

ANASTAT : Un système dédié à la gestion et à l'analyse de données non paramétriques¹

1. Présentation

ANASTAT est un programme qui permet de manipuler des données structurées (appelées aussi objets par la suite) et de calculer quelques statistiques simples à partir de ces données. Les objets les plus courants sont les vecteurs et les tableaux. D'autres objets, dits complexes, peuvent être construits sous la forme de couples ou de listes d'objets plus simples. ANASTAT est conçu de telle manière qu'il soit aisé d'ajouter de nouvelles structures et de nouveaux traitements. Il permet également de fabriquer des objets satisfaisant à certaines contraintes (effectifs de groupes donnés). Il fournit ainsi un environnement didactique qui permet de mesurer l'« espace » fourni par certaines structures, de simuler des plans expérimentaux ou encore de calculer des probabilités exactes selon diverses hypothèses concernant les contraintes.

Le développement de ANASTAT est lié aux travaux du Séminaire de Psychologie. Ce sont les « outils » utilisés dans certaines recherches qui ont été rassemblés petit à petit sous une forme « logicielle ». Les exemples de cette présentation mode d'emploi s'inspirent également de plusieurs recherches menées par des collaborateurs du Séminaire dont ils reprennent parfois les données. La prétention de ce document se limite toutefois à la description des outils. Il ne s'agit pas ici d'interpréter des résultats, ni même de faire des propositions méthodologiques.

L'analyse des interactions entre usage et développement dépasse le cadre de cette présentation. Notons toutefois que celles-ci ont conduit à un produit qui permet des opérations intermédiaires entre celles des logiciels statistiques classiques, par exemple STATISTICA (Statsoft, 1994) ou SPSS (Samuel, 1994), et celles de systèmes d'analyse textuelle comme "The Ethnograph" (Seidel & al, 1995).

Cette présentation traitera tout d'abord des manipulations de base: aperçu des objets utilisés, principe de fonctionnement, quelques fonctions disponibles. Dans cette partie, des exemples montreront les types de statistiques que l'on peut effectuer. Une présentation plus détaillée des manipulations d'objets est faite dans la troisième partie qui traite des objets complexes, des modèles, des filtres, de l'utilisation de dictionnaires et des macro-commandes. La

¹ Ce manuel concerne la version 4.02. Il complète celui de la version 3.2 parue dans le dossier de psychologue 48 (1997). Voici un bref historique des versions. Les premières versions étaient écrites en Lisp et ne fonctionnaient pas sur ordinateur personnel. La version 2 est la première version sur PC. Dès la version 2.3 toutes les procédures décrites dans le Dossier de psychologie No 38, mai 1991 pouvaient être effectuées. Les versions suivantes ont augmenté les possibilités d'analyse de variance non paramétrique (n x 2 critères ou jusqu'à 2 x 4 critères). La version 3.0 est la première fonctionnant sur Windows. Par rapport à la version 2.4, cette version permet de structurer des données brutes, telles que celles obtenues lors du dépouillement de protocoles. La version 3.1, par rapport à la version 3.0 intègre totalement le programme ANAPROT. Elle permet d'utiliser des dictionnaires dans le dépouillement de données. D'autres rubriques de moindre importance ont été ajoutées. La version 3.12 permet de définir des macro-commandes. La version 3.2 introduit la possibilité de générer des données brutes. Les versions ultérieures concernent le passage à Windows 32 bits (XP) et s'attachent à corriger quelques « bugs ». Le programme est téléchargeable à l'adresse: <http://www.irdp.ch/methodo/anastat.htm>.

quatrième partie présente quelques exemples sous la forme de questions-réponses. L'annexe résume les fonctions disponibles. Elle esquisse d'autres possibilités offertes par ANASTAT, mais dont le détail du fonctionnement dépasse le cadre de cette présentation.

2. Manipulations de base

2.1. Les objets simples

Les objets simples sont des « atomes » de structures. Ils sont prédéfinis. Il y en a de deux types. Les nombres et les symboles sont des objets simples élémentaires. Les autres objets simples prédéfinis sont les vecteurs (suite de nombres), les tableaux (données organisées selon deux critères) et les boîtes (trois critères).

Vecteur

Un vecteur est une liste de valeurs entières avec, en option, une liste d'étiquettes. Une étiquette doit toujours être constituée de lettres, de chiffres et du symbole `_`. Elle doit commencer par une lettre. Mais on peut écrire "1P" !

| | | | | | |
|-------------|---------|---|---------|----|---------|
| Exemple 1 : | | | | | |
| | groupe1 | | groupe2 | | groupe3 |
| | 23 | | 24 | | 12 |
| Exemple 2 : | | | | | |
| | 12 | 0 | 1 | 21 | 23 |

Tableau 1: Deux exemples de vecteurs; avec ou sans étiquettes.

Tableau

C'est un tableau au sens habituel. Il faut mettre obligatoirement des étiquettes aux lignes.

| | | | | | |
|-------------|---------|----|---|---|---|
| Exemple 1 : | | | | | |
| | | nc | i | c | |
| | filles | 3 | 2 | 2 | |
| | garçons | 1 | 0 | 4 | |
| Exemple 2 : | | | | | |
| | adulte | 2 | 0 | 1 | 4 |
| | enfant | 2 | 0 | 1 | 1 |

Tableau 2: Deux exemples de tableaux; avec ou sans étiquettes. Les étiquettes nc, i, c représentent l'échelle ordinale: " non conservant, intermédiaire, conservant", fréquemment utilisée dans les épreuves opératoires piagésiennes.

Boîte

Une boîte est une structure contenant des données organisées en « plan factoriel ».

| | | | |
|-----------|-----|---------|-----------|
| Exemple : | | | |
| | | progrès | stabilité |
| -coop | ds | 8 | 6 |
| | sds | 4 | 6 |
| -comp | ds | 2 | 7 |
| | sds | 1 | 11 |

Tableau 3: Un exemple de boîte qui reprend un dispositif de Nicolet (1995) avec des données fictives. Cette structure permet d'étudier la progression de sujets dans quatre types de conditions expérimentales obtenues par le croisement de deux critères: mode coopératif ou compétitif, avec ou sans "droit social".

2.2. Les objets complexes et les autres structures

D'autres objets peuvent être construits par assemblage d'objets simples sous forme de listes ou de couples. L'écriture d'un objet complexe doit respecter une certaine syntaxe. Les objets simples doivent être précédés des mots-clés: VECTEUR, TABLEAU ou BOITE. Les couples sont formés de deux objets entre parenthèses, séparés par une virgule. Une liste est constituée d'une suite d'objets entre crochets [] et séparés par des point-virgules.

| | | | | |
|-----------|---|-----|-----|------|
| Exemple : | | | | |
| (| TABLEAU | | | |
| | c0 | c10 | c50 | c100 |
| m | 2 | 10 | 1 | 3 |
| c | 1 | 10 | 2 | 6 |
| e | 0 | 3 | 3 | 3 , |
| [| ("K de Kruskal =" , 2.93) ; ("facteur de correction =" , 0.8345) ; ("p =" , 0.1725)]) | | | |

Tableau 4: Objet complexe formé d'un couple contenant un tableau et une liste. La liste est elle-même constituée de couples: (texte, valeur).

Les objets complexes servent surtout à conserver des données « brutes » classiques, par exemple: [fille ; cond1 ; 2 ; 4 ; 8]. Les objets complexes peuvent aussi être définis à partir de modèles, c'est-à-dire d'objets génériques ou classes d'objets.

A part les objets et les modèles, il existe encore une troisième structure: les filtres, c'est-à-dire des structures qui présentent un profil (pattern) donné et qui permettent de sélectionner une famille d'objets. Par la suite le terme objet désignera parfois aussi bien un objet au sens strict qu'un modèle ou un filtre.

La possibilité d'utiliser des filtres est une particularité de ANASTAT. Les filtres permettent de dénombrer des observations en fonction de patterns déterminés. Par exemple, le nombre d'arguments de type logique dans une séquence argumentative émise par un « sujet

expérimenté » ou encore le nombre de fois où un argument de type logique a complété un argument de type prescriptif. Cette opération est souvent difficile à effectuer avec des logiciels statistiques standard qui, la plupart du temps, présupposent que toutes les observations ont le même nombre d'items et qui, d'autre part, permettent difficilement des traitements horizontaux des variables.

2.3. Principe de fonctionnement

Tous les objets disponibles sont affichés sous la forme de vignettes dans des fenêtres. Les analyses s'effectuent en sélectionnant les objets à l'aide de la souris et en leur appliquant une fonction choisie dans le menu ou, pour les opérations les plus fréquemment utilisées, dans la barre d'outils.

La figure 1 présente le bureau d'ANASTAT. D'autres figures illustreront de façon plus détaillée les fenêtres des objets, des modèles et des filtres.

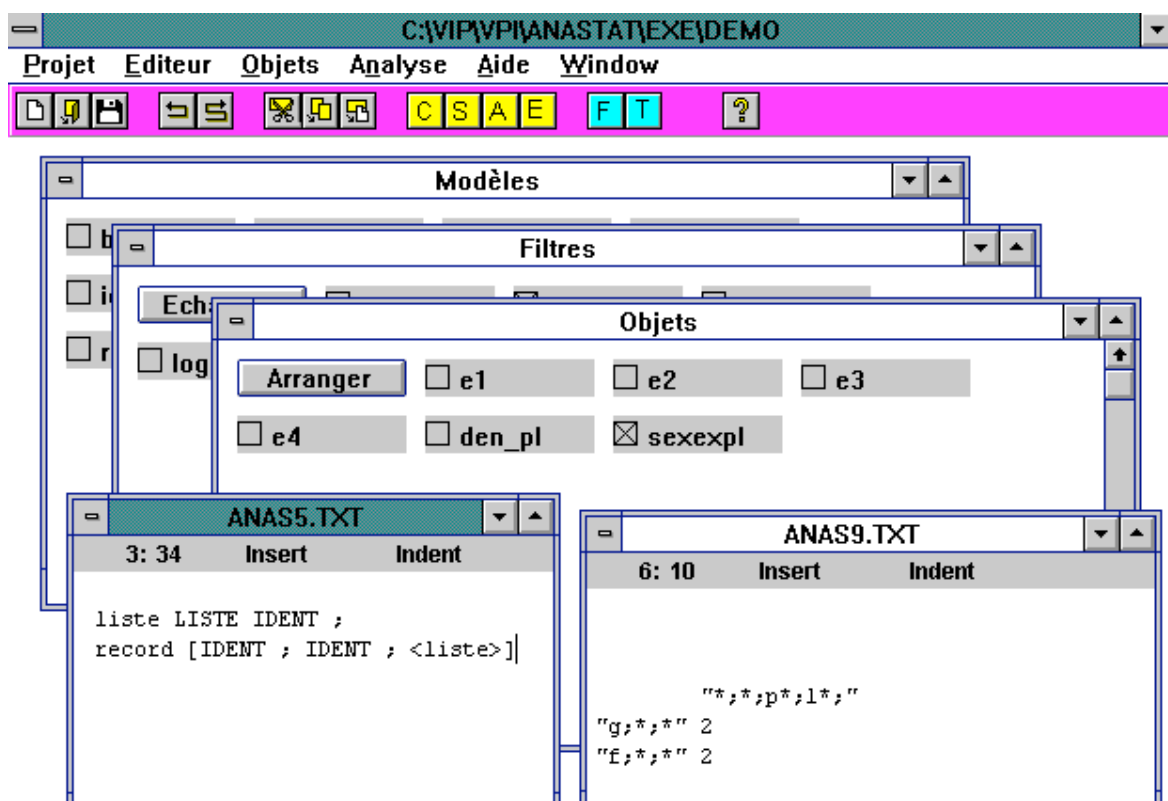


Figure 1: Le bureau d'ANASTAT. On aperçoit la fenêtre contenant les différents objets disponibles, celle contenant des objets, celle des modèles et celle des filtres. La fenêtre ANAS5.TXT est un éditeur de travail qui a servi à définir un modèle. La fenêtre à droite au sommet de la pile contient un résultat d'analyse. Sur la barre d'outils on trouve les fonctions les plus fréquemment utilisées. Outre les opérations concernant le projet et l'éditeur, on y trouve des opérations concernant les objets: compilation, sélection, affichage et effacement. Les deux boutons suivants servent à effectuer les opérations de filtrage ou à calculer des totaux marginaux.

Tous les objets à disposition sont rassemblés et visibles sur un panneau (figure 2). Le système n'est donc pas fait pour traiter des milliers d'enregistrements « anonymes », mais bien plutôt pour travailler de façon plus « personnalisée » avec quelques dizaines d'enregistrements. Les objets qui doivent être soumis à un traitement sont sélectionnés. A la fin du traitement seuls restent sélectionnés les objets auxquels le traitement a été réellement appliqué (pour les autres le traitement s'est révélé inapproprié comme par exemple l'application d'un test simple à des données organisées en plan factoriel).

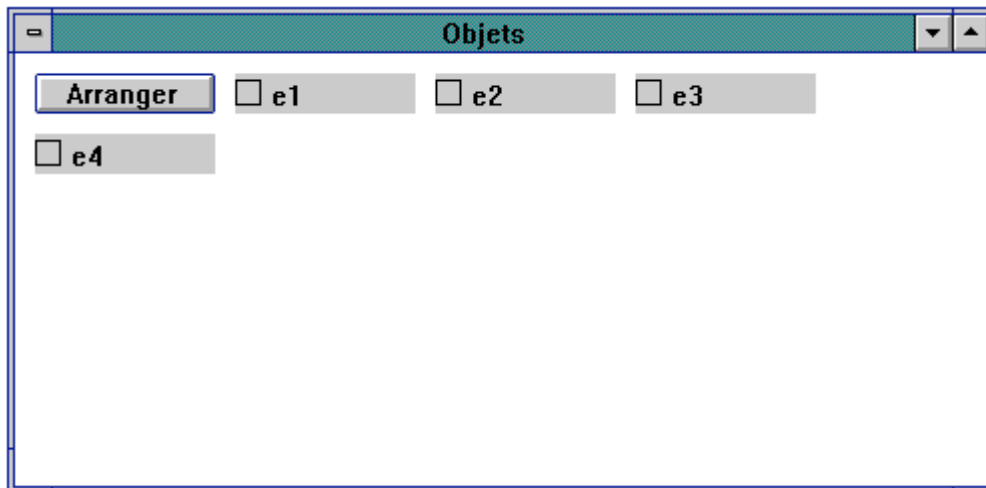


Figure 2: La fenêtre des objets. Le bouton « Arranger » permet de réorganiser la fenêtre lorsque des objets ont été effacés.

Le traitement de chaque objet donne naissance à un nouvel objet: le résultat du traitement. L'adoption de cette technique a l'avantage de permettre de travailler par affinements successifs des données. Elle présente l'inconvénient de multiplier les objets. Il s'agira donc d'effacer régulièrement les objets inutilisés, ce qui suppose avoir une bonne sauvegarde des objets primitifs au cas où une erreur de manipulation (sauvetage inopiné) ou une saturation du système. Une option permet de remplacer les anciens objets plutôt que d'en rajouter de nouveaux. Parfois, on souhaiterait rassembler un ensemble de résultats pour une analyse globale. Pour cela trois options ont été prévues. La première (*Compléter les objets*) ajoute le résultat obtenu à l'objet traité. Cela correspond à l'ajout d'une variable. Cette option a le désavantage « d'alourdir » des objets et de modifier leur structure. Les deux autres sont plus utiles: l'option *Regrouper* fabrique un objet contenant tous les résultats. L'option *Regrouper et résumer* fait de même puis rassemble tous les objets ayant la même structure en un seul. L'exemple ci-dessous montre comment trois tableaux sont regroupés en un seul.

| | | | |
|-------------------|---------------|------------|------------|
| tableaux initiaux | nc i c | nc inc i c | nc i ic c |
| | g1 1 2 0 | g1 2 1 2 3 | g1 1 0 1 1 |
| | g2 2 1 1 | g2 0 2 2 4 | g2 0 1 2 3 |
| regroupement | nc inc i ic c | | |
| | g1 4 1 4 1 4 | | |
| | g2 2 2 4 2 8 | | |

Tableau 5: Regroupement de tableaux.

2.4. La création d'objets

Toutes les données brutes sont tapées dans un « éditeur » en suivant une syntaxe donnée. Avant de les utiliser, il faut les transformer dans un format interne par une opération de « compilation ». La commande se trouve dans le menu *Objet*.

Plusieurs façons de présenter les objets existent. Un tableau peut simplement s'écrire dans sa forme habituelle.

| | nc | i | c |
|---|----|---|----|
| f | 2 | 4 | 6 |
| d | 3 | 7 | 10 |

Tableau 6: Un tableau dans l'éditeur. f et d représentent deux conditions expérimentales.

Lors de la compilation, il faudra toutefois indiquer qu'il s'agit d'un tableau. Le système demandera également de le nommer.

Il est également possible de compiler plusieurs objets en même temps (option *automatique* de la rubrique *compiler*). Dans ce cas, les objets doivent être accompagnés d'informations complémentaires, leur type et leur nom éventuel. Dans l'exemple ci-dessous, on notera que deux objets ont des noms. Le troisième en est dépourvu. ANASTAT demandera de l'introduire lors de la compilation.

| | | | |
|------------|----|----|----|
| o1 VECTEUR | | | |
| | nc | i | c |
| | 5 | 11 | 16 |
| ; | | | |
| TABLEAU | | | |
| | nc | i | c |
| f | 2 | 4 | 6 |
| d | 3 | 7 | 10 |
| ; | | | |
| o2 BOITE | | | |
| | nc | i | c |
| -garçons | | | |
| f | 2 | 4 | 6 |
| d | 3 | 7 | 10 |
| -filles | | | |
| f | 3 | 6 | 6 |
| d | 6 | 4 | 3 |

Tableau 7: Une suite d'objets dans l'éditeur, avec indication du type et séparés par des ;. Deux objets différents doivent avoir des noms différents. Dans le cas où des noms se répètent, ANASTAT demandera de les modifier, proposera d'ajouter un numéro de version ou ne gardera que l'objet le plus récemment créé.

2.5. Le traitement statistique des objets simples

Une fois des objets sélectionnés on peut leur appliquer un traitement. Certains concernent des mises en forme (permutation des lignes et colonnes), des calculs simples (totaux marginaux) ou des calculs d'indices statistiques (S de Kendall, L de Meddis, K de Kruskal). Pour certaines de ces statistiques, il est possible de travailler sur des tableaux selon les techniques décrites dans Pochon (1991, pages 42, 45 et 47) ou des boîtes (plan factoriel). La description des résultats avec plan factoriel est donnée ci-dessous.

| | nc | i | c |
|----------|----|---|----|
| -garçons | | | |
| f | 2 | 4 | 6 |
| d | 3 | 7 | 10 |
| -filles | | | |
| f | 3 | 6 | 6 |
| d | 6 | 4 | 3 |

S1 = 2 (influence des conditions expérimentales sur le niveau opératoire, chez les garçons)
 S2 = -57 (idem pour les filles)
 SI = 59 (interaction)
 p1 = 0.483 ; p2 = 0.08191 ; pI = 0.172 (probabilités correspondantes)

Tableau 8: Analyse d'interaction basée sur le S (Leach, 1979, Pochon, 1991, page 50 et annexe 6)

| | nc | i | c |
|----------|----|---|-----|
| -garçons | | | |
| f | 2 | 4 | 6 |
| d | 3 | 7 | 10 |
| -filles | | | |
| f | 3 | 6 | 6 |
| d | 6 | 4 | 3 ; |

L1 = -103 (L de Meddis pour effet sexe)
 M1 = -183 (moyenne attendue)
 sig 1 = 125.71 (écart type attendu, (z s'obtient par la formule: (L1 - M1)/sig 1)
 L2 = 331 (L de Meddis pour effet situation expérimentale)
 M2 = 122
 sig 2 = 126.066
 LI = -446 (L de Meddis pour l'effet d'interaction)
 MI = -305
 sig I = 124.58
 p1 = 0.262 ; p2 = 0.0487 ; pI = 0.129 (probabilités correspondantes)

Tableau 9: Analyse d'interaction basée sur la procédure de Meddis (Meddis, 1984, Pochon, 1991, page 51). Le système peut traiter des plans factoriels jusqu'à concurrence de 2x4 caractères.

| | | | |
|--|----|----|------|
| | nc | i | c |
| -s | | | |
| f | 1 | 4 | 8 |
| d | 3 | 6 | 5 |
| -m | | | |
| f | 13 | 9 | 12 |
| d | 11 | 12 | 10 |
| -i | | | |
| f | 35 | 15 | 6 |
| d | 22 | 25 | 10 ; |
| Ktotal = 21.997043351 (variancetotale) | | | |
| K1 = 16.581689077 (variance due à l'effet classe sociale) | | | |
| K2 = 1.2668664187 (variance due à l'effet situation expérimentale) | | | |
| KI = 4.1484878551 (interaction) | | | |
| facteur de correction = 0.87547488392 | | | |
| Probabilités correspondantes | | | |
| ptotal = 0.0001 ; p1 = 0.0008 ; p2 = 0.229 ; pI = 0.0936 | | | |

Tableau 10: Analyse de variance non paramétrique (Meddis, 1984, Pochon, 1991, page 52). Les données sont reprises de Grossen (1988). Le système peut traiter des plans factoriels jusqu'à nx2 caractères. Il est également possible de présenter les tableaux sous la forme 2x3 ou 2x4.

3. Manipulations avancées

3.1. Objets complexes et modèles

A partir d'objets simples, il est possible de construire des objets plus complexes. Les différents types d'objets et les constructeurs disponibles sont contenus dans le tableau 11.

| Désignation | Exemple | Remarque |
|--------------------------|--|--|
| Elément de base | 1 ; 12.5 ; cond1 ; « S⇒ » | |
| Vecteur | VECTEUR a b c 1 2 3 | La disposition n'a pas d'importance. |
| Tableau | TABLEAU nc i c f 3 4 8 g 5 3 3 | |
| Boîte | BOITE nc i c -c1 f 3 4 8 g 5 3 3 -c2 f 4 4 5 g 5 4 4 | le tiret (-) introduit chaque couche ou plaque de la boîte. |
| Couple | (« p = » , 0.05) | |
| Liste | [f ; c1 ; [p1 ; l3]] | Une liste peut être composée d'éléments de différents types. |
| Liste d'éléments de base | {3 ; 2 ; 12 } | liste uniquement composées d'éléments semblables au premier ! {2 ; p1 } conduit à une erreur, {p1 ; 2 } est admis. |
| Insertion | <adresse> | l'objet 'adresse' est inséré. |

Tableau 11: Syntaxe utilisée pour écrire des objets.

Il est donc possible d'utiliser des objets dont la structure est relativement complexe. Par exemple, en imaginant une observation comprenant trois types de tests, on peut structurer les données recueillies sous la forme suivante:

```
e11 [(identification, [f ; c1 ; 8]); (test1, [1 ; 0 ; 1]); (test2 , [j ; j ; j ; f]); (test3 , [l1 ; p2 ; l3])]
```

e11 correspond au nom de l'objet (et souvent l'identification du sujet). Les données elles-mêmes sont formées d'une liste de trois couples. Chaque couple est constitué du nom du test et de la liste des résultats.

Lorsque des données sont saisies, il est possible de commettre des erreurs qui peuvent passer inaperçues tant que la syntaxe est respectée. C'est pour pallier à ce problème que les modèles ont été introduits. Ces modèles permettent de définir des structures d'objets. Ils pourront être utilisés pour vérifier qu'un objet est bien conforme à ce qui est souhaité. Chaque modèle est désigné par son nom et fait l'objet d'une description à l'aide de la syntaxe donnée dans les tableaux 13/1 et 13/2. La figure 3 montre la fenêtre des modèles.

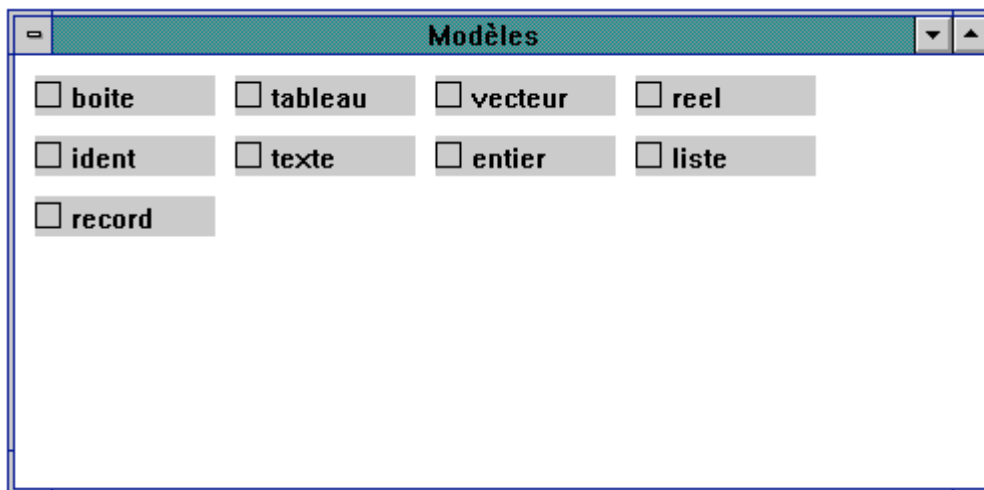


Figure 3: La fenêtre des modèles. La plupart des modèles présentés ici sont prédéfinis. Les modèles 'liste' et 'record' sont des modèles définis par l'utilisateur et font l'objet d'un exemple présenté ultérieurement.

Des données fictives serviront de base à quelques exemples. On suppose que le dépouillement de protocoles fournit les informations résumées dans le tableau 12 donnant une liste de justifications (à caractère logique ou prescriptif) fournies par des sujets lors d'un test.

| Identification | Sexe | Condition expérimentale | Liste de justification |
|----------------|------|-------------------------|------------------------|
| e1 | f | c1 | p1 p2 l2 l1 p1 p3 |
| e2 | f | c2 | l1 l3 p1 p2 p3 p2 l3 |
| e3 | g | c1 | p1 l2 l2 l3 l1 p2 l3 |
| e4 | g | c2 | p1 p2 l1 l1 l2 l3 |

Tableau 12: Données fictives donnant pour 4 sujets (selon le sexe et la condition expérimentale) une liste d'arguments justificatifs à caractère logique (l1, l2, l3) ou prescriptif (p1, p2, p3).

Un modèle est donné par: [IDENT ; IDENT ; LISTE IDENT] ce qui représente une liste constituée de 2 symboles (identificateurs) et d'une liste de symboles, liste de longueur variable. L'identification du sujet n'est pas comprise dans les données. Elle deviendra le nom de l'objet.

On pourra « compiler » ce modèle et y donner un nom. Par exemple: 'record'. Pour être compilées selon ce modèle, les données devront être aménagées selon la structure correspondante:

```
e1 [f ; c1 ; [p1 ; p2 ; l2 ; l1 ; p1 ; p3]] ;
e2 [f ; c2 ; [l1 ; l3 ; p1 ; p2 ; p3 ; p2 ; l3]] ;
e3 [g ; c1 ; [p1 ; l2 ; l2 ; l3 ; l1 ; p2 ; l3]] ;
e4 [g ; c2 ; [p1 ; p2 ; l1 ; l1 ; l2 ; l3]]
```

Ces informations tapées dans l'éditeur, puis compilées avec le modèle 'record' préalablement sélectionné fourniront 4 nouveaux objets de nom e1, e2, e3 et e4 (si une partie du texte est sélectionnée dans l'éditeur, seule cette partie est compilée). A noter qu'il n'est pas nécessaire dans un cas comme celui-ci de réaliser un modèle. Les données peuvent être compilées directement. Sans modèle, les erreurs de syntaxes sont toujours détectées (comme celle commise ici: e1 [f ; c1 ; [p1 ; p2 ; l2 ; l1 ; p1 ; p3]). Par contre les erreurs de structure ne sont pas détectées sans modèle (comme celle que l'on peut noter dans: e1 [f ; c1 ; p1 ; p2 ; l2 ; l1 ; p1 ; p3]).

| Nom | Modèle | Objet correspondant | Remarque |
|---|--------|---|---|
| Modèles d'objets simples, prédéfinis | | | |
| entier | ENTIER | 12 | |
| reel | REEL | 12.5 | Le point est utilisé |
| texte | TEXTE | “ceci est une chaîne“ | |
| ident | IDENT | prot_123 | commence par une lettre, ne contient que des lettres, chiffres et le caractère _ |
| vecteur | | 1 2 3 4 | nombres entiers |
| tableau | | a b c g1 1 2 3 g2 6 1 3 | le nombre de lignes ou de colonnes n'est pas limité. |
| boite | | a b c -c1 g1 1 2 3 g2 6 1 3 -c2 g1 1 2 3 g2 6 1 3 | les modèles vecteur, tableau et boite sont prédéfinis mais ne sont plus des objets élémentaires. C'est pour cela que la syntaxe du modèle n'est pas décrite ici. Si on veut utiliser un de ces modèles, on s'y référera par son nom: <boite> |

Tableau 13/1: Syntaxe utilisée pour décrire des modèles d'objets simples.

| Nom | Modèle | Exemple d'utilisation | Remarque |
|--|----------------|--------------------------|--|
| Modèles d'objets simples paramétrés | | | |
| | VECTEUR n | VECTEUR 3 | |
| | TABLEAU n n | TABLEAU 2 4 | l'utilisateur pourra nommer le modèle TABLEAU 2 4 par tableau2x4 |
| | BOITE n n n | BOITE 3 2 5 | 3 couches ou plaques, 2 lignes et 5 colonnes. |
| Construction de modèles | | | |
| | (mod, mod) | (IDENT , TEXTE) | Un objet correspondant est: (nom , « jacque-andré ») |
| | [mod ; ...] | [ENTIER ; VECTEUR 3] | Un objet correspondant: [3 ; 1 2 3] |
| | LISTE mod | LISTE ENTIER | {1 ; 2 ; 3 }. La liste peut aussi être écrite ainsi [1 ; 2 ; 3] ce qui peut conduire à une transformation des types (3 peut devenir « 3 »). |
| | LISTE * | | [e1 ; (age, 3) ; (sex, m)] |

Tableau 13/2: Syntaxe utilisée pour décrire des modèles d'objets. n représente un nombre entier, mod un modèle.

Des problèmes d'ordre linguistique peuvent créer quelques hésitations. En particulier, il s'agit de distinguer modèle et nom de modèle. Dans le cas des objets simples, le nom et le modèle sont désignés par les mêmes caractères, l'un en majuscule, l'autre en minuscule.

Les autres modèles sont soit des modèles simples avec paramètres, soit des constructions. Ils n'ont donc plus de nom associé. Mais l'utilisateur devra nommer un modèle qu'il a construit: par exemple le modèle (IDENT , TEXTE) peut, lors de la compilation, être appelé 'couple'.

Un modèle peut être construit à partir de modèles déjà existants. Pour faire référence à un autre modèle sans être obligé de le réécrire, on pourra utiliser son nom entre les symboles < et >. Par exemple: <couple>.

Le modèle 'record' défini précédemment [IDENT ; IDENT ; LISTE IDENT] pourrait s'écrire [IDENT ; IDENT ; <liste>] où 'liste' est le modèle LISTE IDENT.

3.2. Les filtres

Les filtres sont des objets particuliers qui permettent de trier les données. L'expérience montre que le dépouillement des protocoles se réalise souvent en résumant le contenu par une liste de symboles. Par exemple: c1 pour condition expérimentale 1; pn2 prise de décision du deuxième type par le novice. Les filtres permettent de sélectionner de telles données organisées selon des modèles. Ces filtres présupposent que les symboles utilisés suivent une certaine logique lexicographique, par exemple: pn1, pn2, .. pour les différents types de décision prises par des novices, pe1, pe2, ... pour celles des experts. Le filtre p* représentera tous les types de décision, pn* tous les types de décision pour les novices, etc.

Un mécanisme supplémentaire a été ajouté, celui de "capture". Ainsi le filtre pn!* ne va pas seulement considérer les différentes prises de décision des novice, mais il va se « souvenir » des symboles passés en revue. Il y aura donc une possibilité de « ventiler », avec un seul filtre, les différents types de prises de décision.

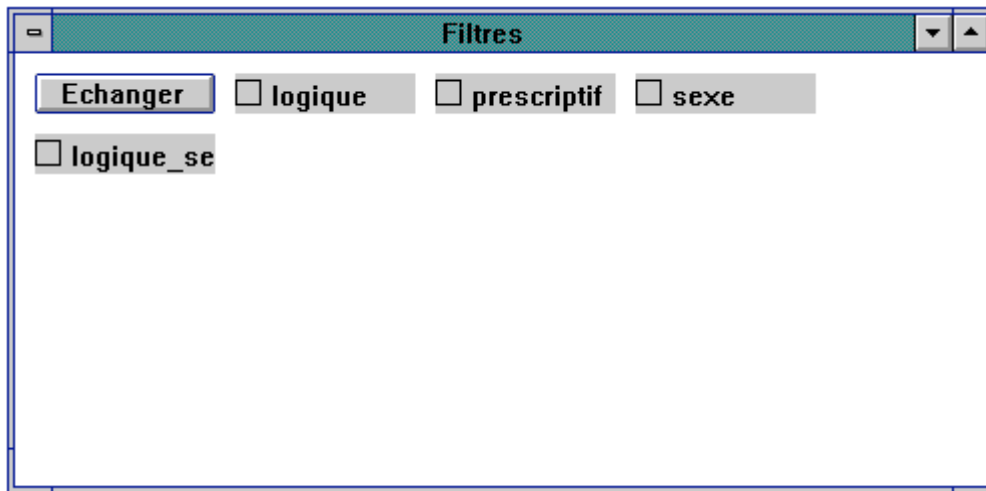


Figure 4: La fenêtre des filtres. Le bouton « Echanger » permet de permuter les deux premiers filtres sélectionnés. Cette fonction est utile lors du filtrage de données. En effet, dans le cas où deux filtres sont sélectionnés, le résultat sera contenu dans un tableau croisé. Le premier filtre sélectionné va définir la variable dont les valeurs seront les étiquettes des lignes. La deuxième correspondra aux étiquettes des colonnes. Lorsque trois filtres sont sélectionnés, l'objet résultant du filtrage sera une boîte (plaques ou couches, lignes puis colonnes).

Du point de vue syntaxique, les filtres sont donnés en reprenant la structure des objets, mais avec la possibilité de remplacer certains éléments par des jokers (tableau 14).

A noter que, après une opération de filtrage, tous les éléments filtrés, et seulement ceux-là, sont sélectionnés. Il est alors possible de ne garder qu'une partie des données en effaçant cette partie sélectionnée ou la partie non sélectionnée.

| Signe utilisé | Explication | Exemple |
|---------------------|--|-----------------------|
| <i>Sans capture</i> | | |
| * | Remplace n'importe quel objet | [* ; * ; *] |
| id*id | joker dans un nom | pe* ou *e1 ou p*1 |
| id?id | ? remplace une lettre | c? |
| N | filtre un nombre donné | 12 |
| <n >n <=n >=n | filtre les nombres inférieurs, supérieurs à une valeur n | <= 12 |
| F1 ; F2 ; ... | opération ou | c1 ; c2 |
| & F1 ; F2 ; ... & | opération et | & e* ; *f & = e*f |
| (F1 , F2) | pour filtrer un couple | (* , !*) |
| [F1 ; F2 ; ...] | pour filtrer une liste | [f ; c* ; [p* ; l*]] |
| \$ | représente un ou plusieurs termes dans une liste | [\$; p* ; l* ; \$] |

| | | |
|-----------------------------|---|--|
| # i = F | permet de filtrer directement l'item i dans une liste | correspond à l'usage des variables classiques |
| Avec capture | | |
| !* | comme * mais avec capture | |
| id!id | comme ? mais avec capture | |
| ! < n ! > n | comme < ou > avec capture | |
| ! <= n ! >=n | idem | |
| ! # i | capture l'item i d'une liste. ! # 3 est la même chose que # 3 = !* | correspond à la sélection classique d'une variable |
| ! @ n | capture les n premiers caractères d'un symbole. | |
| @ F F = * ; id*id ; id?* | compte le nombre d'éléments d'une liste satisfaisant F (F sans capture). Cette valeur est capturée. | @l* |
| @ [F1 ; F2 ; ...] | compte le nombre de fois où le pattern indiqué apparaît dans la liste filtrée. Cette valeur est capturée. | @[\$; l* ; p* ; \$] |

Tableau 14: Syntaxe des filtres. Comme avec les objets ou les modèles, il est possible d'insérer un filtre à l'aide de son nom: <filtre>

Extraction

Si le filtrage permet d'extraire une famille d'enregistrement, la rubrique *Extraire / Un item* du menu *Analyse* permet de ne prendre qu'une partie des informations de chaque enregistrement sélectionné. Dans l'exemple du tableau 12, l'item 3 consiste en la liste des arguments que l'on pourrait traiter indépendamment du reste des données (gain de temps et de place!).

3.3. Dictionnaire

Après dépouillement et normalisation des informations représentant les diverses phases d'une interaction, il est en général procédé à une catégorisation des événements observés. Le problème qui se pose souvent est que plusieurs catégorisations peuvent être faites, soit lors d'un dépouillement multiple, soit par affinements successifs, soit encore dans le cas où des données veulent être analysées selon différents points de vue. Plutôt que de devoir recoder les données à plusieurs reprises, ANASTAT utilise le principe des « dictionnaires ». Les informations sont codées dans une forme aussi riche que possible et les catégories sont réalisées de façon « virtuelle ».

Par exemple dans le cas d'une expérience où le sujet est amené à écrire des mots qu'évoquent certains inducteurs, les mots pourront être entrés in extenso (après une certaine normalisation

orthographique). Puis les catégories se réalisent en notant les mots qui appartiennent à une même famille: math : [mathématique ; calcul ; opération ; nombre]

Plusieurs dictionnaires peuvent se superposer ce qui permet des affinements successifs. Cette possibilité d'utilisation récursive de la notion nécessite une précaution: les symboles utilisés pour désigner les catégories à un niveau donné, ne doivent pas se retrouver dans le vocabulaire de ce même niveau (à éviter: math : [mathématique ; calcul ; opération ; math ; nombre]) !

3.4. Quelques exemples

Analyse d'un tableau

On reprend les données du tableau 3. Ce tableau est entré dans un éditeur (*Projet / Ouvrir éditeur*) sans préoccupation de mise en forme. Puis la fenêtre de l'éditeur étant active, on sélectionne le menu *Objets/Compiler/Objet standard*. Dans le menu déroulant qui s'affiche alors, on sélectionne *boite*. Demande est faite par ANASTAT de nommer l'objet, soit 'tableau1', par exemple. On peut aussi ne compiler qu'une partie de l'information contenue dans un éditeur en sélectionnant cette partie.

Ensuite on sélectionnera l'objet 'tableau1' dans la fenêtre des objets (attention à désélectionner les autres objets). En activant le menu *Analyse/ Coefficients statistiques /L de Meddis/Plan factoriel* un nouvel objet est créé: 'tableau1_Fmeddis'. On peut sélectionner cet objet et l'afficher. Ce qui conduit au résultat suivant:

```
[ ( "L1 =", -111)
  ( "M1 =", 69);
  ( "sig1 =", 71.79516317);
  ( "L2 =", -82);
  ( "M2 =", 23);
  ( "sig2 =", 71.93747285);
  ( "LI =", 146);
  ( "MI =", 161);
  ( "sigI =", 71.079341072);
  ( "p1 =", 0.0060857666715);
  ( "p2 =", 0.072199934456);
  ( "pI =", 0.41643126523)]
```

Le tableau et le résultat auraient pu ne constituer qu'un seul objet si l'option *Compléter les objets* avec été préalablement activée.

Utilisation de filtres

Nous reprenons les données (fictives) du tableau 12. On suppose défini un modèle de nom 'liste': LISTE IDENT et un modèle de nom 'record': [IDENT ; IDENT ; <liste>]. Les données sont écrites dans l'éditeur et compilées, le modèle 'record' étant actif.

```
e1 [f ; c1 ; [p1 ; p2 ; l2 ; l1 ; p1 ; p3]];
```

e2 [f ; c2 ; [l1 ; l3 ; p1 ; p2 ; p3 ; p2 ; l3]] ;

e3 [g ; c1 ; [p1 ; l2 ; l2 ; l3 ; l1 ; p2 ; l3]] ;

e4 [g ; c2 ; [p1 ; p2 ; l1 ; l1 ; l2 ; l3]]

Nous aurons donc quatre objets de noms e1, e2, e3 et e4. Chaque sujet devient donc un objet!
Dans toute cette section, on supposera l'option *Regrouper et résumer* activée.

Filtres sans capture

On définit le filtre 'fille' par: [f ; * ; *] (on pourrait aussi le résumer par [f;\$]). L'opération de filtrage va créer un objet (en l'occurrence '[fille];1') consistant en:

| |
|-------|
| f;*,* |
| 2 |

Mais cette opération va aussi sélectionner automatiquement toutes les filles. En effaçant la sélection on peut ainsi garder que les garçons. En effaçant les objets non sélectionnés, c'est le contraire qui se passe.

Filtres avec capture

Trois filtres seront définis: le filtre 'sexe': [!* ; * ; *] (on pourrait plus simplement écrire [!* ; \$]), le filtre 'cond_exp': [* ; c!* ; *] et le filtre 'logique' : [* ; * ; @l*]

Si on sélectionne le filtre 'sexe' et les 4 objets e1, e2, e3 et e4, le résultat du filtrage sera le nouvel objet (vecteur):

| | |
|-------|-------|
| g;*,* | f;*,* |
| 2 | 2 |

Le filtre !* a capturé les symboles représentant le sexe (le discours est statistique et non psychologique).

Si on sélectionne le filtre 'cond_exp' et les 4 objets e1, e2, e3 et e4, le résultat du filtrage sera un nouveau vecteur dont les étiquettes contiennent, par capture, la condition expérimentale:

| | |
|--------|--------|
| *,c2;* | *,c1;* |
| 2 | 2 |

Si on sélectionne le filtre 'logique' et les 4 objets e1, e2, e3 et e4, le résultat du filtrage sera le nouvel objet:

| | | | |
|----------|----------|----------|----------|
| *,*;!*_4 | *,*;!*_5 | *,*;!*_3 | *,*;!*_2 |
| 1 | 1 | 1 | 1 |

Cela signifie que 1 observation contient 2 arguments logiques, de même pour 3 arguments, 4 ou 5. Il n'y a pas d'observation contenant moins de deux arguments logiques. Le filtre @l* a

compté le nombre d'éléments filtrés par l* (arguments logiques) et a utilisé ce décompte comme élément capturé.

Si on sélectionne les filtres 'sexe' et 'logique' (dans cet ordre) et les 4 objets e1, e2, e3 et e4, le résultat du filtrage sera le nouvel objet de type 'tableau' qui donne le nombre d'arguments logiques utilisés respectivement par les filles et les garçons.

| | *;*_l*_2 | *;*_l*_3 | *;*_l*_5 | *;*_l*_4 |
|-------|----------|----------|----------|----------|
| f;*,* | 1 | 1 | 0 | 0 |
| g;*,* | 0 | 0 | 1 | 1 |

Après remise en ordre et "toilettage", ce tableau être peut utilisé pour des analyses ultérieures

Nous allons maintenant considérer le filtre 'pl' : [* ; * ; @[\$; p* ; l* ; \$]] qui compte le nombre de fois où un argument logique suit un argument prescriptif. Les signes \$ sont là pour indiquer que ces paquets, argument prescriptif - argument logique, peuvent se trouver séparés par d'autres éléments.

L'utilisation de ce filtre sur notre jeu de données conduit au vecteur suivant:

| *;*_l[\$;p*;l*;\$]_2 | *;*_l[\$;p*;l*;\$]_1 |
|----------------------|----------------------|
| 1 | 3 |

Pour 1 sujet, le pattern « argument prescriptif-argument logique » s'observe deux fois. Pour les 3 autres sujets, il n'est présent qu'une seule fois.

L'utilisation simultanée des filtres 'sexe' et 'pl' donne les mêmes informations ventilées entre le groupe des filles et celui des garçons.

| | *;*_l[\$;p*;l*;\$]_2 | *;*_l[\$;p*;l*;\$]_1 |
|-------|----------------------|----------------------|
| g;*,* | 1 | 1 |
| f;*,* | 0 | 2 |

A noter que l'usage des filtres 'sexe' et d'un filtre similaire à 'pl' : [* ; * ; [\$; p* ; l* ; \$]] ne procède pas au comptage, mais signale simplement l'existence ou l'absence du profil (pattern) indiqué.

| | *;*_p*;l* |
|-------|-----------|
| g;*,* | 2 |
| f;*,* | 2 |

Usage de dictionnaires

Un dictionnaire est un ensemble de catégories de mots. Chaque catégorie possède un nom. Par exemple:

math {calcul ; opération ; addition ; nombre } ;
 eval { note ; « travaux écrits » ; carnet ; moyenne } ;
 cogni {penser ; réfléchir ; intelligence } ;
 socio {amis ; camarades ; copains } ;
 systè {branches ; horaire }

Et voici des données que l'on compile avec le modèle 'obs': [IDENT ; LISTE IDENT]

e1 [f ; {calcul ; note ; opération }] ;
 e2 [f ; {penser ; amis ; nombre }] ;
 e3 [f ; {note ; nombre ; "travaux écrits" ; carnet ; moyenne}] ;
 e4 [g ; {penser ; calcul ; intelligence ; carnet }] ;
 e5 [g ; {amis ; camarades ; copains ; moyenne ; réfléchir}] ;
 e6 [g ; {branches ; horaire ; nombre ; copains }]

On crée également le filtre 'math' : [* ; #math]

Le filtrage (sans chargement du dictionnaire) montre que le mot 'math' ne se trouve dans aucun enregistrement:

| |
|------------|
| *,[math]_0 |
| 6 |

Après chargement du dictionnaire, par contre, on constate que cette notion apparaît 0 fois dans un enregistrement, 1 fois à 4 reprises et 1 fois dans 2 enregistrements.

| | | |
|------------|------------|------------|
| *,[math]_0 | *,[math]_1 | *,[math]_2 |
| 1 | 4 | 1 |

3.5. Les macro-commandes

Les macro-commandes ou macros sont des nouvelles commandes créées par assemblages de commandes élémentaires. Lorsque des mêmes analyses sont effectuées à plusieurs reprises, il est économique d'utiliser cette possibilité qui permet de gagner du temps et d'éviter des erreurs. Une macro peut être mise au point avec quelques objets avant d'être appliquée à un ensemble plus large.

Les macros peuvent être créées de façon automatique à l'aide de la rubrique *Enregistrer* du menu *Macro*. Elles peuvent aussi être écrites dans un éditeur et compilées. En principe, une méthode mixte doit souvent être employée.

Le tableau 15 présente un exemple de macro qui appliquée aux données ci-dessus fait un tableau avec la fréquence d'apparition des différents mots utilisés.

| Macro | Commentaires |
|--|--|
| vocabulaire [/ « ceci est un commentaire » / ; SELECTIONNER TOUT ; CHACUN extraire 2 ; EFFACER SEL ; OPTION 2 on ; SELECTIONNER TOUT ; CHACUN répétitions ; DESELECTIONNER TOUT] | Nom de la macro COMMANDE Ce qui suit CHACUN, jusqu'au prochain point-virgule concerne tous les objets sélectionnés. Les anciens objets sont effacés. L'option Regrouper et résumer est activée. Tous les nouveaux objets sont sélectionnés. Un décompte des contenus (qui se réduit alors à une liste de mots) est effectué. |

Tableau 15: Exemple de macro-commande.

Le tableau 16, résume la syntaxe des macros. Tous les traitements disponibles dans les menus sont réalisables à l'aide de macro-commandes. Mais des commandes supplémentaires existent. En particulier des opérations classiques telles que # I1 + #I2 qui remplace l'item I1 par la somme des items I1 et I2. Ces opérations s'effectuent au premier niveau, c'est-à-dire sur des données « classiques » dont un modèle peut être [IDENT ; IDENT ; ENTIER ; ENTIER ; ENTIER].

| Syntaxe | Explication, exemples |
|--|---|
| Forme générale | |
| nom [commande 1 ; commande 2 ; ...] | Une macro est donnée par un nom puis une suite de commandes entre crochets, séparées par des points-virgules. |
| Commandes (en majuscules) | |
| CHARGER nom d'un fichier | CHARGER « donnees.apr » |
| SELECTIONNER TOUT | Sélectionne tous les objets. |
| DESELECTIONNER TOUT | |
| SELECTIONNER filtre nom | Peut aussi s'utiliser avec objet ou modele. |
| DESELECTIONNER filtre nom | |
| EFFACER SEL | Effacer tous les objets sélectionnés. |
| EFFACER DESEL | Effacer tous les objets non sélectionnés. |
| EFFACER nom | Effacer l'objet nom |
| / « ceci est un commentaire » / | Les commentaires sont entre barres obliques et guillemets. |
| < nom > | Execution d'une macro. |
| CHACUN suite de traitements | Attention les traitements ne sont pas séparés par un point-virgule ! |

| Traitements (en minuscules) | |
|---|--|
| extraire n | Extrait l'item n si n est positif. Prend le reste si n est négatif. |
| Filtrer | Effectue un filtrage sur la base des données et des filtres sélectionnés. |
| pattern fixe libre fixe 2 | Tous les patterns possibles sont créés en supposant la somme des lignes fixes ou libre. Avec deux degrés de liberté concernant les boites. |
| distribuer s k fixe libre | On obtient la distribution de s ou k en supposant les totaux sur les lignes fixes ou libres. |
| sommer ligne colonne plaque | |
| \ ; \- | Diagonale et anti-diagonale |
| # n [m] | L'item n est recodé sur un intervalle de largeur m. |
| # n + r | On additionne r à l'item n. |
| # n * r | L'item n est multiplié par r |
| log# n | |
| #n + #m | Somme de deux items. Le résultat constitue le premier item mentionné. |
| #n - #m | Idem pour la différence. |
| #n * #m | Idem pour le produit. |
| #n / #m | Idem pour le quotient. |
| tri compilation copie decoupe transpose | Modification d'objets. |
| répétitions dim totaux | Calcul d'information complémentaire. |
| kendall | S avec information complète. |
| meddis | L avec information complète. |
| pearson | Chi-2 complet. |
| kruskal | K avec information complète. |
| wilcoxon r | W avec information complète. |
| binomiale p | Distribution binomiale avec valeur de coupure (probabilité). |
| w r | Calcul de la matrice associée à Wilcoxon avec valeur de coupure. |
| W | Calcul de Wilcoxon à partir d'une matrice. |
| s k d g c chi_2 | Calcul de coefficients. |
| Options | |
| OPTION no on off | Selon la valeur de no: 1) Regrouper 2) Regrouper et résumer 3) Compléter les objets 4) Remplacement des objets. |

Tableau 16 : Syntaxe des macro-commandes.

4. Questions et réponses

4.1. Notions de modèle et d'objet

Pourquoi est-il possible dans les modèles de se référer aux nombres entiers par ENTIER ou <entier> alors que cela n'est pas le cas pour les vecteurs (seul <vecteur> est admis)?

Cela est dû au fait que le nom vecteur recouvre toute une famille d'objets (des vecteurs de toutes les dimensions). Cela est également le cas pour les autres objets non élémentaires mais prédéfinis tableau, boîte, hyper.

Ce travail est donc à la charge de l'utilisateur. Si, par exemple, les tableaux de dimension 2x8 sont souvent utilisés, on pourra créer un modèle TABLEAU 2 8 et le nommer tableau28.

Quelle relation alors avec le mot VECTEUR que l'on écrit dans les données d'objet?

Dans ce cas, le mot VECTEUR représente une étiquette permettant d'orienter le contrôle des données. Cette étiquette n'est d'ailleurs pas nécessaire lorsque la compilation est effectuée à l'aide d'un modèle.

Quelle différence entre ces deux manières de présenter une famille d'objets?

| | | |
|------------|-----------|---------|
| o1 VECTEUR | [| VECTEUR |
| c i c | c i c | |
| 5 11 16 | 5 11 16 | |
| ; | ; | |
| TABLEAU | TABLEAU | |
| nc i c | nc i c | |
| f 2 4 6 | f 2 4 6 | |
| d 3 7 10 | d 3 7 10 | |
| ; | ; | |
| o2 BOITE | BOITE | |
| nc i c | nc i c | |
| -garçon | -garçon | |
| f 2 4 6 | f 2 4 6 | |
| d 3 7 10 | d 3 7 10 | |
| -fille | -fille | |
| f 3 6 6 | f 3 6 6 | |
| d 6 4 3 | d 6 4 3] | |

Dans le cas de la colonne de gauche, 3 objets seront créés. La colonne de droite ne donne lieu qu'à un seul objet, consistant en une liste de trois composantes.

Quelle méthode adopter pour entrer une famille de tableaux?

Ouvrir l'éditeur et entrer les tableaux sous la forme:

<le nom du tableau> TABLEAU <les données> ; (la mise en page n'a pas d'importance. Par exemple:

```
tab1 TABLEAU  f m e C1 2 0 4 C2 0 4 0 C3 0 0 4 ;
tab2 TABLEAU  r1 m p1 C1 1 3 2 C2 1 1 2 C3 1 1 2 ;
tab3 TABLEAU  m p1 C1 3 2 C2 4 0 C3 3 1 ;
tab4 TABLEAU  f m e C1 0 0 4 C2 0 2 0 C3 2 2 2 C3 0 0 2 ;
```

Puis de choisir dans le menu "objets" la rubrique "Compiler/Objet standard"

4.2. Filtres

Comment obtenir une information sur le nombre d'arguments donnés dans l'exemple du tableau 11?

Il y a plusieurs solutions possibles. En particulier on peut utiliser le filtre [`*` ; `*` ; `@*`]. On voit ici que le résultat n'est pas un nombre, mais un symbole. C'est une caractéristique du système.

Est-il possible de compiler en même temps plusieurs filtres, modèles, etc.

Non dans cette version. Mais c'est une bonne idée. Une version ultérieure pourrait travailler avec des fichiers sources contenant des sections (objet, filtre, modèle, etc.). De plus, il faudra pouvoir ajouter le nom des filtres et des modèles.

La notion de filtre est difficile à saisir. Peut-on faire des tableaux croisés classiques avec cet outil ?

Oui, cela ce présenterait de la façon suivante: les données sont constituées de suites régulières avec chaque fois, par exemple, le sexe (codé 1 ou 2), la condition expérimentale (codée 1 ou 2) et trois réponses (codées a, b ou c).

```
e1 [1; 1; a; b; c];
e2 [1; 2; b; c; a];
e3 [2; 1; c; a; b];
e4 [2; 2; a; b; c];
e5 [2; 2; b; c; a]
```

Nous voulons, par exemple, regarder les réponses à la première question en fonction du sexe. Les deux filtres nécessaires sont:

```
[!*,*,*,*,*]
```

```
[*,*,!**,*]
```

Après avoir activé l'option 'Regrouper et résumer' et utiliser la rubrique 'Filtrage d'objets' du menu 'Analyse' on obtient le tableau croisé:

```
sexe x q1  "*,*,c,*,*"  "*,*,b,*,*"  "*,*,a,*,*"
"2,*,*,*,*"  1          1          1
"1,*,*,*,*"  0          1          1
```

C'est moins lisible que dans les logiciels "classiques". Il manque en particulier l'information des filtres utilisés (**sexe x q1** a été ajouté après coup :). Une prochaine version pourrait

combler cette lacune. L'avantage des filtres est de pouvoir également utilisé pour la sélection une partie de l'information enregistrée dans la cellule et de pouvoir travailler sur des séquences de longueurs différentes.

Les données se trouvent dans 'standard.apt'.

4.3. Dictionnaire

Comment obtenir l'ensemble des mots utilisés sur un ensemble de données? Cela peut être utile avant de créer un dictionnaire.

La première idée est d'utiliser le filtre: [* ; @!*]. Mais ce traitement ne convient pas. Les filtres sont réservés à compter des enregistrements (à sélectionner des enregistrement présentant un profil donné) et non à comptabiliser *des* éléments contenus dans ces enregistrements. Il faudra utiliser ici la rubrique *répétition* et globaliser le résultat sur l'ensemble des enregistrements. On va traiter un exemple avec des enregistrements relativement riches qui suivent le modèle suivant:

[(IDENT , LISTE IDENT) ; (IDENT , LISTE IDENT) ; (IDENT , LISTE IDENT) ; (IDENT , LISTE IDENT) ; (IDENT , LISTE IDENT) ; (IDENT , LISTE IDENT)]

Voici un enregistrement *correspondant* à ce modèle avec tout d'abord une identification du sujet, puis un certain nombre de mots induits par des mots inducteurs.

[(identification , {p f;"4";"2";f;"15";i;si;ch;a;l;m15});

(maître,

{emmerdeur; rabâcher; étudier; travailler; réfléchir; écrire; discipline; rigolade; variété; sévère; interrogations; "il nous apprend"; corrections; punitions; encourager});

(école,

{profs; devoirs; règlement; camarades; congé; travail; discipline; interrogations; classe; cahiers; cours; entraide; bancs;rencontre;disputes});

(français,

{orthographe; rédactions; langage; profs; devoirs; interrogations; travail; méthodique; apprendre; effort; écouter; "être attentif"; mémoriser; conjugaison; application});

(math,

{profs; suivre; apprendre; mémoire; calculs; interrogations; explications; persévérance; comprendre; travail; application; chiffres; addition; multiplication; division});

(élève,

{apprendre; "être attentif"; "il doit être à l'heure"; punitions; discussion; habillement; travailler; "s'amuser"; discipliné; indiscipliné; adolescent; application; camaraderie; obéir}}]

Supposons que nous voulons faire la liste de tous les mots liés à maître. Les opérations à effectuer sont les suivantes: *sélectionner tout / extraire item 2 / effacer les objets sélectionnés / sélectionner tout / extraire item 2*. Dès ce moment les objets ne contiennent plus que la liste des mots associés à maître (attention à ne pas sauver la base si une sauvegarde de l'ensemble n'a pas été préalablement effectuée!). Ensuite, il suffit de cocher l'option *Regrouper et résumer* et de lancer l'opération *répétitions*.

4.4. Généralités

L'ancienne version demandait de choisir (pour le calcul de K, par exemple) entre plan simple et plan factoriel. Ce n'est plus le cas de la nouvelle version, pourquoi?

ANASTAT procède automatiquement à une analyse simple si l'objet sélectionné est un tableau et à une analyse selon le plan factoriel si l'objet est une "boîte".

Comment utiliser les données qui sont produites par un autre programme (SPSS, par exemple) ?

Il y a plusieurs méthodes possibles à partir d'un fichier (format ASCII) produit par un autre programme.

Méthode 1: Chargez le fichier dans l'éditeur (rubrique *Charger texte* du menu *Projet*). Il est possible de procéder dans cet éditeur à une mise au point. Puis il suffit de sélectionner *Compiler* du menu *Objets*.

Méthode 2: Par "copier/coller" à partir d'un texte écrit dans un autre logiciel (Word, par exemple) dans un éditeur ouvert par *Ouvrir éditeur* du menu *Projet*.

Quel nombre maximum d'objets le système est-il capable de traiter?

Ces limites ne sont pas clairement définies, cela va dépendre de la mémoire et de la rapidité d'exécution de l'ordinateur utilisé. Des travaux ont été réalisés avec 120 observations, chacune ayant environ 1000 caractères. Une autre limite est aussi liée au nombre de mots de vocabulaire dans le dictionnaire. Ces mots étant indexés, la table d'indexage fournit une certaine limite qui n'a pas encore été atteinte. L'éditeur, quant à lui, n'accepte que les textes qui ne dépassent pas 64000 caractères. Dans le cas de gros travaux, il ne sera pas possible d'afficher directement les résultats, il faudra d'abord les sauver dans un fichier avant de les reprendre par un autre système. Le tri de grands tableaux peut également poser quelques problèmes de mémoire.

4.5. Fonctions statistiques

Comment obtenir la distribution exacte du S liée à un tableau ?

Sélectionnez le tableau et utilisez la rubrique *K ou S possibles* puis *S* du menu *Analyse*. Un nouvel objet est créé auquel il suffit d'appliquer le traitement *Analyse / Répétitions*. Il ne vous reste plus qu'à afficher la sélection. Ensuite il faut encore calculer les multiplicités.

Pourquoi est-il possible de chercher toutes les valeurs de K ou S possibles alors que cela n'est pas possible pour d'autres coefficients?

C'est possible en suivant une voie moins directe que celle offerte par ce menu qui concerne les coefficients principaux et qui se trouve également être un vestige d'une version précédente. Il faut pour cela commencer par compiler un tableau (à faibles effectifs !) et le sélectionner. Puis chercher tous les patterns possibles. Puis sélectionner tous les tableaux obtenus (en désélectionnant le tableau initial). Puis cocher l'option *regrouper* et enclencher la fonction *Coefficients seuls* avec le coefficient souhaité. Le résultat est une liste des valeurs possibles du coefficient choisi avec les contraintes indiquées.

Quelle différence peut-on observer entre les résultats donnés par le test du Chi-2 et un test basé sur le S de Kendall ?

Première remarque, les tests basés sur le S de Kendall et celui du Chi-2 ne mesurent pas le même effet. Le Chi-2 cherche à mettre en évidence une différence de profil entre deux groupes, le S de Kendall mesure un effet de diagonale. Toutefois, en général, "Kendall significatif" implique "Chi carré significatif", mais il y a des exceptions.

Dans le cas ci-dessous, le test de Jonkheere (basé sur le S de Kendall) donne une valeur de p égale à 0,26, alors que le Chi-2 produit un p de 0,57. Il y a un certain effet diagonal mais les profils sont semblables.

| | | | | |
|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 1 |

Dans ce deuxième cas, c'est le contraire qui se produit: $S=0$ et donc $p > 0,5$ alors que le Chi-2 conduit à $p = 0,29$.

| | | | | |
|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 1 |
| 0 | 1 | 0 | 1 | 0 |

Est-il possible que le gamma soit petit alors qu'un test basé sur le S est significatif ?

La logique des "p-value" et celle des coefficients "gamma" n'est pas la même. Le rapport de l'un à l'autre fait intervenir l'écart type et le nombre d'observations.

Une "p-value" n'est valable que pour un tableau donné. Alors que le "gamma" permet de dire "combien de fois" un effet d'un tableau est plus important que celui d'un autre. Si le nombre d'observations est élevé et que l'écart type est assez important on peut avoir un Kendall significatif bien que l'effet observé soit faible (gamma petit).

Si l'effet observé est grand ($|\text{gamma}| > 0,70$), le test sera-t-il significatif?

Le tableau suivant montre le contraire: $\text{gamma} = -0,755$, $p = 0,11$:

| | | | |
|---|----|---|---|
| 2 | 11 | 4 | 4 |
| 1 | 3 | 0 | 0 |

La contradiction apparente est liée au grand nombre de sujets pour la deuxième case de la première ligne. La liberté pour obtenir des patterns du même type n'est pas très grande d'où un S peu significatif. Mais gamma prend en compte l'égalité comme un effet positif. D'où un effet important. Le coefficient "delta", presque équivalent à "gamma" mais qui ne prend pas en

compte les cas d'égalité, vaut ici -0,44. Il est nettement plus faible. Du point de vue méthodologique, il est tout à fait concevable qu'un grand effet soit observé, mais que, vu la distribution particulière des résultats, il ne peut pas être considéré comme « significatif »!

Des sujets effectuent un pré-test puis un post-test. Comment apprécier la progression des sujets d'un test à l'autre?

Les données seront du type: [2 ; 3]. Le traitement à appliquer va dépendre de l'hypothèse. Si une hypothèse de progression est faite, on pourra utiliser le test de Wilcoxon sur la différence entre le premier score (item 1) et le deuxième (item 2) avec 0 comme valeur de coupure. Si l'hypothèse est plutôt la stabilité, on peut utiliser un test basé sur le S de Kendall sur le tableau obtenu par croisement item1 x item2.

Des sujets effectuent un pré-test, puis sous deux conditions expérimentales différentes un post-test. Comment apprécier l'effet de la condition expérimentale sur la progression des sujets d'un test à l'autre?

Les données seront du type: [c1 ; 3 ; 4] avec la condition expérimentale suivie des deux résultats (considérés comme des rangs). Il y a plusieurs méthodes (outre les analyses de variances conventionnelles si le nombre de sujets et la mesure donnée par les tests le permettent) qui vont dépendre des hypothèses faites. On peut, par exemple: effectuer la différences entre l'item 3 et le 2, établir le tableau croisé « condition expérimentale x différence » et analyser ce tableau à l'aide d'un test basé sur le S. Une autre méthode consiste à analyser la boîte « condition expérimentale x item 1 x item 2 » à l'aide du test basé sur le S de Kendall.

Comment utiliser le S de Kendall dans le cas d'un plan factoriel ?

Exemple d'utilisation avec des données tirées d'une recherche de Alain Brossard, 1990) qui étudie le lien entre niveau atteint par des sujets dans les épreuves opératoires piagétienne (échelle ordinale) leur sexe ou le sexe de l'expérimentateur.

Effet regard (global)

| | nc | inc | ic | c |
|---------------------------------|----|-----|----|----|
| regards normaux | 15 | 15 | 12 | 10 |
| regards centrés sur la tâche | 8 | 17 | 12 | 17 |

S de Kendall = 539 ; correction = 27 ; sigma = 305.88 ; z = 1.674 ; p = 0.0471

Commentaire: Un effet avec $p < 0.05$ est constaté.

Effet regard dans des sous-groupes constitués selon le sexe des sujets

| | nc | inc | ic | c |
|---------------------------------|----|-----|----|---|
| -garçons | | | | |
| regards normaux | 9 | 10 | 4 | 5 |
| regards centrés sur la tâche | 6 | 12 | 7 | 5 |

| | | | | |
|-----------------|---------|---|---|----|
| | -filles | | | |
| regards normaux | 6 | 5 | 8 | 5 |
| regards centrés | 2 | 5 | 5 | 12 |
| sur la tâche | | | | |

$S1 = 99$; $S2 = 187$; $SI = -88$; $p1 = 0.2105$; $p2 = 0.0224$; $pI = 0.2854$

Commentaire: l'effet observé précédemment, apparaît avec un coefficient significatif dans le sous-groupe de filles et non significatif pour les garçons. Il n'y a pas d'effet "croisé".

Effet regard dans des sous-groupes constitués selon le sexe de l'expérimentateur

| | | | | | |
|-----------------|-------|-----|----|---|---|
| | nc | inc | ic | c | |
| | -masc | | | | |
| regards normaux | 8 | 7 | 4 | 6 | |
| regards centrés | | 4 | 7 | 6 | 9 |
| sur la tâche | | | | | |
| | -fem | | | | |
| regards normaux | 7 | 8 | 8 | 4 | |
| regards centrés | | 4 | 10 | 6 | 8 |
| sur la tâche | | | | | |

$S1 = 143$; $S2 = 120$; $SI = 23$; $p1 = 0.0816$; $p2 = 0.1475$; $pI = 0.4405$

Commentaire: il n'y a pas d'effet significatif dans aucun des deux sous-groupes. Cette "bizarrerie" est certainement due à la combinaison d'un effet faible et à la diminution du nombre des états possibles.

Quelle est l'utilité des distributions exactes ?

En principe, une distribution exacte permet de calculer des probabilités sans passer par l'artefact que constitue le passage par des distributions approchées.

Cette question en appelle une autre quant à la façon de calculer une distribution exacte. Par exemple la table suivante a pour statistique de Kendall:

| | | |
|----|-------|---|
| | a | b |
| c1 | 1 3 4 | |
| c2 | 3 0 3 | |
| | 4 3 7 | |

$S = -9$; correction = 3.5; $\sigma = 4.9$; $z = 1.1227$; $p = 0.131$

Une démarche exacte nous conduit à considérer les 4 tables possibles (qui ont mêmes totaux marginaux) :

| | | | | |
|---|-----|-----|-----|-----|
| | 1 3 | 2 2 | 3 1 | 4 0 |
| | 3 0 | 2 1 | 1 2 | 0 3 |
| S | -9 | -2 | 5 | 12 |

Il apparaît donc une distribution du S à 4 valeurs (distribution 1 1 1 1) et donc une p-value $p=1/4$.

Toutefois, cette démarche, parfois rencontrée ne prend en compte les permutations internes dans les groupes et donne une valeur exagérée de p. Ainsi, il y aurait 4 manières possibles de choisir l'élément figurant dans la case supérieure gauche du premier tableau. La distribution exacte du S serait donc plutôt:

| | | | | |
|------------------|-----|----------|----------|-----|
| | 1 3 | 2 2 | 3 1 | 4 0 |
| | 3 0 | 2 1 | 1 2 | 0 3 |
| S | -9 | -2 | 5 | 12 |
| Nb permutations: | 4 | $6*3=18$ | $4*3=12$ | 1 |

Ce qui conduit à une valeur de $p=4/35=0.11$. Elle est inférieure à la valeur approchée. Dans ce cas, on a fait l'hypothèse d'une répartition fixe entre c1 et c2 et aléatoire entre a et b.

4.6. Problème de l'unité d'analyse

On a examiné le nombre d'arguments logiques donné par deux groupes à l'aide du tableau ci-dessous qui donne le nombre de sujets avec 2 arguments, 3 arguments, etc.

| | | | | |
|-------|----------|----------|----------|----------|
| | *,*;l*_2 | *,*;l*_3 | *,*;l*_4 | *,*;l*_5 |
| f,*,* | 1 | 1 | 0 | 0 |
| g,*,* | 0 | 0 | 1 | 1 |

Pourquoi ne pas avoir plus simplement comptabilisé les arguments logiques par groupe:

| | |
|-------|--------|
| | *,*;l* |
| f,*,* | 5 |
| g,*,* | 9 |

Cela semble permettre de gagner une dimension et de ventiler les résultats selon le types d'arguments:

| | | | |
|-------|--------|--------|--------|
| | *,*;l1 | *,*;l2 | *,*;l3 |
| f,*,* | 2 | 1 | 2 |
| g,*,* | 3 | 3 | 3 |

On notera que l'on a changé d'unité d'analyse, dans le premier cas, il s'agit de la comparaison de deux groupes (et le nombre des cas répertoriés est bien le nombre total de sujets). Dans le deuxième cas, ce sont les arguments qui constituent l'unité d'analyse. Mais la répartition de ces arguments à l'intérieur de ces groupes est perdue. On analysera donc la différence de types d'arguments utilisés, mais on ne pourra rien dire entre les groupes.

4.7. Bugs ou limitations connues

Attention: Lorsque l'on sort du programme, ANASTAT devrait émettre un avertissement si les modifications apportées aux objets n'ont pas été sauvegardées. Mais dans une période de test pensez à sauver vos données avant de commettre l'irréversible. Il faut distinguer à ce

propos le sauvetage des objets et le sauvetage du contenu des éditeurs qui ne contiennent souvent que des données transitoires.

Compilation des filtres et des modèles: Si ANASTAT est capable d'afficher tous les modèles ou filtres sélectionnés, il n'est par contre capable que de les compiler un à un (pour le moment).

Précision des calculs: Pour accélérer la rapidité des calculs, des approximations sont utilisées. Pensez à jeter un coup d'oeil sur des tables (notamment dans le cas du calcul de la probabilité associée au chi-2) pour vérifier si le degré d'approximation vous semble suffisant !

Toutes les remarques, suggestions et questions sont les bienvenues !

5. Bibliographie

Brossard, A. (1990) Regards, interactions sociales et développement cognitif chez l'enfant de 6 à 10 ans dans les épreuves opératoires piagésiennes. *Dossiers de psychologie*, 39. Neuchâtel: Université de Neuchâtel, Séminaire de Psychologie.

Grossen, M. (1988) *L'intersubjectivité en situation de test*. Cousset FR: Editions Delval.

<http://www.lib.umich.edu/chdocs/statistics/spss.html>. *A Guide to Statistical Computing Resources on the Internet*. The University of Michigan, School of Information and Library Studies.

Leach, C. (1979) *Introduction to statistics, a non parametric approach for the social sciences*. New York: John Wiley.

Meddis, R. (1984) *Statistics using ranks, a unified approach*. New York: Basic Blackwell Publisher.

Nicolet, M. (1995) *Dynamiques relationnelles et processus cognitifs. Etude du marquage social chez les enfants de 5 à 9 ans*. Neuchâtel: Delachaux et Niestlé.

(à remplacer) Pochon, L.-O. (1991) *Statistiques et sciences humaines, notes de travail. Dossiers de psychologie*, 38. Neuchâtel: Université de Neuchâtel, Séminaire de Psychologie.

Seidel, J., Friese, S. & Leonard D.C. (1995) *The Ethnograph v4.0: A User's Guide*. Amherst MA: Qualis Research Associates.

Samuel, J. (1994) *Getting started with SPSS for Windows*. Indiana University, UCS, Center for Statistical and Mathematical Computing.

Statsoft (1994) *STATISTICA*. Tulsa OK.

ANASTAT.MAN/août 1991/janvier, juillet 1993/LOP
ANASTA3.MAN/juin 1996/juillet 1997/LOP
ANASTAT3-man2.doc/janvier 2007/LOP
ANASTAT4-man.doc/avril 2008/LOP

Annexe 1: Description des menus²

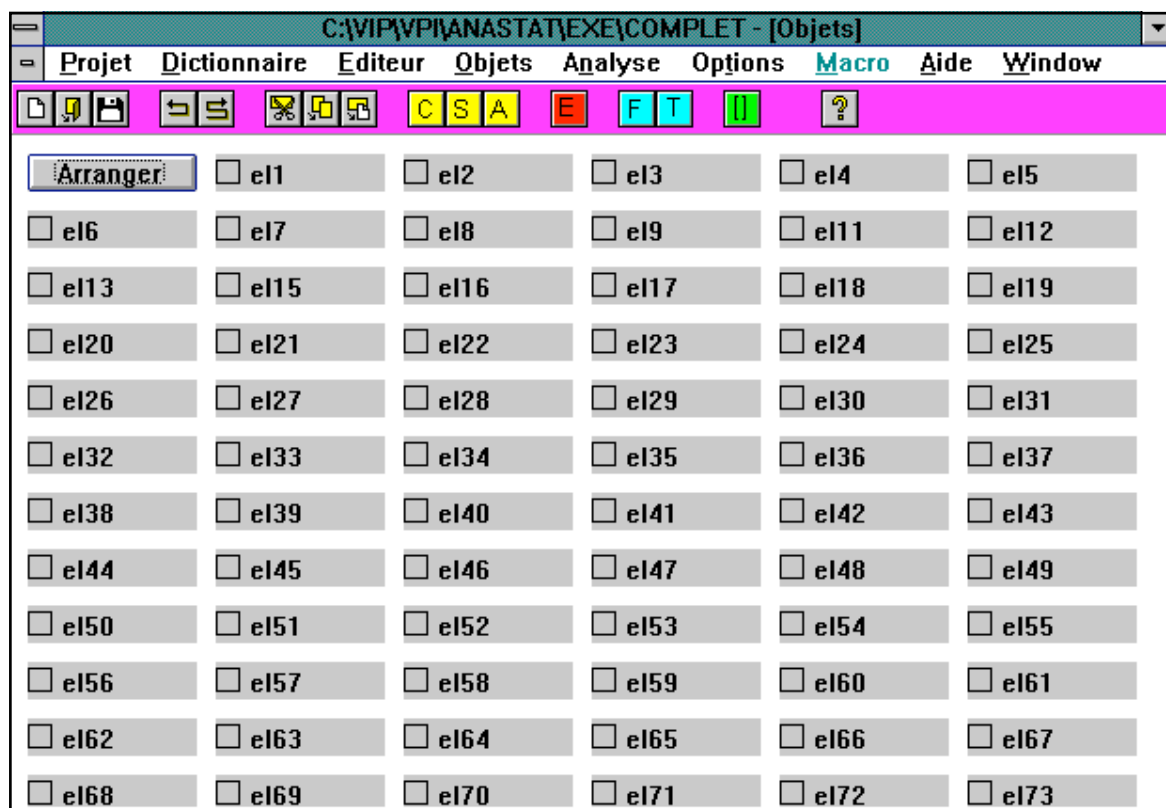


Figure 5: interface de la version 3.1. La fenêtre des objets a été maximisée.

Menu : Projet



*Nouveau** : permet de commencer un nouveau projet.

*Ouvrir**, *Sauver**, *Sauver sous* : permettent de manipuler des fichiers contenant des objets.

Fusionner : permet de charger les objets d'un nouveau fichier sans détruire le travail en cours (contrairement à *Nouveau*).

² Les rubriques marquées d'un astérisque sont également activables à l'aide d'un bouton de la barre d'outils.

La deuxième partie du menu concerne la manipulation de données brutes:

Ouvrir l'éditeur : lorsqu'on veut entrer des nouvelles données qui peuvent aussi être transférées par des opérations de copier/coller.

Charger texte, Sauver texte, Sauver texte sous : permettent de manipuler des fichiers de textes. En particulier, on peut charger des données produites par un autre logiciel (format ascii). Ainsi les données peuvent être conservées de deux manières distinctes : sous la forme d'un texte en mode caractères (dans des fichiers d'extension TXT) et sous la forme structurée (fichiers d'extension APT).

Imprimer : imprime le contenu de l'éditeur actif.

Menu : Dictionnaire

Nouveau, Charger, Fusion, Sauver, Sauver sous : ces rubriques fonctionnent de la même manière que pour le menu projet. Les fichiers destinés aux dictionnaires ont l'extension .DIC.

Afficher, Compiler : permettent d'une part de visualiser un dictionnaire et d'autre part d'intégrer un dictionnaire à partir d'un texte. Combinées, ces deux fonctions permettent d'éditer (modifier) un dictionnaire).

Dans l'éditeur, la syntaxe utilisée pour représenter un dictionnaire est la suivante: tout d'abord le nom d'une catégorie puis une liste des mots ou expressions de la catégorie, séparés par des points virgules et entre accolades. Deux catégories sont séparées par des points-virgules.

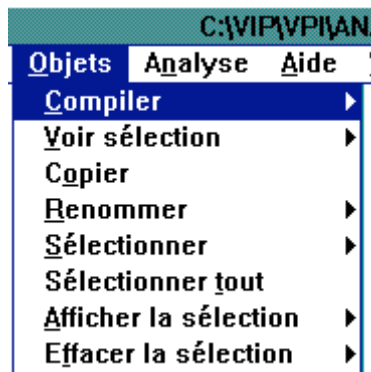
```
math {calcul ; opération ; addition ; nombre } ;  
eval { note ; « travaux écrits » ; carnet ; moyenne } ;  
cogni {penser ; réfléchir ; intelligence } ;  
socio {amis ; camarades ; copains } ;  
systè {branches ; horaire }
```

Dico -> Objet, Objet -> Dico : permet de passer d'une catégorie à un objet (liste de textes) et inversement. Cette fonction permet également d'éditer une catégorie de mots. La façon de décrire les catégories montrent clairement l'équivalence qu'il y a. Toutefois, de façon interne, les objets ne sont pas stockés de la même manière qu'un ensemble de mots.

Menu : Editeur

Dans ce menu, on trouve les rubriques habituelles des éditeurs: copier/couper/coller, etc.

Menu : Objets



*Compiler/Objets standard** : cette opération va transformer les données brutes de l'éditeur actif en objets. Il faut tout d'abord désigner le type d'objet: vecteur, tableau ou boîte. Une quatrième possibilité existe, auto, qui permet de transformer, en un seul passage, plusieurs objets. Dans ce cas, les données brutes doivent obéir à la syntaxe du tableau 11. Le cas le plus simple est une liste d'objets simples. Dans ce cas, chaque objet doit être précédé de son type, éventuellement de son nom. Ils sont séparés par des points-virgules (;).

Compiler/Objets selon modèle : dans ce cas l'objet est contrôlé en fonction d'un modèle, ceci pour vérifier l'exactitudes de données. Cela peut être utile pour s'assurer que les tableaux ont tous la dimension requise. Cette rubrique est aussi utilisée lorsqu'il s'agit de traiter des données complexes.

Compiler/Modèle : c'est cette opération qui permet de définir des modèles. La syntaxe adoptée est décrite dans les tableaux 13.

Compiler/Filtre : les filtres permettront des dénombrements à partir d'enregistrements. Cette rubrique permet de réaliser des filtres selon une syntaxe qui est décrite dans le tableau 14.

Dans tous les cas de compilation, la compilation prend en compte le texte sélectionné dans l'éditeur actif. Si aucune partie n'a été sélectionnée, tout le texte contenu dans l'éditeur est utilisé.

Information / Objet, Modèle, Filtre, Générale : cette rubrique indique le nombre d'objets sélectionnés de la famille désignée. Les noms des objets sélectionnés sont affichés dans un éditeur.

Copier / Objet, Modèle, Filtre : pour avoir une copie de l'objet dont on pourra ensuite modifier le nom.

Renommer / Objet, Modèle, Filtre : comme son nom l'indique.

Sélectionner / Objet, Modèle, Filtre* : les objets à disposition sont présentés dans une fenêtre. Cette rubrique permet d'activer la fenêtre correspondante dans laquelle on pourra sélectionner les objets en cliquant sur les cartouches les représentant.

Sélectionner tout : permet de sélectionner tous les objets. Cette opération peut constituer un gain de temps lorsque, en début de traitement, beaucoup de données sont à traiter simultanément (dépouillement de questionnaires).

Désélectionner tout. opération inverse à la précédente.

*Afficher la sélection / Objet: format court** : un nouvel éditeur est créé dans lequel le ou les objets sélectionnés sont mis en forme. Il est possible de sauver cette mise en forme dans un fichier, l'imprimer ou encore par copier/coller le transférer dans un autre logiciel (traitement de texte, par exemple).

Afficher la sélection / Objet: format long : par rapport au cas précédent la mise en page est complétée par des indications permettant une compilation ultérieure en mode automatique. Cette rubrique permet donc de modifier des objets.

Afficher la sélection / Modèle : affiche la liste des modèles sélectionnés dans l'éditeur.

Afficher la sélection / Filtre : même opération pour les filtres.

Afficher objets / Editeur : identique à *Afficher la sélection / Objet: format long*.

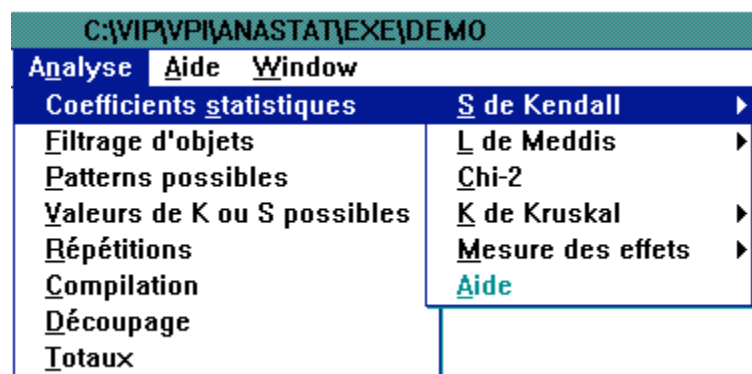
Afficher objets / Imprimer: impression des objets sélectionnés sur l'imprimante (utile puisque l'éditeur est limité à 64 k.B).

Afficher objets / Fichier: mise en forme des objets sélectionnés dans un fichier (peut être repris avec un logiciel de traitement de texte).

Effacer la sélection / Objet, Modèle, Filtre* : les objets sélectionnés de la catégorie choisie sont effacés de la mémoire.

Effacer les objets non sélectionnés : les objets non sélectionnés sont effacés de la mémoire.

Menu : Analyse



Les analyses sont toujours effectuées sur l'ensemble des objets sélectionnés. Les quatre premières rubriques concernent le traitement de données élaborées, par exemple des tableaux.

Coefficients statistiques : calcul de statistiques pour les objets sélectionnés. Les coefficients disponibles sont: S de Kendall, L de Meddis, Chi-2, K de Kruskal (analyse de la variance non paramétrique), W de Wilkerson, Binomial, Mesure des effets (coefficients delta, gamma et de contingence). Une aide (extension future) fera des suggestions en fonction des données à disposition.

Les calculs effectués par les trois rubriques: S de Kendall, L de Meddis, K de Kruskal vont dépendre des données (plan simple ou plan factoriel).

Coefficients seuls : cette rubrique conduit à calculer un coefficient (S, K, etc.) sans présenter tous les détails offerts par la rubrique *Coefficients statistiques*.

K ou S possibles : donne la liste des K ou S de tous les objets soumis aux mêmes contraintes qu'un objet donné (sélectionné).

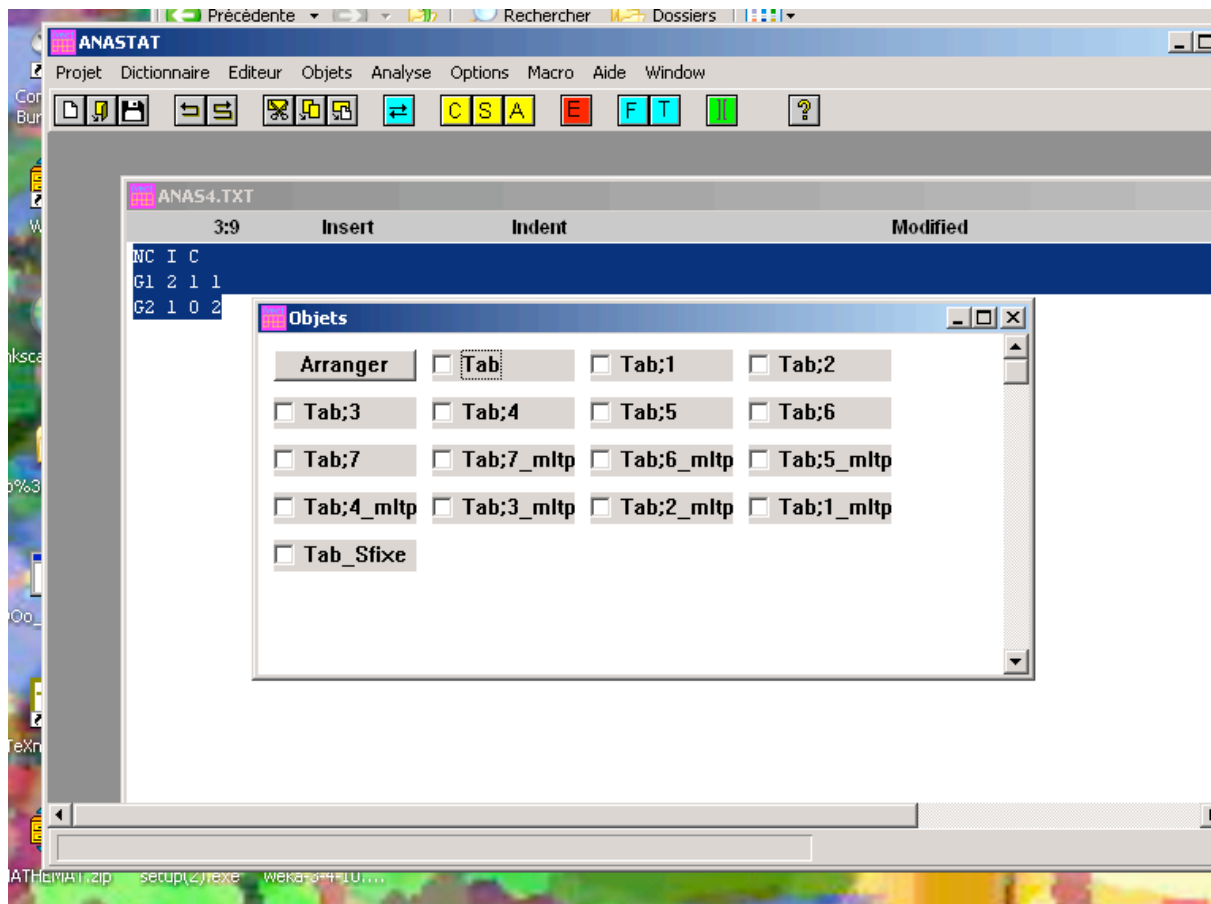
Patterns possibles / Tableaux: étant donné un tableau, le système propose tous les tableaux possibles ayant même somme pour les lignes. Deux options sont possibles : totaux des lignes identiques à l'original (u Fixés) ou non (u Libres). Dans le cas des "boîtes", l'égalité de la somme des lignes peut être imposée pour chaque sous-groupe (u Fixés Strict), simplement pour une facette (u Fixés), seulement pour l'ensemble (u Libres). La rubrique u Libres peut aussi s'utiliser avec un vecteur. Le résultat de l'opération consistera en un tableau dont les sommes des composants de chaque ligne sera égale à la somme des composants du vecteur donné.

Il faut savoir que le nombre de patterns possibles peut être très grand.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|---|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|--|--|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|--|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|--|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|-----|--|--|--|---|---|---|---|---|---|---|---|
| <p>BOITE</p> <table style="margin-left: 20px;"> <tr><td></td><td>a</td><td>b</td><td>c</td></tr> <tr><td>-c1</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>f</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>-c2</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>f</td><td>1</td><td>0</td><td>1</td></tr> </table> | | a | b | c | -c1 | | | | e | 1 | 0 | 1 | f | 1 | 1 | 1 | -c2 | | | | e | 1 | 1 | 1 | f | 1 | 0 | 1 | <p>Avec l'option u Fixés Strict on trouve 5 boîtes présentant les mêmes totaux que la boîte ci-contre. Exemple:</p> <table style="margin-left: 20px;"> <tr><td></td><td>a</td><td>b</td><td>c</td></tr> <tr><td>-c1</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>0</td><td>0</td><td>2</td></tr> <tr><td>f</td><td>2</td><td>1</td><td>0</td></tr> <tr><td>-c2</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>2</td><td>1</td><td>0</td></tr> <tr><td>f</td><td>0</td><td>0</td><td>2</td></tr> </table> <p>Avec l'option u Fixés on en trouve 25. Exemple:</p> <table style="margin-left: 20px;"> <tr><td></td><td>a</td><td>b</td><td>c</td></tr> <tr><td>-c1</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>2</td><td>0</td><td>0</td></tr> <tr><td>f</td><td>0</td><td>1</td><td>2</td></tr> <tr><td>-c2</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>f</td><td>1</td><td>0</td><td>1</td></tr> </table> <p>Avec l'option u Libres, il y en a 149. Exemple:</p> <table style="margin-left: 20px;"> <tr><td></td><td>a</td><td>b</td><td>c</td></tr> <tr><td>-c1</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>0</td><td>0</td><td>2</td></tr> <tr><td>f</td><td>2</td><td>0</td><td>1</td></tr> <tr><td>-c2</td><td></td><td></td><td></td></tr> <tr><td>e</td><td>2</td><td>0</td><td>1</td></tr> <tr><td>f</td><td>0</td><td>2</td><td>0</td></tr> </table> | | a | b | c | -c1 | | | | e | 0 | 0 | 2 | f | 2 | 1 | 0 | -c2 | | | | e | 2 | 1 | 0 | f | 0 | 0 | 2 | | a | b | c | -c1 | | | | e | 2 | 0 | 0 | f | 0 | 1 | 2 | -c2 | | | | e | 1 | 1 | 1 | f | 1 | 0 | 1 | | a | b | c | -c1 | | | | e | 0 | 0 | 2 | f | 2 | 0 | 1 | -c2 | | | | e | 2 | 0 | 1 | f | 0 | 2 | 0 |
| | a | b | c | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 1 | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 1 | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | a | b | c | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 0 | 0 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 2 | 1 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 2 | 1 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 0 | 0 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | a | b | c | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 2 | 0 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 0 | 1 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 1 | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 1 | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | a | b | c | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 0 | 0 | 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 2 | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| -c2 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| e | 2 | 0 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| f | 0 | 2 | 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Patterns possibles / Multiplicité: étant donné un tableau ou plusieurs tableaux, le système en calcule la multiplicité (voir XX).

La figure ci-dessous montre les éléments nécessaires au calcul d'un S exact : le tableau original (*Tab*), les patterns possibles ; les pattern accompagnés de leur multiplicité et les valeurs de S calculées séparément. Elle suivie d'un affichage des différents objets.



| | |
|---|---|
| <pre>"Tab;7_mntp" (TABLEAU NC I C G1 3 1 0 G2 0 0 3, 1);</pre> | <pre>"Tab;3_mntp" (TABLEAU NC I C G1 1 1 2 G2 2 0 1, 9);</pre> |
| <pre>"Tab;6_mntp" (TABLEAU NC I C G1 3 0 1 G2 0 1 2, 3);</pre> | <pre>"Tab;2_mntp" (TABLEAU NC I C G1 1 0 3 G2 2 1 0, 3);</pre> |
| <pre>"Tab;5_mntp" (TABLEAU NC I C G1 2 1 1 G2 1 0 2, 9);</pre> | <pre>"Tab;1_mntp" (TABLEAU NC I C G1 0 1 3 G2 3 0 0, 1);</pre> |
| <pre>"Tab;4_mntp" (TABLEAU</pre> | <pre>Tab_Sfixe VECTEUR</pre> |

| | | | | | | | | | |
|----|----|---|--------|-----|----|----|---|---|---|
| | NC | I | C | -12 | -8 | -4 | 0 | 4 | 8 |
| G1 | 2 | 0 | 2 | | 12 | | | | |
| G2 | 1 | 1 | 1, 9); | | | | | | |

Patterns possibles / Objets complexes : la même possibilité est offerte pour tous les types d'objets. Cette génération se fait à partir d'un modèle nécessairement appelé "germe" et d'un objet nécessairement appelé "var2. Chaque élément terminal du modèle est accompagné du nom d'une variable. Quant à l'objet, il est constitué d'une liste de couples. Le premier élément de chaque couple est le nom d'une variable. L'autre élément est la liste des valeurs. Des règles permettent de préciser des liens entre les variables. Elles sont contenues dans l'objet "règle" constitué d'une liste: [LVI ; LIEN1 ; LIEN2 ; ...] où LVI est la liste des variables indépendantes. LIENi peut avoir deux formes: [[VAR1 ; VAR2 ; ...] ; VAR ; R ; COEFF] signifie que les variables VARi (nécessairement ordinales) ont un lien avec VAR de coefficient R (compris entre -1 et 1). Ce lien est établi avec une certitude COEFF (compris entre 0 et 100). La deuxième forme est: [(VAR1,VAL1) ; (VAR2,VAL2); ...] ; (VAR,VAL) ; COEFF]. Elle signifie que si la valeur de VARi est VALi alors la valeur de VAR est VAL avec une certitude mesurée par COEFF. Par exemple:

germe : [TEXTE : var1 ; ENTIER : var2 ; ENTIER : var3] ;
var : [(var1 , { a ; b ; c ; d }) ; (var2 , { 1 ; 2 ; 3 ; 4 }) ; (var3 , { 1 ; 2 ; 3 ; 4 })] ;
règle: [[var1 ; var2] ; [[var1] ; var3 ; 0.5 ; 50] ; [[var1 ; var2] ; var3 ; -0.3 ; 50]] .

Avec ces informations, ANASTAT va créer une famille d'objets (a , 1, 2) ; (d , 2, 3) ; etc. ;

Patterns possibles / Jeux de données : cette rubrique permet de générer des données fictives sans variable indépendante ou bien avec 1 ou 2 variables indépendantes.

Le choix "1 variable indépendante" conduit au formulaire suivant:

Le premier groupe précise le type de la variable indépendante et ses valeurs. Le deuxième fait de même pour la variable dépendante. Le groupe « données » indique le nom des données et leur nombre. Le groupe suivant, « Effet », permet de favoriser une liaison entre les données (R signifie random, sinon cet effet est donné par un nombre compris entre -1 et 1). Quant au dernier groupe, il indique si les deux variables sont compatibles (par exemple: résultat de pré-test, et post-test) auquel cas une valeur de coupure peut être indiquée.

Les boutons ont les actions suivantes: « OK » enregistre le formulaire, « Générer » procède à la génération des données et « Analyser » propose une analyse des données générées.

Pattern possibles / Initialiser un nouveau formulaire : en cas de modification des données brutes, cette rubrique permet de recalculer un formulaire adapté aux nouvelles données.

La rubrique suivante permet de modifier des données brutes.

Traitement : cette rubrique permet des calculs simples, selon un traitement classique de variables, dans le cas d'objets peu élaborés simples, couples ou listes. Il est possible d'ajouter une constante à une variable, de multiplier une variable par une constante, de prendre le logarithme et de recoder une variable en fonction d'une longueur d'intervalle donnée. La version 4.02 ajoute la possibilité de calculer l'entropie à partir d'un vecteur ou d'une liste d'effectifs.

Les trois rubriques suivantes permettent principalement le passage de données brutes à des données plus élaborées, par exemple la création de tableaux.

*Filtrage d'objets** : cette rubrique permet de procéder à des dénombrements à partir de données brutes ou de constituer des tableaux à double entrée ou des boîtes. Si un seul filtre est sélectionné, il s'agit de dénombrement. Si deux filtres sont sélectionnés, c'est un tableau à double entrée qui va être réalisé. Trois filtres vont conduire à la création d'une boîte.

Répétitions : permet d'obtenir la fréquence d'apparition de chaque élément d'un vecteur, d'une liste de nombres ou d'une liste de symbole.

compilation : permet de passer d'un objet non structuré (texte) en un objet structuré. Cette rubrique est avantageusement remplacée par les éditeurs. C'est un vestige de l'ancienne version.

Les rubriques suivantes permettent de décomposer des données.

Découpage : chaque élément d'une liste devient un objet à part entière. Le même résultat peut s'obtenir en extrayant les items les uns après les autres.

Extraire / Premier élément : l'extraction permet de travailler plus aisément sur une partie des enregistrements seulement. Le premier élément d'un couple d'une liste est clairement défini. Pour un vecteur ou un tableau, il s'agit de la première ligne. Pour une boîte du premier tableau.

Extraire / Reste : Ce qui reste lorsqu'on a ôté le premier élément.

Extraire / Un item : l'ordre de l'élément à extraire est demandé. La valeur 1 correspond naturellement à premier. L'item 0, pour les vecteurs, tableaux et boîtes, dénote la liste des

étiquettes des colonnes. L'item -1 correspond à Reste. Exception, pour un vecteur, l'item n+1 où n est la longueur du vecteur, va donner la liste des composantes du vecteur.

Extraire / Sauf un item : Ce qu'il reste lorsqu'on a ôté l'item mentionné.

Les dernières rubriques traitent des données de façon globale (principalement des données élaborées).

Tri : les vecteurs, tableaux et boîtes sont réordonnés en fonction de l'ordre des étiquettes. Pour les listes, cette commande met les éléments d'une liste par ordre alphabétique ou par ordre croissant.

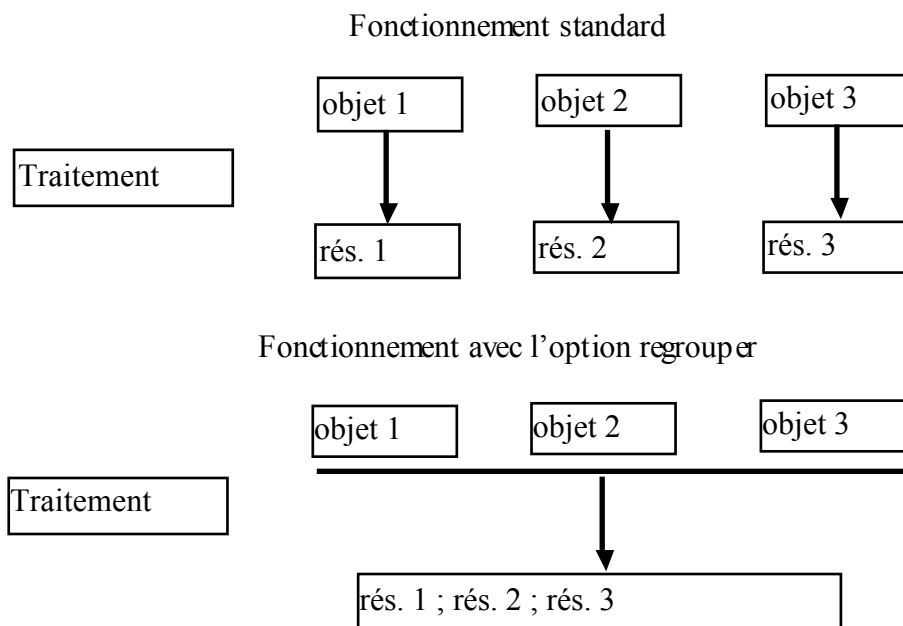
Calcul sur tableau : permet de faire la somme des lignes, des colonnes ou des plaques ou encore de transposer les tableaux (permutation des lignes et colonnes).

Dimension : donne la dimension des tableaux et la longueur des listes.

*Total** : fournit les totaux marginaux de vecteurs, de tableaux ou de boîtes.

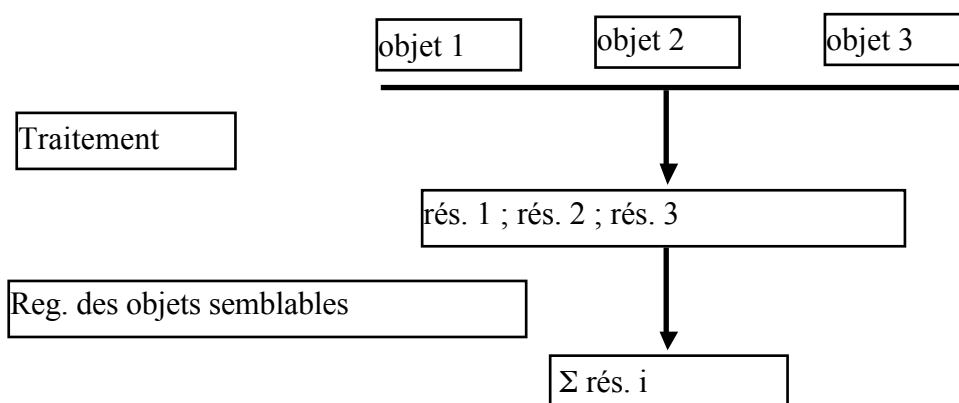
Menu : Options

Regrouper : lorsque cette rubrique est activée, l'ensemble des résultats des opérations menées sur un ensemble d'objets sont regroupés dans une liste.



Regrouper et résumer : cette option procède tout d'abord de la même façon que la précédente. Par la suite les éléments du même type sont regroupés.

Fonctionnement avec Regrouper et résumer



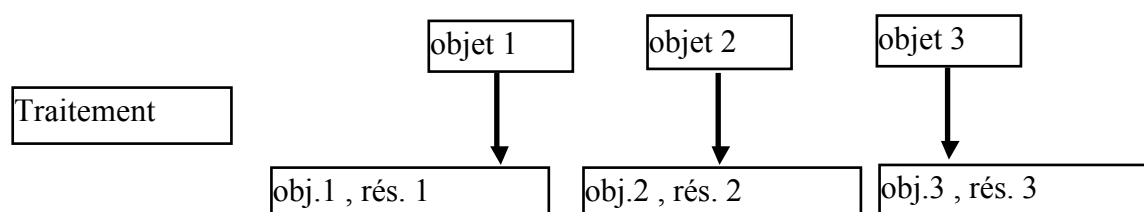
Précision : les nombres décimaux sont présentés avec quelques chiffres significatifs seulement.

Contrôler les noms : chaque fois que le système enregistre un nouvel objet, le nom est revu par l'utilisateur.

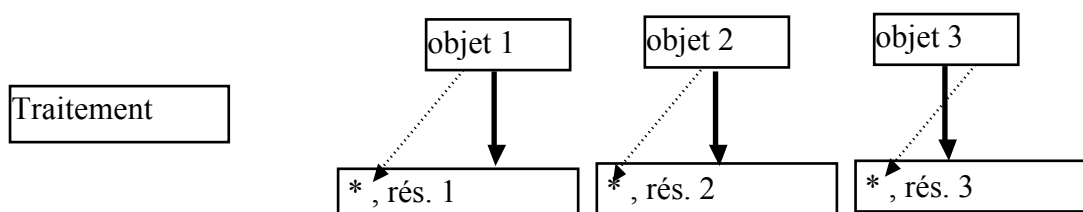
Compléter les objets : Lorsque cette rubrique est activée, l'objet et le résultat de l'opération sont couplés.

Insertion réelle : cette rubrique est activée seulement si la précédente l'est. En cas d'insertion réelle une copie de l'objet initial est effectuée (on peut donc l'effacer). Sinon, l'objet initial n'est qu'en référence.

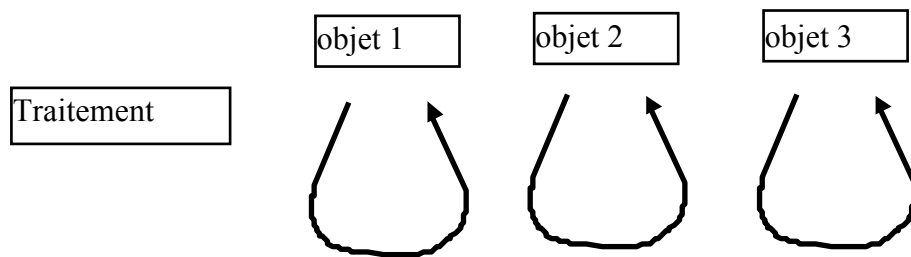
Insertion réelle



Insertion « virtuelle »



Remplacement des objets : lorsque cette option est activée, le résultat d'un traitement se substitue à l'objet qui a subi ce traitement.



*Oter tous les regroupements** : cette option est aussi activable à l'aide d'un bouton. Elle désactive tout type de regroupements, opération que l'on oublie souvent d'effectuer !

Menu : Macro

Ce menu se comprend de lui-même, cinq rubriques (*Nouveau, Charger, Sauver, Sauver sous, Fusionner*) concernent la gestion des fichiers contenant les macro-commandes. Trois rubriques (*Editer, Compiler, Effacer*) permettent de modifier, créer et effacer des macro-commandes. *Exécuter* permet d'utiliser une macro-commande. Quant à la rubrique *Enregistrer*, elle permet de créer une macro automatiquement en gardant la trace des opérations effectuées de façon directe.

Annexe 2: Fichiers d'exemples

demo0.apt: contient simplement un tableau et une boîte qu'il est possible de sélectionner, d'afficher, éventuellement de modifier et de compiler.

demo1.apt: contient 4 objets, 4 filtres et 2 modèles (non prédéfinis). Les objets e1, e2, e3, e4 sont construits selon le modèle 'record' qui lui-même utilise le modèle 'liste'. Il est possible de sélectionner les 4 objets et un filtre (ou deux filtres) et de procéder à une opération de filtrage. L'effet des options 'Regrouper' et 'Regrouper et résumer' peut-être observé sur cet exemple.

demo2.apt: ce fichier est à utiliser avec le dictionnaire demo.dic. On peut observer l'effet du filtrage des 6 objets avec le filtre 'math' (l'option 'Rgrouper et résumer' étant activée) selon que le dictionnaire est chargé ou non. Utiliser la macro 'vocabulaire' de demo0.mac.

demo3.apt: le tableau sert à des macros.

demo4.apt: jeu de données destiné aux macros de demo4.mac.

demo5.apt: contient un objet 'var', un objet 'règle' et un modèle 'germe' permettant de créer un jeu de données.

demo6.apt: contient un formulaire permettant de créer un plan expérimental factoriel et trois filtres pour organiser ces données en une boîte.

demo7.apt :

standard.apt :

demo0.mac: quelques macros utiles.

mcnemar: effectue un test de McNemar sur un tableau 2x2.

stest: effectuer un test des signes sur une liste ou un vecteur de données.

vocabulaire: travaille sur les données de demo2.apt.

sexexlogique: travaille sur la base des données de demo1.apt. Elle établit le nombre d'arguments logiques utilisés selon le sexe du sujet.

s: travailler à partir du tableau contenu dans demo3.apt. Elle établit la distribution du S en cherchant tous les tableaux avec les mêmes totaux marginaux.

kendall: cette macro travaille à partir d'un l'objet contenu dans demo0.apt. Sa seule fonction est de procéder à l'application du test de kendall.

demo4.mac: ces macros travaillent avec les données de demo4.apt (qui sont chargées automatiquement). Elles permettent de comparer des résultats du style « pré-test, post-test » avec le test de Wilcoxon (hypothèse de progrès) ou avec un test basé sur le S (hypothèse d'équivalence).

wilcox: procède au test de Wilcoxon.

kendall: produit un tableau sur lequel il faut effectuer un test de Kendall après y avoir remis de l'ordre.

effet: pour calculer l'influence d'une condition expérimentale calcul le tableau croisé des progrès. A faire un test de Kendall sur le résultat après avoir remis en ordre.

sous: effectue la différence de l'item 3 et de l'item 2 (utilisée dans les autres macros).

demo.dic : voir demo2.apt