

Licence de Psychologie - Semestre N° 5 - TD n° 3

Organiser son travail avec Statistica

Tests statistiques non paramétriques

9 Transférer dans Statistica des données provenant d'un autre logiciel

9.1 Copier - coller des données d'Excel dans Statistica

Il est tout à fait possible de copier des données dans une feuille de calcul Excel pour les coller dans Statistica. Attention cependant au comportement de Statistica sur les valeurs de type "texte", comme le montre le petit exemple qui suit.

Ouvrez le classeur Excel Apprentissage.xls.

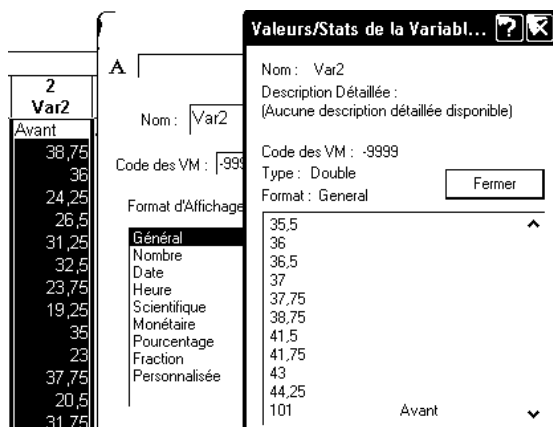
Chargez Statistica et affichez une nouvelle feuille de données.

Revenez à Excel ; sélectionnez la plage de cellules A1:C33 de la feuille "Données" (c'est-à-dire, l'ensemble des données, y compris les noms de variables figurant en première ligne) et copiez-la.

Retournez dans Statistica et collez-la dans la feuille de données.

Statistica a ajusté automatiquement le nombre de lignes de la feuille, ce qui est pratique. Cependant...

Affichez les caractéristiques de la variable Var2 (par exemple), en faisant un double-clic sur la tête de la colonne correspondante :



On s'aperçoit que la valeur de Var2 sur la première observation est "Avant" et que cette valeur est représentée de façon interne par le nombre 101. Evidemment, tous les calculs que nous pourrions faire sur cette variable conduiraient à des résultats inexacts...

Ouvrez une nouvelle feuille de données Statistica et refaites l'opération, en vous limitant cette fois aux seules données numériques, puis renommez les variables "Sujet", "Avant", "Après".

Calculez les statistiques descriptives (moyenne et écart type) relatives aux deux variables "avant" et "après". Vérifiez vos résultats en les comparant à ceux figurant dans la feuille "Corrigé" du classeur Excel.

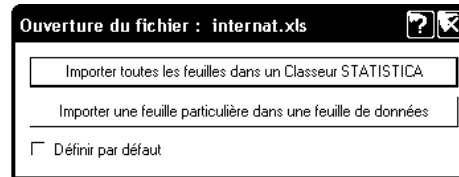
9.2 Importer un fichier Excel

Statistica peut également importer des données en provenance d'une feuille de calcul Excel, convenablement structurée.

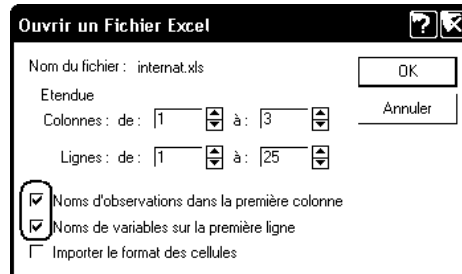
Ouvrez à l'aide d'Excel le fichier Internat.xls. Notez la façon dont les données sont organisées : données dans la feuille 2, noms des variables en ligne 1, noms des observations en colonne A, données jusqu'en ligne 25.

Revenez à Statistica et utilisez le menu Fichier-Ouvrir, sélectionnez "Fichiers Excel (*.xls)" dans la zone d'édition "Fichiers de type".

Sélectionnez alors le fichier Internat.xls, puis cliquez sur le bouton "Importer une feuille particulière dans une feuille de données" :



Dans le dialogue suivant, sélectionnez la feuille 2, puis indiquez les rôles de la première ligne et de la première colonne :



Enregistrez ensuite cette feuille de données sous le nom Internat.sta.

10 Gérer les sorties

10.1 Modifier le comportement de Statistica

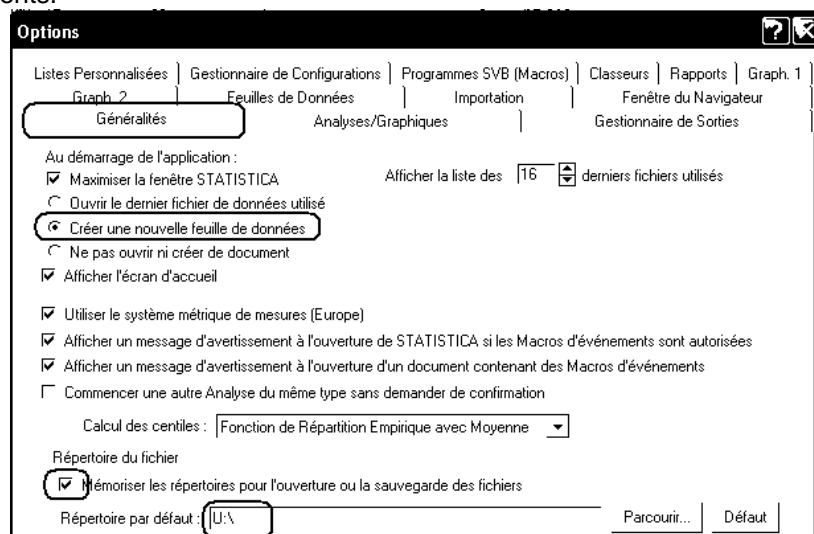
Le comportement de Statistica peut être modifié en intervenant dans la fenêtre de dialogue affichée par le menu Outils - Options.

Par exemple, nous souhaitons :

- que Statistica n'ouvre plus systématiquement la dernière feuille de données utilisée lors du chargement du logiciel ;
- que Statistica nous propose par défaut le volume U: pour enregistrer nos documents, au lieu du répertoire "Mes Documents".

Exécutez le menu Outils - Options. Sous l'onglet Généralités, activez le bouton radio "Créer une nouvelle feuille de données".

Désactivez la boîte à cocher "mémoriser les répertoires pour l'ouverture ou la sauvegarde des fichiers". Complétez la zone d'édition "Répertoire par défaut" en indiquant U:\, puis réactivez la boîte à cocher (N.B. Bien que l'option soit en apparence désactivée, Statistica proposera par défaut le répertoire U:\ pour l'enregistrement de nouveaux documents).



10.2 Gérer les sorties

Lorsqu'on utilise Statistica sans se préoccuper des options de sortie des résultats, on se retrouve vite à la tête d'une quantité de fenêtres (classeurs, feuilles de données de résultats, fenêtres de graphiques...). Pour réaliser

un travail que l'on souhaite conserver et reprendre au cours de plusieurs séances de travail, il paraît indispensable d'organiser correctement son espace de travail et ses sauvegardes.

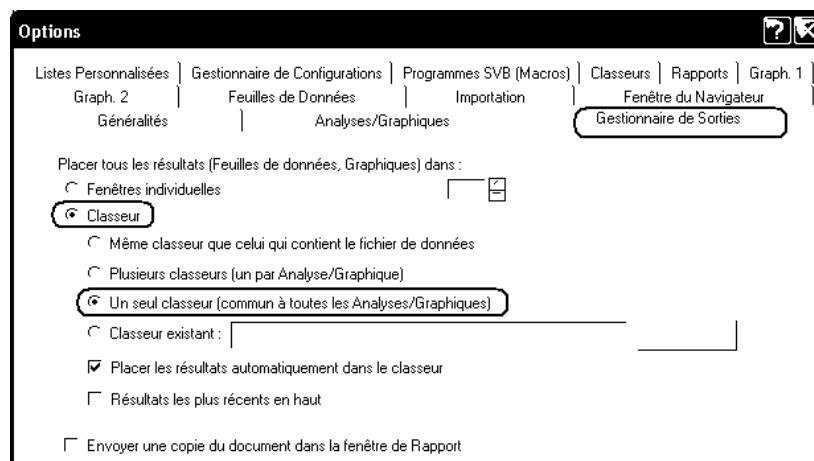
En fait, plusieurs méthodes de travail sont envisageables avec Statistica :

10.2.1 Première méthode : utiliser un fichier de données et un classeur de résultats

C'est la méthode que nous avons utilisée jusqu'à présent, pour la plupart des traitements que nous avons effectués :

- Les données se trouvaient dans une feuille de données séparée (fichier *.sta)
- Les résultats des traitements étaient produits dans un classeur (fichier *.stw) et Statistica produisait un seul classeur pour l'ensemble d'une session de travail.

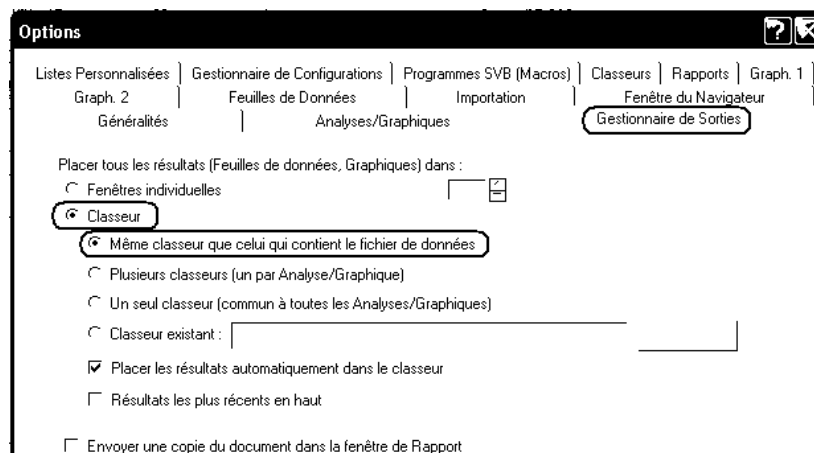
Ce comportement correspond aux options "par défaut" de Statistica. Mais ces options ne sont pas toujours adaptées au travail à réaliser. Ces options correspondent aux réglages suivants dans le menu Outils - Options - Onglet Gestionnaire de Sorties :



10.2.2 Deuxième méthode : enregistrer données et résultats dans un seul classeur

Cette méthode consiste à enregistrer les données, les résultats de traitements, et les commentaires éventuels comme objets d'un même classeur. Ainsi, un unique fichier du disque rassemble l'ensemble de notre travail sur un cas donné.

Ce comportement correspond aux réglages suivants dans le menu Outils - Options - Onglet Gestionnaire de Sorties :



Remarque : Le réglage ne sera actif que si la feuille de données se trouve effectivement dans un classeur. Or, ce ne sera pas le cas si la feuille de données a été ouverte à partir d'un fichier *.sta, ou importée à partir d'une feuille Excel. Dans ce cas, vous devez insérer la feuille de données dans le classeur comme il a été indiqué au paragraphe précédent.

10.2.3 Indiquer quelle est la feuille de données active

Lors des premières manipulations avec Statistica, nous n'avons pas eu besoin de nous préoccuper de la notion de "feuille de données active", les choix par défaut faits par Statistica nous convenant parfaitement. Cependant, cette notion permet de résoudre plusieurs problèmes :

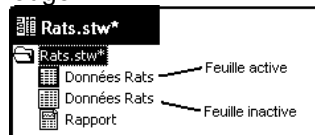
- Ouvrir plusieurs fichiers .sta et effectuer un travail sur l'un d'eux (pas nécessairement le dernier ouvert)
- Utiliser une feuille de résultats comme feuille de données pour des traitements ultérieurs.
- Lorsque l'on travaille avec une feuille de données insérée dans un classeur, il arrive couramment que Statistica ne retrouve pas la feuille à partir de laquelle les traitements doivent être effectués. Mais on peut éviter ce comportement en spécifiant la propriété "feuille de données active" pour l'objet du classeur qui contient nos données.

Pour spécifier comme feuille de données active une feuille d'un classeur :

- Cliquez avec le bouton droit de la souris sur l'icône de la feuille de données dans le volet gauche du classeur.
- Utilisez l'item Feuille de données active du menu local.

On peut également utiliser le menu Données - Feuille de données active.

Remarquez que le volet gauche d'un classeur indique si une feuille insérée dans le classeur est active ou non : l'icône d'une feuille active est encadrée en rouge :



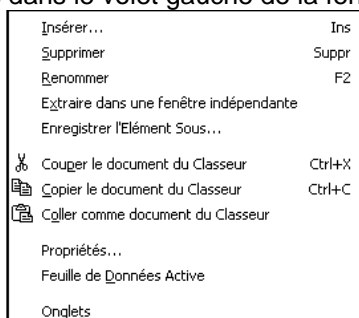
10.3 Enregistrer les données et l'ensemble des traitements réalisés dans un même classeur

Ouvrez le fichier de données Internat.sta créé dans le paragraphe 9. Réalisez un ou plusieurs traitements relatifs à ces données (par exemple, des statistiques descriptives et un graphique). Si vous avez gardé les options par défaut de Statistica, les résultats de tous ces traitements se trouvent dans un classeur.

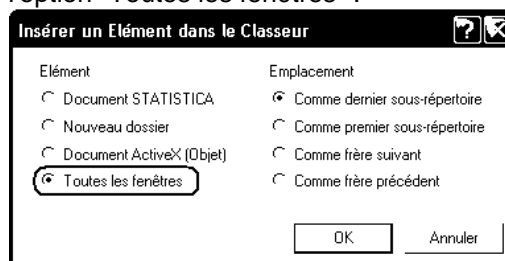
Pour enregistrer données, traitements et rapport dans un seul classeur :

Affichez la fenêtre du classeur contenant les résultats.

Cliquez avec le bouton droit de la souris dans le volet gauche de la fenêtre du classeur.



Sélectionnez l'item Insérer..., puis l'option "Toutes les fenêtres" :



N'oubliez pas, ensuite, de spécifier la feuille Internat.sta du classeur comme feuille active.

Après avoir refermé toutes les fenêtres autres que celle du classeur, poursuivez le traitement en effectuant une comparaison de moyennes sur groupes appariés. Rassemblez au besoin les fenêtres de résultats dans le classeur et enregistrez-le.

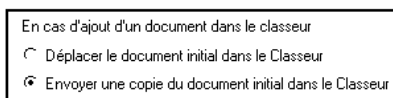
10.4 Manipuler les objets contenus dans un classeur

10.4.1 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".



10.4.2 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.

Mise en œuvre : Réalisez différents traitements complémentaires (statistiques descriptives, graphiques, comparaison de moyennes, etc) sur les données "Internat" puis rassemblez la feuille de données et ces résultats dans un classeur, en éliminant les traitements redondants. Enregistrez le document final dans un classeur.

10.5 Travail avec 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.

Ouvrez le document Wagner.stw. Ce classeur Statistica contient des données relatives à 107 sujets et différents traitements sur ces données, correspondant approximativement aux questions posées dans l'énoncé.

Mettez en forme le rapport en supprimant les traitements figurant en double et en ajoutant, sous forme de titres, les questions correspondant à chaque traitement.

Réalisez une extraction dans une fenêtre indépendante d'une copie du rapport (clic droit sur l'icône du rapport dans le volet gauche de la fenêtre du classeur). Visualisez l'aperçu avant impression du rapport.

Copiez certains éléments du rapport et collez-les dans un document Word. Essayez plusieurs méthodes. Par exemple, pour un tableau de résultat :

- Utilisez Edition - Collage Spécial - Texte mis en forme. Observez la forme du résultat.
- Utilisez Edition - Collage Spécial - Statistica Feuille de données objet et étudiez le comportement de l'objet ainsi collé vis à vis du texte.
- Utilisez Edition - Collage Spécial - Statistica Feuille de données objet et cochez l'option "Dissocier du texte".

Enregistrez le rapport au format .rtf et ouvrez-le sous Word.

Remarque 1 : Dans le menu Outils - Options, l'onglet Gestionnaire de sorties permet d'obtenir une copie des résultats des traitements dans un rapport. Mais, même si l'option "Placer tous les résultats dans le même classeur que celui qui contient les données" est active, le rapport n'est pas automatiquement inséré dans le classeur des données et traitements. Il faut donc d'utiliser la méthode du paragraphe précédent pour insérer le rapport dans le classeur à un moment quelconque de la session. C'est ce rapport qui continuera à être utilisé pour les traitements ultérieurs.

11 Réaliser un travail de synthèse

11.1 Le cas "Intervention brève".

Des chercheurs ont réalisé une expérience visant à mettre en évidence l'effet d'une séance d'intervention motivante brève sur le comportement relatif à la consommation d'alcool.

Soixante sujets, qui ont déclaré avoir bu occasionnellement à 2 reprises ou plus au cours du mois précédant l'expérience ont été affectés au hasard soit dans un groupe contrôle, sans traitement (31 sujets) soit dans un groupe expérimental dit "groupe d'intervention brève" ou "groupe IB" (29 sujets).

Le comportement des sujets est mesuré par la variable "nombre de verres bus par semaine".

Les sujets sont évalués avant l'expérience (condition de référence). Chacun des sujets du groupe d'intervention brève bénéficie d'un entretien personnalisé relatif aux problèmes liés à l'alcool. Six semaines après l'entretien, l'ensemble des sujets est de nouveau évalué.

Les observations sont regroupées dans le classeur Excel Intervention-breve.xls.

Réalisez un classeur et un rapport Statistica.

Vous placerez dans le classeur une feuille de données, avec les données convenablement organisées.

Vous effectuerez ensuite différents traitements sur ces données : moyennes, écarts types, etc, comparaison des différentes conditions à l'aide de graphiques adaptés, tests de comparaison de moyennes. Vous devrez au moins faire les comparaisons suivantes :

- groupe IB v/s groupe contrôle en situation de référence,
- groupe IB v/s groupe contrôle lors de l'étude de suivi,
- situation de référence v/s étude de suivi pour le groupe IB).

Vous veillerez à ce que les résultats des traitements soient également enregistrés dans un rapport.

Vous devrez ensuite compléter le rapport par des textes explicatifs, indiquant notamment le résultat de chacun des tests statistiques.

Vous enregistrerez d'une part le classeur, et d'autre part le rapport produit par Statistica.

11.2 Exercice

Réalisez de même une étude au niveau descriptif et inférentiel concernant les données contenues dans le fichier Enfants-PRN.xls. L'énoncé du cas se trouve dans la première feuille du classeur.

12 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. 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.

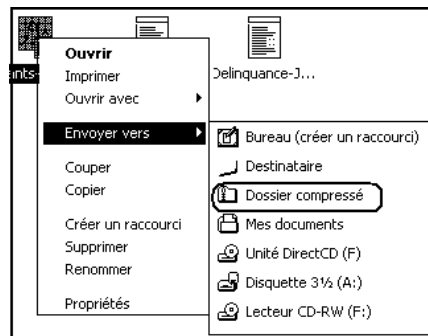
12.1 Pour compresser un fichier...

On veut, par exemple, compresser le fichier Enfants-PRN.stw.

Cliquez sur l'icône du fichier 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 Enfants-PRN.zip.

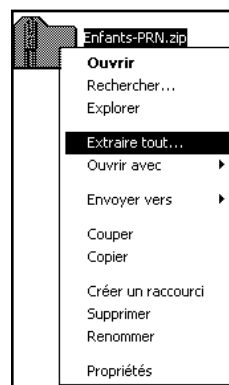
Affichez les propriétés des deux fichiers et comparez leurs tailles.



Pour compresser l'ensemble des fichiers contenus dans un dossier, on procédera de la même façon, en agissant sur l'icône du dossier.

12.2 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 Enfants-PRN.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 fichier Enfants-PRN.stw. 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.



Tests non paramétriques

13 Test de la médiane

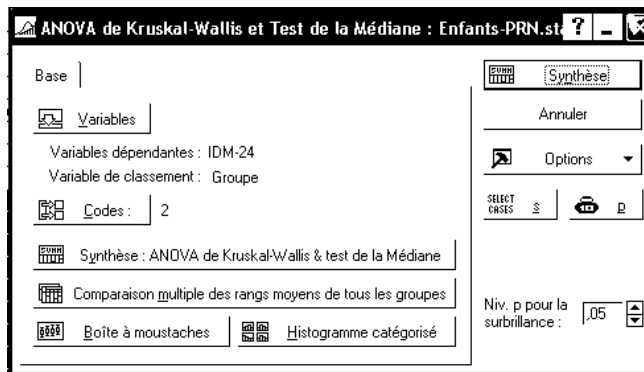
Ouvrez le classeur Statistica Enfants-PRN.stw.

On veut comparer l'IDM à 24 mois dans le groupe témoin et dans le groupe expérimental à l'aide d'un test de la médiane.

Rappel de la méthode : on construit un tableau de contingence en croisant les variables "Groupe" et "Position par rapport à la médiane" et on réalise un test du khi-deux sur le tableau de contingence obtenu.

En utilisant, par exemple, le menu Statistiques - Tests non paramétriques - Statistiques ordinales, vérifiez que la médiane des IDM à 24 mois est égale à 111,5.

Dans le cours, le test de la médiane a été présenté avec une variable "Groupe" à deux modalités. Cependant, la méthode peut s'étendre sans difficultés au cas où la variable "Groupe" comporte plus de deux modalités. C'est pourquoi Statistica range ce test dans le menu : Statistiques - Tests non paramétriques - Comparaison de plusieurs échantillons indépendants :



Spécifiez la variable dépendante et la variable de classement, puis cliquez sur le bouton "Synthèse : ANOVA de Kruskal-Wallis & test de la Médiane". On obtient le résultat suivant :

Test Médiane, Méd. Globale = 111,500; IDM-24 (Enfants-PRN.sta) Var. indépendante (classement) : Groupe Chi-Deux = 3,540645 dl = 1 p = ,0599			
Dépendant : IDM-24	Témoin	Expérimental	Total
<= Médiane : observ.	19,00000	9,00000	28,00000
théorique	15,50000	12,50000	
obs.-thé.	3,50000	-3,50000	
> Médiane : observée	12,00000	16,00000	28,00000
théorique	15,50000	12,50000	
obs.-thé.	-3,50000	3,50000	
Total : observé	31,00000	25,00000	56,00000

Remarque : Le test de la médiane ne met pas en évidence de différence entre les deux groupes. En revanche, un test unilatéral de comparaison de moyennes établit une différence au bénéfice du groupe expérimental. Mais le test de la médiane est moins puissant, et c'est nécessairement un test bilatéral.

14 Protocoles de rangs et test de Wilcoxon Mann Whitney

14.1 Le test de Wilcoxon Mann Whitney - Groupes indépendants

La comparaison précédente peut être reprise à l'aide d'un test de Wilcoxon Mann Whitney.

Reprenez le menu Statistiques - Tests non paramétriques. Sélectionnez l'item "Comparaison de deux échantillons indépendants". Si nécessaire, spécifiez de nouveau la variable dépendante et la variable de classement, puis cliquez sur le bouton "Test U de Mann-Whitney". Vous devriez obtenir comme résultat :

Test U de Mann-Whitney (Enfants-PRN.sta)									
Par var. Groupe									
Tests significatifs marqués à p <,05000									
variable	SommeRgs Témoïn	SommeRgs Expérimental	U	Z	niv. p	Z ajusté	niv. p	N Actif Témoïn	N Actif Expérimental
IDM-24	731,5000	864,5000	235,5000	-2,50522	0,012238	-2,51430	0,011927	31	25

La statistique calculée par Statistica est-elle la même statistique que celle indiquée en cours ?
 Statistica calcule les sommes des rangs W_1 et W_2 . On peut vérifier que la valeur Z indiquée (-2,505) correspond bien à la formule du cours :

$$Z = \frac{\bar{R}_1 - \bar{R}_2}{E} \quad \text{avec :} \quad E^2 = \frac{(n_1 + n_2 + 1)(n_1 + n_2)^2}{12n_1n_2}$$

En effet, on a ici : $\bar{R}_1 = \frac{731,5}{32} = 23,60$ et $\bar{R}_2 = \frac{864,5}{25} = 34,58$;

$$E^2 = \frac{(31 + 25 + 1)(31 + 25)^2}{12 \times 31 \times 25} = 19,22, \quad E = 4,38 \quad \text{et enfin :} \quad Z = \frac{23,60 - 34,58}{4,38} = -2,505$$

En revanche, Statistica calcule aussi une autre statistique : le U de Mann-Whitney.

14.2 Détermination du protocole des rangs

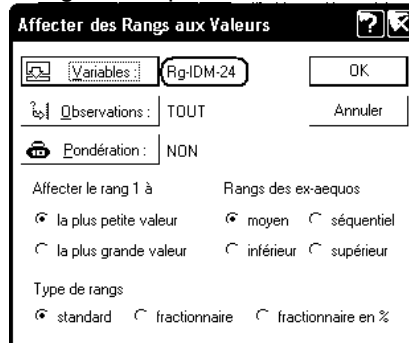
Pour la mise en œuvre du test de Mann Whitney, la détermination préalable du protocole des rangs n'est pas nécessaire. Cependant, il peut être intéressant de le déterminer pour contrôler, par exemple, que les ex-æquo ne sont pas trop nombreux...

Le menu Données - Affecter les rangs... permet de déterminer le protocole des rangs. Mais, le protocole obtenu remplace le protocole observé à partir duquel il a été déterminé. Si nous voulons conserver à la fois le protocole des rangs et le protocole observé, nous devons au préalable faire une copie de ce dernier.

Insérez une nouvelle variable après la variable IDM-24. Cette nouvelle variable sera nommée Rg-IDM-24.

Recopiez les données de la colonne IDM-24 dans la colonne Rg-IDM-24.

Utilisez le menu Données - Affecter les rangs... en spécifiant comme variable : Rg-IDM-24 :



Au besoin, modifiez les caractéristiques de la variable Rg-IDM-24 de façon que les données s'affichent avec au moins une décimale.

On observe la présence d'assez nombreux ex-æquo dans ce protocole.

14.3 Le test de Mann Whitney sur de petits échantillons

On considère les données contenues dans le classeur : Delinquance-Juvenile.stw

On compare les deux groupes "Maison des parents" et "Foyer".

Réalisez un test de Mann-Whitney : vous devriez obtenir le résultat suivant :

Test U de Mann-Whitney (Délinquance dans Delinquance-Juvenile.stw)										
Par var. Domicile										
Tests significatifs marqués à p <,05000										
variable	SommeRgs Maison des parents	SommeRgs Foyer	U	Z	niv. p	Z ajusté	niv. p	N Actif Maison des parents	N Actif Foyer	2*(1-p) p exact
Absentéisme	101	70	25	1,37	0,17	1,37	0,17	9	9	0,190251

Compte tenu de la faible taille des échantillons, ce n'est pas le niveau de la statistique Z qu'il faut ici prendre en compte, mais la valeur indiquée dans la dernière colonne : $2*(1-p) - p$ exact.

Pour vérifier cette valeur, on peut se servir des "tables statistiques en ligne" accessibles à l'adresse <http://geai.univ-brest.fr/~carpenti/statistiques/table1.php> :

Test des rangs de Wilcoxon (groupes indépendants)
Calcul de W critique :

Alpha :

N.B. : Prendre l'échantillon le plus petit comme 1er échantillon

Taille 1er éch. :

Taille 2nd éch. :

Nature du test :

Test unilatéral

Test bilatéral

W critique "à gauche" : 71
N.B : H1 retenue pour W strictement inférieur à W critique

W critique "à droite" : 100
N.B : H1 retenue pour W strictement supérieur à W critique

Calculer W | Annuler

Pour le niveau de significativité calculé par Statistica, les valeurs indiquées par les tables en ligne (calculées par le logiciel de Statistiques R) sont compatibles avec les sommes de rangs observées.

En revanche, on remarquera que Statistica ne fait pas de correction pour tenir compte des ex-æquo. D'autres logiciels (Statgraphics, Minitab) font cette correction, et affichent $W=56$, avec un niveau de significativité de 0,1844.

Exercice : Procéder de même pour effectuer les deux autres comparaisons de groupes pris deux à deux. La seule comparaison qui nous conduit à accepter l'hypothèse alternative est la troisième : les enfants placés en foyer sont moins souvent absents que les enfants placés en famille adoptive.

15 Test du khi-2 de Mac Nemar

On reprend l'exemple traité en cours : reconnaissance d'une série de portraits à deux semaines et à 1 an. Les observations sont résumées par le tableau suivant :

		1 an	
		Reconnu	Non reconnu
Deux semaines	Reconnu	81	46
	Non reconnu	8	49

Utilisez le menu Statistiques - Tests non paramétriques - Tables 2x2.

Indiquez les effectifs ci-dessus dans la fenêtre de dialogue et cliquez sur le bouton "Synthèse".

Statistica nous sert en vrac différents résultats : khi-deux "classique", phi-deux, khi-deux de Mac Nemar... A nous de savoir choisir le résultat qui nous intéresse (et qui a un sens par rapport à nos données) :

	Table 2 x 2 (Feuille de données59)		
	Colon. 1	Colon. 2	Totaux Bruts
Effectifs, ligne 1	81	46	127
%age du total	44,022%	25,000%	69,022%
Effectifs, ligne 2	8	49	57
%age du total	4,348%	26,630%	30,978%
Totaux colonne	89	95	184
%age du total	48,370%	51,630%	
Chi-deux (dl=1)	38,98	p= ,0000	
V-deux (dl=1)	38,77	p= ,0000	
Chi ² corrigé de Yates	37,02	p= ,0000	
Phi-deux	,21186		
p exact Fisher, unilatéral		p= ,0000	
bilatéral		p= ,0000	
Chi ² de McNemar (A/D)	7,39	p= ,0066	
Chi-deux (B/C)	25,35	p= ,0000	

16 Test du signe - Groupes appariés

On reprend le classeur Enfants-PRN.stw et on se propose de comparer l'IDM à 6 mois et l'IDM à 24 mois dans le groupe témoin.

On veut essayer de montrer que le nombre de différences négatives est significativement grand, ou, de manière symétrique, que le nombre de différences positives est suffisamment faible pour montrer une baisse de l'IDM entre 6 et 24 mois, dans la population dont est tiré l'échantillon.

On va donc utiliser un test du signe pour comparer les scores des enfants du groupe témoin à 6 mois et à 24 mois.

Utilisez le menu Statistiques - Tests non paramétriques - Comparaison de deux échantillons appariés.

Indiquez IDM-6 et IDM-24 comme variables et cliquez sur le bouton "Test des signes".

Vous devriez obtenir le résultat suivant :

Couple de variables	Test des Signes (Enfants-PRN.sta) Tests significatifs marqués à p <,05000			
	Nbe Non ex-aequo	%age v < V	Z	niv. p
IDM-6 & IDM-24	31	38,70968	1,077632	0,281198

Statistica nous indique que 38,71% des paires sont telles que IDM-6 est inférieur à IDM-24. Il calcule l'approximation par une loi normale donnée par :

$$Z = \frac{2D - 1 - N}{\sqrt{N}} \quad \text{où} \quad D = \text{Max}(D_+, D_-)$$

et indique que le niveau de significativité de cette statistique est de 28% pour un test bilatéral.

Conclusion : on n'a pas démontré de différence significative entre l'IDM à 6 mois et l'IDM à 24 mois pour la population d'où a été tiré l'échantillon d'enfants du groupe témoin.

Remarques.

1. Nos données comprennent 31 observations pour IDM-6 (le groupe témoin seul), mais 56 pour IDM-24 (groupe témoin et groupe expérimental). Remarquez que Statistica réalise le test en ne considérant que les 31 paires "complètes" : les valeurs manquantes sont ignorées.

2. Statistica ne prévoit ici aucune procédure pour traiter le cas des petits échantillons, et l'aide renvoie à l'ouvrage de Siegel et Castellan pour traiter les cas où $n < 20$...

17 Le test de Wilcoxon - Groupes appariés

17.1 Le test des rangs signés de Wilcoxon

La comparaison des scores IDM-6 et IDM-24 peut également être effectuée à l'aide d'un test de Wilcoxon (test des rangs signés).

Utilisez le menu Statistiques - Tests non paramétriques - Comparaison de deux échantillons appariés.

Indiquez IDM-6 et IDM-24 comme variables et cliquez sur le bouton "Test de Wilcoxon, échantillons appariés".

Vous devriez obtenir le résultat suivant :

Couples de variables	Test de Wilcoxon pour Ech. Appariés (Enfants-PRN.sta) Tests significatifs marqués à p <,05000			
	N	T	Z	niv. p
IDM-6 & IDM-24	31	177,5000	1,381556	0,167109

On vérifie que la statistique calculée par Statistica est :

$$Z = \frac{T - \frac{N(N+1)}{4}}{E} \quad \text{avec} \quad E^2 = \frac{N(N+1)(2N+1)}{24}$$

où T est le maximum des deux sommes de rangs des différences positives et négatives. Cette statistique est pratiquement celle donnée en cours.

Il faut également remarquer que, comme précédemment :

- Il n'est tenu compte des ex-æquo : Statgraphics, qui introduit une correction pour les ex-æquo, trouve $Z=1,48944$ au lieu de $1,3815$.
- Il n'est pas prévu de procédure pour traiter le cas des petits échantillons.

Remarque.

Le traitement des différences nulles est particulièrement mal pris en compte par le test de Wilcoxon mis en œuvre dans Statistica. On pourra, par exemple, reprendre le fichier de données précédent et modifier les données en introduisant de plus en plus de différences nulles. Celles-ci ne seront éliminées (N actifs inférieur à 31) que dans de rares occasions.

17.2 Calcul du protocole des rangs signés

Il est possible de calculer le protocole des rangs signés à l'aide de Statistica. Pour cela :

- Ajoutez 4 colonnes supplémentaires en fin de tableau de données. Ces colonnes pourront être nommées : Diff, RgDiffAbs, RgPlus, RgMoins.
- Pour la colonne Diff, introduisez la formule : = V3-V4
- Pour la colonne RgDiffAbs, introduisez la formule =abs(Diff), puis transformez les valeurs en rangs.
- Pour la colonne RgPlus, introduisez la formule : = iif(Diff >0; RgDiffAbs; -9999)
- Pour la colonne RgMoins, introduisez la formule : = iif(Diff <0; RgDiffAbs; -9999).

Remarques.

1. La formule de calcul de la colonne RgDiffAbs reste mémorisée avec la feuille de données. Si on doit demander un recalcul des autres colonnes, il faudra éviter que le recalcul concerne cette colonne, ou refaire le calcul des rangs.
2. Remarquez l'utilisation de la valeur -9999 comme code pour les valeurs manquantes.
3. Le calcul précédent est correct en l'absence de différences nulles. S'il y avait des différences nulles, la formule de la colonne RgDiffAbs devrait être remplacée par : = iif(Diff<>0;abs(Diff);-9999)

Exercice

Ouvrez le fichier Wilcoxon.xls à l'aide d'Excel.

Recopiez les données dans une nouvelle feuille de données Statistica et faites un test de Wilcoxon pour étudier s'il existe une différence significative entre l'aîné et le cadet du point de vue de la variable étudiée.

Vous devriez obtenir une statistique Zobs égale à 2,18, et donc conclure à une différence significative entre les deux conditions.