

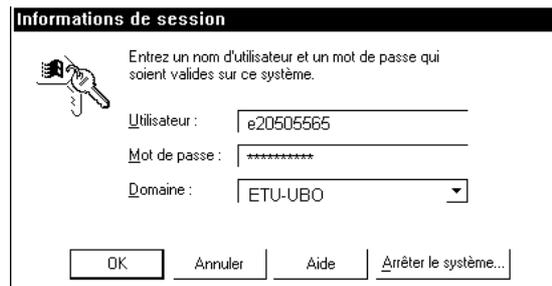
Master de Psychologie - 1^{ère} année

PSY73B : Informatique : traitement des données - TD N°1

Test de Student : mise en oeuvre et conditions d'application avec Statistica

1. Organiser son espace de travail sous Statistica:

Affichez le dialogue d'ouverture de session en appuyant simultanément sur les trois touches Ctrl+Alt+Suppr. Complétez le dialogue en ouvrant la session à l'aide de vos identifiants ENT :



N.B. Pour des raisons de confidentialité, le mot de passe ne s'affiche pas "en clair".

Remarque 1. Si vous ne disposez pas encore de votre identifiant ENT, ou si votre mot de passe n'est pas reconnu, vous pouvez ouvrir une session en utilisant le compte :

Utilisateur : LETA20xPyy\etudiant
Mot de passe : ubo

Cependant, la configuration de nos appareils impose que la session soit ouverte par un utilisateur identifié, dans le domaine ETU-UBO pour que le logiciel Statistica soit disponible. Si vous avez ouvert la session avec l'identifiant étudiant, le mot de passe ubo, vous devez ensuite effectuer le montage du disque réseau contenant le logiciel Statistica (ainsi que, de préférence, de celui qui contient les fichiers utilisés en TD) à l'aide du menu Monter un volume réseau du poste de travail. Pour ces montages, vous devez utiliser les coordonnées ENT d'un utilisateur reconnu par l'ENT (l'enseignant, un collègue étudiant, etc.).

Paramètres pour le montage du volume réseau contenant le logiciel Statistica :

- Volume réseau : \\servsciences\statistica
- Lettre de lecteur : S:
- Utilisateur : login et mot de passe d'un utilisateur reconnu sur l'ENT.

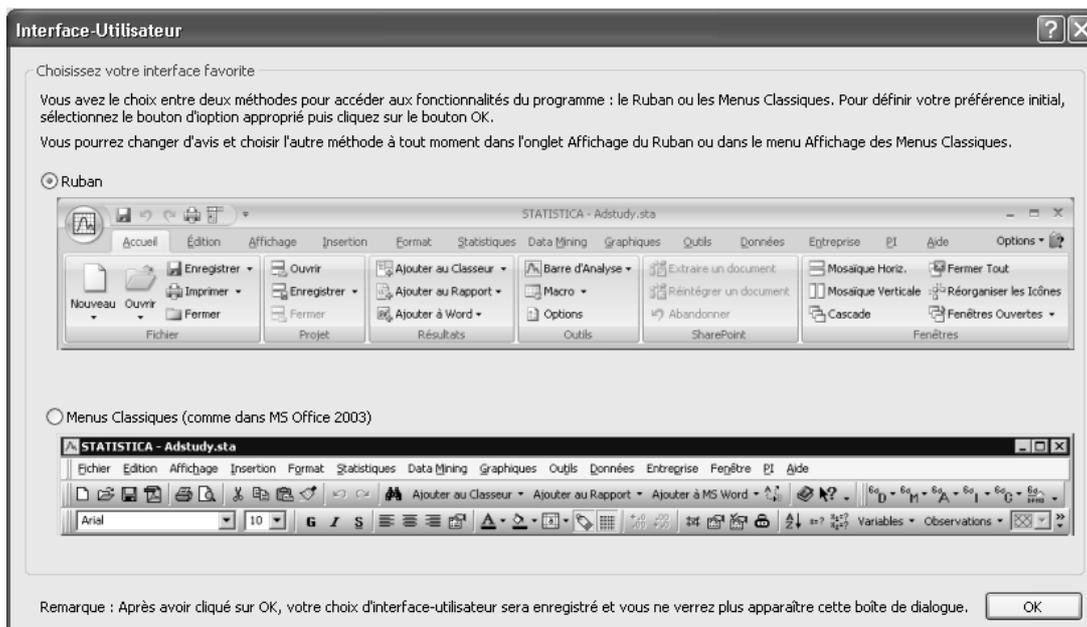
Paramètres pour le montage du volume contenant les fichiers utilisés en TD :

- Volume réseau : \\serv-bu\tdlettres ou \\172.18.127.1\tdlettres
- Lettre de lecteur : W:
- Utilisateur : login et mot de passe d'un utilisateur reconnu sur l'ENT.

Chargez le logiciel Statistica en double-cliquant sur l'icône présente sur le bureau. La configuration par défaut du logiciel n'est pas vraiment satisfaisante. Nous allons donc commencer par adapter la configuration à nos besoins.

1.1. Au premier chargement du logiciel

Le logiciel propose deux options possibles pour l'affichage des menus. Vous pouvez choisir celle que vous préférez. Notez toutefois que les copies d'écran de ce polycopié utilisent l'option "Menus Classiques".



N.B. Il sera toujours possible de basculer d'une interface à l'autre à l'aide des menus Affichage - Ruban des menus classiques ou Affichage - Menus Classique de l'autre option;

Le logiciel demande ensuite si l'on souhaite installer les composants permettant au logiciel d'interagir avec R. Vous pouvez décliner la proposition et cocher l'option "Ne plus me proposer cette boîte de dialogue" car vous n'avez pas les droits d'administration permettant de faire cette installation.

1.2. Le menu Outils - Options

Le menu Outils - Options contient de nombreuses possibilités de paramétrage de Statistica. Heureusement, seules quelques-unes d'entre elles méritent d'être retouchées.

Ouvrez la fenêtre de dialogue accessible par le menu Outils-Options et explorez les différents onglets qui y sont rassemblés.

N.B. Les options ainsi choisies sont enregistrées dans le profil de l'utilisateur lorsque l'on quitte le logiciel. // n'y a aucun enregistrement si le compte est verrouillé ou si Statistica se plante en cours de travail.

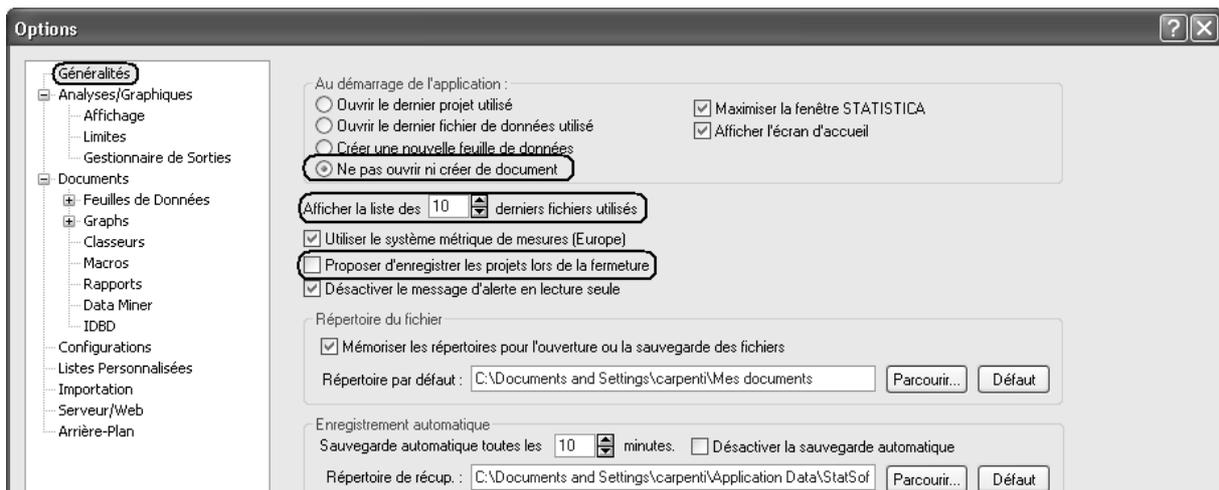
1.2.1 Spécifier le répertoire d'enregistrement par défaut

Affichez les options disponibles sous l'onglet Généralités.

- Choisissez de préférence l'option : *Au démarrage de l'application, ne pas ouvrir ni créer de document.*, les autres options étant plutôt déroutantes.

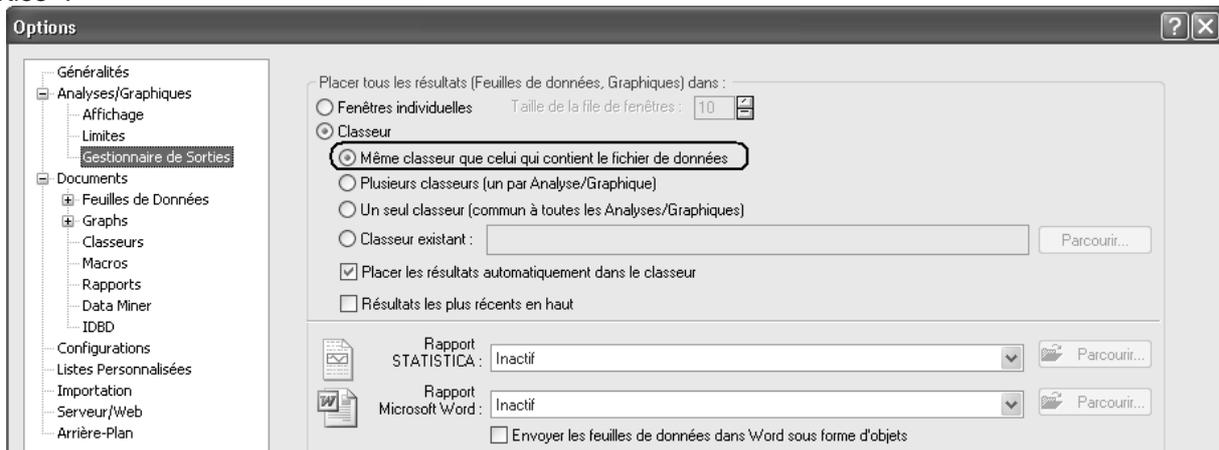
- Par défaut, Statistica affiche les noms des 16 derniers fichiers utilisés. On peut modifier ce comportement. Par exemple, on peut aussi, sans inconvénient, réduire la longueur de la liste à 10 au lieu de 16.

- L'option : *Proposer d'enregistrer les projets lors de la fermeture* peut également être déroutante si on ignore ce qu'est un fichier de projet Statistica (cf. § 1.3.3) et peut sans inconvénient être désactivée.

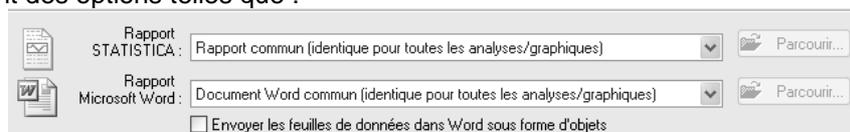


1.2.2 Gérer les sorties

La manière la plus commode de gérer nos documents avec Statistica consiste à rassembler dans un même classeur la ou les feuilles de données et les résultats de traitements concernant ces données. Ce comportement sera obtenu à l'aide du réglage suivant, sous l'onglet "Analyses/Graphiques - Gestionnaire de sorties" :



Il peut également être commode de demander à Statistica de placer une copie des résultats dans un rapport, en utilisant des options telles que :



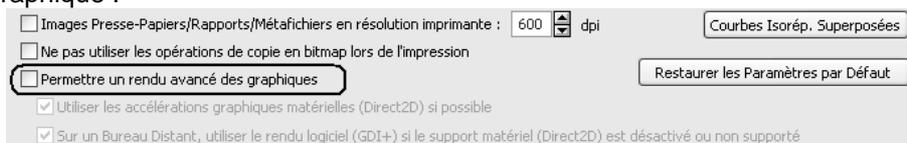
En effet un rapport peut être enregistré au format .rtf pour être relu sur une autre machine par un logiciel de traitement de textes, même si Statistica n'est pas installé sur l'appareil. *Cependant, cette pratique présente plus d'inconvénients que d'avantages.* En effet :

- Les rapports produisent rapidement des fichiers très volumineux. Un rapport, ou un classeur contenant un ou des rapports devra être compressé (zippé) avant d'être envoyé par mail. Et par ailleurs, un rapport trop volumineux semble provoquer des plantages du logiciel dans certains cas.
- Si plusieurs séances de travail sont nécessaires pour réaliser le traitement, un nouveau rapport sera créé à chaque séance, ce qui est assez peu pratique.

En revanche, on pourra utiliser un rapport pour y taper de courts commentaires textuels, l'interprétation du résultat d'un traitement par exemple.

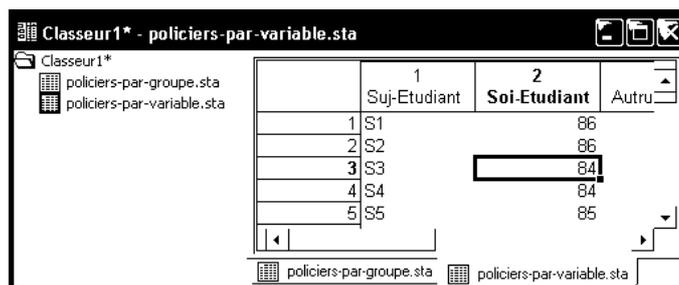
1.2.3 Réglages concernant les graphiques

Sous l'onglet Documents > Graphs, veillez à désactiver l'option "Permettre un rendu avancé des graphiques". En effet, sur nos postes, lorsque cette option est active, Statistica se plante dès que l'on essaie de réaliser un graphique :



1.2.4 La feuille de données active

Les traitements demandés via les menus s'appliquent à la fenêtre de données **active**. Dans le cas de données rassemblées dans plusieurs fenêtres indépendantes, la feuille active est celle qui se trouve au premier plan sur l'écran. Dans le cas d'un classeur, la feuille active est repérée par un liseré rouge :



Dans le classeur ci-dessus, la feuille active est "policiers-par-variable.sta"

On peut rendre active une feuille, ou changer de feuille active :

- soit en cliquant sur l'icône de la feuille et en utilisant le menu : Classeur - Feuille de données active;
- soit en cliquant avec le bouton droit sur l'icône de la feuille et en utilisant l'item "Feuille de données active" du menu local.

N.B. Il faut parfois rendre inactive la feuille actuellement active (mêmes menus que ci-dessus) pour pouvoir en activer une autre.

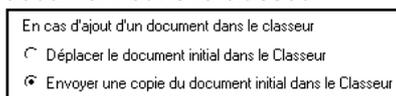
1.3. Manipulations de base sur un classeur

1.3.1 Copier - coller entre classeurs, entre un classeur et un objet Statistica

Pour déplacer un objet d'un classeur à un autre, il suffit de déplacer son icône depuis le volet gauche du premier classeur dans le volet gauche du second. On peut également utiliser les menus locaux Copier et Coller obtenus à l'aide d'un clic droit dans le volet gauche de chaque classeur.

Le menu local "Insérer" du volet gauche d'un classeur permet également d'insérer dans ce classeur un document contenu dans une fenêtre indépendante. Il suffit de choisir les options : Document Statistica - Créer à partir d'une fenêtre.

L'opération faite par Statistica est soit une copie (l'original de l'objet est conservé) soit un déplacement (l'original de l'objet n'est pas conservé) selon le paramétrage choisi dans le menu Outils - Options - Onglet Classeurs - Item "En cas d'ajout d'un document dans le classeur".



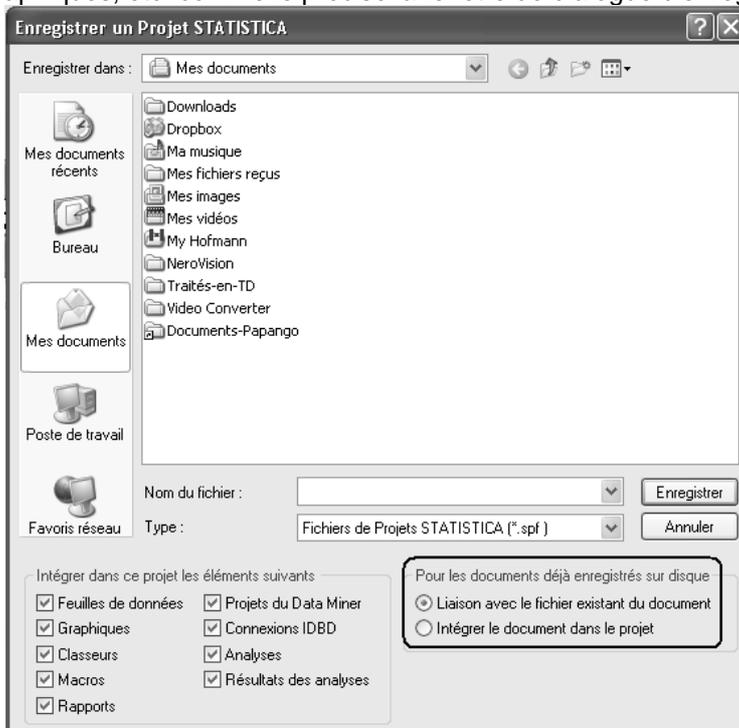
1.3.2 Supprimer un objet d'un classeur

Il est également possible de supprimer un objet d'un classeur, à l'aide d'un clic droit et de l'item de menu Supprimer. Cela permet notamment de ne garder, pour un traitement donné, que le résultat le plus abouti. Attention cependant : lorsque l'on supprime un objet qui n'est pas une feuille de la hiérarchie, on supprime en même temps tous les objets qui en dépendent.

1.3.3 Qu'est ce qu'un fichier de "Projet"

On peut enregistrer un projet soit en réponse à une fenêtre de dialogue, si l'option : *Proposer d'enregistrer les projets lors de la fermeture* est active soit en utilisant le menu Fichier - Enregistrer le projet....

Un fichier de projet permet de mémoriser un "instantané" au cours d'une séance de travail : feuilles de données et de résultats, analyses actives, etc. Mais, avec les options par défaut, le fichier lui-même ne contient pas les données proprement dites, il contient seulement des liens vers les classeurs, feuilles de données, feuilles de graphiques, etc. comme le précise la fenêtre de dialogue d'enregistrement :



Avec l'option par défaut, ce format de fichier ne permet donc pas de recopier votre travail sur un autre compte, ou de transmettre votre travail à un autre utilisateur. Le logiciel enverra également des messages d'erreur si vos fichiers sont déplacés ou renommés après l'enregistrement du projet.

2. Tests de Student avec Statistica

2.1. Quelques traitements sur le cas "Policiers"

Ouvrez à partir du serveur les deux feuilles [policiers-par-groupe.sta](#) et [policiers-par-variable.sta](#). Définissez un nouveau classeur et insérez les deux feuilles précédentes dans ce classeur. Enregistrez ce classeur sur votre compte, sous le nom [Policiers.stw](#). Rendez active l'une des deux feuilles.

2.1.1 Enoncé du cas :

Siiter et Ellison (1984) se sont proposé d'étudier les stéréotypes de la "personnalité du policier" Aux dires de ces chercheurs, il existe une croyance communément admise selon laquelle les fonctionnaires chargés de l'application de la loi constituent un groupe socio-professionnel à part. Beaucoup de personnes croient que les représentants de la police sont en général plus autoritaires que les autres gens. Siiter et Ellison ont testé ce stéréotype dans deux groupes de sujets de sexe masculin: un groupe d'étudiants et un groupe de fonctionnaires de police. Tous les sujets ont, à deux reprises, rempli une échelle standard d'autoritarisme: une fois par rapport à leurs propres croyances, et l'autre fois en imaginant comment répondrait un membre typique de l'autre groupe (facteur "Cible" à deux modalités: "soi-même" et "autrui").

Dans une reprise de cette expérience, on utilise 62 sujets dans chacun des deux groupes. Les données observées sont rassemblées, sous deux formes différentes, dans les fichiers de données Statistica [policiers-par-groupe.sta](#) et [policiers-par-variable.sta](#).

2.1.2 Comment faut-il structurer les données ?

Nous disposons ici de données relatives à deux groupes indépendants de sujets, et de deux observations pour chaque sujet. Autrement dit, ces données correspondent à un plan d'expérience du type $S < A > * B$. Quelle est alors la meilleure façon de structurer ces données dans une feuille de données Statistica ?

Par groupe :

	1 Groupe	2 Sujet	3 Soi	4 Autrui
1	Etudiant	S1	86	107
2	Etudiant	S2	86	105
3	Etudiant	S3	84	124
4	Etudiant	S4	84	135
5	Etudiant	S5	85	121
78	Officier	S78	78	97
79	Officier	S79	107	86
80	Officier	S80	113	105
81	Officier	S81	91	129

Par variable :

	1 Suj-Etudiant	2 Soi-Etudiant	3 Autrui-Etudiant	4 Suj-Policier	5 Soi-policier	6 Autrui-policier
1	S1	86	107	S63	117	134
2	S2	86	105	S64	94	83
3	S3	84	124	S65	112	93
4	S4	84	135	S66	94	112
5	S5	85	121	S67	123	106
6	S6	84	135	S68	85	94
7	S7	104	96	S69	135	62

Il n'existe malheureusement pas de réponse évidente à cette question :

- La structuration "par groupe" attribue une ligne du fichier de données à chaque individu statistique, et semble donc plus satisfaisante d'un point de vue théorique. Cependant, pour de nombreux traitements, nous devons avoir recours à des filtres, assez peu commodes sous Statistica.

- La structuration "par variable" rend plus simples certains traitements (étude de la normalité de la VD pour chaque combinaison des deux facteurs, par exemple), mais ne permet pas d'effectuer d'autres traitements, tels que les tests d'homogénéité de variances.

Finalement, la meilleure solution est peut-être de garder à disposition les données sous les deux formes.

2.1.3 Affichage d'un fichier de données comportant de nombreuses observations : possibilité de filtrer l'affichage

Il est possible de n'afficher qu'une partie des données. Pour cela :

- sélectionner une ou plusieurs colonnes
- utiliser le menu Données - Filtre automatique - Filtre automatique.

Les têtes des colonnes sélectionnées comportent alors des listes déroulantes permettant de gérer l'affichage selon différents critères.

Par exemple, activez le filtre automatique pour les colonnes de la feuille policiers-par-groupes, puis limitez l'affichage aux sujets pour lesquels l'évaluation de soi vaut 84.

	1 Groupe	2 Sujet	3 Soi	4 Autrui
3	Etudiant	S3	84	124
4	Etudiant	S4	84	135
6	Etudiant	S6	84	135
103	Officier	S103	84	61

2.1.4 Utiliser les tests de Student sur groupes indépendants et sur groupes appariés pour traiter les données proposées

Lors d'une étude sur une partie de ces données, la conclusion obtenue était la suivante :

Les étudiants et les policiers sont d'accord pour estimer les policiers plus autoritaires que les étudiants. En effet, d'une part, les estimations de soi vont dans ce sens, d'autre part les comparaisons "soi-même" v/s "autrui" confirment le résultat. D'autre part, s'agissant des étudiants, les estimations de soi (réalisées par les étudiants) et les estimations d'autrui (réalisées par les policiers) concordent. L'autoritarisme des étudiants est estimé de la même façon par les étudiants et les policiers: les deux échelles sont donc comparables. Enfin, s'agissant des policiers, les estimations de soi (réalisées par les policiers) sont significativement plus faibles que les estimations d'autrui. (réalisées par les étudiants). Combiné avec le résultat précédent, ce résultat semble montrer que l'autoritarisme des policiers est surestimé par les étudiants.

A l'aide de tests de Student convenablement choisis, étudier si les données fournies sont en accord avec les affirmations ci-dessus. Réaliser les tests d'une part sur les données structurées "par groupe" et d'autre part sur les données structurées "par variable".

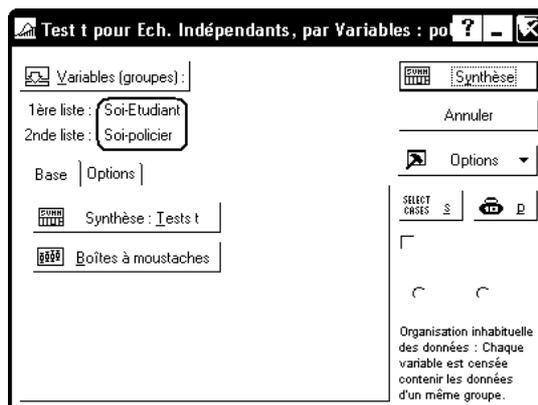
2.1.5 Réalisation des tests pour des données structurées "par variable"

Pour réaliser ces tests, il faut au préalable rendre active la feuille "policiers-par-variable".

Affirmation N° 1 : *Les étudiants et les policiers sont d'accord pour estimer les policiers plus autoritaires que les étudiants. En effet, d'une part, les estimations de soi vont dans ce sens*

On compare les scores d'autoritarisme que se sont attribués les étudiants à ceux que se sont attribués les policiers. Il s'agit d'une comparaison de moyennes sur deux groupes indépendants.

Utilisez le menu Statistiques - Statistiques élémentaires - Test t pour échantillons indépendants - par variable.



Vous devriez obtenir le résultat suivant :

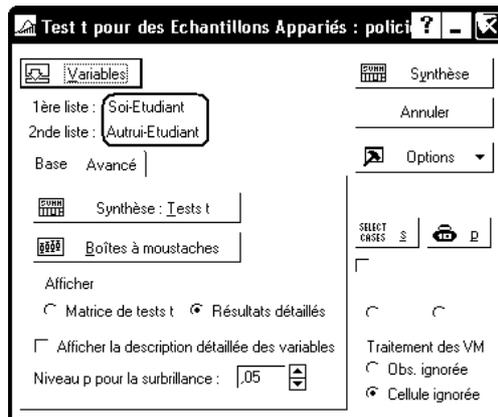
Test t pour des Echantillons Indépendants (policiers-par-variable.sta dans Classeur										
Note : Variables traitées comme des échantillons indépendants										
	Moyenne Groupe 1	Moyenne Groupe 2	valeur t	df	p	N Actifs Groupe 1	N Actifs Groupe 2	Ec-Type Groupe 1	Ec-Type Groupe 2	Ratio 'arian
Groupe1 vs. Groupe2	1	2				1	2	1	2	
Soi-Etudiant vs. Soi-policier	87,15	94,97	-2,33	122	0,0212	62	62	19,02	18,29	1

Le niveau de significativité du test est ici $p=0,0212$ ou 2,12%. La différence est donc significative au seuil de 5%.

Affirmation N° 2 : *les étudiants et les policiers sont d'accord pour estimer les policiers plus autoritaires que les étudiants (...)* d'autre part les comparaisons "soi-même" v/s "autrui" confirment le résultat

Les étudiants estiment-ils les policiers plus autoritaires qu'ils ne s'estiment eux-mêmes ? Il s'agit là d'une comparaison de deux moyennes sur des groupes appariés (dans les deux cas, les scores ont été attribués par les étudiants).

Utilisez le menu Statistiques - Statistiques élémentaires - Test t pour des échantillons appariés



Test t pour des Echantillons Appariés (policiers-par-variable.sta dans Classeur1) Différences significatives marquées à $p < ,05000$								
Variable	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
Soi-Etudiant	87,1452	19,02358						
Autrui-Etudiant	118,9516	18,34231	62	-31,8065	26,75011	-9,36236	61	0,000000

Même traitement pour les policiers ; vous devriez obtenir :

Test t pour des Echantillons Appariés (policiers-par-variable.sta dans Classeur1) Différences significatives marquées à $p < ,05000$								
Variable	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
Soi-policier	94,96774	18,28998						
Autrui-policier	88,14516	18,93808	62	6,822581	24,09171	2,229856	61	0,029448

Affirmation N° 3 : *D'autre part, s'agissant des étudiants, les estimations de soi (réalisées par les étudiants) et les estimations d'autrui (réalisées par les policiers) concordent.*

On compare ici les variables "Soi-Etudiant" et "Autrui-Policier", définies sur deux groupes indépendants. Vous devriez obtenir :

Test t pour des Echantillons Indépendants (policiers-par-variable.sta dans Classeur1) Note : Variables traitées comme des échantillons indépendants										
Groupe1 vs. Groupe2	Moyenne Groupe 1	Moyenne Groupe 2	valeur t	dl	p	N Actifs Groupe 1	N Actifs Groupe 2	Ec-Type Groupe 1	Ec-Type Groupe 2	Ratio /variance
Soi-Etudiant vs. Autrui-polici	87,15	88,15	-0,29	122	0,7698	62	62	19,02	18,94	

On obtient là un niveau de significativité $p = 77\%$. Il n'existe donc pas de différence significative entre les moyennes des deux variables.

Affirmation N° 4 : *Enfin, s'agissant des policiers, les estimations de soi (réalisées par les policiers) sont significativement plus faibles que les estimations d'autrui. (réalisées par les étudiants).*

On compare les variables Autrui-Etudiant et Soi-Policier. On obtient :

Test t pour des Echantillons Indépendants (policiers-par-variable.sta dans Classeur1) Note : Variables traitées comme des échantillons indépendants										
Groupe1 vs. Groupe2	Moyenne Groupe 1	Moyenne Groupe 2	valeur t	dl	p	N Actifs Groupe 1	N Actifs Groupe 2	Ec-Type Groupe 1	Ec-Type Groupe 2	Ratio /variance
Autrui-Etudiant vs. Soi-polici	118,95	94,97	7,29	122	0,0000	62	62	18,34	18,29	1,

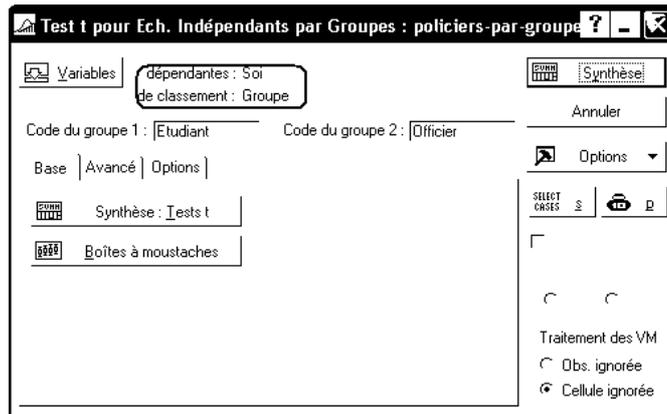
2.1.6 Réalisation des tests pour des données structurées "par groupe"

Rendons maintenant active la feuille "policiers-par-groupe".

Affirmation N° 1 : *Les étudiants et les policiers sont d'accord pour estimer les policiers plus autoritaires que les étudiants. En effet, d'une part, les estimations de soi vont dans ce sens*

On compare les scores d'autoritarisme que se sont attribués les étudiants à ceux que se sont attribués les policiers. Il s'agit d'une comparaison de moyennes sur deux groupes indépendants.

Utilisez le menu Statistiques - Statistiques élémentaires - Test t pour échantillons indépendants - par groupes.



Vous devriez retrouver le résultat précédent sous la forme suivante :

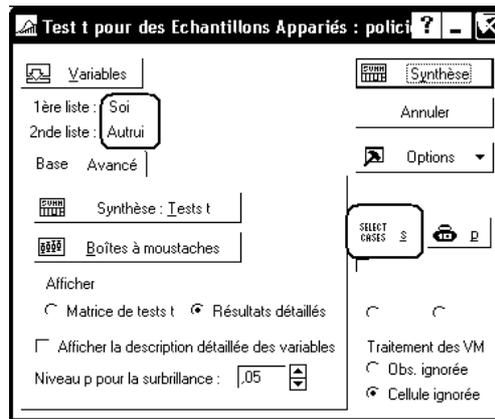
Tests t ; Classmt : Groupe (policiers-par-groupe.sta dans Classeur1)										
Groupe1: Etudiant										
Groupe2: Officier										
Variable	Moyenne Etudiant	Moyenne Officier	Valeur t	dl	p	N Actifs Etudiant	N Actifs Officier	Ecart-Type Etudiant	Ecart-Type Officier	Ratio /arian
Soi	87,15	94,97	-2,33	122	0,0212	62	62	19,02	18,29	1

Affirmation N° 2 : *les étudiants et les policiers sont d'accord pour estimer les policiers plus autoritaires que les étudiants (...)* d'autre part les comparaisons "soi-même" v/s "autrui" confirment le résultat

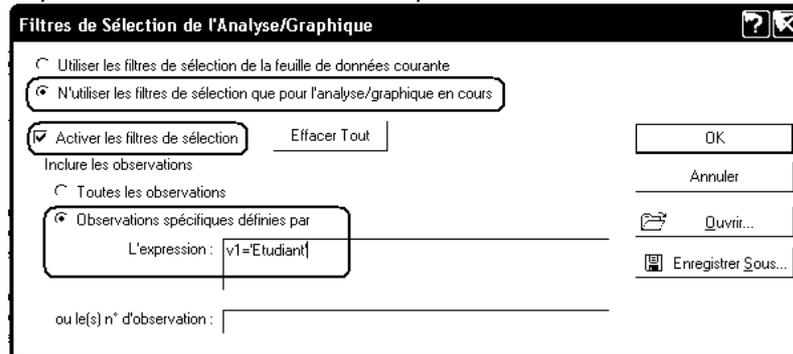
Les étudiants estiment-ils les policiers plus autoritaires qu'ils ne s'estiment eux-mêmes ?

Il s'agit encore d'une comparaison de deux moyennes sur des groupes appariés, mais pour retrouver les résultats précédents, on doit ajouter un filtre pour indiquer à Statistica que seules les données relatives aux étudiants doivent être prises en compte.

On utilise donc encore le menu Statistiques - Statistiques élémentaires - Test t pour des échantillons appariés. On complète la fenêtre de dialogue en indiquant les variables concernées, et on utilise le bouton "Select Cases" :



La fenêtre de dialogue permettant de définir le filtre se présente ainsi :



On retrouve alors le même résultat que précédemment.

Test t pour des Echantillons Appariés (policiers-par-groupe.sta dans Policiers-correction-2006.st)								
Différences significatives marquées à $p < ,05000$								
Filtres de Sélection - 'Inclure' : v1='Etudiant'								
Variable	Moyenne	Ec-Type	N	Différ.	Ec-Type Différ.	t	dl	p
Soi	87,15	19,02						
Autrui	118,95	18,34	62	-31,81	26,75	-9,36	61	0,000000

Dans l'en-tête du tableau de résultats figure la définition du filtre qui a été utilisé. C'est là une nouveauté de la version 7 de Statistica, les versions précédentes omettaient complètement d'indiquer ces informations.

On peut procéder de même pour comparer les variables "Soi" et "Autrui" chez les policiers.

Affirmation N° 3 : D'autre part, s'agissant des étudiants, les estimations de soi (réalisées par les étudiants) et les estimations d'autrui (réalisées par les policiers) concordent.

Il faudrait ici comparer les variables "Soi" en se limitant au groupe "Etudiant" et "Autrui" en se limitant au groupe "Policier". Cela ne relève plus d'un filtre, puisque cet outil sélectionne ou élimine des lignes entières. On ne peut donc pas réaliser l'équivalent des traitements réalisés dans le paragraphe précédent. Notez en revanche qu'un autre outil (l'analyse de variance sur un plan $S < G > * A$) permettrait d'obtenir un résultat à partir de données structurées sous cette forme.

3. Normalité des distributions parentes - Droite de Henry

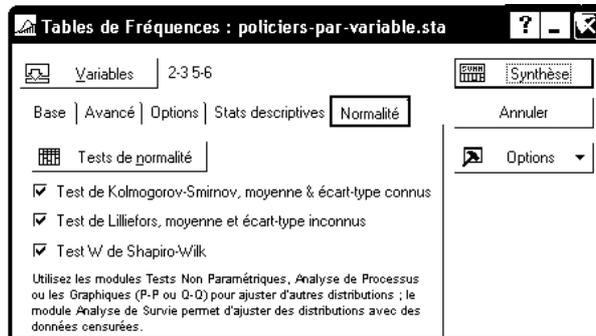
Les tests qui figurent dans ce paragraphe et le suivant auraient dû être réalisés avant les tests de Student, pour vérifier que les conditions d'application de ces derniers étaient respectées.

3.1. Tests de normalité sur des données structurées par variable

Est-il légitime de supposer que la VD est distribuée selon une loi normale dans chacune des 4 conditions obtenues en croisant les 2 facteurs ?

Le traitement le plus simple sera obtenu à partir de la feuille policiers-par-variable.

Plusieurs menus peuvent produire le résultat. Par exemple, on peut utiliser le menu Statistiques Élémentaires - Statistiques Descriptives puis l'onglet Normalité. Mais, il est plus simple d'utiliser le menu Statistiques Élémentaires - Tables de Fréquences et l'onglet Normalité.



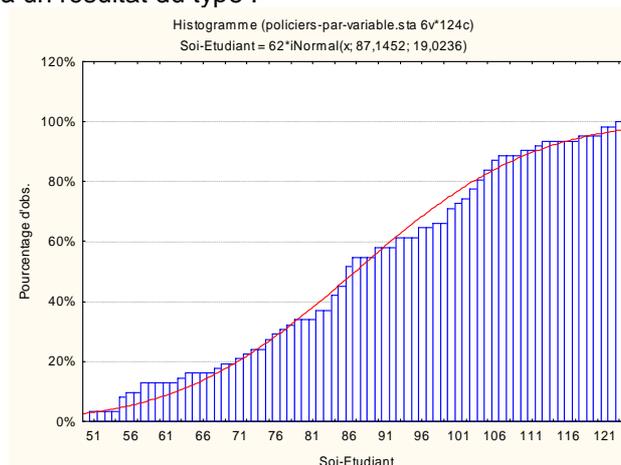
Sélectionnez les 4 variables numériques et les trois tests vus en cours, puis cliquez sur le bouton "Tests de normalité". Vous devriez obtenir les résultats suivants :

Variable	Tests de Normalité (policiers-par-variable.sta)					
	N	D max	K-S p	Lillief. p	W	p
Soi-Etudiant	62	0,089104	p > .20	p > .20	0,970911	0,147712
Autrui-Etudiant	62	0,066099	p > .20	p > .20	0,992622	0,972342
Soi-policier	62	0,083396	p > .20	p > .20	0,983030	0,546797
Autrui-policier	62	0,068903	p > .20	p > .20	0,991816	0,954717

Interprétez les résultats obtenus.

3.2. Droite de Henry relative aux 4 échantillons

Une illustration graphique de la normalité des données peut être obtenue en affichant des graphiques de type "droite de Henry". Le but de ce graphique est de représenter sur un même dessin la fonction de répartition observée et celle de la loi normale de mêmes paramètres. Mais, une représentation classique de ces deux fonctions conduit à un résultat du type :

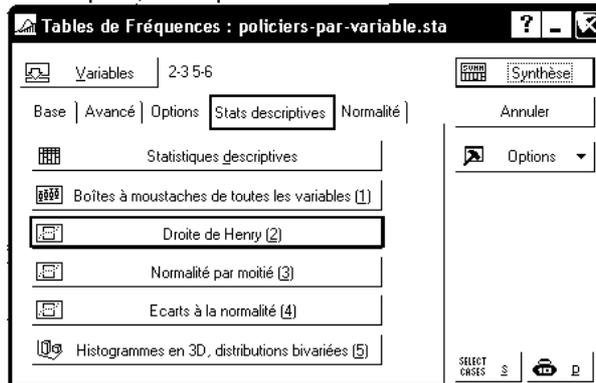


Pour mieux comparer la fonction de répartition de la distribution observée et celle de la variable normale de mêmes moyenne et variance, on construit le graphique selon les règles suivantes :

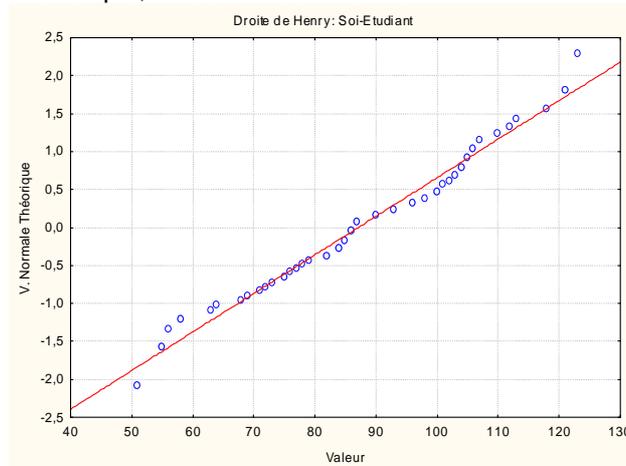
- l'axe horizontal est muni d'une graduation régulière, selon les valeurs prises par la variable.
- l'axe vertical correspond à des fréquences cumulées, comprises entre 0 et 1 ; mais sa graduation est faite de manière que la fonction de répartition de la loi normale soit représentée par une droite.
- de plus, Statistica gradue l'axe vertical en indiquant les valeurs centrées réduites correspondantes.

Pour réaliser ce graphique avec Statistica :

- Réaffichez le dialogue de l'analyse "Statistiques - Statistiques Élémentaires - Tables de fréquences", et sélectionnez l'onglet "Stats descriptives".
- Sélectionnez les variables numériques, et cliquez sur le bouton "Droite de Henry".



Pour la première variable, par exemple, le résultat est le suivant :



3.3. Tests de normalité sur des données structurées par groupe

Dans les anciennes versions de Statistica, il n'existe pas de menu donnant seulement les résultats relatifs aux tests de normalité pour des données structurées par groupe. En revanche, on peut obtenir ces résultats en commentaire des graphiques relatifs de type "droite de Henry".

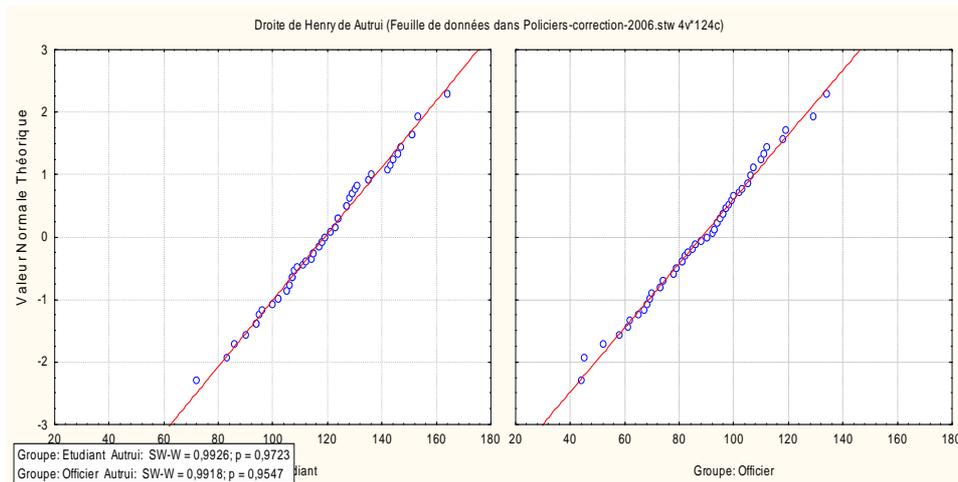
- Rendez active la feuille Policiers-par-groupe.sta.
- Choisissez le menu Graphiques - Graphiques catégorisés - Tracés de normalité.
- Indiquez "Soi" et "Autrui" comme variables et "Groupe" comme "Catégorie X".
- Sous l'onglet "Avancé", cochez "Test de Shapiro-Wilk".

En plus des graphiques de normalité relatifs aux deux variables, on obtient en commentaire :

Groupe: Etudiant Soi: SW-W = 0,9709; p = 0,1477
 Groupe: Officier Soi: SW-W = 0,983; p = 0,5468

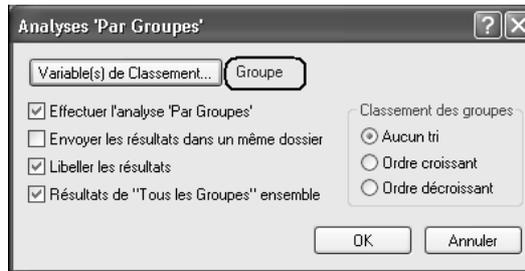
et

Groupe: Etudiant Autrui: SW-W = 0,9926; p = 0,9723
 Groupe: Officier Autrui: SW-W = 0,9918; p = 0,9547



Avec les versions récentes (versions 9 et supérieures), il est possible d'utiliser le bouton "Par Groupes" présent dans de nombreuses fenêtres de dialogue. Par exemple :

- Rendez active la feuille Policiers-par-groupe.sta.
- Choisissez le menu Statistiques - Statistiques élémentaires - Tables de fréquences.
- Indiquez "Soi" et "Autrui" comme variables et cliquez sur le bouton "Par Groupes...". Indiquez "Groupe" comme variable de classement :



- Affichez l'onglet "Normalité" et cliquez sur le bouton "Tests de normalité".

Vous obtenez les résultats suivants :

Groupe=Etudiant Tests de Normalité						
Variable	N	D max	K-S p	Lillief. p	W	p
Soi	62	0,089104	p > .20	p > .20	0,970911	0,147712
Autrui	62	0,066099	p > .20	p > .20	0,992622	0,972342

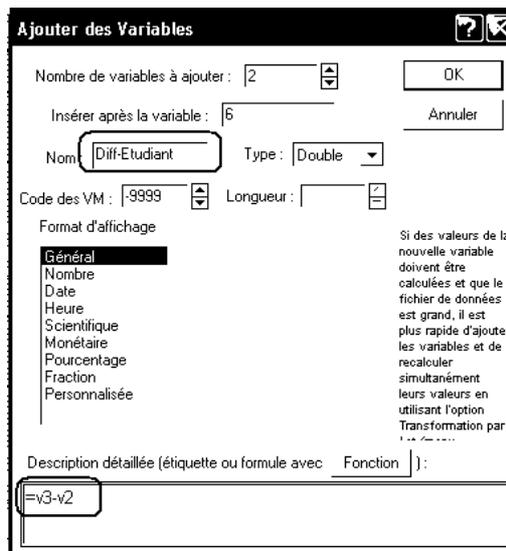
Groupe=Officier Tests de Normalité						
Variable	N	D max	K-S p	Lillief. p	W	p
Soi	62	0,083396	p > .20	p > .20	0,983030	0,546797
Autrui	62	0,068903	p > .20	p > .20	0,991816	0,954717

3.4. Normalité des protocoles dérivés des différences individuelles

L'étude utilise également le test de Student sur deux groupes appariés. La condition d'application de ce test est la normalité de la variable "différences individuelles" dans chacun des deux groupes.

Dans la feuille de calcul policiers-par-variable, créez deux nouvelles variables nommées Diff-étudiant et Diff-policier. Introduisez des formules pour calculer les valeurs prises par ces deux variables.

Pour cela, utilisez le menu Données - Variables - Ajouter et complétez le dialogue de la façon suivante :



Procédez de même pour la seconde variable.

Réalisez ensuite les tests de normalité sur ces deux variables. Vous devriez obtenir :

Variable	Tests de Normalité (policiers-par-variable.sta)					
	N	D max	K-S p	Lillief. p	W	p
Diff-Etudiant: =v3-v2	62	0,073745	p > .20	p > .20	0,984578	0,627045
Diff-policier: =v6-v7	62	0,081932	p > .20	p > .20	0,978900	0,362088

3.5. Remarque importante concernant les variables calculées

Comment les variables calculées se comportent-elles lorsque l'on demande un recalcul après avoir modifié la feuille ?

On sait que, avec une configuration standard de Statistica, le recalcul n'est pas automatique : si on modifie une ou plusieurs des valeurs observées, il faut passer par le menu Données - Recalculer les formules... pour réactualiser les valeurs calculées.

Il faut également noter que les formules de calcul ne sont pas remises à jour lorsqu'on modifie la structure de la feuille, par exemple en insérant une variable (colonne). Par exemple, si on insère une variable entre les colonnes 2 et 3, un recalcul de la variable Diff-Etudiant, toujours calculée à partir de la formule =v3-v2, prendra en compte la "nouvelle" variable v3 au lieu de la variable Autrui-Etudiant, conduisant ainsi à des résultats incorrects.

On peut également écrire les formules de calcul en utilisant les noms des variables au lieu de leurs noms génériques v1, v2, etc. Ainsi, pour la variable calculée Diff-Etudiant, la formule de calcul pourrait s'écrire :

= 'Autrui-Etudiant' - 'Soi-Etudiant'

Notez bien l'usage de cotes autour des noms de variables, rendu nécessaire par la présence de tirets dans les noms de variables (ambiguïté avec le signe moins).

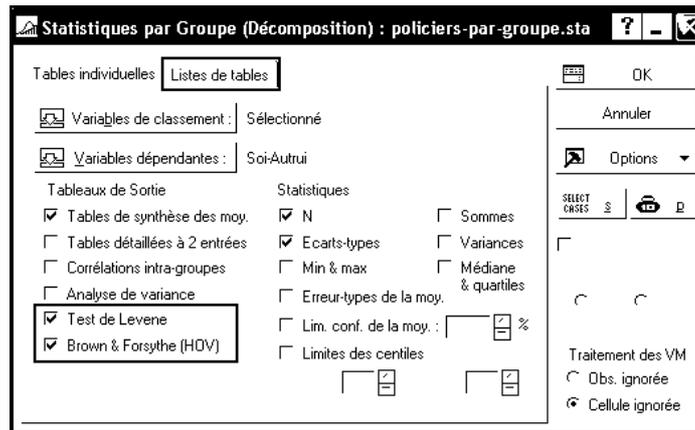
Cette solution permet de s'affranchir du problème précédent. Il vaut donc mieux utiliser cette syntaxe, même si la saisie des formules est plus fastidieuse. Mais on retrouve un problème analogue si l'une des variables intervenant dans les formules est renommée.

4. Tests d'homogénéité des variances

4.1. Tests d'homogénéité sur les données structurées "par groupe"

Les tests de Student que nous avons faits supposent notamment l'égalité des variances pour chacune des variables "soi-même" et "autrui" dans les deux groupes ? Est-il légitime de faire cette supposition ?

Ici, c'est la feuille de données policiers-par-groupe qui est la plus commode. Activez cette feuille, puis utilisez le menu Statistiques - Statistiques Élémentaires - Décompositions et ANOVA à un facteur.



Activez l'onglet "listes de tables". Sélectionnez "groupe" comme variable de classement, et les deux variables "soi" et "autrui" comme variables dépendantes. Cochez enfin les deux tests : Levene et Brown & Forsythe. Vous devriez obtenir les résultats suivants :

Test de Levene d'Homogénéité des Variances (policiers-par-groupe.sta)								
Effets significatifs marqués à $p < ,05000$								
Variable	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
Soi	8,15847	1	8,15847	13729,11	122	112,5337	0,072498	0,788189
Autrui	26,76812	1	26,76812	15549,02	122	127,4510	0,210027	0,647562

Test d'Homogénéité des Variances de Brown-Forsythe (policiers-par-groupe.sta)								
Effets significatifs marqués à $p < ,05000$								
Variable	SC Effet	dl Effet	MC Effet	SC Erreur	dl Erreur	MC Erreur	F	p
Soi	6,78226	1	6,78226	13967,18	122	114,4851	0,059241	0,808108
Autrui	23,51613	1	23,51613	15877,58	122	130,1441	0,180693	0,671527

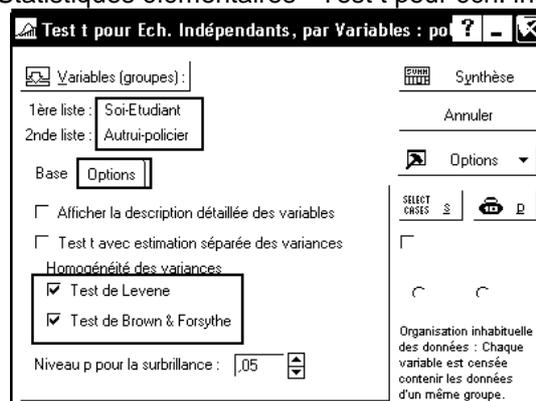
Interprétez ces résultats.

4.2. Test d'homogénéité sur les données structurées "par variable"

L'étude s'appuie également sur des tests de Student "croisés", par exemple : "soi-même" chez les étudiants v/s "autrui" chez les policiers. Il faudrait donc s'assurer de la validité de l'hypothèse d'homogénéité des variances correspondantes.

Le menu précédent ne permet plus d'obtenir directement le résultat. En revanche, on pourra l'obtenir à partir de la feuille de données policiers-par-variable, et du menu "Test t pour éch. indépendants, par variable".

Utilisez le menu Statistiques - Statistiques élémentaires - Test t pour éch. indépendants, par variable.



Indiquez les deux variables étudiées (par exemple Soi-étudiant et Autrui-policier).
 Activez l'onglet Options et sélectionnez les des tests relatifs aux variances.

Outre les résultats du test de Student, Statistica nous affiche les résultats relatifs aux tests sur les variances :

Test t pour des Echantillons Indépendants (policiers-par-variable.sta)																	
Note : Variables traitées comme des échantillons indépendants																	
	Moyenne Groupe 1	Moyenne Groupe 2	valeur t	dl	p	N Actifs Groupe 1	N Actifs Groupe 2	Ec-Type Groupe 1	Ec-Type Groupe 2	Ratio F Variances	p Variances	Levene F(1,dl)	dl Levene	p Levene	Brn-Fors F(1,dl)	dl Brn-Fors	p Brn-Fors
Groupe1 vs. Groupe2																	
Soi-Etudiant vs. Autrui-policier	87,14516	88,14516	1,293335	122	0,769764	62	62	19,02358	18,93808	1,009050	0,972048	0,024375	122	0,876192	0,025696	122	0,872910

5. Exercice

Sur des échantillons expérimentaux prélevés au hasard, on a observé une variable dépendante nommée Average_Res_Weigth dans deux conditions expérimentales différentes nommées CP+ et CP-. Les valeurs observées se trouvent dans le classeur SequencesCP.stw.

- 1) Etudiez au niveau descriptif la variable Average_Res_Weight dans les deux conditions.
- 2) On veut étudier si la condition expérimentale a un effet sur la variable dépendante.
 - a) Commencez par faire un test de Student sur les deux groupes. Quel niveau de significativité obtient-on ? Quelle serait la conclusion au seuil de 5% ?
 - b) Faites ensuite un test de Student, en spécifiant "Estimation séparée des variances" dans l'onglet "Options" du dialogue Test de Student. Quel est alors le niveau de significativité obtenu ?
 - c) Faites enfin un test de Wilcoxon Mann Whitney (test non paramétrique s'appliquant à ce type de situation). Quel niveau de significativité obtient-on ? Ce niveau de significativité est-il proche de ceux obtenus en a) et b) ?
- 3) Les résultats trouvés précédemment semblent manquer de cohérence. Afin de mieux comprendre la situation, étudier la normalité des distributions parentes des échantillons fournis et l'homogénéité des variances. En conclusion, que pensez-vous de l'effet de la condition expérimentale sur la variable dépendante étudiée ?

6. Présenter les résultats obtenus à l'aide de Statistica

6.1. Présenter les résultats dans un rapport

Les rapports sont des documents "texte" contenant les résultats des traitements. Pour un certain nombre d'usages, ils sont préférables aux autres objets de Statistica.

- En vue d'une impression : lorsqu'il imprime un classeur, Statistica imprime chaque objet sur une page séparée. Au contraire, le contenu du rapport pourra être imprimé séquentiellement, et en indiquant des en-têtes, pieds de page, numéros de page, etc.
- Pour insérer des commentaires, ou des titres, entres les différents traitements. En effet, un rapport est fondamentalement un objet de type "texte" dans lequel l'utilisateur peut insérer du texte libre et le mettre en forme.
- En vue d'une importation des objets Statistica dans Word, à l'aide des menus Copier et Coller. En effet, lorsqu'un objet est copié à partir d'un rapport, sa taille est mieux ajustée.
- En vue d'une exploitation des résultats de traitement sous Word. En effet, un rapport peut être enregistré au format *.rtf, puis ouvert à l'aide de Word.

Poursuivre l'exercice précédent en créant une nouvelle fenêtre de rapport et en y collant les différents résultats obtenus précédemment. Insérer des titres et des phrases indiquant les résultats des différents tests.

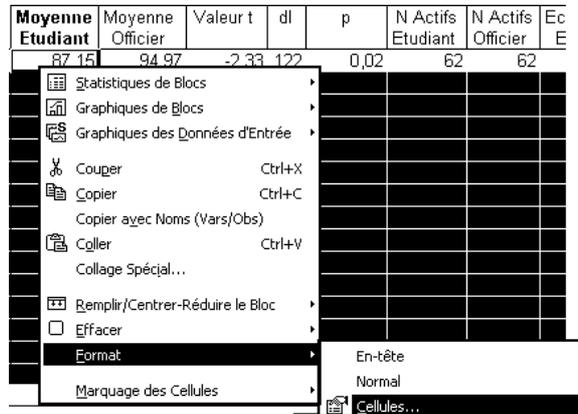
6.2. Utiliser les résultats dans un document Word

Il est généralement plus pratique de coller les résultats obtenus directement dans un document Word. On veut, par exemple, coller le résultat du premier test de comparaison dans un document Word.

Sans quitter Statistica, ouvrez un nouveau document Word.

Revenez à Statistica, affichez le tableau de résultats correspondant à ce premier test.

On peut, en cliquant avec le bouton droit de la souris dans les cellules concernées et en utilisant le menu Format - Cellules - Nombre, modifier le nombre de décimales affichées dans les résultats.



Cliquez dans le volet de droite du classeur, puis utilisez les menus Edition - Tout sélectionner et Edition - Copier avec noms.

Réaffichez le document Word. Pour coller le résultat dans Word, plusieurs solutions sont possibles :

Si on utilise le menu Edition - Coller, le tableau de résultats est collé sous forme de texte, avec des tabulations pour passer d'une colonne à l'autre.

Variable

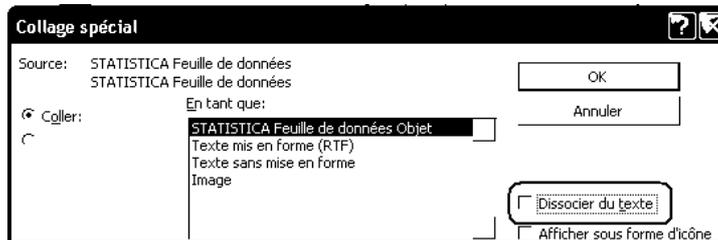
Tests t ; Classmt : Groupe (policiers-par-groupe.sta)

Groupe1: Etudiant

Groupe2: Officier

	Moyenne Etudiant	Moyenne Officier	Valeur t	dl	p	N Actifs Etudiant	N Actifs Officier	Ecart-Type Etudiant	Ecart-Type Officier
Soi	87,15	94,97	-2,33	122	0,02	62	62	19,02	18,29

Si on utilise le menu Edition - Collage spécial - Statistica Feuille de données Objet, en désactivant l'option "Dissocier du texte" :



on obtient un résultat du type suivant :

Tests t ; Classmt : Groupe (policiers-par-groupe.sta)									
Groupe1: Etudiant									
Groupe2: Officier									
Variable	Moyenne Etudiant	Moyenne Officier	Valeur t	dl	p	N Actifs Etudiant	N Actifs Officier	Ecart-Type Etudiant	Ecart-Type Officier
Soi	87,15	94,97	-2,33	122	0,02	62	62	19,02	18,29

6.3. Compresser un fichier ou un dossier

Les classeurs Statistica deviennent vite très volumineux dès qu'ils contiennent des rapports ou des graphiques. Leur taille peut alors rapidement dépasser les capacités de stockage d'une disquette (ou la taille limite autorisée pour les envois de documents joints par mail). Cependant, Windows XP comporte en standard un programme de compression de fichiers, très efficace sur les fichiers de ce type.

Les manipulations suivantes concernent évidemment le système d'exploitation XP et non le logiciel Statistica, qu'il sera d'ailleurs préférable de quitter avant de les entreprendre.

6.3.1 Pour compresser un dossier...

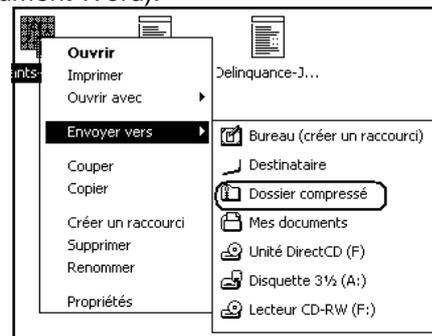
On veut, par exemple, rassembler dans un même dossier compressé les différents fichiers correspondant au traitement du cas "Policiers" (feuille de données, classeur, document Word).

Créez d'abord un dossier Policiers dans lequel vous ferez glisser les différents fichiers.

Cliquez sur l'icône du dossier à l'aide du bouton droit de la souris et utilisez l'item "Envoyer vers ... - Dossier compressé".

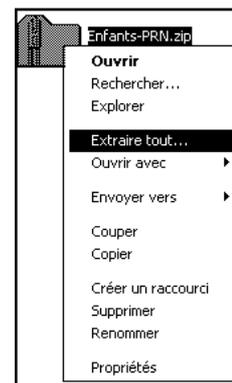
Windows XP crée alors sur le disque un fichier dont le nom est Policiers.zip.

Affichez les propriétés des deux dossiers et comparez leurs tailles.



6.3.2 Pour décompresser un dossier

Les fichiers contenus dans un dossier compressé sont utilisables de façon "transparente" : un double clic sur l'icône de l'objet Policiers.zip affichera son contenu, comme s'il s'agissait d'un dossier ordinaire. Cependant, pour différentes raisons (vitesse d'exécution par exemple), on peut souhaiter recréer une version décompressée du dossier Policiers.zip. Pour cela, cliquez avec le bouton droit de la souris et utilisez le menu Extraire tout.... Un assistant vous guide alors pour le reste de la démarche.



7. Travail à rendre par mail

N.B. Vous pouvez vérifier certains des résultats que vous obtenez ou les interprétations que vous en faites en consultant le document accessible à l'adresse:

<http://geai.univ-brest.fr/~carpenti/2012-2013/Correction-Exam2013-01L.pdf>

Réf. Navarro J., L'essentiel de la statistique en Psychologie, chapitre 4, Editions Ellipses 2012, exercice faisant référence à Charron C., Festoc A., Hiron A. & Petibon A.-C., "Child pedestrians' deliberately risk-taking behavior: experimental studies on simulator and developmental outcomes", Recherche Transports Sécurité, 101, 2008, pp. 239-251.

Des chercheurs ont mené une expérience dont l'objectif était de montrer comment l'enjeu, indépendamment des contraintes objectives de temps, produit un accroissement de la pression temporelle même chez des enfants. Cette étude a été menée auprès d'enfants de 10 ans. Il était expliqué aux enfants que, dans le cadre d'un jeu concours, deux objectifs étaient à atteindre : un cinéma puis une boîte aux lettres. Pour atteindre ces objectifs, les enfants devaient réaliser des traversées de rue dans un environnement simulé avec du trafic routier. Les enfants ont été répartis en trois groupes d'effectif 10 :

- Un groupe ayant la consigne de remplir les deux objectifs pour gagner le jeu (groupe A1) ;
- Un groupe ayant la consigne de remplir les deux objectifs et de ne pas se faire renverser pour gagner le jeu (groupe A2) ;
- Un groupe ayant la consigne de remplir les deux objectifs, de ne pas se faire renverser et de mettre le moins de temps possible pour gagner le jeu (groupe A3).

La variable dépendante recueillie est le temps de parcours en secondes. On obtient les résultats suivants :

A1	A2	A3
100	107	90
134	123	110
151	115	91
117	117	125
135	98	108
161	112	102
111	102	120
180	149	95
152	157	122
111	130	105

- 1) Créez un nouveau classeur Statistica et saisissez ces données dans une feuille de ce classeur, en les structurant de manière à pouvoir réaliser les traitements statistiques envisagés ci-dessous.
- 2) Réalisez les tests de comparaison de moyennes suivants (seuil retenu : 5%) :
 - Comparaison deux à deux des temps de parcours observés dans les conditions A1, A2 et A3.
 - Comparaison de la condition A3 à l'ensemble formé par les deux autres conditions.
- 3) Testez l'hypothèse de normalité des distributions parentes dans chacun des trois groupes ainsi que dans le groupe formé par la réunion des conditions A1 et A2.
- 4) Testez l'hypothèse d'homogénéité des variances pour chacune des situations correspondant aux quatre tests réalisés.
- 5) Réalisez un document Word dans lequel vous commenterez les résultats de chacun des tests effectués (en y collant au besoin les résultats obtenus sous Statistica).

Travail à rendre par mail à votre enseignant (Francois.Carpentier@univ-brest.fr) :

- Le classeur Statistica contenant les données et les tests effectués.
- Le fichier Word commentant les résultats (cf. question 5).

Faites attention à la taille des fichiers que vous envoyez. Tous les serveurs de messagerie imposent une taille limite pour les messages, et, dans le cas de la passerelle Web-mail de l'ENT de l'UBO, cette taille est assez basse : 2 Mo. En cas de problème pour envoyer les fichiers par mail, vous pouvez aussi déposer votre travail sur le serveur du département, en procédant de la manière suivante :

- Affichez dans un navigateur Web la page :
<http://geai.univ-brest.fr/%7Ecarpentier/> ou <http://geai.univ-brest.fr/~carpentier/>
 - Cliquez sur le lien "Pour déposer vos fichiers d'examen : cliquez ici"
 - Cliquez sur le bouton "Créer un compte de dépôt". Un login et un mot de passe vous seront demandés. Choisissez par exemple votre nom ou e+n° étudiant comme login.
- N.B. Cette fonctionnalité n'est active que si le poste de travail fait partie du réseau de l'UBO.

Une fois le compte de dépôt créé, vous pouvez l'utiliser, à partir de n'importe quel appareil connecté à l'internet, en utilisant le bouton "Déposer un fichier", pour déposer votre travail.

N.B. La taille limite des fichiers que l'on peut déposer est limitée à 15 Mo, ce qui devrait être suffisant.

Si vous utilisez cette boîte de dépôt, envoyez un mail (sans pièces jointes) à votre enseignant, en lui indiquant le login utilisé, pour le prévenir du dépôt.