'aujourd'hui

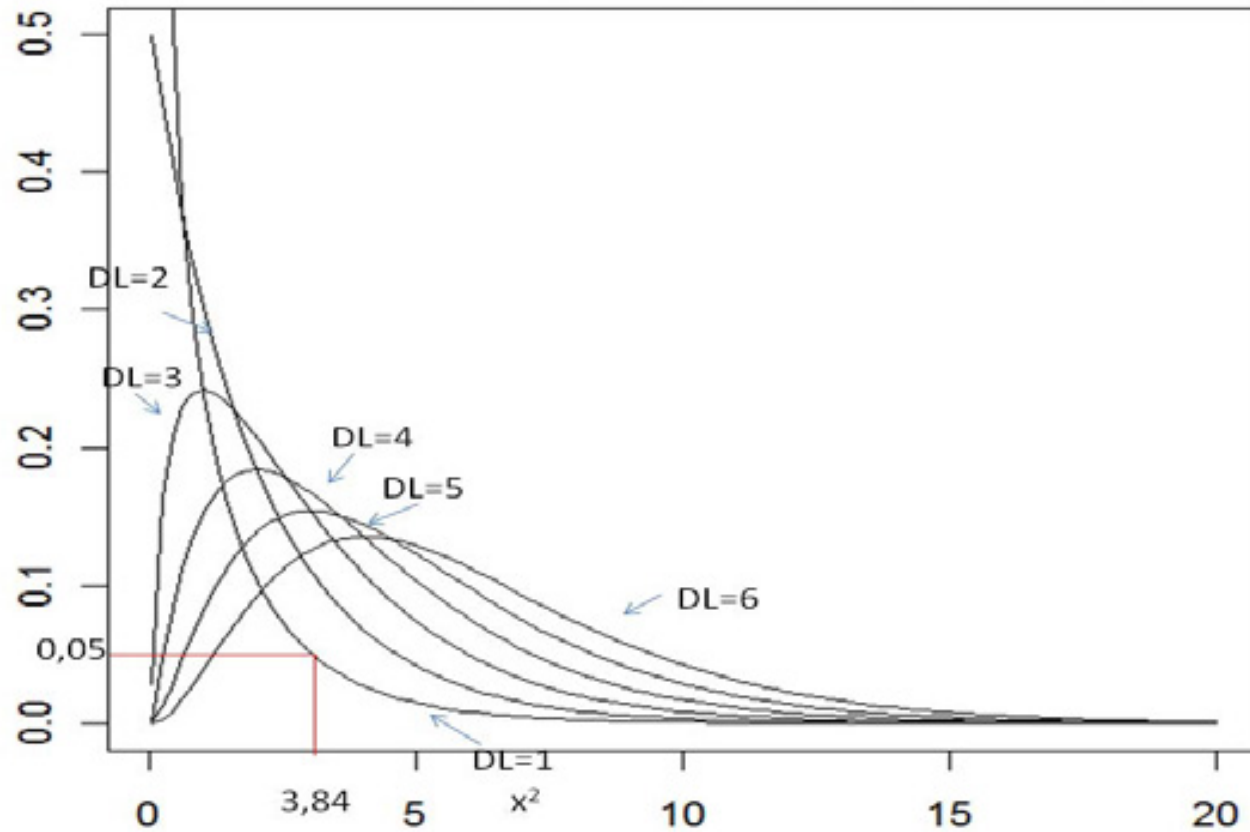
## Distribution T

# Distribution T



Distribution Chi-carrée

# Distribution de $\chi^2$



Degrés de Liberté(DL) de Chi carré (Colonne-1)X(rangée-1)  
Ex: Tableau 2X2 : DL=1

$$\sum_{i=1}^K \frac{(O_i - E_i)^2}{E_i}$$

O=Fréquence Observé  
E=Fréquence Escompté