

ADONIS
A DYNAMIC APL INTERACTIVE PACKAGE
FOR MULTIVARIATE ANALYSIS

Josiane Confais
I.S.U.P.-B.U.R.O
Tour 45-55, 4 Place Jussieu
75230, Paris Cedex 05, France

Sylviane Bonnefous
Jean Brenot
Jean-Pierre Pagès
CENFAR
DPr/SEGP/LSEES
BP n°6, 92260, Fontenay-aux-Roses
France

Abstract

ADONIS is a portable data analysis package using the interactive facilities of APL. The system consists of modules which can be easily connected to investigate fastly and progressively the data set. The modules are multivariate analysis oriented and emphasize french practices. The dynamic organization of the system makes it easy to introduce new options. The package is described, a complete documentation is available.

A general overview

ADONIS is an APL statistical package devoted to multivariate analysis ; it is written in VSAPL (under the VSPC, IBM system). It gives a better version of some functions already published in 1976 and 1978 (13,4) and completes them. It has been presented for the first time in Compstat 80 (6). There are several APL statistical packages with multivariate analysis ; to our knowledge, Smillie's statpack (15), the IBM advanced statistical library (10), Anscombe's programs (1) and Lebart's ones (12). Among them, only the IBM-library uses the conversational facility of APL. In ADONIS we take full advantage of the conversational mode : in a session, the man-machine dialogue has been particularly studied and is expressed in natural french language. The computer "talks" much more than the user who has to

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

select among options by replying "oui", "non" and sometimes by giving numbers and "S.O.S." when it is "foggy" ; to break out properly, many exists are available.

The modular structure of ADONIS allows extensions and modifications without difficulties.

The data set which is analysed by a function of ADONIS can be extracted from a data base. But ADONIS has no function for creating, editing, listing, updating a data file. To do so, one has to know a "bit" of APL. ADONIS does not use a file-system and it is therefore less machine dependent ; it is really efficient for data sets which are not too big : for example, in a 256 k bytes workspace, it is possible to analyse by the principal components technique a 300 cases x 20 variables data matrix.

Using ADONIS is easy and makes informatics quite invisible ; we think that the package can be activated after an extremely short training.

ADONIS : reflection of statistical practices.

Many multivariate techniques and computer programs have been developed by french statisticians during these these last twenty years (2,3,7,11,12,14) ; the basic methods are well known : principal components, canonical, discriminant, correspondence and regression analysis in the multivariate analysis domain and cluster analysis, either hierarchical or not. Nevertheless, the french practices are more original and this is widely taken into account in the package. With qualitative data, we prefer, in the exploratory phase, to describe the data set with a multivariate technique rather than to test log-linear models. For these data,

correspondence analysis -"a neglected multivariate method"- (Hill (9)) is quite useful and broadly used in France through J.P. BENZECRI's impulse (2). This technique gives instructive graphs and results for two-way contingency tables and its multi-way generalization allows to study 0-1 data and tables generated by cross-tabulation of more than two variables (8).

Also, we specially care about the observations (or cases) in the whole descriptive process : in plots, each case appears with its own name, an alphanumeric symbol ; coefficients are associated with the plots, such as cosines, inertia index, distances to the factorial planes,...

Moreover, every time we can, we use the "extra-points" technique. We obtain then, better interpretation firstly by correlating the factors with other variables, secondly by projecting onto the factorial planes either new observations, continuous variables and nominal or ordinal variables, the categories of which appear on the plots as group means.

ADONIS : structure

For each multivariate analysis, five modules (or groups of functions) are available :

- 1 - Input of the active data set
- 2 - Unidimensional description of this data set :

summary statistics, histograms, and transformations on the variables when they are needed.

- 3 - Multivariate analysis :
eigenvectors are found either by Jacobi's method or by the deflation one ; plots of observations and variables are superimposed : observations and variables appear with their names ; a "zoom" facility is offered by deleting points on these plots : numerical outputs concern both the observations and the variables : all of these outputs are optional ; factor's scores and principal components are stored for any further handling, the extra-points technique for example.
- 4 - Input of the supplementary data set (also noted the passive data)
- 5 - Supplementary data analysis :
extra-points (cases and/or variables) can be projected onto the factor's spaces.

Usually, these modules are linked but each one is self contained.

Figure 1 is a flow chart of five modules used to realize completely a discriminant analysis.

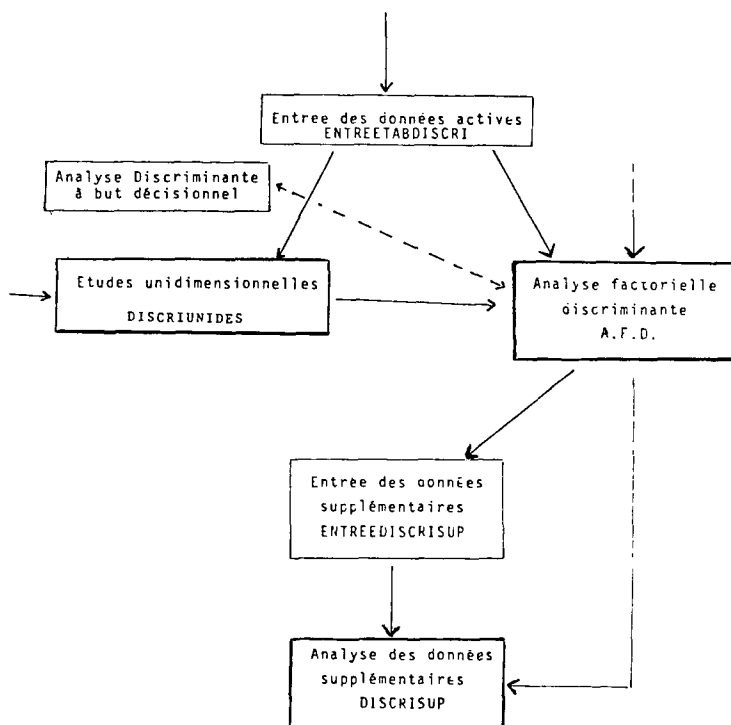


Figure 1

Content and description of ADONIS

Actually, four basic multidimensional techniques are available in ADONIS : principal components, correspondence, discriminant analysis and the linear model. Canonical analysis, forward and stepwise regression, MDSCAL (Kruskal's method), multiple correspondence analysis and hierarchical clustering are not yet included. For each one, a workspace contains the five modules. These workspaces are titled ADONIS1 through ADONIS8. They are widely documented in the "user's manuals" (5).

In the user's manual concerned by a technique one finds :

- a theoretical presentation of the technique with the usual formulas,
- a short description of its modules and their tree-structure,
- a detailed description of each module,
- an example with its complete output.

There is also a workspace titled ADONIS0 which is devoted to the internal documentation of ADONIS and which contains some elementary functions or groups :

- the LISTROG-function gives interactively a list of the available functions by type (plot, probability, linear algebra...), where they are and what they do. If the function FONC for example call some others, then a group is created and named FONCGRP ; in the output of LISTPROG a such function FONC is located by the symbol Δ and a conversational function is pointed out by a C symbol. A partial output is given in figure 2 in Appendix.
- a global description of the package is also given by the LIST function. An output is given in figure 3 in Appendix.

Examples

The first example, figure 4 in Appendix, gives the "man-machine" dialogue for the data set input ; the second example, figure 5 in Appendix is concerned by a small part of the uni-dimensional description module, that is the histograms' phase.

Bibliography

- (1) Anscombe F.J. (1973), ASP : A statistical package in APL, Technical report n° 29, Yale University, New Haven, Conn. 06520.
- (2) Benzecri J.P. (1973), L'Analyse des Données, Tmes 1 et 2, Dunod, Paris.
- (3) Bertier P., Bouroche J.M. (1975), Analyse des Données Multidimensionnelles, PUF, Paris.
- (4) Bonnefous S., Cloutier E., Roy E., (1978), Analyse Multidimensionnelle des Proximités : techniques, programmes APL, Exemples, Rapport CEA R-4910, C.E.N. Saclay, B.P. n° 2, 91190, Gif-sur-Yvette.
- (5) Bonnefous S., Confais J., ADONIS : Ed. 1981 Bibliothèque APL d'Analyse de Données Multidimensionnelles (Tome 0: Introduction, Tome 1: Analyse en composantes principales, Tome 2: Analyse factorielle discriminante, Tome 3: Analyse factorielle des correspondances, Tome 4: Modèle linéaire).
- (6) Bonnefous S., Brenot J., Confais J., Pagès J.P., ADONIS : an APL interactive package for data analysis. Compstat 80, proceedings in computational statistics, physica-verlag-wien.
- (7) Cailliez F., Pagès J.P., (1976), Introduction à l'Analyse des Données, S.M.A.S.H., Paris.
- (8) Cazes P., Baumerder A., et al. (1977), Codage et analyse des tableaux logiques ; Introduction à la pratique des variables qualitatives, Cahier du B.U.R.O. n° 27, Université Pierre et Marie Curie, Paris.
- (9) Hill M.O. (1974), Correspondence Analysis : a neglected multivariate method, Applied Statistics, vol. 23, p. 340-354.
- (10) I.B.M. (1978), APL Advanced Statistical Library, Program Description/Operations Manual - SH20 - 1948-0.
- (11) Lebart L., Morineau A., Tabard N., (1977), Techniques de la Description Statistique : méthodes et logiciels pour l'analyse des grands tableaux, Dunod, Paris.
- (12) Lebart L., Morineau A., Fenelon J.P., (1979), Traitement des Données Statistiques ; méthodes et programmes, Dunod, Paris.
- (13) Mailles J.P., Mailles D., Bonnefous S., (1976), Analyse de Données et APL : techniques, bibliothèque, exemples, Rapport CEA R-4753, C.E.N. Saclay, B.P. n° 2, 91190, Gif-sur-Yvette.
- (14) Saporta G., (1978), Théories et Méthodes de la Statistique, Technip, Paris.
- (15) Smillie K.W., (1969), Statpack 2 : an APL Statistical Package, University of Alberta, Edmonton.

Appendix

LISTPHOG

PROGRAMMES DISPONIBLES SOUS ADONIS
.....

*****MISE EN PAGE :

N° DE RUBRIQUE OU '?' : 7

POUR OBTENIR LA LISTE DES FONCTIONS RELATIVES AUX RUBRIQUES :

+ UTILITAIRES	Tapez 1
+ EDITION	Tapez 2
+ ALGEBRE LINEAIRE	Tapez 3
+ STATISTIQUES DESCRIPTIVE	Tapez 4
+ STATISTIQUE MATHEMATIQUE	Tapez 5
+ PROBABILITES	Tapez 6
+ ANALYSE DE DONNEES	Tapez 7
+ TRACES GRAPHIQUES	Tapez 8

POUR ARRETER

Tapez +

DANS CES LISTES UNE FONCTION :

- CONVERSATIONNELLE EST REPERE PAR LE SYMBOLE C
- POUR LAQUELLE UN GROUPE EXISTE EST REPEREE PAR LE SYMBOLE A
(SI LA FONCTION EST NOM ,LE GROUPE EST NOMGRP)

*****MISE EN PAGE :

N° DE RUBRIQUE OU '?' : 3

FONCTIONS D'ALGEBRE LINEAIRE (17 JUILLET 1981)

SYNTAXE	OBJET	LIEU	N°FICHE
Z+M PUISSANCE N	PUISSANCE N D'UNE MATRICE	ADONIS0	AL01
C+TRIDIAGI X	TRIDIAGONALISATION D'UNE MATRICE SYMETRIQUE	ADONIS0	AL02
C+HSHLDR X	TRIANGULARISATION PAR LA METHODE DE HOUSHOLDER	ADONIS0	AL03
R+DETR A	DETERMINANT D'UNE MATRICE A	ADONIS0	AL04
Y+JCOBI X	DIAGONALISATION PAR LA METHODE DE JACOBI	ADONIS1	AL05
Z+CV DIAG S	RECHERCHE DE CV VECTEURS ET VALEURS PROPRES	ADONIS3	AL06
T+TRIANG A	TRIANGULARISATION PAR LA METHODE DE DOOLITTLE	ADONIS2	AL07

*****MISE EN PAGE :

N° DE RUBRIQUE OU '?' : 1

FONCTIONS UTILITAIRES (17 JUILLET 1981)

SYNTAXE	OBJET	LIEU	N°FICHE
R+N ENTREER1 NC	ENTREE DE N IDENTIFICATEURS ALPHANUMERIQUES	ADONIS3	UTIL01 C A
A+P DP X	CADRES A DROITE SUR C CARACTERES	ADONIS2	UTIL02
A+P DPOI X	PRE MULTIPLICATION DE X PAR LA MATRICE DP	ADONIS2	UTIL03
T+L THANSRANG 'X'	POST MULTIPLICATION DE X PAR LA MATRICE DP	ADONIS3	UTIL04 A
	TRANSFORMATION DE CERTAINES COLONNES DE X		
	PAR LES RANGS		
T+L TRANSLOG 'X'	TRANSFORMATION DE CERTAINES COLONNES DE X	ADONIS3	UTIL05
	PAR LES LOG		
T+L THANSRANGNORM 'X'	TRANSFORMATION DE CERTAINES COLONNES DE X	ADONIS3	UTIL06 A
	PAR UNE TRANSFORMATION ROBUSTE		
Y+CENTREFR X	TABLAEU CENTRE DEDUIT DE X	ADONIS0	UTIL07
Y+CENTRED X	TABLAEU CENTRE REDUIT DEDUIT DE X	ADONIS0	UTIL08
V+BORN N	DETERMINATION DE BORNES DANS UN DECOUPAGE	ADONIS0	UTIL09
C+CENTREC N	DETERMINATION DE CENTRE DE CLASSE	ADONIS0	UTIL10
R+C RECOD X	CODAGE D'UNE VARIABLE QUANTITATIVE	ADONIS0	UTIL11
DESS+CAGD DES	CADRAGE A GAUCHE D'UN TABLEAU ALPHANUMERIQUE	ADONIS0	UTIL12

Figure 2 (partial output)

LIST
*****MISE EN PAGE :

* BIBLIOTHEQUE D'ANALYSE DE DONNEES : ADONIS *

17 JUILLET 1981

	GROUPE	PROGRAMME	TAILLE	LIEU

A.C.P.				

→ENTREE DES DONNEES ACTIVES	ENTREETABACGRP	ENTREETABACP	6K	ADONIS1
→DESCRIPTION UNIDIMENSIONNELLE	UNIDESACGRP	UNIDESACP	17K	ADONIS1
→ANALYSE EN COMPOSANTES PRINCIPALES	ACPGRP	ACP	22K	ADONIS1
→ENTREE DES DONNEES SUPPLEMENTAIRES	ENTREEACPSUPGRP	ENTREEACPSUP	10K	ADONIS1
→ANALYSE DES DONNEES SUPPLEMENTAIRES	SUPACGRP	SUPACP	20K	ADONIS1

A.P.D.				

→ENTREE DES DONNEES ACTIVES	ENTREETABDISCRIGRP	ENTREETABDISCRI	8K	ADONIS2
→DESCRIPTION UNIDIMENSIONNELLE	DISCRIUNIDESGRP	DISCRIUNIDES	20K	ADONIS2
→ANALYSE FACTORIELLE DISCRIMINANTE	APDGRP	APD	32K	ADONIS2
→ENTREE DES DONNEES SUPPLEMENTAIRES	ENTREEDISCRISUPGRP	ENTREEDISCRISUP	10K	ADONIS8
→ANALYSE DES POINTS SUPPLEMENTAIRES	DISCRISUPGRP	DISCRISUP	21K	ADONIS8

A.P.C.				

→ENTREE DES DONNEES ACTIVES	ENTREETABAFGRP	ENTREETABAF	7K	ADONIS3
→ETUDE PRELIMINAIRE	PREPARAFGRP	PREPARAF	10K	ADONIS3
→ANALYSE FACTORIELLE DES CORRESPONDANCES	AFGRP	AF	22K	ADONIS3
→ENTREE DES DONNEES SUPPLEMENTAIRES	ENTREEAFCSUPGRP	ENTREEAFCSUP	8K	ADONIS3
→ANALYSE DES DONNEES SUPPLEMENTAIRES	AFCSUPGRP	AFCSUP	8K	ADONIS3

MODELE LINEAIRE				

→DESCRIPTION BIDIMENSIONNELLE	BIDESGRP	BIDES	11K	ADONIS6
→ETUDE D'UN MODELE LINEAIRE	REGREGRP	REGRE	22K	ADONIS6
→EXPLORATION RAPIDE DE REGRESSIONS	REGPASAPASGRP	REGPASAPAS	20K	ADONIS6

Figure 3

ENTREETABACP

ACP : ENTREE DU TABLEAU DES DONNEES ACTIVES ET DES TABLEAUX D'IDENTIFICATEURS

ENTREZ LE NOM DE VOTRE TABLEAU DES DONNEES QUANTITATIVES

TAB

TAB EST UN TABLEAU A 91 LIGNES ET 12 COLONNES

ENTREZ LE NOM DU TABLEAU DES IDENTIFICATEURS DES LIGNES DE TAB

IDI

ENTREZ LE NOM DU TABLEAU DES IDENTIFICATEURS DES COLONNES DE TAB

IDV

DESIREZ-VOUS L'IMPRESSION DU TABLEAU TAB ?

OUI

	T1	T2	T3	T5	T6	T7	T8	T9	T11	T12	T13	T14
1	9	3	5	7	8	2	5	5	7	4	3	3
2	7	1	5	8	5	2	7	2	5	4	1	6
3	6	8	6	10	10	4	6	3	2	6	7	9
4	5	7	3	6	7	6	6	1	5	10	5	7
5	5	5	9	6	7	2	6	4	5	9	1	2
6	6	3	6	10	10	4	6	6	2	6	4	1
7	5	8	5	10	9	4	9	4	1	8	7	5
8	5	1	5	5	7	5	7	2	7	6	8	4
9	6	2	4	6	5	1	5	4	10	9	5	2
10	4	7	3	5	5	4	6	5	5	9	8	1
[												1
1												1
1												1
1												1
1												1

MODIFICATION DU TABLEAU TAB

VOULEZ-VOUS MODIFIER CERTAINES VALEURS DU TABLEAU DES DONNEES ?
NON

Figure 4

