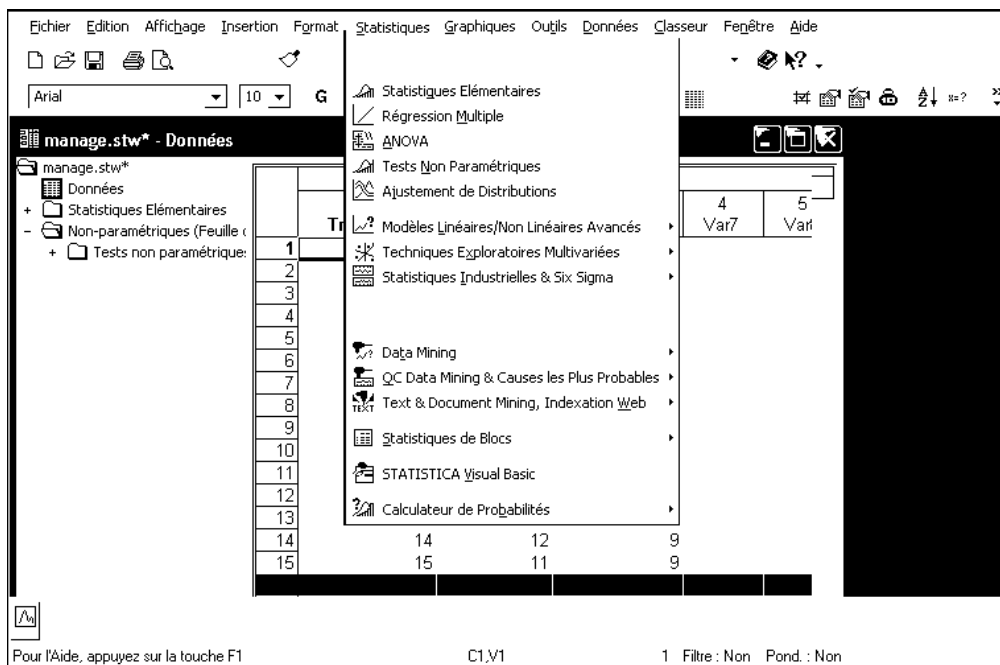


Master de Psychologie sociale des représentations
PSR S73C : Informatique - TD N°1
Configuration du courrier électronique dans les salles de TD
Utilisation d'un logiciel de traitement statistique : Statistica

1. Statistica : l'interface utilisateur

1.1 L'écran de travail

Statistica 6.1 est un logiciel dédié aux traitements statistiques. C'est également la "brique" de base des logiciels proposés par Statsoft, et ses possibilités d'interaction avec d'autres logiciels (tableurs, systèmes de gestion de bases de données, traitements de textes, ...) sont nombreuses. En revanche, l'interface utilisateur pourra sembler un peu déconcertante au premier abord.



1.2 Les objets manipulés par Statistica

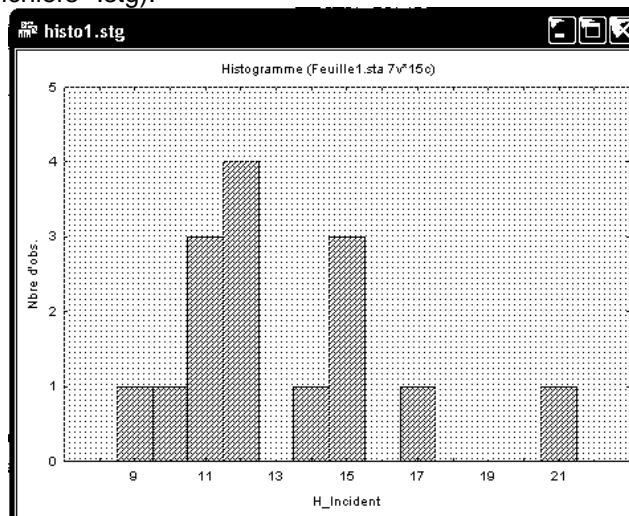
La **feuille de données** est organisée en variables et observations. Les colonnes sont les variables. Chaque ligne représente un individu statistique, appelé observation.

	1 Trimestre	2 H_Incident	3 D_Incident	4 Var7
1	1	11	11	8
2	2	11	11	13
3	3	14	14	12
4	4	21	21	17
5	5	12	12	14
6	6	10	10	9
7	7	15	15	10
8	8	15	15	12
9	9	17	17	13
10	10	9	9	10
11	11	12	12	8
12	12	12	12	13
13	13	15	15	12
14	14	12	12	9
15	15	11	11	9

Les feuilles de données peuvent être enregistrées comme fichiers autonomes (fichiers *.sta). Elles contiennent les données d'entrée sur lesquelles s'effectuent les traitements statistiques. Les résultats de ces traitements s'affichent dans un document de sortie. Plusieurs possibilités sont offertes.

Fenêtre de rapport : C'est la méthode traditionnelle pour gérer les résultats produits par le logiciel. Un rapport se comporte plus ou moins comme un document produit par un traitement de textes. On peut insérer des commentaires, modifier la mise en forme, spécifier la mise en page, la numérotation des pages, l'en-tête et le pied de page en vue de l'impression. Les rapports peuvent être enregistrés comme fichiers autonomes (fichiers *.str).

Les résultats de sortie peuvent également être dirigés vers des fenêtres individuelles. Les résultats numériques sont alors affichés dans des fenêtres de données. Les graphiques sont affichés dans des **fenêtres de graphiques** (fichiers *.stg).



Les classeurs : les données d'entrée et de sortie peuvent également être stockées comme onglets dans un classeur. Un classeur est un "conteneur" accueillant d'autres objets, organisés sous forme hiérarchique. Ils correspondent aux fichiers de type *.stw.

Variable	N Actifs	Moyenne
H_Incident	15	13,13333
D_Incident	15	11,26667

Traitements statistiques

Statistica est organisé en modules, accessibles à partir du menu Statistiques. Chaque module contient un groupe de procédures statistiques reliées entre elles. Par exemple, le module "Statistiques élémentaires" se présente comme suit :



2 Quelques traitements statistiques sur le cas "policiers"

Enoncé :

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 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 devrons 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.

2.2 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".

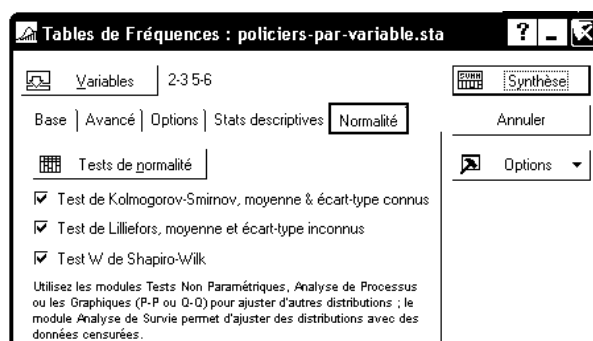
3. Normalité des distributions parentes - Droite de Henry

3.1 Tests de normalité

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 du fichier policiers-par-variable.sta.

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.



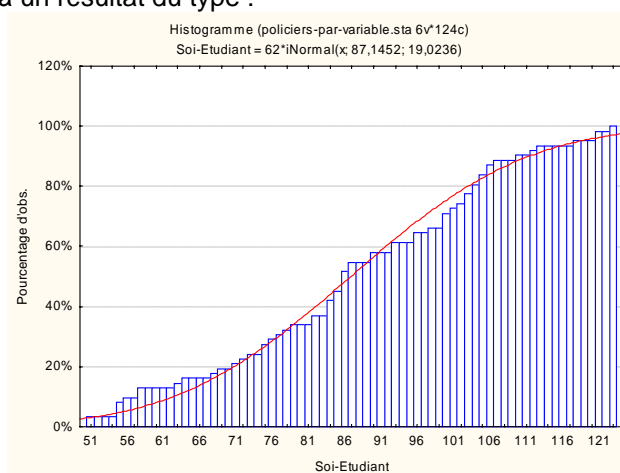
Sélectionnez les 4 variables numériques et les trois tests vus en cours, puis cliquez sur le bouton "Tests de normalité". Vous devriez obtenir les résultats suivants :

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 :

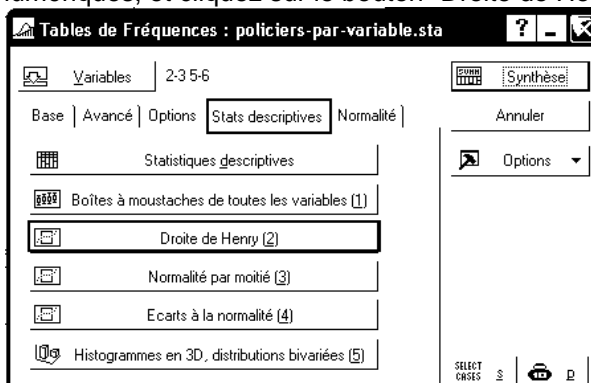


Pour mieux comparer la fonction de répartition de la distribution observée et celle de la variable normale de mêmes moyenne et variance, on construit le graphique selon les règles suivantes :

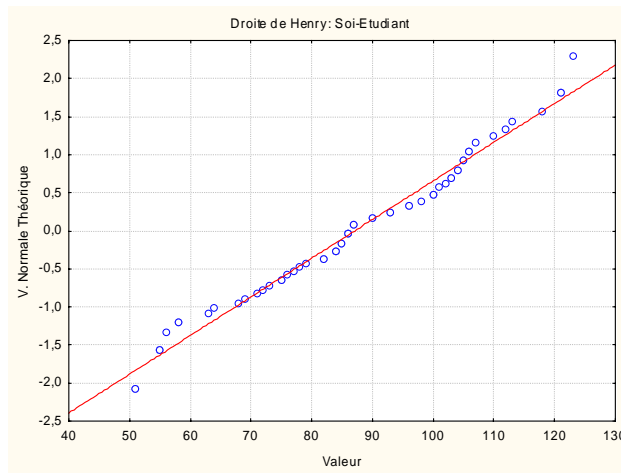
- l'axe horizontal est muni d'une graduation régulière, selon les valeurs prises par la variable.
- l'axe vertical correspond à des fréquences cumulées, comprises entre 0 et 1 ; mais sa graduation est faite de manière que la fonction de répartition de la loi normale soit représentée par une droite.
- 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 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.sta, créez deux nouvelles variables nommées Diff-étudiant et Diff-policier. Introduisez des formules pour calculer les valeurs prises par ces deux variables. Réalisez ensuite les tests de normalité sur ces deux variables. Vous devriez obtenir :

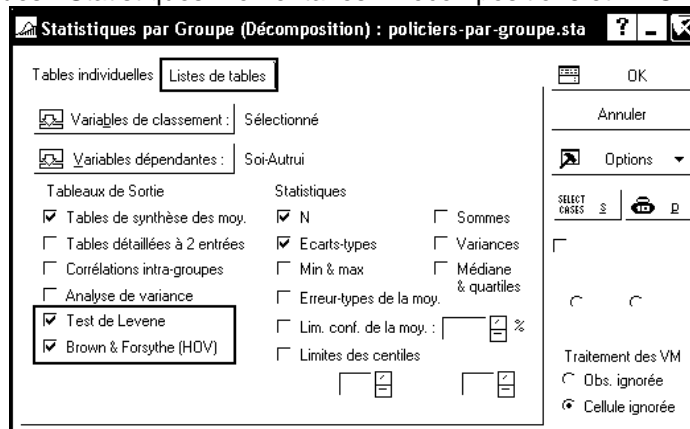
Variable	Tests de Normalité (policiers-par-variable.sta)					
	N	D max	K-S p	Lillief. p	W	p
Diff-Etudiant: =v3-v2	62	0,073745	p > .20	p > .20	0,984578	0,627045
Diff-policier: =v6-v7	62	0,081932	p > .20	p > .20	0,978900	0,362088

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.sta 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.

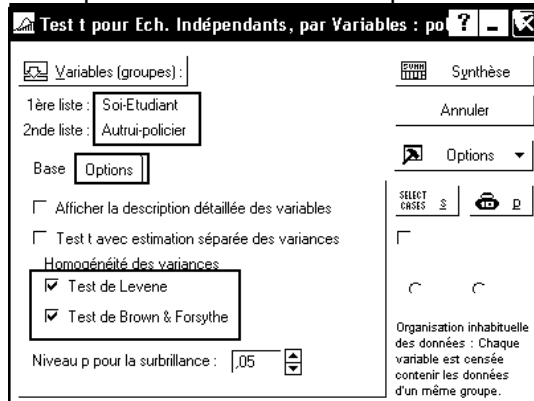
4.2 Test d'homogénéité sur les données structurées "par variable"

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 de normalité correspondante.

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

Affichez le fichier policiers-par-variable.sta.

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																	
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 F Variances	p Variances	Levene F(1,dl)	dl Levene	p Levene	Brn-Fors F(1,dl)	dl Brn-Fors	p Brn-Fors
Soi-Etudiant vs. Autrui-policier	87,14516	88,14516	293335	122	0,769764	62	62	19,02358	18,93808	1,009050	0,972048	0,024375	122	0,876192	0,025696	122	0,872910