

**L. Lebart A. Morineau  
J.-P. Fénélon**

**traitement  
des  
données  
statistiques**

**méthodes et programmes**

**dunod**

TP274  
L3

8065734

**Ludovic Lebart**

Maître de Recherche au CNRS  
Professeur à l'ISUP

**Alain Morineau**

Chargé de Recherche au CEPREMAP  
Professeur à l'ISUP

**Jean-Pierre Fénélon**

Chargé de Recherche au CNRS  
Professeur à l'ISUP

**traitement  
des  
données  
statistiques**

**méthodes et programmes**



E8065734

**dunod**

© BORDAS, Paris, 1979  
ISBN 2-04-010787-8

“ Toute représentation ou reproduction, intégrale ou partielle, faite sans le consentement de l'auteur, ou de ses ayants-droit, ou ayants-cause, est illicite (loi du 11 mars 1957, alinéa 1<sup>er</sup> de l'article 40). Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal. La loi du 11 mars 1957 n'autorise, aux termes des alinéas 2 et 3 de l'article 41, que les copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective d'une part, et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration ”

**traitement  
des  
données  
statistiques**

## Avant-propos

Depuis la première parution de l'ouvrage « Statistique et Informatique Appliquées » en 1971 <sup>(1)</sup>, les deux disciplines évoquées dans ce titre ont évolué assez vite. Une refonte totale de la 3<sup>e</sup> édition de l'ouvrage, concernant aussi bien les développements statistiques que les logiciels, s'imposait. Il était également nécessaire d'en compléter le contenu en tenant compte de l'importance croissante de techniques nouvelles (comme par exemple les méthodes de classification) et d'outils nouveaux (comme le langage conversationnel APL).

L'étendue du travail de refonte, le recadrage des sujets, le caractère nouveau de certains développements ont justifié la transformation de la dernière édition en un nouveau manuel. Le présent ouvrage tente donc de dresser un nouveau bilan de la Statistique Appliquée après plusieurs années d'innovation, d'effervescence et de stimulation informatique. Il constitue pour l'essentiel la matière de l'enseignement du CYCLE DE STATISTIQUE APPLIQUÉE <sup>(2)</sup> de l'Institut de Statistique des Universités de Paris.

Le public auquel s'adresse ce manuel est censé posséder les éléments de statistique habituellement acquis dans les premières années de l'enseignement supérieur, et un niveau en mathématique comparable à celui de la licence en Sciences Economiques, ou à celui des Ecoles d'Ingénieurs.

Devant la rapidité de l'évolution des techniques statistiques, il est difficile d'enseigner la statistique, même appliquée, sous forme de recettes, sans que la personne qui utilise les méthodes puisse les critiquer et éventuellement les remettre en question.

Pour bien des disciplines, la statistique est le seul point de contact possible avec la réalité ; le statisticien doit donc savoir ce qu'il fait. L'itinéraire qui le conduit de l'analyse d'une situation concrète à des conclusions pratiques lui impose les conditions suivantes :

1) Il doit avoir de bonnes connaissances de calcul des probabilités, qui peuvent cependant être limitées aux notions et aux résultats ayant des rapports directs avec les applications statistiques. Le chapitre I doit rappeler ou compléter ces bases.

---

<sup>(1)</sup> L. Lebart, J. P. Fénelon. Dunod.

<sup>(2)</sup> Fondé en 1969 à l'initiative de M. Georges Morlat, le Cycle de Statistique Appliquée est destiné à donner en une ou deux années un enseignement complémentaire en statistique à des étudiants déjà avancés dans d'autres disciplines (économie, sciences humaines, médecine, agronomie, sciences de l'ingénieur, etc.) ou à des personnes déjà engagées dans la vie professionnelle (Formation Continue).

2) Il doit connaître les principales méthodes statistiques non comme une juxtaposition de techniques spécifiques, mais à un niveau plus général, de façon à pouvoir choisir, comparer, et éventuellement approfondir. Ici encore, un compromis devait être trouvé entre la généralité et l'accessibilité. Nous avons tranché de façon à assurer une bonne compréhension en vue des applications. Un petit nombre de résultats (signalés explicitement) ne sont pas démontrés lorsque leur portée ne justifie pas des développements volumineux ou difficiles. Les chapitres II (Méthodes non paramétriques), III (Modèle linéaire), IV (Analyse des données, méthodes factorielles) et V (Analyse des données, classification) tentent d'exposer les principales méthodes donnant lieu à des applications.

3) Il doit pouvoir mettre en œuvre ces méthodes, ou dialoguer efficacement avec celui qui les mettra en œuvre. Pour cela il doit connaître un ou plusieurs langages de programmation, et savoir écrire (ou lire) les programmes d'application. C'est pourquoi chaque chapitre est suivi d'une annexe où figurent de tels programmes. Ces programmes sont écrits de façon à rester très près de l'exposé statistique, sans souci d'optimiser l'exploitation. Les programmes optimaux sont en effet toujours écrits pour un type particulier de machine, ou pour un type particulier d'application (les spécialistes intéressés par de telles versions savent dans quelles programmathèques ou auprès de quelle organisation d'utilisateurs les chercher).

Le lecteur trouvera en Annexe Générale un rappel des notions et des résultats d'algèbre linéaire et de programmation nécessaires à la compréhension de l'ouvrage. (Il pourra trouver de véritables exposés pédagogiques de ces deux matières dans les manuels spécialisés.) On a également précisé dans cette annexe les options adoptées pour la rédaction des programmes, et rassemblé certains programmes intéressant plusieurs chapitres.

La programmation APL a été réalisée avec le concours et les conseils de Josiane Confais, professeur et responsable de l'enseignement de l'APL à l'Institut de Statistique (ISUP).

Hormis les exemples traités dans chaque chapitre, ce manuel ne comporte pas d'exercices d'application. On trouvera quelque 200 textes (et leurs corrigés détaillés) d'exercices de statistique et d'informatique correspondant au contenu de ce manuel, dans l'ouvrage publié par l'équipe des enseignants du Cycle de Statistique Appliquée sous le pseudonyme de Ronald Cehessat <sup>(1)</sup>.

Ce travail, réalisé au CREDOC <sup>(2)</sup> et au CEPREMAP <sup>(3)</sup>, avec à l'origine l'appui bienveillant de MM. Edmond Lisle et Georges Rösch, n'aurait pu être effectué sans le soutien amical et la collaboration de Nicole Tabard, de

---

<sup>(1)</sup> Ronald Cehessat (1976) : *Exercices commentés de Statistique et Informatique Appliquées* ; Dunod.

<sup>(2)</sup> Centre de Recherche pour l'Etude et l'Observation des Conditions de Vie.

<sup>(3)</sup> Centre d'Etudes Prospectives d'Economie Mathématique Appliquées à la Planification.

Marie-Odile Lebeaux, ni sans relations étroites avec le laboratoire du Professeur Jean-Paul Benzecri.

Le CREDOC et le CEPREMAP sont actuellement dirigés par MM. A. Babeau et C. Fourgeaud que nous sommes heureux de remercier ici.

La frappe du manuscrit a été réalisée par Jacqueline Jean, à qui nous adressons nos plus vifs remerciements.

LL, AM, JPF.

*Juillet 1979*



# Index des 70 programmes FORTRAN

|       | notice | listage |
|-------|--------|---------|
| AIDAC | 362    | 371     |
| AMPEG | 81     | 90      |
| ANACO | 360    | 367     |
| ANACP | 358    | 364     |
| ARMIN | 407    | 413     |
| AVAR2 | 252    | 261     |
| BINOM | 83     | 93      |
| BOFIX | 81     | 90      |
| BORNS | 430    | 445     |
| BOXMU | 84     | 93      |
| CALFC | 361    | 368     |
| CALFP | 359    | 366     |
| CHLON | 428    | 439     |
| CHLSK | 424    | 434     |
| CLASH | 408    | 413     |
| CLAS4 | 410    | 416     |
| COMAJ | 421    | 431     |
| COMBI | 83     | 92      |
| COORC | 362    | 370     |
| COORP | 360    | 367     |
| CORRL | 80     | 89      |
| CRUSK | 83     | 92      |
| DEWAW | 253    | 263     |
| DFISH | 166    | 172     |
| DISTK | 409    | 414     |
| DONNI | 250    | 258     |
| DONNK | 408    | 414     |
| DPLOI | 251    | 258     |
| DTRIG | 409    | 415     |
| EDCHA | 412    | 418     |
| EDITR | 252    | 260     |
| EDIVP | 358    | 364     |
| EFEGO | 81     | 90      |
| EPUR4 | 429    | 444     |
| FISHM | 167    | 173     |

|       | notice | listage |
|-------|--------|---------|
| FISHS | 168    | 175     |
| FISHW | 168    | 174     |
| FTINV | 82     | 92      |
| GPLAN | 429    | 441     |
| GRAPH | 428    | 440     |
| HISTV | 79     | 87      |
| IGSDP | 425    | 435     |
| KLICO | 428    | 438     |
| LIRAF | 357    | 363     |
| LISTP | 427    | 437     |
| MLCOV | 251    | 260     |
| MOMER | 80     | 88      |
| MOME2 | 80     | 88      |
| PERMX | 427    | 436     |
| PINLG | 82     | 91      |
| POISS | 83     | 93      |
| POUSS | 358    | 363     |
| PREDE | 250    | 257     |
| PREMC | 249    | 257     |
| QTCOR | 86     | 95      |
| QUANT | 79     | 88      |
| RECOD | 81     | 91      |
| REGMC | 248    | 256     |
| RETIR | 85     | 94      |
| RNKSF | 422    | 432     |
| SEN3A | 423    | 433     |
| SHELK | 426    | 436     |
| SIMFM | 169    | 175     |
| SIMFS | 170    | 177     |
| SIMFW | 169    | 176     |
| SIMUR | 84     | 94      |
| STROB | 85     | 94      |
| TADEP | 80     | 89      |
| TESTF | 166    | 171     |
| VVPRO | 424    | 433     |

## Index des 52 fonctions APL

|       | notice | listage |       | notice | listage |
|-------|--------|---------|-------|--------|---------|
| AMPEG | 99     | 105     | FTINV | 99     | 105     |
| ANACO | 374    | 378     | HISTV | 96     | 103     |
| ANACP | 373    | 376     | ITERM | 179    | 181     |
| ANAVA | 265    | 269     | JCOBI | 448    | 452     |
| ARMIN | 420    | 420     | KLICO | 446    | 449     |
| BINOM | 100    | 106     | LISTP | 448    | 452     |
| BOFIX | 99     | 105     | MLCAL | 264    | 268     |
| BOXMU | 101    | 106     | MLVAR | 266    | 271     |
| CALFC | 375    | 378     | MOYEN | 97     | 103     |
| CALFP | 373    | 376     | PINLG | 99     | 105     |
| COEFR | 97     | 104     | PLANF | 446    | 449     |
| COMAJ | 447    | 451     | POISS | 100    | 106     |
| COORC | 375    | 379     | QTCOR | 102    | 107     |
| COORP | 374    | 376     | QUANT | 97     | 103     |
| CORRL | 98     | 104     | RETIR | 101    | 107     |
| CRUSK | 100    | 106     | RNKSF | 447    | 450     |
| DEWAW | 267    | 272     | SEN2A | 447    | 451     |
| DFUSE | 447    | 450     | SIMFM | 180    | 182     |
| DPLOI | 265    | 269     | SIMFS | 181    | 183     |
| ECTYP | 97     | 103     | SIMFW | 180    | 182     |
| EDITK | 266    | 270     | SIMUR | 101    | 107     |
| EDITR | 265    | 268     | STROB | 101    | 107     |
| EFEGO | 99     | 105     | TADEP | 98     | 104     |
| FISHM | 179    | 181     | UNIFO | 100    | 106     |
| FISHS | 180    | 182     | XXCOR | 98     | 104     |
| FISHW | 180    | 181     | XXCOV | 98     | 104     |

Chez le même éditeur :

**R. CÉHESSAT**

**Exercices commentés de statistique  
et informatique appliquées**

432 pages, 15,5 x 24, broché

ISBN 2-04-001625-2

**L. LEBART, A. MORINEAU, N. TABARD**

**Techniques de la description statistique  
Méthodes et logiciels pour l'analyse des grands tableaux**

360 pages, 15,5 x 24, cartonné

ISBN 2-04-010090-3

**J.-P. BENZÉCRI et collaborateurs**

**L'analyse des données**

**Tome 1 – La taxinomie**

640 pages, 15,5 x 24, 3<sup>e</sup> édition, relié

ISBN 2-04-010891-2

**Tome 2 – L'analyse des correspondances**

624 pages, 15,5 x 24, 3<sup>e</sup> édition, relié

ISBN 2-04-010964-1

**LES CAHIERS DE L'ANALYSE DES DONNÉES**

Revue trimestrielle

Chaque numéro : 128 pages

Direction : Professeur J.-P. Benzécri

ISSN 0339-3097

**Renseignements et abonnements :**

Centrale des Revues, B.P. n° 119

93104 Montreuil Cedex. France

Tél. : 374-12-45

**dunod**

数据处理  
统计学.

63-84

1989年11月 | 7日 统计

# Sommaire

|  |           |
|--|-----------|
| INTRODUCTION .....   | 1         |
| <b>I. ÉLÉMENTS DE LA THÉORIE DES PROBABILITÉS<br/>ET APPLICATIONS STATISTIQUES</b> |           |
| <b>I.1. PROBABILITÉ SUR UN ENSEMBLE FINI .....</b>                                 | <b>5</b>  |
| 1. Probabilité et théorie mathématique .....                                       | 5         |
| 2. Algèbre des événements .....  | 6         |
| 3. Probabilité sur un ensemble fini .....  | 9         |
| 4. Equiprobabilité et dénombrement .....   | 11        |
| 5. Probabilité conditionnelle et indépendance .....                                | 14        |
| <b>I.2. VARIABLES ALÉATOIRES SUR UN ENSEMBLE FINI .....</b>                        | <b>18</b> |
| 1. Définition et propriétés .....  | 18        |
| 2. Couple de variables aléatoires ; indépendance .....                             | 20        |
| 3. Moyennes et moments .....   | 23        |
| 4. Exemples : variable binomiale et vecteur multinomial .....                      | 28        |
| <b>I.3. VARIABLES ALÉATOIRES GÉNÉRALES .....</b>                                   | <b>33</b> |
| 1. Espaces probabilisés généraux .....   | 33        |
| 2. Variables aléatoires générales .....  | 34        |
| 3. Fonction de répartition .....   | 35        |
| 4. Couple de variables aléatoires ; indépendance .....                             | 37        |
| 5. Espérance, variance et covariance .....   | 39        |
| 6. Une distribution discrète : la loi de Poisson .....                             | 40        |
| 7. Une distribution absolument continue : la loi de Laplace-Gauss .....            | 43        |
| 8. Loi de Laplace-Gauss à plusieurs dimensions .....                               | 45        |
| 9. Lois dérivées de la loi de Laplace-Gauss .....                                  | 49        |
| <b>I.4. THÉORÈMES LIMITES .....</b>  | <b>51</b> |
| 1. Inégalité de Bienaymé-Chebychev .....   | 51        |
| 2. Loi faible des grands nombres .....   | 52        |
| 3. Théorème de la limite centrale .....  | 54        |
| <b>I.5. APPLICATIONS AU RAISONNEMENT STATISTIQUE .....</b>                         | <b>62</b> |
| 1. Échantillons ; statistiques .....   | 62        |
| 2. Estimateur ; estimation .....   | 64        |
| 3. Application 1 : Monte Carlo et simulations .....                                | 67        |
| 4. Application 2 : Estimateur de Quenouille-Tukey (Jackknife) .....                | 70        |
| 5. Application 3 : Réponse à une question indiscrete .....                         | 73        |
| <b>A.1. ANNEXE : LOGICIELS POUR LE CHAPITRE I</b>                                  |           |
| 1. Présentation des programmes FORTRAN .....                                       | 77        |
| 2. Présentation des fonctions APL .....  | 96        |



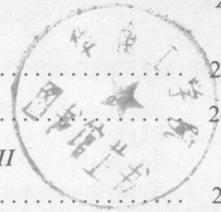
## II. MÉTHODES NON PARAMÉTRIQUES

|   |     |
|---|-----|
| <b>II.1. TESTS FISHÉRIENS</b> .....   | 109 |
| 1. Principe des tests fishériens .....  | 109 |
| 2. Test sensible à un écart des valeurs centrales de deux distributions .....   | 112 |
| 3. Test de symétrie de la distribution d'une différence .....                   | 114 |
| 4. Test d'une dépendance monotone entre deux variables aléatoires .....         | 116 |
| 5. Traitement des grands échantillons .....                                     | 120 |
| 6. Note sur les tirages sans remise .....                                       | 124 |
| <b>II.2. TESTS BINOMIAUX</b> .....  | 127 |
| 1. Test des signes .....  | 127 |
| 2. Applications des tests binomiaux .....                                       | 130 |
| 3. Test multinomial et CHI-2 .....  | 134 |
| 4. Puissance et types d'erreur .....  | 136 |
| <b>II.3. TESTS FISHÉRIENS APPLIQUÉS AUX RANGS</b> .....                         | 138 |
| 1. Test sensible à un écart des valeurs centrales (Wilcoxon-Mann-Whitney) ..... | 138 |
| 2. Symétrie de la distribution d'une différence (Wilcoxon) .....                | 145 |
| 3. Test de dépendance monotone (Spearman) .....                                 | 149 |
| 4. Test de dépendance monotone (Kendall) .....                                  | 152 |
| 5. Procédures attachées aux fonctions de répartition (Kolmogorov-Smirnov) ..... | 157 |
| <i>Notice bibliographique</i> .....   | 163 |
| <b>A.II. ANNEXE : LOGICIELS POUR LE CHAPITRE II</b>                             |     |
| 1. <i>Présentation des programmes FORTRAN</i> .....                             | 165 |
| 2. <i>Présentation des fonctions APL</i> .....                                  | 178 |

## III. MODÈLE LINÉAIRE

|   |     |
|---|-----|
| <b>III.1. AJUSTEMENT PAR LA TECHNIQUE DES MOINDRES-CARRÉS</b> .....     | 185 |
| 1. Définition d'un modèle linéaire .....                                | 185 |
| 2. Calcul et propriétés de l'ajustement des Moindres-Carrés .....       | 188 |
| 3. Calculs particuliers isolant le terme constant .....                 | 192 |
| <b>III.2. LE THÉORÈME DE GAUSS-MARKOV</b> .....                         | 196 |
| 1. Les hypothèses de Gauss-Markov .....                                 | 196 |
| 2. Théorème de Gauss-Markov .....                                       | 198 |
| <b>III.3. LE MODÈLE LINÉAIRE CLASSIQUE</b> .....                        | 201 |
| 1. Hypothèses de Laplace-Gauss et loi des estimateurs .....             | 201 |
| 2. Test sur un sous-ensemble de coefficients .....                      | 208 |
| <b>III.4. ANALYSE DE LA VARIANCE</b> .....                              | 215 |
| 1. Analyse de la variance à un critère .....                            | 215 |
| 2. Analyse de la variance à plusieurs critères (sans interaction) ..... | 223 |
| 3. Analyse de la variance à deux critères avec interaction .....        | 230 |

|   |     |
|---|-----|
| <b>III.5. ANALYSE DE LA COVARIANCE</b> .....                              | 237 |
| <i>Notice bibliographique</i> .....                                       | 245 |
| <b>A.III. ANNEXE : LOGICIELS POUR LE CHAPITRE III</b>                     |     |
| 1. <i>Présentation des programmes FORTRAN</i> .....                       | 247 |
| 2. <i>Présentation des fonctions APL</i> .....                            | 264 |
| <br><b>IV. ANALYSE DES DONNÉES : MÉTHODES FACTORIELLES</b>                |     |
| <b>IV.1. RÉSULTATS GÉNÉRAUX</b> .....                                     | 273 |
| 1. Les différentes méthodes .....   | 273 |
| 2. Analyse générale .....   | 275 |
| <b>IV.2. ANALYSE EN COMPOSANTES PRINCIPALES</b> .....                     | 283 |
| 1. Généralités .....  | 283 |
| 2. Analyse dans $\mathbb{R}^p$ .....                                      | 284 |
| 3. Analyse dans $\mathbb{R}^n$ .....                                      | 286 |
| 4. Exemple numérique et règles d'interprétation .....                     | 287 |
| 5. Variables et individus supplémentaires .....                           | 291 |
| 6. Exemple d'application .....  | 292 |
| <b>IV.3. ANALYSE DES RANGS</b> .....                                      | 295 |
| 1. Tabulation approchée des pourcentages de variance .....                | 295 |
| 2. Exemple d'application .....  | 296 |
| <b>IV.4. ANALYSE DES COVARIANCES ET DES CORRÉLATIONS PARTIELLES</b> ..... | 300 |
| 1. Définition du coefficient de corrélation partielle .....               | 300 |
| 2. Calcul des covariances et corrélations partielles .....                | 302 |
| 3. Analyse du nuage résiduel .....  | 304 |
| <b>IV.5. ANALYSE DES CORRESPONDANCES</b> .....                            | 305 |
| 1. Généralités .....  | 305 |
| 2. Nuage, masses et distances .....                                       | 308 |
| 3. Analyse dans $\mathbb{R}^p$ .....                                      | 310 |
| 4. Analyse dans $\mathbb{R}^n$ .....                                      | 312 |
| 5. Relation entre les deux espaces $\mathbb{R}^p$ et $\mathbb{R}^n$ ..... | 313 |
| 6. Reconstitution du tableau de fréquences .....                          | 314 |
| 7. Positionnement d'éléments supplémentaires .....                        | 315 |
| 8. Aides à l'interprétation .....   | 316 |
| 9. Exemple numérique .....  | 317 |
| 10. Loi limite d'un vecteur multinomial .....                             | 320 |
| 11. Autre présentation de l'analyse des correspondances .....             | 324 |
| 12. Exemples illustratifs .....   | 326 |



|   |     |
|---|-----|
| <b>IV.6. MÉTHODES DIVERSES ET COMPLÉMENTS</b> .....       | 329 |
| 1. Analyse en facteurs communs et spécifiques .....       | 329 |
| 2. Régression orthogonale .....                           | 333 |
| 3. Régression sur composantes principales .....           | 334 |
| 4. Analyse canonique .....                                | 336 |
| 5. Analyse discriminante .....                            | 340 |
| 6. Validité des résultats .....                           | 348 |
| <b>A.IV. ANNEXE : LOGICIELS POUR LE CHAPITRE IV</b> ..... | 355 |
| 1. <i>Présentation des programmes FORTRAN</i> .....       | 355 |
| 2. <i>Présentation des fonctions APL</i> .....            | 373 |

## V. ANALYSE DES DONNÉES : NOTIONS DE CLASSIFICATION AUTOMATIQUE

|  |     |
|--|-----|
| <b>V.1. INTRODUCTION</b> .....   | 381 |
| <b>V.2. LA CLASSIFICATION ASCENDANTE HIÉRARCHIQUE</b> .....                            | 384 |
| 1. Généralités .....   | 384 |
| 2. Classification ascendante selon le saut minimal et arbre de longueur minimale ..... | 388 |
| 3. Agrégation selon la variance et techniques apparentées .....                        | 396 |
| <b>V.3. CLASSIFICATION NON ASCENDANTE</b> .....  | 399 |
| 1. Agrégation autour de centres mobiles .....  | 399 |
| 2. Deux autres méthodes .....  | 404 |
| <i>Note sur les algorithmes et les programmes</i> .....                                | 405 |
| <b>A.V. ANNEXE : LOGICIELS POUR LE CHAPITRE V</b> .....                                | 407 |
| 1. <i>Présentation des programmes FORTRAN</i> .....                                    | 407 |
| 2. <i>Présentation des fonctions APL</i> .....   | 420 |

## ANNEXE GÉNÉRALE : UTILITAIRES GÉNÉRAUX POUR LA STATISTIQUE

|  |     |
|--|-----|
| <i>AG1 : PRÉSENTATION DES PROGRAMMES FORTRAN</i> ..... | 421 |
| <i>AG2 : PRÉSENTATION DES FONCTIONS APL</i> .....      | 446 |

## APPENDICES

|  |     |
|--|-----|
| <b>I. PRÉSENTATION SOMMAIRE DU LANGAGE FORTRAN</b> ..... | 453 |
| 1. Exemple .....   | 453 |
| 2. La définition des données mises en jeu .....          | 455 |
| 3. Le déroulement du programme .....                     | 458 |
| 4. Les sous-programmes .....                             | 460 |
| 5. Les opérations de calcul .....                        | 462 |
| 6. Les entrées-sorties .....                             | 463 |
| 7. La présentation des programmes dans les annexes ..... | 465 |
| 8. La programmation du statisticien .....                | 466 |

|  |     |
|--|-----|
| <b>II. PRÉSENTATION SYNTHÉTIQUE DES OPÉRATEURS APL</b> .....     | 468 |
| 1. Opérateurs monadiques ou dyadiques scalaires .....            | 470 |
| 2. Autres opérateurs monadiques scalaires .....                  | 471 |
| 3. Autres opérateurs dyadiques scalaires .....                   | 471 |
| 4. Opérateurs dyadiques circulaires .....                        | 472 |
| 5. Opérateurs dyadiques logiques .....                           | 472 |
| 6. Opérateurs dyadiques de relation .....                        | 472 |
| 7. Les listes ; opérateurs monadiques .....                      | 473 |
| 8. Les listes ; opérateurs dyadiques .....                       | 474 |
| 9. Les matrices .....  | 477 |
| <b>III. RAPPELS DE CALCUL MATRICIEL</b> .....                    | 480 |
| 1. Définitions et opérations élémentaires sur les matrices ..... | 480 |
| 2. Longueur, produit scalaire, forme quadratique .....           | 482 |
| 3. Indépendance linéaire, rang d'une matrice .....               | 483 |
| 4. Matrice inverse .....   | 484 |
| 5. Éléments propres .....  | 485 |
| 6. Trace .....   | 486 |
| 7. Matrice symétrique et idempotente .....                       | 486 |
| 8. Inverse d'une matrice symétrique partitionnée .....           | 487 |
| 9. Dérivation vectorielle et matricielle .....                   | 487 |
| 10. Problèmes d'extréma .....                                    | 488 |
| <b>TABLES STATISTIQUES</b> .....                                 | 491 |
| <b>BIBLIOGRAPHIE</b> .....                                       | 503 |
| <b>INDEX ALPHABÉTIQUE DES PROGRAMMES</b> .....                   | 511 |
| <b>INDEX DES 52 FONCTIONS APL</b> .....                          | 513 |

## Introduction

La statistique, souvent définie comme l'étude des ensembles de données numériques, devait de façon privilégiée être aidée, enrichie, remodelée par le développement et la diffusion de l'informatique.

*Des techniques anciennes dans leur principe mais peu utilisées sont devenues faciles à mettre en œuvre.*

Ainsi, l'obstacle du calcul a pendant longtemps détourné l'attention des statisticiens des méthodes de statistique descriptive multidimensionnelle dont les principes théoriques étaient acquis de longue date. C'est également la dissuasion exercée par le volume des calculs qui a fait négliger les techniques robustes et les procédures de simulation, au profit de méthodes plus aisées à mettre en œuvre, mais au prix d'hypothèses de commodité parfois irréalistes.

*Des champs d'observation, autrefois jugés trop vastes et trop complexes, peuvent maintenant être étudiés.*

Les grands tableaux fournis par les données d'enquêtes peuvent dans une certaine mesure être traités de façon globale et donner lieu à des synthèses. Dans les domaines les plus divers, on peut procéder à des relevés de mesures extensifs sans réduction *a priori* du champ de l'observable, grâce aux possibilités actuelles de traitement.

*Des techniques nouvelles sont conçues et éprouvées.* C'est le cas des méthodes de classifications automatiques, et plus généralement de toutes les procédures de type algorithmique.

*Enfin, dans les méthodes mêmes de travail et d'investigation, on assiste à un déplacement des centres d'intérêt* : les questions d'adéquation à la réalité, d'analyse descriptive, de fidélité aux observations ont tendance à jouer un rôle de premier plan.

Dans le domaine de l'analyse des données multidimensionnelles, on ne peut manquer de citer l'influence et la stimulation exercées en France par les travaux de Jean-Paul Benzecri sur ce recentrage d'une partie de l'activité des statisticiens (Benzecri, 1964, 1973, 1976). Dans le domaine de l'analyse exploratoire des données (unidimensionnelles), on doit citer en particulier l'important travail de John W. Tukey (1977). Bien qu'ils soient de nature et d'inspiration très diverses, ces travaux ont en commun un certain nombre de caractères, le premier d'entre eux étant d'être « réactifs », c'est-à-dire issus d'une réaction par rapport à une certaine pratique de la statistique. On sait en effet que, conjointement aux hypothèses générales impliquées dans toute activité scientifique, la statistique fait intervenir un certain nombre

d'hypothèses de commodité (cette expression semble due à G. Morlat) qui ne sont pas là pour être confirmées ou infirmées, mais pour permettre l'exécution d'un test ou l'ajustement d'un modèle (un exemple classique en est l'hypothèse de normalité des résidus dans le modèle linéaire).

Test ou ajustement ne sont réalisables que si le réseau d'hypothèses permet une formulation analytique relativement simple, ou donne accès à des tables statistiques existantes. Or ces dernières conditions ne sont plus nécessaires avec la disponibilité des puissants moyens de calcul.

Les techniques de classification (chapitre V) ne donnent lieu à aucune formulation analytique élaborée ; les techniques d'estimation robuste, de validation par simulation, ne sont pas hypothéquées par des assertions ou des modèles très restrictifs, de même que les techniques d'analyse factorielle (chapitre IV) qui fournissent des représentations pour les associations entre lignes et colonnes de tableaux numériques.

On peut donc s'affranchir de certaines des hypothèses de commodité, et l'on peut aussi mettre à l'épreuve des hypothèses relatives au contenu du travail ou de la recherche beaucoup plus complexes et mieux adaptées grâce aux procédures de simulation. Cette capacité d'intervention du statisticien suppose une connaissance et un contrôle accrus des outils dont il dispose.

Les méthodes utilisées par le statisticien ne font pas toutes intervenir au même degré l'outil informatique. Les tests de rangs non paramétriques présentés au chapitre II n'impliquent que des calculs légers, pouvant être réalisés la plupart du temps à la main, ou avec l'aide d'une calculatrice « de poche ». Les techniques fishériennes présentées dans le même chapitre, qui font intervenir des permutations de l'ensemble fini des données numériques, nécessitent déjà un véritable ordinateur. Les calculs demandés par les applications du modèle linéaire (chapitre III) ne sont pas aisés à réaliser à la main ; l'ordinateur apporte cependant confort, rapidité et sécurité. Par contre, les méthodes d'analyse factorielle (chapitre IV) et de classification (chapitre V) ne peuvent donner lieu à des applications « en vraie grandeur » sans l'aide d'une calculatrice scientifique relativement importante.

L'utilisation de l'outil mental qu'est l'ordinateur peut décupler la puissance d'investigation et d'analyse du statisticien, pourvu qu'une division excessive du travail ne le transforme pas en consommateur passif de produits informatiques : lorsqu'il s'agit de connaissance et non de production, le partage des tâches peut en effet avoir des conséquences néfastes, car le produit de l'activité ne peut être contrôlé ni évalué facilement. L'autonomie du statisticien suppose donc l'aptitude à la mise en œuvre des méthodes, et la capacité d'analyse critique des produits logiciels.

L'empiètement des domaines de compétence des statisticiens et des informaticiens ne vise pas à subordonner l'une des professions à l'autre, mais doit permettre un dialogue efficace entre les spécialistes mis en cause. L'évolution rapide des matériels suffirait à elle seule à justifier la publication des programmes au niveau des algorithmes ou des petits modules dont la connaissance permettra de faire face aux adaptations nécessaires.