



Guide SPSS **(pour les nuls, mais pas trop quand même)**

Sophie Nidegger & Pascal Wagner-Egger

TABLE DES MATIERES

1.	Constitution d'une base de données SPSS.....	4
1.1.	Lancer SPSS (PC ou MAC)	4
1.1.1.	Options de lancement de SPSS	5
1.2.	Fenêtre principale	6
1.3.	Définition d'une variable et de ses propriétés	7
1.4.	Enregistrement des données	12
1.5.	Quitter SPSS.....	12
2.	Importation d'un fichier de données à partir d'Excel.....	12
2.1.	Fichier de données Excel.....	12
2.2.	Importation des données.....	13
3.	Fonctions de base de SPSS.....	14
3.1.	Insérer/déplacer une variable.....	15
3.2.	Options d'affichage des variables	17
3.3.	Option d'affichage des données.....	19
4.	Manipulation des données.....	20
4.1.	Créer une variable à partir d'une ou de plusieurs variable(s) existante(s) ...	20
4.2.	Recoder des variables	22
4.3.	Fragmenter la base de données	25
4.4.	Travailler sur une partie des données.....	26
5.	Statistiques Descriptives.....	29
5.1.	Calcul des fréquences.....	29
5.2.	Tableaux croisés ou tableaux de contingence (2 variables nominales).....	33
5.3.	Moyennes (variables quantitatives).....	35
5.4.	Observation d'une variable par rapport à une autre.....	36
5.5.	Graphiques.....	39
5.5.1.	Histogramme	39
5.5.2.	Box-plots.....	40
5.5.3.	Graphiques de moyennes	47
5.6.	Valeurs extrêmes ou aberrantes	52
5.6.1.	Détection par boxplot.....	52
5.6.2.	Détection par histogramme	53
5.6.3.	Détection par tableau	53
5.6.4.	Elimination des analyses	55
5.7.	Examen de la normalité	57
5.7.1.	À l'aide du boxplot	57
5.7.2.	À l'aide du test de Kolmogorov-Smirnov	58
5.8.	Transformations mathématiques.....	59
6.	Statistiques inférentielles	61
6.1.	Tableau de décision statistique.....	61

6.2.	Khi-carré.....	62
6.2.1.	Deux variables nominales	62
6.2.2.	Une variable nominale : Khi-carré d'ajustement.....	65
6.3.	T-test	67
6.3.1.	Une moyenne empirique vs. une moyenne théorique.....	67
6.3.2.	T-test à groupes indépendants.....	69
6.3.3.	T-test à mesures répétées.....	71
6.4.	ANOVA.....	73
6.4.1.	ANOVA à un facteur intersujets (groupes indépendants).....	73
6.4.2.	ANOVA à plusieurs facteurs intersujets (groupes indépendants)	75
6.4.3.	ANOVA à mesures répétées (+ ANOVA mixte).....	79
6.4.4.	Comparaisons <i>a priori</i> (contrastes) et <i>a posteriori</i> (<i>post hoc</i>).....	84
6.4.5.	MANOVA.....	90
6.5.	Tests non paramétriques	93
6.5.1.	<i>U</i> de Mann-Whitney : 2 groupes indépendants	93
6.5.2.	Kruskal-Wallis : <i>n</i> groupes indépendants	95
6.5.3.	Wilcoxon : 2 groupes appariés	97
6.5.4.	ANOVA de Friedman : <i>n</i> groupes appariés	99
6.6.	Corrélation.....	100
6.7.	Alpha de Cronbach	102
6.8.	Régression	106
6.9.	Analyse en Composantes Principales (ACP).....	113

Merci à Elena Bayboerek pour sa relecture attentive !

1. Constitution d'une base de données SPSS

1.1. Lancer SPSS (PC ou MAC)

Sous PC :

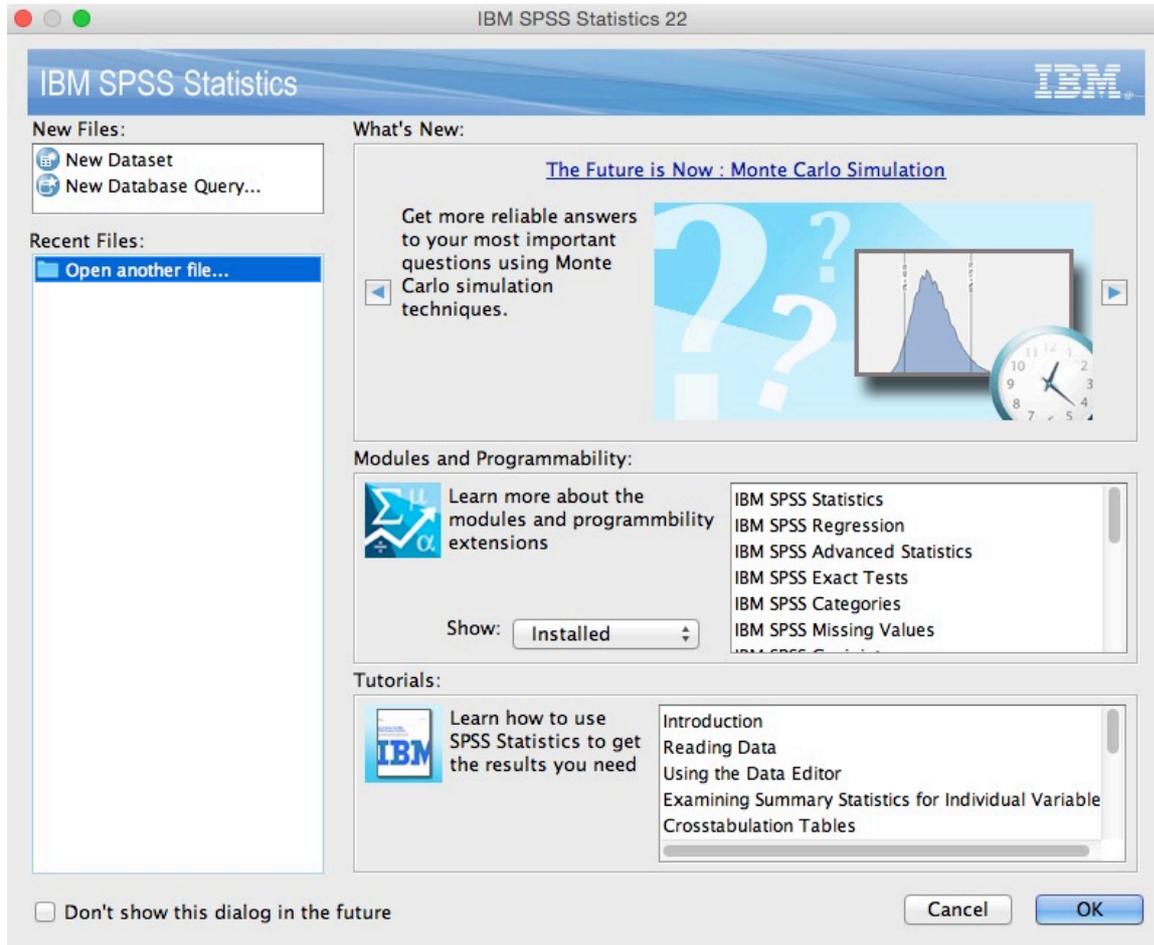
- 1) Aller dans le menu **Démarrer** puis dans **Programmes**, choisir SPSS parmi la liste des programmes.

Sous Mac :

Dans le menu **Applications**, vous trouverez le dossier SPSS et il suffit de cliquer sur l'icône de celui-ci.

1.1.1. Options de lancement de SPSS

Lorsque l'application est lancée, deux fenêtres s'ouvrent. En premier plan la fenêtre d'options de lancement de SPSS et en arrière-plan la fenêtre principale des données.



Cette fenêtre vous propose différentes options.

La plupart du temps, on choisira :

1) L'option **New Dataset** pour faire la saisie des données récoltées à l'aide des questionnaires.

ou

2) L'option **Open another file...** pour ouvrir un fichier déjà existant.

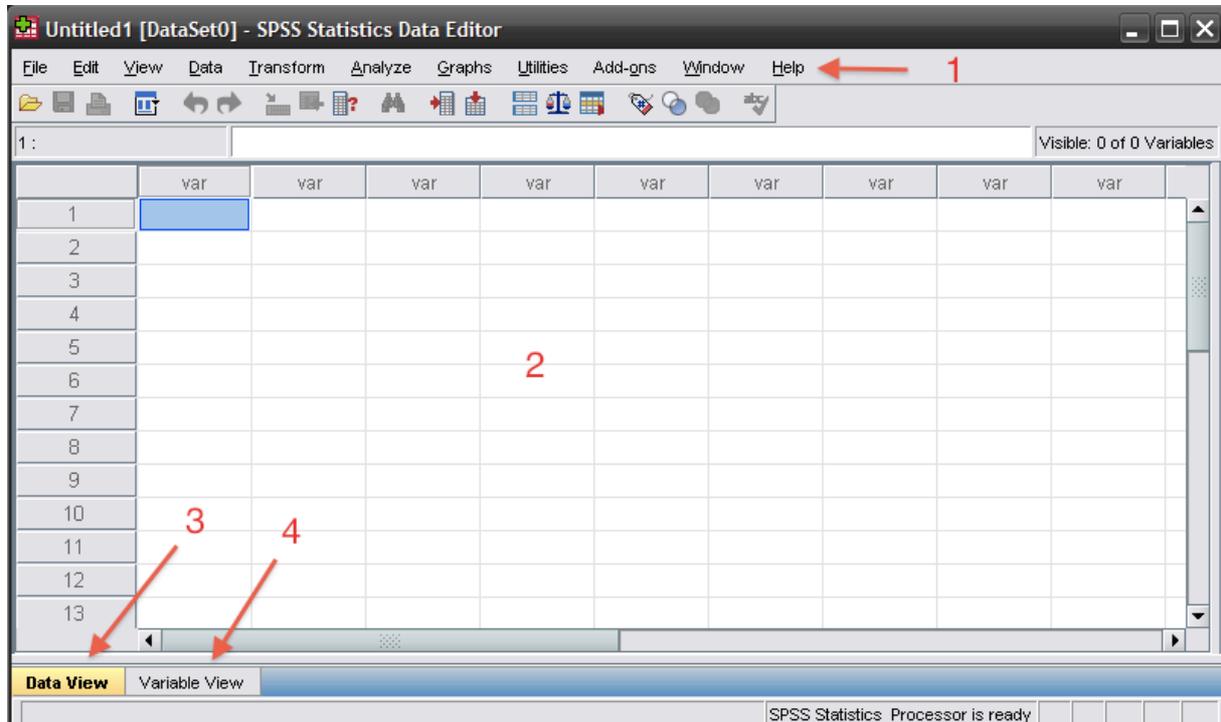
Lorsque vous sélectionnez l'option **Type in data**, vous voyez apparaître au premier plan la fenêtre qui se trouvait en arrière plan (la fenêtre des données) et qui représente la fenêtre principale de SPSS.

Lorsque vous ouvrez une **base de données déjà existante**, il se peut qu'une autre fenêtre s'ouvre automatiquement : l'**Output** (fenêtre d'édition des résultats). Pour l'instant nous ne nous occuperons que de la fenêtre principale.

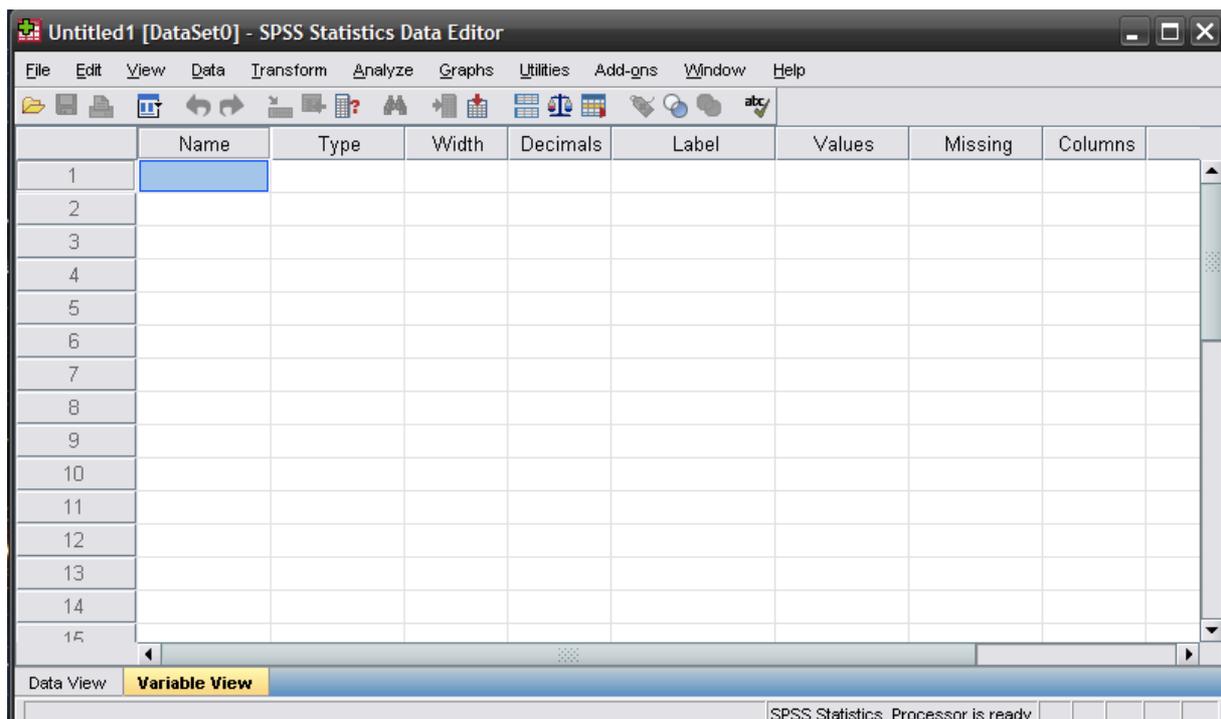
On peut également fermer cette fenêtre (**Cancel**) et travailler directement dans les menus du programme (pour un nouveau fichier : **File** → **New** → **Data** et pour ouvrir des données déjà existantes : **File** → **Open** → **Data**).

1.2. Fenêtre principale

Elle se compose de plusieurs parties :



- 1) La barre des menus et des boutons de commande
- 2) La fenêtre principale de SPSS pour l'entrée et le traitement des données
- 3) Ici nous sommes sur la fenêtre « **Data View** », c'est-à-dire la fenêtre des données
- 4) Si vous cliquez sur « **Variable View** », vous vous trouverez alors sur la fenêtre de résumé des variables :



Pour la saisie des données, il faut dans un premier temps définir le nom des variables que l'on utilise.

1.3. Définition d'une variable et de ses propriétés

Cette opération s'effectue dans la fenêtre **Variable View**. Vous pouvez passer d'une fenêtre à l'autre en cliquant sur les onglets correspondants dans la barre en bas à gauche de la fenêtre.

Sous la colonne **Name** on indique le nom de la variable (nom sans espaces et sans accents), par exemple **age** pour l'âge des participant-e-s. On appuie ensuite sur **Enter** pour valider cette entrée. Des propriétés par défaut s'inscrivent alors sur la ligne qui concerne cette variable :

	Name	Type	Width	Decimals	Label	Values	Missing	(
1	sujet	Numeric	8	2		None	None	8
2	sexe	Numeric	8	2	sexe	{1.00, homme}...	None	8
3	national	Numeric	8	2	national	{1.00, suisse}...	None	8
4	langue	Numeric	8	2	langue	{1.00, français}...	None	8
5	civil	Numeric	8	2	état civil	{1.00, célibataire}...	None	8
6	pb_med	Numeric	8	2	Pbs médicaux	None	97.00 - 99.00	8
7	pb_sco	Numeric	8	2	Pbs scolaires	None	97.00 - 99.00	8
8	pb_soc	Numeric	8	2	Pbs sociaux	None	97.00 - 99.00	8
9	pb_fam	Numeric	8	2	Pbs familiaux	None	97.00 - 99.00	8
10	pb_leg	Numeric	8	2	Pbs légaux	None	97.00 - 99.00	8
11	Conso_Dro	Numeric	8	2	Gravité conso. ...	None	97.00 - 99.00	8
12	Conso_OH	Numeric	8	2	Gravité conso. ...	None	97.00 - 99.00	8
13	psy	Numeric	8	2		{1.00, L}...	None	8
14								
15								

1) **Type** (par défaut, SPSS affiche **Numeric**) :

Variable Type

Numeric Width: 8

Comma Decimal Places: 2

Dot

Scientific notation

Date

Dollar

Custom currency

String

OK Cancel Help

On obtient la fenêtre **Variable Type** en double cliquant sur le rectangle gris figurant dans la case **type** de chaque variable. Ici il s'agit d'indiquer quelle est la nature de la variable pour laquelle le type est défini.

Si vous entrez des chiffres ayant une valeur numérique (**variable quantitative**) → cliquez sur **Numeric**. Si vous entrez une **variable nominale avec un nombre de modalités fini ou peu élevé** (sexe, condition, filière d'étude, etc.), gardez le type de votre variable **Numeric** et attribuez dans la case **Values** un nombre (c'est égal lequel) à chaque modalité, comme dans l'image ci-dessous 1.00 homme, 2.00 femme.

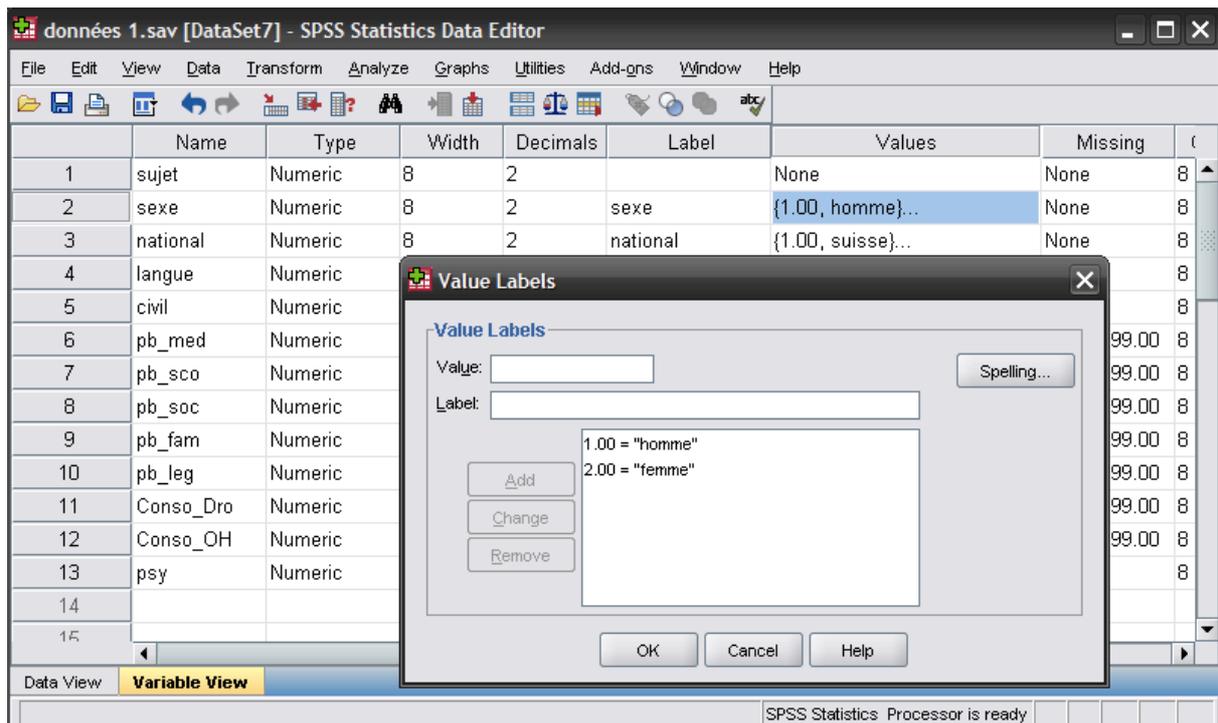
Si votre **variable nominale a beaucoup de modalités**, vous pouvez entrer des chiffres sans valeur numérique, des lettres, des mots avec le type **String**.

Vous définissez également le nombre de caractères que vous pouvez inscrire (pour des variables nominales) et le nombre de décimales (pour les variables quantitatives).

Ces deux informations vont également s'inscrire dans les colonnes **Width** et **Decimals**. **Width** correspond à l'étendue de la case (le nombre de lettres ou de chiffres que vous pouvez inscrire). Par défaut, ces valeurs sont 8 et 2 respectivement. Vous pouvez les modifier à tout moment en changeant les chiffres qui apparaissent dans les cases correspondantes.

2) **Label** : le label correspond au descriptif de la variable. C'est un aide-mémoire qui peut être utile pour y inscrire des informations supplémentaires sur la variable, telle que la question entière (dans le cas d'un questionnaire). Le label est entré dans la case après avoir double cliqué dessus. Vous avez à disposition autant de caractères que vous le désirez. Pour agrandir physiquement cette case, sélectionnez la limite droite, gardez sélectionné et déplacez le curseur à droite ou gauche selon que vous voulez agrandir ou rétrécir la case.

3) **Values** : cette option est surtout utilisée pour les **variables nominales avec un nombre fini ou peu élevé de modalités**. Par exemple pour la variable **sexe**, on va attribuer le code 1 aux hommes et 2 aux femmes (ou l'inverse, peu importe). Autre exemple, pour une question posée, on notera les réponses selon une échelle allant de 1 à 6 pour laquelle 1 correspondra à « pas d'accord » et 6 correspondra à « tout à fait d'accord ». Les codages sont entrés sous **Values** afin de pouvoir lire plus facilement les tableaux de résultats qui autrement ne vous donneraient que les valeurs numériques (p.ex. 1, 2, ...).



Pour entrer une valeur il faut cliquer sur le rectangle gris qui apparaît dans la case **Values** une fois celle-ci sélectionnée. La fenêtre ci-dessus s'affiche alors. Il suffit alors d'entrer dans l'onglet **Value** le premier code (par exemple 1 pour la modalité homme de la variable sexe) et dans **Label** le nom de la modalité correspondante (par exemple homme). On clique alors sur **Add** pour valider la manœuvre. Attention, si vous ne cliquez pas sur **Add**, votre label ne sera pas pris en compte. Vous pouvez modifier labels et valeurs à tout moment. Pour tout nouveau label ajouté, répétez la procédure précédente. Pour enlever un label, cliquez sur la ligne que vous voulez supprimer. L'onglet **Remove** apparaît alors. Cliquez dessus. Pour modifier un label, cliquez sur la valeur ou le label à modifier, effectuez les modifications dans les deux rectangles supérieurs (**Value** et **Label**) et au lieu de cliquer sur **Add**, cliquez sur **Change**.

Pour une variable continue (par exemple la variable *politique* qui va de 1 = extrême gauche à 8 = extrême droite), vous pouvez n'indiquer que le code correspondant à extrême gauche et le code correspondant à extrême droite sans définir toutes les valeurs de l'échelle (cela vous permet de vous rappeler les bornes de l'échelle) :



NB : les valeurs et labels indiqués dans cette fenêtre ne sont qu'indicatifs ! Modifier les labels et valeurs ne modifiera pas les données entrées dans la fenêtre Data View ! Il s'agit d'un aide-mémoire.

4) **Missing** : pour indiquer une non réponse (ou valeur manquante). Cela ne s'utilise plus guère, lorsque vous remplissez votre base de données, si des participant-e-s n'ont pas répondu à certaines questions, vous pouvez laisser les cases correspondantes vides.

5) **Columns** : permet de contrôler la largeur physique de la case de la variable. Laissez 8 par défaut.

6) **Align** : permet de choisir si le texte dans les colonnes sera aligné à droite, à gauche ou sera centré. Laissez à droite par défaut.

7) **Measure** : permet d'indiquer si la variable que vous définissez est nominale (**Nominal**), ordinale (**Ordinal**) ou quantitative (**Scale**). Laissez **Scale** par défaut. Le type d'analyse que vous allez demander sur une variable va définir implicitement le type de variable que vous utilisez (si vous effectuez un Khi-carré, SPSS présume que les variables sont nominales). Cela ne sert pas à grande chose de les redéfinir ici.

Si plusieurs colonnes doivent avoir exactement les mêmes caractéristiques, c'est-à-dire le même type, avec le même label, le même format de colonne, les mêmes **Values**, etc., utilisez dans **Variable View** les fonctions **Copy** et **Paste** du menu **Edit** (ou les touches copier-coller). Le principe est le même que dans Word : vous allez sur la case dont vous voulez copier le contenu, vous le copiez, vous allez sur la case dans laquelle vous désirez copier le contenu, et vous collez le contenu de la première case. Attention : vous devez avoir nommé une variable (la case à l'extrême gauche sur **Variable View**) pour copier les autres cases, car la variable est « créée » seulement après avoir été nommée. Au fur et à mesure que vous définissez les variables et leurs propriétés dans **Variable View**, la feuille des données (**Data View**) se forme automatiquement :

données 1.sav [DataSet7] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : sujet Visible: 13 of 13 Variables

	sujet	sexe	national	langue	civil	pb_med	pb_sco	pb_soc	pt
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									

Data View Variable View

SPSS Statistics Processor is ready

Les noms des variables s'affichent en haut des colonnes et chaque ligne représente un-e participant-e.

Pour résumer:

- 1) Dans **Data View** : une ligne = un-e participant-e et une colonne = une variable.
- 2) Dans **Variable View** : une ligne = une variable et une colonne = les caractéristiques de cette variable.

Pour rentrer les données il faut donc remplir en ligne les informations correspondantes à chaque participant-e dans **Data View** :

données 1.sav [DataSet7] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : sujet Visible: 13 of 13 Variables

	sujet	sexe	national	langue	civil	pb_med	pb_sco	pb_soc	pt
1	1.0	1.00	1.00	1.00	1.00	7.00	8.00	5.00	
2	2.00	1.00	1.00	1.00	1.00	0.00	3.00	1.00	
3	3.00	1.00	1.00	1.00	1.00	2.00	6.00	6.00	
4	4.00	2.00	1.00	1.00	1.00	1.00	5.00	3.00	
5	5.00	2.00	1.00	1.00	1.00	5.00	7.00	7.00	
6	6.00	2.00	1.00	1.00	1.00	8.00	6.00	6.00	
7	7.00	2.00	45.00	2.00	1.00	4.00	7.00	6.00	
8	8.00	2.00	1.00	1.00	1.00	2.00	9.00	8.00	
9	9.00	2.00	1.00	1.00	1.00	7.00	4.00	7.00	
10	10.00	2.00	1.00	1.00	1.00	6.00	2.00	3.00	
11	11.00	1.00	21.00	1.00	1.00	2.00	7.00	9.00	
12	12.00	1.00	1.00	1.00	1.00	7.00	3.00	2.00	
13	13.00	2.00	1.00	1.00	1.00	4.00	8.00	9.00	

Data View Variable View

SPSS Statistics Processor is ready

1.4. Enregistrement des données

La première fois que vous enregistrez vos données (même principe que dans Word) : On enregistre les données en exécutant l'option **Save As** du menu **File**. Dans un premier temps, on choisit l'emplacement où l'on veut enregistrer le fichier (**Look in** :). Dans un deuxième temps, on entre un nom dans la fenêtre **File name** ; ensuite on clique sur le bouton **Save**. L'extension d'un fichier de données SPSS est « **.sav** » et cette extension s'inscrit automatiquement à la suite du nom de votre fichier. Si vous voyez un fichier du type « nom.sav » sachez que c'est un fichier de données SPSS. Il est important de vérifier de bien avoir enregistré le fichier « **.sav** » depuis la fenêtre du fichier de données SPSS, et non seulement le fichier de résultats (Output) dont l'extension est « **.spv** ».

Pour ajouter des données à un fichier déjà existant (même principe que dans Word) : lorsque votre document est ouvert, cliquez simplement sur **Save** (pas **Save as**).

1.5. Quitter SPSS

Sur PC : On exécute l'option **Exit** du menu **File** pour quitter SPSS. Le logiciel vous demande si vous voulez vraiment quitter l'application, cochez **Yes**.

Sur MAC : On exécute l'option **Quit SPSS Statistics** du menu **SPSS Statistics** (qui se trouve à gauche du menu **File**). Le logiciel vous demande si vous voulez sauver les données contenues dans chaque fenêtre avant de quitter, répondez **Yes** pour ne rien perdre.

2. Importation d'un fichier de données à partir d'Excel

2.1. Fichier de données Excel

Le fichier aura comme extension « **.xlsx** ».

Les variables correspondent aux colonnes et les participant-e-s aux lignes, comme dans SPSS.

Le nom des variables doit figurer sur la première ligne de la grille. N'utilisez pas d'espaces, de caractères spéciaux ou d'accents.

	A	B	C	D	E
1	sujet	sexe	national	langue	civil
2	1	1	1	1	1
3	2	1	1	1	1
4	3	1	1	1	1
5	4	2	1	1	1
6	5	2	1	1	1
7	6	2	1	1	1
8	7	2	45	2	1
9	8	2	1	1	1
10	9	2	1	1	1
11	10	2	1	1	1
12	11	1	21	1	1
13	12	1	1	1	1
14	13	2	1	1	1
15	14	2	1	1	1

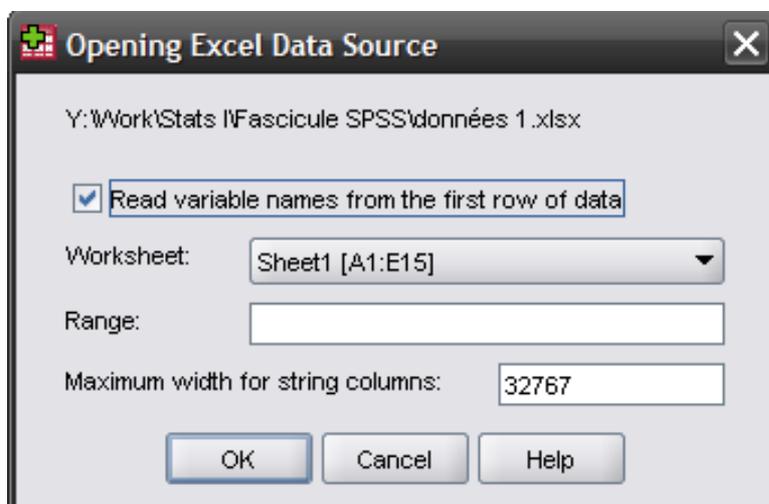
2.2. Importation des données

Dans SPSS, il faut suivre les étapes suivantes:

Depuis la fenêtre d'options de lancement (cf. paragraphe 1.1.1) lancer de SPSS (première fenêtre qui s'ouvre), spécifiez que vous désirez ouvrir un fichier déjà existant (**Open another file...**). Indiquez dans la fenêtre **Open File** que le fichier à importer (**File of type:**) est de type **Excel (.xls)**. Vous trouvez dans vos répertoires votre fichier Excel : sélectionnez-le et cliquez sur **Open**.

Depuis SPSS directement (si vous n'avez plus la fenêtre d'options de lancement), passez par les menus **File → Open → Data**.

Cochez la case **Read variable names from the first row of data** dans la fenêtre **Opening Excel Data Source** et cliquez sur **OK**.



Cette fonction permet au logiciel de reconnaître la première ligne de votre feuille Excel comme étant le nom des variables. Attention les caractéristiques de vos variables sont définies par défaut par SPSS, puisque le logiciel ne dispose d'aucune information sur les données depuis Excel. Il faudra peut-être changer cela manuellement.

Le fichier de données s'ouvre alors dans SPSS et vous pouvez utiliser les fonctionnalités de ce dernier. N'oubliez pas de sauvegarder le nouveau fichier.

3. Fonctions de base de SPSS

Le logiciel SPSS travaille sur plusieurs feuilles (fenêtres) en même temps :

1) La feuille de données sur laquelle vous entrez vos données **Dataset**. Cette feuille a une extension **.sav** (elle comprend la page **Data View** et la page **Variable View**). Vous pouvez lui donner un nom pour la sauver, sinon le nom par défaut reste (**Untitled1** p.ex.).

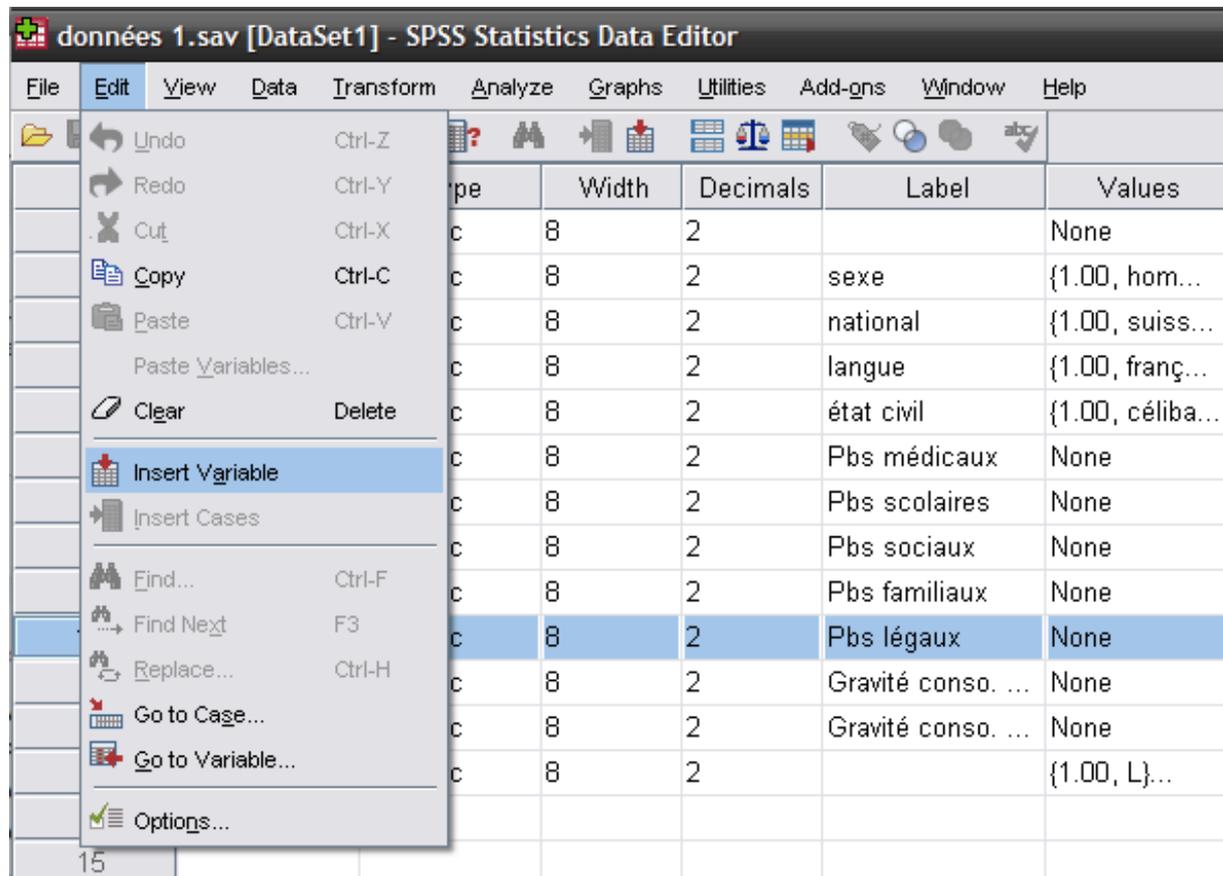
2) La (les) feuille(s) de sortie ou **Output** sur laquelle apparaissent les résultats de vos analyses statistiques. Cette feuille a une extension **.spv**. Vous pouvez ouvrir plusieurs fenêtres de ce type en même temps. Sur ces fenêtres, le logiciel présentera les résultats de vos analyses sous forme de tableaux ou de graphiques.

3) La (les) feuille(s) de syntaxe (**Syntax**) sur laquelle vous pouvez programmer des analyses sans utiliser les menus. Cette feuille a une extension **.sps**. Vous pouvez ouvrir plusieurs fenêtres de ce type en même temps. Sur ces fenêtres, le logiciel vous présentera les ordres que vous avez soit écrit manuellement en « langage SPSS » (programmation), ou alors copié puis collé dans la feuille de syntaxe depuis les fenêtres d'analyses statistiques de la feuille de données. La feuille de syntaxe vous permet de sauvegarder les analyses sans devoir passer par l'output (qui est un fichier « lourd », du fait des tableaux et des graphiques). De plus, sauvegarder la syntaxe peut être utile pour garder une trace des nouvelles variables que vous avez définies lors de vos analyses (voir chapitre 4), ou alors pour refaire une série d'analyse déjà faites sur de nouvelles données (ou avec des données supplémentaires).

Que vous soyez sur ces trois fenêtres différentes ne change rien à la barre de menu, vous pouvez lancer toutes les analyses souhaitées depuis n'importe quelle fenêtre.

3.1. Insérer/déplacer une variable

Pour insérer une nouvelle variable dans une base de données existante, sélectionnez une ligne dans **Variable View** (ou une colonne dans **Data View**), puis suivez le chemin suivant : **Edit** → **Insert Variable** (ou **Clic droit** → **Insert Variable**).



Apparaît alors une nouvelle ligne dans **Variable View** ou une colonne vide dans **Data View**, précédant celle que vous avez sélectionnée, prédéfinie par défaut, mais sans nom.

données 1.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values
1	sujet	Numeric	8	2		None
2	sexe	Numeric	8	2	sexe	{1.00, hom...
3	national	Numeric	8	2	national	{1.00, suiss...
4	langue	Numeric	8	2	langue	{1.00, franç...
5	civil	Numeric	8	2	état civil	{1.00, céliba...
6	pb_med	Numeric	8	2	Pbs médicaux	None
7	pb_sco	Numeric	8	2	Pbs scolaires	None
8	pb_soc	Numeric	8	2	Pbs sociaux	None
9	pb_fam	Numeric	8	2	Pbs familiaux	None
10	VAR00001	Numeric	8	2		None
11	pb_leg	Numeric	8	2	Pbs légaux	None
12	Conso_Dro	Numeric	8	2	Gravité conso. ...	None
13	Conso_OH	Numeric	8	2	Gravité conso. ...	None
14	psy	Numeric	8	2		{1.00, L}...

Pour déplacer une variable dans la base de données, insérez une ligne comme ci-dessus à l'endroit voulu, puis sélectionnez la ligne (variable) à déplacer. Suivez alors le chemin suivant : **Edit** → **Cut** (ou **Clic droit** → **Couper**). Puis sélectionnez la colonne vide et **Edit** → **Paste** (ou **Clic droit** → **Coller**).

Il est aussi possible de cliquer sur le numéro de la variable (par exemple ci-dessous, dans le cas de la variable « psy », cliquez sur le 14) et de tirer la variable vers le haut ou vers le bas. On voit son emplacement futur à la ligne rouge.

données 1.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values
1	sujet	Numeric	8	2		None
2	sexe	Numeric	8	2	sexe	{1.00, hom...
3	national	Numeric	8	2	national	{1.00, suiss...
4	langue	Numeric	8	2	langue	{1.00, franç...
5	civil	Numeric	8	2	état civil	{1.00, céliba...
6	pb_med	Numeric	8	2	Pbs médicaux	None
7	pb_sco	Numeric	8	2	Pbs scolaires	None
8	pb_soc	Numeric	8	2	Pbs sociaux	None
9	pb_fam	Numeric	8	2	Pbs familiaux	None
10	VAR00001	Numeric	8	2		None
11	pb_leg	Numeric	8	2	Pbs légaux	None
12	Conso_Dro	Numeric	8	2	Gravité conso. ...	None
13	Conso_OH	Numeric	8	2	Gravité conso. ...	None
14	psy	Numeric	8	2		{1.00, L}...
15						

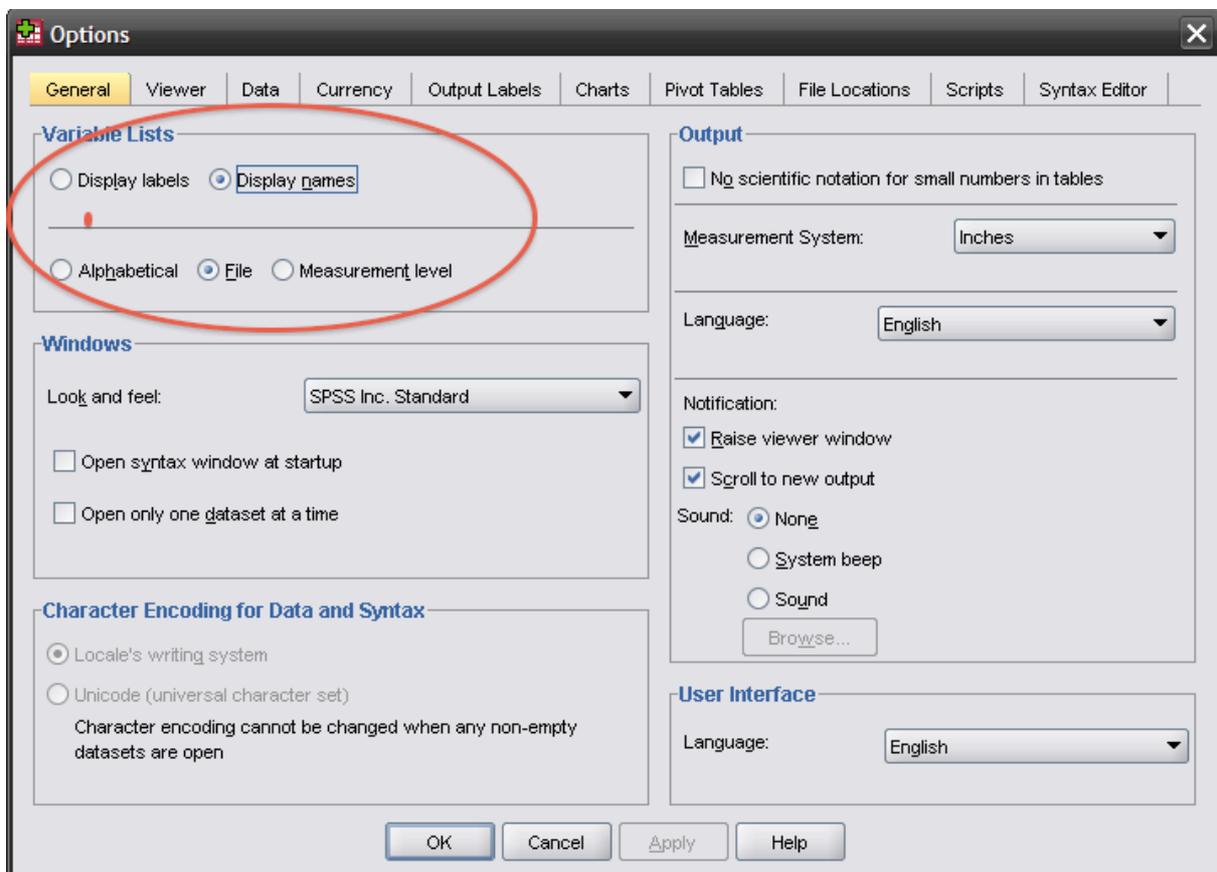
3.2. Options d'affichage des variables

Lorsque vous entrez vos variables, particulièrement dans le cas d'un questionnaire comportant plusieurs parties distinctes, il n'est pas très pratique de nommer chaque variable en fonction de la question à laquelle elle se réfère. Il est beaucoup plus pratique d'utiliser le nom de la partie de laquelle la question fait partie (par ex. « Estimeso1 », « Estimeso2 »...; « Paranoia1 », « Paranoia2 », ...) et d'inscrire plus d'informations dans le **Label** de la variable. Ainsi, lorsque vous voudrez regarder les résultats d'une seule variable, vous pourrez la situer d'un coup d'œil dans votre questionnaire et vos catégories, sans avoir besoin de regarder dans la structure du questionnaire.

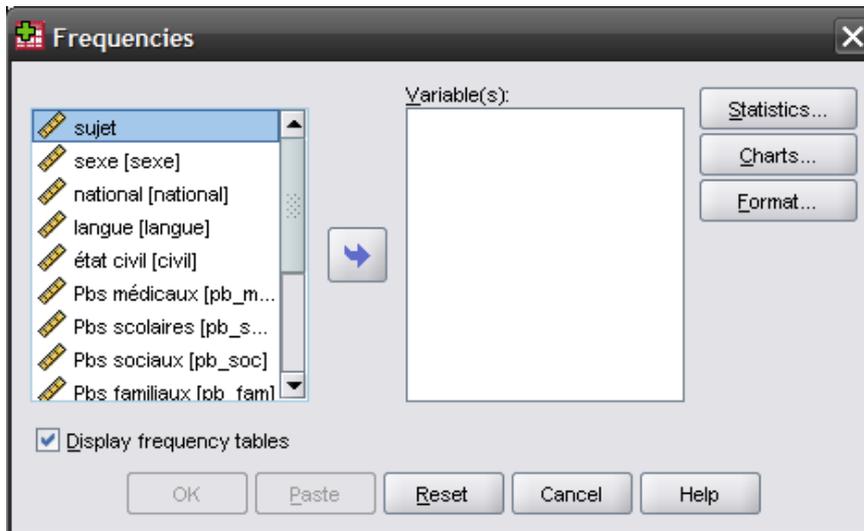
Vous pouvez inscrire un nombre infini de caractère dans le Label, il est donc possible d'y copier entièrement la question (par ex. « Je me sens inutile » ; « Les gens parlent dans mon dos » etc...) ou d'écrire des mots clés (par ex. « inutile » ; « gens parlent » etc...). Cela dépendra de vos goûts personnels et aussi du temps que vous voulez/pouvez passer à définir vos variables.

À noter que ces options ne servent qu'au confort d'utilisation des données et ne modifient en rien les analyses effectuées.

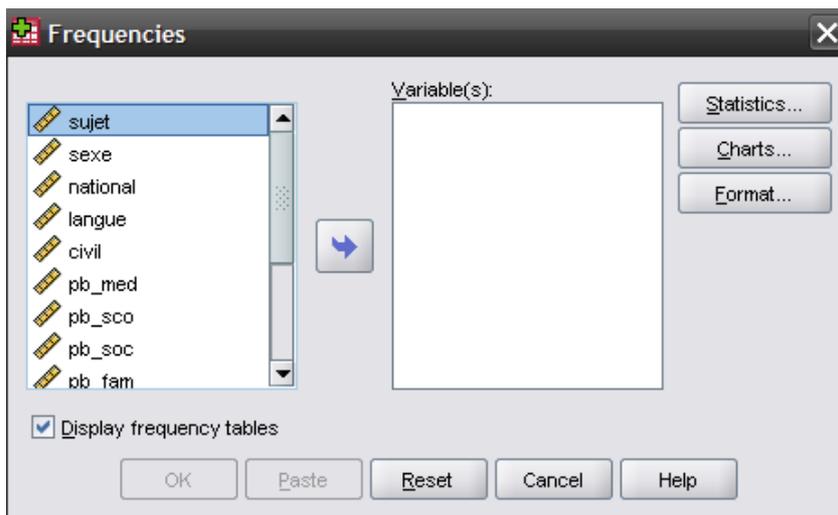
Dans **Edit** → **Options** (ou **SPSS Statistics** → **Preferences**), puis dans l'onglet **General**, il est possible de modifier les options d'affichage des variables dans les boîtes de dialogue SPSS.



Si vous choisissez l'option **Display labels**, vous obtiendrez les labels des variables, suivis de leurs noms :

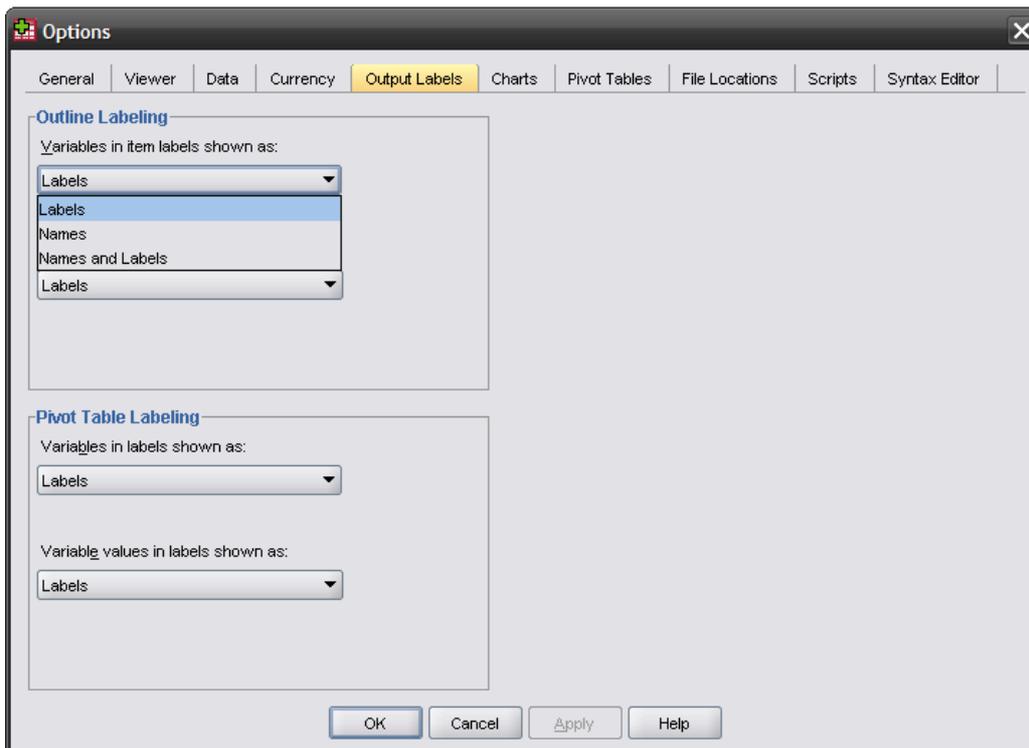


Si vous choisissez l'option **Display names**, vous n'obtiendrez que leurs noms :



L'option à choisir dépendra de vos préférences personnelles et aussi de la taille des informations que vous avez mises dans la colonne pour chaque variable. Par exemple, il sera difficile de différencier vos variables si elles commencent pour la plupart par « Est-ce que... » (dans le cas où vous avez copié les questions complètes dans les labels).

Dans **Edit** → **Options** (ou **SPSS Statistics** → **Preferences**), puis dans l'onglet **Output**, il vous est possible de modifier la façon dont les variables sont présentées dans les outputs, particulièrement dans les tableaux de résultats.

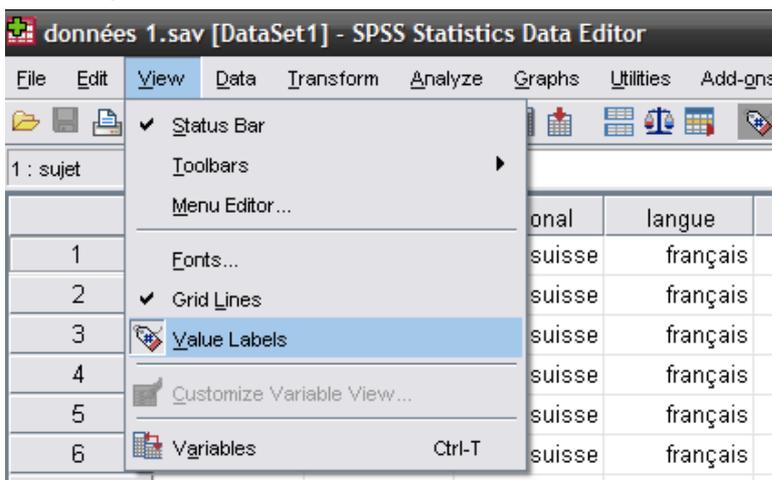


Vous pouvez choisir de n'afficher que le nom, que le label ou les deux dans la présentation des variables (**Outline Labeling**) et dans les tableaux (**Pivot Table Labeling**).

3.3. Option d'affichage des données

Dans **Data View**, pour les variables nominales, vous pouvez soit afficher les valeurs numériques (ex. sexe, 1 et 2) ou les noms que vous avez attribués à ces valeurs (ex. sexe, homme et femme).

Cette option se trouve dans **View → Value Labels**.



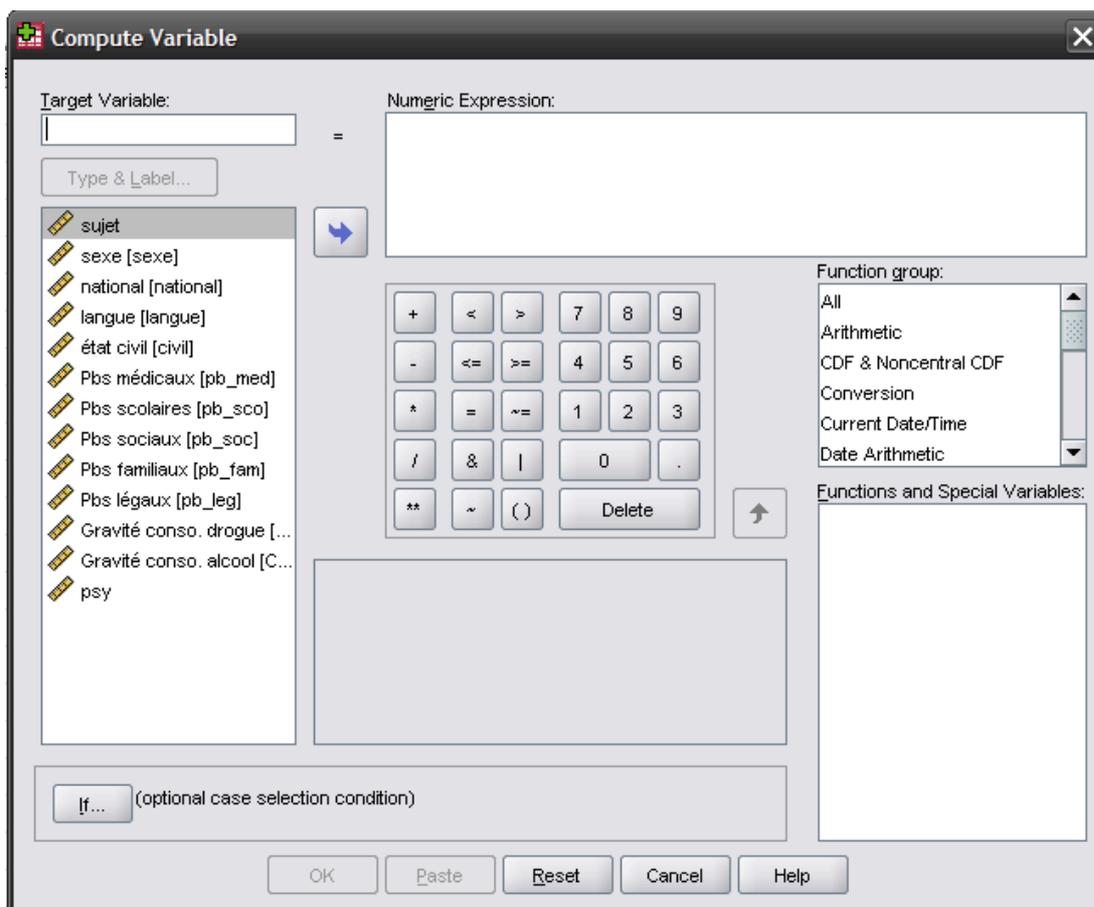
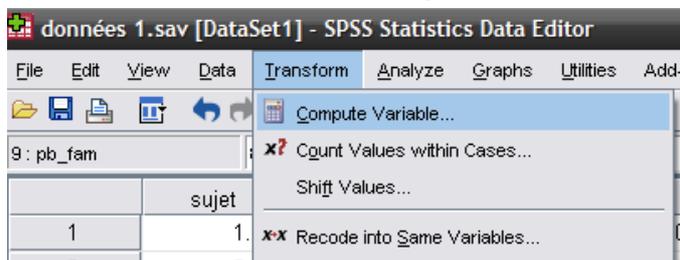
Comme vous pouvez le voir ci-dessus, les noms attribués aux valeurs sont activés, c'est pourquoi il est noté « français » au lieu de « 1 » pour la variable « langue ».

4. Manipulation des données

4.1. Créer une variable à partir d'une ou de plusieurs variable(s) existante(s)

But : créer une nouvelle variable à partir de variables existantes, soit en les modifiant (par addition, soustraction, etc.), soit en les « agglomérant » (moyenne d'un ensemble de variables). Normalement, ces opérations sont effectuées sur des variables quantitatives.

Chemin : **Transform** → **Compute Variable...**

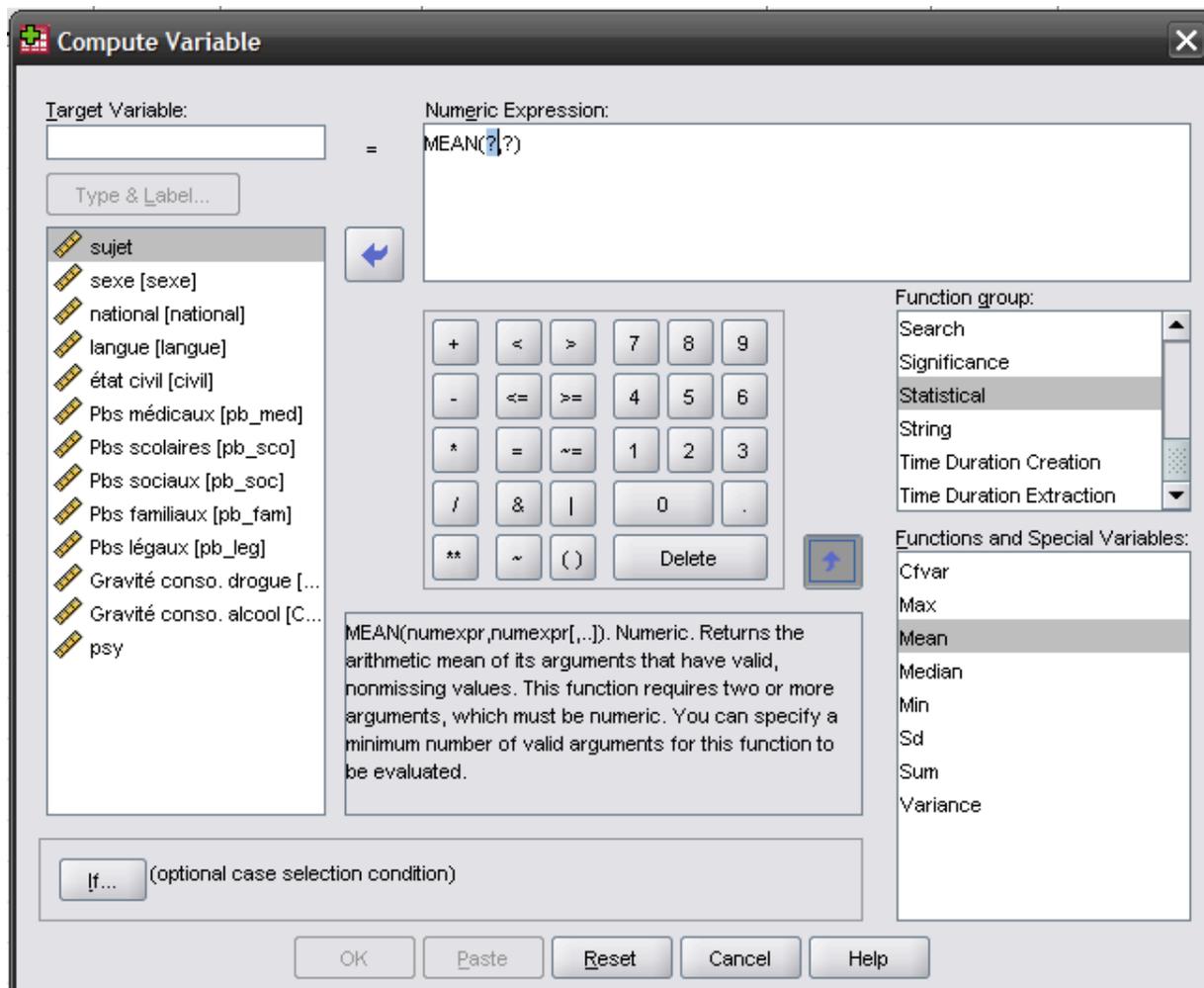


Exemple : à l'aide d'entretiens, vous avez déterminé le degré de problèmes (médicaux, scolaires, sociaux, familiaux et légaux) d'adolescent-e-s, sur une échelle

à 10 niveaux. Vous souhaitez maintenant faire le score moyen de problèmes de vos sujets.

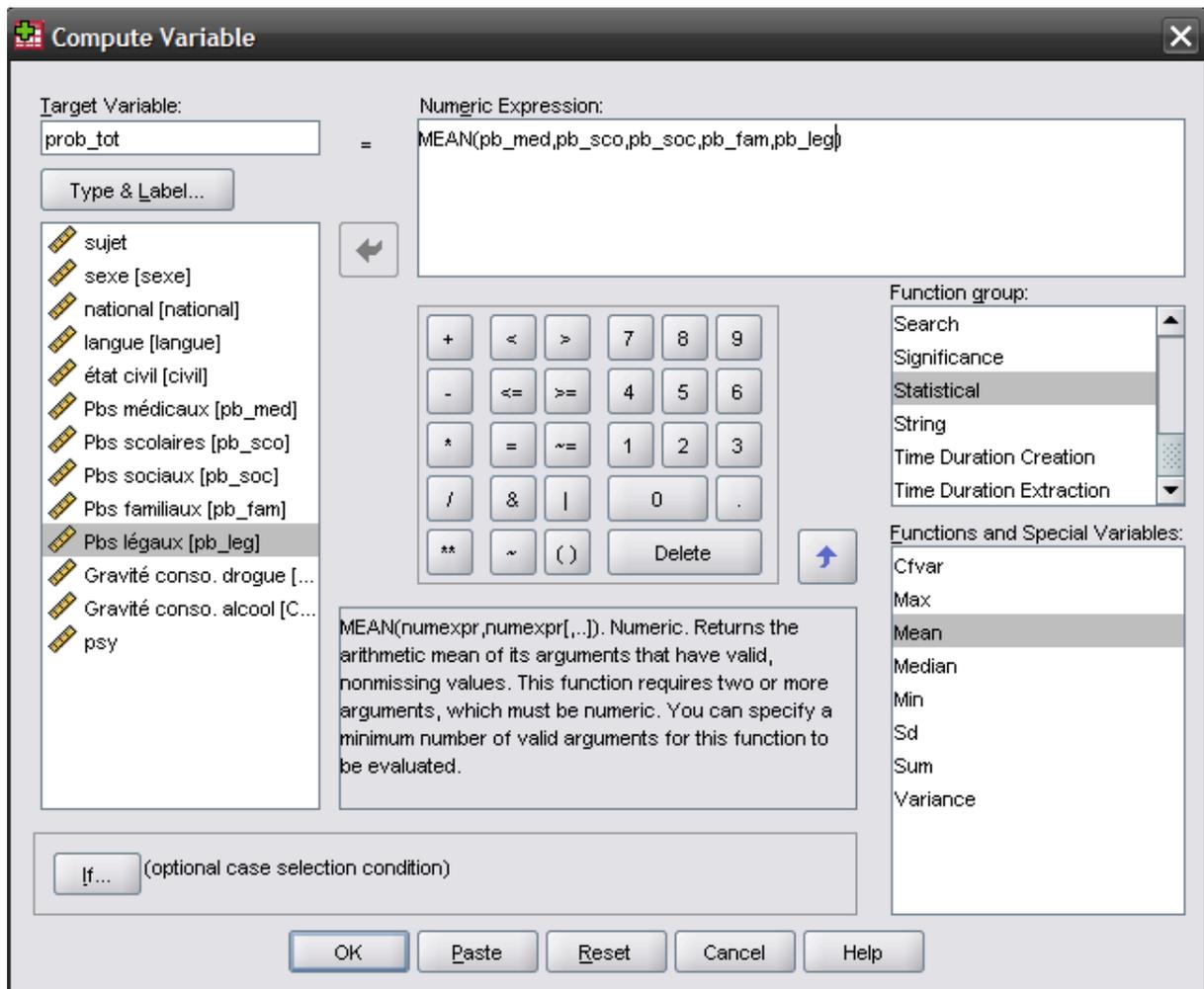
Dans **Target Variable** :, inscrivez le nom de la nouvelle variable que vous voulez créer (par exemple « probtot » ou « prob_tot »), sans espaces et sans accents.

Dans **Numeric Expression** : inscrivez la formule correspondante au calcul qui vous permet d'obtenir votre nouvelle variable. Soit vous pouvez écrire la formule vous-même à l'aide du pavé numérique, ou vous pouvez chercher la fonction correspondante en cliquant **All** dans l'onglet **Fonction group** et la fonction recherchée dans **Functions and Special Variables**, puis soit double cliquer sur la fonction que vous souhaitez utiliser, ou à l'aide de la flèche montante la faire passer dans la case **Numeric Expression**. Dans notre exemple, nous allons utiliser la moyenne, que l'on trouve sous **Statistical** (ou **All**) → **Mean**.



Pour entrer les variables, vous avez deux possibilités : soit vous inscrivez vous-mêmes le nom des variables dans la case **Numeric Expression**:, soit pour éviter les risques d'erreur vous pouvez les chercher dans la case de gauche. Avec la flèche, vous les déplacez dans la case **Numeric Expression**: après les avoir sélectionnées (une à la fois).

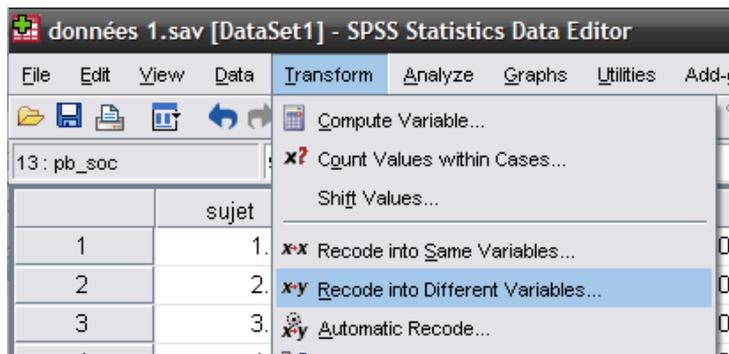
Dans notre exemple, pour calculer la moyenne, il faut remplacer les « ? » par les variables et les séparer par des virgules.



Une fois que votre nouvelle variable est correctement paramétrée, cliquez sur **OK**.
La nouvelle variable ainsi créée viendra s'ajouter à la fin de la base de données, comme dernière variable.

4.2. Recoder des variables

But : Recoder différemment les valeurs que vous avez donné à des catégories de réponse. Cela signifie que vous allez manipuler les étiquettes que vous avez déterminées pour vos données. Cette fonction vous permet aussi de créer des groupes : vous allez regrouper différemment des catégories de réponse pour en créer d'autres. Souvent, ce genre de manipulation porte sur des variables nominales, mais cela peut également concerner des variables quantitatives. Par exemple, si vous voulez former des classes d'âge (variable ordinale) à partir de l'âge des participant-e-s (variable quantitative). Vous devez aussi utiliser cette fonction lorsque vous êtes amenés à inverser l'ordre de l'échelle utilisée pour mesurer une variable (items inversés dans une échelle).

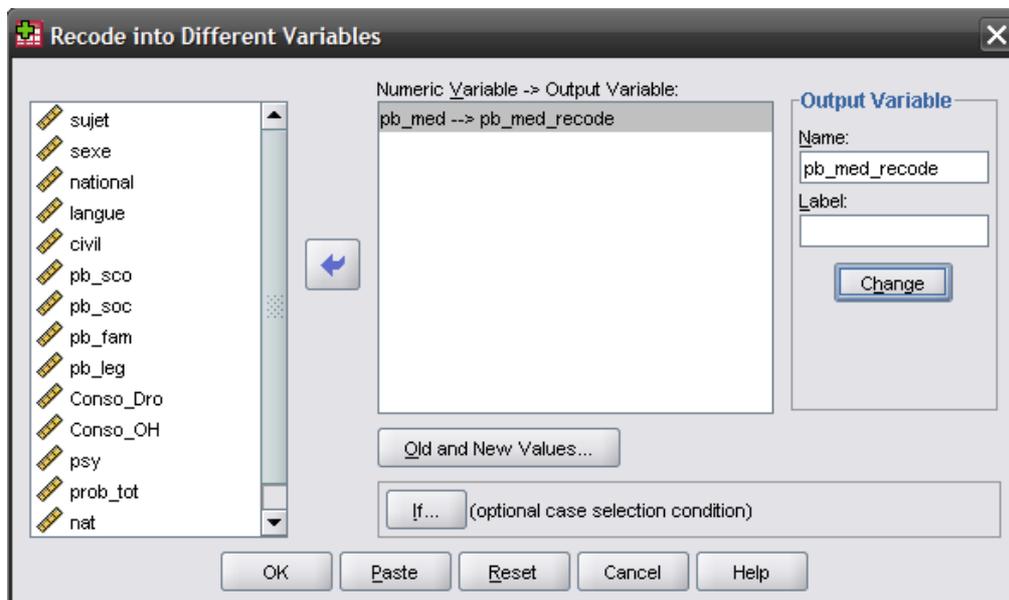


Chemin : **Transform → Recode into Different Variables.**

Sélectionnez toujours **Recode Into Different Variables**. Cela signifie que le logiciel crée une nouvelle variable qui s'ajoutera à la fin de votre base de données (si vous cliquez sur *Recode Into Same Variables* le logiciel modifiera l'ancienne variable sans en modifier le nom. Le problème est que de cette manière, vous ne gardez pas de trace de cette transformation et il devient difficile de se rappeler si une variable est encore sous sa forme originale ou pas. De plus, si vous faites une erreur, vous perdez les données de la variable originale).

Exemple : la variable « problèmes médicaux » (pb_med) est cotée de 0 (aucun problème) à 9 (problèmes très graves). Pour des raisons de facilité d'interprétation, nous souhaitons la recoder de -4 (pas de problème) à +5 (problèmes très graves).

Sélectionnez dans le cadre de gauche le nom de la variable à recoder (« pb_med ») et cliquez sur la flèche. Le nom de cette variable apparaît dans le cadre **Input Variable → Output Variable**. Écrire dans la case de droite, sous **Name**: le nom de la future variable recodée (« pb_med_recode »). Cliquez sur **Change**. Si vous ne cliquez pas dessus, vous ne pourrez pas procéder à la création de la nouvelle variable. Un conseil : lorsque vous nommez la nouvelle variable, reprenez le nom de l'ancienne en lui rajoutant quelque chose qui vous rappelle qu'elle a été recodée. Par exemple, l'ancienne variable s'appelle v30 et la nouvelle v30rec.

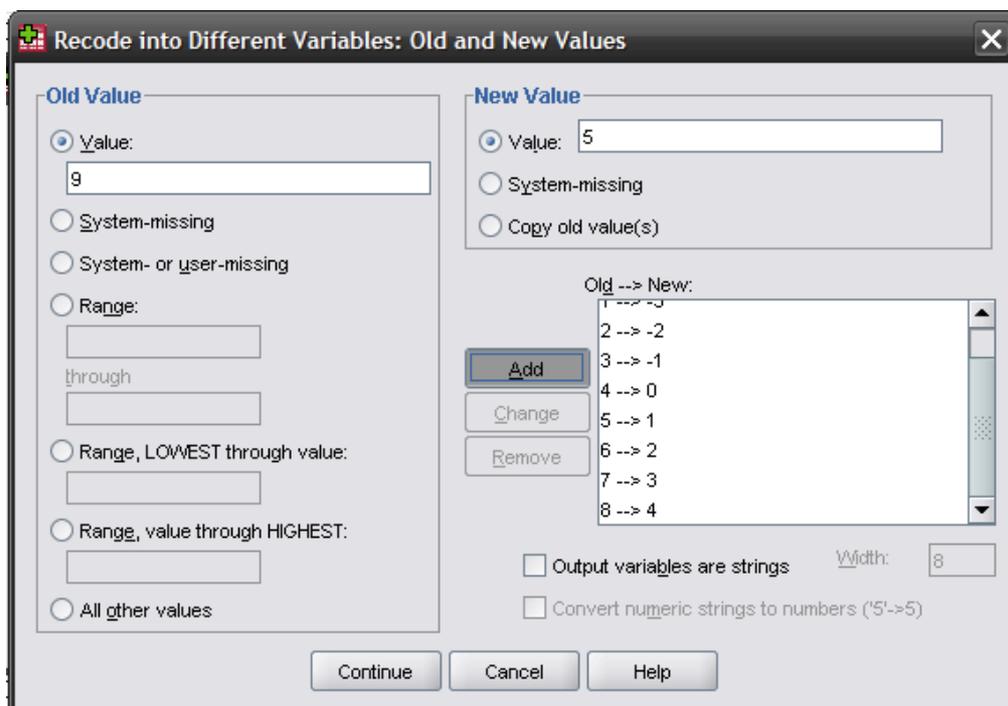


Cliquez ensuite sur **Old and New Values...** Une nouvelle boîte de dialogue apparaît. Celle-ci vous permet de déterminer quelles nouvelles valeurs doivent être assignées aux anciennes.

Dans le cadre **Old Value** et sous **Value**, écrire la valeur à recoder et à droite, dans le cadre **New Value** et sous **Value**, la nouvelle valeur recodée, puis cliquez sur **Add**.

Dans notre cas, il faut entrer « 0 », qui deviendra « -4 », « 1 », qui deviendra « -3 » etc. Dans la case **Old → New**, vous voyez les valeurs déjà entrées. Si vous voulez changer une valeur, sélectionnez-la, faites vos modifications et cliquez sur **Change**. Pour éliminer un codage, sélectionnez-le et cliquez sur **Remove**.

Une fois toutes les valeurs à modifier inscrites dans cette case, cliquez sur **Continue** puis cliquez sur **OK**. La première boîte de dialogue se ferme et le logiciel aura créé votre nouvelle variable à la suite des autres dans la base de données.



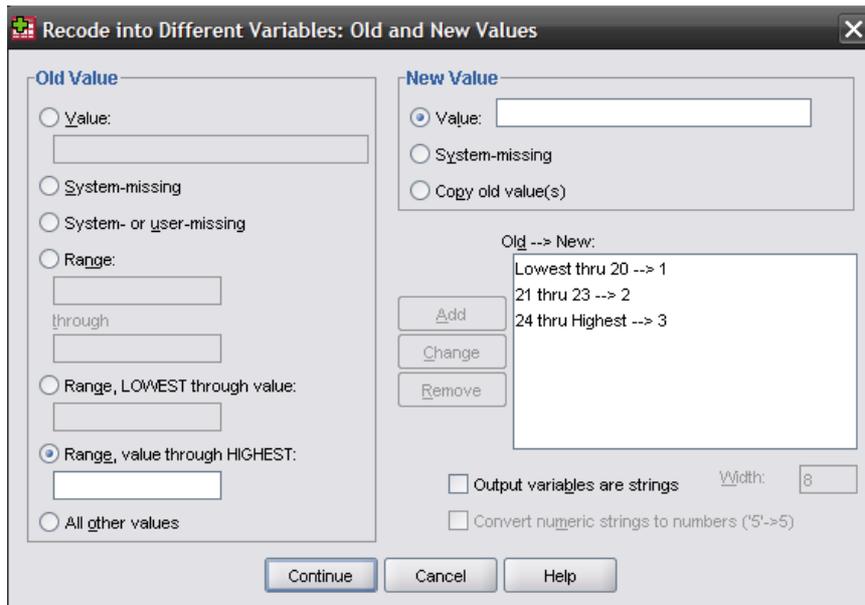
Si vous voulez faire des groupes (par exemple des groupes d'âge), il est laborieux de recoder toutes les valeurs une à une. Imaginons que vous voulez séparer les participant-e-s selon leur âge, le 1^{er} groupe allant jusqu'à 20 ans, le 2^{ème} de 21 à 23 et le 3^{ème} 24 ans et plus.

Dans la partie gauche (**Old Value**), sélectionnez **Range, LOWEST through value** (plus petit que...) et entrez le chiffre 20, écrivez le chiffre 1 dans **New Value**, puis cliquez sur **Add**.

À gauche, cliquez sur **Range**, entrez les chiffres 21 **through** 23, puis 2 dans **New Value** et enfin **Add**.

À gauche, cliquez sur **Range, value through HIGHEST** (plus grand que...), écrivez 24, 3 dans **New Value**, puis **Add**.

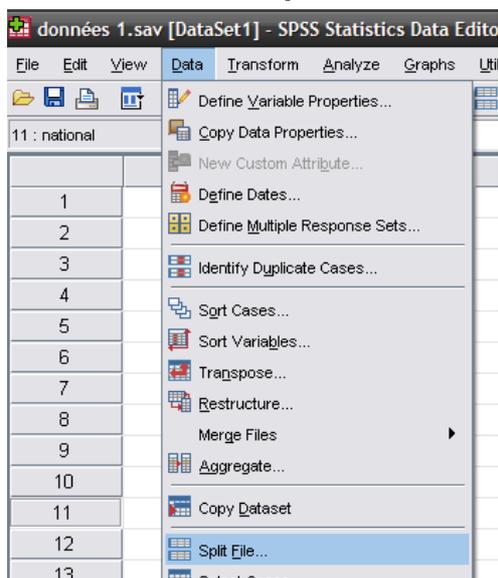
Notez à gauche l'option parfois très utile pour **Old Value** : vous pouvez recoder toutes les autres valeurs non inscrites (**All other values**).



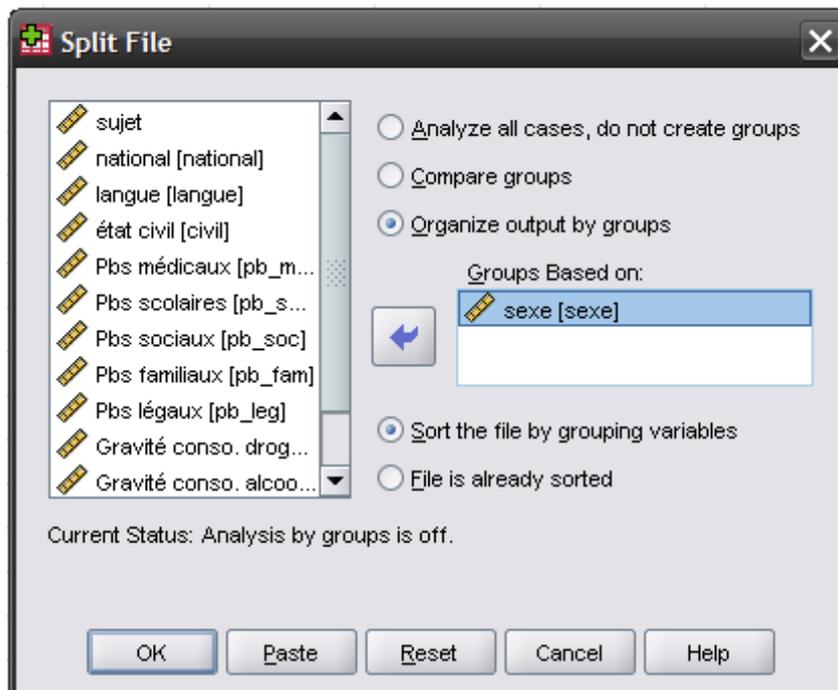
4.3. Fragmenter la base de données

But : Travailler sur l'ensemble des données, mais en obtenant des informations distinctes sur divers groupes en parallèle – comme les hommes et les femmes. Les résultats apparaissent alors séparément pour les groupes concernés dans l'output. Cette option vous permet de pouvoir considérer des groupes différents en même temps. C'est-à-dire que le logiciel, au lieu de travailler sur l'ensemble des individus, fera la même analyse sur plusieurs groupes d'individus et vous présentera les résultats pour chaque groupe séparément. Souvent la variable qui définit la fragmentation des données est une variable nominale (qui identifie des groupes de participant-e-s).

Chemin : **Data → Split File...**



Cliquez sur **Organize output by groups** (ou **Compare groups** si vous désirez avoir l'information dans un seul tableau à la place de plusieurs. Le résultat est le même, uniquement la présentation change). Sélectionnez dans la case de gauche la variable en fonction de laquelle vous voulez séparer vos résultats (ici « sexe ») et faites-la glisser dans la case de droite (**Groups Based on:**) à l'aide de la flèche.



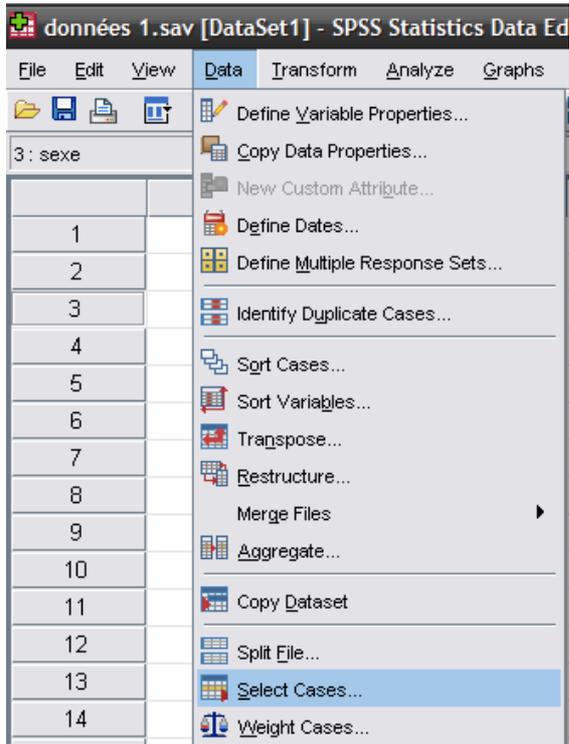
N'oubliez pas de désactiver le **Split File** quand vous n'en aurez plus besoin. Il ne se désactive pas seul et le logiciel ne vous rappelle pas qu'il est activé. Cette fonction peut être désactivée par:

Data → Split File → Analyze all cases, do not create groups et OK.

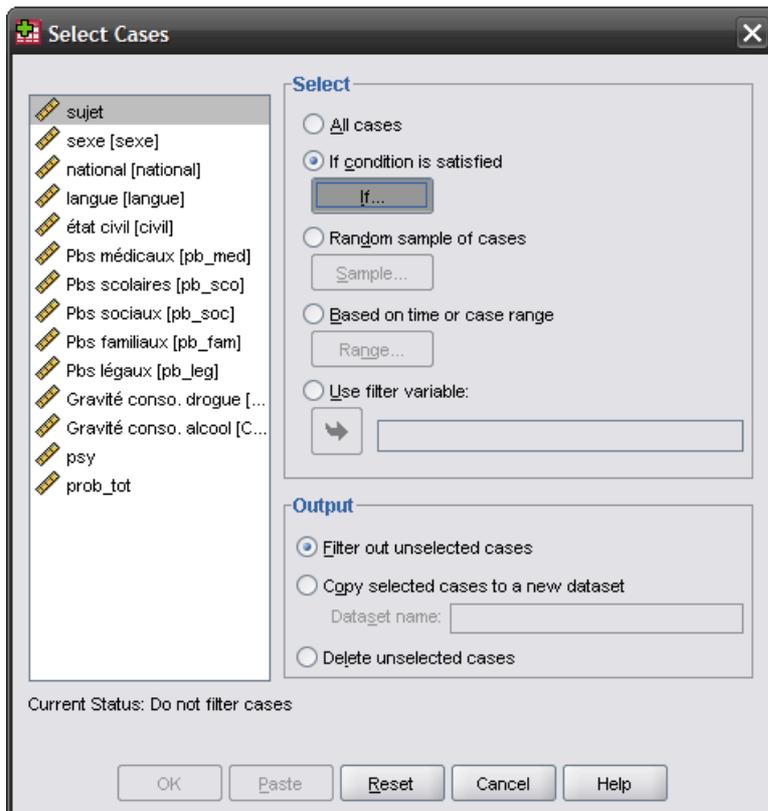
4.4. Travailler sur une partie des données

But : Lorsqu'une seule partie de l'échantillon vous intéresse, vous pouvez demander de n'avoir les résultats que pour cette partie, pourvu qu'elle soit définie par des variables incluses dans la base de données (par exemple : résultats pour les femmes). Encore une fois, la variable qui vous permet de sélectionner les individus est, la plupart du temps, une variable nominale. La différence entre cette fonction et la précédente est qu'ici vous n'avez pas les résultats pour tous les groupes (par exemple les hommes et les femmes), mais uniquement pour une catégorie de participant-e-s (les femmes).

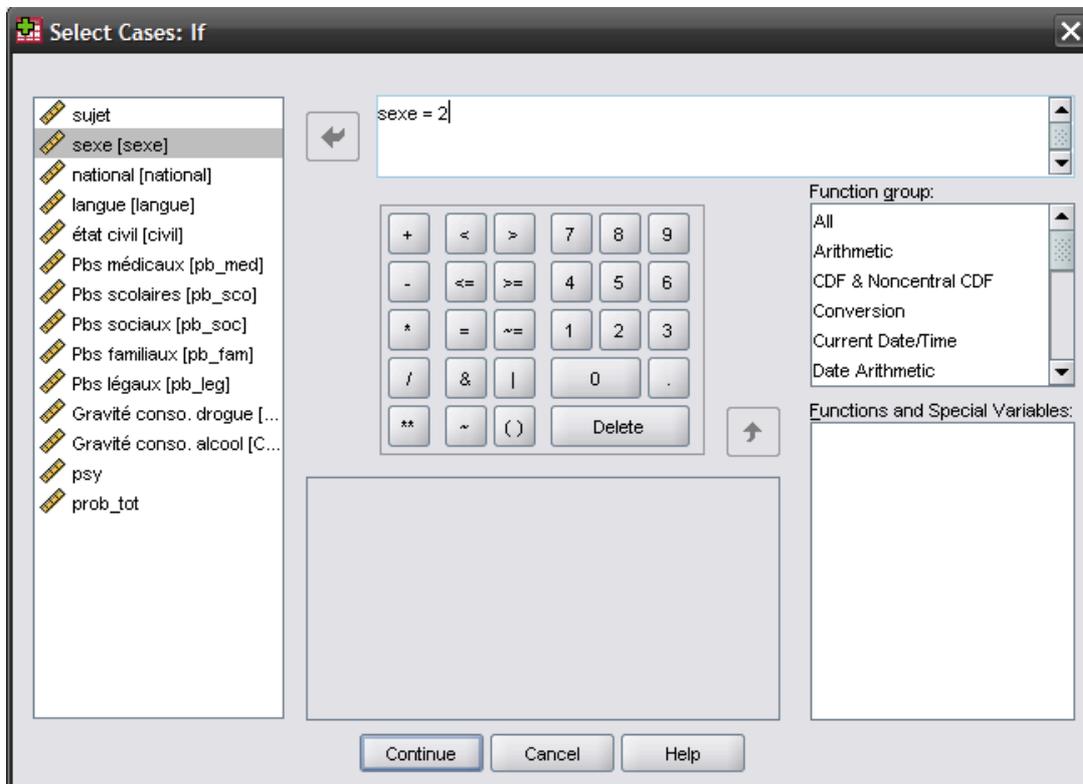
Chemin : **Data** → **Select Cases...**



Cliquez sur **If condition is satisfied**, puis l'onglet **If...**



Ecrire l'expression numérique nécessaire. Par exemple, pour traiter uniquement les données qui portent sur les femmes (codés 2), écrire : `sexe = 2`. Puis cliquer sur **Continue** et **OK**.



Si vous allez voir dans la fenêtre des données, sous **Data View**, vous verrez que les participant-e-s qui n'ont pas eu « 2 » à la variable **sexe** (donc les hommes), sont barrés, ils ne seront donc plus pris en compte par SPSS pour les analyses ultérieures.

	sujet	sexe
1	1.00	1.00
2	2.00	1.00
3	3.00	1.00
4	4.00	2.00
5	5.00	2.00
6	6.00	2.00
7	7.00	2.00
8	8.00	2.00

Cette sélection est active pour toute analyse faite par la suite jusqu'à ce que vous la désactiviez ou quittiez l'application. Pour désactiver :

Data → Select cases → All cases et OK.

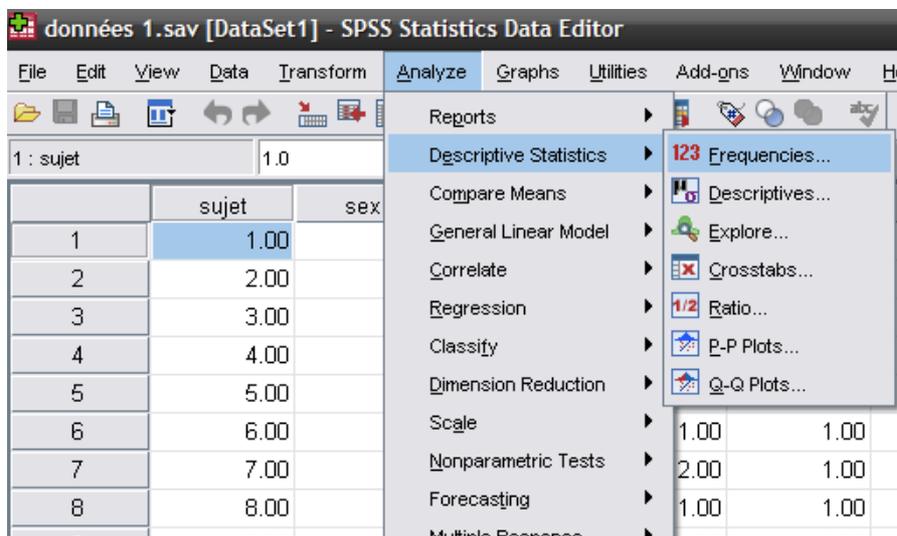
5. Statistiques Descriptives

Jusqu'à présent, il a été question de manipuler la base de données et les variables que celle-ci comprend. Nous passons maintenant à la phase de description des données.

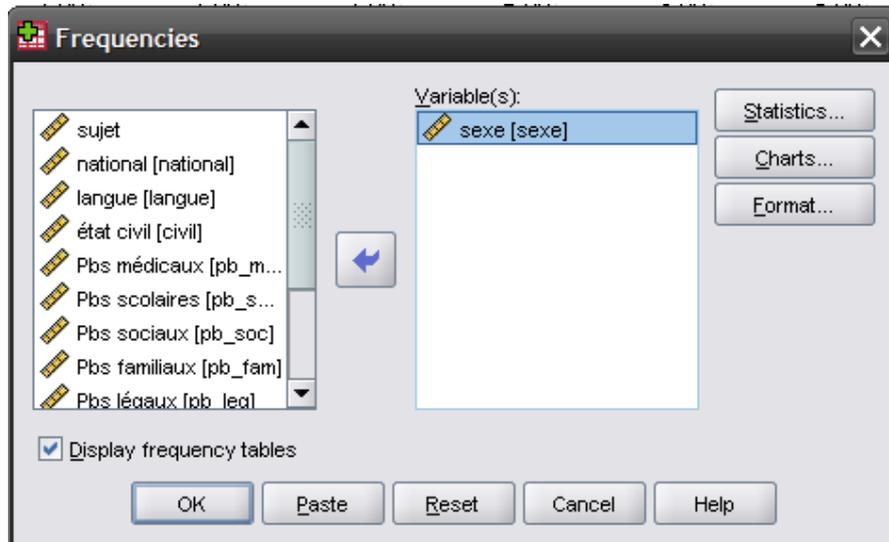
5.1. Calcul des fréquences

But : Calculer la fréquence (nombre et/ou pourcentage) des participant-e-s qui sont regroupé-e-s dans chaque modalité d'une variable (nominale, ordinale ou quantitative).

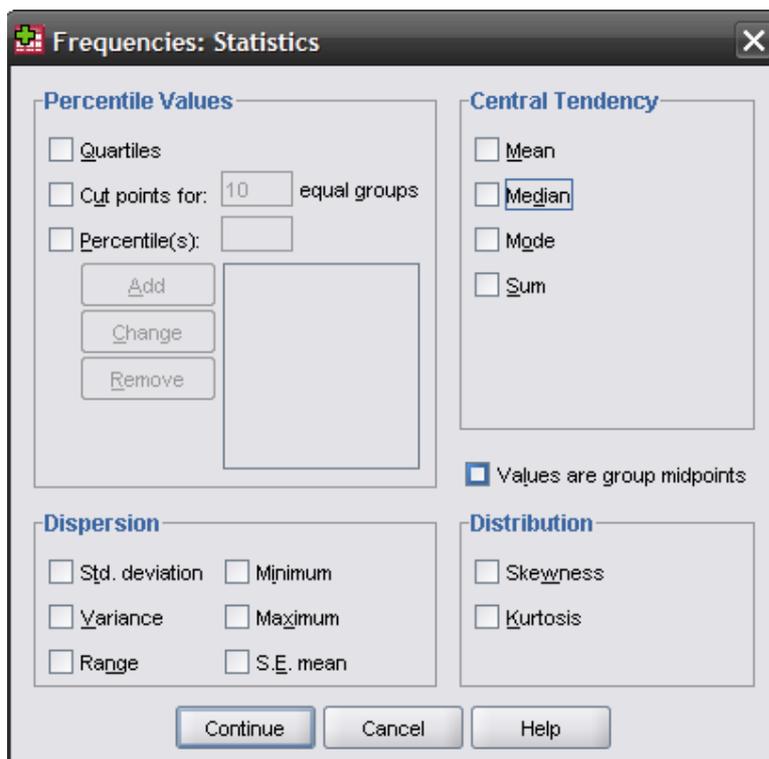
Chemin : **Analyze → Descriptive Statistics → Frequencies...**



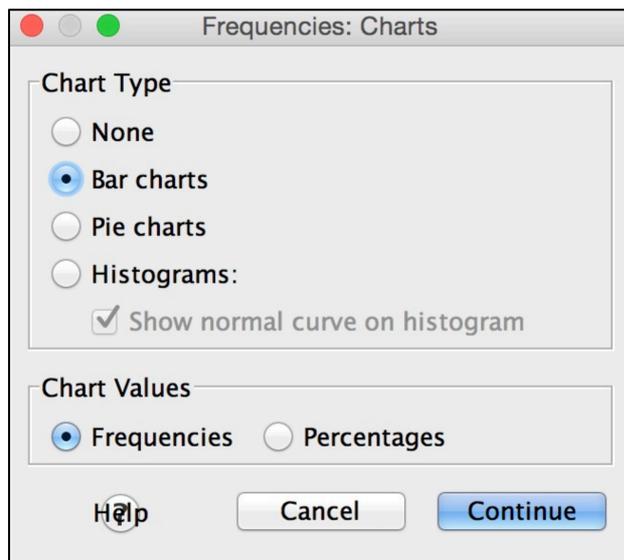
Sélectionnez dans la case de gauche les variables que vous souhaitez analyser. (vous pouvez demander les tableaux pour plusieurs variables à la fois, si elles se suivent dans la liste, vous sélectionnez la première variable, appuyez sur la touche Majuscule en maintenant appuyé et cliquez sur la dernière variable ; si elles ne se suivent pas, vous cliquez en maintenant la touche cmd appuyée), puis glissez-les à droite en utilisant la flèche.



Dans l'onglet **Statistics**, on peut choisir l'information que l'on désire avoir sur les données. Ainsi vous pouvez avoir des indices de tendance centrale (moyenne, médiane et mode), obtenir la valeur pour laquelle l'échantillon sera divisé en quartiles (**Quartiles**), en x sous-échantillons composés du même nombre de participant-e-s (**Cut points for: x equals groups**) ou certains pourcentiles (**Percentile(s):**), et d'autres informations.



Dans l'onglet **Charts...** vous pouvez demander le diagramme en bâtonnets (**Bar charts**) qui correspond à la version graphique du tableau de fréquence, ou l'histogramme (**Histograms:**) pour les variables ordinales ou quantitatives qui comptent beaucoup de modalités (avec l'histogramme, SPSS regroupe automatiquement des catégories pour que le graphique soit plus lisible, par exemple si vous avez des âges de 20 à 60 ans, plutôt que de créer 41 bâtonnets, l'histogramme va regrouper les âges de 2 en 2 pour garder 21 catégories, 20-22, 22-24, etc.).



Cliquez sur **Continue** puis **OK** pour lancer votre analyse. Le logiciel créera alors une nouvelle fenêtre de résultats (**Output**).

SPSS OUTPUT :

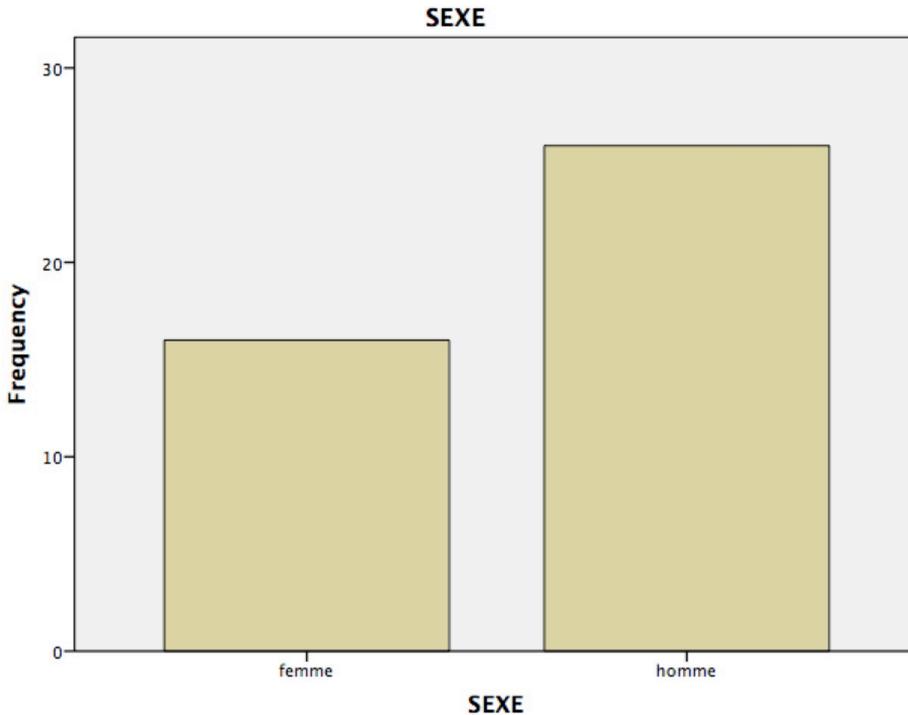
Statistics

SEXE

N	Valid	42
	Missing	2

SEXE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	femme	16	36,4	38,1	38,1
	homme	26	59,1	61,9	100,0
	Total	42	95,5	100,0	
Missing	System	2	4,5		
Total		44	100,0		



Dans cet échantillon, il y a 26 hommes (59.1 % de l'échantillon) et 16 femmes (36.4 % de l'échantillon). Vous remarquez qu'en ligne, vous avez les catégories qui composent votre variable (modalités). Si vous avez inscrit des labels lors de la constitution des variables et que vous avez demandé leur affichage dans les options (voir point 3.3), ces noms apparaissent dans le tableau, autrement vous avez des chiffres.

En colonne, vous voyez les fréquences (**Frequency** : le nombre de participant-e-s dans chaque catégorie). Les participant-e-s qui n'ont pas répondu sont indiqués-e-s comme **Missing System** (ici 2 personnes). Suivent (de gauche à droite), les pourcentages des participant-e-s dans chaque catégorie (**Percent**) : attention, parmi ces pourcentages se trouvent également les participant-e-s qui n'ont pas répondu. Si vous voulez indiquer les pourcentages de participant-e-s dans les catégories qui ont répondu à la question, vous devez faire référence à la colonne qui suit (**Valid Percent**). La dernière colonne est la somme des pourcentages des individus à chaque nouvelle catégorie (**Cumulative Percent**).

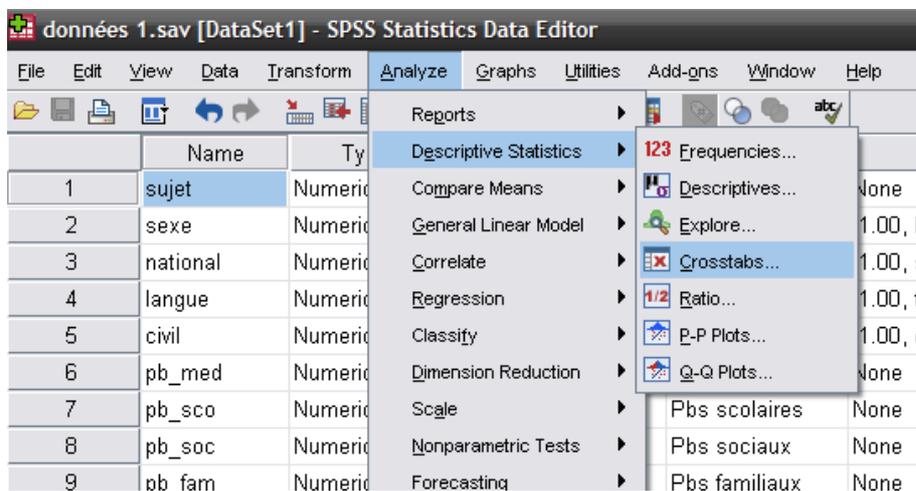
5.2. Tableaux croisés ou tableaux de contingence (2 variables nominales)

But : Calculer la fréquence (nombres et/ou pourcentage) des participant-e-s qui sont compri-se-s dans les modalités d'une variable nominale, croisée avec les participant-e-s qui sont compri-se-s dans les modalités d'une seconde variable nominale.

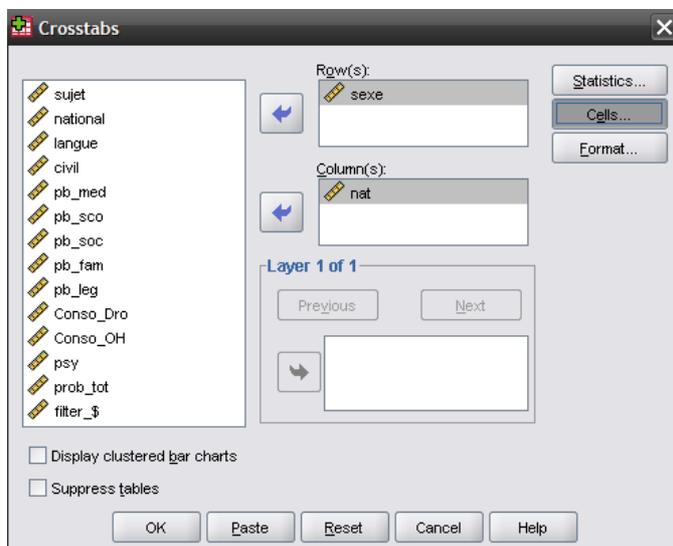
Attention : vous n'allez utiliser que des variables nominales pour cette fonction.

Exemple : ici nous allons regarder le nombre d'hommes et de femmes (variable « sexe ») qui sont de nationalité suisse ou d'une autre nationalité (variable « nat »).

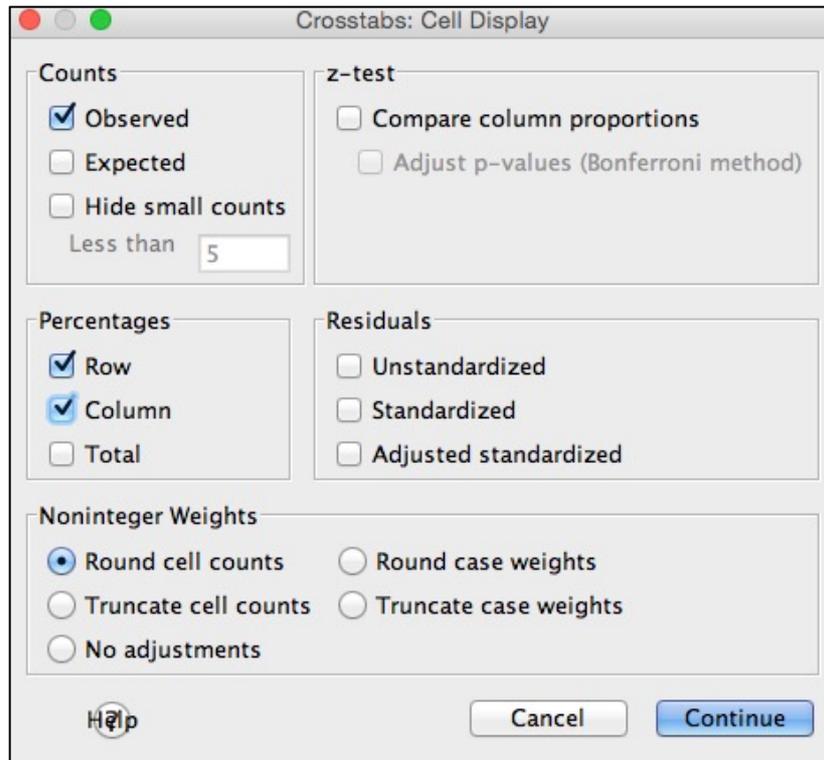
Chemin : **Analyze** → **Descriptive Statistics** → **Crosstabs...**



Entrez une première variable dans la case **Row(s):** et une deuxième variable sous **Column(s):** toujours en les sélectionnant et en cliquant sur la flèche qui sépare les deux cases. L'ordre et la position que vous utilisez pour entrer les variables ne sont pas importants pour le résultat, seule la présentation change.



Ensuite, cliquez sur l'onglet **Cells...** pour demander les pourcentages (en ligne et en colonne).



SPSS OUTPUT :

Si vous avez entré la variable « nat » en colonne (**column**) cela signifie que les catégories « suisse » et « autre » seront sur le haut du tableau et définissent les colonnes de celui-ci. Les pourcentages en colonne se lisent alors comme le pourcentage des deux groupes dans les catégories de la variable en ligne. Pour savoir quelle valeur regarder, faites attention à quels chiffres donnent une somme de 100% en colonne.

Si vous avez entré la variable « sexe » en ligne (**row**) cela signifie que les catégories seront sur le côté gauche du tableau et définiront les lignes de celui-ci. Les pourcentages en ligne se lisent alors comme le pourcentage d'hommes ou de femmes dans les catégories de la variable en colonne (« nat »). Pour savoir quelle valeur regarder, faites attention à quels chiffres donnent la somme de 100% en ligne.

sexe * nat Crosstabulation

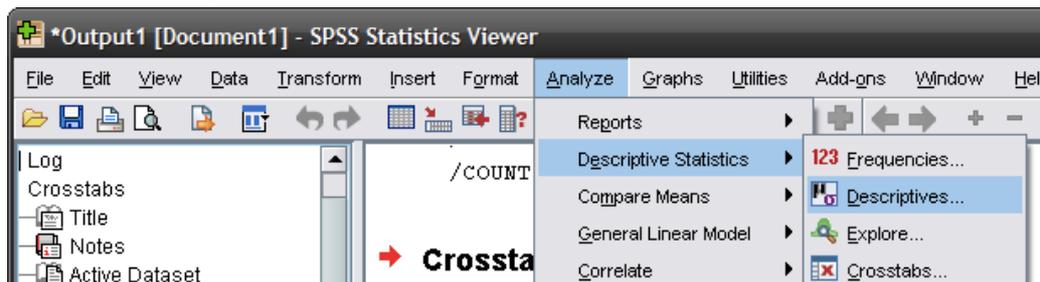
			nat		Total
			suisse	autre	
sexe	homme	Count	48	15	63
		% within sexe	76,2%	23,8%	100,0%
		% within nat	65,8%	57,7%	63,6%
	femme	Count	25	11	36
		% within sexe	69,4%	30,6%	100,0%
		% within nat	34,2%	42,3%	36,4%
Total	Count	73	26	99	
	% within sexe	73,7%	26,3%	100,0%	
	% within nat	100,0%	100,0%	100,0%	

On peut voir que l'échantillon se compose de 25 femmes suisses et qu'elles représentent 69,4% des femmes de l'échantillon et 34.2% des suisses de l'échantillon. Remarque : l'échantillon est l'ensemble des participant-e-s qui constituent la base des données.

5.3. Moyennes (variables quantitatives)

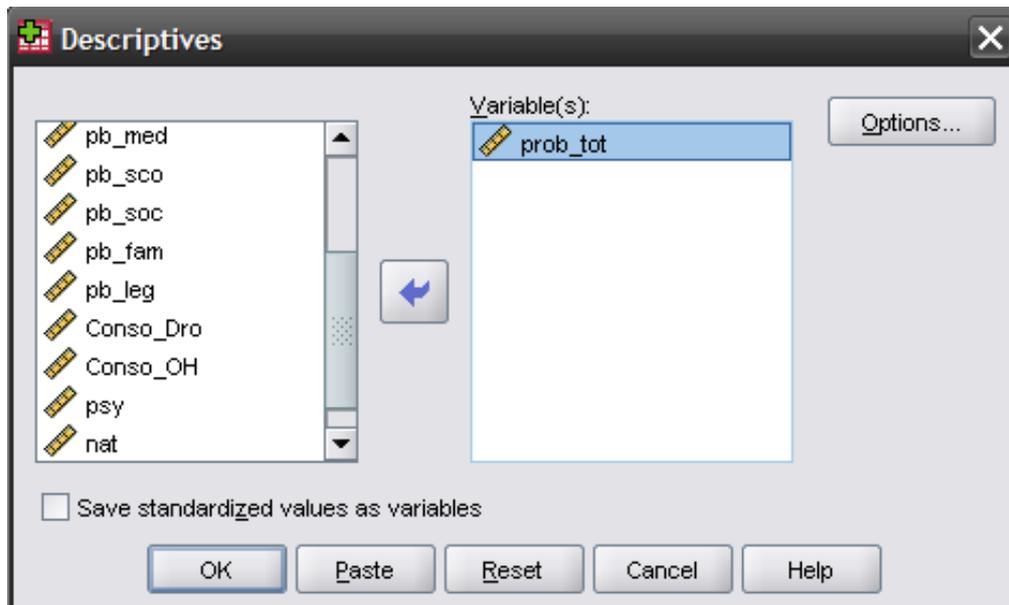
But : Obtenir la moyenne d'une ou de plusieurs variable(s).

Chemin : **Analyze** → **Descriptives statistics** → **Descriptives...**

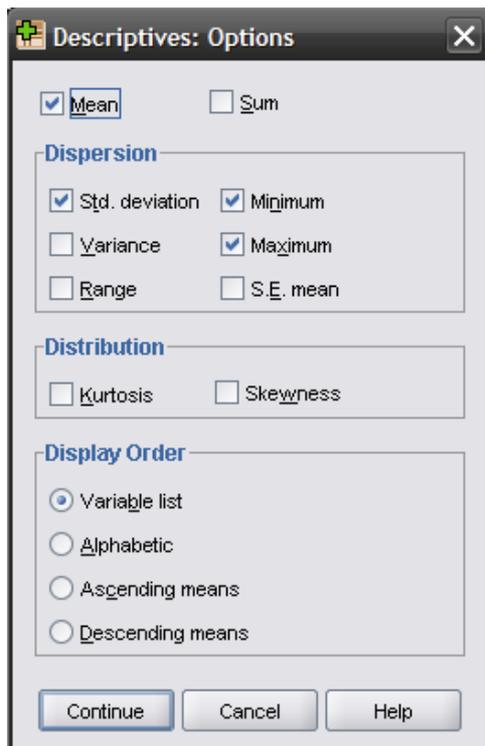


Exemple : nous voulons connaître la moyenne des problèmes totaux (prob_tot) rencontrés par nos participant-e-s.

Entrez la variable en question dans **Variable(s)**: (il est possible d'obtenir ces statistiques pour plusieurs variables).



Dans l'onglet **Options**, il est possible de demander plusieurs indices statistiques descriptifs concernant la variable sélectionnée : la moyenne (**Mean**), l'écart-type (**Std. deviation**), la somme (**Sum**), la valeur minimale (**Minimum**) et la valeur maximale (**Maximum**), etc.



SPSS OUTPUT :

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
prob_tot	99	.80	7.40	4.4061	1.57932
Valid N (listwise)	99				

Le tableau montre que le taux de problème moyen des participant-e-s est de 4.41, avec un écart-type de 1.58. Le/la participant-e avec le moins de problèmes est à .80 et celui avec le plus de problèmes à 7.40.

5.4. Observation d'une variable par rapport à une autre

But : Obtenir les données descriptives d'une variable quantitative pour différents sous-groupes de participant-e-s, donc en fonction d'une **variable nominale**. Dans le chapitre 5.3, vous avez calculé la moyenne sur l'ensemble des participant-e-s qui constituent l'échantillon. Dans ce cas-ci, vous voulez calculer la moyenne de certains sous-groupes de votre échantillon.

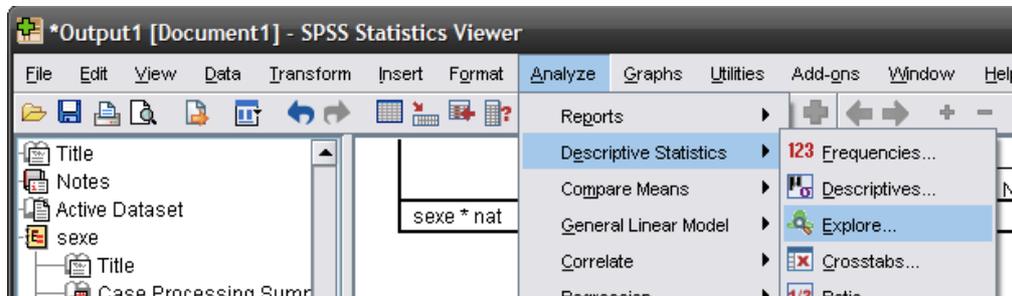
Exemple : quelle est la moyenne des problèmes des femmes et des hommes ?

Vous pouvez procéder de plusieurs manières. Aux points 4.3 (Split File) et 4.4 (Select Cases), nous vous avons montré comment diviser vos participant-e-s en plusieurs sous-groupes ou sélectionner seulement les participant-e-s qui vous intéressent. Vous pouvez utiliser cette méthode, mais il y a une manière plus simple

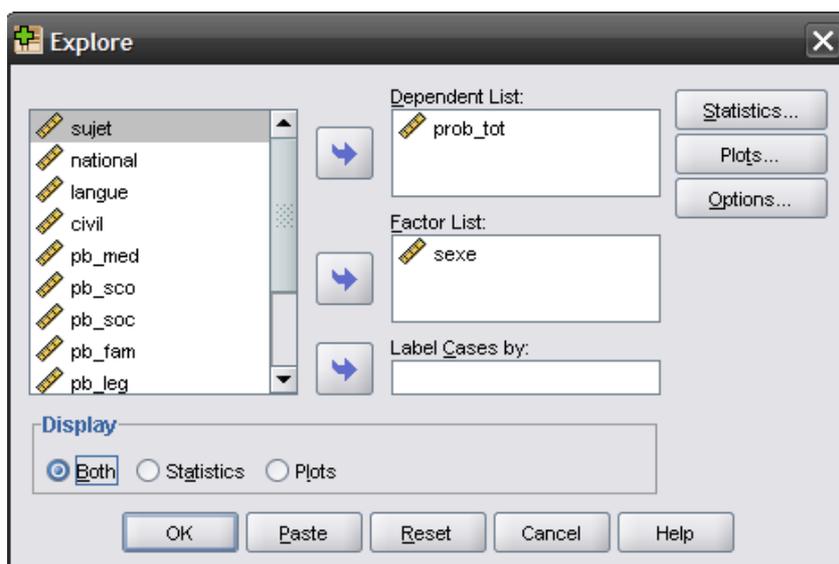
de procéder, qui consiste à créer des tableaux avec l'information que vous cherchez. Il s'agit « d'explorer » une variable en fonction des modalités d'une autre.

Attention : il est question d'une variable quantitative et d'une variable nominale !

Chemin : **Analyze** → **Descriptive Statistics** → **Explore...**

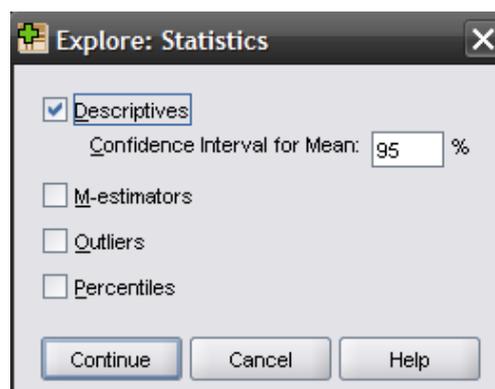


Entrez la variable quantitative dans la case **Dependent List:** et la (ou les) variable(s) nominale(s) dans la case **Factor List.** À noter que si vous introduisez plusieurs variables dans **Factor List**, vous aurez un tableau pour chaque variable nominale.



Dans la case **Display**, vous pouvez choisir si vous souhaitez les données statistiques, les graphiques (**Plots**) ou les deux (**Both**).

Dans l'onglet **Statistics**, vous cochez **Descriptives** pour avoir les indices descriptifs.



SPSS OUTPUT :

			Descriptives		
sexe			Statistic	Std. Error	
prob_tot	homme	Mean	4.1905	.20327	
		95% Confidence Interval for Mean	Lower Bound	3.7841	
			Upper Bound	4.5968	
		5% Trimmed Mean	4.1970		
		Median	4.2000		
		Variance	2.603		
		Std. Deviation	1.61342		
		Minimum	.80		
		Maximum	7.20		
		Range	6.40		
		Interquartile Range	2.40		
		Skewness	.013	.302	
		Kurtosis	-.639	.595	
		femme		Mean	4.7833
95% Confidence Interval for Mean	Lower Bound			4.2881	
	Upper Bound			5.2786	
5% Trimmed Mean	4.7753				
Median	4.6000				
Variance	2.143				
Std. Deviation	1.46375				
Minimum	2.40				
Maximum	7.40				
Range	5.00				
Interquartile Range	2.45				
Skewness	.048			.393	
Kurtosis	-1.151			.768	

Par défaut, SPSS nous donne de nombreuses valeurs pour chaque variable, mais les plus importantes (moyenne et écart-type), sont ici entourées.

Vous pouvez voir que les femmes ont une moyenne des problèmes plus élevée ($M = 4.78$, $SD = 1.61$) que les hommes ($M = 4.19$, $SD = 1.46$).

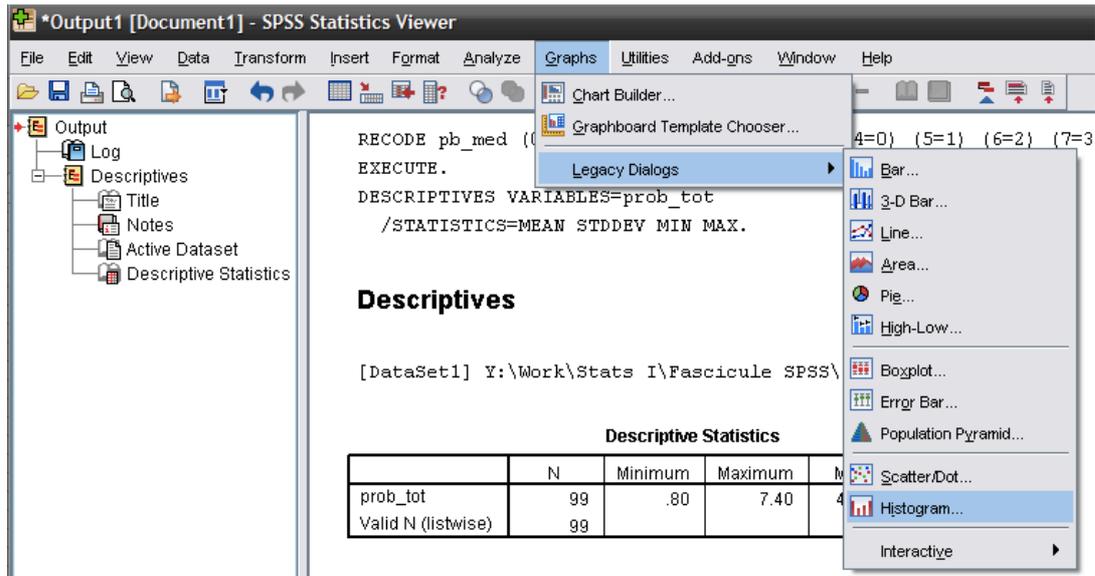
A noter pour cette commande **Explore** que dans **Plots**, vous pouvez choisir en plus du box-plot (cf. chapitre 5.5.2) donné par défaut l'histogramme, en cochant la case correspondante, ainsi que le test de normalité Kolmogorov-Smirnov (en cochant **Normality plots with tests**, cf. chapitre 5.7.3)

5.5. Graphiques

Chemin : **Graphs** → **Legacy Dialogs**, puis le nom du graphe, (ou par **Graphs** → **Chart Builder** mais c'est plus compliqué) ou encore dans depuis la commande **Explore** pour les histogrammes et boxplots.

5.5.1. Histogramme

Chemin : **Graphs** → **Legacy Dialogs** → **Histogram...**



The screenshot shows the SPSS Statistics Viewer interface. The 'Graphs' menu is open, and the 'Legacy Dialogs' option is selected. A sub-menu is displayed, showing 'Histogram...' as the chosen option. The main window displays the following commands and a table of descriptive statistics:

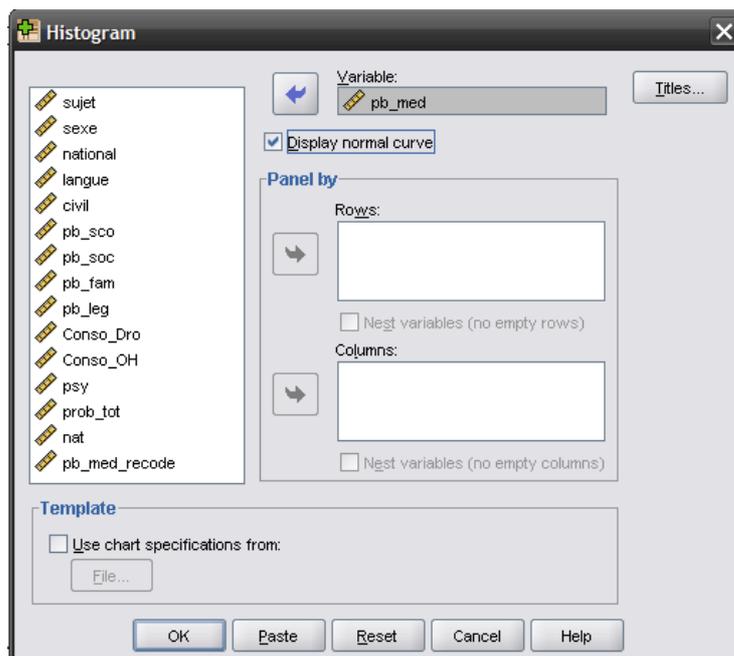
```
RECODE pb_med (0=0) (5=1) (6=2) (7=3)
EXECUTE.
DESCRIPTIVES VARIABLES=prob_tot
  /STATISTICS=MEAN STDDEV MIN MAX.
```

Descriptives

[DataSet1] Y:\Work\Stats I\Fascicule SPSS\

Descriptive Statistics				
	N	Minimum	Maximum	Mean
prob_tot	99	.80	7.40	4.40
Valid N (listwise)	99			

Puis, entrez la variable souhaitée dans la case **Variable**, grâce à la flèche, puis cliquez sur **OK**. Cocher la case **Display normal curve** permet d'afficher la courbe normale, c'est-à-dire la courbe que devrait suivre les données si la variable est distribuée normalement. Cette option est très pratique pour estimer la normalité des variables et aussi détecter d'éventuelles valeurs extrêmes.



The screenshot shows the 'Histogram' dialog box. The 'Variable' field contains 'pb_med'. The 'Display normal curve' checkbox is checked. The 'Panel by' section is empty. The 'Template' section has the 'Use chart specifications from:' checkbox unchecked.

Variable: pb_med

Display normal curve

Panel by

Rows:

Columns:

Nest variables (no empty rows)

Nest variables (no empty columns)

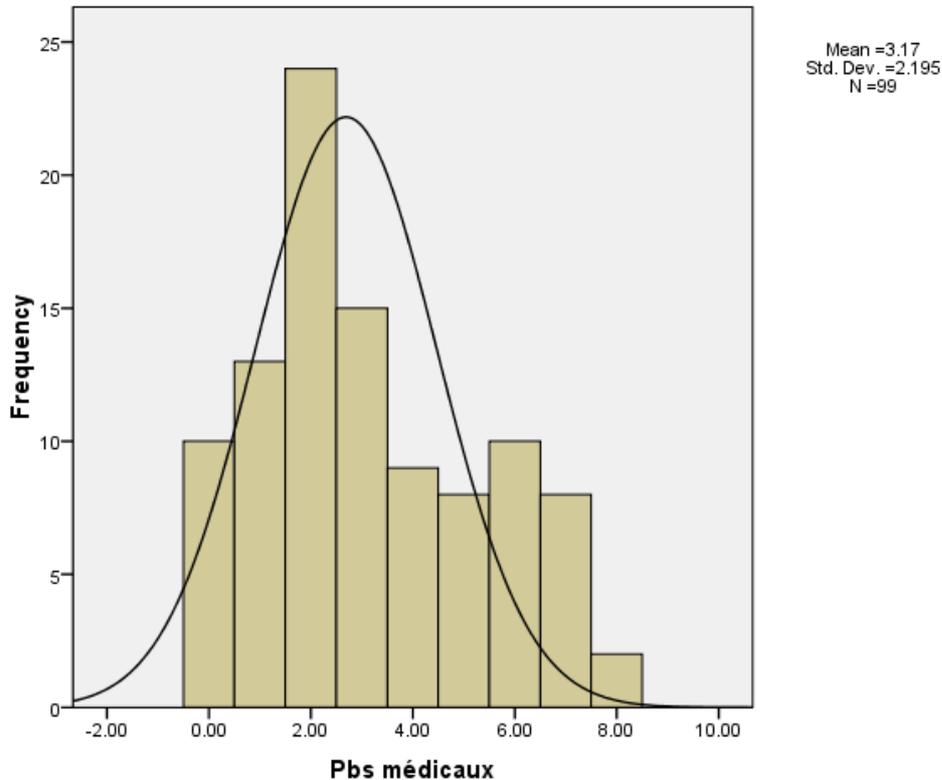
Template

Use chart specifications from:

File...

OK Paste Reset Cancel Help

SPSS OUTPUT :



L'histogramme s'affiche dans la fenêtre **Output**, en vous donnant également la moyenne et l'écart-type de la variable en haut à droite du graphique.

Vous pouvez également obtenir l'histogramme par la fonction **Explore** (voir plus loin).

5.5.2. Box-plots

Chemin : **Graphs** → **Legacy Dialogs** → **Box-plot...**

```
RECODE pb_med (0=0) (1=1) (2=2) (3=3) (4=4) (5=1) (6=2) (7=3)
EXECUTE.
DESCRIPTIVES VARIABLES=prob_tot
  /STATISTICS=MEAN STDDEV MIN MAX.
```

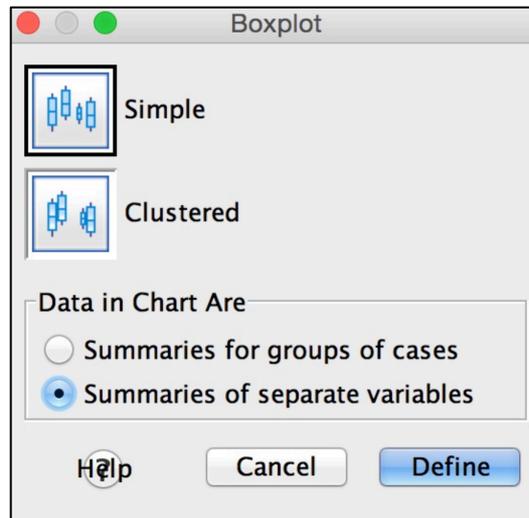
Descriptives

[DataSet1] Y:\Work\Stats I\Fascicule SPSS\

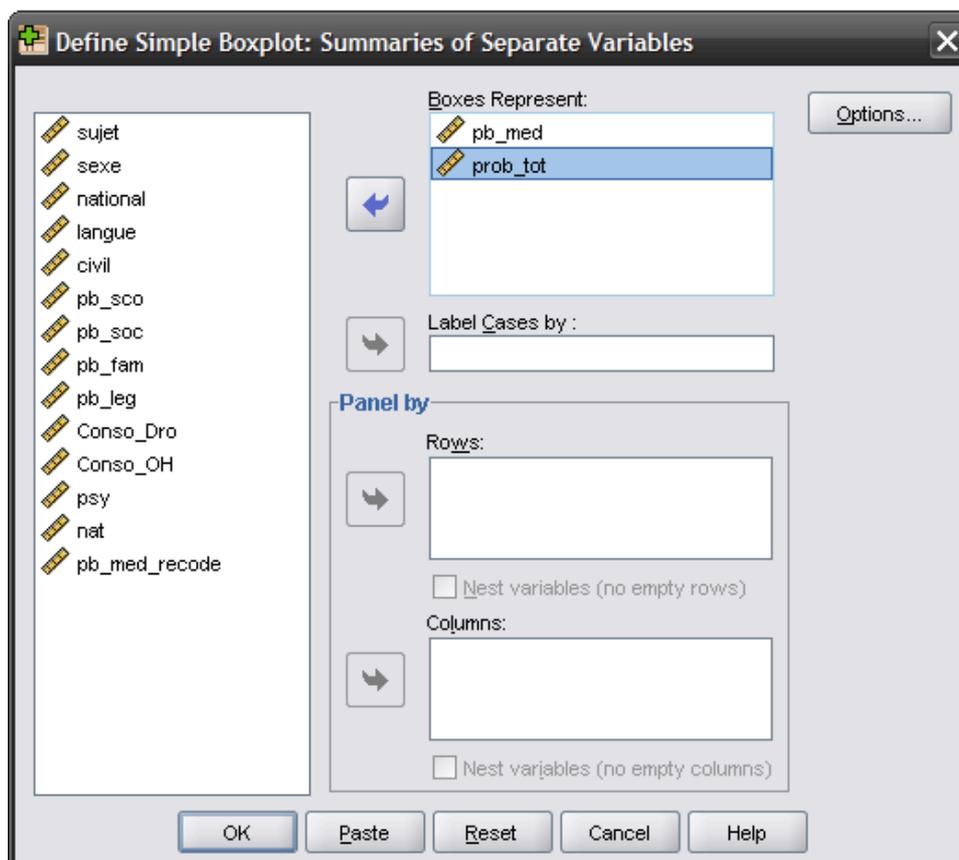
Descriptive Statistics				
	N	Minimum	Maximum	Mean
prob_tot	99	.80	7.40	3.17
Valid N (listwise)	99			

Vous obtenez d'abord une fenêtre d'options, depuis laquelle il est possible de générer quatre types de box-plot. À noter que plus on introduit de variables dans un graphique, moins il est lisible. Veillez à bien réfléchir à quelles informations vous souhaitez pouvoir voir dans votre box-plot avant de choisir le type.

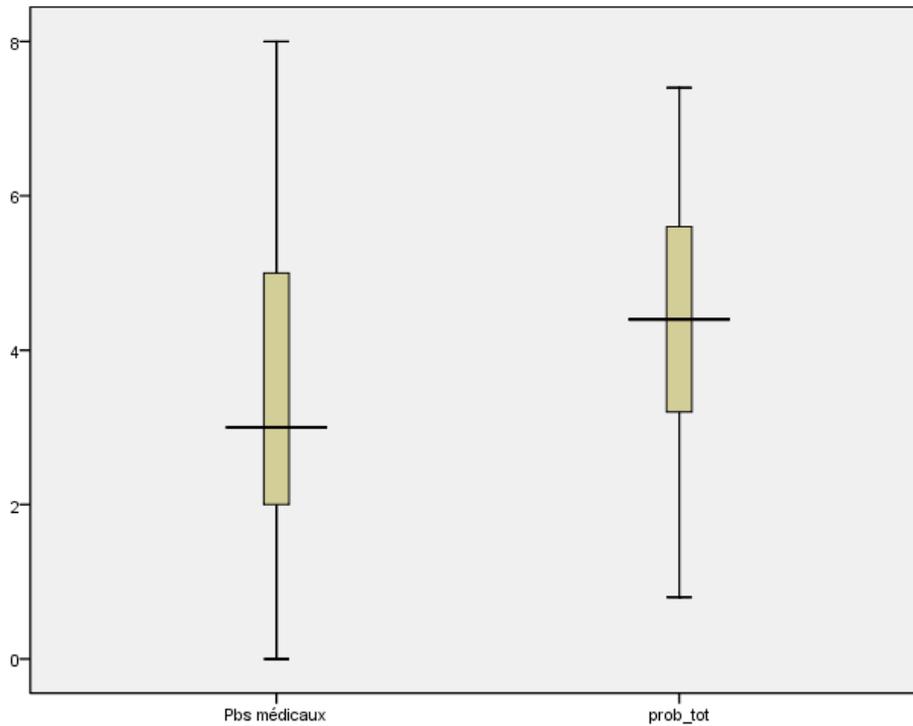
Cas 1 : Pour générer un box-plot représentant plusieurs variables, mais présentées individuellement, choisir **Simple** et **Summaries for separate variables**, puis presser **Define**.



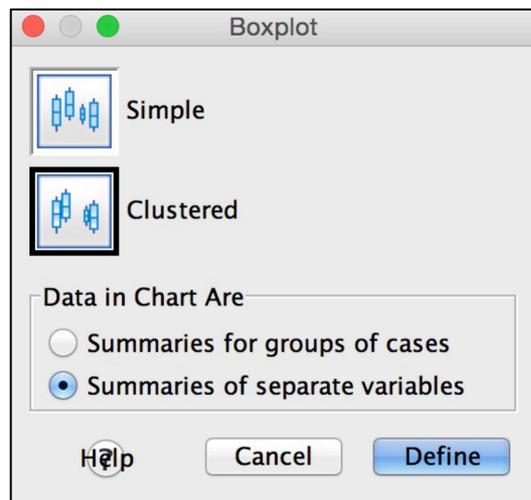
Dans la case **Boxes represent**, entrez toutes les variables que vous souhaitez voir affichées dans le box-plot. Attention toutefois aux échelles, si elles sont trop différentes entre les variables, vous aurez de la peine à lire les informations.



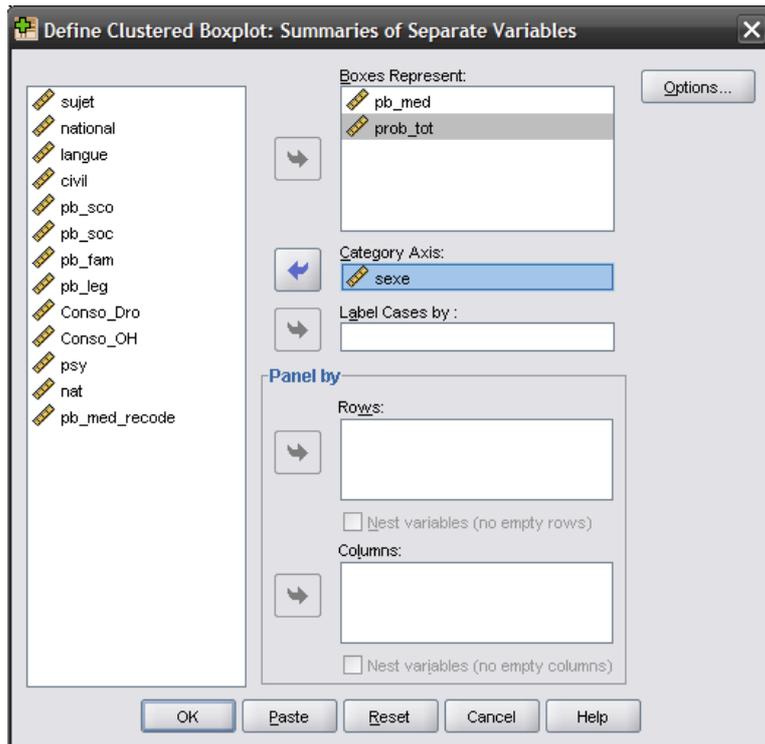
SPSS OUTPUT :



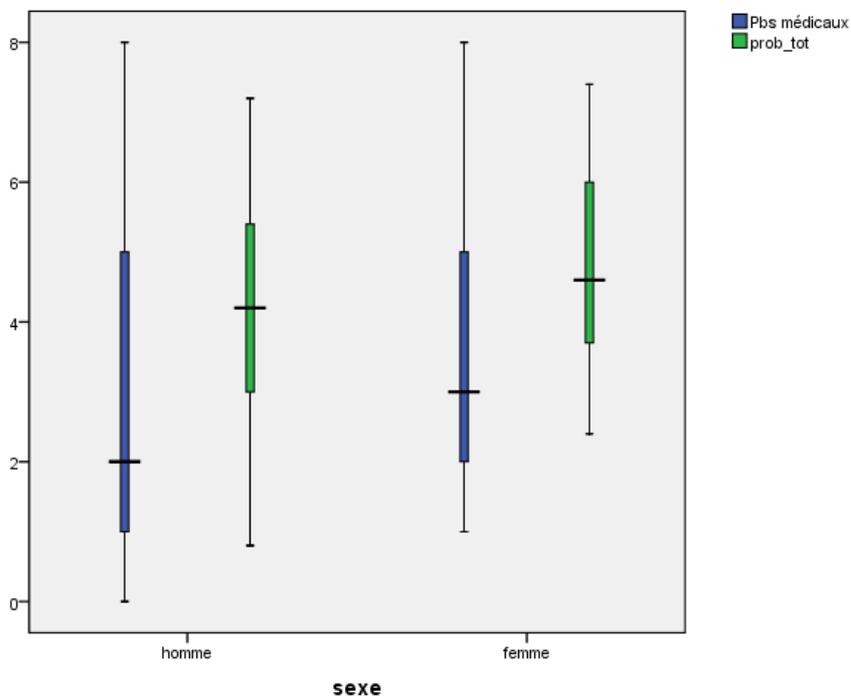
Cas 2 : En choisissant **Clustered**, puis **Summaries for separate variables**, vous obtiendrez un box-plot représentant deux variables séparées, mais en fonction d'une troisième variable nominale.



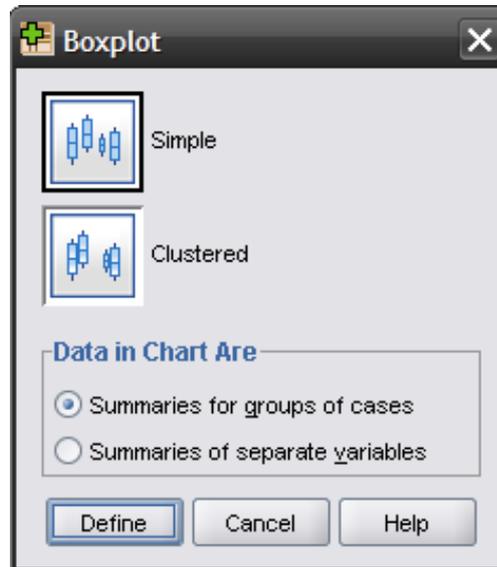
Entrez toutes les variables souhaitées dans **Box Represent:** (encore une fois, attention aux échelles), la variable nominale dans **Category Axis**, puis **OK**.



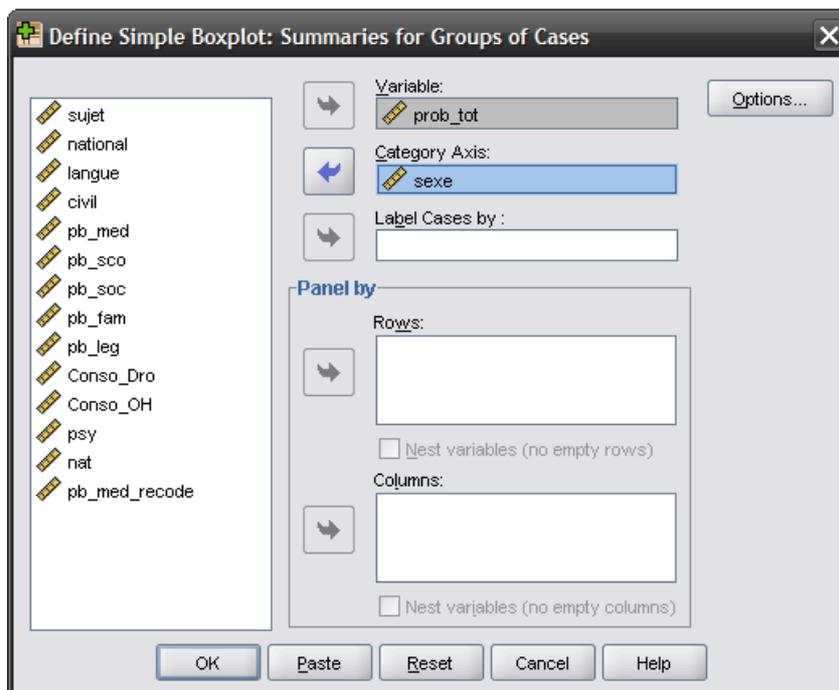
SPSS OUTPUT :



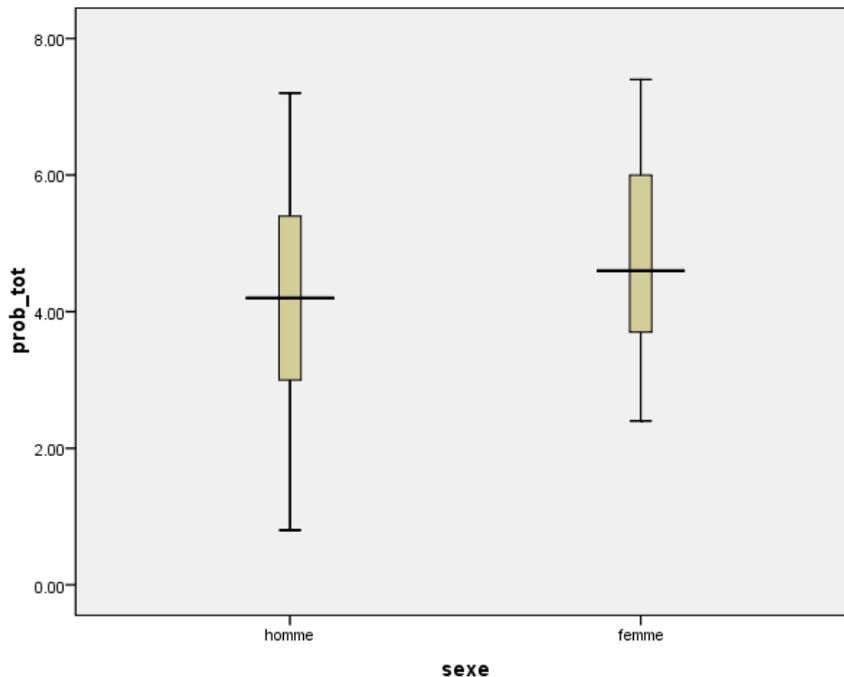
Cas 3 : Pour générer un box-plot représentant deux sous-groupes, choisir **Simple** et **Summaries for groups of Cases**. Ce type de box-plot est plutôt adapté pour une variable quantitative et une variable nominale.



Entrez votre variable quantitative dans la case **Variable**, et la variable nominale, dans la case **Category Axis**, puis **OK**.

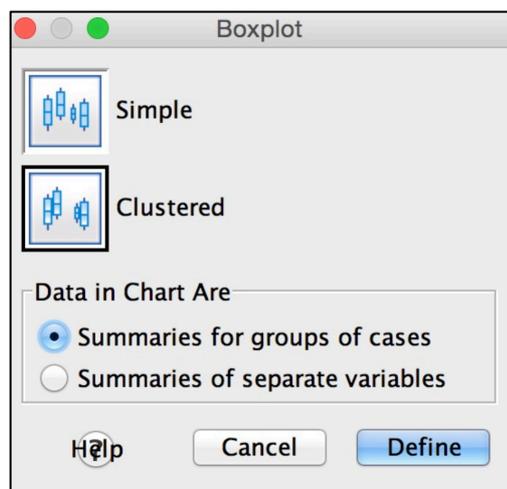


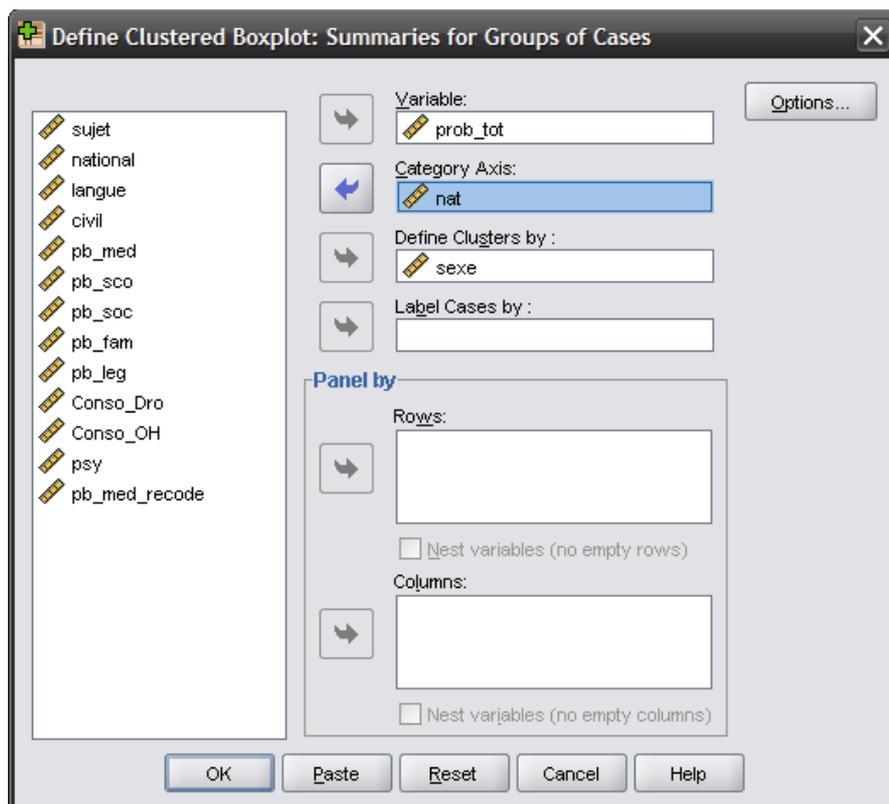
SPSS OUTPUT :



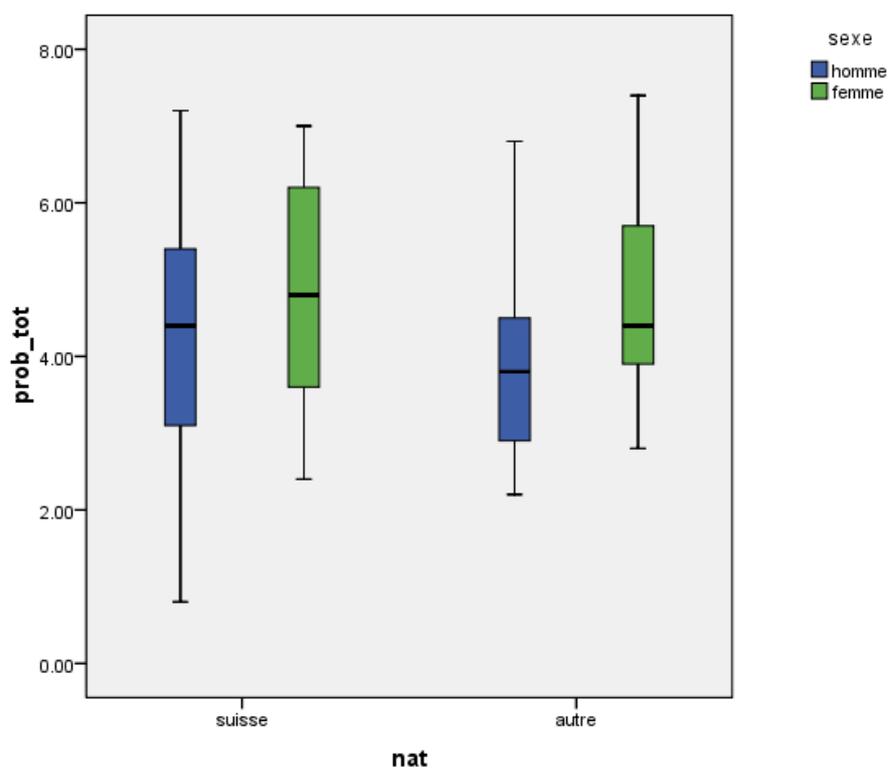
À noter que c'est le même box-plot obtenu en choisissant **Plots** dans la fonction **Explore** (voir point 5.4).

Cas 4 : Choisissez **Clustered**, puis **Summaries for groups of cases**, si vous souhaitez représenter une variable quantitative en fonction de deux variables nominales. Ici, nous avons choisis *prob_tot* en fonction du sexe et de la nationalité. La variable nominale que vous entrez dans **Category axis**, sera la variable représentée sur l'axe X et celle dans **Define Clusters by** représentera les différentes couleurs.





SPSS OUTPUT :



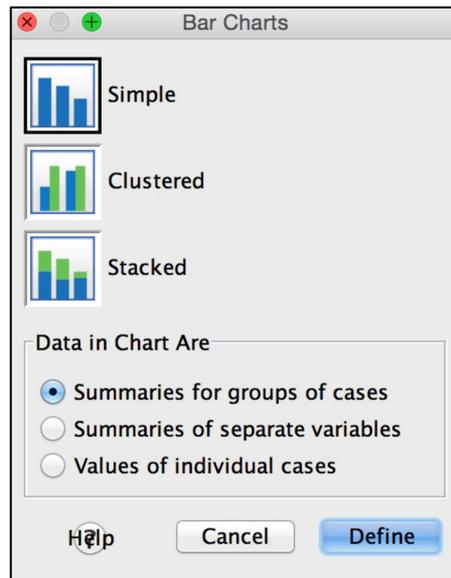
5.5.3. Graphiques de moyennes

But : Les graphiques de moyenne sont utiles pour représenter les résultats lorsque l'on veut représenter des moyennes de variables quantitatives dans des sous-groupes de l'échantillon définis par des variables nominales.

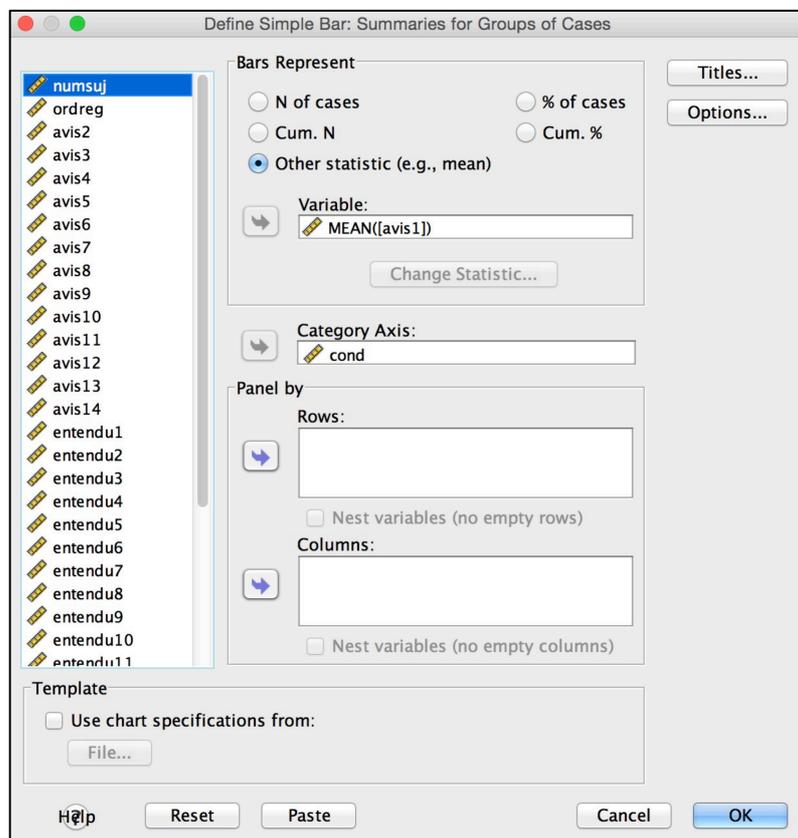
Chemin : **Graphs** → **Legacy Dialogs** → **Bar...**

Cas 1 : Une variable nominale et une variable quantitative.

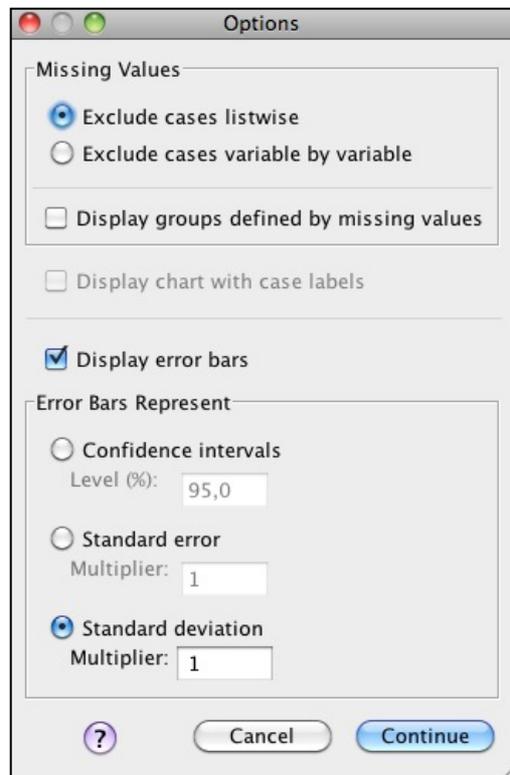
Choisir **Simple** et **Summaries for groups of cases**, puis presser **Define**.



Cochez **Other statistic** (e.g., mean) dans la case **Bar represent:**, entrez la variable quantitative dans **Variable:** et la variable nominale dans **Category axis**.

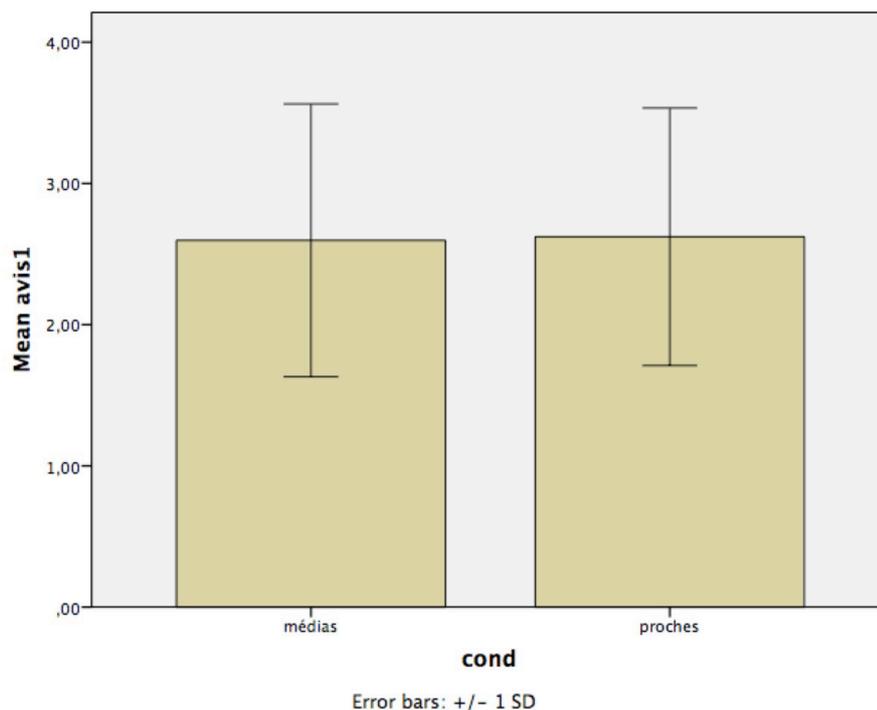


Dans l'onglet **Options...**, vous pouvez représenter les barres d'erreur (soit l'intervalle de confiance généralement de 95%, soit l'erreur standard ou l'écart-type, en fixant le **Multiplieur** à 1 : barre d'erreur = 1 x le *SD* ou le *SE* au dessus et au-dessous de la moyenne représentée par la hauteur du bâtonnet) :



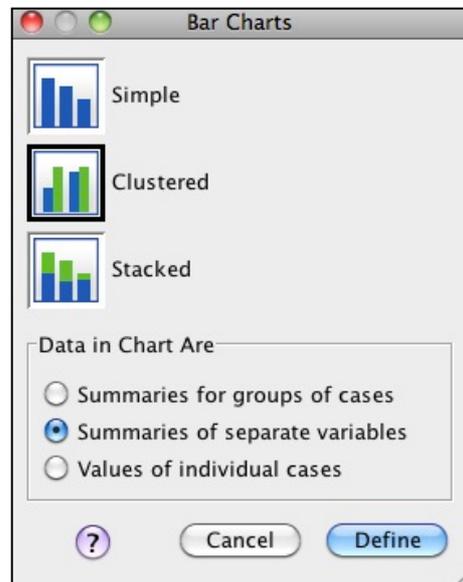
SPSS OUTPUT :

Graph

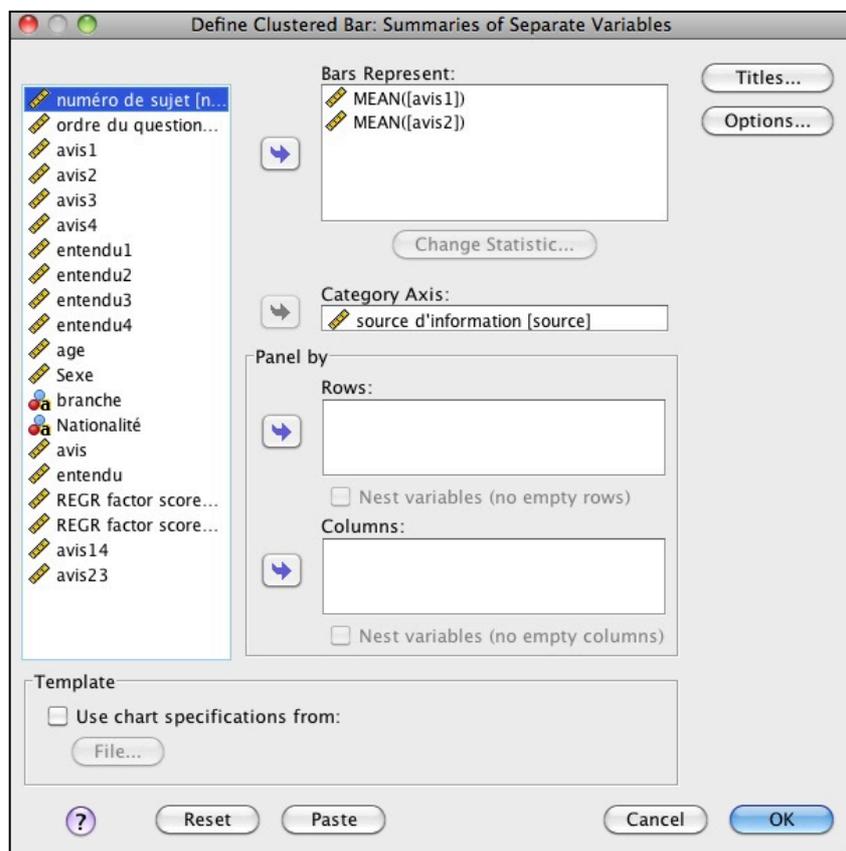


Note : En cliquant deux fois sur le graphique, puis deux fois sur les différentes parties du graphique (axes, barres, etc.), vous pouvez l'éditer pour le rendre plus lisible (nom des axes, grandeur des polices, couleurs, etc.).

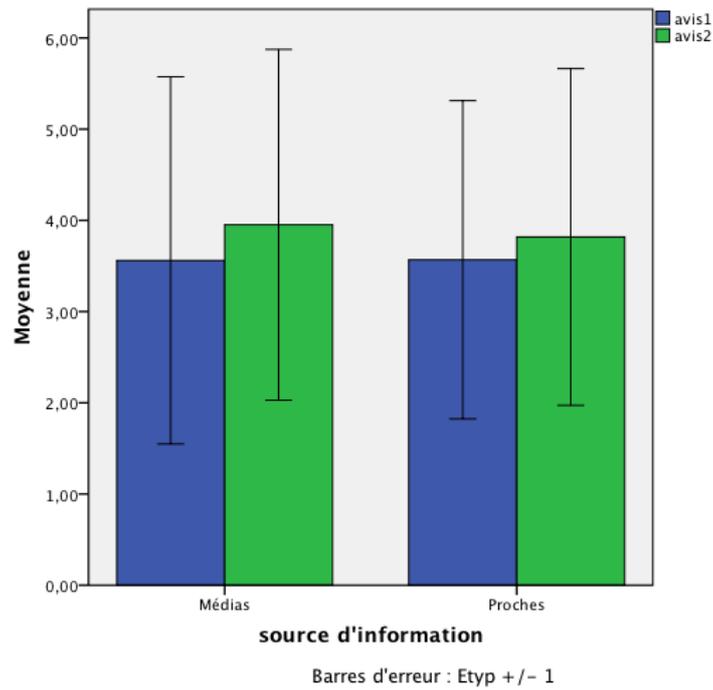
Cas 2 : Une variable nominale et plusieurs variables quantitatives.
Choisir **Clustered** et **Summaries of separate variables**.



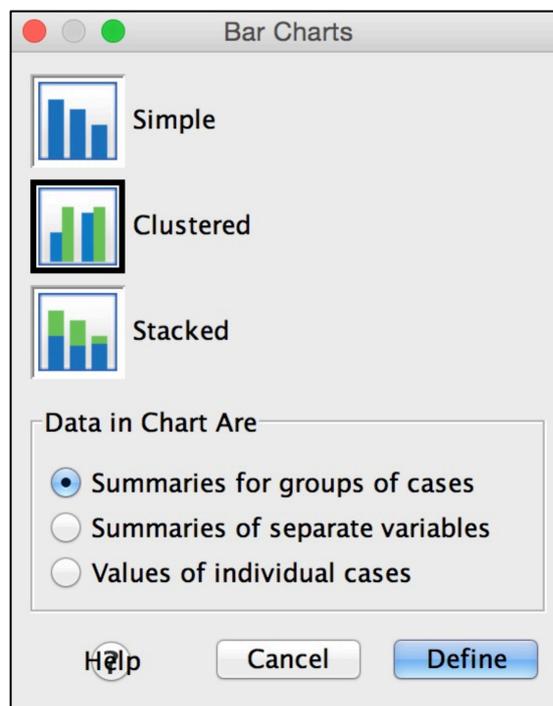
Les variables quantitatives sont à placer dans la case **Bar represent:** et la variable nominale dans **Category axis**. Les barres d'erreur sont dans l'onglet **Options...** comme précédemment décrit.



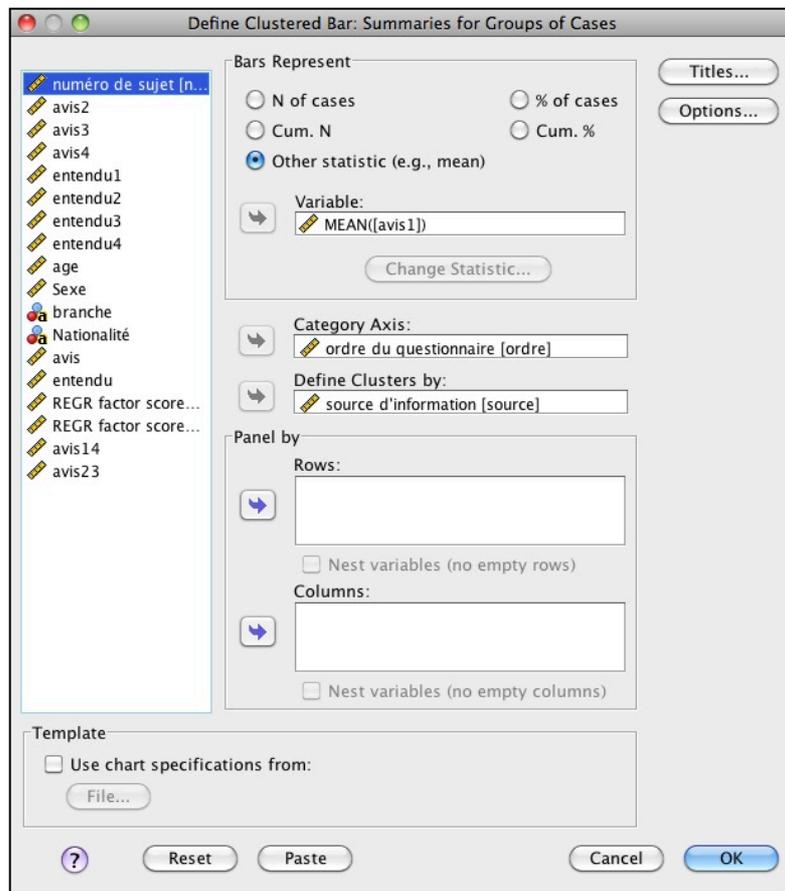
SPSS OUTPUT :



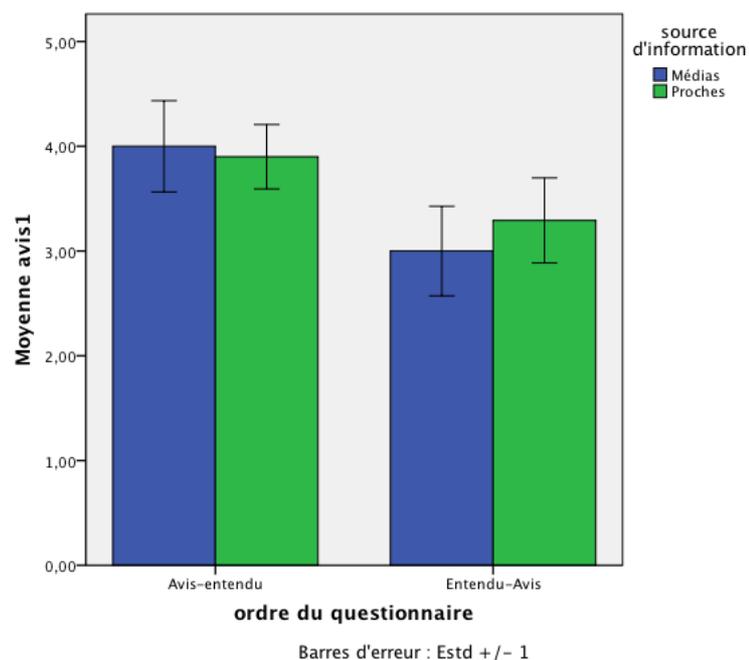
Cas 3 : Une variable quantitative et plusieurs variables nominales. Choisir **Clustered** et **Summaries for groups of cases**.



La variable quantitative est à placer dans la case **Variable:** (choisir généralement **Other statistic (e.g., mean)**) et les variables nominales dans **Category axis:** (axe des x) et **Define clusters by:** (barres de couleur différente).



SPSS OUTPUT :



5.6. Valeurs extrêmes ou aberrantes

5.6.1. Détection par boxplot

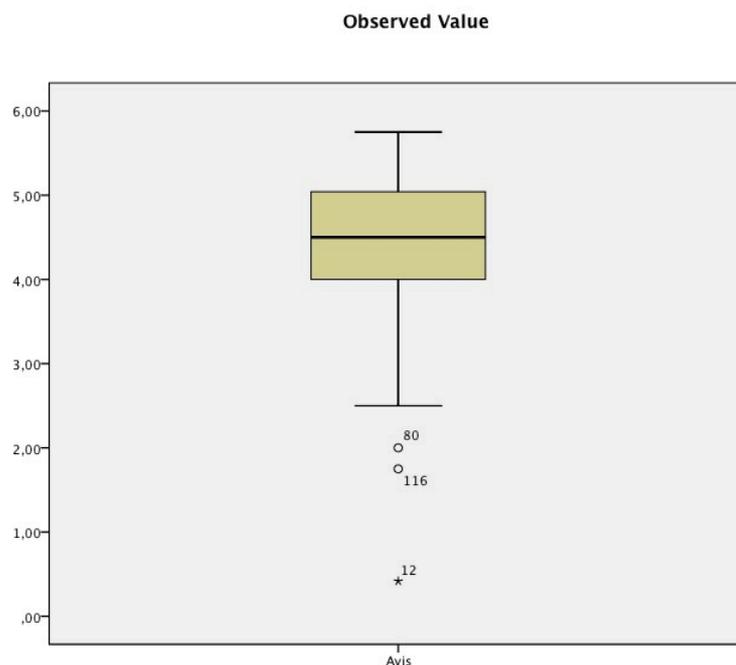
Les valeurs aberrantes (en anglais "extreme values", distantes de plus de 3 intervalles interquartiles au-dessous du 1^{er} quartile ou au-dessus du 3^{ème} quartile) faussent considérablement les données, il est donc judicieux de les éliminer. Dans le boxplot, elles sont représentées par une étoile.

Concernant les valeurs extrêmes (en anglais "outliers", distantes de plus de 1,5 intervalle interquartile au-dessous du 1^{er} quartile ou au-dessus du 3^{ème} quartile), le chercheur ou la chercheuse doit déterminer s'il/elle veut les enlever ou non. Si cela ne fausse pas la normalité de la distribution de la variable (cf. chapitre 5.7. ci-dessous), ou si elles sont trop nombreuses, il faut les garder. Dans le boxplot, elles sont représentées par un rond.

Pour détecter les valeurs extrêmes et aberrantes, il existe plusieurs solutions : soit à l'aide de graphiques (boxplot ou histogramme), soit à l'aide d'un tableau fourni par SPSS. Dans le box-plot, les ronds ou étoiles sont accompagnées du numéro de la **ligne** du sujet dans le fichier de données SPSS (et **pas** son numéro attribué par le chercheur).

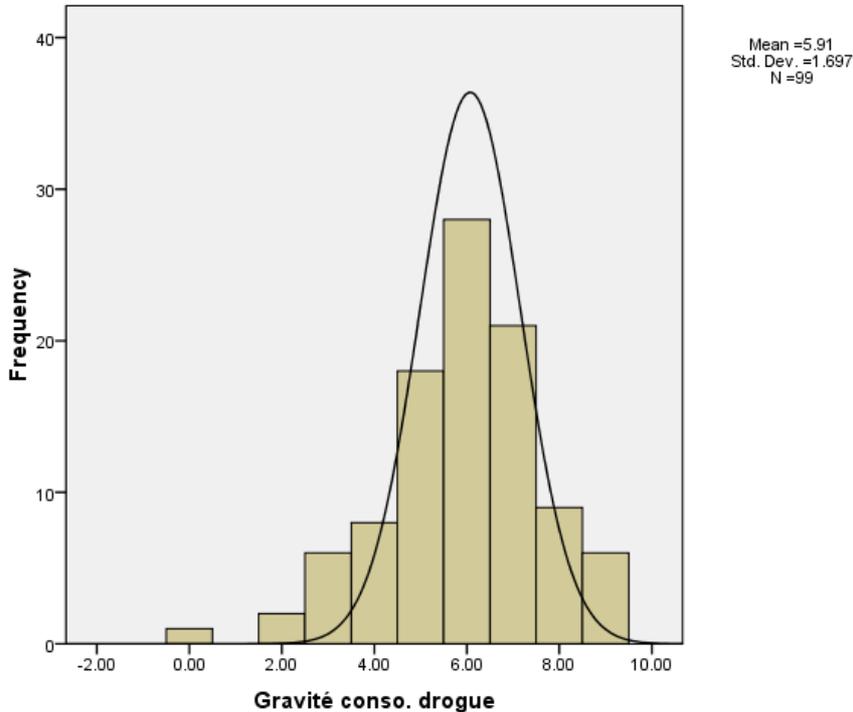
Exemple : dans l'exemple ci-dessous, le sujet représenté par la ligne 12 dans le tableau de données SPSS a une valeur aberrante sur la variable « Avis ». Les sujets représentés par les lignes 80 et 116 dans le fichier de données sont des valeurs extrêmes (mais attention au-dessous de ces sujets il peut y avoir d'autres sujets ayant le même score qui n'apparaîtront pas sur ce graphique, mais sur le tableau **Extreme Values**, cf. pages suivantes).

SPSS OUTPUT :



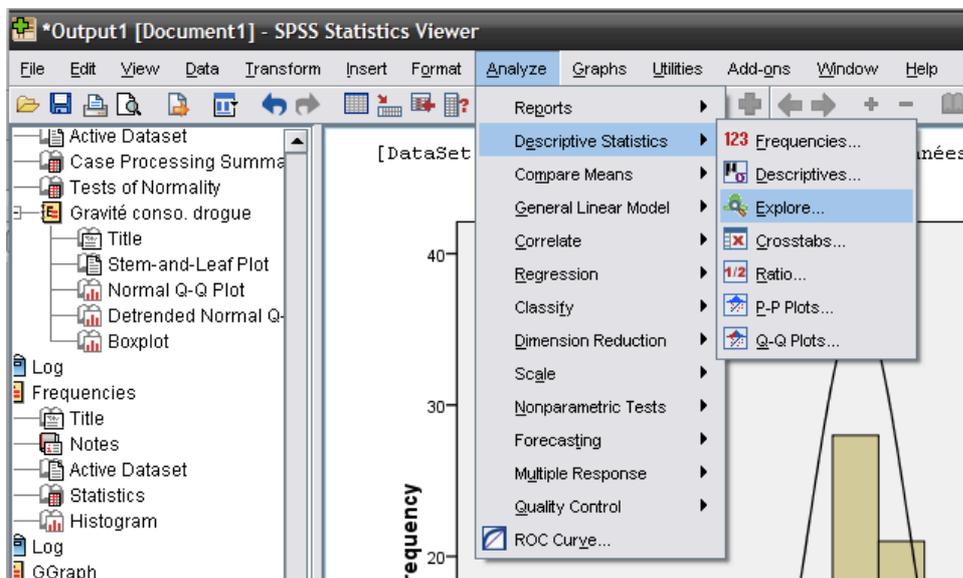
5.6.2. Détection par histogramme

Dans un histogramme, le sujet extrême est représenté par un point éloigné des autres. De plus, si la courbe de normalité est affichée, nous pouvons voir que ce sujet n'est pas sous la courbe, donc qu'il dévie de la normalité.

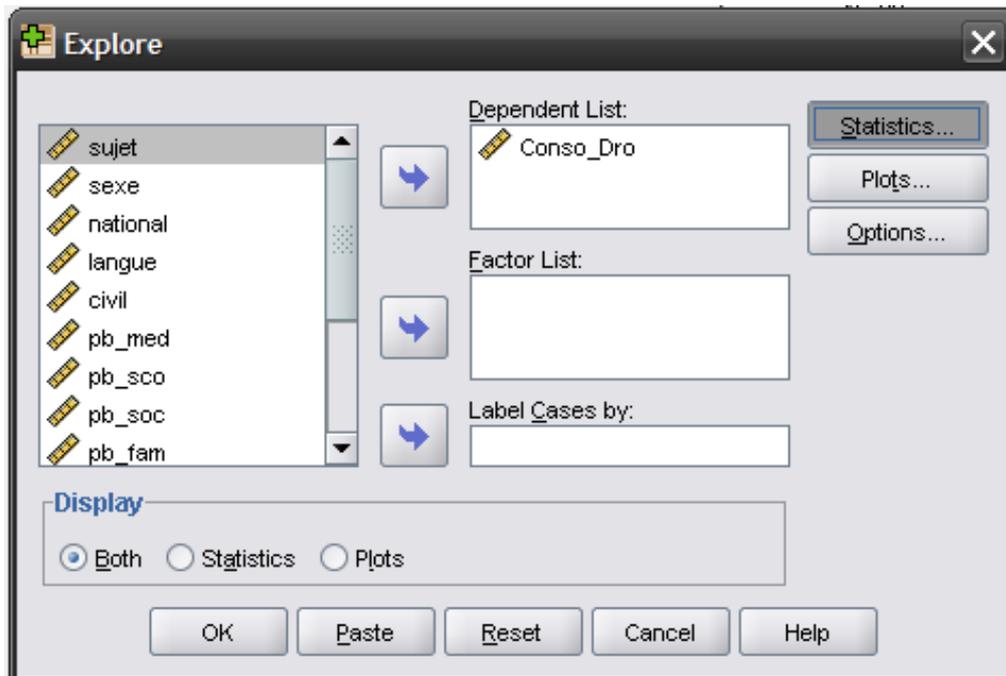


5.6.3. Détection par tableau

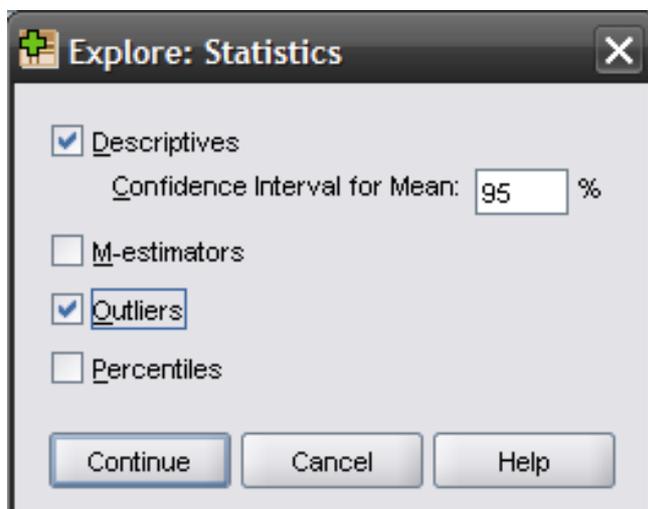
Il est possible de générer un tableau représentant les 5 valeurs les plus élevées et les 5 valeurs les plus basses. Allez sous **Analyse** → **Descriptive Statistics** → **Explore...**



Entrez la variable souhaitée dans **Dependent List:**, puis cliquez sur **Statistics**.



Puis, choisir l'option **Outliers**.



SPSS OUTPUT :

Extreme Values

			Case Number	Value
Gravité conso. drogue	Highest	1	5	9.00
		2	6	9.00
		3	9	9.00
		4	13	9.00
		5	40	9.00 ^a
	Lowest	1	2	.00
		2	76	2.00
		3	55	2.00
		4	97	3.00
		5	90	3.00 ^b

a. Only a partial list of cases with the value 9.00 are shown in the table of upper extremes.

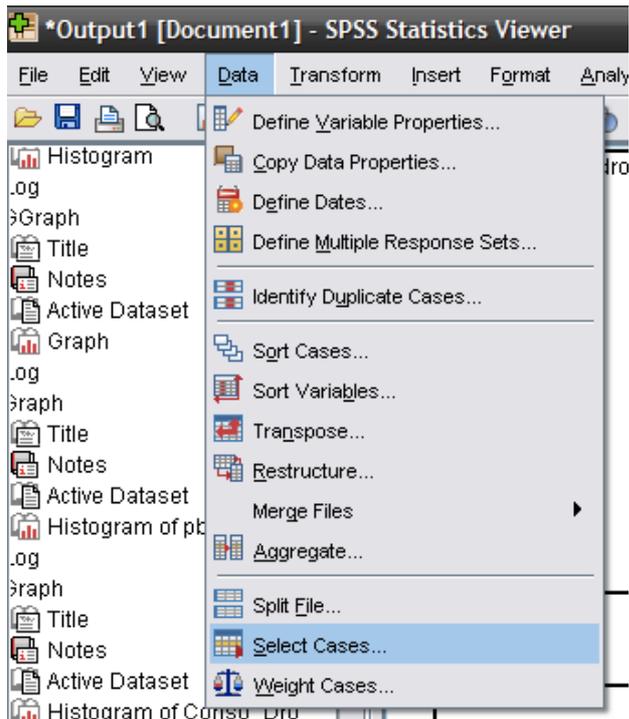
b. Only a partial list of cases with the value 3.00 are shown in the table of lower extremes.

Le tableau **Extreme Values** donne donc les 5 valeurs supérieures et les 5 valeurs inférieures. Il est difficile de déterminer si les valeurs sont extrêmes uniquement à l'aide du tableau, il faut donc s'aider d'un histogramme ou d'un box-plot. L'avantage du tableau est qu'il vous donne le numéro de la **ligne** des sujets dans le fichier de données, ce qui permet de les retrouver afin de les éliminer des analyses. Si plusieurs sujets ont le même score, ils apparaîtront dans ce tableau (mais pas dans le boxplot).

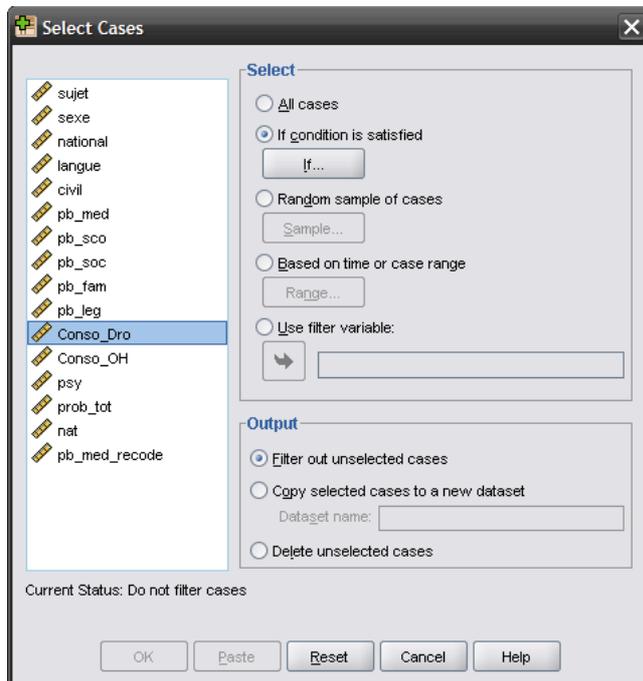
5.6.4. Elimination des analyses

Pour éliminer les valeurs extrêmes ou aberrantes, deux solutions s'offrent à nous. La première manuelle : si ces valeurs sont peu nombreuses, on peut les effacer directement dans le fichier de données (rappel : le numéro figurant dans le boxplot à côté de l'étoile ou du rond, = "case number" dans le tableau "Extreme values", est le numéro de la **ligne** du tableau de données SPSS. Il suffit de se rendre à la colonne de la variable et n'effacer que cette valeur (pas toutes les autres données du sujet qui peuvent être valides). Si les valeurs extrêmes ou aberrantes sont nombreuses, nous pouvons utiliser la commande "Select cases" :

Data → Select Cases

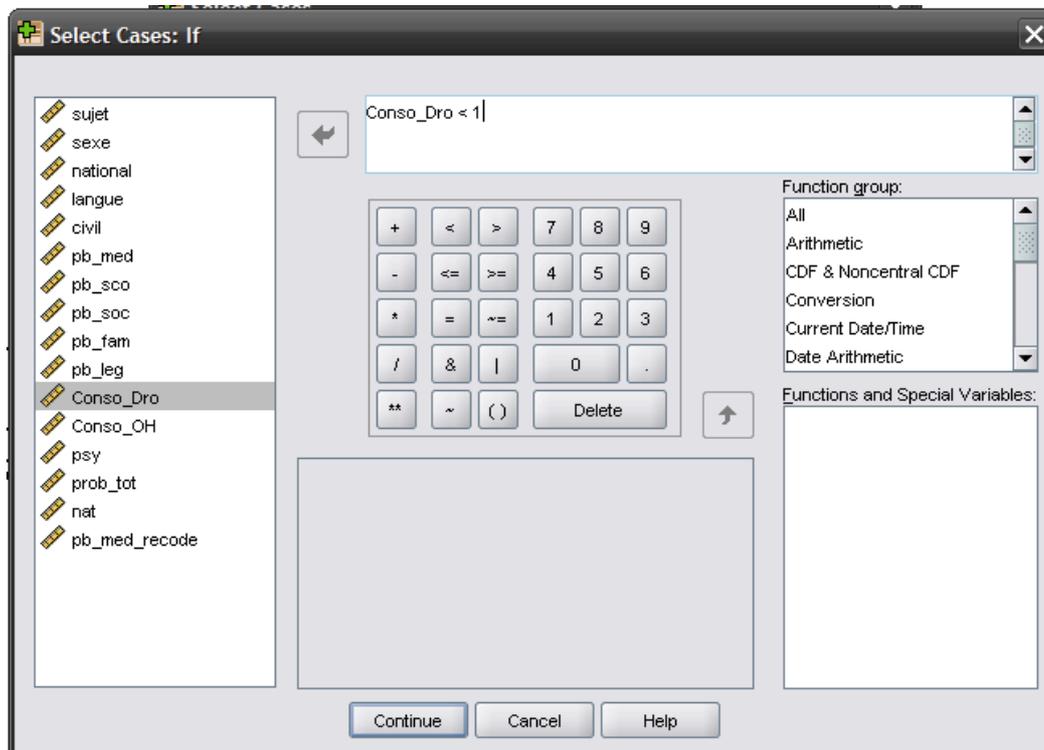


Dans la case **Select**, choisir **If condition is satisfied**, puis cliquer sur **If...**



Dans la nouvelle fenêtre, entrez la variable pour laquelle vous avez détecté une valeur extrême d'un sujet dans la case. Cette fonction sélectionne les participant-e-s à **garder**, donc si, comme dans notre exemple, le sujet extrême a une valeur de 0, nous pouvons entrer $\text{Conso_Dro} > 1$, ce qui le fera choisir tous les sujets ayant une valeur supérieure à 1 pour la variable « Gravité de la consommation de drogue ». Puis pressez **Continue**.

Attention : cette option est à désactiver lorsque vous effectuez des analyses sans cette variable-là. Surtout lorsque vous avez peu de sujets, vous perdez de précieuses données qui ne sont pas extrêmes.



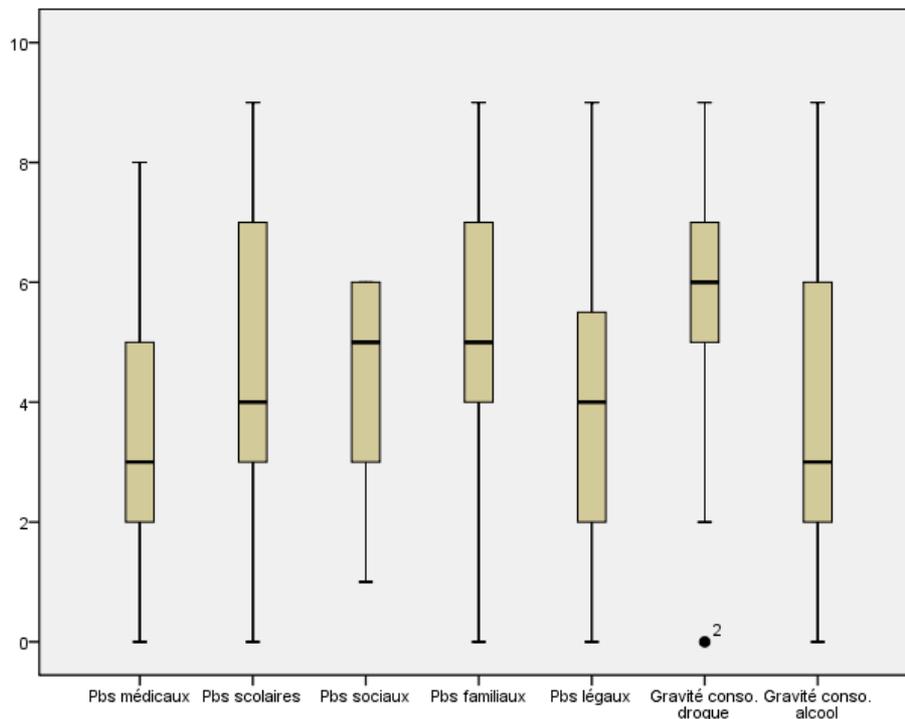
5.7. Examen de la normalité

L'examen de la normalité est une étape importante avant de commencer des analyses statistiques. En effet, les conditions d'application des tests paramétriques nécessitent des variables distribuées normalement.

Il y a plusieurs façons de vérifier la normalité : soit en observant le boxplot et l'histogramme, soit grâce au test de Kolmogorov-Smirnov.

5.7.1. À l'aide du boxplot

Dans **Graphs** → **Legacy Dialogs** → **Boxplot**, nous avons demandé un boxplot pour toutes nos variables. **Attention** aux différences d'échelle ! Ici elles ont toutes la même, ce qui nous permet de toutes les mettre en même temps.



Pour être symétrique (et donc peut-être normale), la médiane doit être à peu près au centre de la boîte, les "moustaches" (antennes) doivent être plus ou moins égales, et il ne doit pas y avoir (trop) de sujets extrêmes. Les critères d'observation de la normalité dans un boxplot sont à juger globalement :

Ici, nous pouvons détecter une asymétrie marquée sur la variable « Pbs sociaux », puisqu'il n'y a pas de « moustache » supérieure. Cette variable n'est donc pas considérée comme distribuée normalement. La variable « Gravité conso. Drogue » non plus ne peut pas être considérée comme parfaitement normale, de par la valeur extrême du sujet 2, mais il est possible de la rendre sa distribution plus normale en éliminant ce dernier des analyses.

5.7.2. À l'aide du test de Kolmogorov-Smirnov

Toujours dans **Analyse** → **Descriptive** → **Explore**, il est possible de demander le test de normalité de Kolmogorov-Smirnov, dans l'onglet **Plots**, en cliquant sur **Normality plots with tests**. Dans ce menu, vous pouvez également demander l'histogramme, et le boxplot.

SPSS OUTPUT :

Tests of Normality

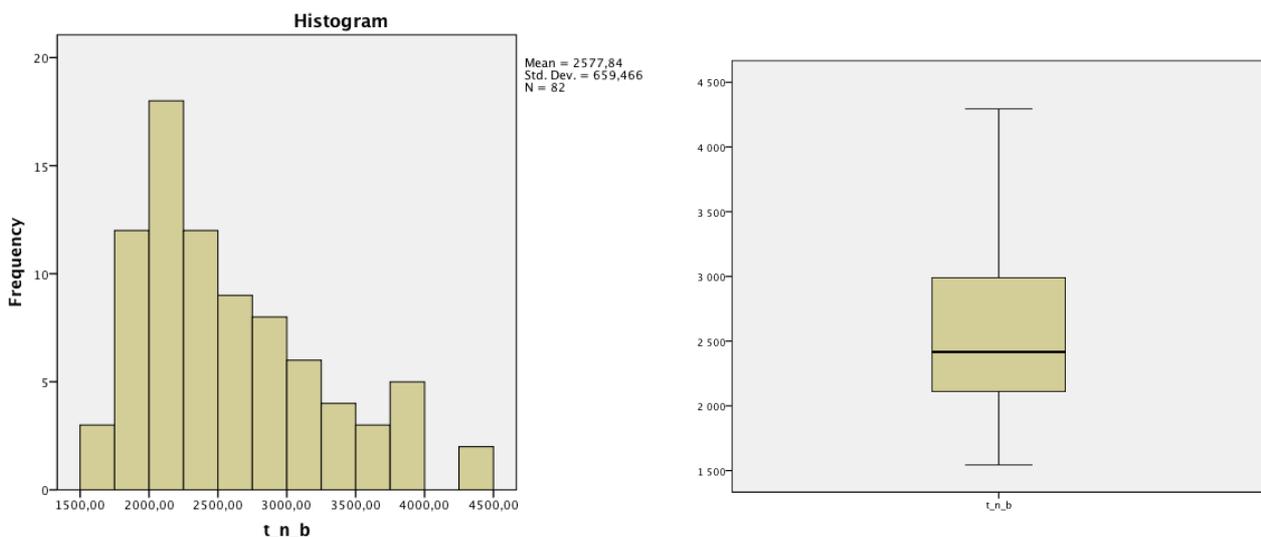
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pbs sociaux	.222	99	.000	.853	99	.000
Pbs médicaux	.178	99	.000	.929	99	.000

a. Lilliefors Significance Correction

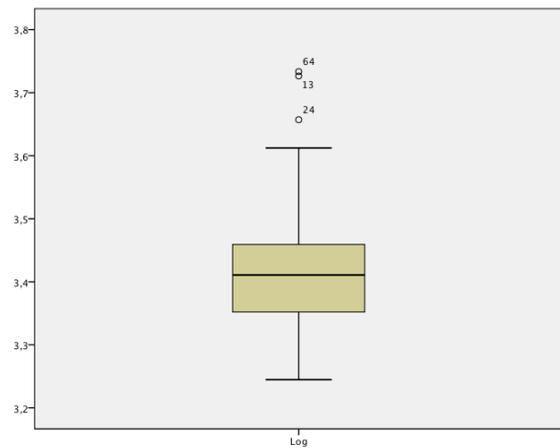
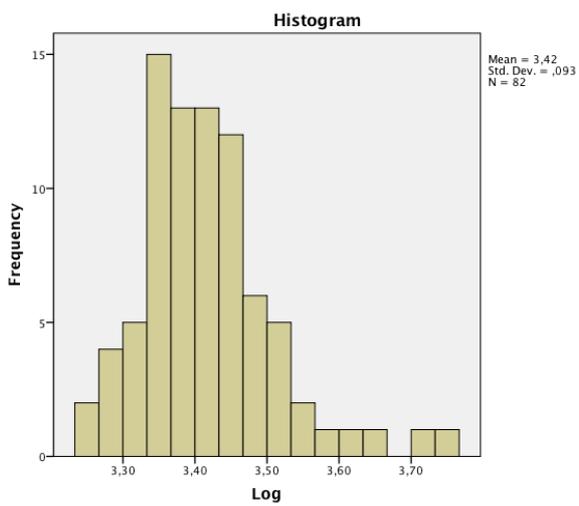
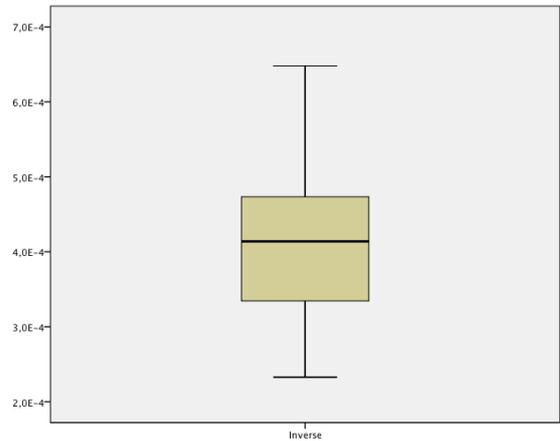
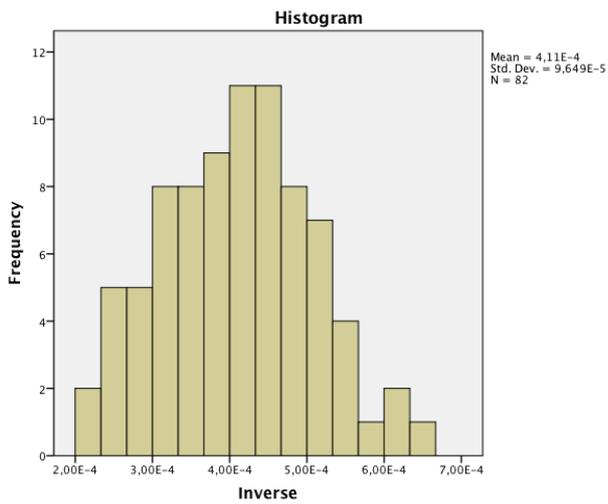
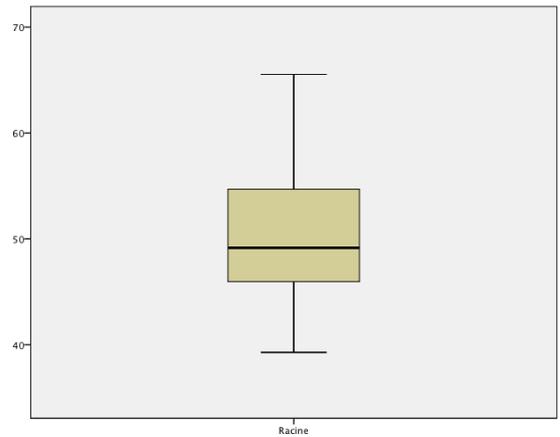
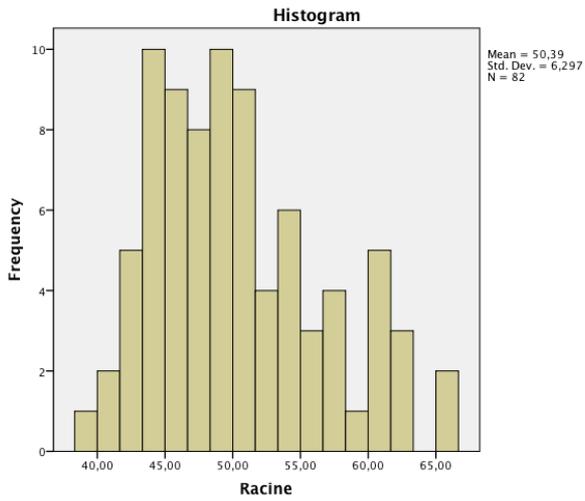
Si le test de Kolmogorov-Smirnov est significatif (donc $p < .05$), cela veut dire que la distribution n'est pas normale (H_0 correspondant à une distribution normale dans la population). Dans notre cas, nous avions pensé que la variable « Pbs sociaux » n'était pas distribuée normalement, mais le test indique que « Pbs médicaux » ne l'est pas non plus. Cependant, il faut utiliser le test de Kolmogorov-Smirnov avec circonspection. En effet, à partir de 50 participant-e-s environ et plus, il a tendance à devenir trop facilement significatif et donc à rejeter l'hypothèse nulle trop facilement. Dès lors, si le test est non significatif, nous pouvons considérer que la variable suit une distribution normale (peu importe le nombre de participant-e-s). Mais si le test est significatif et qu'il y a plus de 50 participant-e-s, il est préférable d'aller voir le boxplot et l'histogramme pour tirer une conclusion quant à la normalité de la distribution. (D'après les boxplots, nous pouvons ici considérer que « Pbs sociaux » n'est pas distribuée normalement, mais que « Pbs médicaux » l'est davantage).

5.8. Transformations mathématiques

Si vous ne pouvez enlever les valeurs extrêmes ("outliers") parce qu'elles sont trop nombreuses (vous perdriez ainsi un trop grand nombre d'observations) ou que la distribution est fortement asymétrique (selon la procédure décrite dans le chapitre précédent 5.7), vous pouvez essayer des transformations mathématiques de tous les scores de votre variable, comme par exemple : racine x , $1/x$, ou encore $\log x$, afin de voir si la nouvelle distribution obtenue est plus proche de la normale. Pour ce faire, vous créez une nouvelle variable à l'aide du menu **Transform** → **Compute variable** (cf. chapitre 4.1), avec les commandes respectivement **SQRT**(nom variable), **1/nom variable**, et **Lg10**(nom variable). Ensuite vous gardez la transformation qui améliore le mieux la normalité de votre variable, en refaisant les analyses décrites au point 5.7. Prenons par exemple la variable « t_n_b » représentant le temps de réaction mesuré dans une tâche informatisée. Sa distribution est fortement asymétrique :



La meilleure des trois transformations est l'inverse $1/x$, la distribution devient plus normale que l'originale :



6. Statistiques inférentielles

Dans le chapitre 5, il a été question de la description de vos données. Dans cette partie, il est question de tester des hypothèses. Avant de présenter les tests, nous proposons un tableau qui vous aidera à décider quelle analyse est la plus adéquate selon le type de variables (nominale, ordinale, ou quantitative) dont il est question dans votre base de données, et selon le rôle de chacune de celles-ci (variable dépendante ou indépendante).

6.1. Tableau de décision statistique

Récapitulation du choix des analyses statistiques (il existe des analyses pour les cases en gris, mais elles ne sont pas abordées dans ce guide).

VI VD	<u>Variable(s) nominale(s)</u> (ex : sexe, section d'études)	<u>Variable(s) quantitative(s)</u> (ex : âge)
<u>Variable nominale</u>	<i>Khi²</i> : (deux variables nominales)	
<u>Variable(s) ordinale(s)</u>	Tests non paramétriques : <i>U de Mann-Whitney</i> (2 groupes indépendants) <i>Kruskal-Wallis</i> (<i>n</i> groupes indépendants) <i>Wilcoxon</i> (2 groupes appariés) <i>ANOVA de Friedman</i> (<i>n</i> groupes appariés)	
<u>Variable(s) quantitative(s)</u>	<i>t-test</i> (comparaison de 2 moyennes) <i>(M)ANOVA</i> (comparaison de <i>n</i> moyennes, une ou plusieurs VI(s), une ou plusieurs VD(s))	<i>Corrélation</i> (deux variables quantitatives. Toutes les variables sont de même niveau, il n'y a pas de VI ou de VD) <i>Régression simple</i> (deux variables quantitatives. VI = prédicteur, VD = critère) <i>Régression multiple</i> (plusieurs variables quantitatives. VIs = plusieurs prédicteurs, VD = critère)

6.2. Khi-carré

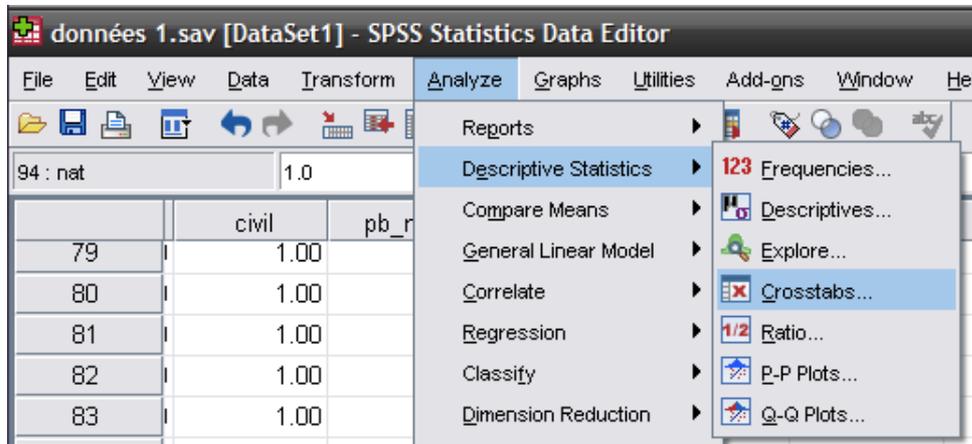
6.2.1. Deux variables nominales

But : Tester s'il y a un lien entre deux variables nominales.

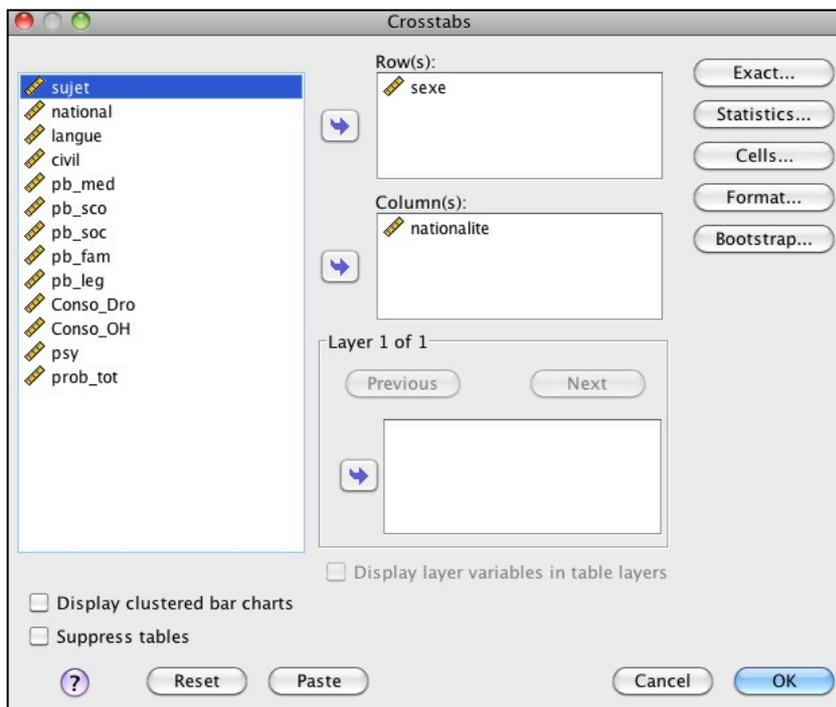
Exemple :

Tester s'il existe une relation entre le fait d'être un homme ou une femme (variable X) et la nationalité (variable Y). H_0 : Il y a la même proportion d'hommes et de femmes chez les suisses et chez les personnes d'autres nationalités (ou il n'y pas de lien entre la variable « sexe » et la variable « nationalité »).

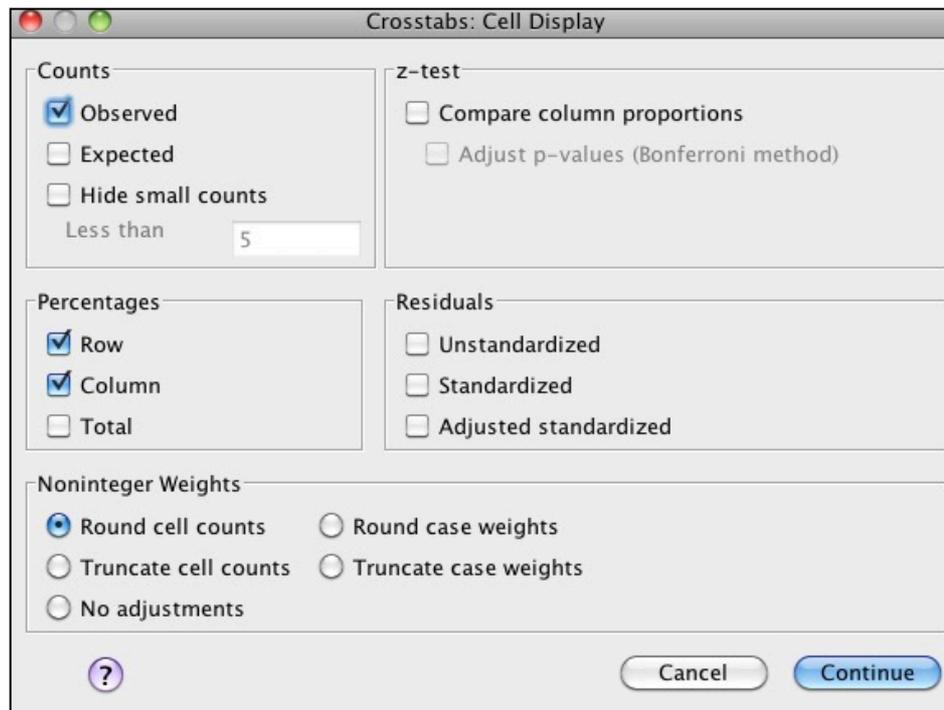
Chemin : **Analyze** → **Descriptive Statistics** → **Crosstabs...**



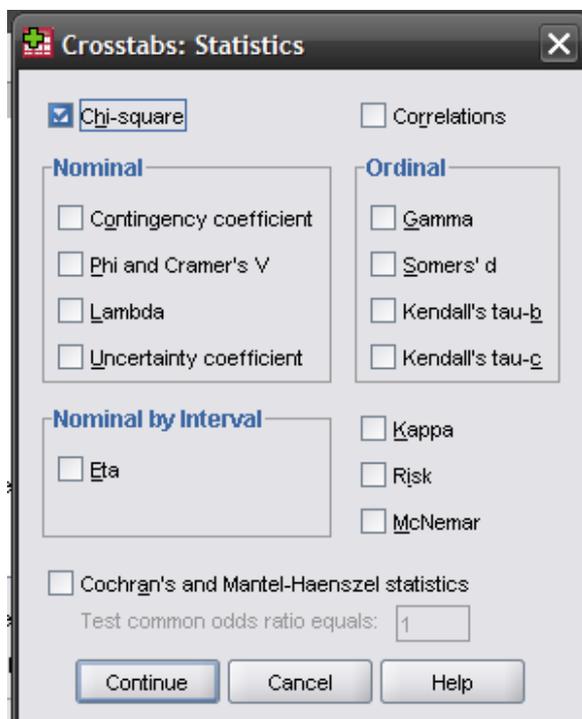
Dans **Row(s)**:, ajoutez une des variables (sexe). Dans **Column(s)**:, ajoutez l'autre variable (nationalite), ou l'inverse, cela ne change rien au résultat.



Dans l'onglet **Cells**, demandez les pourcentages (**Row, Column**).



Dans l'onglet **Statistics** cochez **Chi-square**, puis cliquez sur **Continue**, puis **OK**.



SPSS OUTPUT :

sexe * nationalite Crosstabulation

			nationalite		Total
			suisse	autre	
sexe	homme	Count	48	15	63
		% within sexe	76,2%	23,8%	100,0%
		% within nationalite	65,8%	57,7%	63,6%
	femme	Count	25	11	36
		% within sexe	69,4%	30,6%	100,0%
		% within nationalite	34,2%	42,3%	36,4%
Total		Count	73	26	99
		% within sexe	73,7%	26,3%	100,0%
		% within nationalite	100,0%	100,0%	100,0%

On obtient le tableau croisé avec le nombre et le pourcentage de participant-e-s correspondant à chaque case. Dans la phrase de notation des résultats, vous pouvez utiliser deux pourcentages, entourés de la même couleur dans le tableau ci-dessus (plusieurs possibilités). Dans la phrase de conclusion ci-dessous nous avons choisi les pourcentages en bleu.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,538 ^a	1	,463		
Continuity Correction ^b	,246	1	,620		
Likelihood Ratio	,532	1	,466		
Fisher's Exact Test				,485	,307
Linear-by-Linear Association	,533	1	,465		
N of Valid Cases	99				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 9,45.

b. Computed only for a 2x2 table

Les trois valeurs entourées sont le χ^2 , le degré de liberté et la probabilité associée à ce test. Comme $p > .05$, le test n'est pas significatif (pas de lien entre les deux variables, la proportion d'hommes et de femmes est la même pour les nationalités suisse et autre. La flèche bleue indique la réponse à l'une des conditions d'application du test (ici 0% de valeurs attendues inférieures à 5, donc la condition est remplie).

Notation : "La proportion d'hommes et de femmes n'est pas significativement différente dans les groupes de nationalité suisse (66% d'hommes) et nationalité autre (58% d'hommes), $\chi^2(1) = .54, p > .05$."

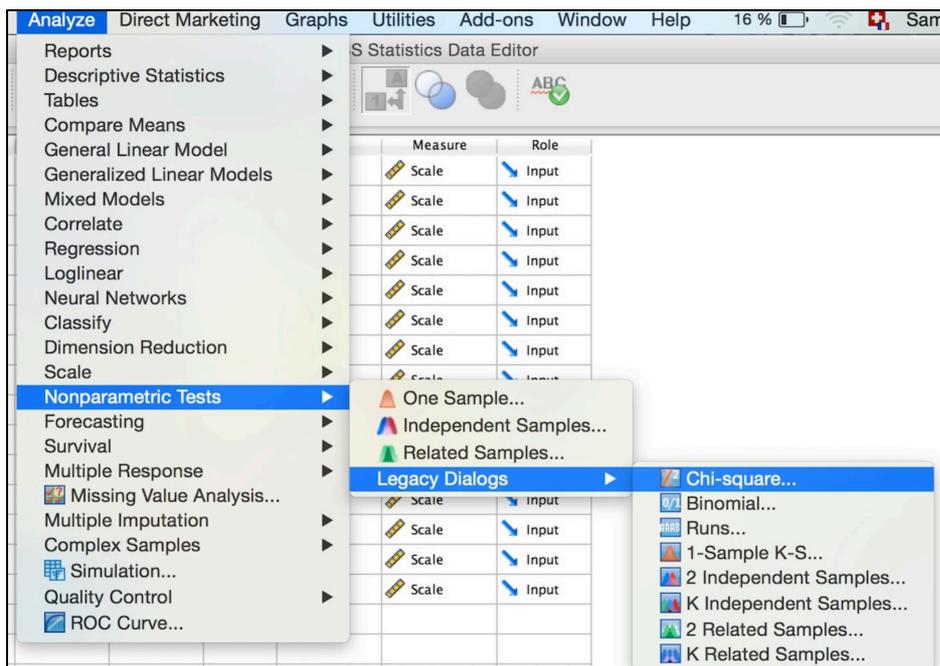
6.2.2. Une variable nominale : Khi-carré d'ajustement

But : Tester si les proportions (fréquences relatives) des modalités d'une variable nominale sont équivalentes (ou égales à certaines valeurs précises).

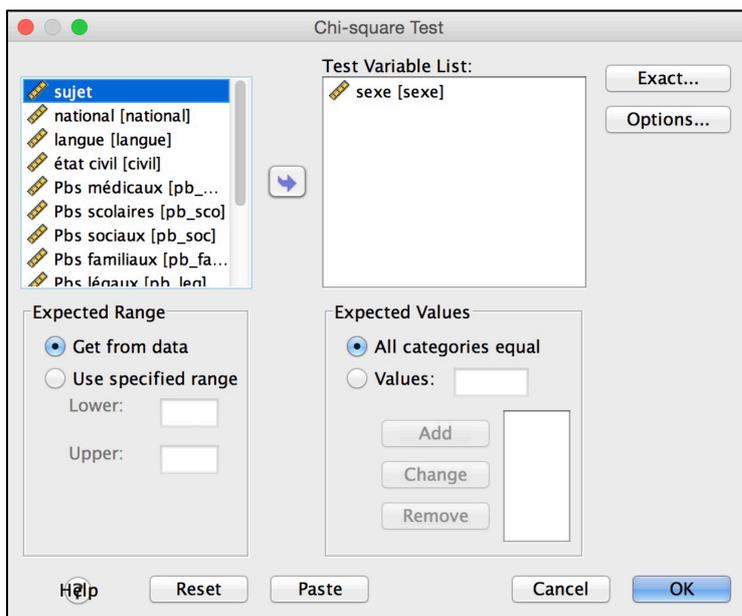
Exemple :

Tester si la proportion des hommes et des femmes est égale (50%-50%) dans la population.

Chemin : **Analyze** → **Nonparametric Tests** → **Legacy Dialogs** → **Chi-square...**



Entrez la variable dans la case **Test Variable List**: puis cliquez sur **OK**.



SPSS OUTPUT :

Chi-Square Test

Frequencies

		sexe		
	Observed N	Expected N	Residual	
homme	63	49,5	13,5	
femme	36	49,5	-13,5	
Total	99			

Test Statistics

		sexe
Chi-Square		7,364 ^a
df		1
Asymp. Sig.		,007

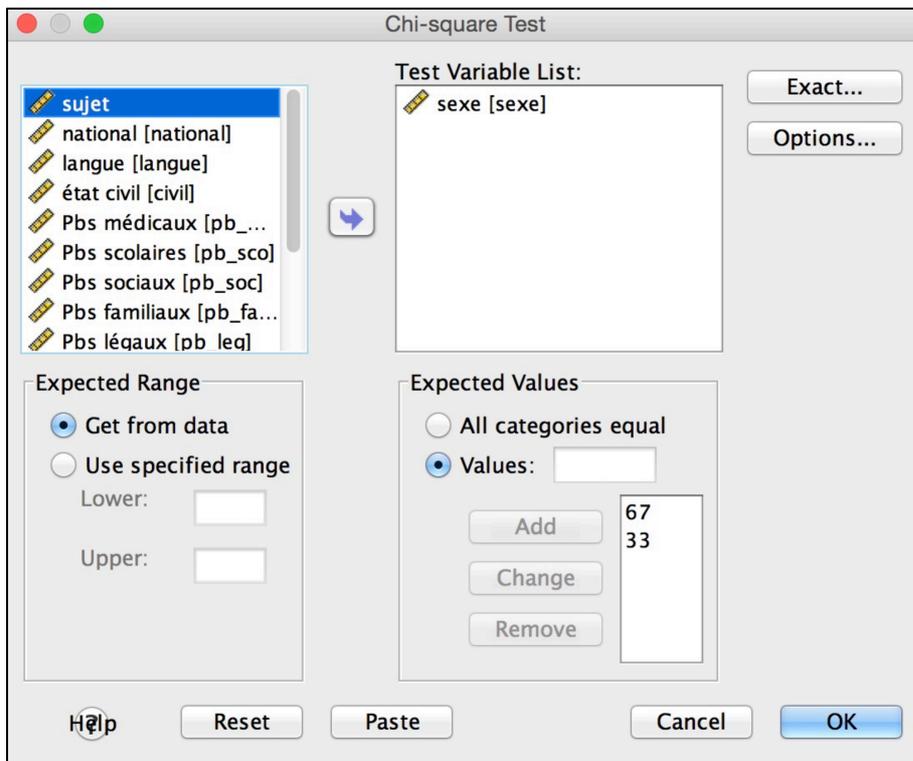
a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 49,5.



Ici comme $p < .05$, le test est significatif, nous rejetons l'hypothèse nulle que l'échantillon proviendrait d'une population où la proportion d'hommes et de femmes est égale (dans l'échantillon, nous avons effectivement plus d'hommes que de femmes, 64% contre 36%). La flèche bleue indique la réponse à l'une des conditions d'application du test (ici 0% de valeurs *attendues* inférieures à 5, donc la condition est remplie).

Notation : "La proportion d'hommes et de femmes est significativement différente (64% d'hommes contre 36% de femmes), $\chi^2(1) = 7.36, p < .05$."

N.B. : Vous pouvez changer la valeur par défaut **All categories equal** (50%-50%, 33%-33%-33%, etc.) si vous voulez tester d'autres proportions. Essayons par exemple de tester l'idée qu'il y aurait deux tiers d'hommes (67%) et un tiers de femmes (33%), en cochant la case **Values** et en entrant les valeurs 67 et 33 puis **Add** et **OK** :



SPSS OUTPUT :

Chi-Square Test

Frequencies

sexe			
	Observed N	Expected N	Residual
homme	63	66,3	-3,3
femme	36	32,7	3,3
Total	99		

Test Statistics

sexe	
Chi-Square	,507 ^a
df	1
Asymp. Sig.	,477

a. 0 cells (0,0%) have expected frequencies less than 5. The minimum expected cell frequency is 32,7.

Ici comme $p > .05$, le test n'est pas significatif, nous pouvons garder l'hypothèse que l'échantillon proviendrait d'une population où la proportion d'hommes et de femmes serait de 2/3-1/3.

Notation : "La proportion d'hommes et de femmes est d'environ deux tiers d'hommes pour un tiers de femmes, $\chi^2(1) = .51$, $p > .05$."

6.3. T-test

But : Tester des différences de moyennes d'une variable quantitative (plusieurs cas possibles : moyenne empirique contre moyenne théorique, moyennes provenant de deux groupes appariés ou indépendants constitués par une variable nominale).

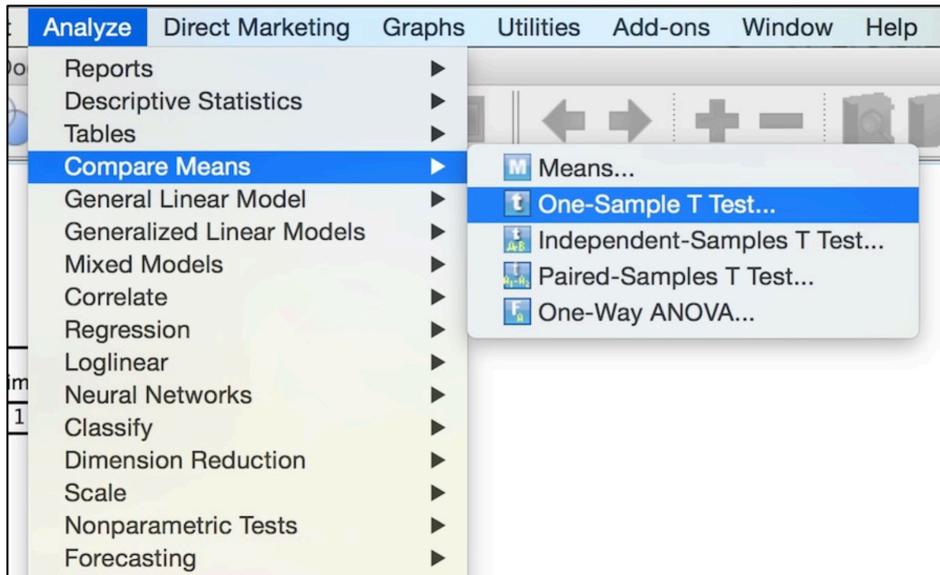
6.3.1. Une moyenne empirique vs. une moyenne théorique

But : Tester s'il y a une différence significative entre une moyenne empirique d'une variable quantitative et une moyenne théorique (fixée d'avance).

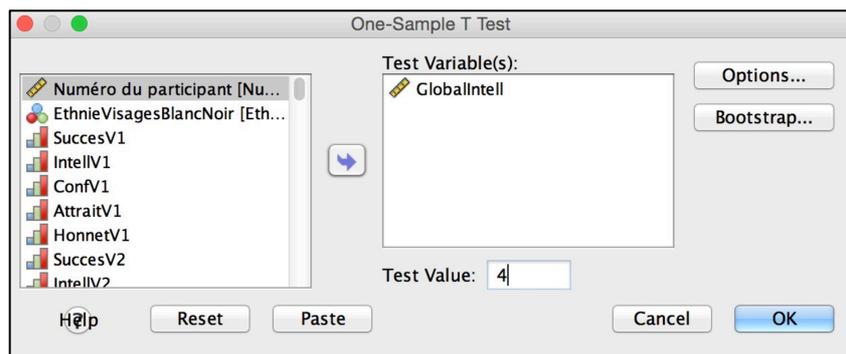
Exemple :

Nous pouvons tester l'hypothèse suivante : est-ce que la moyenne du score global des jugements d'intelligence des visages (VD quantitative) est différente du milieu de l'échelle du questionnaire (échelle de 1 à 7, milieu = 4) ?

Chemin : **Analyze** → **Compare Means** → **One-Sample T Test...**



Déplacez votre VD (pour notre exemple, « GlobalIntell ») de la liste de gauche vers l'onglet **Test Variable(s)**: à l'aide de la petite flèche du haut. Spécifiez la moyenne théorique dans la case **Test Value** (4 dans notre exemple). Cliquez sur **OK**.



SPSS OUTPUT :

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
GlobalIntell	103	4,6569	,72478	,07141

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
GlobalIntell	9,198	102	,000	,65690	,5152	,7986

Dans le premier tableau se trouvent les indices descriptifs, le deuxième tableau donne le résultat du test : la valeur du t est de 9.20, et la valeur du p associée est

de .000 (en réalité une valeur $<.001$), ce qui signifie que le test est significatif, et donc que la différence entre les deux moyennes est significative.

Notation : La moyenne du score global des jugements d'intelligence des visages est significativement différente du milieu de l'échelle du questionnaire $t(102) = 9.20$, $p < .05$.

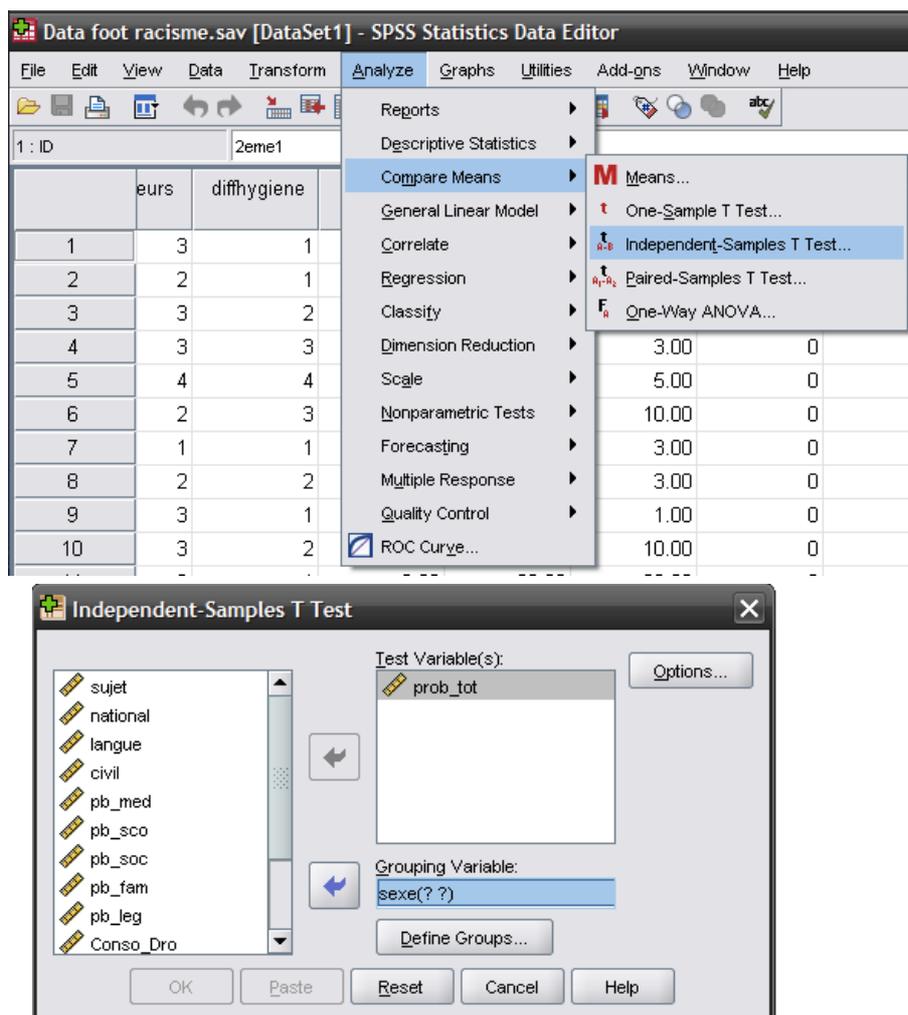
6.3.2. T-test à groupes indépendants

But : Tester s'il y a une différence significative entre deux moyennes (provenant de deux groupes formés par une variable nominale = VI) d'une variable quantitative (VD).

Exemple :

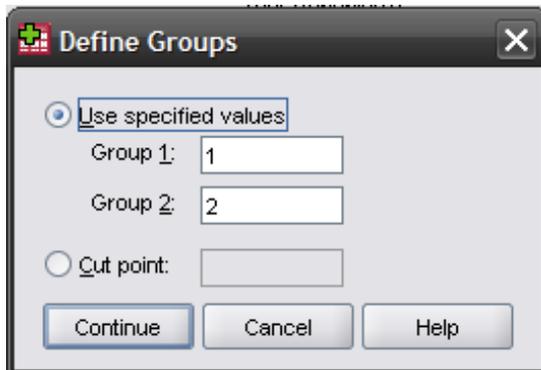
Nous pouvons tester l'hypothèse suivante : les femmes ont moins de problèmes que les hommes (H_1). Notre VI est donc le sexe (variable nominale à 2 modalités) des participant-e-s et notre VD est l'indice total de problèmes (variable quantitative).

Chemin : **Analyze** → **Compare Means** → **Independent Samples T Test...**



Déplacez votre VD (pour notre exemple, « prob_tot ») de la liste de gauche vers l'onglet **Test Variable(s)** à l'aide de la petite flèche du haut. Déplacez ensuite votre VI (pour notre exemple, « sexe ») de la liste de gauche vers l'onglet **Grouping Variable(s)** à l'aide de la petite flèche du bas.

Le logiciel vous demande alors de définir les groupes de votre VI dont vous voulez comparer les moyennes. Cliquez sur **Define Groups...**



Une fenêtre apparaît alors et vous devez spécifier que le groupe 1 correspond à la modalité 1 de votre VI (les hommes) et que le groupe 2 correspond à la modalité 2 de votre VI (les femmes). Lorsque vous entrez vos données dans SPSS, il vous est tout à fait possible de donner d'autres valeurs aux groupes, comme 0 et 1, 1 et 2, etc. C'est pour cela que le logiciel vous demande de spécifier les valeurs attribuées à vos données.

Cliquez sur **Continue** puis **OK** :

SPSS OUTPUT :

		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
prob_tot	Equal variances assumed	.055	.815	-1.818	97	.072	-.59286	.32615	-1.24018	.054
	Equal variances not assumed			-1.867	78.979	.066	-.59286	.31755	-1.22492	.039

Dans ce premier tableau se trouvent les informations qui concernent la significativité de la différence entre les deux moyennes.

Le résultat du test **Levene's test for equality of variance** vous indique quelle ligne il faudra consulter pour vos résultats :

- 1) Si le **F** est **non significatif** (ici $F = .055$, $p = .82$, *ns.*), on regarde les résultats sur la ligne **Equal variances assumed**.
- 2) Si le **F** est **significatif** (il faudrait que $p < .05$, ce qui n'est pas le cas ici), on regarde les résultats de la ligne **Equal variances not assumed**.

Dans notre exemple, on considère la première ligne (**Equal variances assumed**). Dans la colonne **t** on peut lire la valeur -1.82 et dans la colonne **Sig (2-tailed)** on peut lire .072. Cela veut dire que la différence entre hommes et femmes concernant

le nombre de problèmes n'est pas significative. Les degrés de liberté sont notés dans la colonne **df**.

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
prob_tot	homme	63	4.1905	1.61342	.20327
	femme	36	4.7833	1.46375	.24396

Dans ce second tableau (mais qui arrive en premier dans l'output SPSS), vous pouvez lire les effectifs pour chaque groupe (63 hommes et 36 femmes), ainsi que les moyennes correspondantes. Comme le test indique seulement si les deux moyennes sont significativement différentes l'une de l'autre et non pas laquelle est la plus élevée, il faut considérer ce tableau.

Notation : La moyenne du nombre total de problèmes des hommes ($M = 4.19$, $SE = .20$) et celle des femmes ($M = 4.78$, $SE = .24$) ne diffère pas significativement $t(97) = -1.82$, $p > .05$.

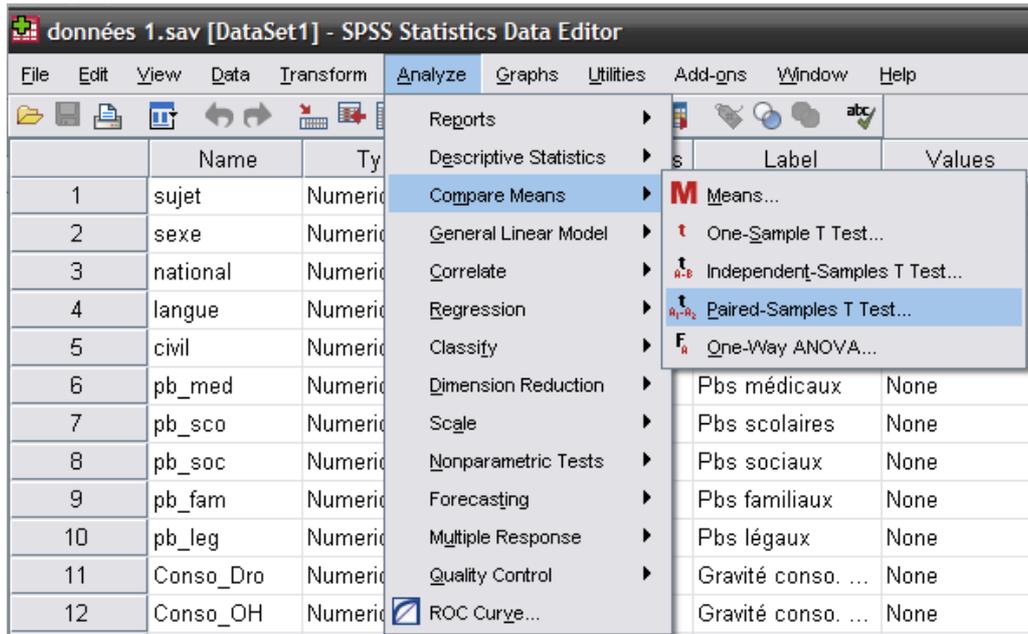
6.3.3. T-test à mesures répétées

But : Tester s'il y a une différence significative entre deux moyennes mesurées chez les mêmes sujets. Dans cette analyse, il est question de deux variables dépendantes (VD) quantitatives, sans aucune variable indépendante (ou alors, on peut considérer qu'il y a une VI à mesures répétées, et une VD).

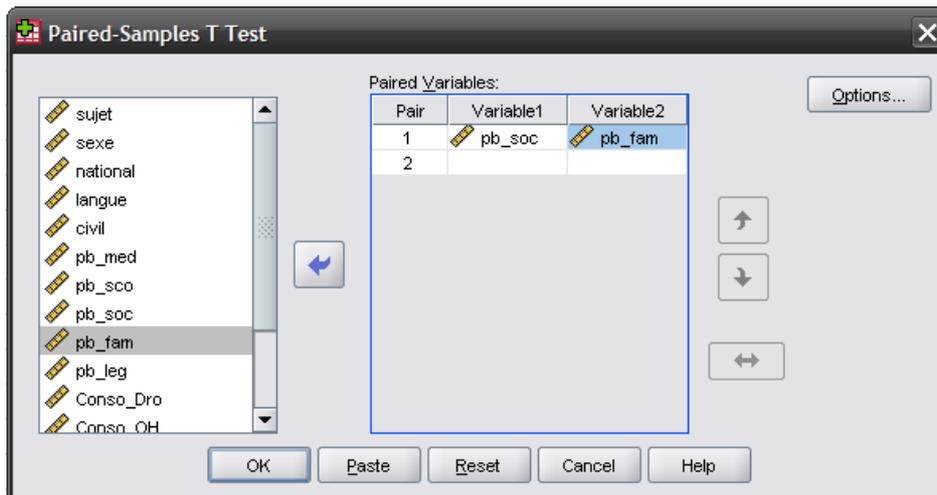
Nous prenons, par exemple, les problèmes sociaux et familiaux des participant-e-s, mais nous voulons seulement savoir si les participant-e-s ont des niveaux différents pour ces deux types de problèmes (on peut considérer qu'il s'agit de 2 VDs, ou alors une VI qui est le type de problèmes à deux modalités, sociaux et familiaux, et une VD qui est le nb de problèmes).

L'hypothèse qui nous intéresse est la suivante : les participant-e-s ont plus de problèmes familiaux que sociaux.

Chemin : **Analyze** → **Compare Means** → **Paired-Samples T Test...**



Entrez vos deux variables dans la case **Paired Variables** (elles se mettent automatiquement dans les colonnes Variable 1 et Variable 2), puis cliquez sur **OK**.



SPSS OUTPUT :

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pbs sociaux - Pbs familiaux	-.52525	1.74580	.17546	-.87345	-.17706	-2.994	98	.003

Vous trouvez dans le tableau **Paired Samples Test** le résultat du test. La valeur du t est de -2.99, et la valeur du p associée est de .003, ce qui signifie que le test est significatif, et donc que la différence entre les moyennes des deux VD est

significative. Ensuite, vous allez regarder les moyennes dans le tableau **Paired Samples Statistics** pour savoir dans quel sens va la différence.

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pbs sociaux	4.8384	99	2.05390	.20642
	Pbs familiaux	5.3636	99	2.24270	.22540

Notation : Les sujets ont dans l'ensemble plus de problèmes familiaux ($M = 5.36$, $SD = 2.24$) que de problèmes sociaux ($M = 4.84$, $SD = 2.05$), $t(98) = -2.99$, $p < .05$.

6.4. ANOVA

But : Tester des différences de moyennes d'une ou plusieurs variable(s) quantitative(s) (VDs), en fonction d'une ou plusieurs variable(s) nominale(s) (= facteurs, VIs) qui peuvent comporter plus de deux modalités, contrairement au t-test. C'est en quelque sorte une extension du t-test.

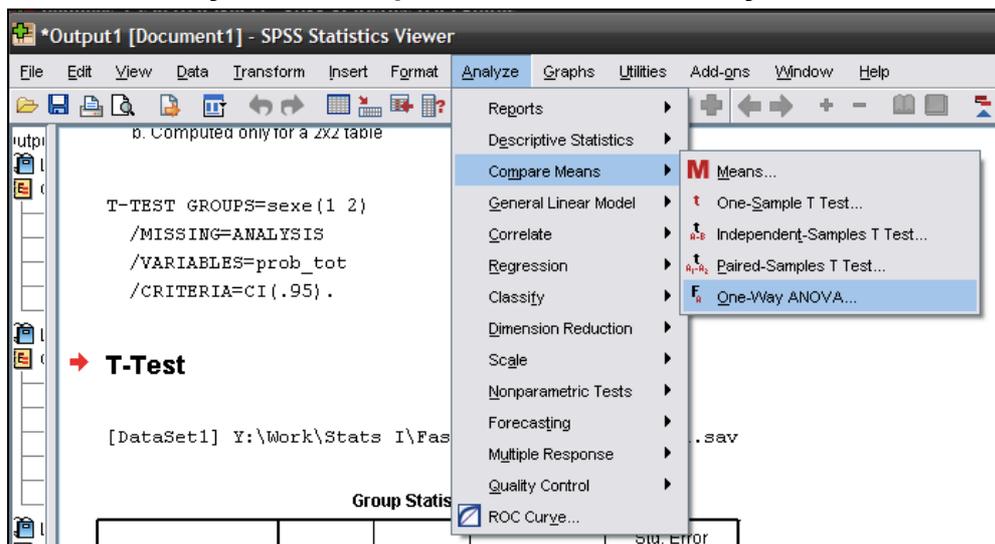
6.4.1. ANOVA à un facteur intersujets (groupes indépendants)

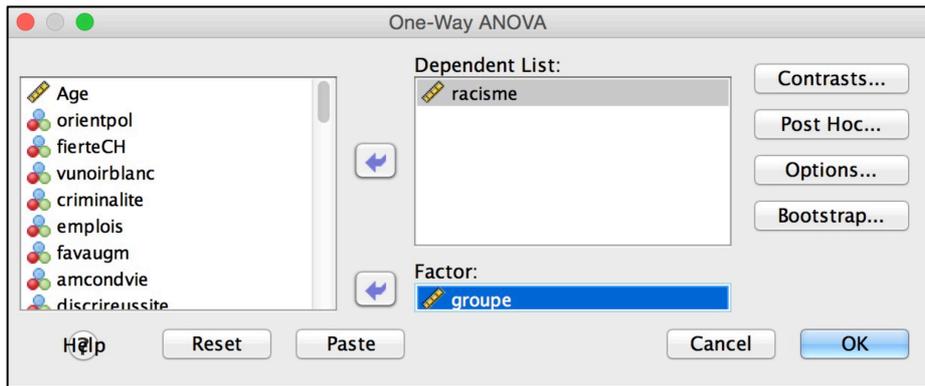
But : Tester s'il y a une différence significative entre plusieurs (2 ou plus) moyennes d'une variable quantitative (VD) provenant de plusieurs groupes formés par une variable nominale (VI = facteur).

Exemple :

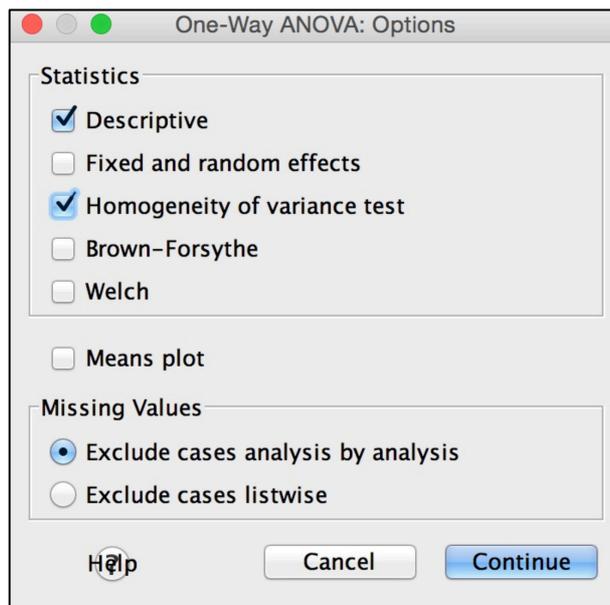
H_1 : Les joueurs, arbitres et fans de football ont un score moyen de racisme différent. Notre VI est donc le groupe des participant-e-s (variable nominale à 3 modalités = facteur) et notre VD est le score de racisme (variable quantitative).

Chemin : **Analyze** → **Compare Means** → **One-Way ANOVA...**





Entrez la variable quantitative dans la case **Dependent List:** et la variable nominale dans la case **Factor:** Puis cliquez sur **Options...** Une nouvelle boîte de dialogue s'ouvre et vous pouvez demander que les statistiques descriptives et le test de Levene (homogénéité des variances) apparaissent dans l'output. Sélectionnez les cases **Descriptive** et **Homogeneity of variance test.** Puis **Continue** et la deuxième boîte de dialogue se ferme. Cliquez sur **OK.**



SPSS OUTPUT :

ANOVA

racisme					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,731	2	,866	2,583	,082
Within Groups	25,140	75	,335		
Total	26,872	77			

Descriptives

racisme								
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
footballeurs	42	2,1786	,59542	,09188	1,9930	2,3641	1,00	3,50
arbitres	14	2,1964	,51121	,13663	1,9013	2,4916	1,50	3,00
fans	22	1,8523	,58584	,12490	1,5925	2,1120	1,00	3,25
Total	78	2,0897	,59075	,06689	1,9566	2,2229	1,00	3,50

Le F qui nous intéresse dans le tableau est celui qui est associé à notre facteur intergroupe (**Between Groups**). Dans notre exemple, il est de 2.58, avec une probabilité p associée de .082.

Dans le tableau **Descriptives** se trouvent les moyennes des trois groupes sur la VD (dans notre exemple, le score de racisme).

Test of Homogeneity of Variances

racisme			
Levene Statistic	df1	df2	Sig.
,202	2	75	,818

Le tableau **Test of Homogeneity of Variances** vous indique si la condition d'application de l'ANOVA (égalité des variances dans les groupes) est respectée. Ici le test est non significatif, donc les variances sont homogènes, $F(2, 75) = .20, p > .05$.

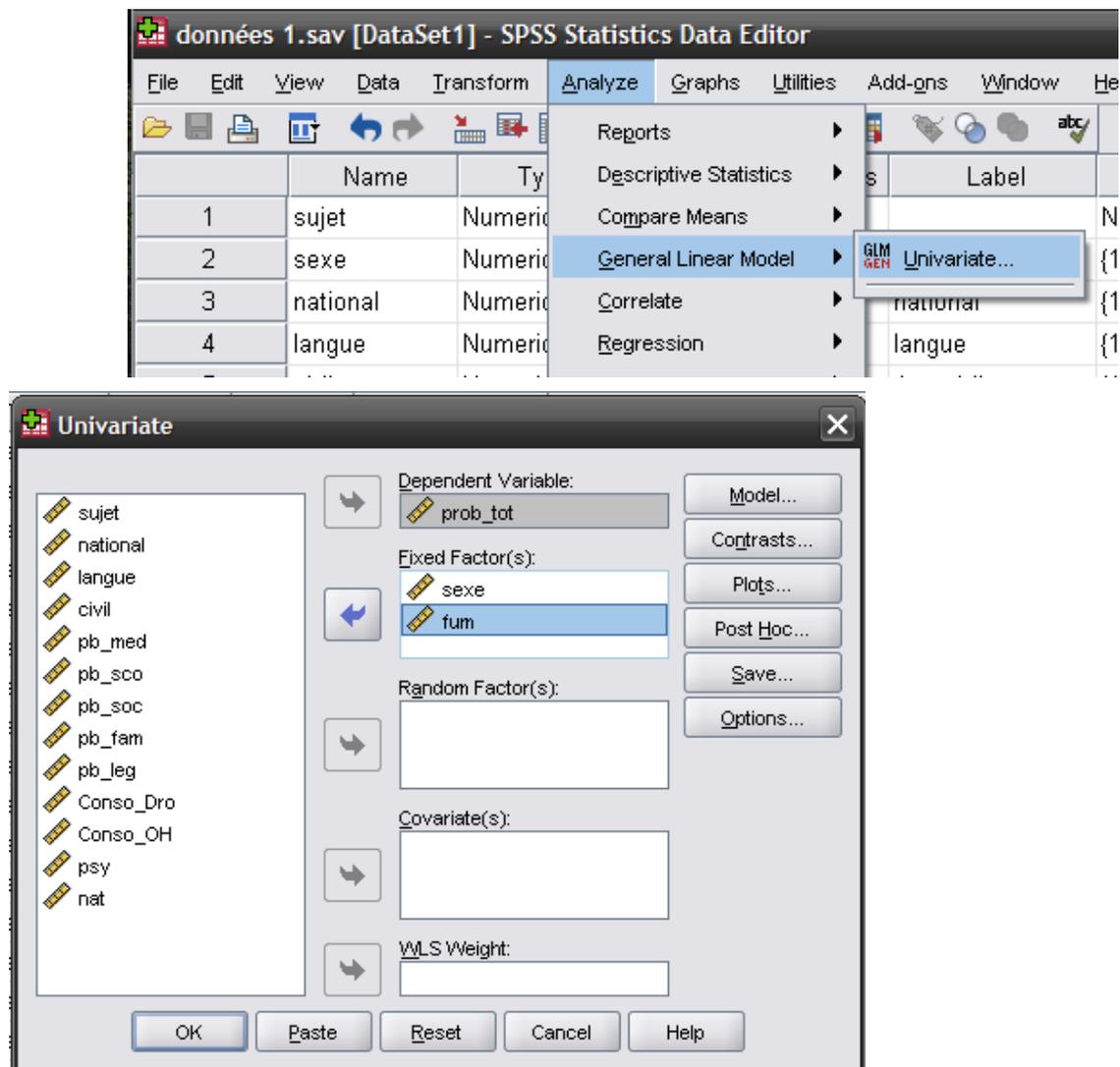
Notation : Les footballeurs ($M = 2.18, SD = .59$) et les arbitres ($M = 2.20, SD = .51$) ont un score de racisme supérieur aux fans ($M = 1.85, SD = .59$), mais la différence n'est pas statistiquement significative, $F(2, 75) = 2.58, p > .05$.

6.4.2. ANOVA à plusieurs facteurs intersujets (groupes indépendants)

But : Tester s'il y a une différence significative entre plusieurs moyennes d'une variable quantitative (VD) provenant de plusieurs groupes formés par plusieurs variables nominales (VIs = facteurs). La différence par rapport au test présenté au chapitre précédent est que vous évaluez l'effet de plusieurs VI à la fois sur la même VD et non plus l'effet d'une seule VI. Le test nous donne ainsi les effets principaux de chaque VI, ainsi que le ou les effets d'interaction entre VIs.

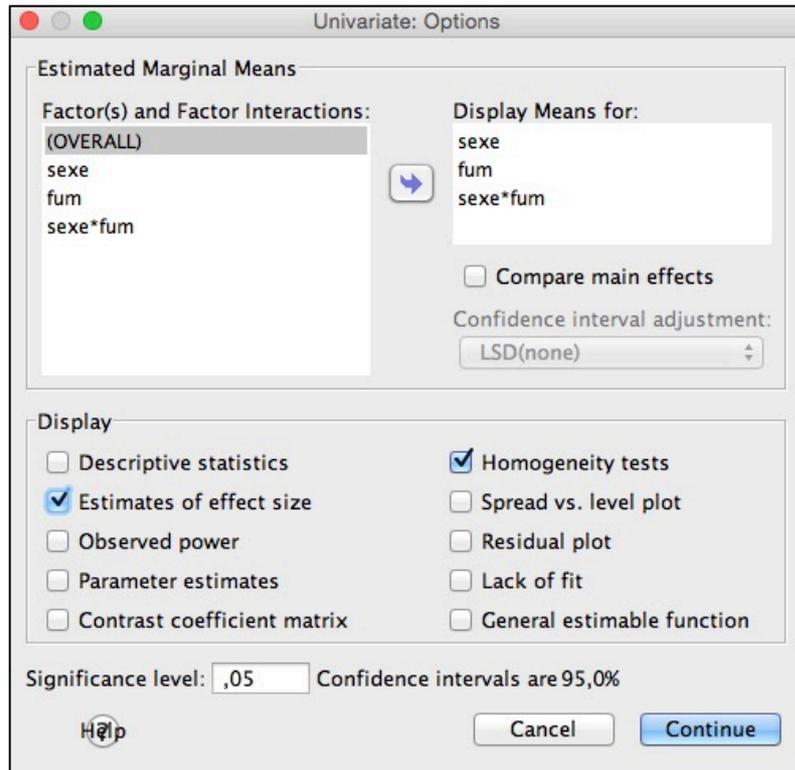
Exemple : Nous cherchons à savoir s'il existe un effet conjoint du sexe (VI1) et du fait de fumer (VI2) sur le nombre total de problèmes chez des adolescent-e-s. Nos deux VI sont donc le sexe et le statut de fumeur, toutes deux nominales. Notre VD est le nombre total de problèmes, et est une variable quantitative.

Chemin : **Analyze** → **General Linear Model** → **Univariate...**



Entrez la variable quantitative (VD) dans la case **Dependent Variable:** et les variables nominales (VI) dans la case **Fixed Factor(s):** (remarque : si vous ne mettez qu'un facteur dans cette case, l'analyse sera la même qu'avec **One-Way ANOVA**).

Dans **Options**, demandez les moyennes pour chaque effet testé (**Display Means for:** Dans notre exemple « sexe », « fum » et l'effet d'interaction « sexe*fum »). Demandez également le test d'homogénéité des variances **Homogeneity tests** (test de Levene), ainsi que les tailles d'effet **Estimates of effects sizes**, puis cliquez **Continue** puis **OK**.



L'analyse donne à la fois les effets principaux comme auparavant (différences éventuelles entre hommes et femmes, et entre les fumeurs et les non-fumeurs) et en plus l'effet d'interaction entre les deux variables (par exemple différence éventuelle entre hommes et femmes, mais seulement parmi les fumeurs, ou seulement parmi les non fumeurs). Un effet d'interaction représente le croisement des effets du sexe et du statut de fumeur sur la VD. Par exemple, il se peut qu'il n'y ait pas de différence significative entre les hommes et les femmes, ni entre les fumeurs et les non-fumeurs, mais que les femmes fumeuses se différencient des femmes non-fumeuses, et les hommes fumeurs ne se différencient pas des hommes non-fumeurs. Plusieurs cas de figure sont possibles. Vous pouvez formuler des hypothèses sur les effets principaux et sur les effets d'interaction, cela dépend de votre cadre théorique.

SPSS OUTPUT :

Between-Subjects Factors

		Value Label	N
sexe	1,00	homme	63
	2,00	femme	36
fum	,00	non-fumeur	17
	1,00	fumeur	82

Levene's Test of Equality of Error Variances^a

Dependent Variable: prob_tot

F	df1	df2	Sig.
,743	3	95	,529

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + sexe + fum + sexe * fum

Le premier tableau donne une indication des codes utilisés pour identifier les groupes et leurs labels (homme ou femme, fumeur ou non-fumeur). Le second tableau vous renseigne au sujet d'une condition d'application de l'ANOVA, l'homogénéité des variances dans les groupes. Ici, l'homogénéité peut être assumée, $F(3, 95) = .74, p > .05$.

Tests of Between-Subjects Effects

Dependent Variable: prob_tot

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	37,902 ^a	3	12,634	4,714	,004	,130
Intercept	595,612	1	595,612	222,240	,000	,701
sexe	1,738	1	1,738	,648	,423	,007
fum	18,522	1	18,522	6,911	,010	,068
sexe * fum	2,794	1	2,794	1,042	,310	,011
Error	254,604	95	2,680			
Total	2001,270	99				
Corrected Total	292,505	98				

a. R Squared = ,130 (Adjusted R Squared = ,102)

Le troisième tableau est nommé **Tests of Between-Subjects Effects**. Dans ce tableau, on trouve les effets, sur la VD, de chaque variable indépendante (effets principaux « sexe » et « fum »), puis de l'interaction (« sexe*fum »).

Sur la ligne « sexe », vous avez l'effet principal du sexe sur la VD avec le F et le p associé, on voit ici que le test n'est pas significatif, donc que le sexe n'a pas d'influence sur le degré de problèmes totaux.

Sur la ligne « fum », vous avez l'effet du statut de fumeur sur la VD (avec F et p). Le test est quant à lui significatif, ce qui veut dire qu'il y a une différence entre fumeurs et non-fumeurs par rapport au nombre total de problèmes rencontrés.

Sur la ligne « sexe*fum », vous avez l'effet d'interaction entre des deux variables, auquel est associé un F et un p . Le test n'est pas non plus significatif, il n'y a donc pas d'effet d'interaction entre ces deux variables.

En plus de ces 3 tests d'inférence, il convient pour les interpréter de regarder les tableaux de moyenne sur lesquels portent les tests, qui sont données pour chacun des trois effets testés dans **Estimated Marginal Means**.

Estimated Marginal Means

1. sexe

Dependent Variable: prob_tot

sexe	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
homme	3,549	,255	3,043	4,055
femme	3,955	,434	3,093	4,816

2. fum

Dependent Variable: prob_tot

fum	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
non-fumeur	3,090	,468	2,161	4,020
fumeur	4,414	,185	4,046	4,782

3. sexe * fum

Dependent Variable: prob_tot

sexe	fum	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
homme	non-fumeur	2,631	,454	1,729	3,532
	fumeur	4,468	,232	4,008	4,928
femme	non-fumeur	3,550	,819	1,925	5,175
	fumeur	4,359	,289	3,785	4,934

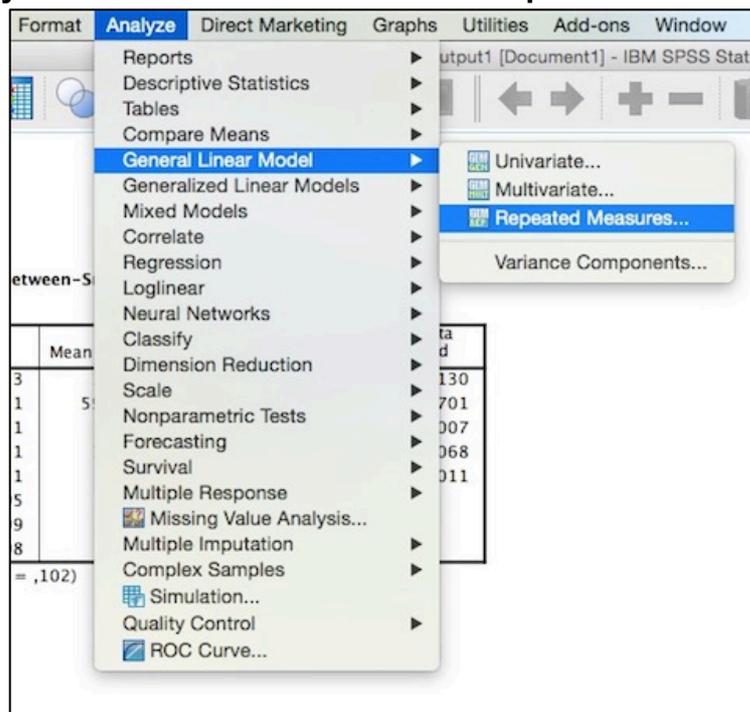
Notation : Il n'y a pas de différence entre les femmes ($M = 3.96$, $SE = .43$) et les hommes ($M = 3.55$, $SE = .26$) quant au nombre total de problèmes rencontrés, $F(1, 95) = .65$, $p > .05$, $\eta^2 = .01$, mais les fumeurs ($M = 4.41$, $SE = .19$) disent rencontrer plus de problèmes que les non fumeurs ($M = 3.09$, $SE = .47$), $F(1, 95) = 6.91$, $p < .05$, $\eta^2 = .07$. L'effet d'interaction non significatif indique que l'effet principal de la variable *fum* est valable tant pour les hommes que pour les femmes, $F(1, 95) = 1.04$, $p > .05$, $\eta^2 = .01$.

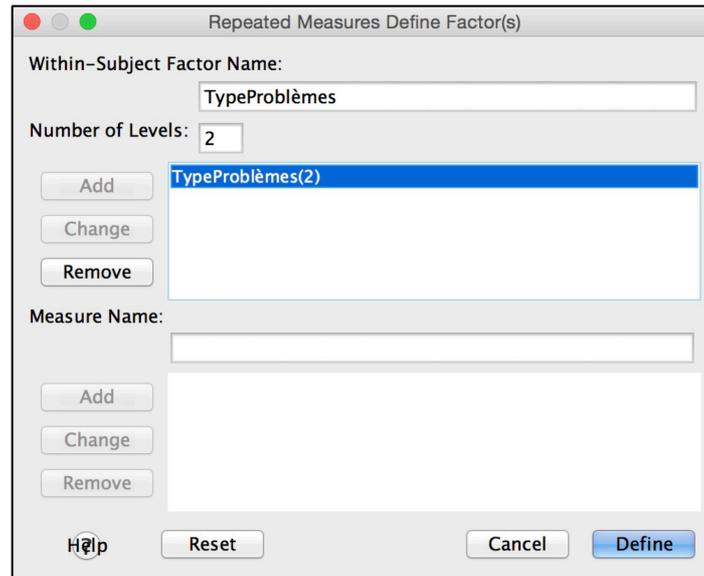
6.4.3. ANOVA à mesures répétées (+ ANOVA mixte)

Lorsque vous avez plusieurs VD, donc des mesures répétées, ou plus exactement une ou plusieurs VI nominale(s) à mesure répétées (= un ou plusieurs facteur(s) intrasujet(s)), et une ou plusieurs VI nominale(s) à groupes indépendants (= un ou plusieurs facteur(s) intersujet(s)), vous pouvez faire une ANOVA mixte (à groupes indépendants ET mesures répétées), qui dans SPSS se trouve dans le menu **Repeated Measures...** (Si vous n'avez que des facteurs intrasujets, faites également l'ANOVA à mesures répétées en laissant la case **Between-Subjects Factor(s)**: vide, cela fera une ANOVA à mesures répétées).

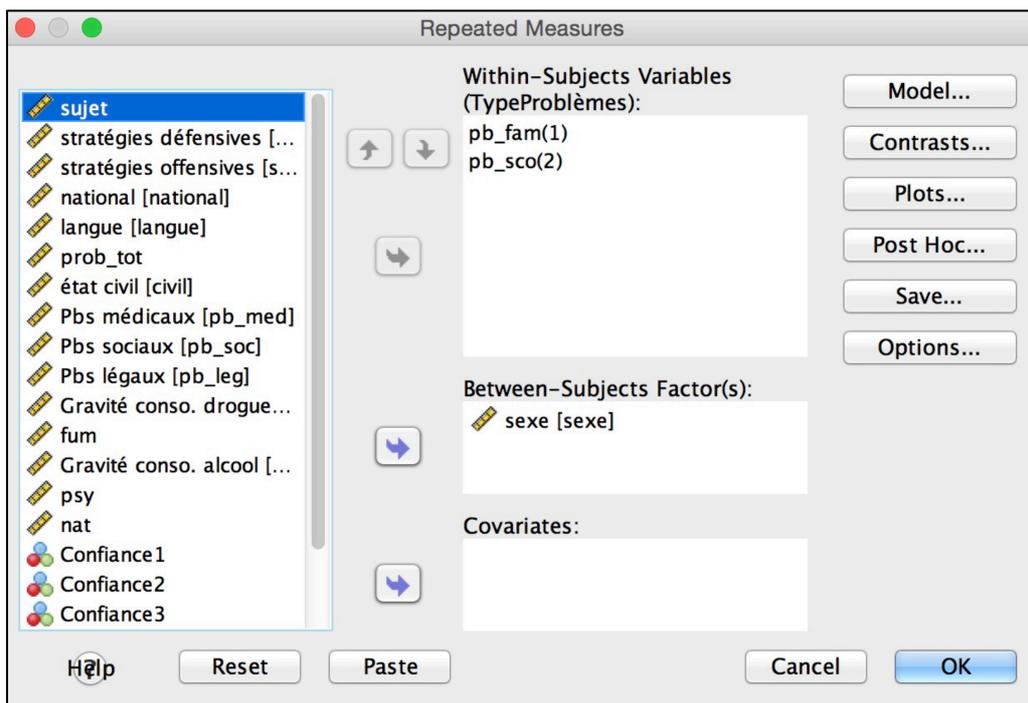
Exemple : Nous allons tester les différences entre filles et garçons adolescent-e-s (= facteur intersujets) sur l'évaluation de l'importance de deux types de problèmes, les problèmes scolaires et les problèmes familiaux (codés par un-e psychologue sur une échelle de 10 niveaux allant de 0 = absence de problèmes, à 9 = problèmes très graves). On peut considérer qu'il s'agit de deux VDs (importance des problèmes familiaux, importance des problèmes scolaires), ou alors une VD (importance des problèmes) avec un facteur intrasujets (type de problème à deux modalités : problèmes familiaux vs. problèmes scolaires = mesures répétées).

Chemin : **Analyze** → **General Linear Model** → **Repeated Measures...**



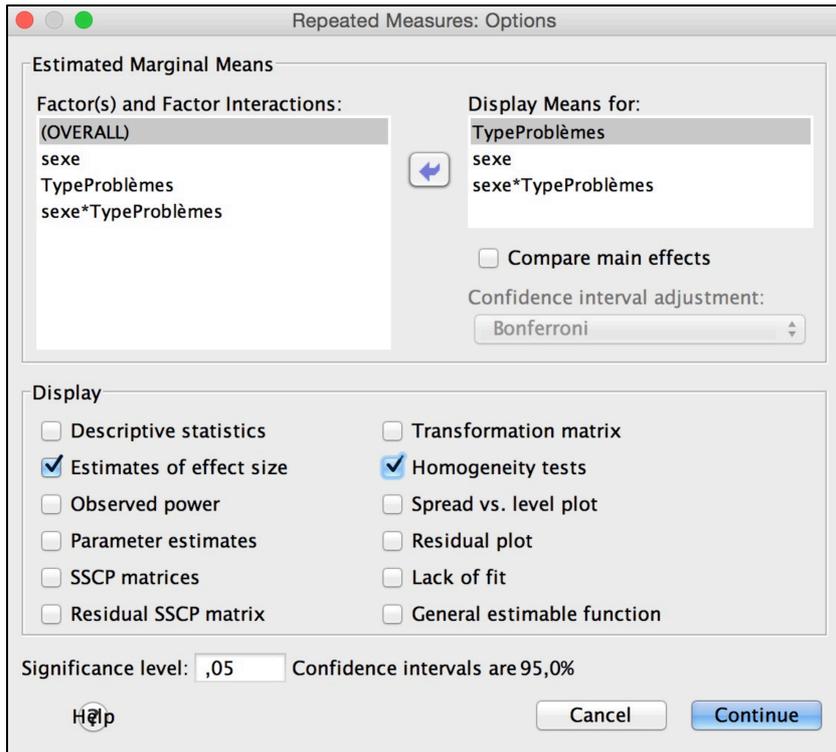


D'abord, définissez le facteur intrasujets (la mesure répétée) en entrant un nom dans **Within-Subject Factor Name**: (ici « TypeProblèmes ») et le nombre de modalités dans **Number of Levels**: (ici 2), puis cliquez **Add** pour que cela apparaisse dans la case au-dessous, puis **Define**.



Entrez les deux VDs (ou les deux niveaux de la VI intrasujets) dans **Within-Subjects Variables** (ici « pb_fam » et « pb_sco »), et la VI intersujets (ici « sexe ») dans **Between-Subjects Factor(s)**:

Dans les **Options**, demandez comme précédemment l'ANOVA (6.4.2.) les moyennes, le test d'homogénéité des variances et les tailles d'effet, **Continue** puis **OK**.



SPSS OUTPUT :

General Linear Model

Within-Subjects Factors

Measure: MEASURE_1

factor1	Dependent Variable
1	pb_fam
2	pb_sco

Between-Subjects Factors

	Value Label	N
sexe 1,00	homme	63
sexe 2,00	femme	36

Box's Test of Equality of Covariance Matrices^a

Box's M	2,067
F	,672
df1	3
df2	167425,867
Sig.	,569

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept +
sexe
Within Subjects
Design: factor1

Les premiers tableaux résument les facteurs (VIs) de l'analyse. Le test de Box doit être non significatif (égalité des covariances des VDs dans les groupes) pour pouvoir faire l'ANOVA, ce qui est le cas ici.

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
TypeProblèmes	Pillai's Trace	,117	12,817 ^b	1,000	97,000	,001	,117
	Wilks' Lambda	,883	12,817 ^b	1,000	97,000	,001	,117
	Hotelling's Trace	,132	12,817 ^b	1,000	97,000	,001	,117
	Roy's Largest Root	,132	12,817 ^b	1,000	97,000	,001	,117
TypeProblèmes * sexe	Pillai's Trace	,058	6,023 ^b	1,000	97,000	,016	,058
	Wilks' Lambda	,942	6,023 ^b	1,000	97,000	,016	,058
	Hotelling's Trace	,062	6,023 ^b	1,000	97,000	,016	,058
	Roy's Largest Root	,062	6,023 ^b	1,000	97,000	,016	,058

a. Design: Intercept + sexe
Within Subjects Design: TypeProblèmes

b. Exact statistic

Le tableau **Multivariate Tests** n'est à considérer que si la condition d'application de sphéricité (**Test de Mauchly** décrit ci-dessous) n'est pas remplie. Quand elle est remplie, ou que le facteur intrasujets n'a que deux modalités comme dans cet exemple, ce tableau donne les mêmes résultats que le tableau à reporter (**Tests of Within-Subjects Effects**).

Mauchly's Test of Sphericity^a

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TypeProblèmes	1,000	,000	0	.	1,000	1,000	1,000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + sexe
Within Subjects Design: TypeProblèmes

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Ici le **Test de Mauchly** n'est pas effectué puisqu'il n'y a que deux modalités du facteur intrasujets, la sphéricité n'est donc pas un problème.

Autre exemple: Lorsque vous avez plus de deux modalités répétées, le même tableau peut devenir :

Mauchly's Test of Sphericity^a

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TypeProblèmes	,985	1,456	2	,483	,985	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: TypeProblèmes

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Ici on garderait H_0 (= sphéricité, égalité des variances du facteur intrasujets) puisque p (**Sig.**) est $> .05$.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
TypeProblèmes	Sphericity Assumed	29,879	1	29,879	12,817	,001	,117
	Greenhouse-Geisser	29,879	1,000	29,879	12,817	,001	,117
	Huynh-Feldt	29,879	1,000	29,879	12,817	,001	,117
	Lower-bound	29,879	1,000	29,879	12,817	,001	,117
TypeProblèmes * sexe	Sphericity Assumed	14,041	1	14,041	6,023	,016	,058
	Greenhouse-Geisser	14,041	1,000	14,041	6,023	,016	,058
	Huynh-Feldt	14,041	1,000	14,041	6,023	,016	,058
	Lower-bound	14,041	1,000	14,041	6,023	,016	,058
Error (TypeProblèmes)	Sphericity Assumed	226,121	97	2,331			
	Greenhouse-Geisser	226,121	97,000	2,331			
	Huynh-Feldt	226,121	97,000	2,331			
	Lower-bound	226,121	97,000	2,331			

Le tableau **Tests of Within-Subjects Effects** vous donne l'effet principal du facteur intrasujets (ici « TypeProblèmes ») ainsi que ses effets d'interaction avec les autres facteurs (inter ou intra s'il y en a d'autres). Ici il y a l'effet d'interaction « TypeProblème*sexe ».

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
Pbs familiaux	,703	1	97	,404
Pbs scolaires	,024	1	97	,877

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + sexe
Within Subjects Design: TypeProblèmes

On peut ignorer le tableau **Test of Within-Subjects Contrasts** pour s'intéresser au **Test de Levene** qui contrôle la condition d'application d'égalité des variances entre les groupes du facteur intersujets (ici « sexe ») pour les deux VD (problèmes familiaux et problèmes scolaires). Le test doit être non significatif pour garder H_0 (= égalité des variances), ici c'est le cas puisque les deux p (**Sig.**) sont $> .05$.

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	4890,485	1	4890,485	638,644	,000	,868
sexe	48,586	1	48,586	6,345	,013	,061
Error	742,788	97	7,658			

Enfin nous considérons dans le tableau **Tests of Between-Subjects Effects** les effets (principaux et d'interaction) des facteurs intersujets, ici l'effet principal du facteur sexe.

Comme dans l'ANOVA précédente (6.4.2.), les effets principaux et d'interaction des facteurs inter ou intra doivent être interprétés avec l'aide des tableaux de moyennes (**Estimated Marginal Means**) :

Estimated Marginal Means

1. sexe

Measure: MEASURE_1

sexe	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
homme	4,651	,247	4,162	5,140
femme	5,681	,326	5,033	6,328

2. Typeproblèmes

Measure: MEASURE_1

Typeproblèmes	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	5,569	,224	5,125	6,014
2	4,762	,243	4,280	5,244

3. sexe * Typeproblèmes

Measure: MEASURE_1

sexe	Typeproblèmes	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
homme	1	4,778	,270	4,242	5,313
	2	4,524	,293	3,943	5,105
femme	1	6,361	,357	5,652	7,070
	2	5,000	,387	4,231	5,769

Notation : Une ANOVA mixte 2x2 à un facteur intersujets (Sexe : hommes vs. femmes) et un facteur intrasujets Type de problèmes : familiaux vs. scolaires) a été calculée. Elle a révélé un effet significatif du Type de problèmes, $F(1, 97) = 12.82$, $p < .05$, $\eta^2 = .12$, selon lequel les jeunes gens interrogé-e-s rapportent avoir davantage de problèmes familiaux ($M = 5.57$, $SE = .22$) que scolaires ($M = 4.76$, $SE = .24$). Un effet significatif du sexe, $F(1, 97) = 6.35$, $p < .05$, $\eta^2 = .06$, montre que les femmes ($M = 5.68$, $SE = .33$) disent globalement rencontrer plus de problèmes (familiaux et scolaires) en moyenne que les hommes ($M = 4.65$, $SE = .25$). L'effet d'interaction significatif entre Type de Problèmes et Sexe indique que la différence entre femmes ($M = 6.36$, $SE = .36$) et hommes ($M = 4.78$, $SE = .27$) est plus forte concernant les problèmes familiaux qu'en ce qui concerne les problèmes scolaires (femmes : $M = 5.00$, $SE = .39$; hommes : $M = 4.52$, $SE = .29$), $F(1, 97) = 6.02$, $p < .05$, $\eta^2 = .06$.

6.4.4. Comparaisons *a priori* (contrastes) et *a posteriori* (*post hoc*)

Dans l'ANOVA, rejeter H_0 signifie qu'il y a au moins une différence entre les moyennes. Quand le facteur (intra ou inter) n'a que deux moyennes, il suffit de les calculer avec SPSS pour savoir de quelle différence il s'agit. Néanmoins, quand vous avez plus de deux modalités, rejeter H_0 ne nous dit pas où se situent exactement les différences entre les moyennes. Pour ce faire, vous pouvez définir des contrastes (*a*

priori) si vous avez des hypothèses préalables sur le sens des différences de moyennes que vous devriez observer. Si vous n'avez pas d'hypothèses particulières, vous pouvez faire des comparaisons *a posteriori* (ou tests *post hoc*).

Exemple Contrastes : Les joueurs, arbitres et fans de football ont un score moyen de racisme symbolique différent ; on s'attend à ce que les fans (plus jeunes) aient un score inférieur aux deux autres groupes, qui ne seront pas différents entre eux. Le résultat de l'ANOVA est le suivant :

Univariate Analysis of Variance

Between-Subjects Factors		
groupe	Value Label	N
1,00	footballeurs	43
2,00	arbitres	15
3,00	fans	22

Levene's Test of Equality of Error Variances^a

Dependent Variable: racismesymb			
F	df1	df2	Sig.
,715	2	77	,493

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + groupe

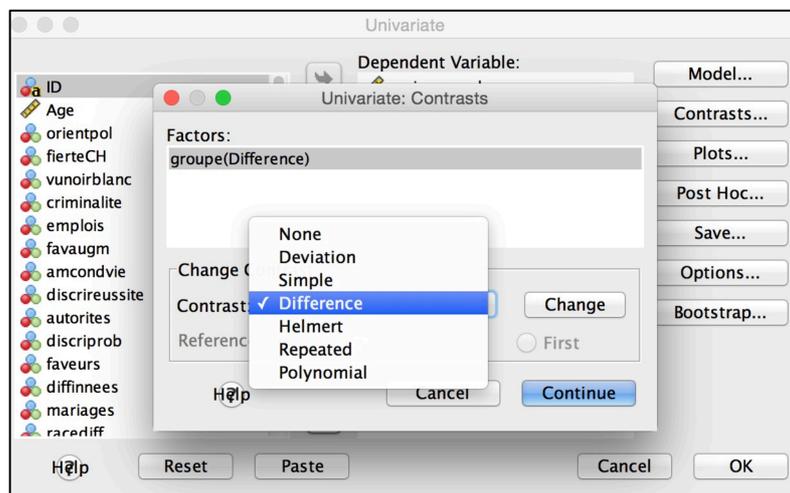
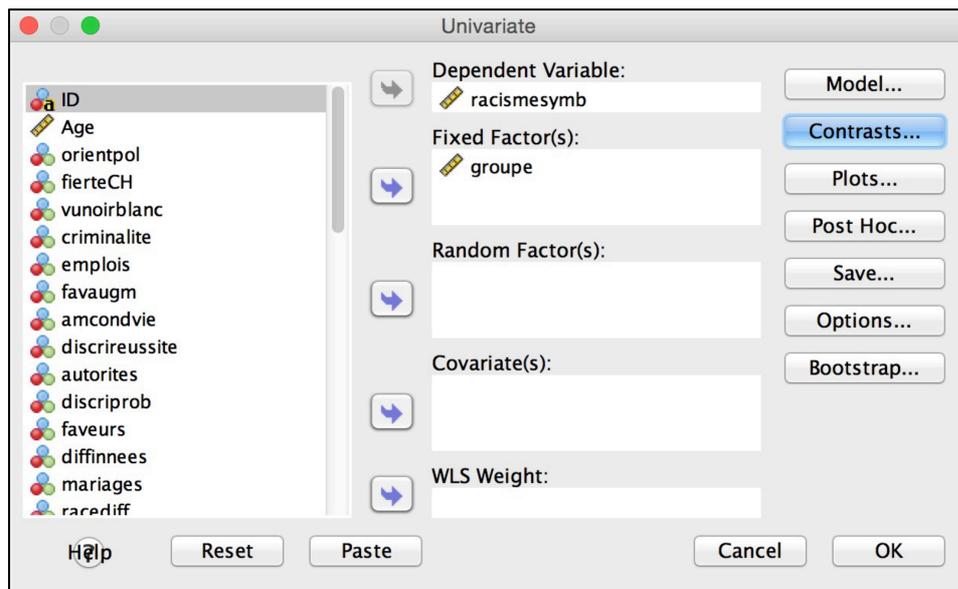
Tests of Between-Subjects Effects

Dependent Variable: racismesymb

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7,600 ^a	2	3,800	7,478	,001	,163
Intercept	492,579	1	492,579	969,287	,000	,926
groupe	7,600	2	3,800	7,478	,001	,163
Error	39,130	77	,508			
Total	629,931	80				
Corrected Total	46,731	79				

a. R Squared = ,163 (Adjusted R Squared = ,141)

Nous rejetons donc H_0 de l'égalité entre les trois moyennes, mais nous ne savons pas entre quelles moyennes se situe(nt) la ou les différence(s).



Pour tester des contrastes, cliquez sur l'onglet **Contrasts...** dans le menu de l'ANOVA (*idem* pour les ANOVA à un ou plusieurs facteur(s), mixte, MANOVA) et choisissez en cliquant dessus pour chaque facteur l'un des contrastes proposés par SPSS (cf. cours pour une description des différents contrastes). Ici nous choisissons par exemple le contraste **Difference**, ensuite cliquez sur **Change** et **Continue**, ce qui produit l'Output suivant :

Custom Hypothesis Tests

groupe Difference Contrast		Dependent Variable
		racismesymb
Level 2 vs. Level 1	Contrast Estimate	,362
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	,362
	Std. Error	,214
	Sig.	,095
	95% Confidence Interval for Difference	Lower Bound Upper Bound
Level 3 vs. Previous	Contrast Estimate	-,708
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	-,708
	Std. Error	,186
	Sig.	,000
	95% Confidence Interval for Difference	Lower Bound Upper Bound

Estimated Marginal Means

groupe

groupe	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
footballeurs	2,777	,109	2,561	2,994
arbitres	3,139	,184	2,772	3,505
fans	2,250	,152	1,947	2,553

Notation : Les arbitres, les footballeurs et les fans varient quant à leur score moyen de racisme symbolique, $F(2, 77) = 5.48$, $p < .05$, $r^2 = .16$. Les contrastes (méthode *Différence*) ont montré que comme attendu le groupe 2 (arbitres) ($M = 3.14$, $SE = .18$) n'est pas différent du groupe 1 (footballeurs) ($M = 2.78$, $SE = .11$), $p > .05$. Par contre, le groupe 3 (fans) ($M = 2.25$, $SD = .59$) a comme prévu une moyenne inférieure aux deux autres groupes, $p < .05$.

Pour les facteurs intrasujets (ici Type de Problèmes à 4 modalités, familiaux, scolaires, médicaux et sociaux), la procédure est exactement la même (les facteurs intrasujets apparaissent dans la même fenêtre que les facteurs intersujets), seul le tableau d'Output diffère un peu :

Source	TypeProblèmes	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
TypeProblèmes	Level 1 vs. Level 2	42,677	1	42,677	8,707	,004	,082
	Level 2 vs. Level 3	230,313	1	230,313	26,163	,000	,211
	Level 3 vs. Level 4	302,313	1	302,313	37,565	,000	,277
Error (TypeProblèmes)	Level 1 vs. Level 2	480,323	98	4,901			
	Level 2 vs. Level 3	862,687	98	8,803			
	Level 3 vs. Level 4	788,687	98	8,048			

Estimated Marginal Means

TypeProblèmes	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	5,370	,240	4,894	5,846
2	4,652	,249	4,158	5,147
3	3,149	,221	2,710	3,588
4	5,007	,217	4,577	5,437

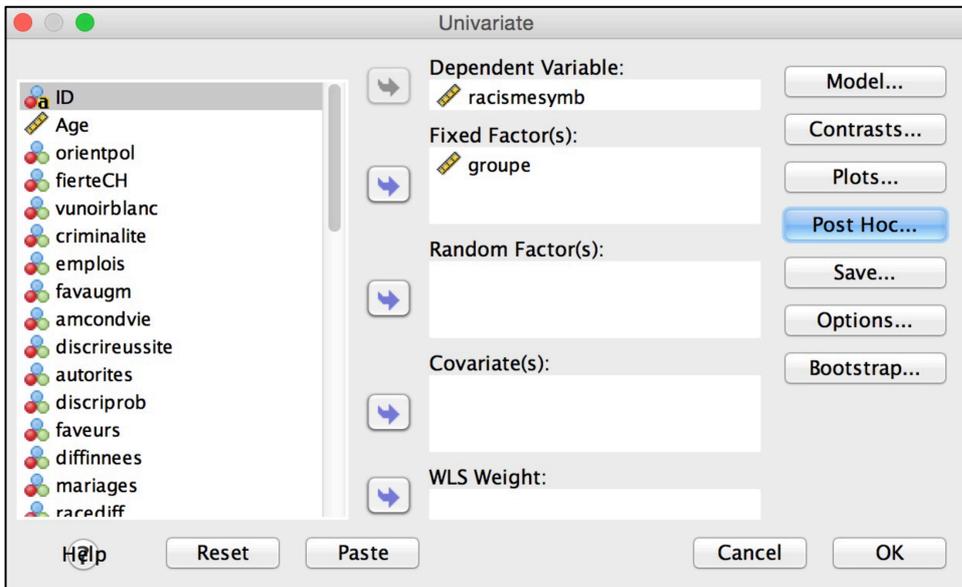
Tests of Within-Subjects Effects

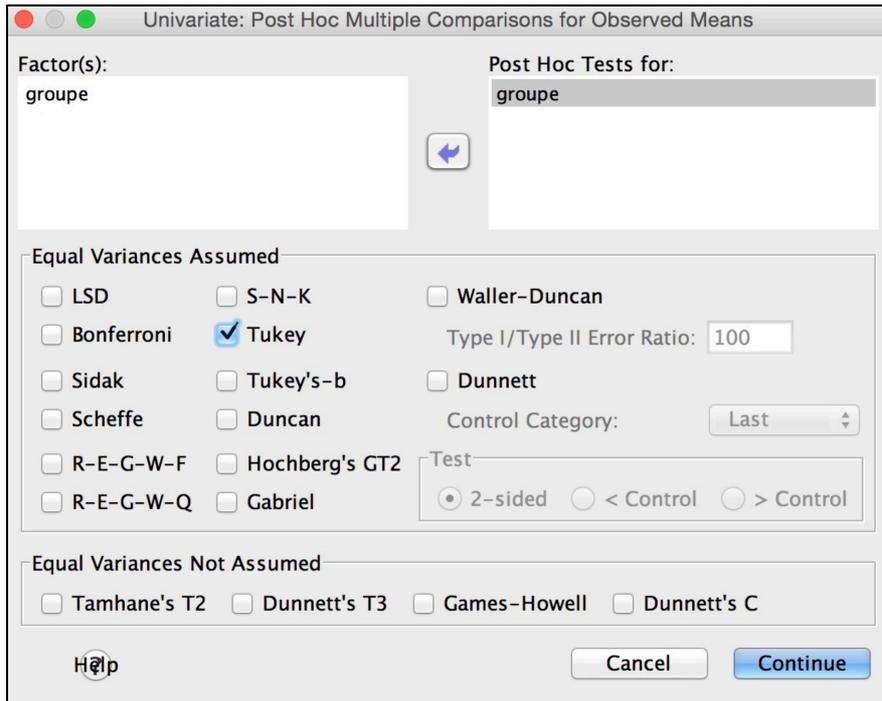
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
TypeProblèmes	Sphericity Assumed	267,535	3	89,178	27,716	,000	,220
	Greenhouse-Geisser	267,535	2,507	106,711	27,716	,000	,220
	Huynh-Feldt	267,535	2,578	103,767	27,716	,000	,220
	Lower-bound	267,535	1,000	267,535	27,716	,000	,220
Error (TypeProblèmes)	Sphericity Assumed	945,965	294	3,218			
	Greenhouse-Geisser	945,965	245,695	3,850			
	Huynh-Feldt	945,965	252,666	3,744			
	Lower-bound	945,965	98,000	9,653			

Notation : Les moyennes des quatre types de problèmes mesurés ne sont pas identiques, $F(3, 294) = 27.72$, $p < .05$, $r^2 = .22$. Des contrastes (méthode Répétée) ont montré que le nombre moyen de problèmes familiaux ($M = 5.37$, $SE = .24$) est plus élevé que le nombre moyen de problèmes scolaires ($M = 4.65$, $SE = .25$), $p < .05$, qui est à son tour plus élevé que le nombre moyen de problèmes médicaux ($M = 3.15$, $SE = .22$), $p < .05$. Celui-ci est également significativement inférieur au nombre moyen de problèmes sociaux ($M = 5.01$, $SE = .22$), $p < .05$.

Exemple Post Hoc : Les joueurs, arbitres et fans de football ont un score moyen de racisme symbolique différent, mais nous n'avons pas d'hypothèse préalable, nous allons donc comparer tous les groupes entre eux.





Pour faire les tests *a posteriori* sur les facteurs intersujets, cliquez sur **Post Hoc...** dans le menu de l'ANOVA (idem pour les ANOVA à un ou plusieurs facteur-s, mixte, MANOVA) et mettez chaque facteur inter que vous voulez tester dans la case **Post Hoc Tests for:** Cochez la case **Tukey** au dessous (qui est le test post hoc le plus puissant), puis cliquez **Continue** et **OK**.

Post Hoc Tests

groupe

Multiple Comparisons

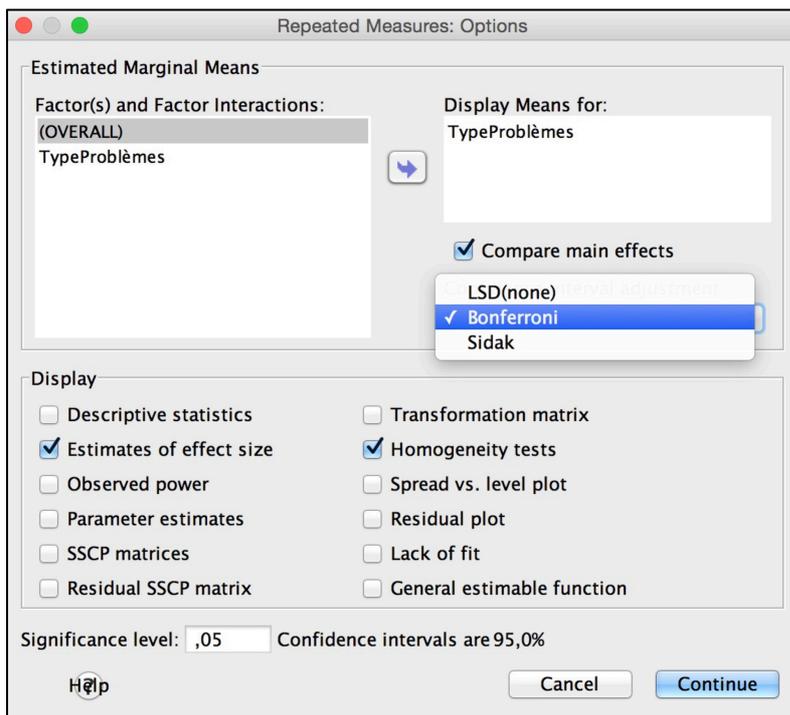
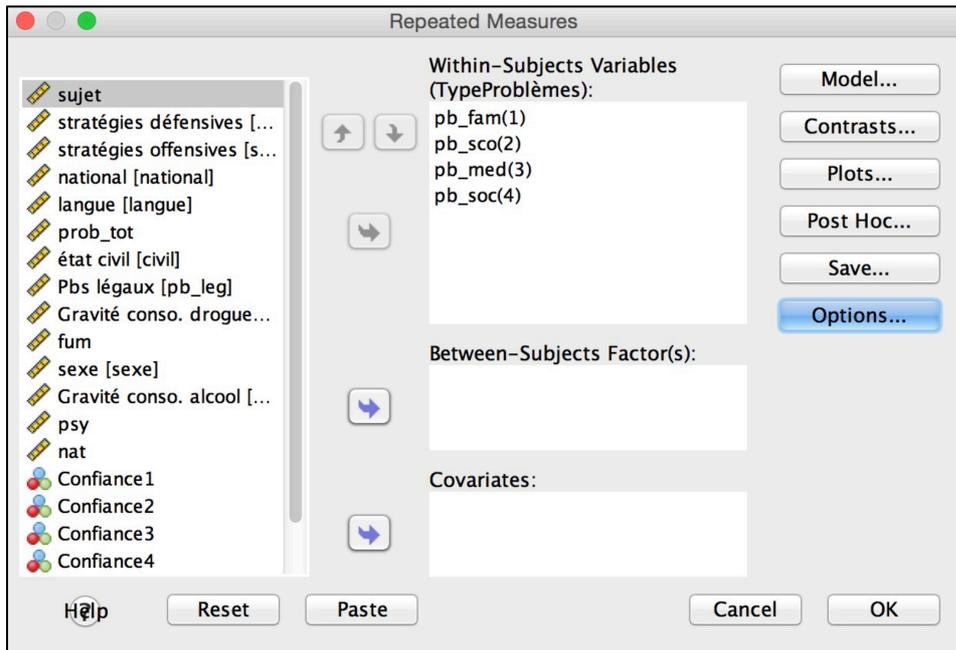
Dependent Variable: racismesymb
Tukey HSD

(I) groupe	(J) groupe	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
footballeurs	arbitres	-,3618	,21377	,215	-,8726	,1491
	fans	,5271*	,18686	,017	,0806	,9737
arbitres	footballeurs	,3618	,21377	,215	-,1491	,8726
	fans	,8889*	,23870	,001	,3184	1,4594
fans	footballeurs	-,5271*	,18686	,017	-,9737	-,0806
	arbitres	-,8889*	,23870	,001	-1,4594	-,3184

Based on observed means.
The error term is Mean Square(Error) = ,508.
*. The mean difference is significant at the

Notation : Les arbitres, les footballeurs et les fans varient quant à leur score moyen de racisme symbolique, $F(2, 77) = 5.48, p < .05, r^2 = .16$. Des tests *post hoc* (HSD de Tukey) ont montré que le groupe des fans ($M = 2.25, SD = .59$) a une moyenne inférieure au groupe des arbitres ($M = 3.14, SE = .18$), $p < .05$, et à celui des footballeurs ($M = 2.78, SE = .11$), $p < .05$. Par contre, ces deux derniers groupes ne diffèrent pas significativement entre eux, $p > .05$.

Pour les tests *a posteriori* sur les facteurs intrasujets, cliquez sur **Options...** dans le menu de l'ANOVA **Repeated Measures**.



Cochez la case **Compare main effects** et choisissez dans la case **Confidence interval adjustment** l'indice de **Bonferroni** de préférence, puis cliquez **Continue** et **OK**. Le tableau est semblable à celui des facteurs intersujets.

Pairwise Comparisons

Measure: MEASURE_1

(I) TypeProblèmes	(J) TypeProblèmes	Mean Difference (I- J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	,657*	,223	,024	,057	1,256
	3	2,182*	,292	,000	1,397	2,967
	4	,434	,200	,194	-,104	,973
2	1	-,657*	,223	,024	-1,256	-,057
	3	1,525*	,298	,000	,722	2,328
	4	-,222	,213	1,000	-,795	,351
3	1	-2,182*	,292	,000	-2,967	-1,397
	2	-1,525*	,298	,000	-2,328	-,722
	4	-1,747*	,285	,000	-2,515	-,980
4	1	-,434	,200	,194	-,973	,104
	2	,222	,213	1,000	-,351	,795
	3	1,747*	,285	,000	,980	2,515

Based on estimated marginal means

*. The mean difference is significant at the

b. Adjustment for multiple comparisons: Bonferroni.

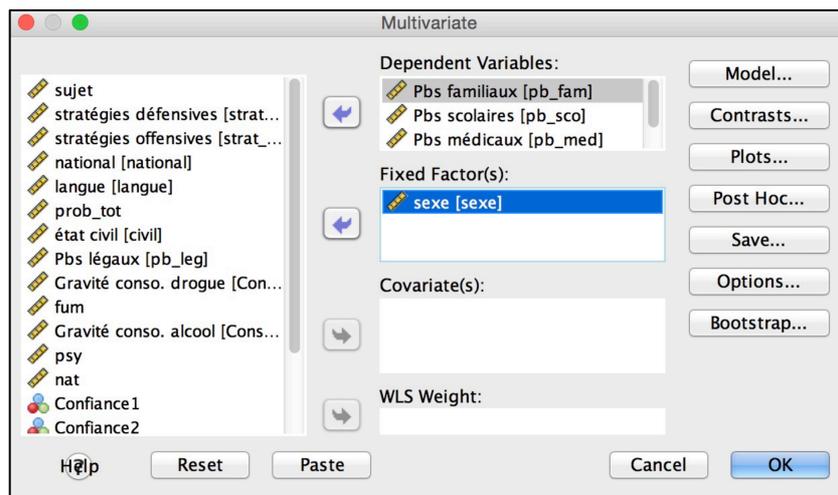
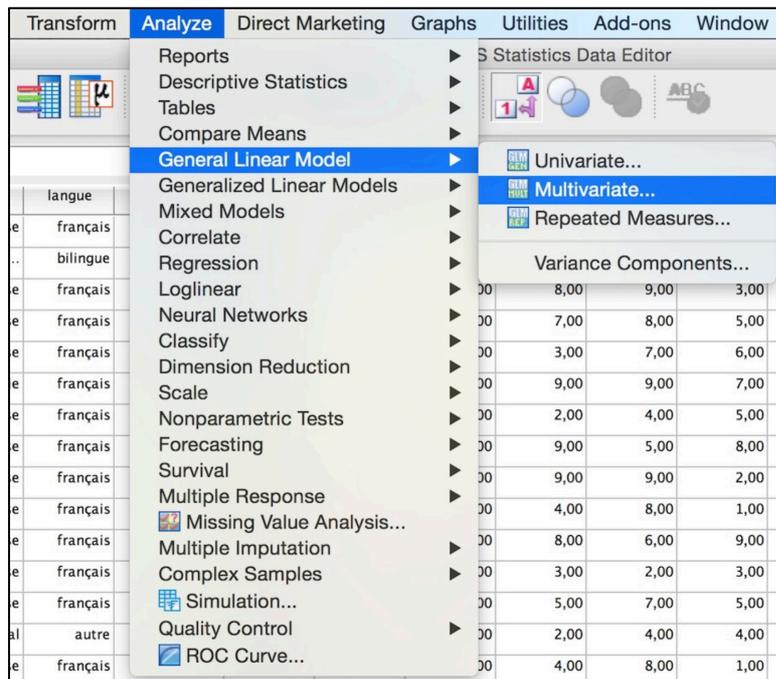
Notation : Les moyennes des quatre types de problèmes mesurés ne sont pas identiques, $F(3, 294) = 27.72$, $p < .05$, $r^2 = .22$. Des tests a posteriori (méthode Bonferroni) ont montré que chez les adolescent-e-s testé-e-s, le nombre moyen de problèmes familiaux ($M = 5.37$, $SE = .24$) est plus élevé que le nombre moyen de problèmes scolaires ($M = 4.65$, $SE = .25$) et médicaux ($M = 3.15$, $SE = .22$), mais pas différent du nombre moyen de problèmes sociaux ($M = 5.01$, $SE = .22$). Le nombre de problèmes scolaires est quant à lui plus élevé que le nombre de problèmes médicaux, mais pas différent du nombre de problèmes sociaux. Finalement, le nombre moyen de problèmes sociaux est plus élevé que celui des problèmes médicaux (toutes les différences reportées le sont à $p < .05$).

6.4.5. MANOVA

Lorsque vous avez une ou plusieurs VIs intersujets et plusieurs VDs, donc des mesures répétées (mais qui ne forment pas plusieurs facteurs intrasujets), vous pouvez si vous ne vous intéressez pas aux différences entre VDs (ou entre les modalités du facteur intrasujets) faire une MANOVA à la place de l'ANOVA à mesures répétées.

Exemple : Nous aimerions connaître les différences entre filles et garçons adolescent-e-s (= facteur intersujets) sur l'évaluation de l'importance de quatre types de problèmes, les problèmes familiaux, scolaires, médicaux et sociaux (codés par un-e psychologue sur une échelle de 10 niveaux allant de 0 = *absence de problèmes*, à 9 = *problèmes très graves*). On ne s'intéresse pas aux différences de moyennes entre problèmes, mais on cherche à savoir s'il y a une différence globale entre filles et garçons (donc sur l'ensemble des VDs), et ensuite où se situe exactement cette différence éventuelle (sur quelles VDs spécifiquement).

Chemin : **Analyze** → **General Linear Model** → **Multivariate...**



Entrez les VDs dans **Dependent Variables:** et les VIs dans **Fixed Factor(s)**.
 Comme pour les autres ANOVAs décrites précédemment, demandez les tableaux de moyennes dans **Options...** (**Display Means for:**), le(s) test(s) de Levene, les tailles d'effets (et contrastes ou tests *post hoc* si besoin), etc.

SPSS OUTPUT :

Les premiers tableaux donnent comme dans les ANOVAs décrites précédemment un résumé des facteurs de l'analyse, le test de Box pour l'égalité des covariances (qui doit être non significatif).

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	,900	211,102 ^b	4,000	94,000	,000	,900
	Wilks' Lambda	,100	211,102 ^b	4,000	94,000	,000	,900
	Hotelling's Trace	8,983	211,102 ^b	4,000	94,000	,000	,900
	Roy's Largest Root	8,983	211,102 ^b	4,000	94,000	,000	,900
sexe	Pillai's Trace	,135	3,678 ^b	4,000	94,000	,008	,135
	Wilks' Lambda	,865	3,678 ^b	4,000	94,000	,008	,135
	Hotelling's Trace	,157	3,678 ^b	4,000	94,000	,008	,135
	Roy's Largest Root	,157	3,678 ^b	4,000	94,000	,008	,135

a. Design: Intercept + sexe

b. Exact statistic

Dans le tableau ci-dessus, nous relevons l'effet multivarié du facteur intersujets sur l'ensemble des VDs.

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Pbs familiaux	57,432 ^a	1	57,432	12,513	,001	,114
	Pbs scolaires	5,195 ^b	1	5,195	,962	,329	,010
	Pbs médicaux	8,335 ^c	1	8,335	1,743	,190	,018
	Pbs sociaux	6,191 ^d	1	6,191	1,468	,229	,015
Intercept	Pbs familiaux	2842,442	1	2842,442	619,318	,000	,865
	Pbs scolaires	2077,922	1	2077,922	384,863	,000	,799
	Pbs médicaux	970,274	1	970,274	202,949	,000	,677
	Pbs sociaux	2281,827	1	2281,827	540,952	,000	,848
sexe	Pbs familiaux	57,432	1	57,432	12,513	,001	,114
	Pbs scolaires	5,195	1	5,195	,962	,329	,010
	Pbs médicaux	8,335	1	8,335	1,743	,190	,018
	Pbs sociaux	6,191	1	6,191	1,468	,229	,015
Error	Pbs familiaux	445,194	97	4,590			
	Pbs scolaires	523,714	97	5,399			
	Pbs médicaux	463,746	97	4,781			
	Pbs sociaux	409,163	97	4,218			
Total	Pbs familiaux	3340,000	99				
	Pbs scolaires	2713,000	99				
	Pbs médicaux	1468,000	99				
	Pbs sociaux	2811,000	99				
Corrected Total	Pbs familiaux	502,626	98				
	Pbs scolaires	528,909	98				
	Pbs médicaux	472,081	98				
	Pbs sociaux	415,354	98				

a. R Squared = ,114 (Adjusted R Squared = ,105)

Dans le tableau ci-dessous, nous relevons l'effet du facteur intersujets sur chaque VD prise séparément. Avec les moyennes (**Estimated Marginal Means**), nous pouvons écrire les résultats.

Estimated Marginal Means

sexe

Dependent Variable	sexe	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Pbs familiaux	homme	4,778	,270	4,242	5,313
	femme	6,361	,357	5,652	7,070
Pbs scolaires	homme	4,524	,293	3,943	5,105
	femme	5,000	,387	4,231	5,769
Pbs médicaux	homme	2,952	,275	2,406	3,499
	femme	3,556	,364	2,832	4,279
Pbs sociaux	homme	4,730	,259	4,217	5,244
	femme	5,250	,342	4,571	5,929

Notation : L'indice Pillai's Trace nous indique que le genre des adolescent-e-s interrogé-e-s a une influence globale sur les variables considérées dans leur ensemble, $V = .14$, $F(4, 94) = 3.68$, $p < .05$, $r^2 = .14$, les jeunes femmes ayant tendance à indiquer rencontrer davantage de problèmes que les jeunes hommes. La différence est notamment significative sur le nombre de problèmes familiaux qui sont en moyenne supérieurs chez les filles ($M = 6.36$, $SE = .36$) que chez les garçons ($M = 4.78$, $SE = .27$), $F(1, 97) = 12.51$, $p < .05$, $r^2 = .11$.

6.5. Tests non paramétriques

Si le t-test et l'ANOVA permettent de tester des hypothèses de différences de moyennes de groupes sur une VD quantitative, les tests non paramétriques permettent de le faire sur des variables ordinales. Quand les conditions d'application du t-test et de l'ANOVA ne sont pas remplies, il est possible de transformer la VD quantitative en VD ordinale (rangs). (Note : Pour des raisons pratiques, car il n'est pas aisé de créer de nouvelles données expressément pour ces tests non paramétriques, nous utiliserons les mêmes données que précédemment, même si elles suivent les conditions d'application du t-test et de l'ANOVA). Ainsi le but de ces tests est le même que le t-test et l'ANOVA appliqué à des variables ordinales.

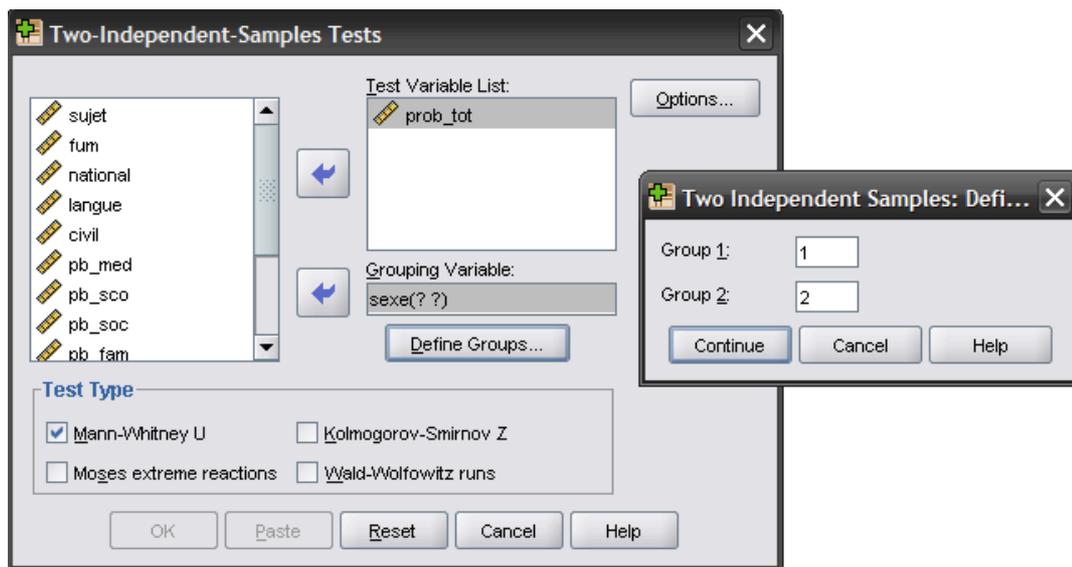
6.5.1. U de Mann-Whitney : 2 groupes indépendants

La statistique du test U réunit les deux échantillons et ordonne les observations par ordre croissant de taille. Le test calcule le nombre de fois où un résultat du groupe 1 précède un résultat du groupe 2, ainsi que le nombre de fois où un résultat du groupe 2 précède un résultat du groupe 1. Contrairement au t ou au F , l'indice U est d'autant plus petit que les groupes sont différents.

Chemin : **Analyze** → **Nonparametric Tests** → **Legacy dialogs** → **2 Independent Samples...**

The screenshot shows the SPSS 'Analyze' menu with 'Nonparametric Tests' selected. A sub-menu 'Legacy Dialogs' is open, and '2 Independent Samples...' is highlighted. In the background, the 'Statistics Data Editor' window is visible, showing a dataset with columns: pb_leg, Conso_Dro, Conso_OH, psy, prob_tot, and nationalite. The data rows show values for these variables across different categories of 'civil' status (célibataire).

pb_leg	Conso_Dro	Conso_OH	psy	prob_tot	nationalite
9,00	6,00	,00	G	6,80	suisse
1,00	2,00	1,00	G	1,20	suisse
3,00	4,00	,00	L	4,60	suisse
1,00	7,00	3,00	P	3,60	suisse
9,00	9,00	8,00	P	7,00	suisse
8,00	9,00	,00	G	6,80	suisse
			L	5,80	autre
			G	6,20	suisse
			P	6,20	suisse
7,00	7,00	2,00			
5,00	7,00	,00			
8,00	9,00	3,00			
2,00	7,00	6,00			
1,00	7,00	3,00			
9,00	5,00	3,00			



Comme pour le t-test, entrez vos variables (VD dans **Test Variable List:**, VI dans **Grouping Variable:**), puis indiquez les valeurs des deux groupes. Ici nous voulons savoir si les femmes et les hommes diffèrent quant au nombre total de problèmes rencontrés.

SPSS OUTPUT :

Ranks

		sexe	N	Mean Rank	Sum of Ranks
prob_tot	homme		63	46.45	2926.50
	femme		36	56.21	2023.50
Total			99		

Test Statistics^a

	prob_tot
Mann-Whitney U	910.500
Wilcoxon W	2926.500
Z	-1.628
Asymp. Sig. (2-tailed)	.104

a. Grouping Variable: sexe

Puisque les tests non paramétriques transforment les données en rangs, pour donner les résultats, nous utilisons plutôt que les moyennes et les écarts-types (ou erreurs standards), les **médianes** et les **intervalles interquartiles**. Ceux-ci peuvent être calculés dans le menu **Explore** en introduisant la VD dans **Dependent List:**

Descriptives

sexe			Statistic	Std. Error
prob_tot	homme	Mean	4,0889	,21657
		95% Confidence Interval for Mean	Lower Bound	3,6560
			Upper Bound	4,5218
		5% Trimmed Mean	4,0899	
		Median	4,2000	
		Variance	2,955	
		Std. Deviation	1,71897	
	Minimum	,80		
	Maximum	7,20		
	Range	6,40		
	Interquartile Range	2,60		
	Skewness	-,054	,302	
	Kurtosis	-,750	,595	
	femme	femme	Mean	4,2694
95% Confidence Interval for Mean			Lower Bound	3,6736
			Upper Bound	4,8653
5% Trimmed Mean			4,2790	
Median			4,4000	
Variance			3,102	
Std. Deviation			1,76114	
Minimum		1,10		
Maximum		7,40		
Range		6,30		
Interquartile Range		2,85		
Skewness		-,141	,393	
Kurtosis		-,771	,768	

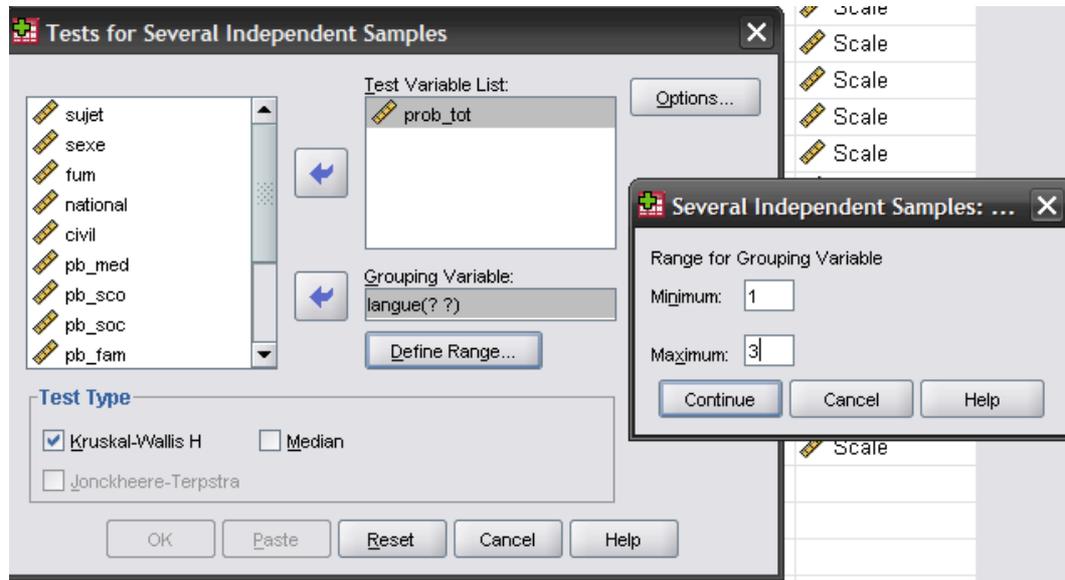
Notation : Les hommes ($Mdn = 4.20$, $IQR = 2.60$) ne diffèrent pas significativement des femmes ($Mdn = 4.40$, $IQR = 2.85$) quant au nombre total de problèmes, $U = 910.5$, $p > .05$.

6.5.2. Kruskal-Wallis : n groupes indépendants

Chemin : **Analyze** → **Nonparametric Tests** → **Legacy dialogs** → **K Independent Samples...**

The screenshot shows the SPSS 'Analyze' menu with 'Nonparametric Tests' selected. The 'Legacy Dialogs' sub-menu is open, and 'K Independent Samples...' is highlighted. The background shows a data editor window with a table of variables including 'pb_leg', 'Conso_Dro', 'Conso_OH', 'psy', 'prob_tot', and 'nationalite'.

Dans cet exemple, nous souhaitons voir si la langue parlée (français, bilingue ou autre) influence le degré de problèmes totaux.



Entrez vos variables, puis définissez les groupes pour la VI. Vous n'avez pas besoin de définir chaque groupe, mais juste le minimum et le maximum.

SPSS OUTPUT :

Ranks

	langue	N	Mean Rank
prob_tot	français	79	50.97
	bilingue	7	55.71
	autre	13	41.00
	Total	99	

Test Statistics^{a,b}

	prob_tot
Chi-Square	1.648
df	2
Asymp. Sig.	.439

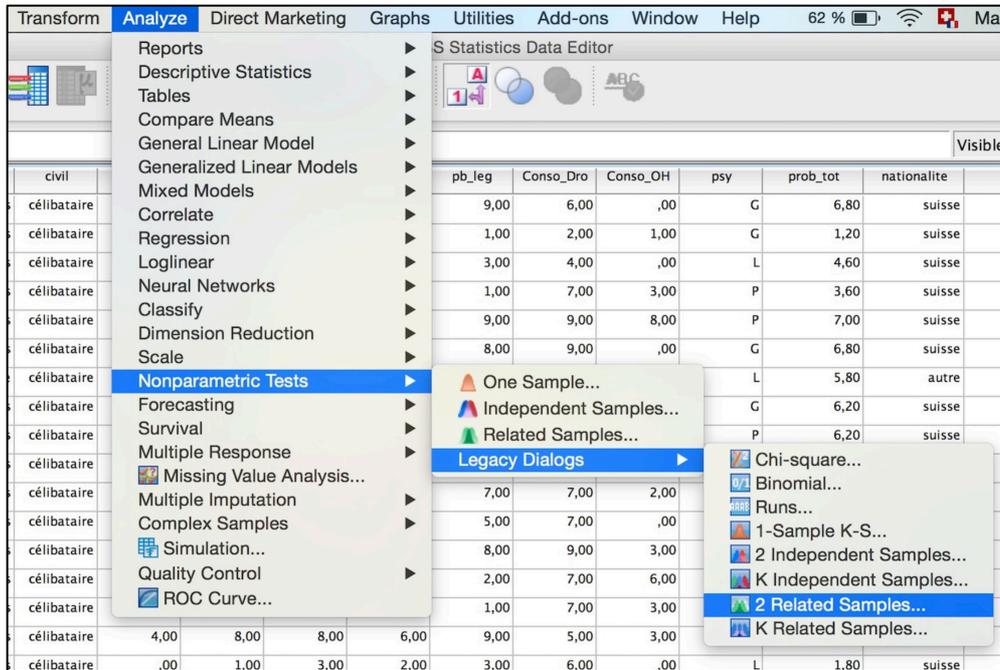
a. Kruskal Wallis Test

b. Grouping Variable: langue

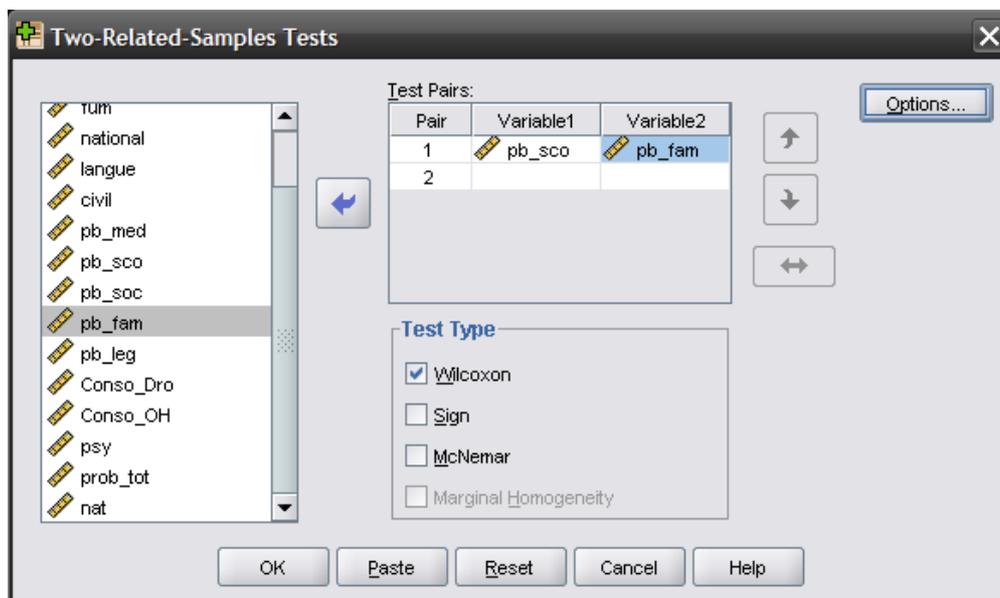
Notation : La langue n'influence pas le degré de problèmes totaux, $Chi2(2) = 1.65$, $p > .05$ (ou alors $H(2) = 1.65$, $p > .05$; vous pouvez bien sûr indiquer dans la phrase les médianes et écarts interquartiles comme dans l'exemple précédent).

6.5.3. Wilcoxon : 2 groupes appariés

Chemin : **Analyze** → **Nonparametric Tests** → **Legacy dialogs** → **2 Related Samples...**



Exemple : nous voulons savoir si les participant-e-s ont en moyenne plus de problèmes scolaires ou de problèmes familiaux.



Entrez les deux variables qui vous intéressent. Remarquez que vous pouvez effectuer plusieurs comparaisons à la fois.

SPSS OUTPUT :

Ranks

		N	Mean Rank	Sum of Ranks
Pbs familiaux - Pbs scolaires	Negative Ranks	26 ^a	33.02	858.50
	Positive Ranks	48 ^b	39.93	1916.50
	Ties	25 ^c		
	Total	99		

- a. Pbs familiaux < Pbs scolaires
- b. Pbs familiaux > Pbs scolaires
- c. Pbs familiaux = Pbs scolaires

Test Statistics^b

	Pbs familiaux - Pbs scolaires
Z	-2.882 ^a
Asymp. Sig. (2-tailed)	.004

- a. Based on negative ranks.
- b. Wilcoxon Signed Ranks Test

A l'aide de la commande **Explore**, vous pouvez demander les médianes et intervalles interquartiles afin d'interpréter le résultat du test :

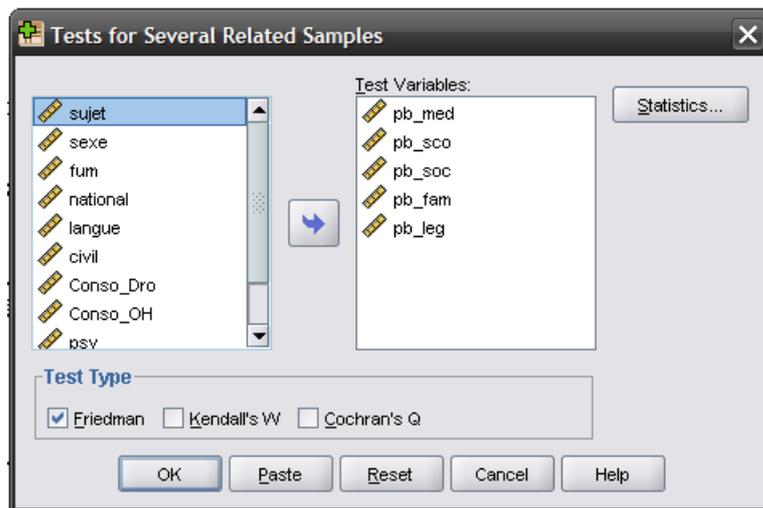
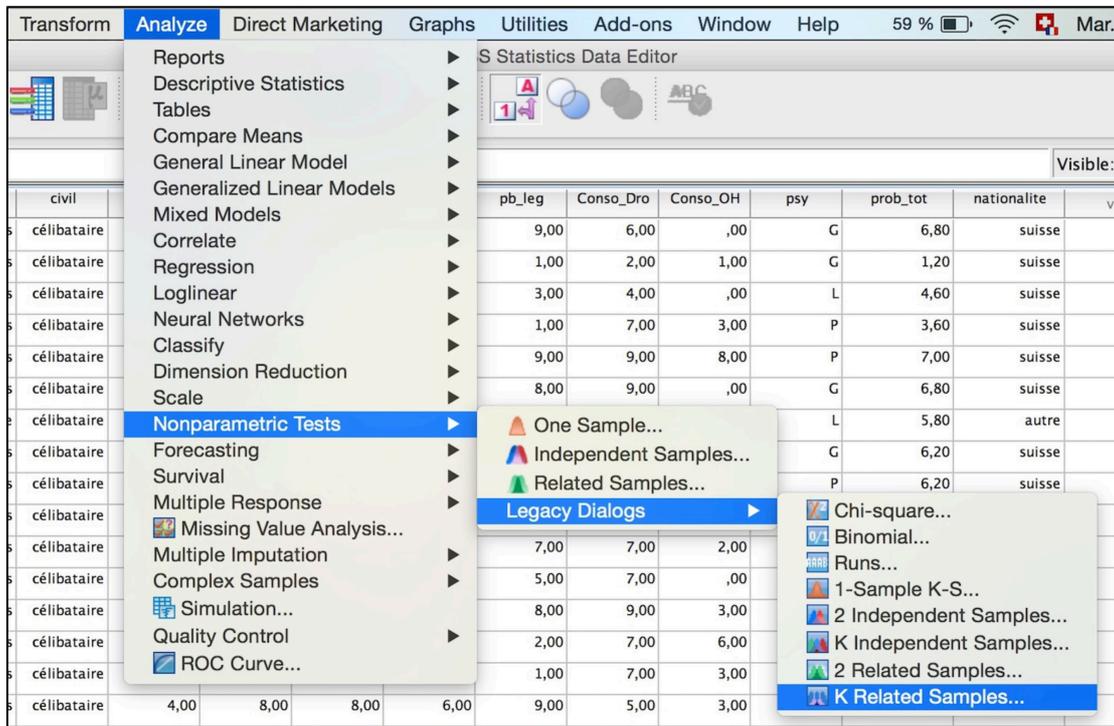
Descriptives

		Statistic	Std. Error	
Pbs scolaires	Mean	4,6970	,23349	
	95% Confidence Interval for Mean	Lower Bound	4,2336	
		Upper Bound	5,1603	
	5% Trimmed Mean	4,6970		
	Median	4,0000		
	Variance	5,397		
	Std. Deviation	2,32315		
	Minimum	,00		
	Maximum	9,00		
	Range	9,00		
	Interquartile Range	4,00		
	Skewness	-,037	,243	
	Kurtosis	-,739	,481	
	Pbs familiaux	Mean	5,3535	,22761
95% Confidence Interval for Mean		Lower Bound	4,9019	
		Upper Bound	5,8052	
5% Trimmed Mean		5,4046		
Median		5,0000		
Variance		5,129		
Std. Deviation		2,26469		
Minimum		,00		
Maximum		9,00		
Range		9,00		
Interquartile Range		3,00		
Skewness		-,144	,243	
Kurtosis		-,712	,481	

Notation : Les problèmes familiaux sont en général plus importants (*Mdn* = 5.00, *IQR* = 3.00) que les problèmes scolaires (*Mdn* = 4.00, *IQR* = 4.00), *T* = 858.50, *p* < .05 (ou *Z* = -2.88, *p* < .05).

6.5.4. ANOVA de Friedman : n groupes appariés

Chemin : **Analyze** → **Nonparametric Tests** → **Legacy dialogs** → **K Related Samples...**



Entrez toutes les variables que vous souhaitez comparer. Dans cet exemple nous voulons voir si le nombre de problèmes de différents types est différent ou semblable.

SPSS OUTPUT :

Ranks

	Mean Rank
Pbs médicaux	2.04
Pbs scolaires	3.25
Pbs sociaux	3.30
Pbs familiaux	3.72
Pbs légaux	2.70

Test Statistics^a

N	99
Chi-Square	73.954
df	4
Asymp. Sig.	.000

a. Friedman Test

Notation : Il y a une différence entre les 5 types de problèmes rencontrés par les participant-e-s, $Chi^2(4) = 73.95$, $p < .05$ (vous pouvez renvoyer à un tableau ou graphique avec les 5 médianes et intervalles interquartiles).

6.6. Corrélation

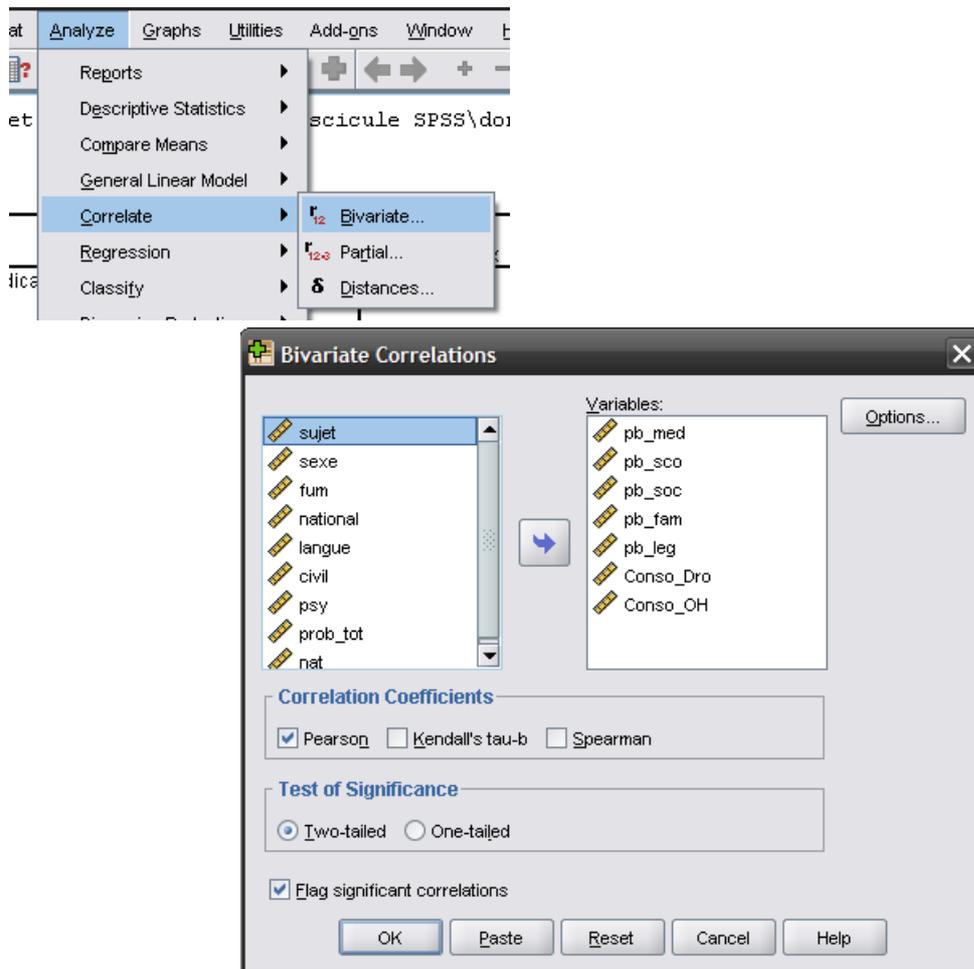
But : L'analyse de corrélation permet de tester s'il existe une relation positive ou négative (ou nulle) entre deux variables quantitatives (r de Bravais-Pearson) ou ordinales (rho de Spearman = ρ).

Exemple :

Nous souhaitons connaître le lien entre différents types de problèmes (médicaux, scolaires, sociaux, familiaux et légaux) et la gravité de consommation de drogue et d'alcool.

Attention : une corrélation ne teste le lien que de deux variables à la fois !

Chemin : **Analyze** → **Correlate** → **Bivariate...**



Comme toujours, vous trouvez la liste des variables dans la case de gauche. Sélectionnez les variables qui vous intéressent et passez-les dans la case **Variables:** à l'aide de la flèche qui se situe entre les deux cases. Attention, vos variables doivent être indiquées comme étant quantitatives par SPSS. Si ce n'est pas le cas, vous ne pourrez pas les transférer.

Par défaut, la case **Pearson** est cochée. Si la normalité de vos variables n'est pas bonne (ou si vous avez des variables ordinales), demandez plutôt le coefficient non paramétrique (calculé sur les rangs) Rho de **Spearman**.

SPSS OUTPUT :

Correlations

		Pbs médicaux	Pbs scolaires	Pbs sociaux	Pbs familiaux	Pbs légaux	Gravité conso. drogue	Gravité conso. alcool
Pbs médicaux	Pearson Correlation	1	.138	.169	.149	.670**	.492**	.084
	Sig. (2-tailed)		.172	.094	.141	.000	.000	.410
	N	99	99	99	99	99	99	99
Pbs scolaires	Pearson Correlation	.138	1	.569**	.536**	.349**	.161	.097
	Sig. (2-tailed)	.172		.000	.000	.000	.112	.340
	N	99	99	99	99	99	99	99
Pbs sociaux	Pearson Correlation	.169	.569**	1	.673**	.303**	.431**	.315**
	Sig. (2-tailed)	.094	.000		.000	.002	.000	.002
	N	99	99	99	99	99	99	99
Pbs familiaux	Pearson Correlation	.149	.536**	.673**	1	.184	.352**	.240*
	Sig. (2-tailed)	.141	.000	.000		.069	.000	.017
	N	99	99	99	99	99	99	99
Pbs légaux	Pearson Correlation	.670**	.349**	.303**	.184	1	.440**	.137
	Sig. (2-tailed)	.000	.000	.002	.069		.000	.178
	N	99	99	99	99	99	99	99
Gravité conso. drogue	Pearson Correlation	.492**	.161	.431**	.352**	.440**	1	.156
	Sig. (2-tailed)	.000	.112	.000	.000	.000		.122
	N	99	99	99	99	99	99	99
Gravité conso. alcool	Pearson Correlation	.084	.097	.315**	.240*	.137	.156	1
	Sig. (2-tailed)	.410	.340	.002	.017	.178	.122	
	N	99	99	99	99	99	99	99

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

SPSS calcule toutes les corrélations possibles deux à deux et les présente sous forme de tableau croisé. On ne s'intéresse donc qu'aux résultats en dessus (ou en dessous) de la diagonale, puisque le tableau est symétrique. Vous remarquez que la diagonale présente des corrélations de valeur $r = 1$: une variable est parfaitement corrélée avec elle-même.

Dans le tableau, les corrélations significatives sont accompagnées d'étoiles *. Plus il y a d'étoiles, plus le p est petit.

Les variables entourées en traitillés rouges sont des corrélations moyennes (de .30 à .50) et en traits pleins sont des corrélations fortes ($> .50$).

Notation : Il y a un lien positif fort entre les problèmes scolaires et les problèmes sociaux, $r = .57$, $p < .05$.

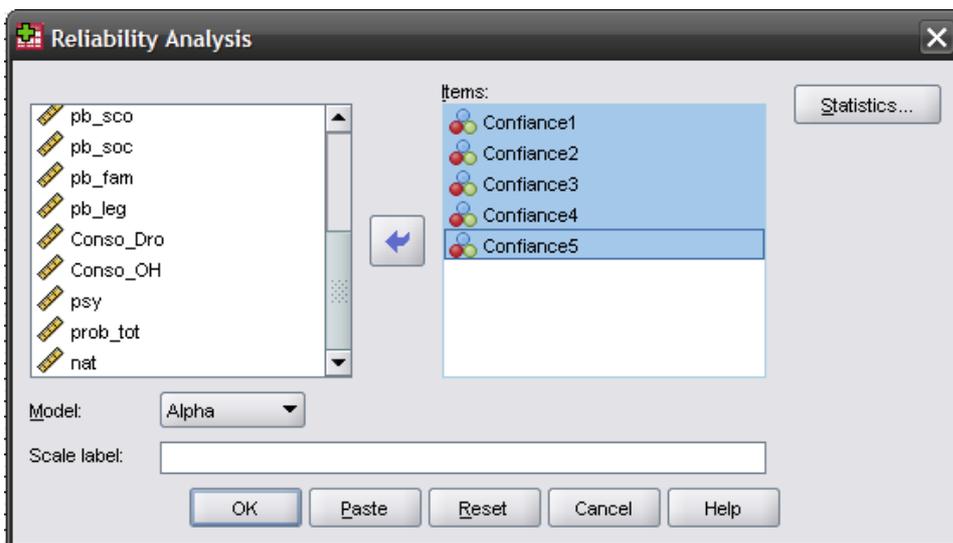
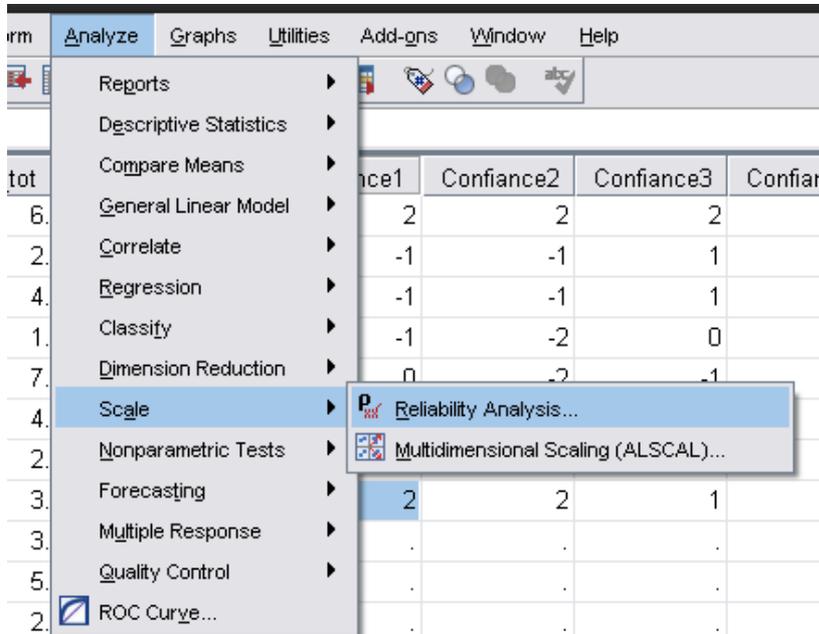
6.7. Alpha de Cronbach

But : L'alpha de Cronbach permet de dire si les items d'une échelle sont suffisamment corrélés positivement entre eux pour créer un seul score de moyenne de ces items (avec la commande **Compute**, cf. paragraphe 4.1).

Exemple : Dans un questionnaire, 5 variables sont censées mesurer le degré de confiance en soi. Pour les participant-e-s en question, ces 6 variables vont-elles bien ensemble ? Est-ce que les participant-e-s les perçoivent comme relevant de la même

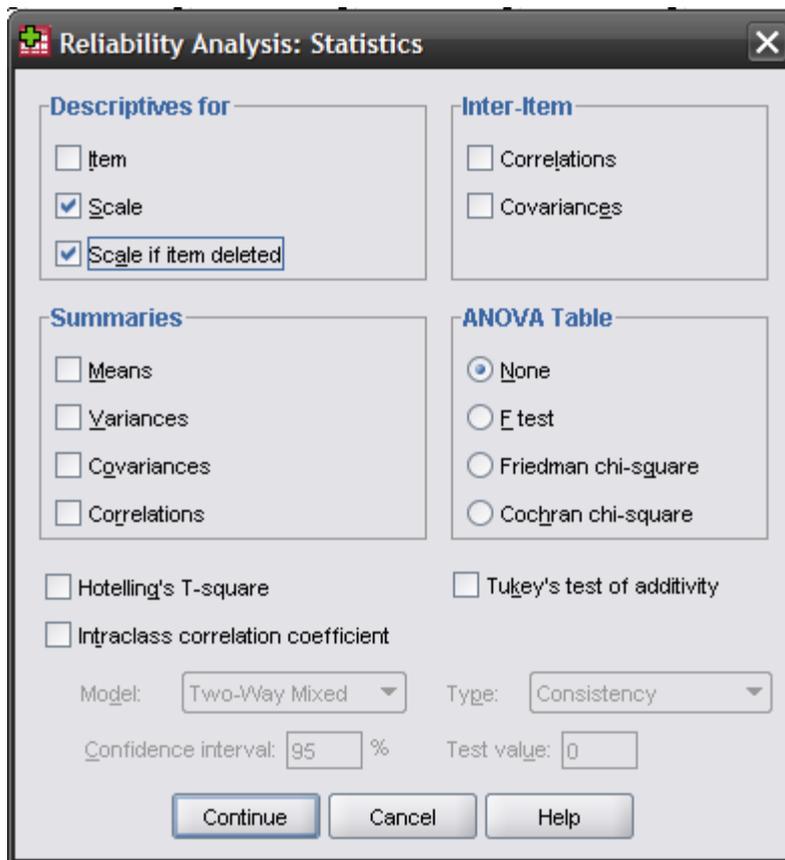
sous-dimension (sont-elles toutes plus ou moins positivement corrélées), ou les considèrent-ils comme indépendantes les unes des autres (corrélations nulles) ?

Chemin : **Analyze** → **Scale** → **Reliability Analysis...**



Sélectionnez dans la case de gauche les variables qui sont censées corrélérer (mesurer la même sous-dimension) et faites-les glisser dans la case **Items**: grâce à la flèche entre les deux cases. Dans **Model**, le test alpha est sélectionné par défaut. Ne changez rien.

Cliquez sur l'onglet **Statistics...** Une nouvelle fenêtre s'ouvre :



Dans la rubrique **Descriptives for**, cochez les cases **Scale** (celle-ci vous donne l'indication de la valeur de l'alpha pour l'ensemble des variables que vous avez sélectionné) et **Scale if item deleted** (celle-ci vous donne l'indication de la valeur de l'alpha si vous enlevez une variable particulière de sa composition). Le fait de cocher cette dernière case vous permet d'évaluer la contribution de chaque variable à la mesure globale et de pouvoir éliminer, le cas échéant, celle ou celles qui réduisent la valeur de l'alpha de la mesure globale.

Cliquez sur **Continue** puis **OK**.

SPSS OUTPUT :

Reliability Statistics

Cronbach's Alpha	N of Items
.677	5

Ce premier tableau vous indique que l'alpha global des 5 variables est de .68, ce qui représente une bonne valeur de l'alpha. On peut donc créer une variable unique en calculant la moyenne des 5 variables.

Si la valeur de l'alpha est très satisfaisante (> .70), vous pouvez vous arrêter là et créer votre indice à partir des 7 variables. Pour l'exercice, nous considérons

également si la valeur de l'alpha peut être augmentée (améliorée) en ne prenant pas en compte certaines variables.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Confiance1	2.91	13.514	.385	.651
Confiance2	3.34	12.795	.462	.613
Confiance3	2.30	13.605	.596	.565
Confiance4	2.66	13.751	.434	.625
Confiance5	2.91	15.775	.315	.672

Ce second tableau indique que la variable qui contribue le moins à la mesure globale est .315 (cf. colonne **Corrected Item-Total Correlation**).

Si elle n'avait pas été entrée dans le test, l'alpha aurait été de .672 (cf. la colonne **Cronbach's Alpha if Item Deleted**).

Ainsi si notre alpha n'avait pas été bon, nous aurions pu ne pas inclure la variable *Confiance5* pour faire remonter cet indice. Ce qui veut dire que lorsqu'on créera la mesure globale on effectuera la moyenne sans cette variable-là.

Important : lorsque vous travaillez sur des échelles qui comportent plusieurs variables et qui ont déjà été validées par d'autres études, vous n'avez pas intérêt à enlever des variables de la mesure globale juste pour « remonter » la valeur de l'alpha. Il se peut que le fait d'enlever une variable modifie la signification du tout ! Dans ces cas, une fois que la cohérence interne est satisfaisante (et si c'est une échelle bien validée, ce sera le cas), créez la mesure globale en comprenant toutes les variables. Autrement... posez-vous la question de la validité de l'échelle elle-même !

ATTENTION : lorsqu'on calcule un alpha, il faut que les échelles de mesure de toutes les variables aillent dans le même sens ! Si votre alpha comporte des valeurs négatives, cela signifie qu'il y a des variables qui s'opposent à d'autres dans votre analyse et il faudra probablement inverser certaines d'entre elles à travers un recodage (voir paragraphe 4.2).

<p>Notation : Les cinq items de l'échelle de confiance ont une bonne consistance interne, alpha de Cronbach = .68.</p>

6.8. Régression

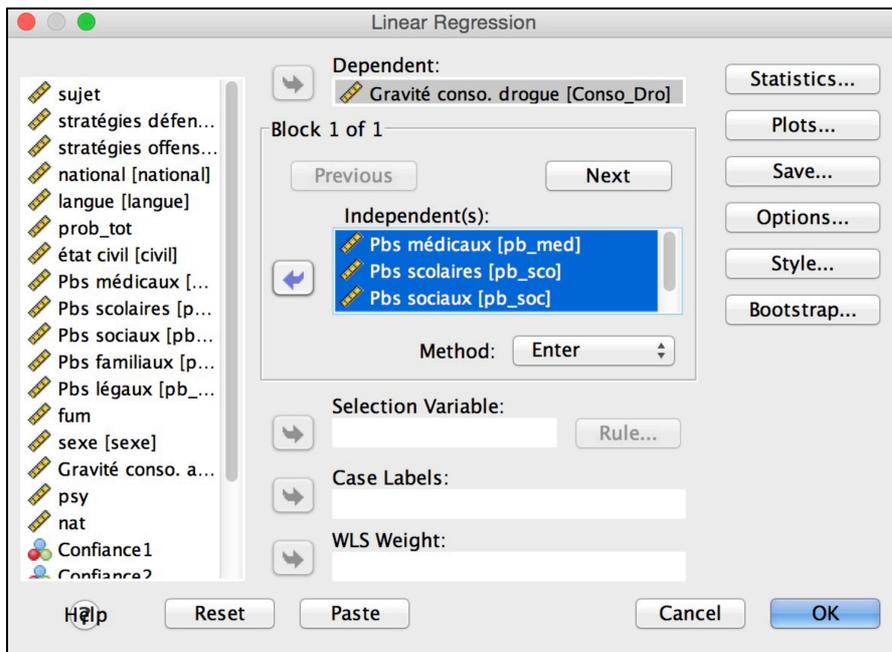
But : Savoir si une variable quantitative (régression simple) ou plusieurs variables quantitatives (régression multiple), appelées les prédicteurs (ou VI quantitatives), sont passible(s) de prédire les valeurs d'une variable quantitative appelée critère (ou VD).

Exemple : Nous aimerions savoir quels types de problèmes (familiaux, scolaires, médicaux ou sociaux) prédisent la consommation de drogue chez des adolescent-e-s. C'est une régression multiple avec les quatre types de problèmes comme prédicteurs et la consommation de drogue comme critère (= régression multiple).

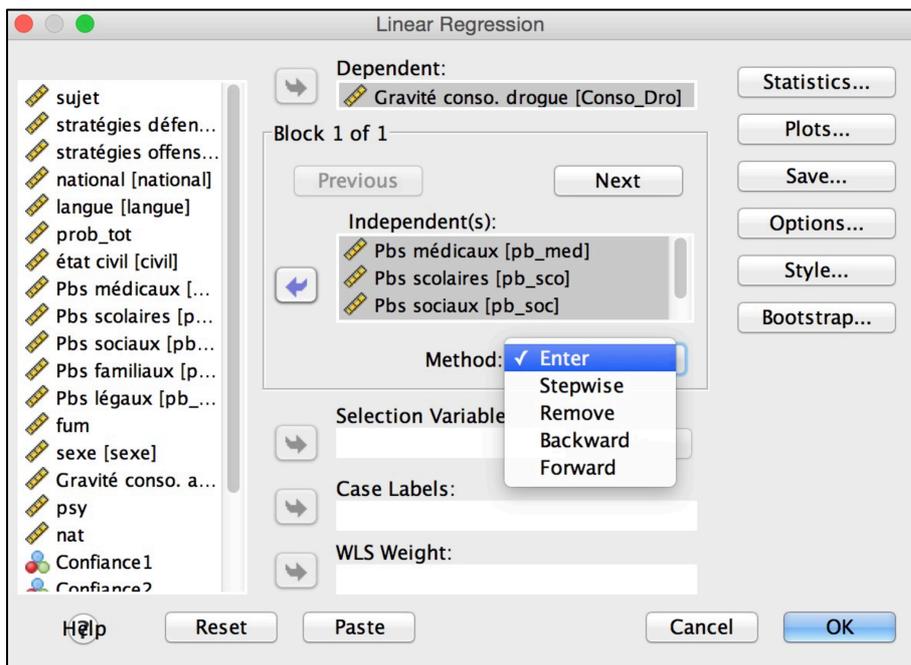
Chemin : **Analyze → Regression → Linear...**

The screenshot shows the SPSS 'Analyze' menu with the 'Regression' option selected. The 'Linear...' option is highlighted. The background shows a data editor window with columns for 'pb_soc', 'pb_fam', 'pb_leg', and 'Conso_Dro'.

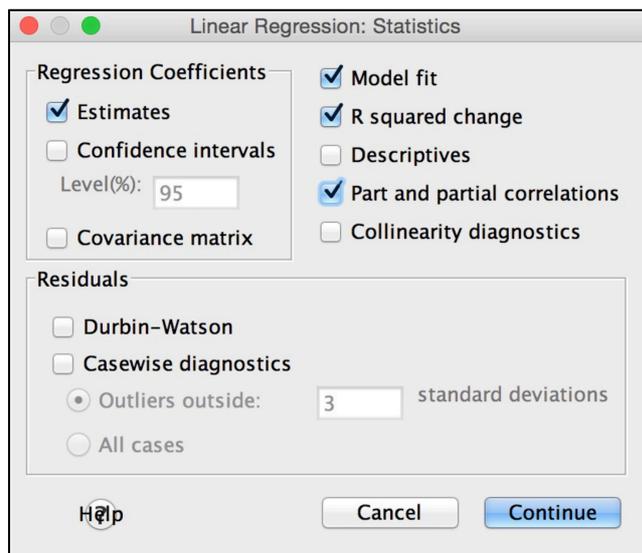
pb_soc	pb_fam	pb_leg	Conso_Dro
6,00	6,00	8,00	9,00
			5,00
			7,00
			9,00
			8,00
			7,00
			7,00
			9,00
			7,00
			7,00
			5,00
			6,00
			7,00
			7,00
			4,00
			4,00
			5,00



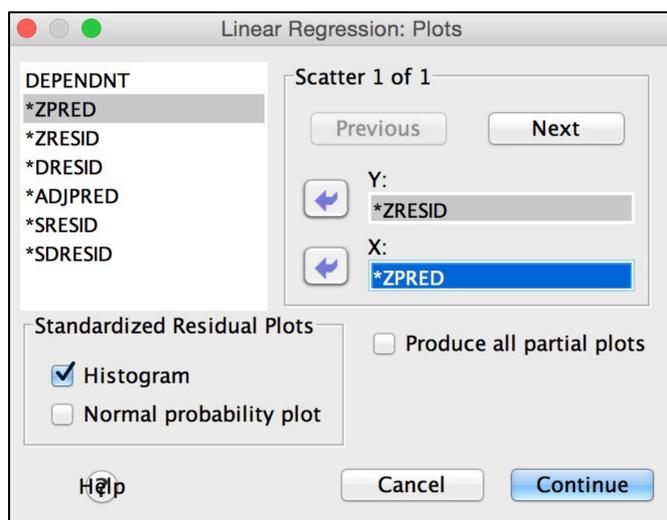
Entrez la VD à prédire (= critère) dans la case **Dependent:** et les VIs quantitatives (= prédicteurs) dans **Independent(s):** (si vous faites une régression simple, vous n'entrez qu'une variable dans la case **Independent(s)**).



Dans le Menu **Method:**, choisissez la méthode d'entrée des prédicteurs dans l'analyse de régression multiple (cf. cours pour la description des différentes méthodes); en cas de régression simple, peu importe la méthode, le résultat sera toujours pareil.



Dans l'onglet **Statistics...**, cochez en plus d'**Estimates** et **Model fit** (par défaut) **R squared change** et **Part and partial correlations**, puis cliquez **Continue**.



Dans l'onglet **Plots...** de **Linear Regression**, mettez *ZPRED (scores z prédits) dans l'axe des X et *ZRESID (scores z résiduels) dans l'axe des Y pour obtenir le scatterplot avec lequel vous pourrez juger des problèmes éventuels d'hétéroscédasticité (condition d'application). Cochez également **Histogram** des **Standardized Residual Plots** qui vous indiquera si les résidus sont bien distribués normalement (autre condition d'application), puis cliquez **Continue**, puis **OK**.

SPSS OUTPUT :

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Pbs médicaux	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).
2	Pbs familiaux	.	Stepwise (Criteria: Probability-of-F-to-enter <= ,050, Probability-of-F-to-remove >= ,100).

a. Dependent Variable: Gravité conso. drogue

Le premier tableau résume quels prédicteurs ont été retenus dans le modèle (ici « Pbs médicaux », « Pbs familiaux »), quelle est la VD (« Gravité conso. drogue »), et quelle est la méthode choisie (ici « Stepwise »).

Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,492 ^a	,242	,234	1,43217	,242	30,981	1	97	,000
2	,570 ^b	,324	,310	1,35921	,082	11,694	1	96	,001

a. Predictors: (Constant), Pbs médicaux

b. Predictors: (Constant), Pbs médicaux, Pbs familiaux

c. Dependent Variable: Gravité conso. drogue

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	63,546	1	63,546	30,981	,000 ^b
	Residual	198,959	97	2,051		
	Total	262,505	98			
2	Regression	85,150	2	42,575	23,045	,000 ^c
	Residual	177,355	96	1,847		
	Total	262,505	98			

a. Dependent Variable: Gravité conso. drogue

b. Predictors: (Constant), Pbs médicaux

c. Predictors: (Constant), Pbs médicaux, Pbs familiaux

Les deux tableaux suivants vous indiquent les indices utiles pour évaluer les deux modèles retenus ($R^2 = \%$ de variance expliquée, le gain de R^2 apporté par le modèle (= **R Square Change**), ainsi que la significativité de ce changement (**Sig. F Change**). Le tableau d'ANOVA teste si les prédicteurs du modèle prédisent une part significative de variance du critère ($H_0 : R^2 = 0$), ici c'est le cas, on rejette H_0 pour les deux modèles.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	4,766	,254		18,775	,000			
	Pbs médicaux	,367	,066	,492	5,566	,000	,492	,492	,492
2	(Constant)	3,748	,383		9,790	,000			
	Pbs médicaux	,334	,063	,447	5,269	,000	,492	,474	,442
	Pbs familiaux	,210	,061	,290	3,420	,001	,359	,330	,287

a. Dependent Variable: Gravité conso. drogue

Le tableau principal est celui-ci, qui vous indique quels prédicteurs ont été retenus dans les différents modèles, quelle est l'influence de ces prédicteurs (**Beta**) si l'influence de ces variables est significative (**Sig.**), les valeurs des t-tests (**t**), ainsi que les corrélations simples, partielles et semi-partielles entre les prédicteurs et le critère. Les coefficients non standardisés **B (Unstandardized Coefficients)** vous permettent d'écrire l'équation de régression.

Excluded Variables^a

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Pbs scolaires	,094 ^b	1,059	,292	,107	,981
	Pbs sociaux	,283 ^b	3,346	,001	,323	,988
	Pbs familiaux	,290 ^b	3,420	,001	,330	,976
2	Pbs scolaires	-,079 ^c	-,796	,428	-,081	,711
	Pbs sociaux	,175 ^c	1,719	,089	,174	,664

a. Dependent Variable: Gravité conso. drogue

b. Predictors in the Model: (Constant), Pbs médicaux

c. Predictors in the Model: (Constant), Pbs médicaux, Pbs familiaux

Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3,7481	7,7615	5,9293	,93213	99
Residual	-2,37751	2,95536	,00000	1,34527	99
Std. Predicted Value	-2,340	1,966	,000	1,000	99
Std. Residual	-1,749	2,174	,000	,990	99

a. Dependent Variable: Gravité conso. drogue

Les deux tableaux suivants vous indiquent les mêmes éléments pour les prédicteurs exclus du modèle, et les statistiques descriptives des résidus que nous ne commentons pas ici.

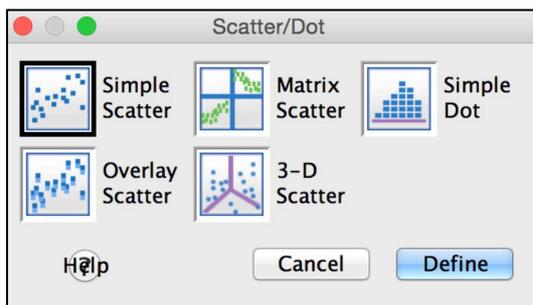
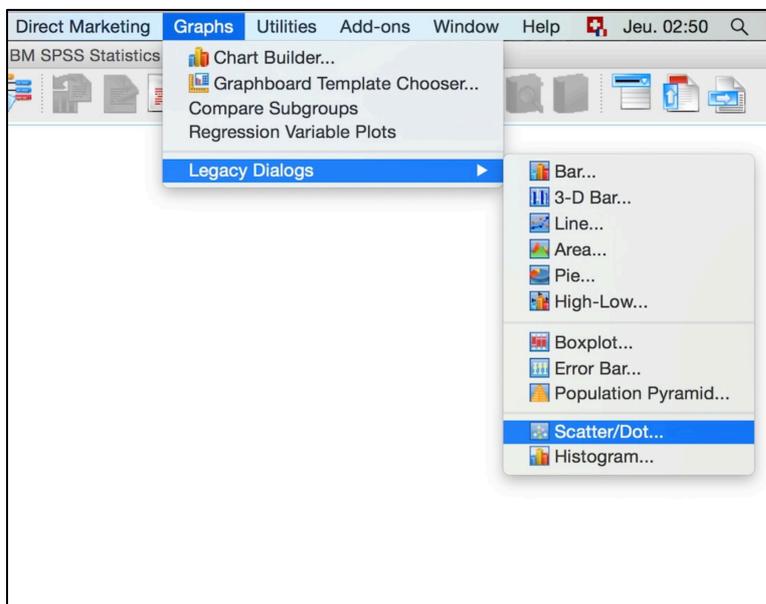
Notation : Une analyse de régression a été menée afin de voir quels types de problèmes (familiaux, scolaires, médicaux ou sociaux) prédisaient la gravité de la consommation de drogue chez des adolescent-e-s. Les prédicteurs retenus par l'analyse expliquent 31% de la variance du critère, $R^2 = .31$, ce qui est significativement différent de 0, $F(2, 96) = 23.04$, $p < .05$. L'importance des problèmes médicaux, $\beta = .45$, $t = 8.92$, $p < .05$, et le nombre de problèmes familiaux, $\beta = .29$, $t = 3.42$, $p < .05$ prédisent significativement et positivement la gravité de la consommation de drogues.

NB. Le lien de causalité entre problèmes médicaux et consommation de drogues est certainement inverse, la consommation de drogue induisant des problèmes médicaux et pas l'inverse, mais nous prenons ces résultats comme un exemple !

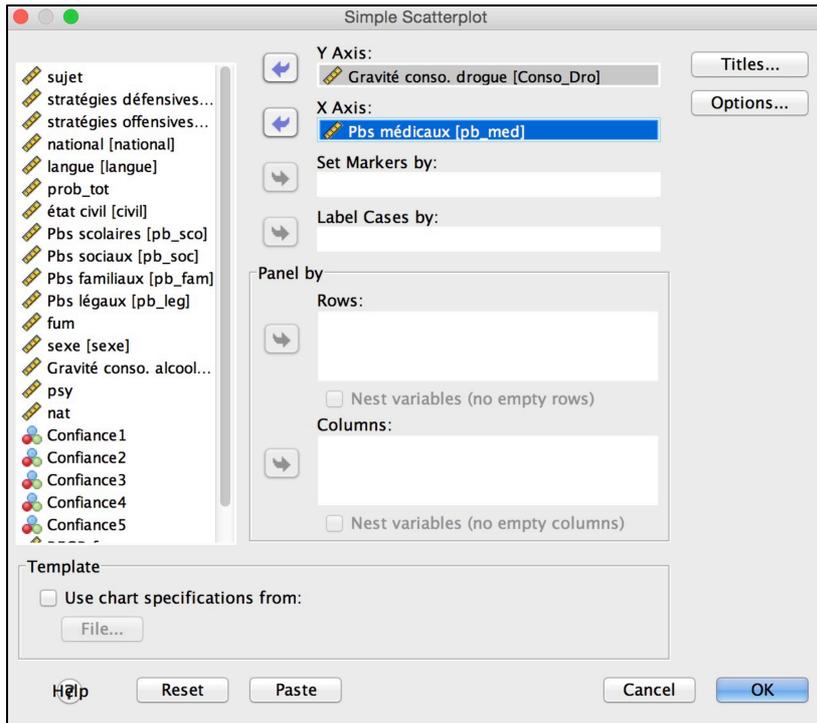
Bonus :

Un graphique très utile pour détecter des valeurs extrêmes bivariées (qui n'apparaissent pas forcément lorsque l'on contrôle la normalité univariée des variables) est le diagramme de dispersion entre un prédicteur et le critère (**Scatterplot**).

Chemin : **Graphs** → **Legacy Dialogs** → **Scatter/Dot...**



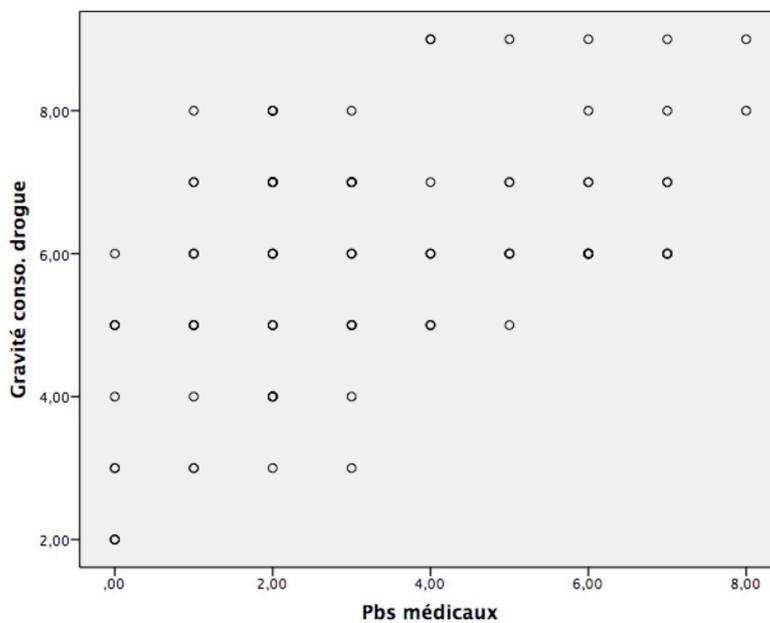
Choisissez **Simple Scatter** puis **Define**.



Entrez le critère dans la case **Y Axis** et le prédicteur dans la case **X Axis** puis cliquez **OK**.

SPSS OUTPUT :

Graph

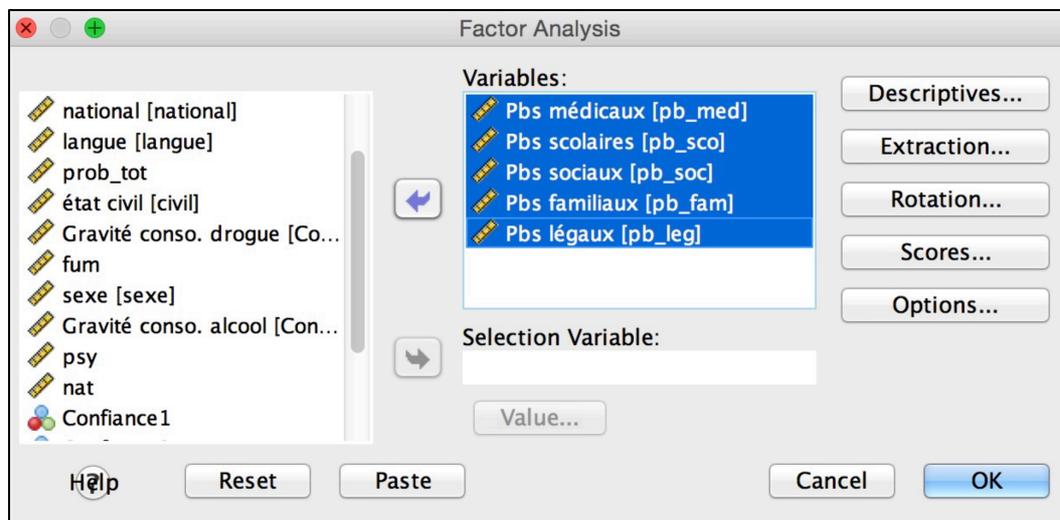
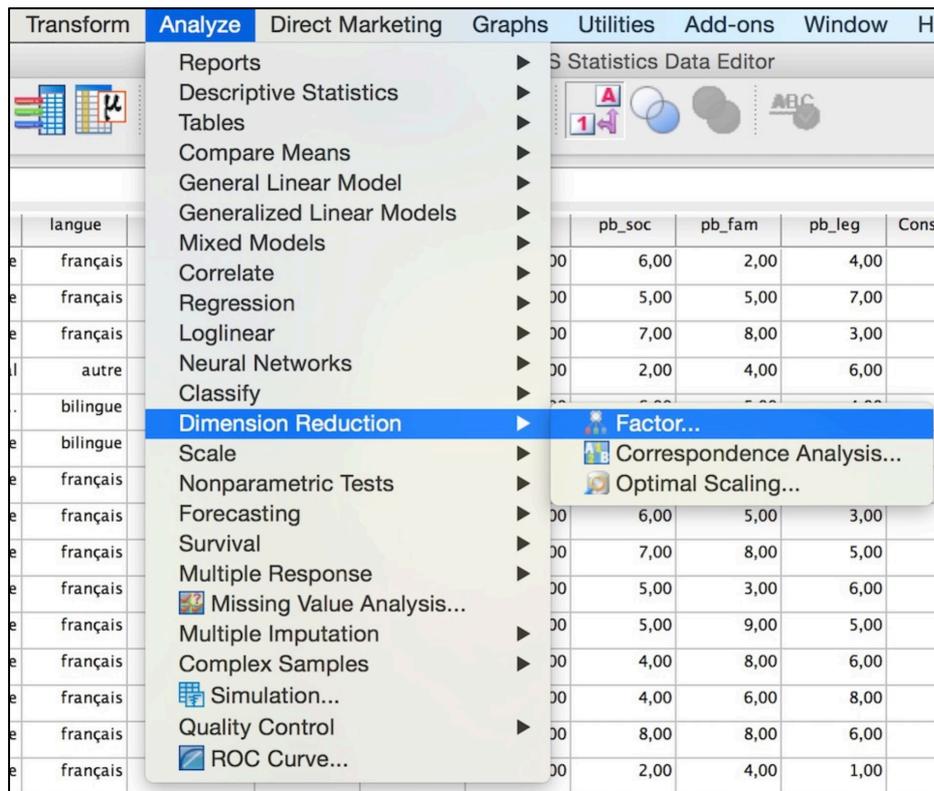


6.9. Analyse en Composantes Principales (ACP)

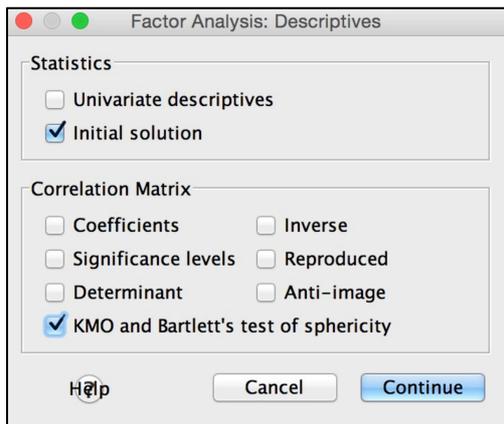
L'Analyse en Composantes Principales a pour but de réduire le nombre de variables quantitatives (par exemple les différents items d'un questionnaire) en un nombre plus réduit de dimensions formées par des variables corrélées entre elles.

Exemple : Nous aimerions savoir si les cinq types de problèmes rencontrés (familiaux, scolaires, médicaux, sociaux ou légaux) par des adolescent-e-s peuvent être réduits en un plus petit nombre de dimensions (appelés facteurs).

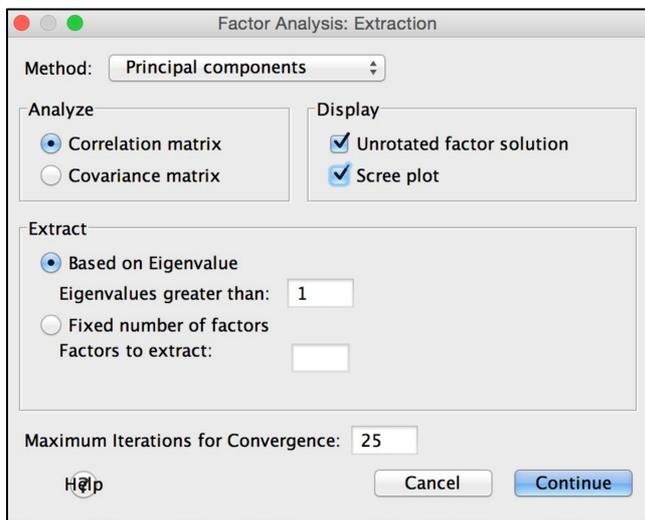
Chemin : **Analyze** → **Dimension Reduction** → **Factor...**



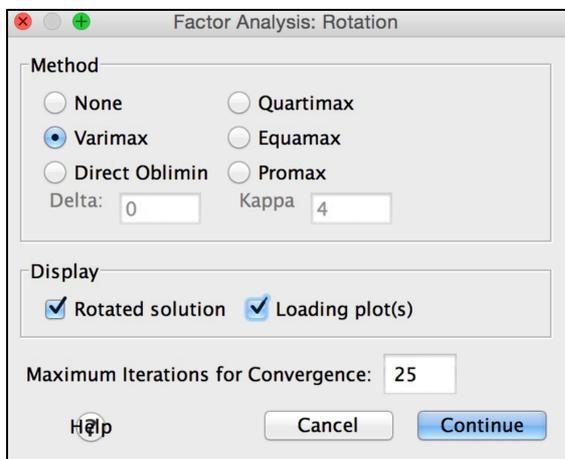
Entrez les variables dans la case... **Variables:**



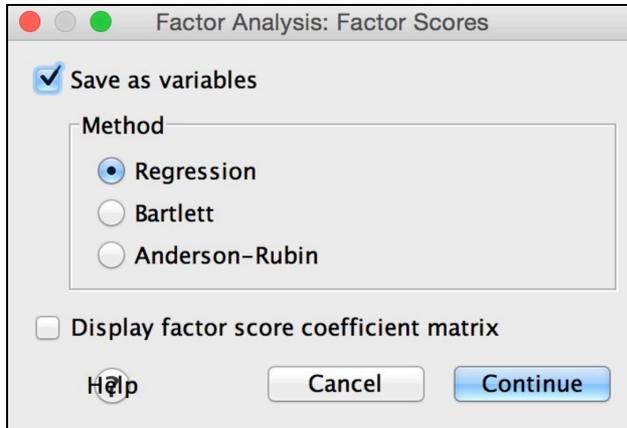
Dans l'onglet **Descriptives...** demandez en plus de **Initial solution** (coché par défaut) le **KMO and Bartlett's test of sphericity**, puis cliquez **Continue**.



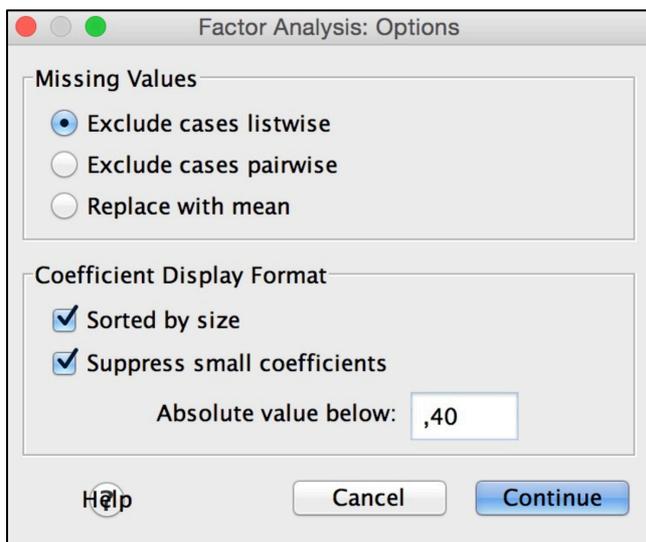
Dans l'onglet **Extraction...** cochez le **Scree plot** dans la case **Display** qui vous donnera une représentation graphique des valeurs propres utiles afin de déterminer le nombre de facteurs à retenir (scree test, critère de Cattell), puis cliquez **Continue**.



Dans l'onglet **Rotation...** vous pouvez choisir une méthode de rotation des facteurs (orthogonale = **Varimax**, oblique = **Direct Oblimin** ou **Promax**) et **Loading plot(s)** pour avoir une représentation graphique des saturations.



Dans l'onglet **Scores...** vous pouvez demander le calcul des scores factoriels (qui seront ajoutés dans le fichier de données) en cochant **Save as variables** (vous pouvez également cocher **Display factor score coefficient matrix** qui vous donnera les coefficients utilisés pour le calcul des scores factoriels), puis cliquez **Continue**.



Enfin dans l'onglet **Options...** il est utile de cocher **Sorted by size** qui ordonnera les saturations par ordre d'importance dans la matrice des saturations (ce qui facilitera la lecture du tableau), ainsi que **Suppress small coefficients** en introduisant la valeur ,40 dans la case **Absolute value below:**, ce qui enlèvera de la matrice des saturations les saturations faibles inutiles pour l'interprétation (et qui rendent la lecture du tableau plus difficile), puis cliquez **Continue**, puis **OK**.

SPSS OUTPUT :

Factor Analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,624
Bartlett's Test of Sphericity	Approx. Chi-Square	156,373
	df	10
	Sig.	,000

Communalities

	Initial	Extraction
Pbs médicaux	1,000	,846
Pbs scolaires	1,000	,680
Pbs sociaux	1,000	,720
Pbs familiaux	1,000	,708
Pbs légaux	1,000	,839

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,428	48,554	48,554	2,428	48,554	48,554	2,112	42,234	42,234
2	1,366	27,314	75,868	1,366	27,314	75,868	1,682	33,634	75,868
3	,508	10,159	86,027						
4	,430	8,605	94,632						
5	,268	5,368	100,000						

Extraction Method: Principal Component Analysis.

Le premier tableau indique le **KMO** et le **test de sphéricité** (qui doit être significatif, on doit rejeter H_0 qui stipule que les corrélations entre les variables sont nulles). Le second tableau est celui des communalités (**Communalities**) qui indique quelle proportion de la variance de chaque variable est expliquée par les facteurs retenus. Le troisième tableau indique combien de facteurs ont été retenus selon le critère de Kaiser (valeurs-propres > 1), et le pourcentage de variance expliquée par ces facteurs, avant et après rotation (si **Varimax**).

Component Matrix^a

	Component	
	1	2
Pbs scolaires	,774	
Pbs sociaux	,751	
Pbs familiaux	,739	-,403
Pbs légaux	,666	,629
Pbs médicaux	,526	,754

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Rotated Component Matrix^a

	Component	
	1	2
Pbs sociaux	,845	
Pbs familiaux	,839	
Pbs scolaires	,804	
Pbs médicaux		,919
Pbs légaux		,891

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Les matrices de saturation avant et après rotation permettent d'interpréter les facteurs retenus (avec rotations obliques, vous obtenez deux matrices de saturation après rotation, **Pattern matrix** et **Structure matrix** ; bien que souvent similaires, interprétez de préférence la première ; les rotations obliques vous donnent également un tableau **Component Correlation Matrix** vous indiquant les corrélations de Pearson entre les facteurs).

Les scores factoriels « FAC1_1 » (score factoriel 1 de l'ACP 1) et « FAC2_1 » apparaissent dans votre fichier de données, deux nouvelles variables quantitatives sur lesquelles vous pouvez faire des tests statistiques (ANOVA, régression, etc.).

Confiance5	FAC1_1	FAC2_1	va
0	,23850	2,16358	
-2	-,23540	-1,46261	
2	,62426	-,59054	
1	,29278	-1,27258	
-1	,95623	1,49108	
1	,25109	2,13125	
0	1,03753	,09294	
1	2,07291	-,77216	
2	,57485	1,08010	
2	-,70182	1,22244	
1	1,91087	,09989	
0	-1,33703	1,40377	
1	1,22167	,99947	
2	1,95637	-1,03325	
-1	,27663	-1,04293	
2	1,19376	1,22587	
1	-1,42967	-,86535	
0	,69673	,12660	
0	-1,16435	-,41085	
-2	1,09713	-1,04624	
1	-,75363	-,90335	

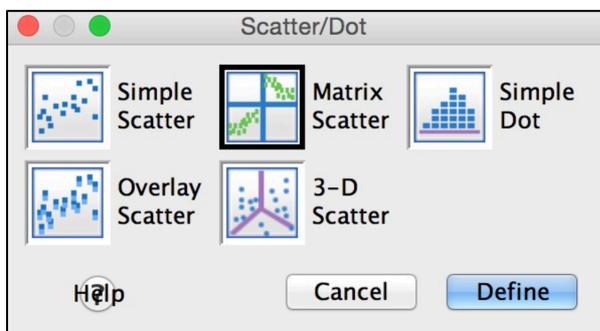
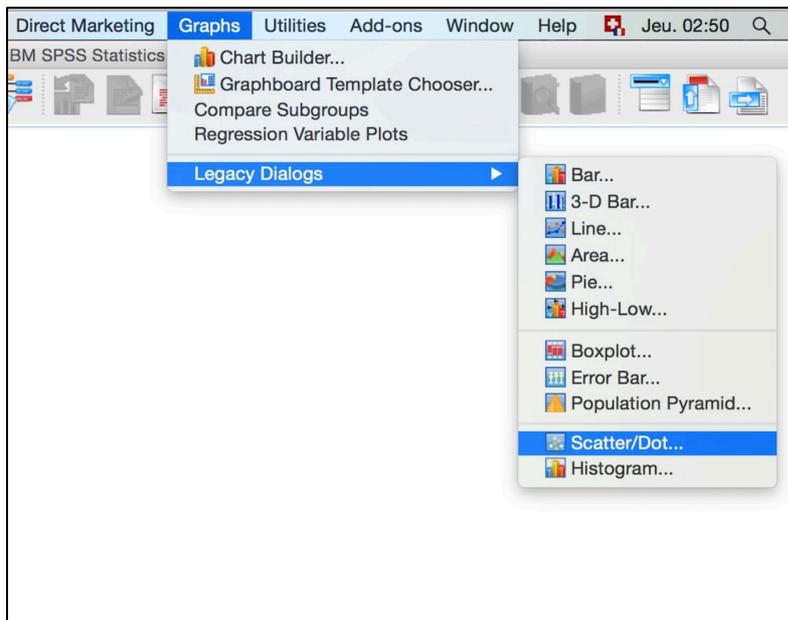
Notation : Afin d'évaluer si les problèmes rencontrés dans différents domaines par des adolescent-e-s sont liés entre eux, et par conséquent pourraient être résumés en un nombre plus réduit de dimensions, une analyse en composantes principales (ACP) avec rotation Varimax a été conduite sur les cinq domaines de problèmes évalués, les prérequis suivants étant remplis : ces variables sont mesurées sur des échelles quantitatives, le nombre total de sujets est plus de cinq fois supérieur au nombre de variables (99 sujets pour 5 variables), et l'analyse de la normalité uni- et bivariée s'est montrée satisfaisante ; de plus, 6 corrélations moyennes à fortes, toutes significatives, ont été observées entre les 5 variables. Le test de Bartlett significatif et un KMO minimal confirmant la pertinence d'une ACP.

Une solution à 2 composantes a été retenue selon le critère de Kaiser-Guttman (valeur-propres > 1). La première composante explique 42% de la variance totale et regroupe les problèmes scolaires, sociaux et familiaux (avec des saturations respectives après rotation de .80, .85 et .84). L'appellation « problèmes relationnels » a été retenue dans la mesure où ce type de difficultés semblant être le noyau des problèmes dans ces différents domaines. La deuxième composante explique 34% de la variance totale et regroupe les problèmes médicaux et légaux (avec des saturations respectives après rotation de .92 et .89), nous la nommons donc « problèmes médico-légaux ». Toutes les variables sont expliquées de manière satisfaisante par la solution retenue, avec des communautés allant de .68 (problèmes scolaires) à .85 (problèmes médicaux).

Bonus :

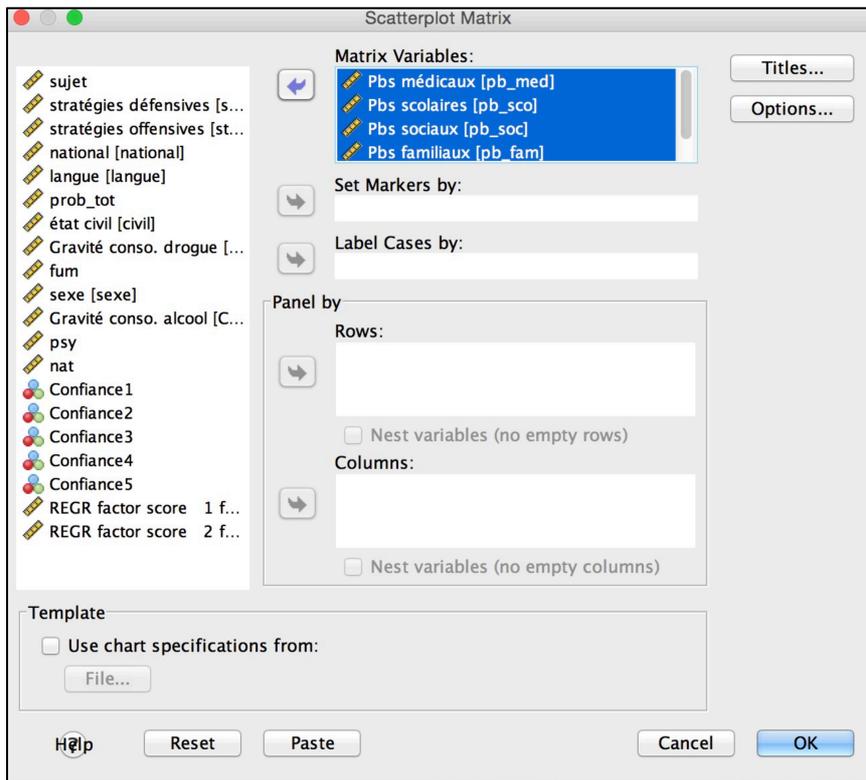
Un graphique très utile pour la condition d'application de l'ACP qui demande la normalité univariée mais aussi bivariée est le **Scattermatrix**, une matrice des diagrammes de dispersion entre les variables.

Chemin : **Graphs** → **Legacy Dialogs** → **Scatter/Dot...**



Choisissez **Matrix Scatter** puis **Define**.

Entrez les variables dans **Matrix Variables**: puis cliquez **OK**.



SPSS OUTPUT :

Graph

