

001734/1

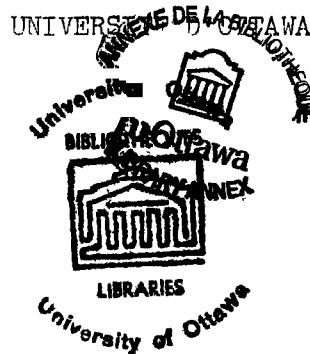
INFORMATIQUE ET ESPACE GEOGRAPHIQUE

LE MODELE S.E.T.G.E.G.

par

Guy Lemay

Baccalauréat ès Arts (Géographie)



Thèse présentée au Département de Géographie de

l'Université d'Ottawa

en vue de l'obtention de la

Maîtrise es Arts

OTTAWA - SEPTEMBRE 1972

UMI Number: EC55795

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform EC55795
Copyright 2011 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

REMERCIEMENTS

Nos remerciements sont offerts en premier lieu et tout spécialement à M. Jean-Bernard Racine, notre directeur de thèse qui, non seulement nous a aidé dans la préparation finale du présent ouvrage, mais en a aussi permis la réalisation à long terme en nous faisant participer à ses recherches, en tant qu'assistant pendant près de trois ans. Nous tenons à remercier également M. Henri Raymond pour ses nombreux conseils et suggestions et André Leclerc du Centre de Calcul pour son assistance du côté informatique de même que tous les professeurs et les étudiants qui de loin ou de près ont contribué indirectement à ce travail par leurs discussions et leurs commentaires.

Guy Lemay

SOMMAIRE

Considérant les exigences de la géographie moderne l'auteur se propose de concilier les techniques informatiques et la géographie en vue de l'obtention d'une méthodologie systématique. Dans un premier temps il discute de la nature des données géographiques et des problèmes liés à leur manipulation. Cette considération l'amène à réfléchir sur la transformation des observations qualitatives aux données quantitatives ainsi qu'à la représentation de l'espace quantifié. En second lieu il s'attarde aux divers algorithmes de l'analyse informatique de l'espace géographique qu'il traite séparément selon leur apport respectif à la détermination des structures spatiales. Ensuite il aborde la question de la modélisation des systèmes spatiaux qui complète la démarche de l'analyse. Enfin il présente un Système Informatique et Globale de l'Espace Géographique (S.E.T.G.E.G.) qui sous-tend la logique de la méthode quantitative en géographie.

TABLE DES MATIERES

	Page
REMERCIEMENTS	ii
SOMMAIRE	iii
TABLE DES MATIERES	iv
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xiii
INTRODUCTION	
0.1 Les objectifs de l'étude	1
0.1 a) Le point de départ	1
i) Les exigences épistémologiques	1
ii) Les exigences scientifiques	2
iii) Les exigences pédagogiques	3
0.1 b) Le problème se situe à trois niveaux	4
0.2 La nature de l'étude	5
0.2 a) Les réponses à trouver	5
0.2 b) Les tâches à réaliser	9
0.3 Le choix d'un exemple	9
0.4 Apport d'une interprétation qualitative de la banlieue	11

CHAPITRE I: LA NATURE DES DONNEES GEOGRAPHIQUES ET LES PROBLEMES LIES A LEUR MANIPULATION INFORMATIQUE	14
1.0 Introduction	14
1.1 De l'observation qualitative aux données quantitatives: une métamorphose virtuelle de l'espace	15
1.1 a) Le niveau potentiel	15
1.1 b) Niveau de l'observation et du recueil d'information	19
1.1 c) Le laboratoire: la reconstitution artificielle de l'espace étudié	20
1.1 d) Niveau analytique	21
1.1 e) Niveau du développement formel de la recherche: géographie fondamentale vs. géographie appliquée	22
1.2 L'espace quantifié: les nombres et leur signification	25
1.2 a) L'échelle nominale	27
1.2 b) L'échelle ordinale ou le rang	27
1.2 c) L'échelle d'intervalle	28
1.2 d) L'échelle des données relatives	28
1.3 L'évaluation et la classification des données brutes	29
1.3 a) Système numérique fermé	29
1.3 b) Système numérique ouvert	31
1.4 Les types de matrice géographique	34
1.4 a) La matrice descriptive graphique ou de convexité	34
1.4 b) La matrice d'information spatiale	35
1.4 c) La matrice de probabilité pré-graphique	39

1.5	La transformation de la matrice chrono-spatiale	41
1.5 a)	L'intérêt de la transposition	41
1.5 b)	L'intérêt de la standardisation	44
1.5 c)	Le problème de la normalisation	48
1.6	Les propriétés de la courbe normale et de ses paramètres constituants	50
1.6 a)	Test de normalité	55
1.6 b)	Mesure de dissymétrie	56
1.6 c)	Test de la kurtose	58
1.6 d)	Programme informatique de la normalisation	59
CHAPITRE II: LES ALGORITHMES DE L'ANALYSE INFORMA- TIQUE DE L'ESPACE GEOGRAPHIQUE		65
2.0	Introduction	65
2.1	La corrélation: point central de la réflexion et de l'analyse géographiques	66
2.1 a)	La linéarité	73
2.1 b)	L'homoscédasticité	74
2.1 c)	Test de t et coefficient de corrélation: niveau de signification du coefficient	77
2.1 d)	Le coefficient de détermination r^2	81
2.1 e)	Phi: ϕ	84
2.2	Les étapes de la réduction des systèmes spatiaux	88
2.2 a)	La réduction des attributs spatiaux: les analyses factorielles, leurs contraintes et leurs problèmes mathématiques	88
2.2 b)	L'analyse-image	109

2.2 c) La rotation Varimax	116
2.2 d) La rotation Quartimax	117
2.2 e) La rotation Equimax	120
2.2 f) La normalisation des facteurs	121
2.2 g) La rotation oblique	124
2.2 h) Les poids locaux	127
2.3 La réduction de la matrice d'information spatiale par l'analyse typologique	130
2.3 a) Le congroup	132
2.3 b) Le "linkage analysis"	139
2.3 c) La fonction discriminatoire	147
i) Les définitions	148
ii) De la matrice originale à la matrice des groupes	152
iii) Les itérations de l'optimisation typologique	154
iv) La structure de l'algorithme itératif dans le programme discrim	160
2.4 Les étapes de la modélisation partielle	176
2.4 a) La régression	177
i) Détermination de la ligne de régression	179
ii) L'analyse de variance comme test de signification pour la régression	182
iii) L'analyse des résidus et son intérêt géographique	185
2.4 b) La covariance multiple	190
2.4 c) L'analyse canonique	196
2.5 Conclusion	203

CHAPITRE III: UN SYSTEME INFORMATIQUE DE L'ANALYSE MULTIVARIEE EN GEOGRAPHIE: LE MODELE S.E.I.G.E.G.	205
3.0 Introduction	205
3.1 Caractéristiques générales du système S.E.I.G.E.G.	206
3.1 a) Niveau de la préparation des données et de la réduction des attributs spatiaux	209
3.1 b) Réduction des observations	210
3.1 c) La simulation partielle et les différents types de régression	212
3.2 Caractéristiques techniques du modèle S.E.I.G.E.G et mode d'emploi	214
3.3 Cartes de contrôle de la première partie du système S.E.I.G.E.G.	217
3.3 a) Cartes générales de contrôle	218
i) Carte de spécificité de la matrice d'information spatiale	218
ii) Carte de titre	219
iii) Carte d'identification des variables	220
iv) Cartes de format	220
v) Cartes de données	222
vi) Cartes de sélection	224
3.3 b) Carte A génération	224
3.3 c) Carte B transposition	229
3.3 d) Carte B typologie	232
3.3 e) Carte C normalisation/standardisation	232
3.3 f) Carte D corrélation	235
3.3 g) Carte E linéaire	235
3.3 h) Carte F factorisation	236

3.4	Les cartes de contrôle des algorithmes de l'analyse typologique - le congroup et le discrim	230
3.4 a)	Caractéristiques et organisation du programme de groupement géographique	239
3.4 b)	Caractéristiques et organisation du programme discrim	246
3.4 c)	Caractéristiques et cartes de contrôle du programme mulcov	249
3.5	Troisième partie du modèle S.F.I.G.E.G.: sous-système de la modélisation partielle	252
3.6	Conclusion	265
	CONCLUSION	267
	APPENDICE A: SIGNIFICATION DES SYMBOLES EMPLOYES DANS LES ORGANIGRAMMES	271
	APPENDICE B: LISTE DES SOUS-PROGRAMMES	272
	APPENDICE C: LES PROGRAMMES INFORMATIQUES DU SYSTEME	277
	BIBLIOGRAPHIE	

LISTE DES TABLEAUX

Numéro		Page
1	Système numérique fermé: utilisation du sol de quatre villes fictives	30
2	Système numérique ouvert: variables socio-économiques de quatre villes fictives	31
3	Matrice d'information spatiale	38
4	Matrice transposée	42
5	Matrice standardisée	46
6	Calcul du coefficient de corrélation	68
7	Matrice des corrélations	71
8	Niveau de signification de la corrélation	80
9	Matrices d'entrée de l'analyse en composantes principales et de l'analyse factorielle	91
10	Matrice standardisée	96
11	Matrice des saturations (douze facteurs)	99
12	Matrice des vecteurs propres	101
13	Facteurs de la méthode factorielle-composantes (Falcomp)	110
14	Saturations: rotation Varimax	115
15	Convergence des cycles de la rotation Quartimax	119
16	Matrice des saturations normalisées	122
17	Matrice des saturations dé-normalisées	123
18	Matrice des corrélations factorielles	126
19	Matrices des poids locaux par la méthode classique et celle de Kaiser	131

20	Matrice des distances	133
21	Matrice de contiguïté	137
22	Correspondances typiques	146
23	Classification initiale de la fonction discriminante	171
24	Matrice des distances à la centroïde des groupes, itération 1	172
25	Matrice des distances à la centroïde des groupes, itération 3	173
26	Bilan des permutations après trois itérations	174
27	Analyse de la distribution des densités résidentielles dans la zone métropolitaine de Montréal au sud du Saint-Laurent (1966)	189

LISTE DES FIGURES

Numéro		Page
1	Carte des 22 municipalités de la rive sud de Montréal	10
2	Démarche méthodologique de la géographie	16
3	Organigramme méthodologico-informatique	26
4	Forme générale de la matrice d'information spatiale	36
5	Forme générale de la matrice des probabilités	39
6	Représentation verticale ponctuelle d'une composition unitaire de l'espace	43
7	Courbes normales de moyennes variables	51
8	Courbes normales de différents échantillons	52
9	Courbes normales de différents écart-types	52
10	Représentation en pourcentage de l'écart-type sous une courbe normale	53
11	Dissymétries de la courbe normale	57
12	Organigramme de la normalisation	63
13	Rapport entre les coefficients r et k	73
14	"Scattergram" de deux variables en corrélation	77
15	Rapport de variation de la corrélation	81
16	Rapport entre les coefficients r et r^2	83
17	Représentation graphique de quatre corrélations	93
18	Représentation graphique d'un facteur	106
19	Représentation graphique des facteurs 1 et 2	114

20	Représentation graphique d'une rotation oblique	125
21	Représentation graphique de l'algorithme de classification congroup	134
22	Dendrogramme de l'utilisation du sol	136
23	Représentation graphique de la fonction discriminatoire	140
24	Représentation graphique de l'erreur-type d'estimation	181
25	Représentation graphique de la variation dans la régression	183
26	Schéma d'une configuration sectorielle	191
27	Lignes de régression à pentes variables	193
28	Lignes de régression à interceptes variables	194
29	Organigramme de la covariance multiple	195
30	Ordinogramme du système S.F.T.C.T.G.	207
31	Schéma général de la soumission d'un programme à l'ordinateur	215
32	Cartes maîtresses du programme	216
33	Exemple de cartes d'identification des variables	221
34	Cartes de format et cartes de données correspondantes	223
35	Cartes de sélection de la première partie du système	225
36	Cartes de contrôle de la sélection A génération	230
37	Schéma de l'organisation du programme congroup	241

3°	Cartes de contrôle du programme sous-groupe	245
3°	Schéma de l'organisation du programme discrim	247
4°	Cartes de contrôle du programme discrim	250
41	Schéma de l'organisation du programme multicor	251
42	Cartes de contrôle du programme multicor	253
43	Schéma de l'organisation des programmes de la modélisation partielle	255
44	Cartes de contrôle des programmes de la modélisation partielle	262

TNTRODUCTION

TNFORMATIQUE ET ESPACE GEOGRAPHIQUE

LE MODELE S.E.T.G.E.G.

0-1) LES OBJECTIFS DE L'ETUDE

0-1 a) Le point de départ:

Depuis les dix dernières années la rigueur scientifique a été de plus en plus en vogue parmi les sciences humaines et la géographie n'a pas fait exception à cette transformation qui s'avérait urgente et nécessaire. Rien que celle-ci ait fait des pas de géant elle est maintenant aux prises avec trois types d'exigences fondamentales. Elles sont d'ordre épistémologique, scientifique et pédagogique.

i) Les exigences épistémologiques

On partira du constat d'une première contradiction excellemment mise en évidence par (Harvey, 1969) dans son livre sur la problématique de la géographie: "Jusqu'à présent les géographes se sont considérés satisfaits des définitions implicites de leur point de vue sans toutefois chercher une théorie spécifique à la discipline". En effet le géographe a besoin d'identifier un champ d'étude particulier ou un ensemble de champs avec lequel il est strictement concerné. De là tient la nature même de la géographie; si on s'accorde à dire que la géographie est une discipline du sol, de l'espace concret, des noms de lieux, que le savoir du géographe est un savoir chrono-spatial (Labasse, 1971) et que la connaissance

géographique doit déboucher sur la découverte des conditions de la prévision et de l'action (George, 1964): il faut en arriver à dégager les traits fondamentaux, en termes dynamiques, de la différenciation et de l'organisation de l'espace. A cette fin le géographe doit se donner les outils qui puissent lui permettre de satisfaire sa curiosité intellectuelle et professionnelle à l'intérieur d'un corpus méthodologique où il se sentira à l'aise et confiant dans sa recherche et dans la présentation de ses énoncés et de ses résultats. Ce qui implique une modélisation méthodologique qui corresponde à la fois à la conceptualisation théorique de la démarche géographique et à la réalité à laquelle elle s'applique. Tel est la dimension première à laquelle notre projet s'intéresse.

ii) Les exigences scientifiques

En admettant que le géographe doive expliquer cet espace il lui faut travailler au niveau de l'ensemble de ses caractéristiques saisies simultanément dans l'espace et dans le temps, (Racine, 1971). On reconnaîtra cependant - tout particulièrement en ce qui concerne les écoles de géographie traditionnelle, dont l'école géographique française - que cette exigence n'a été cultivée qu'au niveau historico-littéraire de l'analyse. Les exigences nouvelles de la recherche scientifique (expériences et mesures) obligent à avoir recours à l'analyse quantitative, (French, Racine, 1971). Mais en rai-

son de la nature multidimensionnelle des problèmes, cette analyse quantitative réclame ~~en outre~~ l'utilisation de la statistique multivariée qui ne se conçoit qu'en ayant recours à l'informatique. Il n'y a pas lieu ici d'en faire une nouvelle démonstration. A cet égard nous renvoyons le lecteur aux travaux de Berry & Marble (1968), à ceux du département de géographie de l'Université d'Ottawa (1971) édités par M.H. French et Racine (1971) ainsi qu'à l'article de J.B. Racine (1971 et 1972): Modèles graphiques et mathématiques en géographie humaine, qui établissent clairement à notre sens le besoin aujourd'hui inéluctable du recours à l'informatique.

iii) Les exigences pédagogiques

Cette contrainte peut sembler moins évidente et impérative pour certains, mais elle représente un problème pour ceux qui doivent enseigner cette nouvelle géographie et assurer la relève. Or la carence de travaux d'orientation pédagogique se fait d'ailleurs sentir aussi bien au sein de la géographie anglophone qui pratique cette nouvelle géographie depuis bientôt vingt ans que dans l'école française qui ne l'aborde que bien timidement, après une longue période d'indifférence, voire de réaction.

Il existe bien quelques ouvrages qui expliquent les différentes techniques et méthodes employées en géographie

mais aucun à notre connaissance n'a osé se poser le problème d'une présentation systématique et opérationnelle, seule susceptible de venir en aide à une population de géographes jeunes ou plus anciens qui sont ensemble au stade de l'initiation, tout particulièrement lorsqu'il s'agit d'aborder la statistique multivariée et l'informatique.

Le problème ne se réduit d'ailleurs pas au seul plan des techniques. Le simple fait d'établir une structure logique et opérationnelle facilement accessible et compréhensible aura pour effet de favoriser la constitution d'un nouveau "paradigme" capable de faire évoluer les diverses théories et modèles géographiques, la nature des problèmes géographiques évoluant peut-être en même temps que les méthodes.

0-1 b) Le problème se situe à trois niveaux

i) Quelles sont les techniques de l'informatique qui permettent de mieux répondre aux questions que se pose le géographe;

ii) quels sont les problèmes que pose leur utilisation et quels en sont les limites et les dangers;

iii) comment peut-on concevoir, en admettant que les travaux de Racine (1971, 1972) et de Racine - Lemay (1972) aient ouvert une voie nouvelle à la définition d'une démarche spécifiquement géographique (en termes épistémologiques), un véritable système informatique de l'analyse de l'espace géo-

graphique? Dans ce cas quel en serait la structure?

0-2) LA NATURE DE L'ETUDE

0-2 a) Les réponses à trouver

Il n'est pas possible dans le cadre d'un travail académique de cet ordre de reprendre pour la discuter la problématique d'ensemble de la géographie. Le lecteur pourra se référer à une quantité de travaux depuis Vidal de la Blache à Pierre George (1970) et Jacqueline Beaujeu Garnier (1971) du côté français et de Harlan Barrows (1923) à Berry (1972) "A new paradigm for modern geography" en passant par les célèbres ouvrages de Harsthorne, (1939) Bunge, (1966) Harvey (1969).

De la lecture de ces travaux nous nous contenterons d'isoler la définition de la géographie à peu près admise par tout le monde et citée au début de notre présentation: la géographie est la science de la différenciation et de l'organisation des éco-systèmes spatiaux, définition qui en soit est largement suffisante pour nous permettre de choisir à l'intérieur de l'immense corpus méthodologique lié à l'informatique, les techniques qui semblent à l'évidence répondre à des besoins que plus personne ne conteste, techniques qui ont en outre déjà été utilisées par des géographes sans que toutefois ces derniers se soient toujours suffisamment préoccupés de la pertinence de l'utilisation d'une méthode en soit parfaite

mais peut-être difficilement applicable à un domaine qui obéit à des spécificités différentes de celles des domaines pour lesquelles les méthodes ont été élaborées. (Un exemple aujourd'hui classique, est celui du problème de la normalisation et de l'autocorrélation tel que mentionné par Racine, Harvey, Gould, Berry).

Dans son article Racine (1972) propose en particulier, après avoir souligné l'importance de se donner dès le départ un langage utilisable par l'ordinateur (cf. discussion sur la discipline de la matrice d'information spatiale), d'utiliser d'abord, (montés "en batterie") un ensemble d'algorithmes de réduction. Cette réduction résulte de la découverte des principaux facteurs de la différenciation de l'espace (FACTO) et de la mise en évidence (CONGROUP, LINKA, DISCRIM) des principaux types de combinaisons spatiales, combinaisons qui se réduisent finalement en un certain type d'organisation de l'espace, la réalité du système observé tel que défini au départ dans la matrice d'information spatiale originale. Dans leur article sur l'analyse discriminatoire des correspondances typologiques Racine & Lemay (1972) ont cependant cherché à aller plus loin et conformément au modèle de recherche globale et informatique proposé par Racine (1972), ils ont cherché les moyens de modéliser les systèmes spatiaux une fois que ceux-ci ont été réduits.

On peut toutefois concevoir la modélisation partielle des systèmes spatiaux. C'est à ce titre que les géographes utilisent les modèles de régression multiple normale ou par étape, qu'ils utilisent les modèles de corrélation canonique. La solution "globale" envisagée par Racine et Lemay l'a été parce qu'elle leur paraissait plus géographique en ce sens qu'elle était plus conforme à la définition de départ: différenciation (facto) et organisation de l'espace. Dans bien des cas cependant, le géographe a besoin de ces modèles partiels. C'est donc l'ensemble de ces différents procédés d'analyse multivariée (volet opérationnel de la notion de système) que nous allons présenter, discuter et tester dans notre travail, suivant un ordre conforme au modèle de démarche globale proposé par Racine.

Cependant les travaux de Racine, et de Racine et Lemay n'ont pas couvert tout le champ du modèle de recherche proposé (pas de régression multiple, pas de régression par étape, pas de régression canonique, pas de covariance multiple surtout pas de matrice TYPOLOGIQUE DEDUITE) mais encore, chacun des algorithmes utilisés possède plusieurs variantes (par exemple Facto offre un choix entre l'analyse de composantes principales, l'analyse factorielle, "l'image analysis") sur les qualités respectives et les problèmes particuliers sur lesquels il faudra nous pencher.

On ne peut enfin concevoir d'analyse géographique véritablement explicative qui fasse abstraction de la prise en considération scientifique du jeu des influences et de la perception des populations concernées. Ce serait, comme l'a remarquablement montré Andrei Rogers (1967) dans sa critique des modèles de structure interne urbaine élaborés par l'école de Chicago, revenir à une sorte de déterminisme social sinon à un déterminisme "scientiste" du type de celui qui a trop longtemps prévalu aux débuts de la géographie et qu'a violemment dénoncé Vidal la Blache. Rogers propose qu'une solution soit cherchée dans le recours systématique à des modèles de type stochastique.

Dans l'état actuel de nos connaissances et des sources de documentation dont nous disposons, nous ne pouvons pas élaborer plus avant ce thème de recherche qui n'est qu'à l'état embryonnaire et sur lequel se penchent encore ensemble les professeurs J.B. Racine et Bryn Greer-Wootten.

L'application de ces techniques élaborées à l'origine pour des besoins totalement différents des besoins des géographes, posent des problèmes d'utilisation qui ne sont pas encore résolus (cf. Comparative Factorial Ecology, Berry et Horton ch. 9, Gould, 1970, Berry, 1972; normalisation, rotation, etc.).

0-2 b) Les tâches à réaliser

La tâche à réaliser est donc encore considérable pour que le modèle de Racine puisse être valablement évalué même si on accepte à priori la validité de sa structure.

En ce qui concerne notre projet - qui reste dans le cadre obligatoirement limité d'une thèse de maîtrise - **trois** tâches nous attendent:

i) Traiter de la nature de l'information quantifiée livrée à l'ordinateur et des phases préliminaires de sa transformation avant que ne soit pratiquée l'analyse quantitative proprement dite.

ii) Compléter la présentation des algorithmes utilisés et utilisables et présenter leurs différentes caractéristiques respectives.

iii) Mettre au point et publier l'ensemble des programmes utilisés selon un système logique et opérationnel, le modèle S.E.I.G.E.G. (Système d'Etude Informatique et Globale de l'Espace Géographique).

0-3) LE CHOIX D'UN EXEMPLE

Il est certain qu'un tel travail se voulant utile et explicite était impensable sans le recours à un exemple relativement simple et versatile pour démontrer avec toutes les exigences qui s'imposent les caractéristiques et les limites de

l'analyse quantitative. Le problème était donc pour nous de trouver un ensemble de variables, d'attributs spatiaux qui réuniraient simultanément la simplicité, l'intelligibilité et le fonctionnel.

De plus cet exemple se devait de contenir une logique proprement géographique susceptible d'inciter de nouvelles voies par le simple biais de l'analogie. Si on est fidèle à la définition de la géographie citée au début il nous faut un espace concret et des caractéristiques propres à cet espace. Nos travaux d'assistant de J.B. Racine nous ont amené à travailler de façon toute particulière avec les vingt-deux municipalités de la rive sud de la région métropolitaine de Montréal. (figure 1). Cette région contient un continuum qui va de l'urbain au rural en passant par au moins une demi douzaine de types suburbains différents. C'est du moins ce qu'ont pu démontrer les algorithmes utilisés par Racine (1972) et Racine et Cavalier (1972). Cette banlieue nous est apparue comme un territoire typique sur lequel l'analyse quantitative était en mesure de fournir des éléments démonstratifs d'une interprétation géographique.

L'aspect sous lequel nous avons étudié cet espace devait renfermer deux qualités: premièrement être simple, facilement compréhensible, afin de ne pas perdre d'énergie à définir avec emphase le problème et deuxièmement caractériser l'espace de façon concrète en montrant l'impact de l'homme

sur son milieu tout en rassemblant les préoccupations de la géographie physique et humaine. A cet effet il nous est vite apparu que les différents types de l'utilisation du sol se sont avérés des critères fondamentaux pour une analyse possédant les deux qualités réclamées ci-dessus. A cet effet douze caractéristiques de l'utilisation du sol (que le lecteur trouvera plus loin) ont fourni les données de base pour présenter, tester et illustrer les procédés de l'analyse quantitative.

Mais avant de passer à la première partie du travail, voyons ce que la géographie qualitative pourrait faire ressortir de l'étude de ces variables. Ceci n'a pas pour but d'abaïsser ou de diminuer les mérites du qualitatif mais plutôt de fournir un point de comparaison entre les deux points de vue. Nous ne voulons pas non plus reprendre un débat qui a déjà fait l'objet de travaux antécédents (French/Racine, 1971) ni agir en défenseur d'une méthode que nous venons à peine de maîtriser. Nous posons simplement les faits. Au lecteur de faire son choix si choix il y a.

0-4) APPORT D'UNE INTERPRETATION QUALITATIVE DE LA BANLIEUE

La préoccupation première du géographe qualitatif est de délimiter et de décrire le secteur étudié. Ceci fait, il tente de décrire le phénomène de dispersion à l'aide des critères des structures démographiques et de la mobilité de

la population. Il traite de la diversité de la vie de la banlieue qu'il rattache à divers types de banlieue prédéfinis tel que la banlieue dortoir, la banlieue industrielle, la banlieue récréation, etc. Il aborde souvent les conditions historiques du développement de la banlieue qui ont, il est vrai, plus d'importance dans les banlieues européennes qu'américaines. L'aspect de l'utilisation du sol n'a que le mérite de la description du paysage accompagné de quelques cartes qui servent à montrer plutôt qu'à expliquer des relations et tester la validité des liaisons.

Sous le couvert des "images" est-ce que le qualitatif saisit adéquatement la réalité des phénomènes qu'il étudie? Ce n'est pas à nous de répondre à cette question. Plusieurs travaux qualitatifs antécédents l'ont fait par leur contenu.

De plus une approche qualitative, tout en saisissant la diversité des paysages demeure en difficulté lorsqu'elle doit établir une typologie. Celle-ci est alors forcément basée sur les critères de l'intuition et de l'esprit de finesse du chercheur. Toutefois une bonne classification demeure en elle-même un très bon point de départ pour des études subséquentes.

Un géographe qualitatif expérimenté pourra après plusieurs sessions sur le terrain définir adéquatement la structure de la banlieue et en percevoir les divergences spatiales.

L'étude descriptive considère les formes d'organisation de la vie humaine par l'intermédiaire des paysages qui en sont le reflet. Ceci débouche sur une typologie descriptive des faits observés qui conduit à la cartographie et à l'interprétation graphique de la configuration péri-urbaine.

La superposition des cartes offre un moyen d'établir des relations entre les différents phénomènes spatiaux mais n'indique avec certitude ni le nombre ni la force (l'intensité) des rapports qui sont à notre avis à la base de toute la philosophie de la géographie contemporaine.

Nous présentons donc aux étudiants et aux chercheurs des méthodes qui sont autant d'outils capables de renforcer l'infra-structure de la géographie qualitative et d'offrir par le fait même toute une gamme de procédés indispensables à la nouvelle géographie.

CHAPITRE I

LA NATURE DES DONNEES GEOGRAPHIQUES ET LES PROBLEMES

LIES A LEUR MANIPULATION INFORMATIQUE

1.0) INTRODUCTION

La gamme des différences individuelles chez l'homme donne lieu à une multiplicité de perceptions subjectives face au monde réel. De ce fait se pose la question suivante, à savoir, comment plusieurs individus peuvent-ils en arriver à étudier un espace et ses caractéristiques en arrivant à des résultats communs ou au moins très proches les uns des autres, condition de leur validité scientifique?

Le géographe doit se donner une méthode qui puisse permettre au chercheur de mettre à profit ses expériences, sa personnalité, son intuition dans des recherches qui seront facilement comparables et intégrables sur le plan multi-disciplinaire. Pour y parvenir il est impératif d'établir premièrement un moyen de mettre sur une base commune, en l'occurrence, les valeurs numériques, les différentes observations d'ordre qualitatif et deuxièmement de développer une méthode de transformation des observations ainsi quantifiées pour être traitées par les différents programmes informatiques utiles aux géographes.

1.1) DE L'OBSERVATION QUALITATIVE AUX DONNEES QUANTITATIVES: UNE METAMORPHOSE VIRTUELLE DE L'ESPACE

La multiplicité des descriptions de paysages ou d'espaces a placé le géographe devant le problème de l'impossibilité de saisir simultanément toutes les variations, parfois subjectives, des constituantes spatiales. La solution, facile en théorie mais ardue en pratique, consiste à choisir un modèle méthodologique commun afin de pouvoir comparer les informations recueillies et les méthodes d'individus qui peuvent fort bien réagir au départ, de façon foncièrement différente face au même espace observé.

Ce modèle de la démarche géographique se doit de procéder à trois niveaux que nous caractérisons de: niveau potentiel, niveau de l'observation et du recueil des informations, niveau de l'analyse enfin (figure 2).

1.1 a) Le niveau potentiel

Pour reprendre une idée précédente, c'est à cette étape qu'entre en ligne de compte la personnalité, l'intuition, la culture géographique et les expériences perceptuelles de chaque individu au début de toute recherche, phénomène déjà remarqué par l'astronome Bessel et plus tard défini comme l'équation personnelle du chercheur.

Ce niveau, dénommé "potentiel" puisqu'il contient toute la problématique et l'orientation de la recherche, existe à l'état

FIGURE 1

LES MUNICIPALITES DE LA ZONE METROPOLITAINE DE MONTREAL

AU SUD DU SAINT-LAURENT

- 1 Boucherville (Définition 1966)
- 2 Brossard
- 3 Candiac
- 4 Caughnawaga
- 5 Chateauguay
- 6 Chateauguay-Centre
- 7 Chateauguay-Heights
- 8 Delson
- 9 Greenfield Park
- 10 Jacques-Cartier
- 11 Laflèche
- 12 Laprairie
- 13 Lemoyne
- 14 Léry
- 15 Longueuil
- 16 Notre-Dame
- 17 Prévilles
- 18 Saint-Bruno-de-Montarville
- 19 Sainte-Catherine d'Alexandrie
- 20 Saint-Constant
- 21 Saint-Hubert
- 22 Saint-Lambert

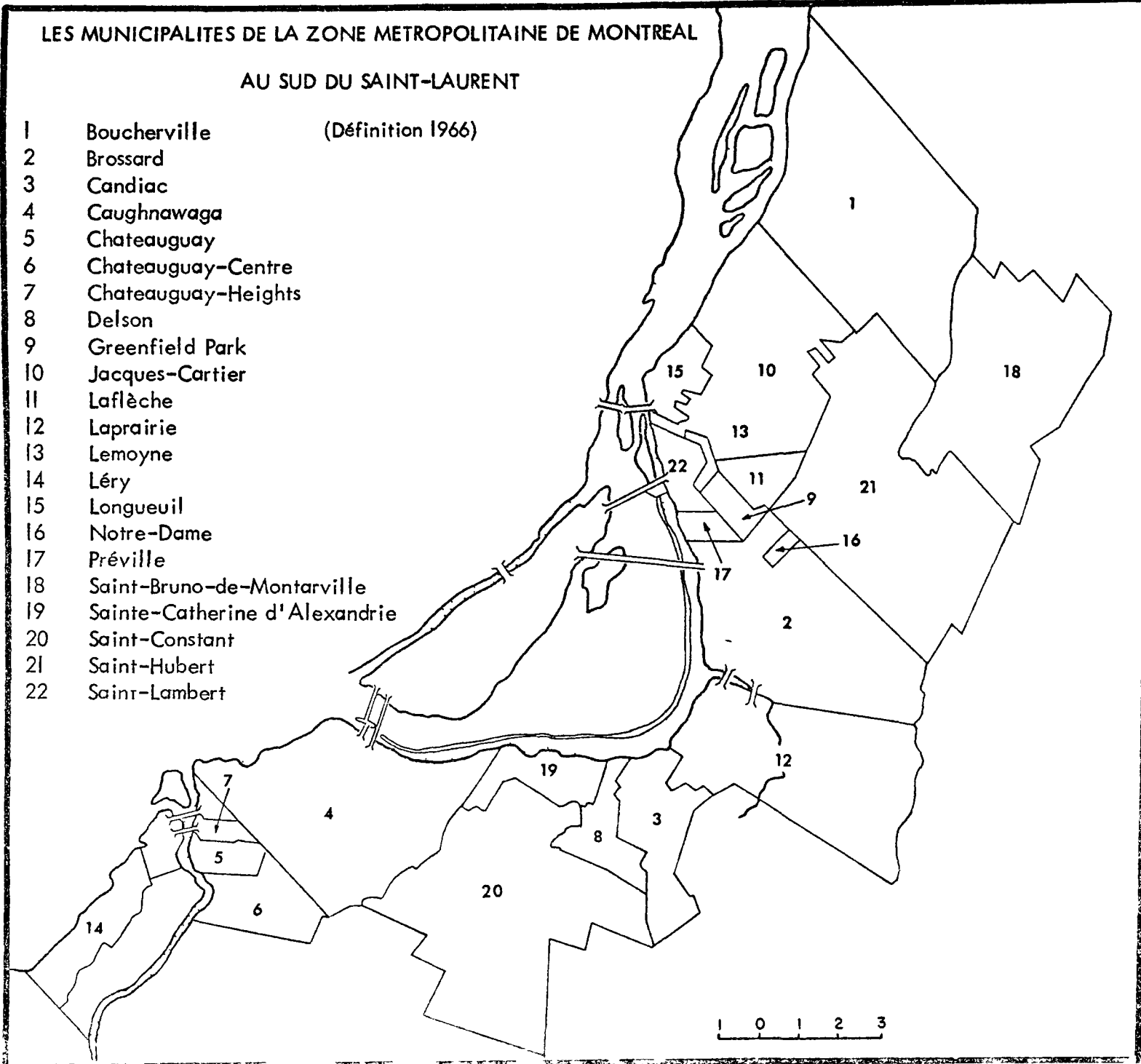
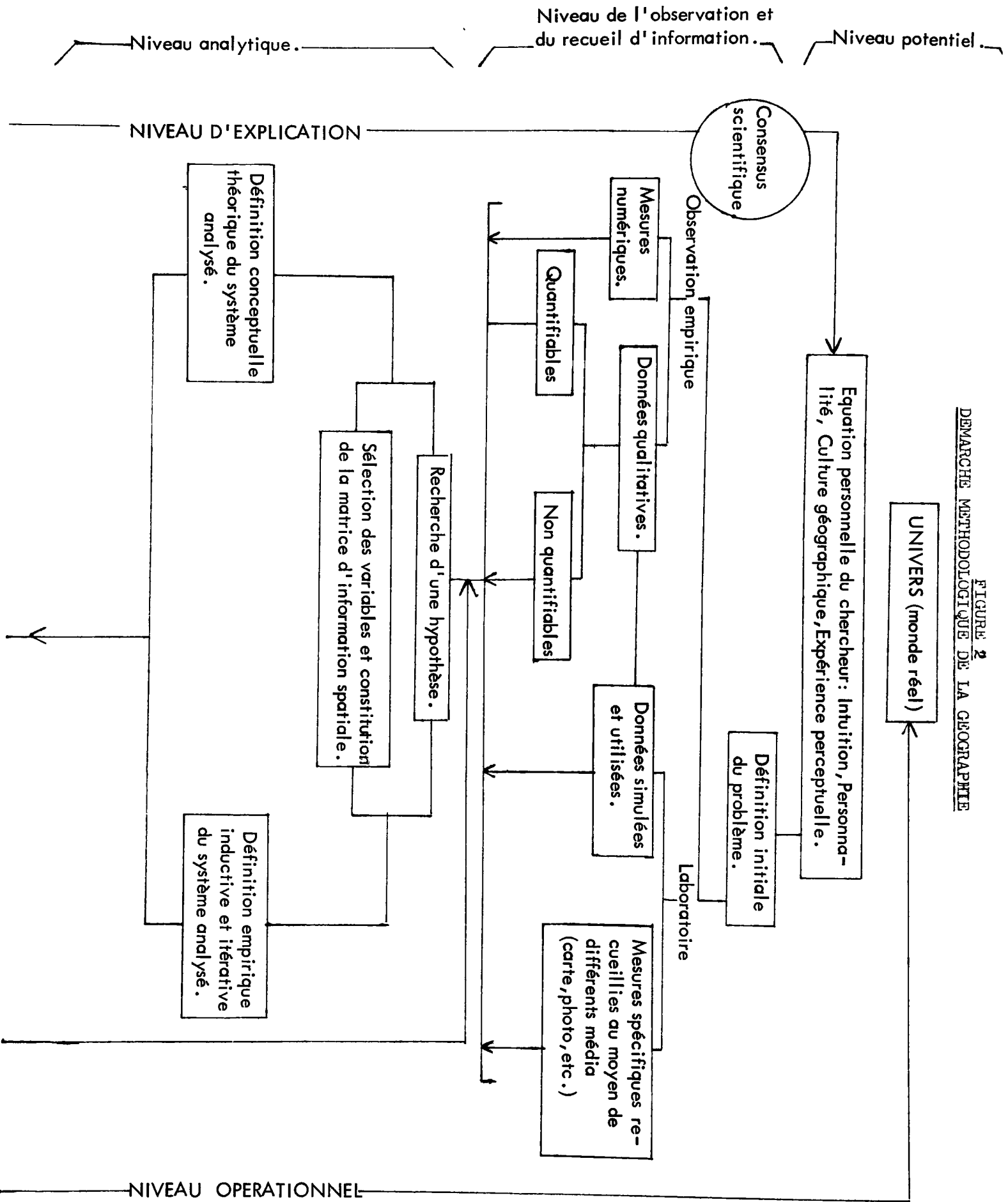
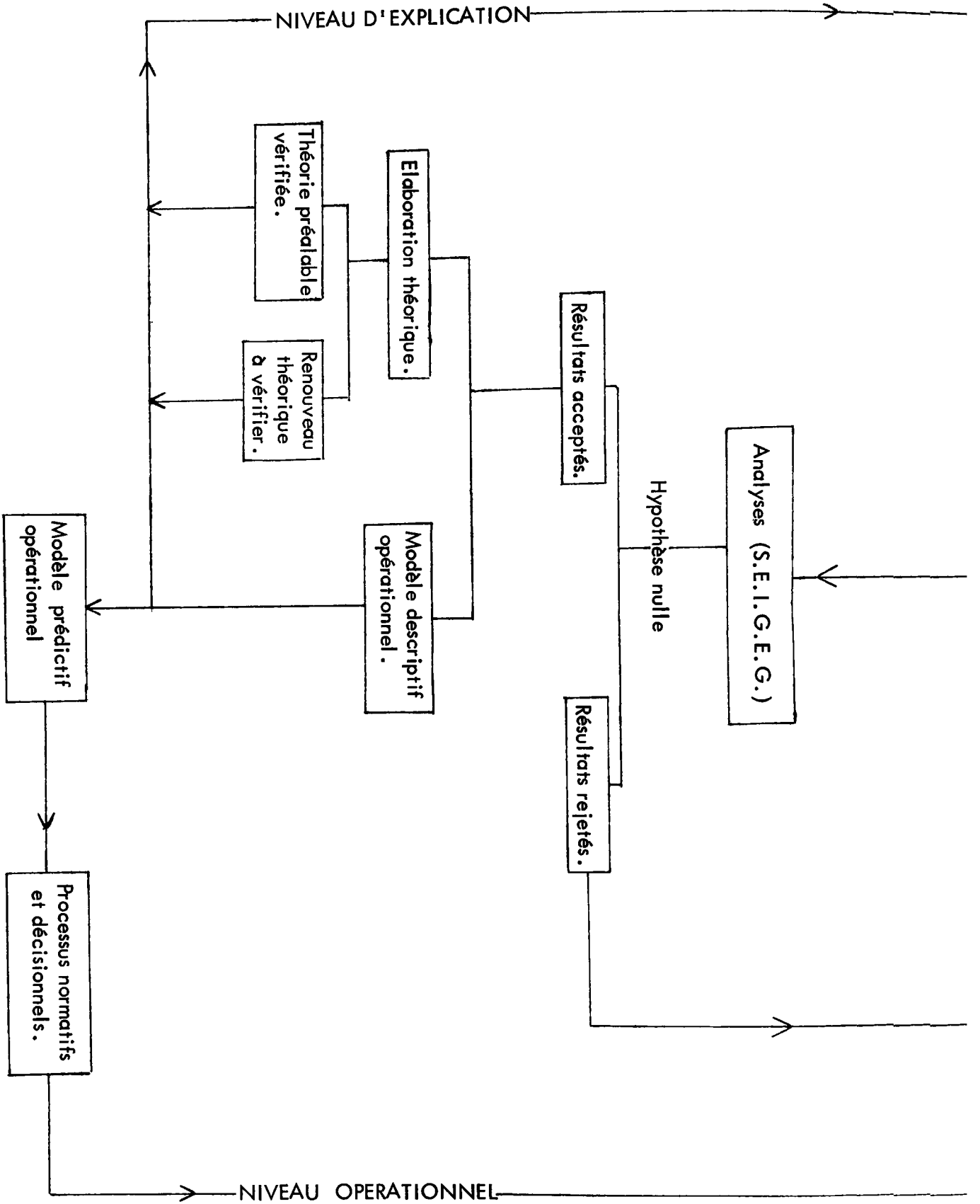


FIGURE 2
 DEMARCHE METHODOLOGIQUE DE LA GEOGRAPHIE



Niveau du développement formel de la recherche



latent, contribue à la définition de la recherche et fournit un cadre original et personnel difficilement réalisable aux autres niveaux, en particulier à celui de l'observation et du recueil d'information.

1.1 b) Niveau de l'observation et du recueil d'information

On partira du constat que les données géographiques s'observent à deux échelles schématiques: le terrain et le laboratoire. L'observation empirique sur le terrain permet au géographe de prendre contact avec le milieu qu'il étudie. "L'esprit de finesse", la perspicacité et l'intuition du géographe l'aideront à établir dès le départ les diverses théories, les modèles ou les lois qu'il se devra de vérifier empiriquement par le traitement informatique et statistique des données.

Cependant certaines observations ne sont pas quantifiables. Prenons un exemple: les valeurs sociales que les gens investissent dans les différents quartiers sont très souvent fonction des attachements sentimentaux plutôt que reliées à des causes bien précises et fonctionnelles. Walter Firey a soulevé la question dans son article: "Sentiment et métabolisme" (1945) et met en garde contre une tentative d'explication à priori qui ne tiendrait pas compte des tendances et des aspects psychologiques de l'homme. Il n'est point question de mettre de côté ces observations incommensurables car elles serviront justement à venir appuyer les aspects qui semblent passer inaperçus à travers une quantification souvent

rigide et ambitieuse, surtout si les études en laboratoire sont disproportionnées par rapport au temps passé sur le terrain.

1.1 c) Le laboratoire: la reconstitution artificielle de l'espace étudié

A ce niveau de notre réflexion, une distinction s'impose: la différence est marquée si la spécialisation du géographe est d'ordre physique ou humaine. Si les techniques de mesure et de manipulation en laboratoire ont été très bien développées pour la géographie physique, il n'en est pas de même pour le géographe humain qui pourtant a lui aussi l'ambition de se plier à l'expérimentation (George, 1970).

Un premier type d'expérimentation consiste à pouvoir mesurer au moyen de divers médiums tels la carte, les photos aériennes, le cadastre, le rôle d'évaluation, un certain nombre de caractéristiques capables de fournir différents éléments d'information. Un second type d'étude en laboratoire, plus délicat et plus controversé, est celui de la simulation. Cette technique, relativement récente, s'est vu ouvrir des horizons nouveaux grâce à la facilité d'accès et à la rapidité des ordinateurs et s'est épanoui avec les travaux de J.W. Forester (1971) dont le retentissement a été mondial. La technique nécessite au préalable que les données de terrain soient assemblées et permet de contrôler les différents paramètres constituant d'un espace pour arriver, au moyen d'itérations successives, à déterminer les diverses inter-relations définies-

sant la structure dynamique du système étudié.

Le second niveau de la figure 1 illustre schématiquement la place respective des différents ordres de mesure à l'intérieur du schéma méthodologique de la géographie contemporaine.

L'étape du recueil d'information terminée, la séquence se poursuit au niveau analytique.

1.1 d) Niveau analytique

L'énoncé d'une hypothèse oblige à sélectionner les variables pour former une matrice d'information spatiale sur laquelle sera effectuée l'analyse statistique.

L'hypothèse et la sélection des variables obéissent aux contraintes d'une définition conceptuelle théorique qui doit satisfaire une idée ou une philosophie mais qui ne peut éviter de tenir compte de la définition empirique et itérative du système.

En complément on retrouve chez S.S. Stevens (1935) cette même préoccupation de définir tout simplement le problème par les opérations concrètes qui contribuent à l'élucider. Cette approche de l'école anglaise, définie comme "opérationisme" (Harvey 1969), est employée pour éclairer des problèmes méthodologiques et philosophiques complexes qui ne peuvent être cernés par la réflexion théorique pure et simple.

L'analyse proprement dite comprend plusieurs techniques et méthodes qui seront expliquées successivement dans la deuxième par-

tie de notre travail. Il convient de noter ici cependant que le choix d'une méthode dépend tout à la fois de la réflexion théorique et des limites mathématiques et opérationnelles de cette méthode. Après analyse, l'acceptation ou le rejet de l'hypothèse nulle* et des résultats, oriente d'une façon différentielle la démarche à suivre. Devant un cas de rejet de l'hypothèse, il est nécessaire de revenir à l'étape de la formulation de l'hypothèse et du choix des variables, ou encore de reformuler l'hypothèse originale pour sélectionner des variables plus appropriées. Par contre l'acceptation de l'hypothèse nulle donne le feu vert pour passer au niveau du développement formel de la recherche.

1.1 e) Niveau du développement formel de la recherche: Géographie fondamentale vs. géographie appliquée

Une fois l'hypothèse confirmée, le chercheur peut s'orienter dans deux directions différentes. La première consiste à prolonger l'élaboration théorique tandis que la deuxième présente l'opportunité de développer un modèle descriptif opérationnel.

L'élaboration théorique a pour but de vérifier ou de modifier une théorie préalable ou encore d'en établir une de toute pièce. Dans le cas d'un renouveau théorique la branche récurrente de notre modèle montre qu'intervient à nouveau l'équation person-

* La notion d'hypothèse nulle sera commentée plus loin (cf. 3e partie partie, ch. 2).

nelle du chercheur. Ce renouveau théorique que G. Olsson (1965) appelle un "niveau d'explication" doit être à l'origine et développer un consensus accepté par la majorité qui permettra de formuler un nouveau paradigme scientifique (Kuhn, 1962). Mais le retour à l'équation personnelle du chercheur oblige à refaire la séquence méthodologique entière pour que soit vérifiée la nouvelle théorie, basée sur les modifications qui s'avéraient nécessaires à la fin du cycle précédent. Ce procédé itératif permet une progression sûre et efficace qui satisfait la rigueur scientifique et accroît le domaine des connaissances sur le problème posé.

Le cheminement du modèle descriptif opérationnel est beaucoup plus récent et fait appel à l'action, basée sur les connaissances acquises au cours du procédé de l'analyse.

S'il demeure au stade descriptif il emprunte la voie du niveau explicatif. Par contre si on peut lui garantir une dimension prévisionnelle, il joue un rôle de premier ordre dans les processus normatifs et "décisionnels" qui s'emploieront, par le biais du niveau opérationnel, à avoir un impact tangible et réel sur la conjoncture spatiale que le modèle s'est attaché à étudier et à simuler. Cette dernière étape est à notre avis une nouvelle ouverture, un défi à la mesure du 20e siècle, sur lequel il faudra se pencher comme l'ont déjà fait les précurseurs en la matière, tels Jay W.S. Forester (1961) et Andrei Rogers (1965).

Mais avant de procéder à l'étude plus approfondie du système, il serait opportun de démontrer brièvement la place de l'or-

dinateur à travers la démarche méthodologique ci-dessus. Au premier niveau, le niveau potentiel, il n'y a place que pour la réflexion et le raisonnement. Nous avons déjà démontré l'importance de ces opérations, qui s'avèrent une condition "sine qua non" de la bonne fin de la recherche. Par contre au second stade, c'est la qualité du processus de cueillette des données qui importe. Le choix des données doit correspondre à la recherche en cause. Il est à conseiller de ne pas essayer de chercher des variables trop hétéroclites, qui encombrant l'analyse sans y apporter de dimension nouvelle mais qui soient théoriquement légitimes. Il ne s'agit pas non plus d'être rigoriste, mais de réfléchir. Si les données sont considérables il est préférable de les rentrer sur bande magnétique et de former une banque de données que l'on pourra facilement modifier, augmenter et même échanger. Pour un petit nombre de données la carte perforée demeure un médium d'entrée peu coûteux et fonctionnel.

Une fois les données possédées, on aborde l'étape du traitement analytique. C'est précisément à ce niveau que le système S.E.I.G.E.G. entre en fonction en mettant au service de l'utilisateur une batterie de programmes, capable de lui fournir des résultats susceptibles de concrétiser et de vérifier ses hypothèses. A la suite du traitement informatique, l'interprétation des résultats et les commentaires sont traduits sous forme de rapports prêts à la diffusion.

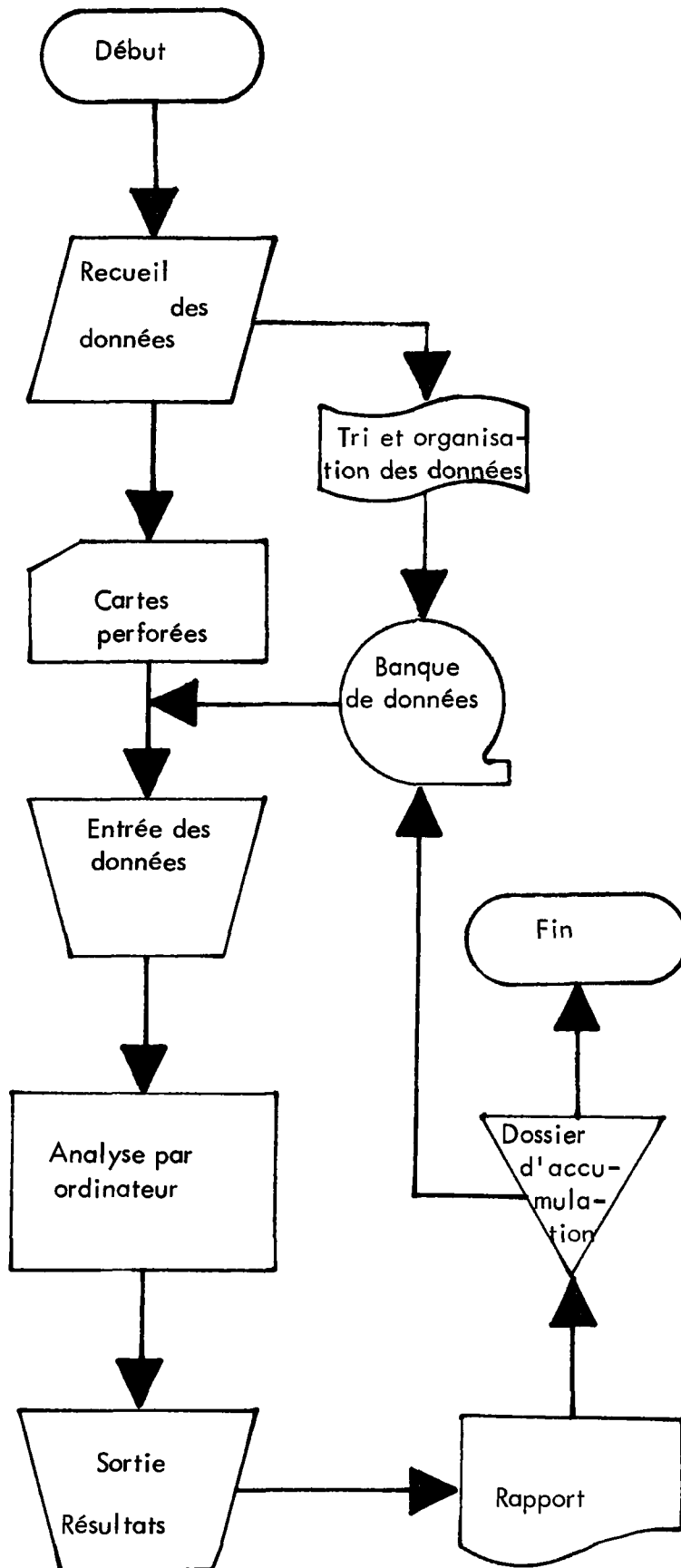
L'amoncellement des recherches et des écrits sur un sujet forme enfin un "dossier d'accumulation" mis en librairie (ordinateur) soit par le biais de l'accès direct, soit sur disque ou autre médium (tambour, cellules à haute performance etc.). Il sera ainsi à la disposition des intéressés. La figure 3 illustre à propos cet organigramme méthodologico-informatique.

1.2) L'ESPACE QUANTIFIE: LES NOMBRES ET LEUR SIGNIFICATION

Mais plutôt que sur le niveau potentiel qui a d'ailleurs déjà été expliqué, quoique de façon moins formelle et en termes légèrement différents par plusieurs géographes, parmi lesquels on retrouve aussi bien les noms de Pierre George (1970) et Jean Labasse (1971) que ceux des jeunes géographes plus familiarisés avec les méthodes informatiques comme J.P. Ferrier (1972) et J.B. Racine (1971, 1972), c'est sur les paliers subséquents que nous voulons attirer ici l'attention: celui de la nature des données quantitatives géographiques par exemple, qui requiert une attention particulière en ce qui concerne les différents types d'échelle de mesure, la classification des données brutes et finalement la mise en oeuvre sous forme de tableaux ou de matrices informatiques directement utilisables par l'ordinateur.

Ainsi, les données brutes et les processus de mesure font appel à quatre différents types d'échelle de calibration quantitative, qui sont d'ordre nominal, ordinal, d'intervalle et de rapport.

FIGURE 3



1.2 a) L'échelle nominale

L'échelle nominale est employée pour classifier des objets selon un code défini, tout en assurant l'égalité des différentes caractéristiques codifiées. A titre d'exemple prenons le cas de 50 villes ayant une population totale comprise entre 1,000 et 100,000 habitants. Les villes qui comptent une population intermédiaire, entre 1,000 et 10,000 habitants sont codées 1, les villes entre 10,000 et 20,000 sont codées 2, etc... Les codes représentent une classification. Remarquons immédiatement qu'ils pourraient aussi être définis par des caractères alphabétiques. Il est important de noter en effet que les codes en eux-mêmes ne représentent aucune valeur, c'est-à-dire qu'on peut substituer 1 à 3 ou vice-versa sans changer la valeur de l'intervalle de classe. Cette échelle permet d'effectuer par la suite des listes et des opérations numériques tel que le calcul du nombre respectif de cas dans chaque classe et/ou de les transformer sous forme de pourcentage et/ou de ramasser l'information sous forme de présence (1) - absence (0).

1.2 b) L'échelle ordinale ou le rang

On emploie cette échelle pour situer dans une hiérarchie de rangs, un attribut quelconque d'un objet ou d'un morceau d'espace. Lorsqu'il y a notion de comparaison tel que plus grand ou plus petit, l'échelle ordinale est appropriée pour déterminer la hiérarchie mais elle ne dit rien de la valeur de cette différence. Si

l'on reprend l'exemple précédent au sujet des 50 villes, et qu'on les ordonne de façon croissante à partir de la moins peuplée, on aura donc une échelle calibrée de 1 à 50 comprenant 50 individus. Avec cette échelle, la médiane est toujours la valeur située au centre de la distribution. Ainsi, dans notre exemple la médiane est de 25 villes.

1.2 c) L'échelle d'intervalle

Celle-ci est caractérisée par l'égalité des intervalles, sans qu'il soit nécessaire que le point de référence soit égal à 0. L'écart-type et le coefficient de corrélation sont des exemples courants en géographie humaine. Les différents types d'échelle de température tel Fahrenheit et Celsius sont aussi des exemples classiques de ce procédé de mensuration.

1.2 d) L'échelle des données relatives

C'est le plus versatile des procédés de mesure. Son utilisation se réfère à l'égalité des rapports en ce qui concerne l'unité observée. La différence d'avec l'intervalle existe du fait qu'elle se réfère à un 0 absolu comme base. A titre d'exemple les notions d'élévation et de distance représentent fidèlement la nature de cette échelle objective et fort utile en géographie quantitative. Plusieurs autres exemples peuvent être tirés de la géographie humaine. La densité, le revenu per capita, le rendement à

l'acre sont autant de points d'intérêt utiles à la détermination d'une structure organisée.

Les catégories de mesure déterminées, il est alors important de pouvoir évaluer et classifier les données afin d'éviter des erreurs qui pourraient être néfastes au moment de tirer des conclusions d'une recherche.

1.3) L'EVALUATION ET LA CLASSIFICATION DES DONNEES BRUTES

Nous empruntons la classification suivante au géologue W.C. Krumbein et au statisticien Franklin A. Graybill dans leur excellent livre sur les modèles statistiques en géologie. Son originalité repose sur leur définition de la nature systématique des nombres.

La distinction est faite en effet entre les systèmes numériques ouverts et les systèmes numériques fermés. Regardons de plus près leurs particularités respectives.

1.3 a) Système numérique fermé

Plusieurs types de données rencontrées auparavant peuvent être exprimées en pourcentage. Considérant que la somme des composantes en pourcentage est toujours égale à 100% il est évident que si l'on modifie une constituante A en l'augmentant de x (exemple 10%) il faudra diminuer une constituante B de $100 - x$ pour garder le total de 100%.

A = 25%	A - x: 25 - 10: 15%
B = 25%	B + x: 25 + 10: 35%
C = 25%	----- 25%
D = 25%	----- 25%
Total: 100%	Total: 100%

Il appert que si on transpose cette notion à plus d'une variable il est possible de former un tableau dont les rangées ont une somme de 100%. De par la définition même du système numérique fermé la somme des moyennes de chaque colonne se devra d'être égale à chaque rangée c'est-à-dire que l'on est en présence d'un référentiel **général** auquel se rapporte la nature même des nombres (cf. tableau 1).

TABLEAU 1

SYSTEME NUMERIQUE FERME: L'UTILISATION
DU SOL DE QUATRE VILLES FICTIVES

Villes	Variables			Espace total
	% de terrain résidentiel	% de terrain industriel	% de terrain vacant	
A	30	30	40	100
B	20	20	60	100
C	80	10	10	100
D	70	20	10	100
Moyennes	50%	20%	30%	100%

(référentiel général)

Etant donné le fait que la somme des rangées et celle des moyennes égalent 100%, il est permis de supposer qu'il existe des contraintes restrictives entre les différentes variables choisies. Tel n'est pas le cas dans le système numérique ouvert.

1.3 b) Système numérique ouvert

Celui-ci va à l'encontre du précédent. L'égalité au niveau du référentiel ne tient plus et les prérogatives telles que les notions d'association et de commutation systémique ne sont plus évocables. En voici un exemple:

TABLEAU 2

SYSTEME NUMERIQUE OUVERT: VARIABLES SOCIO-ECONOMIQUES

DE QUATRE VILLES FICTIVES

Villes	Variables			
	Nombre de services commerciaux en milliers	Population en milliers	Revenu moyen en milliers	Total
A	2	500	8	510
B	1	250	1	252
C	3	750	3	756
D	4	1000	4	1008
Moyennes	2.5	625	4	631.5

Cependant ce système numérique ouvert peut être converti en système numérique fermé par la conversion des éléments en valeurs relatives.

Il n'est cependant pas facile de savoir si on doit utiliser dans une analyse le système numérique ouvert plutôt que le système numérique fermé. Il semble que ceci demeure à la discrétion de l'utilisateur. Certains préfèrent travailler avec un seul système. Jean Labasse par exemple (1971) a nettement indiqué sa préférence pour les données relatives. D'autres utilisent les deux et l'expérience démontre qu'un mélange des deux genres ouvre des perspectives intéressantes et constitue une matrice d'entrée originale, d'autant plus efficace que très souvent la signification géographique des données relatives doit être contrôlée par le niveau de valeurs absolues auxquelles elles correspondent (Thibodeau, 1972).

En effet il démontre avec un exemple que si le nombre d'établissement d'un type donné passe de 1000 à 1100 entre 1962 et 1967 dans la ville de Montréal, le taux de croissance aurait été de 10%. Par contre si pour pour le même type et établissement le nombre passe de 4 à 6 durant la même période dans la zone périphérique le taux de croissance serait de 50%. Une comparaison prématurée des taux de croissance de 10% et 50% nous portent à dire qu'il y a un mouvement important d'exurbanisation de ce genre d'établissement alors que c'est totalement l'inverse.

Pour éviter des interprétations du genre il faut pondérer pour les valeurs absolues dans chacun des espaces et établir des inter-relations. En reprenant l'exemple ci-haut avec l'aide des poids relatifs on obtient:

Pour 1962: Montréal = 1000 entreprises

Périphérie = 4 entreprises

Poids relatif
de Montréal = 99,6%

Pour 1967: Montréal = 1100 entreprises

Périphérie = 6 entreprises

Poids relatif
de Montréal = 99,4%

Source: Thibodeau 1971 p. 28

Contrairement à la conclusion précédente, ces poids relatifs nous mènent à conclure qu'il n'y a en effet aucun mouvement significatif vers la périphérie.

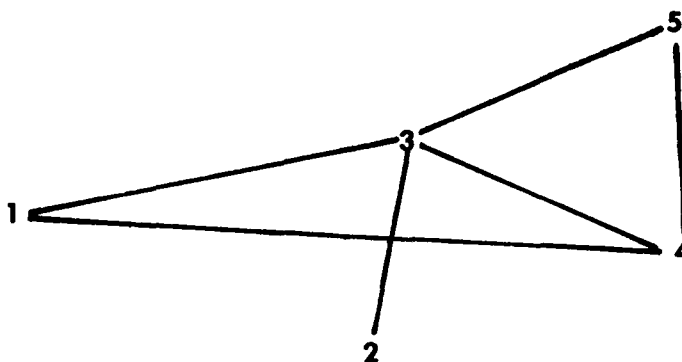
Donc il est important de faire un choix judicieux de ses indices et variables afin qu'il n'y ait pas de distorsion inutile lors de la manipulation mathématique de ces données présentées sous forme de matrices géographiques.

1.4) LES TYPES DE MATRICE GEOGRAPHIQUE

Comment l'ordinateur peut-il manipuler les données géographiques? La réponse à cette question définit par elle-même la nature même de la matrice géographique. Le problème revient en fait à traduire sous forme de chiffres les caractéristiques spatiales, seul langage compréhensible pour l'ordinateur. En géographie la matrice est donc la représentation chiffrée, sous forme de tableaux à double entrée, de phénomènes spatiaux interactifs, de type structuraux, relationnels, dynamiques ou perceptifs (Ress, 1971).

1.4 a) La matrice descriptive graphique ou de convexité

Ce type de matrice que l'on peut qualifier de post-graphique reconstitue graphiquement les liens d'un graphique. Les composantes de cette matrice sont le zéro (0) et le un (1). Le zéro signifie qu'il n'existe aucun lien, par opposition au un qui dénote une relation quelconque. A noter que le un n'a qu'une valeur nominale et qu'il n'indique pas la valeur du lien. A titre d'exemple prenons cinq (5) villes situées sur un territoire formant le réseau routier suivant.



Il est possible de convertir ce graphe sous forme de matrice d'interaction en indiquant la présence (1) ou l'absence (0) du lien. On aura donc le tableau suivant.

Villes	1	2	3	4	5
1	0	0	1	1	0
2	0	0	1	0	0
3	1	1	0	1	1
4	1	0	1	0	1
5	0	0	1	1	0

La diagonale de la matrice est composée de 0 puisqu'il n'existe aucun lien entre une ville et elle-même.

Cette matrice de convexité peut être très utile dans les cas d'une pénurie de données quantitatives sur des sujets complexes tel que le commerce de détail ou les domaines hiérarchisés. Dans la plupart des cas il y a possibilité de faire une matrice présence-absence qui informe au moins le chercheur d'une manière partielle. Cependant ce genre de matrice descriptive graphique est surtout employée dans les études de transport et de réseaux. Par le fait qu'elle ne présente que le lien sans l'assortir d'une valeur de référence, elle demeure unidimensionnelle et offre un net désavantage par rapport à la vraie matrice d'information spatiale.

1.4 b) La matrice d'information spatiale

Ce tableau soumis au traitement informatique comporte deux entrées: les colonnes et les rangées. Les colonnes accueillent

Ces variables comprennent aussi bien des faits mesurables qu'observables, d'ordre statique tel que la population totale, la distance, le réseau, ou dynamique comme la distance-temps, le nombre de mutation annuelle, etc...

A chaque observation i ($i= 1$ à n) correspond une série d'attributs j ($j= 1$ à m). Cette matrice à double entrée peut être lue horizontalement ou verticalement. Elle est la somme de plusieurs vecteurs (colonnes ou rangées). Comme l'a souligné J.B. Racine (1971), chaque vecteur colonne est donc transformable en autant de cartes géographiques. De là vient la force de cette matrice, due au fait qu'elle contient autant de dimensions que de variables. C'est pourquoi J.B. Racine l'a qualifiée "d'image centrale" de l'étude de toute information quantifiée.

L'analyse multivariée exige très souvent que le nombre de variables soit égal ou plus petit que celui des observations, $m < n$. L'expérience démontre qu'un rapport de 2/3: var./observ. offre encore les meilleurs résultats. Ceci n'est cependant pas une règle. (Voir tableau 3 comme exemple de matrice d'information spatiale)

Quoique ce genre de matrice révèle une plus grande flexibilité que la précédente, elle a l'inconvénient d'être statique, c'est-à-dire de décrire un point précis dans le temps, un état plutôt qu'une réalité. Si les données qui rentrent dans la matrice datent de 1972, les résultats n'auront de valeur que pour 1972 et les comparaisons par rapport à d'autres recherches à des intervalles différents ne seront pas équipotentielles. Pour contourner la dif-

TABLEAU 3

MATRICE D'INFORMATION SPATIALE

OBSERVATIONS	VARIABLE 1	VARIABLE 2	VARIABLE 3	VARIABLE 4	VARIABLE 5	VARIABLE 6	VARIABLE 7	VARIABLE 8	VARIABLE 9	VARIABLE 10	VARIABLE 11	VARIABLE 12
1	35.100	34.200	57.200	94.000	4.200	0.200	1.700	1.300	3.400	1.800	0.800	1.500
2	10.100	38.500	50.700	82.800	2.400	0.0	14.800	0.0	1.100	6.800	1.700	1.200
3	17.500	29.900	28.200	99.400	0.800	0.0	0.0	22.500	1.200	1.900	15.500	0.733
4	6.500	15.700	48.200	99.700	0.300	0.0	0.0	0.0	1.800	1.200	7.300	1.000
5	30.400	21.700	63.500	79.500	0.300	0.900	18.900	0.900	2.800	4.000	0.200	2.000
6	17.800	25.800	62.400	98.500	3.000	0.9	0.200	0.0	0.0	0.0	0.500	0.0
7	58.200	34.400	78.500	97.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	99.200	34.000	78.500	97.000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	99.200	21.200	59.000	85.400	0.0	1.000	0.0	0.0	0.800	1.500	48.700	1.400
10	29.900	24.100	49.300	74.000	15.400	0.400	0.0	2.200	4.200	12.100	1.300	0.3
11	75.900	32.300	52.500	64.000	12.000	0.0	19.400	0.500	7.700	4.400	6.900	4.600
12	12.400	12.400	13.900	60.400	32.200	1.300	6.700	1.700	3.800	4.300	3.500	1.733
13	99.200	7.000	63.500	31.900	49.600	8.133	10.400	1.000	3.900	2.503	45.500	2.100
14	20.400	12.500	53.100	34.203	1.100	0.0	65.100	30.300	0.600	9.000	6.000	2.500
15	86.900	6.200	61.200	49.300	42.100	2.633	6.000	2.700	6.500	0.600	0.0	3.330
16	96.900	15.400	78.100	59.200	0.800	0.0	40.000	0.600	12.400	5.600	11.400	3.100
17	46.700	14.500	40.600	99.903	0.0	0.0	0.0	4.000	1.900	0.600	0.0	3.100
18	8.700	17.700	61.203	93.100	0.0	0.800	0.0	13.300	1.400	4.300	0.0	0.233
19	13.400	15.700	58.200	97.700	0.0	0.0	0.0	25.000	3.200	4.200	8.200	0.233
20	12.400	32.400	37.100	89.700	12.000	0.0	18.800	0.0	4.000	1.400	2.800	1.700
21	12.400	24.700	37.100	84.300	6.000	0.0	7.700	2.400	4.900	1.500	1.100	1.700
22	79.200	10.900	59.200	88.400	8.300	0.630	2.300	6.000	6.200	3.500	8.800	3.400

ficulté la matrice de transition introduit le dynamisme par l'intermédiaire de la probabilité.

1.4 c) La matrice de probabilité prégraphique

Cette matrice décrit les liaisons probabilistes entre les attributs d'une variable et la position de cette variable dans une séquence temporelle. On la retrouve à la base de toutes les études stochastiques dont la préoccupation est de fournir une base prévisionnelle. La matrice P_{ij} (figure 5) est un exemple de matrice de transition probabiliste. Ses éléments constitutants sont indépendants de la dimension temps et réfèrent les probabilités à un temps fixe.

FIGURE 5 - FORME GENERALE DE LA MATRICE DES PROBABILITES

	1	2	3	4	5 . . . j m
1	P_{11}	P_{12}		P_{14}	
2		P_{22}			
3			P_{33}		
$P_{ij} = j$	-----				P_{ij}
.					
.					
.					
m					

Chaque élément de la matrice mesure la probabilité d'être déplacé d'un état i vers un autre état j dans un intervalle de temps t .

Donc pour chaque transition il est démontré d'affirmer que les probabilités satisfont les conditions suivantes.

$$p_{ij} \geq 0 \quad j = 1, 2, \dots, m$$

$$\sum_{j=1}^m p_{ij} = 1 \quad i = 1, 2, \dots, m$$

$$j = 1 \quad i = j = m$$

Prenons l'exemple des probabilités de changement de l'utilisation du sol. Voici un exemple de la ville X entre 1962-1972.

	Terrain résiden- ciel	Terrain commer- cial	Terrain indus- triel	Terrain vacant	
Terrain résidentiel	$p_{11} = .3$	$p_{12} = .2$	$p_{13} = .2$	$p_{14} = .3$	$\sum_{j=1}^4 p_{1j} = 1 \quad i = 1$
Terrain commercial	$p_{21} = .0$	$p_{22} = .5$	$p_{23} = .5$	$p_{24} = .0$	$\sum_{j=1}^4 p_{2j} = 1 \quad i = 2$
Terrain industriel	$p_{31} = .3$	$p_{32} = .1$	$p_{33} = .5$	$p_{34} = .1$.
Terrain vacant	$p_{41} = .1$	$p_{42} = .1$	$p_{43} = .7$	$p_{44} = .1$.

Littéralement $p_{12} = .2$ signifie que la probabilité que le terrain résidentiel devienne terrain commercial est de .2 sur 1 ou 2 sur 10 et $p_{11} = .3$ indique la probabilité que le terrain résidentiel demeure terrain résidentiel est de .3 sur 1 ou 3 sur 10.

Le traitement de cette matrice par différents processus stochastiques est un outil indispensable pour un géographe, qui se serait donné pour but de connaître l'effet sur le paysage de la modification de l'un des paramètres constituant du système. Pour l'instant nous ne pousserons pourtant pas plus loin le traitement de cette question, nous réservant d'y revenir au niveau de nos re-

cherches de doctorat.

1.5) LA TRANSFORMATION DE LA MATRICE CHRONO-SPATIALE

A partir de la matrice d'information spatiale le chercheur a le choix de travailler soit avec les attributs sur la structure organisée de l'espace, soit avec les éléments spatiaux par l'intermédiaire de la transposition et/ou de la standardisation. Il peut enfin utiliser successivement les deux techniques.

1.5 a) L'intérêt de la transposition

L'obtention de la matrice transposée (tableau 4) résulte du remplacement de chaque rangée de la matrice d'information spatiale par sa colonne correspondante. Il est à noter que si la matrice originale est rectangulaire, avec un plus grand nombre d'observations que de variables, la transposée aura moins de rangées que de colonnes ce qui deviendra une contrainte limitative pour effectuer certaines analyses multivariées telle que l'analyse factorielle ou la régression multiple. D'autre part la matrice transposée sera d'une grande utilité dans l'analyse des chaînons d'association statistique maximale, ou "linkage analysis", à laquelle nous nous attarderons plus en détail un peu plus loin. Les algorithmes de classification multivariée ordonnent et réduisent la matrice d'information transposée en dégagant les principaux liens structuraux. La réduction de l'information crée une classification typologique

TABLEAU 4

MATRICE TRANSPOSEE

OBSERVATIONS	VARIABLE 1	VARIABLE 2	VARIABLE 3	VARIABLE 4	VARIABLE 5	VARIABLE 5	VARIABLE 5	VARIABLE 7	VARIABLE 8	VARIABLE 9	VARIABLE 10	VARIABLE 11	VARIABLE 12
1	6	B	C	C	C	C	C	C	D	E	J	L	L
2	U	D	N	U	A	A	A	M	L	E	A	P	A
3	C	7	7	7	T	T	T	T	S	E	C	P	P
4	200	50-700	28-200	48-200	93-500	62-800	48-800	95-100	85-000	85-000	76-000	L	P
5	4-200	2-400	0-400	0-300	7-200	3-000	1-000	3-000	16-400	16-400	15-900	L	P
6	0-200	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	C	A
7	1-700	14-800	0-0	16-900	0-0	0-300	1-000	0-0	1-000	0-500	0-400	C	A
8	3-300	0-0	22-500	25-000	0-900	1-000	2-200	0-500	0-0	0-0	0-800	C	A
9	3-400	1-100	1-200	1-400	0-900	1-000	2-400	0-500	0-0	2-200	0-400	C	A
10	0-800	0-800	1-200	7-800	7-800	2-425	12-700	0-300	0-500	12-700	4-400	C	A
11	1-800	0-800	15-400	7-300	0-200	1-300	48-700	0-0	0-500	1-300	6-900	C	A
12	1-500	1-200	0-700	1-000	2-000	0-0	4-600	1-400	0-0	0-0	8-600	C	A
13	35-100	10-100	17-500	4-500	30-400	17-820	29-300	58-200	29-300	99-200	29-900	75-920	12-432
14	39-200	38-500	29-900	15-700	21-700	23-800	26-000	37-400	26-000	21-200	26-100	32-500	12-900
15	57-200	50-700	28-200	48-200	93-500	62-800	48-800	95-100	85-000	85-000	76-000	68-600	60-400
16	4-200	2-400	0-400	0-300	7-200	3-000	1-000	3-000	16-400	16-400	15-900	12-000	32-200
17	0-200	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	0-0	1-353
18	1-700	14-800	0-0	16-900	0-0	0-300	1-000	0-0	1-000	0-500	0-400	0-0	8-700
19	3-300	0-0	22-500	25-000	0-900	1-000	2-200	0-500	0-0	2-200	0-400	0-0	19-800
20	3-400	1-100	1-200	1-400	0-900	1-000	2-400	0-500	0-0	2-200	0-400	0-0	19-800
21	0-800	0-800	1-200	7-800	7-800	2-425	12-700	0-300	0-500	12-700	4-400	0-0	2-500
22	1-800	0-800	15-400	7-300	0-200	1-300	48-700	0-0	0-500	1-300	6-900	0-0	2-500
23	1-500	1-200	0-700	1-000	2-000	0-0	4-600	1-400	0-0	0-0	8-600	0-0	2-100

OBSERVATIONS	VARIABLE 13	VARIABLE 14	VARIABLE 15	VARIABLE 16	VARIABLE 17	VARIABLE 18	VARIABLE 19	VARIABLE 20	VARIABLE 21	VARIABLE 22
1	L	L	L	M	P	S	S	V	V	V
2	R	R	D	O	R	F	E	C	M	L
3	Y	Y	G	F	V	E	E	C	M	L
4			C	E	T	C	A	O	M	L
5			U	L	L	C	A	O	M	L
6			U	D	L	T	Y	Z	B	L
7			U	D	L	T	Y	Z	B	L
8			U	D	L	T	Y	Z	B	L
9			U	D	L	T	Y	Z	B	L
10			U	D	L	T	Y	Z	B	L
11			U	D	L	T	Y	Z	B	L
12			U	D	L	T	Y	Z	B	L
13			U	D	L	T	Y	Z	B	L
14			U	D	L	T	Y	Z	B	L
15			U	D	L	T	Y	Z	B	L
16			U	D	L	T	Y	Z	B	L
17			U	D	L	T	Y	Z	B	L
18			U	D	L	T	Y	Z	B	L
19			U	D	L	T	Y	Z	B	L
20			U	D	L	T	Y	Z	B	L
21			U	D	L	T	Y	Z	B	L
22			U	D	L	T	Y	Z	B	L
23			U	D	L	T	Y	Z	B	L
24			U	D	L	T	Y	Z	B	L
25			U	D	L	T	Y	Z	B	L
26			U	D	L	T	Y	Z	B	L
27			U	D	L	T	Y	Z	B	L
28			U	D	L	T	Y	Z	B	L
29			U	D	L	T	Y	Z	B	L
30			U	D	L	T	Y	Z	B	L
31			U	D	L	T	Y	Z	B	L
32			U	D	L	T	Y	Z	B	L
33			U	D	L	T	Y	Z	B	L
34			U	D	L	T	Y	Z	B	L
35			U	D	L	T	Y	Z	B	L
36			U	D	L	T	Y	Z	B	L
37			U	D	L	T	Y	Z	B	L
38			U	D	L	T	Y	Z	B	L
39			U	D	L	T	Y	Z	B	L
40			U	D	L	T	Y	Z	B	L
41			U	D	L	T	Y	Z	B	L
42			U	D	L	T	Y	Z	B	L
43			U	D	L	T	Y	Z	B	L
44			U	D	L	T	Y	Z	B	L
45			U	D	L	T	Y	Z	B	L
46			U	D	L	T	Y	Z	B	L
47			U	D	L	T	Y	Z	B	L
48			U	D	L	T	Y	Z	B	L
49			U	D	L	T	Y	Z	B	L
50			U	D	L	T	Y	Z	B	L
51			U	D	L	T	Y	Z	B	L
52			U	D	L	T	Y	Z	B	L
53			U	D	L	T	Y	Z	B	L
54			U	D	L	T	Y	Z	B	L
55			U	D	L	T	Y	Z	B	L
56			U	D	L	T	Y	Z	B	L
57			U	D	L	T	Y	Z	B	L
58			U	D	L	T	Y	Z	B	L
59			U	D	L	T	Y	Z	B	L
60			U	D	L	T	Y	Z	B	L
61			U	D	L	T	Y	Z	B	L
62			U	D	L	T	Y	Z	B	L
63			U	D	L	T	Y	Z	B	L
64			U	D	L	T	Y	Z	B	L
65			U	D	L	T	Y	Z	B	L
66			U	D	L	T	Y	Z	B	L
67			U	D	L	T	Y	Z	B	L
68			U	D	L	T	Y	Z	B	L
69			U	D	L	T	Y	Z	B	L
70			U	D	L	T	Y	Z	B	L
71			U	D	L	T	Y	Z	B	L
72			U	D	L	T	Y	Z	B	L
73			U	D	L	T	Y	Z	B	L
74			U	D	L	T	Y	Z	B	L
75			U	D	L	T	Y	Z	B	L
76			U	D	L	T	Y	Z	B	L
77			U	D	L	T	Y	Z	B	L
78			U	D	L	T	Y	Z	B	L
79			U	D	L	T	Y	Z	B	L
80			U	D	L	T	Y	Z	B	L
81			U	D	L	T	Y	Z	B	L
82			U	D	L	T	Y	Z	B	L
83			U	D	L	T	Y	Z	B	L
84			U	D	L	T	Y	Z	B	L
85			U	D	L	T	Y	Z	B	L
86			U	D	L	T	Y	Z	B	L
87			U	D	L	T	Y	Z	B	L
88			U	D	L	T	Y	Z	B	L
89			U	D	L	T	Y	Z	B	L
90			U	D	L	T	Y	Z	B	L
91			U	D	L	T	Y	Z	B	L
92			U	D	L	T	Y	Z	B	L
93			U	D	L	T	Y	Z	B	L
94			U	D	L	T	Y	Z	B	L
95			U	D	L	T	Y	Z	B	L
96			U	D	L	T	Y	Z	B	L
97			U	D	L	T	Y	Z	B	L
98			U	D	L	T	Y	Z	B	L
99			U	D	L	T	Y	Z	B	L
100			U	D	L	T	Y	Z	B	L
101			U	D	L	T	Y	Z	B	L
102			U	D	L	T	Y	Z	B	L
103			U	D	L	T	Y	Z	B	L
104			U	D	L	T	Y	Z	B	L
105			U	D	L	T	Y	Z	B	L
106			U	D	L	T	Y	Z	B	L
107			U	D	L	T	Y	Z	B	L
108			U	D	L	T	Y	Z	B	L
109			U	D	L	T	Y	Z	B	L
110			U	D	L	T	Y	Z	B	L
111			U	D	L	T	Y	Z	B	L
112			U	D	L	T	Y	Z	B	L
113			U	D	L	T	Y	Z	B	L
114			U	D	L	T	Y	Z	B	L
115			U	D	L	T	Y	Z	B	L
116			U	D	L	T	Y	Z	B	L
117			U	D	L	T	Y	Z	B	L
118			U	D	L	T	Y	Z	B	L
119			U	D	L	T	Y	Z	B	L
120			U	D	L	T	Y	Z	B	L
121			U	D	L	T	Y	Z	B	L
122			U	D	L	T	Y	Z	B	L
123			U	D	L	T	Y	Z	B	L
124			U	D	L	T	Y	Z	B	L
125			U	D	L	T	Y	Z	B	L
126			U	D	L	T	Y	Z	B	L
127			U	D	L	T	Y	Z	B	L
128			U	D	L	T	Y	Z	B	L
129			U	D	L	T	Y	Z	B	L
130			U	D	L	T	Y	Z	B	L
131			U	D	L	T	Y	Z	B	L
132			U	D	L	T	Y	Z	B	L
133			U	D	L	T	Y	Z	B	L
134			U	D	L	T	Y	Z	B	L
135			U	D	L	T	Y	Z	B	L
136			U	D	L	T	Y	Z	B	L
137			U	D	L	T	Y	Z	B	L
138			U	D	L	T	Y	Z	B	L
139			U	D	L	T	Y	Z	B	L
140			U	D	L	T	Y	Z	B	L
141			U	D	L	T	Y	Z	B	L
142			U	D	L	T	Y	Z	B	L
143			U	D	L	T	Y	Z	B	L
144			U	D	L	T	Y	Z	B	L
145			U	D						

et renseigne sur les liens de la structure spatiale.

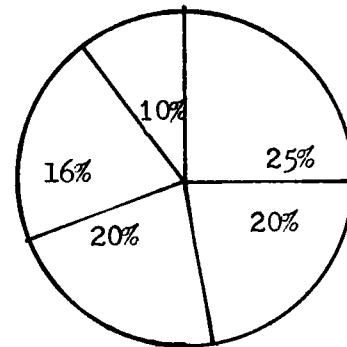
$$\begin{array}{c}
 X_{nm} \\
 \left| \begin{array}{ccc} 4 & 3 & 2 \\ 1 & 2 & 6 \\ 2 & 1 & 4 \\ 0 & 8 & 2 \end{array} \right| \\
 n=4
 \end{array}
 \quad
 \begin{array}{c}
 m=3 \\
 X_{nm}^T \\
 \left| \begin{array}{ccc} 4 & 1 & 2 & 0 \\ 3 & 2 & 1 & 8 \\ 2 & 6 & 4 & 2 \end{array} \right| \\
 n=3
 \end{array}
 \quad
 \begin{array}{c}
 m=4
 \end{array}$$

La matrice de transposition est en quelque sorte une représentation verticale ponctuelle d'une composition unitaire de l'espace, tout comme l'est à sa manière n'importe quel graphique, tel un cercle, divisé par pourcentages qui permet de saisir d'un seul coup d'oeil les composantes relatives de l'unité d'espace (figure 6).

FIGURE 6

QUEBEC*

Terrain résidentiel	20%
Terrain industriel	10%
Terrain transport	15%
Terrain institutionnel	25%
Terrain commercial	20%



Pour le géographe cependant, la transposition n'a pas simplement pour fin d'être une nouvelle matrice d'information. Elle devient un instrument de travail au cours des diverses opérations matricielles.

Mais l'information recueillie, il ne s'agit pourtant pas de se lancer immédiatement dans l'analyse proprement dite. Les

* Les données sont fictives pour les besoins de l'exemple.

quantités élaborées doivent être comparables les unes aux autres. Ceci devient possible après la standardisation et si nécessaire après la normalisation des données.

1.5 b) L'intérêt de la standardisation

Les données relevées varient considérablement dans la plupart des cas, surtout si elles ne sont pas relatives. Les valeurs de taille, de revenu, de surface deviennent difficilement comparables à des indices, à des rapports ou à des pourcentages. La question est donc de ramener sur une base comparative les différentes variables et de trouver un indicateur de variabilité.

Cette mesure de variabilité universellement pratiquée se fait au moyen de l'écart-type, défini comme étant la racine carrée de la moyenne arithmétique des carrés des déviations. Mathématiquement cette définition prend la forme suivante.

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}}$$

n = nombre d'observations

m = nombre de variables

$j = 1$ à m

$i = 1$ à n

\bar{x}_j = moyenne arithmétique de la variable j

x_{ij} = valeur de l'observation i sur l'attribut j

Le procédé de standardisation comme tel est fort simple. Il consiste à remplacer chaque valeur par le rapport existant entre les deux écarts fondamentaux qui caractérisent sa position dans la

distribution dont elle est issue: l'écart à la moyenne de la distribution et l'écart-type de la distribution, ou si l'on préfère:

$$\text{Donnée standardisée} = \frac{\text{écart à la moyenne}}{\text{écart-type}}$$

$$\frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

où x_{ij} = la valeur de l'observation (i) de l'attribut (j)

\bar{x}_j = moyenne de la variable j

σ_j = écart-type de la variable j

Voici le calcul de la standardisation pour la première variable de l'utilisation du sol utilisée par J.B. Racine dans une étude de la banlieue de Montréal. Dans le tableau original les différentes variables ne sont pas directement comparables. En revanche, une fois standardisées, la comparaison structurelle des distributions est immédiate comme le montre le tableau (5) utilisant les distributions de l'importance du terrain urbanisé dans l'espace total et du terrain résidentiel dans le terrain urbanisé.

La standardisation a pour effet de ramener la distribution à une moyenne égale à zéro et une variance et un écart-type égaux à l'unité.

L'avantage de la standardisation c'est qu'elle permet une plus grande lisibilité par rapport à sa matrice originale. En effet d'un seul coup d'oeil on peut repérer les observations qui se situent au-dessus ou en-dessous de la moyenne et déterminer les plus importantes par rapport aux autres de la distribution.

Pour une petite matrice telle que 12×22 il n'existe pas de problème de lecture, mais la plupart du temps les matrices sont beaucoup plus vastes soit du côté des variables (m) soit du côté des observations (n). Tel est le cas pour une étude de la région urbaine de Montréal qui comprend en moyenne 200 à 300 unités d'observations et le recensement qui autorise à lui seul l'utilisation de près de 200 indicateurs structurels, relationnels, dynamiques et perspectifs. La matrice que nous avons travaillé avec J.B. Racine avait déjà 250 attributs, dont moins de 25 seulement étaient de type dynamique ou relationnel. Elle pouvait donc encore être considérablement élargie et le nombre de ses "moments perceptifs" se multiplier d'autant.

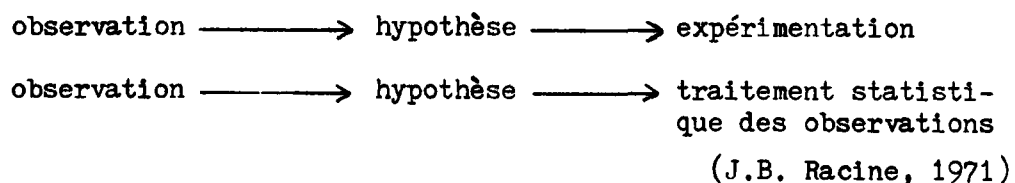
Cet intérêt est d'autant plus évident lorsque après un long processus d'analyse, il convient de tester le degré de "fraîcheur" et le degré de conformité existant entre les êtres mathématiques créés par les algorithmes mathématiques et les valeurs empiriques dont ils sont originellement tirés.

Par exemple la hiérarchie des poids locaux liés à chacune des composantes d'une analyse factorielle reflète-t-elle la hiérarchie des valeurs originales des attributs spatiaux qui ont permis l'identification et le libelle des facteurs? Dans ce cas un coup d'oeil à la matrice standardisée est évidemment beaucoup plus utile que le retour à l'analyse de la matrice originale sur laquelle les valeurs des attributs ne sont pas comparables entre elles.

L'emploi de la standardisation ne demande pas d'être justifié. Ce procédé a sa place dans la presque totalité des analyses statistiques et ne soustrait aucune information. Au contraire une distribution standardisée nous instruit parfois plus que la distribution brute mais malheureusement il n'en est pas de même pour la normalisation.

1.5 c) Le problème de la normalisation

Etant donné que le géographe ne peut pas déterminer à l'aide de critères qualitatifs (à moins une très forte connaissance du milieu étudié) les différences ou les corrélations spatiales significatives ou aléatoires, il doit avoir recours au traitement statistique de l'information, qui fournit des valeurs critiques d'acceptation (test de F, test de T,...), connues de tous les habitués des méthodes de la statistique. C'est dans cet esprit que le géographe tend à substituer, dans son schème méthodologique, le stade de l'expérimentation par celui du traitement statistique des observations.



Cependant le fait d'utiliser les tests d'inférence statistique, au lieu de se contenter des seules situations d'ordre des-

criptif impose au géographe la contrainte de la normalité. Les tests statistiques ont été élaborés sur la notion de la représentativité d'une population "normale". Il en résulte que si on veut utiliser ces mêmes tests en géographie, il faut que la distribution qui décrit l'espace soit également normale. Le postulat de base de tout test statistique est en effet l'hypothèse nulle. Cette hypothèse présuppose qu'il n'existe aucune différence significative entre l'échantillon choisi et la distribution totale, c'est-à-dire que les observations qu'on utilise dans l'analyse multivariée doivent être représentatives de l'espace total étudié. Découvrir une relation, une différence ou une corrélation statistiquement significative dans l'espace géographique revient à dire que l'on peut rejeter l'hypothèse nulle, c'est-à-dire que la différence ou la corrélation ne sont pas dues au hasard. En revanche lorsqu'on choisit de procéder par échantillonnage, la population sondée doit être comparable à la population totale. Dans ce cas l'hypothèse nulle doit pouvoir être acceptée.

Le problème de la normalisation se pose d'ailleurs à un niveau plus général: celui de l'emploi, dans presque toutes les méthodes statistiques subséquentes des notions de tendance centrale (telle la moyenne) et de variabilité (tel l'écart-type). Ces deux valeurs n'ont vraiment de signification à longue portée statistique que si elles représentent une population normale. Ces deux distributions peuvent être complètement différentes, se traduire par deux cartes complètement différentes aussi tout en ayant à peu près la

même moyenne et le même écart-type.

1.6) LES PROPRIETES DE LA COURBE NORMALE ET DE SES PARAMETRES CONSTITUANTS

La notion de "courbe normale" mérite quelques mots d'explication. C'est à la suite de nombreuses mesures, tant dans le domaine de la psychologie, de l'économie et des sciences qu'est apparu le phénomène "normal" de la répartition numérique et graphique. Cette répartition, lorsqu'on en trace la courbe suggère le profil d'une cloche.

Parallèlement à ces recherches, les mathématiciens Gauss et Laplace, indépendamment l'un de l'autre sont arrivés à formuler mathématiquement ce que différentes recherches sur la répartition des variations en grandeur de différents phénomènes avaient démontré empiriquement.

Il y aurait de nombreux points à soulever au sujet de la normalisation en géographie. Nous nous bornerons cependant à démontrer, à l'aide d'exemples les diverses caractéristiques de la courbe normale et leur intérêt, pour aboutir finalement aux techniques de la normalisation par l'ordinateur.

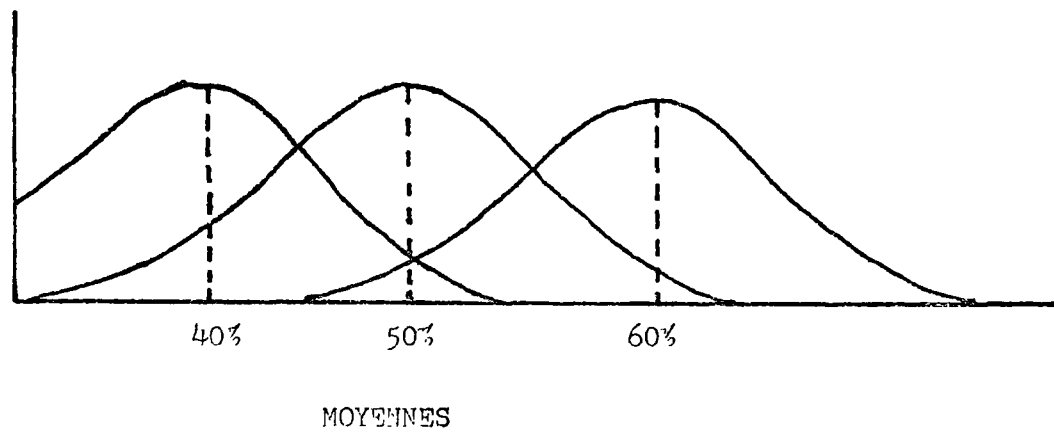
Considérons une courbe, initialement déterminée par les paramètres suivants: le nombre de cas dans l'échantillon, la moyenne et l'écart-type. Une variation quelconque de ces trois paramètres entraîne une modification soit au niveau de la forme soit au

niveau de la position de la courbe,

Prenons un exemple de distribution "normale" et étudions l'effet successif de la variation d'un des paramètres. Si on fait varier la moyenne et que l'on garde fixe l'écart-type et le nombre d'objets dans l'échantillon on constate que la courbe garde la même forme (en cloche) mais il y a déplacement horizontal de celle-ci vers la droite si la moyenne a été augmentée ou vers la gauche dans le cas d'une diminution.

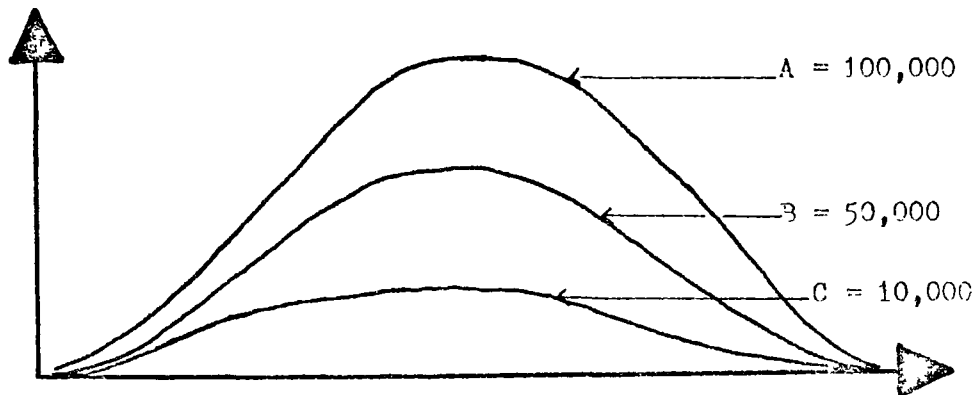
La figure (7) illustre à propos cette variation

COURBES NORMALES DE MOYENNES VARIABLES



Par contre si la moyenne et l'écart-type demeurent fixes et que la population de l'échantillon varie, on s'aperçoit que les extrêmes de la courbe demeurent identiques mais que la partie centrale subit une transformation verticale plus accentuée dans le cas d'un grand nombre dans l'échantillon et inversement pour un petit nombre. Les courbes A, B, et C de la figure (8) concrétisent notre idée à ce sujet.

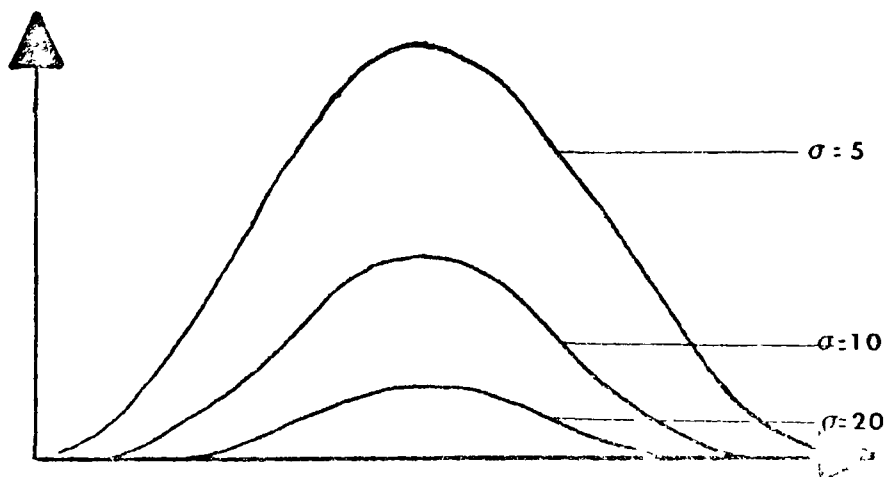
FIGURE 8

COURBES NORMALES DE DIFFERENTS ECHANTILLONS

Le plus grand type de déformation survient avec la variation de l'écart-type tout en gardant stable la moyenne et le nombre dans l'échantillon.

Une moyenne accompagnée d'un écart-type réduit signifie que les données sont relativement concentrées en opposition avec une variable dont la moyenne possède un très grand écart-type. Sur graphique ceci se traduit de la façon suivante:

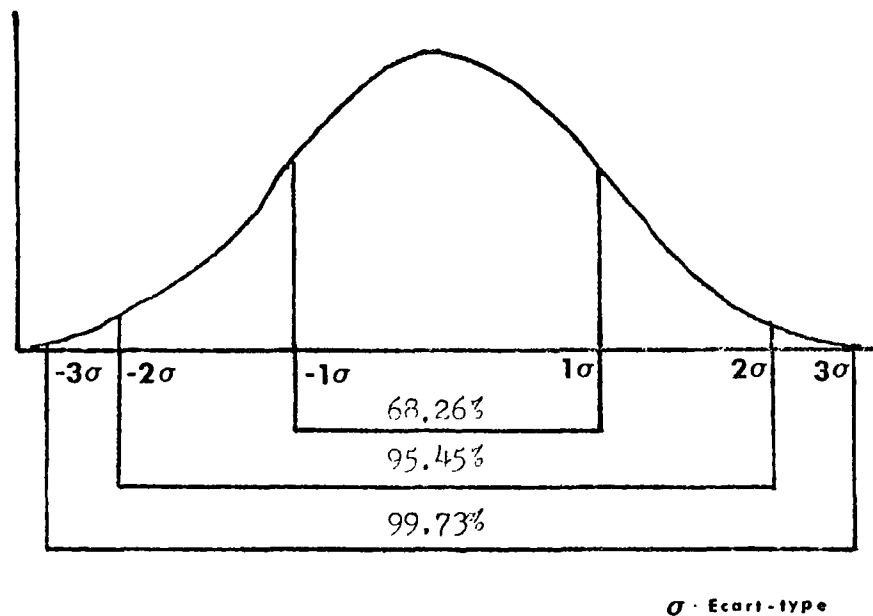
FIGURE 9

COURBES NORMALES DE DIFFERENTS ECART-TYPES

Dans le cas d'une distribution normale les points d'inflexion (c'est-à-dire le point où la pente qui allait en augmentant commence à diminuer) ont la valeur de $+1$ ou -1 écart-type. On démontre que la surface comprise entre $+1$ et -1 écart-type représente 68,26% de l'échantillon; entre -2 et $+2$ elle inclue 95,45% et finalement entre -3 et $+3$ elle couvre la quasi-totalité de la courbe avec 99,73%. (figure 10)

FIGURE 10

REPRESENTATION EN POURCENTAGE DE L'ECART-TYPE SOUS UNE COURBE NORMALE



Cette notion sera reprise plus loin lorsqu'il s'agira de cartographier les résultats en prenant comme unité de mesure et comme base de séparation en classe l'écart-type. Dans le cas présent il est considéré uniquement comme mesure de dispersion.

Maintenant que l'on est renseigné sur les principales composantes de la courbe normale, penchons nous quelques instants sur ses caractéristiques fondamentales. Un premier regard sur la forme de cette courbe permet de constater une symétrie par rapport à l'axe vertical médian. En d'autres mots si on trace une droite verticale de l'axe des x au point le plus élevé de la courbe (au centre) et que cette droite sert de charnière pour rabattre la moitié de la courbe sur l'autre, on découvre une coïncidence parfaite qui par le fait même témoigne de la symétrie et corrobore le fait que les surfaces des champs de dispersion au-dessus et en-dessous de la moyenne sont équivalentes.

Cette dernière observation permet d'affirmer que dans le cas de la normalité, la moyenne, le mode (qui représente le point le plus élevé de la courbe), et la médiane coïncident graphiquement sur l'axe des x.

Telles sont les caractéristiques centrales de la courbe. Pour ce qui est des extrémités elles sont asymptotiques à l'axe des x. Ce fait vient de l'impossibilité d'obtenir une fréquence de 0 pour une valeur de x donnée, le minimum étant nécessairement un. Or toutes ces caractéristiques, qui semblent ne se référer qu'à la configuration de la courbe, sous-tendent les notions de dissymétrie et de dispersion auxquelles se réfèrent les tests de la normalité.

1.6 a) Test de la normalité

Le seul procédé qui permet de cerner efficacement la question de la normalité repose sur la déviation par rapport à la moyenne. Selon qu'elle est exprimée au carré, au cube ou à la quatrième puissance, la déviation par rapport à la tendance centrale revêt les qualificatifs respectifs de deuxième, troisième ou quatrième moment. Le premier moment qui a sans doute été utilisé par tous, sert tout simplement à trouver la moyenne arithmétique avec les données brutes, ou la moyenne standardisée avec les déviations.

Mathématiquement on représente le premier moment de la façon suivante.

Données brutes

$$m_1 = \frac{\sum_{i=1}^n x_{ij}}{n}$$

$j= 1 \text{ à } m$

x_{ij} = valeur de la variable j à l'observation i

\bar{X}_j = moyenne de la variable j

n = nombre d'item

m = nombre de variables

$$m_1 = \frac{\sum_{i=1}^n (x_{ij} - \bar{X}_j)}{n}$$

Le second moment donne un indice qui mesure la variation et les comparaisons de variabilité. Ce coefficient de variation qu'il est commun d'appeler variance se traduit par l'expression:

$$m_2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}$$

Le lecteur attentif remarquera que l'extraction de la racine carrée de la formule précédente est équivalent à la formule désignée pour obtenir l'écart-type.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}}$$

Nous aurons l'occasion de nous servir de la variance à maintes reprises dans différentes formules et procédés comme les tests de signification statistique, et en particulier dans cette analyse bien connue du même nom. Celle-ci permet de comparer deux ou plusieurs groupes en examinant la signification de toutes les différences inter-groupes à la fois.

Le troisième et le quatrième moment contribuent respectivement à mesurer la dissymétrie et la kurtose de la courbe normale.

1.6 b) Mesure de dissymétrie

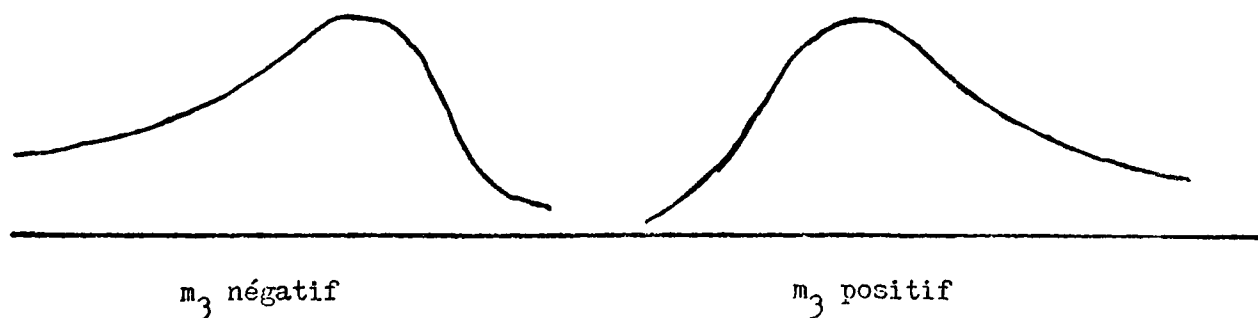
La mesure du degré de dissymétrie s'obtient par la mise au cube des déviations par rapport à la moyenne.

$$m_3 = \frac{\sum_{i=1}^n (x_{ij} - x_j)^3}{n}$$

Il va sans dire que le signe original (+ ou -) de la déviation demeure dû au fait de l'exposant 3 qui est impair. Ceci a pour résultat de ne pas modifier la relativité des valeurs à l'intérieur de la distribution. Si la majorité des valeurs négatives sont situées près de la moyenne et que les valeurs positives sont plus éloignées, la prédominance sera à l'avantage des valeurs au-dessus de la moyenne et vice-versa. Comme règle générale on peut affirmer que lorsque le troisième moment est négatif il y a dissymétrie de la distribution vers la gauche c'est-à-dire vers les plus petites valeurs et inversement lorsqu'il est positif.(figure 11)

FIGURE 11

DISSYMETRIES DE LA COURBE NORMALE



Mais il est préférable de rendre cette mesure indépendante de l'échelle selon laquelle sont enregistrées les données en la

divisant par l'écart-type à la troisième puissance.

$$\begin{aligned}\sigma^3 &= \sigma^2 \sigma \\ &= m_2 \sqrt{m_2} \\ g_1 &= \frac{m_3}{m_2 \sqrt{m_2}}\end{aligned}$$

La mesure de symétrie est alors égale à g_1 : 1,96 (entre +1,96 et -1,96 écart-type, la courbe normale contient 95% des données). Si g_1 est égal à 0 il y a symétrie parfaite dans la distribution. Le champ d'acceptation de g_1 pour la symétrie se situe entre +1,96 et -1,96. Un g_1 positif excédant 1,96 indique un excès d'observations plus petites que la moyenne et une médiane plus petite que la moyenne.

1.6 c) Test de la kurtose

Le degré de convexité représente une autre caractéristique physiologique de la courbe normale. Selon qu'elle aura un degré de courbure accentué, intermédiaire ou faible on dira qu'elle est leptokurtique, mésokurtique ou platikurtique.

Le quatrième moment rentre dans le processus de calcul d'un indice de forme et nous instruit sur la voussure de la courbe.

$$\text{4e moment: } m_4 = \frac{\sum_{i=1}^n (x_{ij} - \bar{X}_j)^4}{n}$$

Encore une fois, pour ne pas faire intervenir l'échelle à laquelle se réfère la distribution, il convient de diviser le quatrième moment par l'écart-type à la quatrième puissance ou si l'on veut, la variance au carré. Il en résulte l'expression suivante.

$$\text{Kurtose} = \frac{m_{44}}{\sigma^4} = \frac{m_{44}}{(\sigma^2)^2} = \frac{m_{44}}{(m_2)^2}$$

Dans le cas d'une courbe normale, le degré de kurtose est égal à 3. Ainsi une valeur supérieure à 3 reflète une accentuation excessive et une valeur inférieure à 3 dénote une courbe relativement peu accentuée. Certains auteurs préfèrent utiliser la base 0 au lieu de 3 à l'aide de l'indice g_2 défini par l'équation suivante: $g_2 = b_2 - 3$. Un g_2 positif correspond au b_2 plus grand que 3 et un g_2 négatif assume l'inverse.

1.6 d) Programme informatique de la normalisation

Le programme de normalisation composé par Bryn Greer-Wootten d'après les travaux de Snedecor (1967), procède d'après les deux tests précédents pour transformer sous forme normale une distribution qui ne l'est pas à l'origine.

La matrice d'entrée du programme correspond à la matrice d'information spatiale déjà définie. L'opération de la normalisation est effectuée sur chacune des colonnes successivement bien

qu'il soit toujours possible à l'utilisateur de demander à l'ordinateur de ne pas introduire une ou plusieurs variables dans le procédé de calcul. Le principe de base est simple: chaque colonne représente une distribution dont on vérifie au départ la normalité à l'aide des tests de symétrie et de kurtose.

Si les tests donnent des résultats qui demeurent dans les limites de la normalité le programme passe à la variable suivante; sinon il effectue une transformation mathématique pour tenter de rendre la distribution "Gaussienne".

Étapes du programme

1. Calcul de la moyenne de la colonne j: somme de la colonne divisée par le nombre d'observations.
2. Calcul des déviations au carré, au cube et à la quatrième puissance.
3. Calcul du deuxième, troisième et quatrième moment.
4. Calcul de $g_1 = m_3/m_2\sqrt{m_2}$
Calcul de $g_2 = m_4/(m_2)^2$
5. Calcul du degré de dissymétrie

$$XS = g_1 / \sqrt{6n(n-1) / (n-2)(n+1)(n+3)}$$
 Calcul du degré de kurtose

$$XT = g_2 / \sqrt{24n(n-1)^2 / (n-3)(n-2)(n+3)(n+5)}$$
6. Si XS se situe entre 1,96 et -1,96 la distribution n'est pas considérée comme asymétrique et aucune transformation n'est nécessaire. Si tel n'est pas le cas une série de transformations sont effectuées sur les données.

La transformation logarithmique est employée pour stabiliser la variance si l'écart-type dans l'échelle originale varie directement avec la moyenne ou en d'autres mots, si le coefficient de variation demeure constant.

La première opération est la mise en logarithme dans la base 10 de chacune des valeurs. Ceci est possible dans la mesure où la distribution originale ne contient aucune valeur plus petite ou égale à 0. Dans le cas contraire, le programme cherche un second moyen de transformation.

En admettant le rejet de la mise en logarithme le programme expérimente l'extraction de la racine carrée. Encore là une valeur inférieure à 0 entraîne le rejet de ce type de transformation qui tente de stabiliser la variance.

Si les deux transformations précédentes ne peuvent être effectuées, le programme optera soit pour la mise au carré qui entraîne la disparition des signes négatifs et par le fait est beaucoup moins bonne, soit pour la mise au cube qui elle par contre, garde le signe original mais hypertrophie les valeurs extrêmes.

7. Après l'une ou l'autre de ces transformations, les degrés de dissymétrie et de kurtose sont recalculés. Si la distribution n'est pas encore normale, le programme refait une ultime tentative de normalisation sur la dernière transformation.
8. Lorsqu'il rencontre la transformation mathématique appropriée il effectue cette dernière sur toute la colonne produisant la nou-

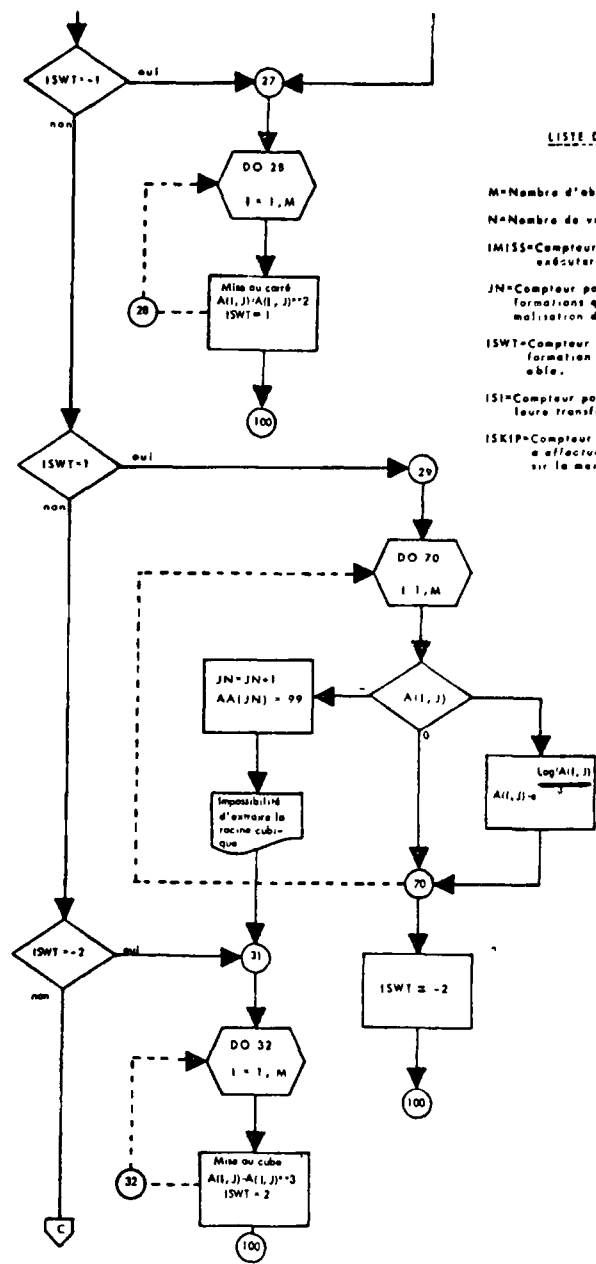
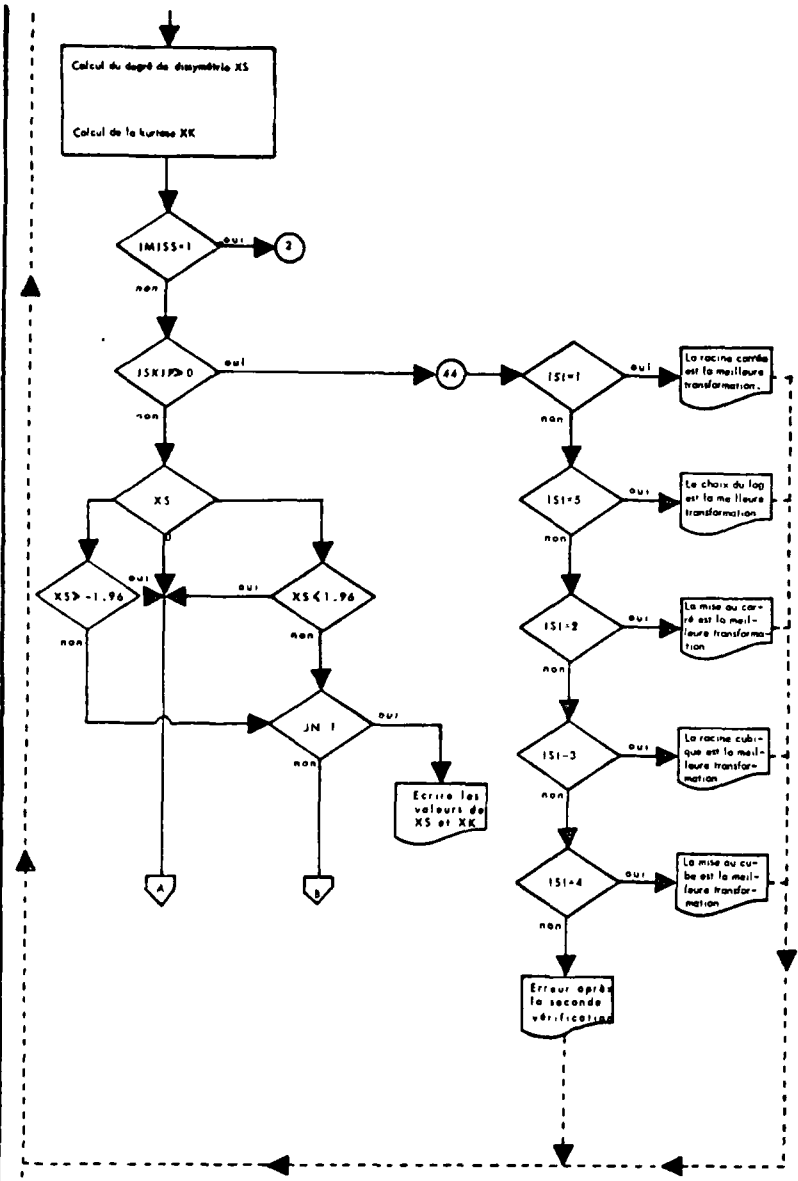
velle variable normalisée.

9. Si après deux itérations il a été impossible d'atteindre l'objectif de normalisation, l'ordinateur choisira comme type de transformation celle pour laquelle la disparité avec la courbe de Gauss est minimale.
10. Le processus recommence avec la variable suivante et ainsi de suite jusqu'à la fin. (figure 12)

L'organigramme qui suit décrit le cheminement logique du programme de la normalisation et complète ce que les mots n'ont pas pu exprimer en la matière.

Ce procédé de normalisation n'est en fait guère sophistiqué. Il mérite d'être réévalué au fur et à mesure que nos connaissances sur la nature même de la normalisation en sciences humaines s'accroîtront. Mais déjà surgit la question de savoir quand il faut normaliser. Les experts diffèrent d'opinion mais la plupart s'accordent à dire qu'il n'en tient qu'à la nature de l'étude, du type de données et de la réflexion appropriée du chercheur (Gould, 1970).

A titre d'exemple, les données relevées sur un continuum spatial subdivisé selon des critères administratifs tel que des limites municipales ne doivent pas nécessairement être normalisées car une telle opération a pour effet de faire disparaître les frontières municipales et n'offrent aucun point de référence. Cependant une étude se fondant sur des échantillons se doit d'avoir des données normalisées pour être vraiment représentative. Dans ce cas le programme entre en jeu à la fois comme procédé de transformation et



LISTE DES COMPTEURS

- M=Nombre d'observations
- N=Nombre de variables à être normalisées
- IMISS=Compteur pour indiquer si l'utilisateur veut exécuter la normalisation
- JN=Compteur pour indiquer le nombre de transformations qui ne concernent pas la normalisation d'une variable donnée
- ISWT=Compteur pour indiquer le choix de la transformation nécessaire pour normaliser la variable.
- ISI=Compteur pour indiquer le choix de la meilleure transformation résiduelle
- ISKIP=Compteur pour indiquer si le programme a effectué un deuxième essai pour choisir la meilleure transformation résiduelle

comme test.

De plus en plus toutefois on tend à utiliser la distribution originale qu'elle soit normale, lognormale, Gamma, circulaire, ou de Poisson et de chercher à l'aide du calcul différentiel et intégral les superficies qui par rapport à la moyenne correspondent à un ou deux ou trois écart-types de la distribution normale. Ceci évite les distortions et est beaucoup plus fidèle à la nature du phénomène étudié. Mais alors la comparaison avec d'autres travaux sur le même espace est impossible.

Malgré toutes les recherches sur la normalisation, les idées sont partagées et le débat reste ouvert. Peu importe que l'on soit en faveur ou non, il s'agit d'être le plus fidèle possible à la réalité (en admettant que l'on sache ce qu'est la réalité) et de pouvoir justifier son option en termes géographiques évidemment. Comme l'a montré J.B. Racine dans son troisième article sur les modèles mathématiques et graphiques en géographie humaine (Racine, 1972), le géographe ne saurait en aucun cas, sous peine de se trahir, sacrifier les impératifs, la logique et les contraintes de ses propres préoccupations méthodologiques. Or les deux ordres de préoccupations ne correspondent pas obligatoirement. Loin de là!

CHAPITRE II

LES ALGORITHMES DE L'ANALYSE INFORMATIQUE
DE L'ESPACE GEOGRAPHIQUE

2.0) INTRODUCTION

La deuxième partie de ce travail sera consacré à l'explication des différentes méthodes de l'analyse multivariée. Chacune sera étudiée individuellement dans le but d'en montrer les caractéristiques. On partira de la corrélation qui est l'indice de base de toutes les études de structure en géographie. On procédera par la suite aux programmes de la réduction des attributs spatiaux pour ensuite passer aux divers algorithmes de classification. Finalement on étudiera les principaux programmes de la modélisation partielle, qui servent très souvent à la simulation.

Chaque programme est basé sur la matrice d'information spatiale de l'utilisation du sol. Nous avons cru bon d'utiliser toujours le même exemple afin que le lecteur puisse suivre l'évolution des données lors de leurs transformations statistiques et établir ainsi des points de comparaison entre les différents résultats obtenus.

2.1) LA CORRELATION: POINT CENTRAL DE LA REFLEXION ET DE L'ANALYSE GEOGRAPHIQUES

Dans la section précédente nous nous sommes uniquement attardés aux caractéristiques qui se rapportent à une seule et même variable géographique. Le géographe doit à la suite de la mise au point de son information, chercher les inter-relations entre les diverses constituantes de l'espace afin de pouvoir ultérieurement en déceler la structure. La recherche des relations a été conduite, avant l'avènement des méthodes quantitatives, par des procédés tel que la superposition des cartes (géologiques, topographiques, utilisation du sol, etc.) et l'utilisation de graphiques à double entrée. Cependant la question est demeurée toujours la même: dans quelle mesure un changement dans la variable X_1 entraîne-t-il une modification de la variable X_2 ?

Quantitativement la relation entre deux variables est calculée à l'aide du coefficient de corrélation r qui est une valeur unique, comprise entre -1 et +1. Elle nous informe sur la nature, le sens, et l'intensité de la relation. Un coefficient de corrélation de -1 indique une relation parfaitement inversement proportionnelle; un r de 0 signifie une indépendance complète entre les deux variables tandis que $r=1$ indique une relation parfaitement positive et fonctionnelle. Plusieurs livres traitent de la corrélation et nous ne tenterons pas de

les répéter, nous contentant de rappeler tout simplement que la formule du coefficient de corrélation entre deux variables X et Y employé dans la plupart des cas est celle de Pearson, qui se fonde sur la prise en considération des écarts à la moyenne.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{X}) (y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 (y_i - \bar{Y})^2}}$$

$j = 1 \text{ à } n$
 \bar{X} = moyenne de la variable X
 \bar{Y} = moyenne de la variable Y
 n = nombre d'observations

Sans reprendre la démonstration de cette formule nous tenons à souligner que le numérateur de l'équation calcule les écarts à la moyenne des variables X et Y et que le dénominateur compare des écarts-types, pondère les deux variables en les plaçant à la même échelle. Pour illustrer le fonctionnement de cette formule prenons un exemple à partir de notre matrice d'information spatiale. En tant que géographe nous voulons connaître le degré de la relation qui existe entre la variable représentant l'importance du terrain urbanisé et celle de l'importance du terrain vacant dans les 22 municipalités de la rive sud de Montréal.

TABLEAU

CALCUL DU COEFFICIENT DE CORRELATION ENTRE L'IMPORTANCE
DU SOL URBANISE ET L'IMPORTANCE DU TERRAIN VACANT.

X	$X - \bar{X}$	$(X - \bar{X})^2$	Y	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(X - \bar{X})(Y - \bar{Y})$
35.1	-5.3	28.9	34.2	13.0	169.0	-69.9
10.1	-30.3	923.0	38.5	17.3	299.2	-525.6
17.5	-22.9	528.1	29.9	8.7	75.6	-199.9
....
....
79.2	38.7	<u>1499.0</u>	10.9	-10.3	<u>106.0</u>	<u>-398.7</u>
		$\Sigma 23887.6$			$\Sigma 2129.5$	$\Sigma -1937.4$

$$r = \frac{-1937.4}{\sqrt{(23887.61)(2129.5)}} = \frac{-1937.4}{7132.3} = -.27$$

Le coefficient $-.27$ indique que la dispersion des points dans un graphique, ayant comme abscisse la variable X et comme ordonnée la variable Y, ne semble pas avoir tendance à se rapprocher d'une ligne droite et par le fait même nous empêche de dire que dans les 22 municipalités de la rive sud, plus il y a de sol urbanisé moins il y a de terrain vacant. On peut cependant supposer qu'il existe d'autres variables en-

trant en jeu à l'intérieur du système de l'utilisation du sol péri-urbain.

Les coefficients de corrélation peuvent être facilement disposés sous la forme de matrice symétrique à partir de laquelle on effectue les analyses factorielles et les régressions. Notre matrice symétrique est un tableau dont le nombre des colonnes est identique à celui des rangées et où la partie supérieure de la diagonale est identique à la partie inférieure.

re. Exemple:
$$\begin{array}{|c|c|c|c|} \hline 1 & 2 & 3 \\ \hline 2 & 1 & 4 \\ \hline 3 & 4 & 1 \\ \hline \end{array}$$

Il est permis d'affirmer que la corrélation d'une variable avec elle-même est parfaite. Chaque élément de la diagonale est donc identique à 1.

	1	2
1	1.00	
2		1.00

On a trouvé, dans l'exemple ci-dessus, que la corrélation entre la variable 1 et 2 était de $-.27$. C'est pourquoi aux jonctions des coordonnées matricielles 1 et 2 on inscrit la valeur de $-.27$.

	1	2
1	1.0	-.27
2	-.27	1.0

On a ainsi une matrice de rang 2 x 2, tirée de la corrélation entre deux seules variables. Si on introduit dans le système une troisième variable, telle que l'importance du terrain résidentiel, et qu'elle est corréllée successivement avec les variables indexant l'importance du sol urbanisé et celle du terrain vacant pour la rive sud de Montréal, on obtient les coefficients de corrélation suivants: 0,49 et -0,14. Ces valeurs, ajoutées à la matrice précédente, donnent une matrice de rang 3 x 3.

	1	2	3
1	1.0	-.27	0,49
2	-.27	1.0	-0,14
3	0,49	-0,14	1.0

Avec douze variables, celles de l'utilisation du sol de la banlieue de Montréal on obtient les résultats que le lecteur trouvera au tableau no.7.

Après l'obtention de la matrice de corrélation, l'étape de l'interprétation s'impose. L'interprétation est parfois compliquée en raison du fait que le coefficient de corré-

TABEAU 7

MATRICE DES CORRELATIONS

RANGÉE	COLONNE 1	COLONNE 2	COLONNE 3	COLONNE 4	COLONNE 5	COLONNE 6	COLONNE 7	COLONNE 8	COLONNE 9	COLONNE 10	COLONNE 11	COLONNE 12
1	1.000	-0.272	0.493	-0.384	0.450	0.455	0.369	-0.314	0.426	0.408	-0.188	-0.155
2	-0.272	1.000	-0.144	0.404	-0.440	-0.459	-0.102	-0.410	-0.317	-0.364	-0.153	0.081
3	0.493	-0.144	1.000	-0.284	0.087	0.187	0.281	-0.360	0.317	0.325	-0.753	-0.228
4	-0.384	0.404	-0.284	1.000	-0.669	-0.575	-0.692	0.141	-0.396	-0.272	0.042	-0.004
5	0.450	-0.440	0.087	-0.669	1.000	0.813	-0.071	-0.246	0.694	0.540	0.211	-0.048
6	0.455	-0.459	0.187	-0.575	0.813	1.000	-0.056	-0.199	0.452	0.472	0.138	-0.063
7	0.369	-0.102	0.281	-0.692	-0.071	-0.056	1.000	0.050	-0.131	-0.165	-0.256	0.054
8	-0.314	-0.410	-0.360	0.141	-0.246	-0.199	0.050	1.000	-0.425	-0.452	0.057	-0.188
9	0.426	-0.317	0.317	-0.396	0.694	0.452	-0.131	-0.425	1.000	0.418	-0.076	0.083
10	0.408	-0.064	0.325	-0.272	0.540	0.472	-0.165	-0.452	0.418	1.000	-0.133	-0.191
11	-0.188	-0.153	-0.753	0.042	0.211	0.138	-0.256	0.057	-0.076	-0.133	1.000	-0.036
12	-0.155	0.081	-0.228	-0.004	-0.048	-0.063	0.054	-0.188	0.083	-0.191	-0.036	1.000

lation est une mesure qui doit être considérée en fonction des conditions de la recherche. Un coefficient de corrélation de .65 semblera faible à une certaine échelle tandis qu'un autre de .50, entre les mêmes variables, sera significatif à une autre échelle. Le fait de la grande flexibilité du coefficient de corrélation en fonction du type de recherche donne peut-être l'impression d'une grande subjectivité. Avec l'expérience cependant, il est possible d'établir des degrés de hiérarchie à l'intérieur d'une matrice de corrélation.

En effet, tout comme il est possible de calculer le degré de relation entre deux variables, il est aussi possible de connaître le degré d'absence de relation en utilisant le coefficient d'aliénation (k) représenté par la formule suivante:

$$k = \sqrt{1 - r^2}$$

Ainsi pour la corrélation entre la variable du terrain urbanisé et celle du terrain vacant de l'exemple précédent on obtient:

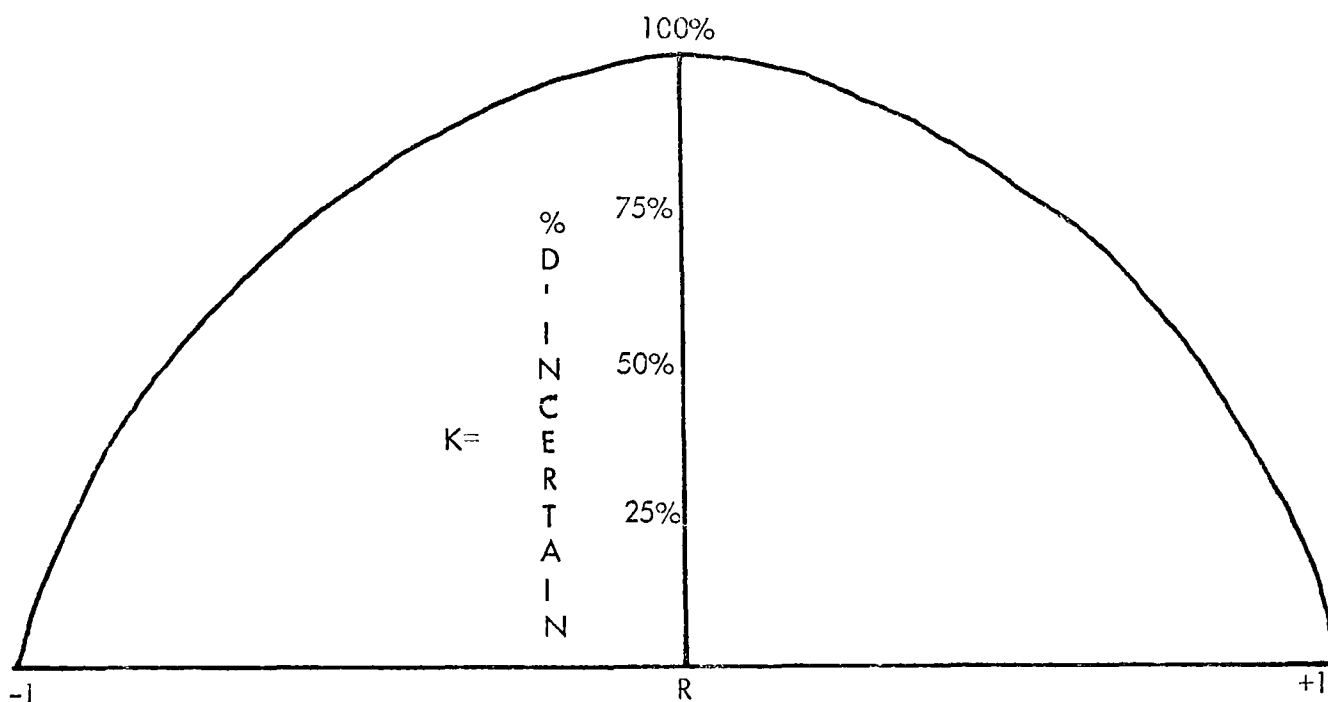
$$k = \sqrt{1 - (-.27)^2} = 30\% \text{ d'incertitude}$$

La relation entre le coefficient de corrélation r et le coefficient d'aliénation k est graphiquement démontrée à la figure no.13 . Nous avons tracé ce graphique afin que le lecteur perçoive d'un seul coup d'oeil le coefficient d'aliénation

sans être obligé de le calculer.

FIGURE 13

RAPPORT ENTRE LES COEFFICIENTS r ET k



Rappelons cependant qu'avant d'utiliser le coefficient de corrélation et le coefficient d'aliénation on doit s'assurer de leur conformité avec les postulats de base de la corrélation: la linéarité et l'homoscédasticité.

2.1 a) La linéarité

Lorsque nous avons tenté de trouver la relation existant entre la variable du sol urbanisé et celle du terrain vacant nous avons essayé de tracer une ligne qui puisse se

placer le plus près possible de tous les points de la distribution. Cette ligne se nomme la ligne de régression, à laquelle nous porterons une attention plus particulière dans la section de la modélisation partielle. Si de prime abord on s'aperçoit que la relation entre les variables ne semble pas linéaire mais plutôt curviligne, il faut abandonner le coefficient de corrélation et opter pour le rapport de corrélation qui tend à diminuer les erreurs commises par les points qui se départissent de la linéarité. Pour savoir si on doit prendre le coefficient de corrélation ou le rapport de corrélation on n'a qu'à calculer les deux et constater que plus l'écart est grand, plus la relation s'écarte de la linéarité.

2.1 b) L'homoscédasticité

Ce terme qui revient souvent dans les tests de variance et de comparaison de plusieurs distributions renferme la condition de l'égalité de la variabilité de chacune des variables. C'est-à-dire que la variance de X_1 soit égale à la variance de X_2 et à celle de X_3 , etc.

$$\sigma_{x_1} = \sigma_{x_2} = \sigma_{x_3}$$

En admettant que la variabilité ne soit pas identique entre les variables, une simple transformation à l'aide de la formule 2.1, du coefficient de corrélation sous la

forme de Z, lequel est normalement distribué, suffit à rétablir l'homoscédasticité.

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r} \quad \text{formule 2,1}$$

Cette solution est très utile en géographie lorsque l'on a l'intention de comparer des coefficients de corrélation à des intervalles de temps différents. Supposons qu'on fasse une étude de l'évolution de l'utilisation du sol entre les années 1950 et 1980. Il est quasi impossible de comparer les matrices de corrélations, puisque les attributs de l'utilisation du sol ont des variabilités spécifiques à chaque époque. E.N. Thomas (1962), a soulevé le problème et a utilisé cette méthode dans ses études sur la stabilité de la relation entre la distance et la population des villes de l'état d'Iowa entre 1900 et 1950. Pour comparer ses coefficients, il les a transformés sous forme de Z afin d'effectuer par la suite la différence entre les coefficients de corrélation. Voici cet exemple que l'on retrouve également dans Leslie King, 1966, p. 132.

Le tableau contient les coefficients de corrélation et la transformation en Z entre la distance entre une ville donnée et sa plus proche voisine de même population et la population de cette ville donnée.

	1900	1910	1920	1930	1940	1950
Coefficient de corrélation	0.30	0.64	0.57	0.57	0.65	0.59
Transformation Z	.3095	0.7582	0.6475	0.6475	0.7763	0.6777

A partir de ces données on obtient la matrice des différences entre les transformations en Z.

	1910	1920	1930	1940	1950
1900	.4487	.3380	.3380	.4658	.3682
1910		.1107	.1107	.0029	.0805
1920			.0000	.1278	.0302
1930				.1278	.0302
1940					.0492

Thomas en est arrivé à la conclusion que si on exclut le cas de 1900, la stabilité demeure remarquable même si on tient compte des changements sociaux et économiques pendant les 50 ans.

Cette comparaison n'aurait point été possible s'il l'avait faite directement à partir de la matrice de corrélation. Donc il faut être très méticuleux lorsqu'il s'agit de comparer des corrélations et s'assurer que les postulats de linéarité et d'homoscédasticité sont inhérents aux distributions de variables en cause. Remarquons cependant que même si la condition de linéarité ne tient pas, il ne faut pas conclure qu'il n'ex-

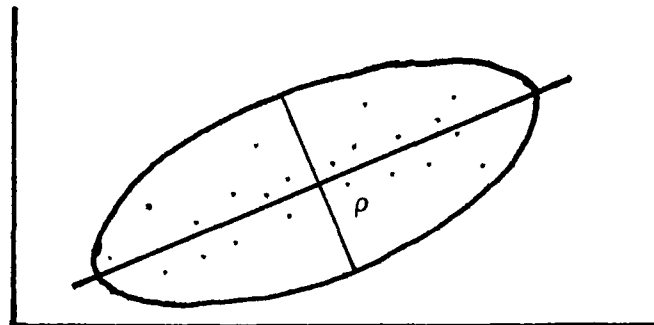
iste aucune relation; rappelons que le coefficient de Pearson détermine dans quelle mesure une ligne droite décrit la relation entre les variables. Un coefficient de corrélation près de 0.0 signifie qu'il n'existe pas de relation linéaire mais il en existe peut-être une autre, circulaire par exemple.

Souvent on fait l'erreur d'éliminer des variables d'une analyse qui ont une faible corrélation. A titre d'exemple on néglige les variables qui ont un coefficient de corrélation situé entre +.40 et -.40. A moins d'avoir une très bonne connaissance du milieu étudié qui permet de tenir compte des coefficients de corrélation marginaux tels que +.36 ou -.39, cette méthode est à déconseiller fortement. Il est préférable de faire un test de t basé sur l'hypothèse nulle que P égale 0 (voir ci-dessous).

2.1 c) Test de t et coefficient de corrélation: niveau de signification du coefficient

ρ est un paramètre qui mesure le degré de contiguité de la relation et l'étroitesse de l'ellipse. La figure 14 contient la majeure partie des valeurs des variables corréllées.

FIGURE 14



L'utilisation du test de t pour calculer le degré de signification du coefficient de corrélation consiste en une comparaison entre le coefficient de corrélation estimé lorsque $\rho = 0$ et le coefficient de corrélation obtenu par la formule de Pearson. Avec une ellipse très allongée ρ tend vers 0; par contre plus elle se rapproche de la forme du cercle, plus la valeur de ρ augmente.

Fisher a démontré que l'hypothèse nulle ($\rho = 0$), peut être vérifiée au moyen du test de t par la formule suivante.

$$t = r \sqrt{(N - 2)/(1 - r^2)}$$

r: coefficient de corrélation
N: nombre d'observations

pour N-2 degrés de liberté

Dans le cas des 22 municipalités de la rive sud de Montréal, avec un coefficient de corrélation de -.80 entre les variables normalisées indexant l'importance respective du résidentiel unifamilial et du résidentiel mixte on obtient:

$$t = -.81 \sqrt{(22 - 2)/(1 - .6481)}$$

= - 3.91984 avec 20 degrés de liberté

De ceci on peut conclure, à partir des tables de t (2.89), que la relation entre ces deux variables de l'utilisation du sol de la rive sud de Montréal est significative à 99.9%.

Ajoutons cependant un autre exemple, pour mettre en garde contre les coupures à une valeur arbitraire telle que $-.40$. La corrélation entre l'importance du terrain résidentiel et le nombre de parcs et de terrains de jeu est de $-.36$ pour des distributions normalisées. Le résultat du test de t est de (-1.72) significatif à 90%. Il est donc tout à fait légitime d'inclure ces variables dans les analyses subséquentes.

Le tableau 8 illustre la matrice de corrélation avec le test de t et le niveau de signification correspondant pour les données de l'utilisation du sol des 22 municipalités de la banlieue de la rive sud de Montréal.

Il nous permet de percevoir, d'un seul coup d'oeil, les relations les plus significatives, tel que le lien étroit ($r = .81$) entre le résidentiel unifamilial (variable 5) et le résidentiel multifamilial (variable 6) mais il offre en plus la possibilité de décortiquer la matrice de corrélation à différents niveaux de précision, $.10$ (90%), $.05$ (95%), $.02$ (98%), $.01$ (99%), $.001$ (99.9%). Si le géographe désire un degré de précision particulier, $.02$ (98%) par exemple, il n'a qu'à choisir les coefficients de corrélation dont le niveau de signification est inférieur à $.02$. L'abréviation NTSG dans la colonne du niveau de signification signifie que le coefficient de corrélation correspondant n'a aucune signification statistique.

TABLEAU 8

NIVEAU DE SIGNIFICATION DE LA CORRELATION

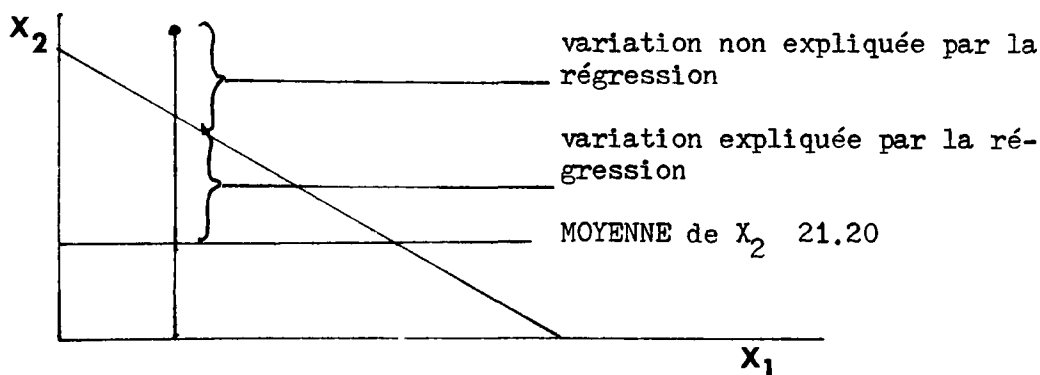
Variables	Corrélations	Somme de X	Somme de X ²	Somme de Y	Somme de Y ²	Somme de X*Y	Test de T	Niveau de Signification
1 2	-0.271641	891.	59941.	467.	12026.	16951.	-1.26228	NTSG
1 3	0.492912	891.	59941.	1122.	62182.	50792.	2.53252	.02
1 4	-0.383991	891.	59941.	1748.	148280.	64983.	-1.85984	.1
1 5	0.450040	891.	59941.	218.	6224.	13268.	2.25377	.05
1 6	0.455248	891.	59941.	16.	77.	1224.	2.28652	.05
1 7	0.068714	891.	59941.	220.	7425.	9670.	0.30883	NTSG
1 8	-0.313902	891.	59941.	204.	5242.	5450.	-1.47854	NTSG
1 9	0.425574	891.	59941.	83.	487.	4231.	2.10319	.05
1 10	0.407732	891.	59941.	77.	456.	3986.	1.99697	.1
1 11	-0.188445	891.	59941.	180.	5182.	5900.	-0.85813	NTSG
1 12	-0.154898	891.	59941.	67.	931.	2050.	-3.70119	NTSG
2 3	-0.143653	467.	12026.	1122.	62182.	23319.	-0.64917	NTSG
2 4	0.404250	467.	12026.	1748.	148280.	38880.	1.97656	.1
2 5	-0.440152	467.	12026.	218.	6224.	3336.	-2.19219	.05
2 6	-0.458889	467.	12026.	16.	77.	173.	-2.30977	.05
2 7	-0.101815	467.	12026.	220.	7425.	4324.	-0.45771	NTSG
2 8	-0.409657	467.	12026.	204.	5242.	3232.	-2.00829	.1
2 9	-0.316893	467.	12026.	83.	487.	1573.	-1.49420	NTSG
2 10	-0.063495	467.	12026.	77.	456.	1600.	-0.28453	NTSG
2 11	-0.152800	467.	12026.	180.	5182.	3382.	-0.69166	NTSG
2 12	0.080954	467.	12026.	67.	931.	1513.	0.36323	NTSG
3 4	-0.283556	1122.	62182.	1748.	148280.	87152.	-1.32238	NTSG
3 5	0.086584	1122.	62182.	218.	6224.	11519.	0.38868	NTSG
3 6	0.186793	1122.	62182.	16.	77.	932.	0.85033	NTSG
3 7	0.281010	1122.	62182.	220.	7425.	12648.	1.30948	NTSG
3 8	-0.359505	1122.	62182.	204.	5242.	8929.	-1.72295	NTSG
3 9	0.316898	1122.	62182.	83.	487.	4536.	1.49422	NTSG
3 10	0.324594	1122.	62182.	77.	456.	4253.	1.53473	NTSG
3 11	-0.753296	1122.	62182.	180.	5182.	5915.	-5.12229	.001
3 12	-0.228470	1122.	62182.	67.	931.	2959.	-1.04951	NTSG
4 5	-0.869279	1748.	148280.	218.	6224.	13199.	-4.02835	.001
4 6	-0.575461	1748.	148280.	16.	77.	836.	-3.14679	.01
4 7	-0.692005	1748.	148280.	220.	7425.	12608.	-4.28697	.001
4 8	0.140782	1748.	148280.	204.	5242.	16998.	0.63593	NTSG
4 9	-0.396306	1748.	148280.	83.	487.	6104.	-1.93040	.1
4 10	-0.272268	1748.	148280.	77.	456.	5781.	-1.26542	NTSG
4 11	0.042433	1748.	148280.	180.	5182.	14527.	0.18994	NTSG
4 12	-0.004151	1748.	148280.	67.	931.	5280.	-0.01857	NTSG
5 6	0.813124	218.	6224.	16.	77.	579.	6.24714	.001
5 7	-0.070640	218.	6224.	220.	7425.	1857.	-0.31670	NTSG
5 8	-0.246076	218.	6224.	204.	5242.	1117.	-1.13540	NTSG
5 9	0.694421	218.	6224.	83.	487.	1406.	4.31584	.001
5 10	0.540100	218.	6224.	77.	456.	1235.	2.87001	.01
5 11	0.211027	218.	6224.	180.	5182.	2602.	0.96548	NTSG
5 12	-0.048459	218.	6224.	67.	931.	577.	-0.21697	NTSG
6 7	-0.055754	16.	77.	220.	7425.	129.	-0.24973	NTSG
6 8	-0.199233	16.	77.	204.	5242.	57.	-0.90922	NTSG
6 9	0.451629	16.	77.	83.	487.	109.	2.26377	.05
6 10	0.472319	16.	77.	77.	456.	109.	2.39642	.05
6 11	0.137699	16.	77.	180.	5182.	200.	0.62173	NTSG
6 12	-0.063252	16.	77.	67.	931.	35.	-0.28344	NTSG
7 8	0.050494	220.	7425.	204.	5242.	2250.	0.22611	NTSG
7 9	-0.131393	220.	7425.	83.	487.	707.	-0.59275	NTSG
7 10	-0.164850	220.	7425.	77.	456.	611.	-0.74746	NTSG
7 11	-0.255949	220.	7425.	180.	5182.	668.	-1.18408	NTSG
7 12	0.054144	220.	7425.	67.	931.	771.	0.24249	NTSG
8 9	-0.424895	204.	5242.	83.	487.	449.	-2.09909	.05
8 10	-0.451770	204.	5242.	77.	456.	361.	-2.26466	.05
8 11	0.056799	204.	5242.	180.	5182.	1867.	0.25442	NTSG
8 12	-0.188308	204.	5242.	67.	931.	323.	-0.85748	NTSG
9 10	0.417581	83.	487.	77.	456.	367.	2.05524	.1
9 11	-0.076032	83.	487.	180.	5182.	619.	-0.34131	NTSG
9 12	0.082838	83.	487.	67.	931.	281.	0.37174	NTSG
10 11	-0.132522	77.	456.	180.	5182.	522.	-0.59793	NTSG
10 12	-0.190982	77.	456.	67.	931.	164.	-0.87012	NTSG
11 12	-0.035993	180.	5182.	67.	931.	485.	-0.16107	NTSG

Ce test de signification sur la corrélation est rarement employé dans les analyses multivariées, et avec grand tort, car il est en mesure d'offrir au chercheur une assurance qui peut parfois pallier à un manque de connaissance du milieu. En complément de ce test, le coefficient de détermination r^2 apporte des renseignements intéressants sur la validité du coefficient de corrélation.

2.1 d) Le coefficient de détermination r^2

Le coefficient de corrélation au carré, r^2 a un champ compris entre 0 et 1. Cet indice fournit au géographe le degré d'explication de la ligne de régression dont nous verrons ultérieurement qu'il est étroitement lié au coefficient de régression. Le coefficient de détermination n'est pas seulement le coefficient de corrélation mis au carré, mais c'est aussi un rapport de variation. La figure (15) résume ce rapport avec le coefficient de corrélation unissant les variables (1) et (2) de notre exemple.

FIGURE 15

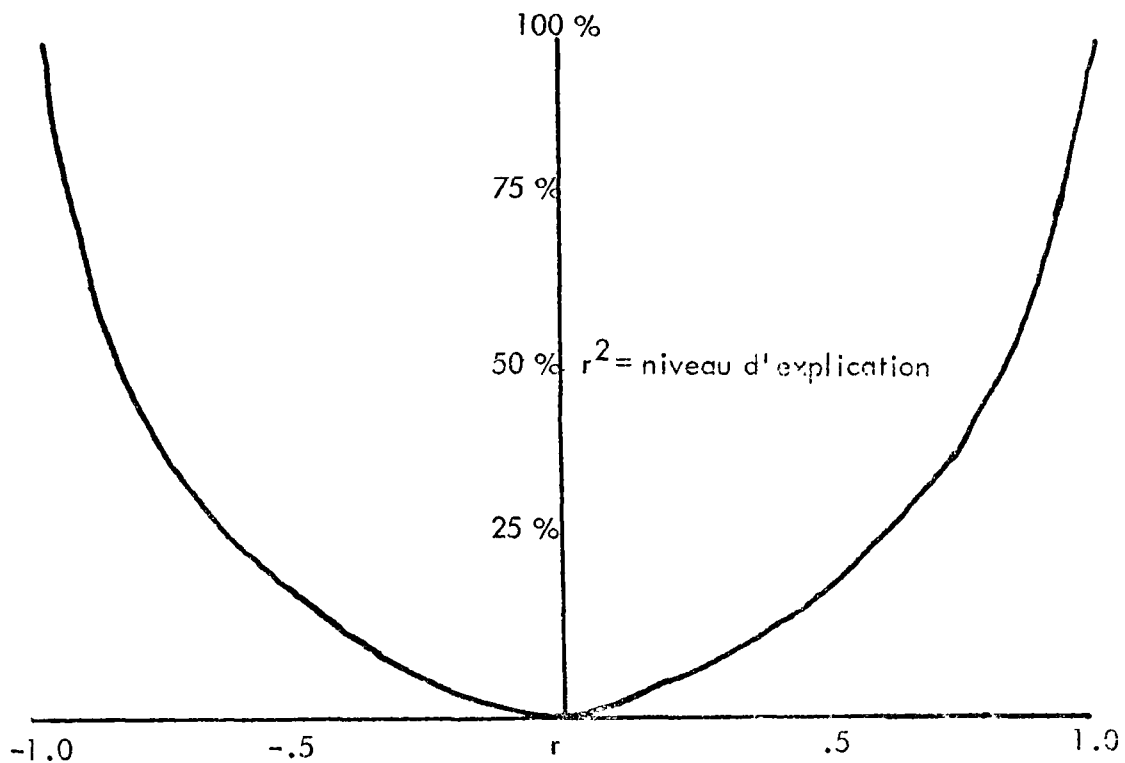


Prenons le cas d'une municipalité, Boucherville par exemple, qui a une valeur de 34,24 pour X_2 et de 35,10 pour X_1 (voir matrice originale). La distance entre le point et la ligne de régression représente la déviation non expliquée tandis que la distance entre la ligne de régression et la moyenne égale le montant de la déviation expliquée. Il va de soi que la somme des deux est égale à la variation totale. Le coefficient de r^2 exprime ainsi le rapport de la somme des déviations expliquées sur la somme des variations totales de chaque observation.

$$r^2 = \frac{\text{variation expliquée}}{\text{variation totale}}$$

On en déduit facilement que plus la relation expliquée tend vers la variation totale, plus précis est le degré de détermination. En multipliant r^2 par 100 on obtient le pourcentage d'explication de la corrélation. Un coefficient de corrélation de .70 représente à peine 49% d'explication, ce qui signifie, en d'autres mots, qu'à peine 49% de la variation totale est expliquée par la relation linéaire. Le géographe se doit d'être en garde contre des coefficients de corrélation qu'il considère élevés et qui ont en fait un très faible pourcentage d'explication. La figure (16) illustre la relation entre le coefficient de corrélation et le coefficient de détermination.

FIGURE 16
RAPPORT ENTRE LES COEFFICIENTS r ET r^2



En géographie il appert qu'on ait besoin d'établir des corrélations entre des variables déterminées par la présence ou l'absence d'un phénomène. Ceci est impensable à l'aide de la formule classique de la corrélation. Pour répondre à ce besoin la formule du coefficient de corrélation Phi ϕ a été mise au point et nous nous empressons d'en démontrer les rudiments.

2.1 e) Phi: ϕ

Même si dans la plupart des cas le géographe explore des variables exprimées en données relatives ou absolues, il arrive souvent que ni les unes ni les autres ne soient disponibles. La seule manière de recueillir alors de l'information repose sur le principe présence-absence, dont nous avons expliqué la base dans le premier chapitre. Toujours intéressé à trouver des relations le géographe voudra établir une matrice de corrélation, mais étant donné que la matrice présence-absence contient uniquement des 0 et des 1 la formule du coefficient de corrélation de Pearson n'est plus adéquate. Il doit donc utiliser une autre formule, basée sur les quatre permutations possibles existant entre 0 et 1, soit 00, 01, 10, et 11. La première permutation, 00, indique qu'une observation ne possède aucune des caractéristiques; avec la deuxième ou la troisième elle possède l'une ou l'autre et avec la dernière 11, les deux caractéristiques sont présentes. Prenons un exemple (*): nous voulons savoir si dans les 22 municipalités de la rive sud il y a corrélation entre la présence du résidentiel appartement et celle des parcs. La valeur 0 signifie absence et la valeur 1 l'inverse, c'est-à-dire la présence.

(*): Nous avons choisi quelques villes seulement pour démontrer l'emploi de la formule; il ne s'agit pas d'une étude complète sur la présence et l'absence de résidentiel appartement et de parcs.

	Présence ou absence de résidentiel appartement	Présence ou absence de parcs
I Boucherville	1	1
I Brossard	0	0
I Candiac	0	1
I Caughnawaga	0	1
I Châteauguay	1	1
I Châteauguay-Centre	0	1
I Châteauguay-Heights	0	1
I Delson	1	0
I Greenfield Park	1	1
I Jacques-Cartier	1	1
I Laflèche	0	1
I Laprairie	1	1
I Lemoyne	1	1
I Lery	0	1
I Longueuil	1	1
I Notre-Dame	0	1
I Prévile	0	1
I St-Bruno	1	1
I Ste-Catherine	0	1
I St-Constant	0	0
I St-Hubert	0	1
I St-Lambert	1	1

L'étape suivante consiste à établir le tableau du nombre de cas dans chaque possibilité,

Possibilités	A	B	C	D
	00	01	10	11
Nombre de cas	2	1	10	9

La formule du coefficient de corrélation Phi s'appuie sur le principe de l'identité et de la différence des couples.

$$\text{Phi} = \text{Produit des couples identiques 00 et 11} - \text{Produit des couples différents 01 et 10}$$

$\sqrt{\text{Produits croisés entre les couples identiques et les couples différents}}$

A l'aide des lettres A B C D employées au tableau précédent la formule devient:

$$\text{Phi} = \frac{(A \times D) - (B \times C)}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}$$

En utilisant les valeurs de notre exemple:

$$\text{Phi} = \frac{(2 \times 9) - (10 \times 1)}{\sqrt{(2+1)(10+9)(10+2)(1+9)}} = \frac{18 - 10}{\sqrt{6840}}$$

$$\phi = \frac{8}{82,70} = 0,096 = \text{coefficient de corrélation}$$

La différence entre le produit des couples identiques et celui des couples dissemblables (numérateur) permet de fournir le sens de la relation négatif ou positif. Le dénominateur sert à établir l'échelle de -1 à +1.

Voici un autre exemple, fictif, qui cette fois démontre qu'à l'aide de cette formule deux variables identiques ont aussi un coefficient de corrélation de 1.

	Variable 1		Variable 2	
Ville 1	1		1	
Ville 2	0		0	
Ville 3	0		0	
Ville 4	1		1	
Ville 5	0		0	
Ville 6	1		1	
Possibilité	A	B	C	D
	00	01	10	11
Nombre de cas	3	0	0	3

$$\text{Phi} = \frac{9 - 0}{\sqrt{3 \times 3 \times 3 \times 3}} = \frac{9}{\sqrt{81}} = 1$$

Ce qui confirme la valeur du coefficient de corrélation entre deux variables qui se distribuent de façon analogue pour une même série d'observations.

Une fois la matrice des corrélations obtenue, qu'elle ordonne des valeurs de r ou des valeurs de Φ , il est relativement aisé de passer à une autre étape de la réflexion géographique en recherchant les divers regroupements de variables qu'elle contient. Il s'agit alors d'identifier le coefficient le plus élevé et d'associer les autres variables qui ont une relation relativement marquée. Cette technique de groupement de variables, parce qu'elle débouche directement sur les problèmes d'analyse typologique sera expliquée en détail dans la section que nous consacrons aux différentes classifications. Mais ceux-ci nécessitent très souvent que la matrice des corrélations soit préalablement réduite. A cette fin, le recours aux techniques factorielles est obligatoire.

2.2) LES ETAPES DE LA REDUCTION DES SYSTEMES SPATIAUX

2.2 a) La réduction des attributs spatiaux: les analyses factorielles, leurs contraintes et leurs problèmes mathématiques

Tel qu'indiqué auparavant le souci constant du géographe est de trouver, décrire et expliquer la structure sous-jacente d'un ensemble de données. La technique des analyses factorielles qui a eu ses débuts en psychologie (Pearson, 1904) a été adoptée par les géographes vers les années

60, (Berry, 1961) parce qu'elle permettait de définir des groupes indépendants de variables qu'on a qualifiés de facteurs pour les analyses factorielles proprement dites et de composantes pour les analyses en composantes principales, l'une des solutions particulières du modèle général de l'analyse factorielle.

Cette technique permet donc de réduire le nombre de variables originales de la matrice d'information spatiale tout en conservant les principaux éléments structurels.

En premier lieu il faut distinguer entre l'analyse factorielle et l'analyse en composantes principales; les deux méthodes sont trop souvent confondues. L'analyse de composantes principales, développée par Hotelling, cherche à extraire le maximum de la variance à l'intérieur d'une matrice des corrélations donnée à partir de l'équation suivante:

$$Z_j = a_{j1} F_1 + a_{j2} F_2 + \dots + a_{jm} F_m$$

$$Z_j = \sum_{p=1}^m a_{jp} F_p \quad \begin{array}{l} j = 1 \text{ à } n \\ i = 1 \text{ à } m \end{array}$$

n = nombre d'observations
 m = nombre de variables
 $a_{jp} F_p$ = la saturation de l'observation j sur le facteur p

Nous reviendrons à la définition et à l'explication des saturations un peu plus loin.

Une propriété importante de cette méthode, c'est qu'elle tient compte de la somme des variances des m variables. Pour un problème particulier quelques composantes seulement ont un pourcentage d'explication.

Contrairement à l'analyse en composantes principales, l'analyse factorielle essaie de reproduire de façon maximale les corrélations. Cette méthode oblige à distinguer entre facteurs communs et facteurs uniques. Le facteur commun intervient dans plusieurs variables tandis que le facteur unique tient compte du reste de la variance en incluant l'erreur d'une variable spécifique. L'équation de l'analyse factorielle est quasiment identique à celle de la composante principale sauf qu'il faut ajouter l'erreur résiduelle.

$$Z_{ji} = \sum_{p=1}^m a_{jp} F_{pi} + d_{j} U_{ji}$$

Harman (1967)

$d_{j} U_{ji}$ = erreur résiduelle
de la variable i

j = 1, 2... N

i = 1, 2... M

p = nombre de facteurs

L'analyse factorielle estime la variance commune ou les communautés qu'elle introduit dans la diagonale de la matrice des corrélations. Par contre l'analyse des composantes utilise l'unité (1,0) dans la diagonale.

Dans le tableau (9) on retrouve deux matrices. La première est destinée à une analyse de composantes princi-

pales et la deuxième à l'analyse factorielle. Le seul point de discordance entre les deux est la diagonale.

TABLEAU 9

MATRICE D'ENTREE DE L'ANALYSE EN COMPOSANTES PRINCIPALES

Variabes	1	2	3	4	5	6
1	1.0	.72	.63	.54	.45	.36
2	.72	1.0	.56	.48	.40	.32
3	.63	.56	1.0	.42	.35	.28
4	.54	.48	.42	1.0	.30	.24
5	.45	.40	.35	.30	1.0	.20
6	.36	.32	.28	.24	.20	1.0

MATRICE D'ENTREE DE L'ANALYSE FACTORIELLE

	1	2	3	4	5	6
1	(.81)	.72	.63	.54	.45	.36
2	.72	(.64)	.56	.48	.40	.32
3	.63	.56	(.49)	.42	.35	.28
4	.54	.48	.42	(.36)	.30	.29
5	.45	.40	.35	.30	(.25)	.20
6	.36	.32	.28	.24	.20	(.16)

Source: Thomson, 1950.

.81 représente une communauté qui est estimée ici de la façon suivante:

	1	3
1	h_1^2	.63
2	.72	.56

$$.56h_1^2 - (.72 \times .63) = 0$$

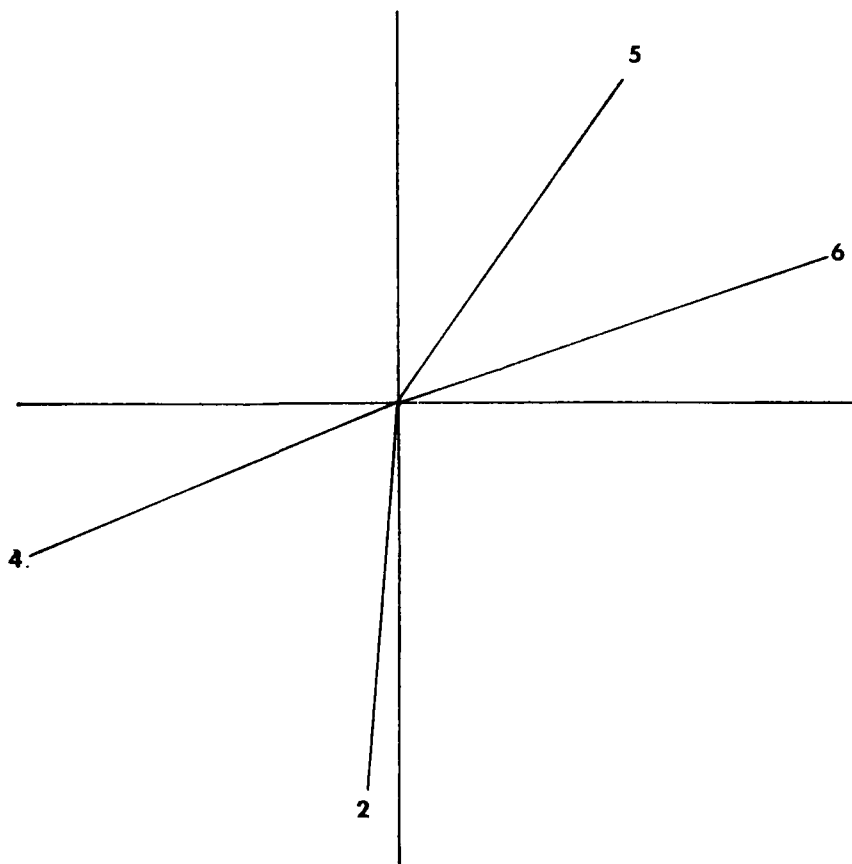
$$h_1^2 = .81$$

Plus loin on constatera que la communauté d'une variable est égale à la somme des carrés de ses saturations. Pour revenir à la matrice de corrélation notons que, étant donné que cette matrice contient toutes les relations existantes entre les données originales il est possible de représenter ces variations à l'aide de vecteurs. Chaque variable est un vecteur dont la longueur et la direction sont déterminées par sa variation. Ainsi deux variables parfaitement corrélées ($r=1$) ont une longueur et une direction identiques et par le fait même sont superposées dans un graphique à deux dimensions. Par ailleurs si elles sont complètement indépendantes ($r=0.0$) on dit qu'elles sont orthogonales, c'est-à-dire perpendiculaires l'une à l'autre. De ces constatations on déduit que le coefficient de corrélation est une fonction des angles existants entre deux variables. La relation est la suivante: le coefficient de corrélation entre deux varia-

bles est identique à la valeur du cosinus de l'angle entre les vecteurs respectifs de chaque variable. Lorsque les deux droites sont superposées, l'angle qu'elles forment, est de 0. En vérifiant dans une table de cosinus on trouve que le cosinus de 0 est égal à 1 ce qui corrobore adéquatement l'existence d'un coefficient de corrélation parfait.

La figure (17) indique graphiquement les corrélations entre quelques variables (2, 4, 5, 6) de notre exemple de l'utilisation du sol.

FIGURE 17



SOUS-MATRICE DES - CORRELATIONS ENTRE - LES
VARIABLES 2-4-5-6 - DE - L'UTILISATION DU SOL

	2	4	5	6
2	1.0(0°)	.40(66°)	-.44 (146°)	-.46(159°)
4		1.0 (0°)	-.67 (160°)	-.58(155°)
5			1.0 (0°)	.81(35°)
6				1.0 (0°)

* Les chiffres entre parenthèses correspondent à l'angle entre les variables.

Le tableau contient les vecteurs de corrélation entre les variables 2 (terrain vacant), 4 (résidentiel unifamilial), 5 (résidentiel bifamilial) et 6 (résidentiel multifamilial). Nous aurions pu tracer les douze variables sur le graphique mais nous avons cru bon de ne pas surcharger le graphique pour qu'il soit plus lisible et plus compréhensible. On voit que ce graphique est composé de deux groupes. Le premier comprend les variables 5 et 6 et le deuxième les variables 2 et 4. De plus ces deux groupes sont situés de part et d'autre de la jonction des axes et présentent une configuration oppositionnelle entre ce qu'on pourrait appeler la banlieue avec ses résidences unifamiliales et la ville avec ses quartiers à loyers multiples.

L'analyse en composantes principales exécutée à l'aide du calcul matriciel ce que nous venons de faire à l'échelle graphique en tenant compte cependant de toutes les variables simultanément. Les graphes de variables obtenus forment donc des "facteurs" pour lesquels la variation est minimisée entre les variables d'un même groupe.

Avec ceci en tête poussons plus loin la démarche. Nous tenons à souligner auparavant qu'il n'est pas possible de faire entrer dans une analyse factorielle une variable qui ne contient que des zéros. Il est admissible toutefois d'ajouter une constante tel que un (1) à chaque observation.

Avant de calculer la matrice de corrélation il est préférable de standardiser les données pour enlever les inconsistances des variables non standardisées. Ceci ne change en rien la matrice de corrélation puisqu'elle représente uniquement la relation entre les variables mais si le géographe veut par quelque moyen que ce soit comparer le comportement de ses observations sur un facteur particulier il devra standardiser ses données avant de faire le calcul des poids locaux.

Résumons: on débute avec une matrice d'information spatiale, (n observations par m attributs), que l'on standardise pour obtenir une seconde matrice de n observations par m attributs avec des moyennes de 0 et des écarts-types de 1. Tableau (10).

TABLEAU 10

MATRICE STANDARDISEE

	MATRICE STANDARDISEE											
I BOUCHERVILL F	-0.1633	1.3204	0.4124	0.7028	-0.4214	-0.3118	-0.5382	-0.5461	-0.1354	-0.5912	-0.5671	-0.2653
I BOSSARD	-0.9220	1.7575	-0.0188	0.1622	-0.5539	-0.4280	0.3117	-0.7514	-0.9583	1.1337	-0.4373	-0.3174
I CANDIAC	-0.6974	0.8833	-1.5111	0.9635	-0.5855	-0.4280	-0.6485	1.0719	-0.9226	-0.5567	0.5643	-0.4043
I CAUGHNAWAGA	-1.0313	-0.5599	-0.1845	0.9780	-0.7085	-0.4280	-0.6485	1.2745	-0.7795	-0.7981	-0.0668	-0.3521
I CHATEAUGUAY	-0.3060	0.0499	0.8303	0.0029	-0.5392	0.0951	0.4479	-0.6785	1.3644	0.1678	-0.5133	-0.1734
I CHATEAUGUAY-CENTRE	-0.6883	0.4666	0.7574	0.8235	-0.5097	-0.4280	-0.6290	-0.5704	-0.4938	0.4783	-0.3593	-0.5258
I CHATEAUGUAY-HIGHTS	0.5377	1.6456	0.3726	0.8525	-0.4435	-0.4280	-0.6485	-0.7109	-0.1722	-1.0395	-0.5285	-0.2479
I DELSON	-0.3393	0.4870	-2.1545	0.9442	-0.7306	0.1532	-0.6485	-0.7514	-1.0655	-0.6946	3.1134	0.2732
I GREENFIELD-PARK	1.7820	-0.0009	0.5318	0.2877	0.3297	-0.1374	-0.6485	-0.5732	0.1494	2.9621	-0.5286	-0.5258
I JACQUES-CARTIER	-0.3211	0.2938	-0.1115	-0.2626	0.4401	-0.1955	-0.0127	-0.7190	1.4001	0.3743	-0.0375	0.6206
I LAFLECHE	1.6749	1.1273	0.1007	-0.5233	0.1530	-0.4280	0.6101	-0.6137	0.0065	0.2713	-0.3593	-0.2305
I LA PRAIRIES	-0.8522	-0.8953	-2.4596	-0.9192	1.6403	0.3276	-0.2138	0.3331	0.0422	-0.3497	2.9732	-0.1611
I LEMOYNE	1.7820	-1.4442	1.0956	-2.2949	2.9215	4.2803	0.0262	-0.6704	0.9713	1.8925	-0.1559	-0.0915
I LFRY	-0.6094	-0.8852	0.1405	-2.1839	-0.6436	-0.4280	3.5749	1.7340	-1.1370	-1.0051	-0.6286	-0.0047
I LONGUEUIL	1.4087	-1.5255	0.6778	-1.4550	2.3693	1.0833	-0.2592	-0.5326	3.0795	0.7197	0.2487	-0.4737
I NOTRE-DANE	1.7122	-0.5701	1.7987	-0.9771	-0.6717	-0.4280	1.9466	-0.7028	-0.6724	-1.0051	-0.5285	0.0125
I PREVILLE	0.1887	-0.6819	-0.6886	0.9876	-0.7306	-0.4280	-0.6485	2.7332	-0.8511	-1.0396	-0.6286	-0.5258
I ST-BRUNO	-0.9045	-1.3731	0.6778	0.6594	-0.2520	-0.0793	-0.6485	0.4884	-0.2079	0.2713	0.0324	-0.4911
I STE-CATHERINE	-0.8097	-0.5599	-0.1845	0.9780	-0.7085	-0.4280	-0.6485	1.2745	-0.7796	-0.7981	-0.0568	-0.3521
I ST-CONSTANT	-1.1557	1.1578	0.2931	-0.4992	0.1603	-0.4280	0.5712	-0.7514	0.0780	0.3059	-0.4747	-0.2305
I ST-HUBERT	-0.8006	0.3548	-0.9208	0.3311	-0.2888	-0.4280	-0.1489	-0.5245	0.2209	-0.5945	-0.3440	4.4073
I ST-LAMBERT	1.1750	-1.0478	0.5451	0.4421	-0.1195	-0.0793	-0.4992	-0.1031	0.8641	-0.0047	0.0486	0.0647

Si la moyenne n'est pas égale à 0 plus ou moins 0,0005 et l'écart-type à 1 plus ou moins 0,0005 il faut supposer qu'il y a erreur dans le procédé de standardisation ou dans la lecture des données par l'ordinateur.

Une fois la matrice standardisée obtenue, le programme calcule et ordonne la matrice des corrélations sur laquelle il est inutile que nous revenions.

C'est à l'aide de la matrice de corrélation qu'est effectuée l'analyse des composantes principales. Il s'agit en fait, ni plus ni moins de résoudre un système de plusieurs équations linéaires. Il y a autant d'équations qu'il y a de variables; si on a douze variables l'ordinateur doit résoudre simultanément douze équations linéaires qui s'écrivent de la façon suivante:

$$\begin{vmatrix} 1 - \lambda & r_{12} & r_{13} & \dots & r_{1p} \\ r_{21} & 1 - \lambda & r_{23} & \dots & r_{2p} \\ r_{31} & r_{32} & 1 - \lambda & \dots & r_{3p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 - \lambda \end{vmatrix} \cdot \begin{vmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \cdot \\ \cdot \\ a_{p1} \end{vmatrix} = 0$$

où r = les coefficients de corrélation, λ = les eigenvalues,
 a = les saturations.

et que l'on résume sous la forme suivante:

$$(R - \lambda I) = 0$$

où R est la matrice de corrélation, λ représente les valeurs propres ou caractéristiques (eigenvalues) qui expriment le pourcentage expliqué de la variance totale et I une matrice unité de travail. Hotelling () a démontré que la solution à trouver était la suivante:

$$F \cdot F^T = R \quad \text{et} \quad F^T \cdot F = \Lambda$$

où F est la matrice des saturations unissant chacune des variables à chacune des composantes. Cette matrice des saturations qui post-multipliée par sa transposée ($F \cdot F^T$) va donner la matrice de corrélation et qui, prémultipliée par cette même transposée ($F^T \cdot F$) procure les racines caractéristiques.

Prenons encore une fois l'exemple de l'utilisation du sol. Le tableau (11) contient les saturations obtenues par l'analyse en composantes principales, les "eigenvalues" respectives à chaque composante et les communautés de chaque variable.

Le nombre de composantes ainsi obtenu est égal au nombre de variables (12). La réduction de cette matrice se fait en coupant le degré de signification des eigenvalues selon le degré de précision voulue, habituellement 1,0; parce qu'à moins de 1,0 chaque nouvelle dimension "reproduit" moins de la variance totale que chacune des variables originales, ce qui a réduit dans le cas présent à 4 le nombre de composantes

délimitées par la ligne pointillée sur le tableau (11).

Les saturations permettent d'établir quelles sont les variables importantes dans la définition de chacun des facteurs, et à quel degré elles sont mathématiquement équivalentes à des coefficients de corrélation et s'interprètent de la même manière. En somme la saturation mesure le degré et la direction de chaque variable avec chaque facteur, (Rummel, 1967).

La relation entre les coefficients de corrélation et la saturation se définit ainsi:

$$r_{xy} = a_{xp1} \cdot a_{yp1} + a_{xp2} a_{yp2} + \dots + a_{xpn} a_{ypn} \quad \text{eq. 2.2}$$

où r_{xy} = le coefficient de corrélation entre deux variables x et y

a_{xp} = la saturation de la variable x sur le facteur p

a_{yp} = la saturation de la variable y sur le facteur p

Si on se rappelle la corrélation de -.27 entre la variable de l'importance du terrain urbanisé (variable no. 1) et celle du terrain vacant (variable no. 2) on note à partir du tableau (11) que ces deux variables ont obtenu les saturations suivantes sur les quatre premiers facteurs-clé (cf. les deux premières lignes du tableau 11, les 4 premières colonnes)

$$a_{xp1} = .70 \quad a_{xp2} = -.19 \quad a_{xp3} = .05 \quad a_{xp4} = -.19$$

$$a_{yp1} = .46 \quad a_{yp2} = -.39 \quad a_{yp3} = .59 \quad a_{yp4} = .28$$

TABLEAU 12

MATRICE DES VECTEURS PROPES

VECTEUR 1	0.34845	-0.22665	0.25689	-0.36364	0.42344	0.39356	0.07127	-0.20589	0.36660	0.32584	-0.06332	-0.05242
VECTEUR 2	-0.13317	-0.27209	-0.52578	-0.00247	0.29215	0.24125	-0.27993	0.21907	0.05267	-0.04605	0.59704	0.02283
VECTEUR 3	0.03552	0.40122	-0.04321	0.39339	0.03622	-0.01219	-0.56021	-0.47344	0.18635	0.31636	0.05831	0.05894
VECTEUR 4	-0.16509	0.24609	-0.21322	-0.28643	0.06464	-0.02459	0.33208	-0.36575	0.10514	-0.14653	0.10719	0.70130

En combinant le tout dans l'équation (2,2) on obtient:

$$\begin{aligned} r &= (.70 \times -.46) + (-.19 \times -.33) + (.05 \times .54) + (-.19 \times .28) \\ &= -3220 + 0741 + 0971 - 0532 \\ &= .2781 \text{ résultat identique au coefficient de corrélation.} \end{aligned}$$

Les saturations proviennent de la matrice des vecteurs propres (tableau 12) qui eux sont calculés à partir de la matrice de corrélation. Les vecteurs propres sont similaires aux saturations puisqu'ils mesurent la direction de la liaison existant entre les variables et les facteurs cependant ils ne prennent pas en considération l'importance relative des différents facteurs. La matrice des saturations est en somme la matrice des vecteurs propres pondérée. Cette pondération est définie par l'équation suivante:

$$a_{ip} = V_{ip} \sqrt{\lambda_p} \quad \begin{array}{l} i = 1 \text{ à } m \\ p = 1 \text{ à } n = \text{facteurs} \end{array}$$

où a_{ip} est la saturation de la variable i sur le facteur p , V_{ip} est le vecteur propre de la variable i sur le facteur p et λ_p est la valeur caractéristique du facteur p . D'après le tableau des saturations, celle de la variable du sol urbanisé .7045 est obtenue en multipliant 0,34, vecteur propre de cette même variable, par la racine carrée de la valeur propre du premier facteur.

Voici le calcul de la première saturation sur la première composante de l'utilisation du sol.

$$\begin{aligned}
 a_{ip} &= v_{ip} \sqrt{\lambda_p} & p = 1 \\
 &= .34845 \sqrt{4.08} \\
 &= .34845 \times 2.08 \\
 a_{ip} &= 0.70451
 \end{aligned}$$

Il est possible de déterminer les communautés à l'aide de la matrice des saturations. On a vu au début du présent chapitre qu'elles correspondent à la diagonale de la matrice de corrélation et qu'elles sont égales à un pour l'analyse en composantes principales et inférieures à un lorsqu'il s'agit de l'analyse factorielle.

La communauté d'une variable révèle la quantité totale de la variation de cette variable particulière qui peut être décrite par toutes les autres dans l'analyse (Rummel, 1967) c'est-à-dire la variation qui est en commun avec toutes les autres variables. Ainsi une variable avec une forte communauté contribue beaucoup à définir les facteurs, en opposition avec une autre, dont la communauté est faible, et qui par conséquent a une participation minime à l'analyse. On obtient en définitive les communautés en additionnant le carré des saturations de chaque rangée de la matrice de saturation.

$$h_i^2 = \sum_{k=1}^p (a_{ik})^2 \quad a_{ip} = \text{saturation de la variable } i \text{ sur le facteur } p$$

Expliquons ce qui précède avec l'exemple de la première rangée de la matrice de saturation :

Communauté de l'importance
 du terrain urbanisé après = $(0.70)^2 + (-.19)^2 + (0.04)^2 + (0.18)^2$
 4 composantes, = 0.57

Avec douze composantes elle devient égale à 1.

De la même manière qu'on a calculé les communautés pour les rangées il est possible de calculer les valeurs propres pour les colonnes de la matrice de saturations en faisant l'addition des carrés des valeurs de chaque colonne. La valeur propre d'un facteur mesure la quantité de la variation dans l'analyse qui est expliquée, ou prise en compte par ce facteur. Tout comme les communautés, les valeurs propres sont divisibles par le nombre total de variables pour obtenir le pourcentage d'explication de chaque facteur. Il est certain que la somme des valeurs propres sera toujours égale à la somme des communautés. Une différence marquée entre les deux sommes dénote une anomalie au niveau du calcul des facteurs.

Voici le calcul de la valeur propre du premier facteur de l'utilisation du sol:

$$= (0.70)^2 + (-.45)^2 + (.51)^2 + (-.73)^2 + (.85)^2 + (.79)^2 \\ + (.14)^2 + (-.41)^2 + (.74)^2 + (.65)^2 + (-.12)^2 + (-.10)^2 \\ = 4.08$$

Pour les autres facteurs on obtient:

$$\lambda_2 = 2.07 \quad \lambda_3 = 1.77 \quad \lambda_4 = 1.30 \quad \lambda_5 = 0.91 \quad \lambda_6 = 0.57 \quad \lambda_7 = 0.50 \\ \lambda_8 = 0.35 \quad \lambda_9 = 0.32 \quad \lambda_{10} = 0.08 \quad \lambda_{11} = 0.0006 \quad \lambda_{12} = 0.00001$$

qui, transformés en pourcentage donnent $\frac{4.08}{12 \text{ variables}} \times 100$

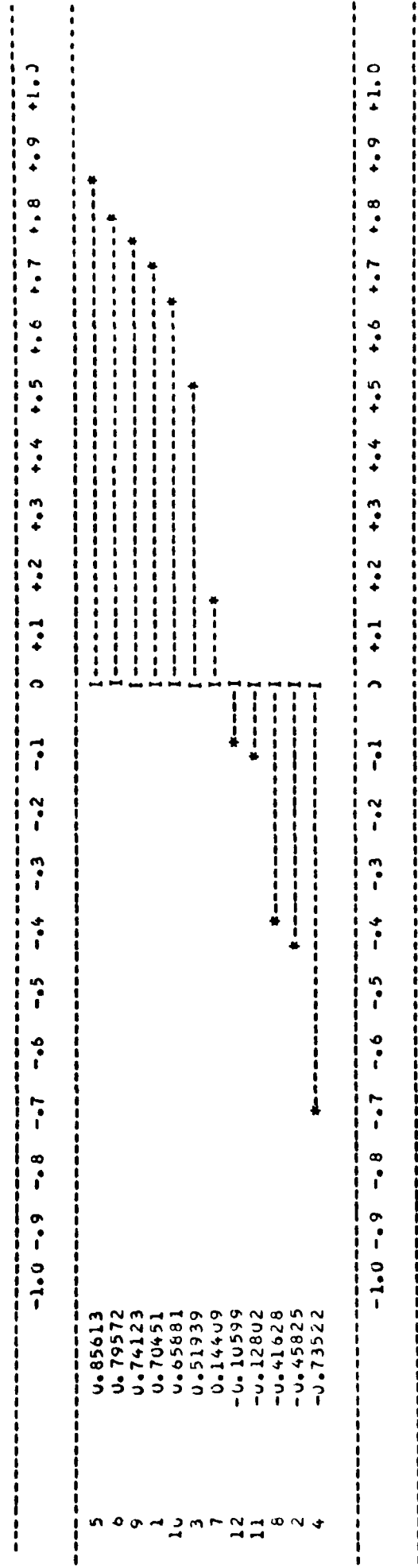
34% pour le premier facteur, 17% pour le deuxième, 15.8% pour le troisième, finalement 10.8% pour le quatrième pour un total de .77 ou 77% d'explication après 4 facteurs. Une question se pose alors: combien de facteurs sont vraiment significatifs? En d'autres mots à quel moment l'addition d'un autre facteur n'apporte presque plus d'explication? Les statisticiens s'accordent à dire que les facteurs qui expliquent moins de 5% de la variation totale ne devraient pas être inclus dans l'analyse. De plus une analyse avec plus de 4 à 5 facteurs devient difficile d'interprétation et pose des problèmes sérieux quant à la dénomination des facteurs.

La somme de toutes les communautés, divisée par le nombre total des variables et multiplié par 100 donne le pourcentage de la variance totale déterminé par toutes les composantes pour toutes les variables. Une analyse sera d'autant meilleure que cette valeur sera plus grande avec un nombre restreint de facteurs.

Il faut savoir tirer le plus d'information possible de la matrice des saturations. Pour le géographe c'est à ce moment précis qu'entre en jeu sa compétence puisqu'il doit trouver un nom ou identifier chaque facteur à partir des variables déterminantes de ce facteur. Prenons le premier facteur et plaçons ses saturations en ordre décroissant (figure 18).

FIGURE 18

PRESENTATION GRAPHIQUE D'UN FACTEUR



Le graphique représente de part et d'autre de l'axe les variables positives et négatives. Son utilité, c'est qu'il nous permet de voir d'un seul coup d'oeil la bipolarité ou l'unipolarité des facteurs. Il y a bipolarité lorsque le graphique tend à être symétrique c'est-à-dire que pour chaque saturation positive il y a une saturation négative relativement proche en valeur absolue. Par contre l'unipolarité présente une asymétrie soit à gauche ou à droite de l'axe, signifiant que le facteur est défini que par des variables positives (à droite) ou négatives (à gauche),

Revenons au premier facteur où on remarque une nette opposition entre le groupe des variables 5 (bifamilial: .85), 6 (multifamilial: .79) et 9 (institutionnel: .74) et le groupe des variables 4 (unifamilial: .73) et 2 (terrain vacant: .43).

En général les saturations comprises entre -40 et 40 ont un très faible poids dans la détermination du facteur et à moins que l'une d'entre elles soit très significative pour donner plus de force au facteur, on les exclut. Avec le premier facteur on remarque une nette opposition entre le bifamilial et l'unifamilial que J.B. Racine (1971) a défini comme étant le niveau de densité et de complexité de l'urbanisation suburbaine et qui renferme de manière latente le niveau d'urbanisation.

Voilà ce qui en est pour l'analyse en composantes principales. Cependant nous avons développé, au cours des trois dernières années une méthode d'analyse qui permette de tenir compte du poids des communautés en les estimant d'une manière plus sûre que dans le modèle traditionnel de l'analyse factorielle. Le procédé est relativement simple. Il s'agit en premier lieu de faire une analyse en composantes principales, de la même manière que nous venons de voir. Les communautés obtenues avec les eigenvalues supérieurs à 1 sont alors introduites dans la diagonale de la matrice de corrélation. Ceci joue donc un rôle de pondération en favorisant les variables qui entrent dans la détermination du facteur. Cette nouvelle matrice de corrélation à diagonale pondérée est soumise de nouveau à l'analyse en composantes principales. Le résultat obtenu met en relief les variables qui au départ s'avéraient significatives et atténue celles qui semblent moins importantes, sans avoir à les éliminer totalement de l'analyse. Cette technique ressemble beaucoup à l'analyse factorielle qui estime les communautés au départ. Mais avec notre méthode nous avons la vraie valeur des communautés puisqu'elles sont calculées. De plus elles sont fonction du nombre de facteurs car après 4 facteurs les communautés ne sont pas identiques à celles obtenues après cinq facteurs. Cette méthode tente ainsi de minimiser la variation spécifique des variables et de maximiser la variation commune de l'ensemble des variables. Ce

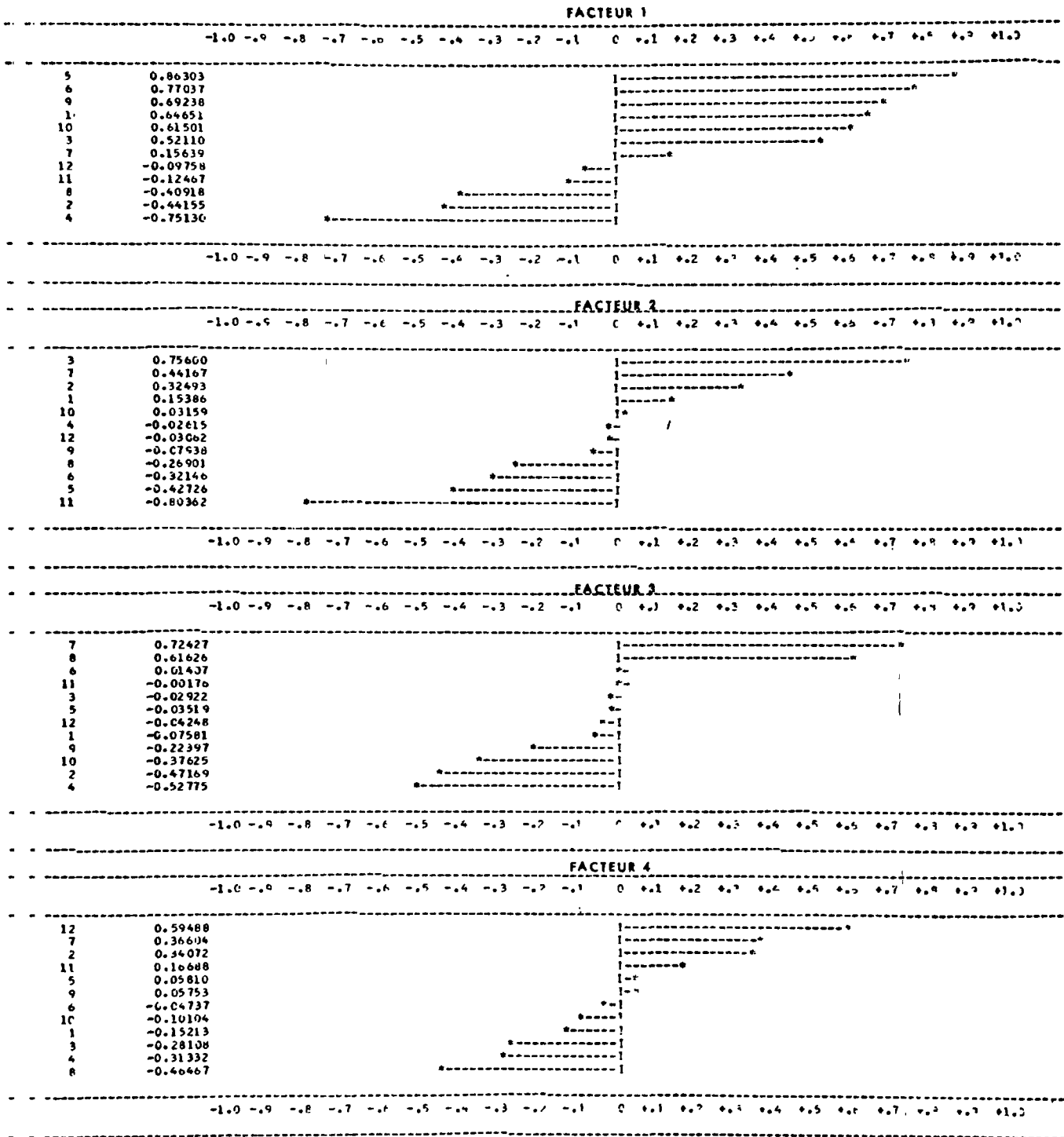
procédé est donc une sorte de filtre qui permet de mieux définir les éléments déterminants de la structure. Le tableau (13) illustre les résultats obtenus avec cette méthode, avec l'exemple des douze variables. Par contre si on veut éliminer totalement la variation spécifique il faut avoir recours à l'analyse-image.

2.2 b) L'analyse-image

C'est à Guttman (1953), Harris (1963) et Kaiser (1963) que revient l'éloge d'avoir développé cette méthode, basée sur le principe de la maximisation des aspects "communs" mesurés par les variables pour enlever complètement la partie spécifique de la variance des variables originales. Géométriquement l'"image analysis" est la projection ou l'image d'une variable dans l'espace déterminé par les autres variables.

Pour solutionner ce problème il faut trouver une matrice qui contient les covariances représentatives des relations existantes seulement entre les parties communes de la variation originale c'est-à-dire les relations partagées entre deux variables ou plus. La quantité totale de la variation pour une variable donnée est le carré de la corrélation multiple, que l'on peut atteindre en le prédisant à partir de toutes les autres variables de l'ensemble. Ces valeurs sont introduites dans la diagonale de la matrice pour chaque va-

FACTEURS DE LA METHODE FACTORIELLE-COMPOSANTES (FALCOMP)



riable respective, toutes les autres entrées de la matrice sont calculées seulement pour représenter la covariation collective.

Voici les étapes du calcul:

1. Calcul de la matrice de corrélation.
2. Calcul de son inverse, c'est-à-dire que si la matrice de corrélation est prémultipliée par son inverse on obtient une matrice identité

$$R^{-1} R = I$$

3. Calcul de la matrice de la covariance: chaque élément de la diagonale est l'anti-image. Le résultat est égal à la division de 1 par son élément correspondant de la matrice inverse.

$$S_i = 1.0/R_{i,j}^{-1} \quad \begin{array}{l} i = 1 \text{ à } m \\ j = i \end{array}$$

4. Calcul des autres éléments de la matrice

$$X_{i,j} = S_i \cdot R_{i,j}^{-1} \cdot S_j$$

5. Calcul de la matrice de la covariance diagonale

$$i,j = R_{i,j} + X_{i,j} - 2S_i \quad i = j$$

6. Calcul des saturations sur cette dernière comme dans le cas de l'analyse en composantes principales.

On a noté que lorsqu'il y a un fort pourcentage de la variation spécifique dû à certaines variables le nombre de facteurs extraits est inférieur au nombre obtenu par l'analyse de composantes. Kaiser (1963) a suggéré, pour remédier à cette lacune, d'extraire un nombre de facteurs égal à la moitié du nombre original des variables en diminuant le niveau de rupture des valeurs propres, (.05 au lieu de 1.0).

L'image analysis a des implications lorsque le chercheur essaie de trouver une structure factorielle dont les facteurs ne sont pas définis par une ou deux variables mais par un ensemble de variables. Dans ce cas ce ne sont pas quelques variables qui servent à identifier le facteur mais l'ensemble. Il arrive parfois, qu'après une analyse en composantes principales, certains facteurs sont difficiles à interpréter et semblent litigieux. Il est souhaitable dans ce cas de faire une "image analysis" pour tenter de trouver une solution pertinente au problème posé.

Cette dernière méthode factorielle complète les principales façons de déterminer une structure à partir d'un ensemble de variables données. Cependant une question demeure: cette structure est-elle optimale?

Les facteurs se doivent d'être orthogonaux c'est-à-dire perpendiculaires l'un à l'autre pour présenter aucune corrélation entre eux. Vérifions ceci avec les deux premiers facteurs de l'analyse en composantes principales en plaçant com-

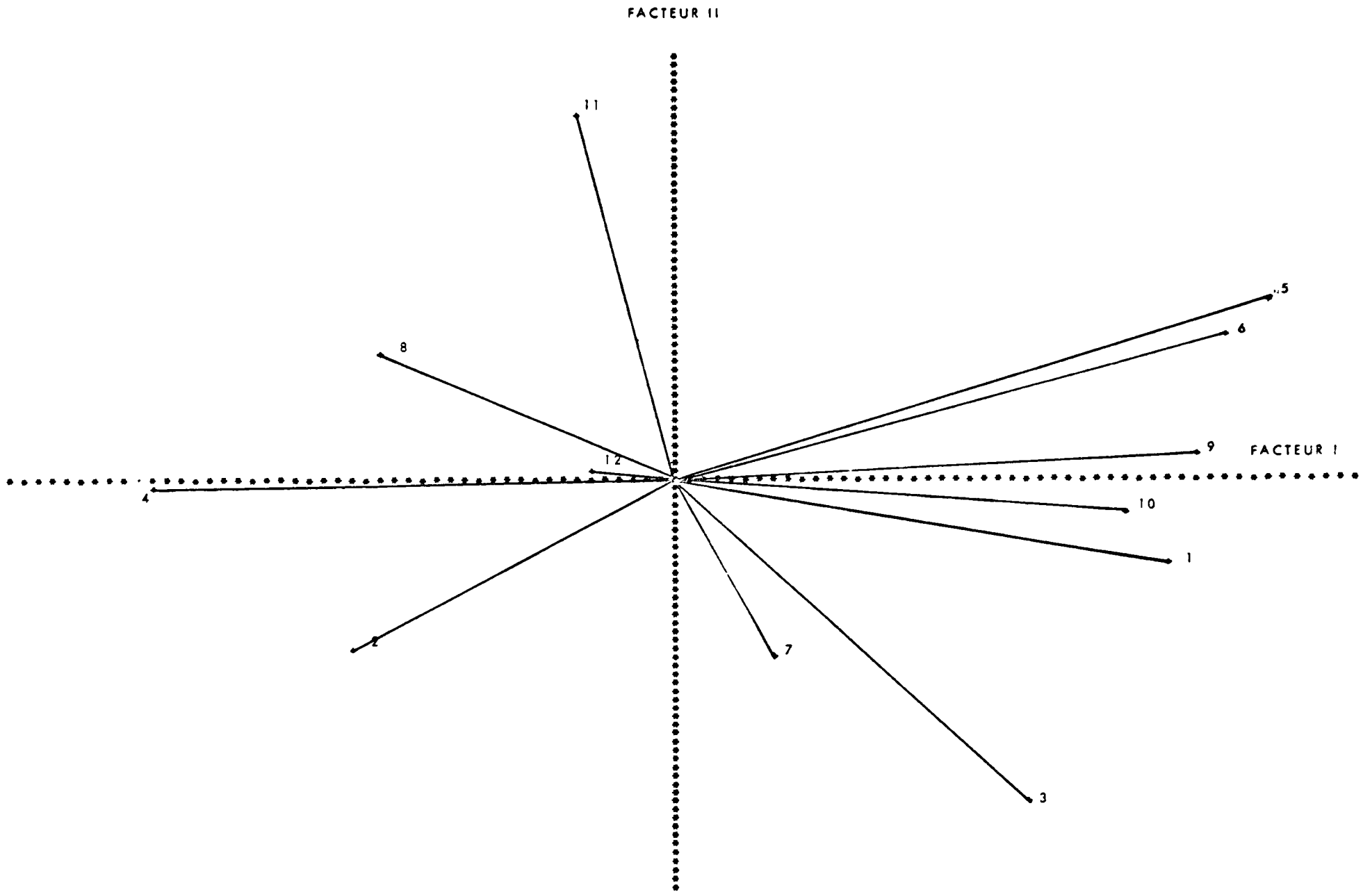
me axe horizontal d'un graphique le premier facteur et comme axe vertical le second. On obtient donc le graphique suivant. (Figure 19)

L'interprétation de ce graphique illustre très bien les deux premiers facteurs. On remarque une nette opposition entre le groupe des variables 5, 6, 9 et le groupe des variables 2 et 4. Ceci détermine donc le premier facteur puisque ce sont les oppositions les plus évidentes; vient par la suite le deuxième facteur défini par l'antagonisme entre le groupe des variables 11 et 2 et celui des variables 1 et 3. Si on avait voulu introduire sur le graphique le troisième facteur il aurait fallu déterminer la position respective de chaque point dans les trois dimensions ce qui devient un peu plus compliqué. L'avantage de l'ordinateur c'est qu'il peut saisir simultanément de manière imaginaire plusieurs dimensions.

A l'observation de ce graphique on s'aperçoit qu'une rotation d'axe pourrait maximiser les variations maximum, et permettrait ainsi de voir si les liens entre les variables tentent de se regrouper. Plusieurs types de rotation ont été développées selon le but poursuivi; nous en avons retenu trois sur lesquelles nous allons nous pencher, la rotation Varimax, Equimax et Quartimax.

FIGURE 19

REPRESENTATION GRAPHIQUE DES FACTEURS 1 ET 2



2.2 c) La rotation Varimax

C'est la plus employée des trois et tout comme son nom l'indique ce procédé ajuste les axes (les facteurs) en séparant de façon maximale les groupes définis par les facteurs non rotationnés. Comme dans la solution initiale les facteurs sont indépendants les uns des autres. De plus le nombre de variables ayant des saturations fortes sur un seul facteur est minimisé de manière à ce qu'on obtienne une plus juste représentativité des variables originales. La solution Varimax nécessite le recours à un processus itératif qui calcule les variances de la matrice factorielle et qui s'arrête lorsque la différence de la variance entre la dernière et l'avant-dernière itération est plus petite ou égale à 10^{-7} .

De plus il est important de souligner que les "valeurs-propres" ne sont plus identiques pour des facteurs rotationnés puisqu'on ne recalcule pas les racines de la matrice. Il faut faire appel au pourcentage d'explication de chaque facteur rotationné qui est calculé en faisant la somme des carrés des saturations des facteurs divisée par le nombre de variables.* Exemple:

$$\% \text{ expliqué} = \sum_{i=1}^n (a_{i,j})^2 \times 100 \quad j = 1$$

* Nous tenons à remercier M. Roberge, professeur au département de géographie de l'Université d'Ottawa qui nous a fait remarquer cette différence et nous a fourni en même temps le programme qui calcule ce pourcentage d'explication.

Au tableau (14) on retrouve la matrice exprimant le résultat de l'analyse de l'utilisation du sol après la rotation Varimax. Les pourcentages d'explication de chaque facteur après la rotation sont également donnés.

La question de savoir quand choisir la solution Varimax fait l'objet de nombreux débats. Toutefois l'expérience de plusieurs analyses nous a démontré que la rotation Varimax n'était pas souhaitable avec des données purement physiques,

TABLEAU 14

SATURATIONS: ROTATION VARIMAX

ROW	1	2	3	4
1	0.6100	-0.4341	-0.0396	-0.0943
2	-0.4339	-0.1451	0.3291	0.6416
3	0.2490	-0.9085	-0.1346	-0.0307
4	-0.5848	0.0889	0.7547	0.0593
5	0.9340	0.1612	-0.1246	-0.0541
6	0.8423	0.0805	-0.1204	-0.1571
7	-0.1316	-0.2682	-0.8925	-0.0057
8	-0.4260	0.2812	-0.1585	-0.7488
9	0.7657	-0.1073	-0.0100	0.1830
10	0.6797	-0.3084	0.2800	0.0951
11	0.1500	0.8513	0.1453	-0.0964
12	-0.0679	0.3157	-0.3316	0.6677

Pourcentage d'explication pour chaque facteur.

1	0.3183603
2	0.1797800
3	0.1465821
4	0.1260115

des données exprimant une réalité brute et immédiate telles que l'élévation, la profondeur de la nappe phréatique, la distance, etc. tandis qu'avec les pourcentages et les différentes variables sociologiques démographiques et économiques la solution de la rotation s'avère plus efficace, tout particulièrement lorsque les données "d'entrées" sont exprimées en pourcentage.

Enfin il faut noter qu'étant donné que les facteurs rotationnés représentent la meilleure approximation d'un noyau de variables, le retrait d'une ou deux d'entre elles n'affecte pas beaucoup la structure finale. Par contre le fait d'exclure une ou plusieurs variables des facteurs initiaux non rotationnés peut modifier sensiblement les résultats et donner une image inadéquate de la réalité.

2.2 d) La rotation Quartimax

Toujours en quête de tirer le meilleur parti des saturations, Neuhaus et Wrigley (1954) proposèrent que l'inégalité parmi les carrés des saturations soit maximisée ou plus précisément que la variance dans la distribution des saturations au carré soit maximum. Cette méthode oblige à monter la matrice des saturations à la quatrième puissance, d'où le nom Quartimax suggéré par Burt (1952).

Littéralement on peut dire que la rotation quartimax est une transformation T sur une matrice de saturations ori-

ginale F pour obtenir une nouvelle matrice de saturation rotationnée R dans laquelle la variance des saturations au carré est maximum.

$$R = F^T$$

Contrairement à la solution Varimax qui mettait l'emphase sur la simplification des colonnes pour trouver la structure optimum, celle-ci s'emploie à simplifier les rangées (les variables) de la matrice factorielle.

On se souviendra que le quatrième moment (cf. normalisation, 1ère partie) représente la kurtose et que cette mesure du degré de courbure réapparaît ici en montant les carrés des saturations à la quatrième puissance. On peut facilement prévoir que si on maximise le rapport suivant,

$$K = \frac{\sum_{i=1}^n \sum_{j=1}^p b_{ij}^4}{\left(\sum_{i=1}^n \sum_{j=1}^p b_{ij}^2 \right)^2} = \frac{4e \text{ moment}}{2e \text{ moment}}$$

il y aurait une augmentation des saturations près de zéro et une diminution des valeurs dans la zone mitoyenne. Sans trop entrer dans les détails mathématiques, le processus de solution quartimax est une transformation cyclique. Lorsque l'ordinateur a trouvé l'angle de rotation qui lui donne une somme maximale des saturations "rotationnées" à la quatrième puissance, il exécute la transformation à condition que l'angle

de rotation soit plus grand qu'une valeur quelconque spécifiée dans le programme selon la capacité de mémoire et de vitesse de l'ordinateur. Les cycles sont exécutés tant que la somme des quatrièmes puissances pour toute la matrice n'augmente plus. La convergence est habituellement rapide dans les premiers cycles et se stabilise par la suite. Tableau (15)

TABLEAU 15

CONVERGENCE DES CYCLES DE LA ROTATION QUARTMAX

CYCLE #	1	INTER #	0.5599257	INTRA #	0.2583218
CYCLE #	2	INTER #	0.5570548	INTRA #	0.2169809
CYCLE #	3	INTER #	0.6048734	INTRA #	0.2345427
CYCLE #	4	INTER #	0.6193374	INTRA #	0.2370718
CYCLE #	5	INTER #	0.6244816	INTRA #	0.2385741
CYCLE #	6	INTER #	0.6253412	INTRA #	0.2385102
CYCLE #	7	INTER #	0.6253678	INTRA #	0.2383769
CYCLE #	8	INTER #	0.6253229	INTRA #	0.2383078
CYCLE #	9	INTER #	0.6252969	INTRA #	0.2382792
CYCLE #	10	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	11	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	12	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	13	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	14	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	15	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	16	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	17	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	18	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	19	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	20	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	21	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	22	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	23	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	24	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	25	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	26	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	27	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	28	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	29	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	30	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	31	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	32	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	33	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	34	INTER #	0.6252861	INTRA #	0.2382684
CYCLE #	35	INTER #	0.6252861	INTRA #	0.2382684

On remarque que les chiffres des deux dernières colonnes se stabilisent après la quatrième rangée. L'ordinateur a calculé 50 cycles et si après ces cycles il n'y a pas de convergence, il y a erreur au niveau de la transformation. La première colonne représente le nombre de cycles, la deuxième la variance entre les facteurs et la troisième la variance à l'intérieur de chaque facteur. La relation entre les deux est inverse c'est-à-dire que le programme cherche à maximiser la "distance" entre les facteurs et à minimiser celle entre les variables à l'intérieur d'un même facteur.

2.2 e) La rotation Equimax

Ce genre de rotation est rarement employée en géographie. Au lieu de chercher la variation maximum cette technique essaie de maximiser l'égalité de la variance entre les facteurs. En d'autres mots, la rotation est calculée de façon à ce que le pourcentage d'explication soit relativement constant pour chaque facteur. En réalité ceci est pratiquement impossible puisque par la définition même de la matrice des saturations il y a des variables qui jouent un rôle plus important que les autres dans la définition de sa structure. Un tableau définirait une telle rotation ainsi que les pourcentages d'explication respectifs pour nos quatre facteurs en exemple. Une telle rotation est sans doute utile pour qui veut comparer plusieurs types de structure tel que l'utilisation du sol et celle

du socio-économique, par exemple. Toutefois la rotation equimax n'en est encore qu'au stade expérimental et les recherches demeurent à faire à son sujet, afin de préciser son rôle et sa valeur géographiques.

2.2 f) La normalisation des facteurs

Kaiser (1959) s'est aperçu que les résultats de la rotation varimax et quartimax présentaient une grande dispersion à l'intérieur des facteurs. La tendance à avoir des grosses valeurs pour les grandes et les petites saturations était prédominante chez les facteurs à pourcentage d'explication élevé. Ces déformations sont attribuables selon Kaiser (1963) à différents poids, implicitement attachés aux variables par l'intermédiaire des communautés montées au carré dans la formule de la solution varimax et equimax. Chaque variable contribue à la fonction par le carré de sa communauté. Ainsi par exemple, une variable avec une communauté deux fois plus grande qu'une autre influence quatre fois plus la rotation. Pour éviter ceci il faut premièrement "normaliser" les facteurs, deuxièmement exécuter la rotation et finalement les dénormaliser.

La normalisation des facteurs s'effectue par pondération de chaque saturation par les communautés respectives en divisant chaque élément des rangées par la racine carrée de sa communauté respective.

$$a_{ip} = \frac{a_{ip}}{\sqrt{(h_i)^2}}$$

a_{ij} = saturation de

h_i = communauté

Le tableau (16) est le résultat de la matrice des facteurs normalisés. Lorsque la rotation est terminée sur la matrice normalisée, la dénormalisation se fait en multipliant cette fois-ci au lieu de diviser, chaque saturation par la racine carrée des communautés respectives.

$$a_{ip} = a_{ip} \sqrt{(h_i)^2}$$

Le tableau (17) contient la matrice dénormalisée.

TABLAU 16

MATRICE DES SATURATIONS NORMALISEES

	1	2	3	4
1	0.8704	0.4823	0.0476	-0.0864
2	-0.5232	0.1623	-0.2901	0.7847
3	0.3427	0.9319	0.1146	-0.0321
4	-0.6317	-0.0248	-0.7723	0.0629
5	0.9637	-0.2395	0.0936	-0.0716
6	0.5681	-0.1453	0.0964	-0.1798
7	-0.0947	0.2639	0.9599	-0.0053
8	-0.4724	-0.2425	0.1291	-0.8375
9	0.9797	0.0924	-0.0306	0.1750
10	0.8703	0.3031	-0.2988	0.1618
11	0.0816	-0.9830	-0.1434	-0.0552
12	-0.1427	-0.4110	0.3941	0.8096

Au point de vue géographique il n'y a que la matrice dénormalisée qui semble être utile, toutefois il est bon de savoir comment les données sont triturées mathématiquement pour pouvoir en connaître la force et la faiblesse.

TABLEAU 17

MATRICE DES SATURATIONS DE-NORMALISEES

	1	2	3	4
1	0.5971	0.3308	0.0326	-0.0502
2	-0.4183	0.1298	-0.2319	0.6273
3	0.3292	0.8952	0.1104	-0.0300
4	-0.6130	-0.0241	-0.7495	0.0611
5	0.9303	-0.2312	0.0906	-0.0491
6	0.8096	-0.1215	0.0806	-0.1503
7	-0.0888	0.2473	0.8994	-0.0049
8	-0.4318	-0.2217	0.1180	-0.7655
9	0.7194	0.0678	-0.0225	0.1285
10	0.6488	0.2209	-0.2177	0.1179
11	0.0677	-0.8161	-0.1197	-0.0066
12	-0.0864	-0.2487	0.2384	0.4899

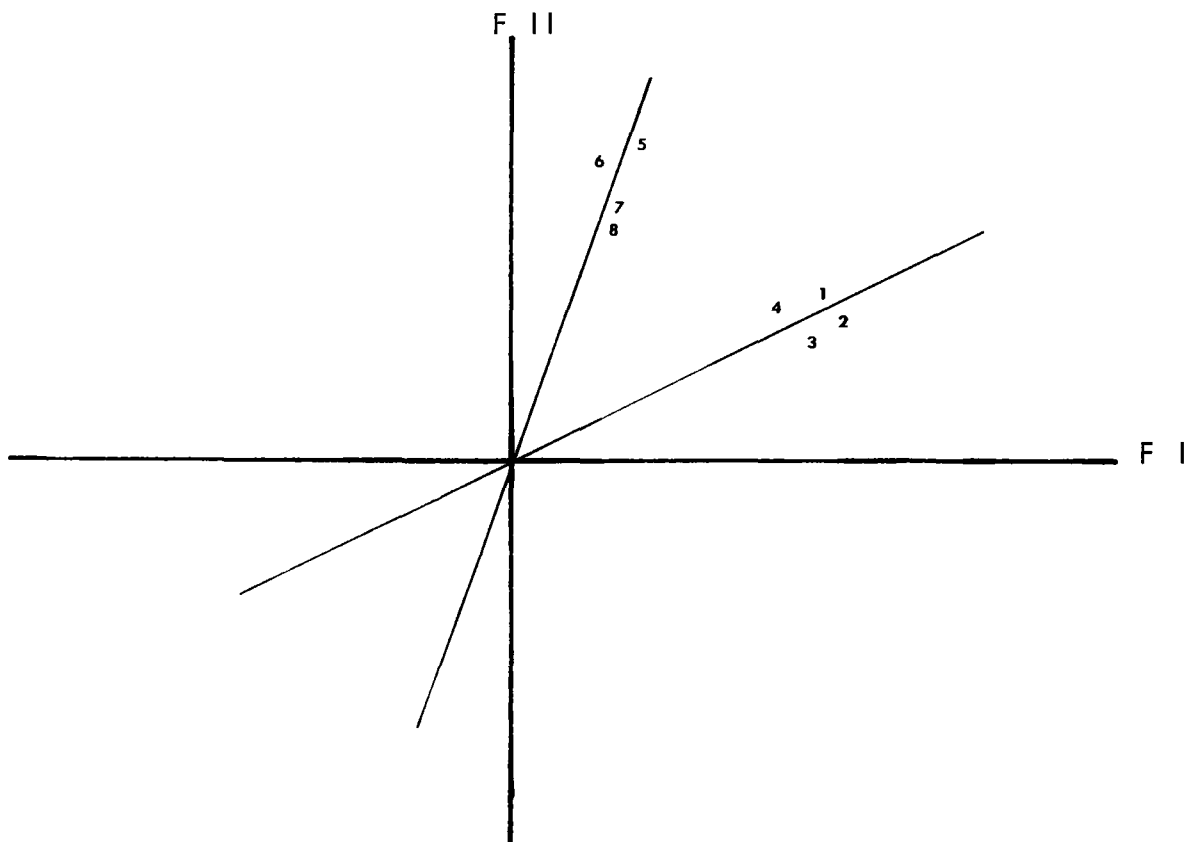
Les rotations que nous avons étudiées concourent toutes à rendre les facteurs orthogonaux l'un à l'autre c'est-à-dire sans aucune corrélation entre eux. Une question demeure: la notion de facteurs indépendants existe-t-elle vraiment dans la réalité? Est-il possible que l'espace soit constitué d'une série de composantes juxtaposées les unes aux autres sans qu'il n'y ait aucune inter-relation? Voilà une question qui vient troubler plusieurs chercheurs et qui remet en cause les études basées sur la rotation orthogonale. Nous croyons cependant que cette dernière est utile pour chercher à définir les caractéristiques spatiales; cependant lorsqu'on veut les replacer dans l'espace et affirmer qu'elles n'ont aucune interaction lors d'un processus de modification spatiale, nous ne sommes plus du même avis. C'est pourquoi nous voulons introduire ici la notion de rotation oblique qui après la rotation orthogonale vient informer le géographe sur les relations possibles entre les constituantes spatiales.

2.2 g) La rotation oblique

Ce type de rotation est effectué après les rotations orthogonales pour servir comme test de la structure factorielle.

Graphiquement la rotation oblique se fait par un déplacement différentiel des axes pour former un angle plus petit que 90°. Voici un exemple: (figure 20.)

FIGURE 20

REPRESENTATION GRAPHIQUE D'UNE ROTATION OBLIQUE

La structure qui était orthogonale au départ devient oblique après la rotation et par le fait même établit une corrélation (tableau 18) entre les facteurs puisque l'angle entre eux est inférieur à 90%.

TABLEAU 18

MATRICE DES CORRELATIONS FACTORIELLES

1	1.00000	-0.13577	-0.05545	-0.04792
2	-0.13577	1.00000	-0.02724	-0.05389
3	-0.05545	-0.02724	1.00000	0.07547
4	-0.04792	-0.05389	0.07547	1.00000

Cette matrice de corrélation factorielle permet donc au géographe de prévoir l'effet sur les autres facteurs en modifiant une composante de son système sans pour cela perdre les éléments structuraux qu'il vient de déterminer soit par l'analyse en composantes principales, soit par l'analyse-image.

Le procédé de calcul en général*est le suivant:

- 1 Un premier facteur est choisi, sur lequel on exécute la minimisation de l'équation de la rotation oblique. Cette opération sur une colonne simple est appelée cycle mineur.
- 2 Lorsque la première étape est terminée sur chaque facteur on obtient un cycle majeur.
- 3 Le cycle majeur est effectué jusqu'à minimisation dans la valeur obtenue.

* Une démonstration mathématique du processus ne serait pas d'une grande utilité pour le géographe. Pour de plus amples informations, consulter Harman (1972, p. 314-314).

Ce critère n'est pas trop utile pour le géographe mais lui montre la "vitesse" à laquelle s'effectue la minimisation de ses facteurs. Une structure sera d'autant plus stable dans la mesure où elle aura utilisé moins de cycles pour se stabiliser.

Maintenant que le géographe a des moyens de saisir et de connaître ces constituantes spatiales il doit les replacer dans l'espace qu'il étudie. Ainsi, après avoir déterminé les principaux facteurs d'utilisation du sol, tels que le niveau d'urbanisation, le type d'utilisation, le type d'habitat et le niveau d'aménagement (Racine, 1972) il lui faut savoir comment varient dans l'espace les poids locaux de ces facteurs, autrement dit quelles sont les unités spatiales qui sont caractérisées par chaque facteur. C'est ici qu'entre en jeu la détermination des "scores", des facteurs, outil indispensable pour le géographe.

2.2 h) Les poids locaux

La façon la plus simple de calculer les poids locaux consiste à multiplier chaque colonne de la matrice des saturations par chaque rangée par la matrice des données originales standardisées. Ceci correspond à l'équation suivante:

$$F_{1i} = a_{11}Z_{1i} + a_{21}Z_{2i} + \dots + a_{p1}Z_{pi} \quad (\text{Coley et Lohnes, 1966})$$

F_{1i} = Le poids local de l'observation i sur le premier facteur

a_{1i} les saturations du premier facteur

Z_{1i} les données standardisées de l'observation i

Voici un exemple numérique à partir de l'utilisation du sol.

Calcul du poids local de Jacques-Cartier sur le premier facteur.

$$\begin{aligned} & (0.70 \times -0.32) + (-0.45 \times 0.29) + (0.51 \times -0.15) + (-0.73 \times -0.26) \\ & + (0.85 \times 0.44) + (0.79 \times 0.19) + (0.14 \times 0.01) + (-0.41 \times -0.71) \times \\ & (0.74 \times 1.4) + (1.03 \times 0.65) + (-0.12 \times -0.09) + (-0.10 \times 0.06) = 1.55 \end{aligned}$$

Cette méthode n'est pourtant valable qu'à condition que l'unité ait été introduite dans la diagonale de la matrice de corrélation. Si tel n'est pas le cas et qu'on place les communautés dans la diagonale, cette solution n'est pas satisfaisante et il faut avoir recours à celle développée par Kaiser.

Selon Kaiser (1958) il est alors impossible de calculer les poids locaux puisque la partie commune des saturations est une variable au hasard, hypothétique, et non observable. Ainsi la meilleure chose à faire est d'estimer les poids locaux à l'aide des moindres carrés. Cependant avec l'analyse image il n'y a rien d'hypothétique, puisque géométriquement l'image est la projection d'un facteur dans l'espace détermi-

né par les autres facteurs. Cette projection n'est pas hypothétique puisque l'espace projeté par le facteur est un sous-ensemble de l'espace total. Il est donc possible de calculer le poids d'un facteur dans l'espace multidimensionnel sur chaque observation à l'aide de la formule suivante:

$$f = (A'A)^{-1} A'WZ$$

A = matrice des saturations

Z = matrice des données standardisées

W = matrice des coefficients de pondération

A^1 = matrice des saturations transposées

L'expression (A'A) correspond aux valeurs propres (eigenvalues) et ainsi l'équation précédente devient

$$f = \frac{A^1 WZ}{A^1 A}$$

f = saturation du premier facteur	x coefficient de pondération	x données standardi- sées
--------------------------------------	---------------------------------	---------------------------------

valeurs propres

f = poids locaux

L'avantage du calcul des poids locaux d'après la méthode de Kaiser c'est qu'elle simplifie la cartographie des poids locaux. Avec cette méthode l'étendue des valeurs des poids locaux s'échelonne habituellement entre 3 et -3 avec

une moyenne zéro contrairement à la méthode classique (la première) qui donne des distributions variant énormément et rend difficile la détermination des classes.

Voici les poids locaux à partir des mêmes données originales calculées selon la méthode classique et celle de Kaiser. (Tableau 19) On remarquera qu'il y a une différence entre les deux matrices attribuable à la non-pondération des saturations dans la première méthode par rapport à la seconde.

Ceci termine la section consacrée à l'analyse factorielle sans épuiser pour autant le sujet de la classification et du regroupement des unités d'espace qui ont en commun les mêmes caractéristiques spatiales. On vient de voir, à l'aide de différents procédés, comment on peut arriver à définir un espace à partir de différents facteurs et à connaître les poids de chacun d'eux sur les unités d'observation. Il reste maintenant à déterminer les types d'espace qui permettront au géographe d'aboutir à une classification à partir de critères non subjectifs. A cette fin les algorithmes du Congroup, du "linkage analysis" et de l'analyse discriminante deviennent des méthodes indispensables.

2.3) LA REDUCTION DE LA MATRICE D'INFORMATION SPATIALE PAR L'ANALYSE TYPOLOGIQUE

Nous allons étudié successivement les techniques des programmes congroup, analyse discriminante et "linkage analysis".

TABLEAU 19

MATRICES DES FONDS LOCAUX PAR LA METHODE CLASSIQUE ET CELLE DE KATSER

FACTEURS SCORES----METHODE CLASSIQUE: NON ROTATIONNES

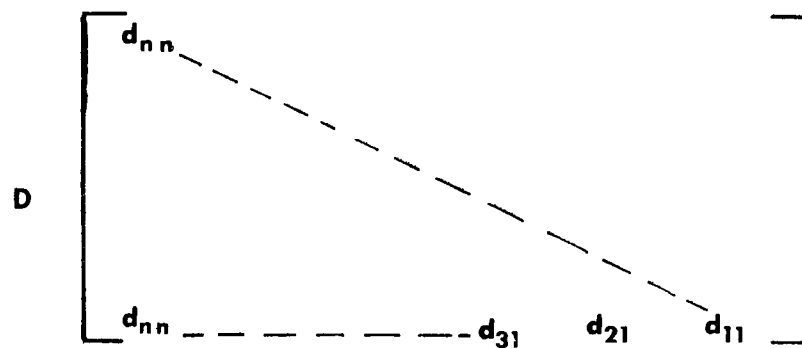
1	-1.830	-1.542	1.491	-0.084
2	-1.907	-1.830	1.382	0.403
3	-4.937	1.530	0.362	-0.572
4	-3.914	0.695	-0.736	-1.307
5	1.371	-1.620	0.390	0.197
6	-1.448	-1.345	1.491	-0.782
7	-2.032	-1.825	1.702	-0.100
8	-4.235	3.931	1.492	1.015
9	3.824	-1.231	2.274	-1.502
10	1.524	-0.055	1.046	1.180
11	1.066	-1.581	0.390	0.499
12	-0.063	6.035	-1.092	0.769
13	12.361	1.700	-0.461	-0.166
14	-0.860	-1.522	-6.186	1.105
15	8.963	1.514	0.126	-0.214
16	1.696	-3.428	-2.516	0.413
17	-4.015	0.864	-1.809	-2.239
18	-0.682	0.442	-0.284	-1.515
19	-3.758	0.653	-0.725	-1.349
20	-0.289	-1.432	0.442	0.905
21	-2.531	0.075	0.880	4.126
22	1.696	-0.028	0.351	-0.784

METHODE DE KATSER

-0.5463	-0.6972	0.7880	-0.0195
-0.3525	-0.9271	0.7908	0.2235
-1.2120	0.7596	0.2107	-0.4096
-0.8212	0.2732	-0.3843	-1.0294
0.3864	-0.8456	0.2218	0.0799
-0.1692	-0.7023	0.8647	-0.7199
-0.6203	-0.9159	0.9666	-0.0317
-1.1400	1.9422	0.8173	0.8577
0.6933	-0.5012	1.2390	-0.9810
0.5701	-0.1190	0.6573	0.7854
0.1565	-0.6507	0.2047	0.5151
0.0289	2.8680	-0.6065	0.5389
3.0321	0.7990	-0.2502	-0.1367
-0.2215	-0.7709	-3.4773	0.8198
2.1447	0.7781	0.0494	-0.1192
0.3033	-1.5510	-1.4258	0.4459
-1.1501	0.5279	-1.0534	-1.5667
0.0275	0.0733	-0.1209	-1.3456
-0.8213	0.2738	-0.3358	-1.1002
0.0846	-0.7434	0.2502	0.9658
-0.6279	0.0414	0.4064	3.1639
0.21	0.0889	0.1566	-0.4660

2.3 a) Le congroup

Le but primordial de l'analyse typologique consiste à grouper logiquement et méthodiquement les observations les plus similaires selon le point de vue étudié. Avec un nombre de variables indépendantes plus petit ou égal à 3 il est possible de faire un groupement en se servant de graphiques; cependant avec plus de 3 variables, par conséquent plus de trois dimensions, il faut avoir recours à l'ordinateur pour manipuler simultanément dans l'espace multidimensionnel l'ensemble des variables indépendantes liées à chaque observation. La mesure de similarité entre les différentes observations est calculée à partir d'une matrice de distances.



Ce moyen de classifier oblige que les axes soient mesurées selon une même échelle si on veut que la distance mesure la similarité (King, 1969, p. 195) et pour cela il est important de choisir des données standardisées (Racine, 1971) tel que facteurs scores de Kaiser. Au tableau (20) on retrou-

ve la matrice des distances obtenues à partir de quatre facteurs scores de l'utilisation du sol pour les vingt-deux municipalités de la rive sud de Montréal.

Les observations sont groupées systématiquement en commençant par les deux plus proches dans l'espace multidimensionnel. Cette paire est remplacée par une centroïde placée au point milieu de leur séparation originale. Ceci a pour effet de modifier la distance entre le nouveau groupe et les autres observations sans changer la distance entre celles-ci.

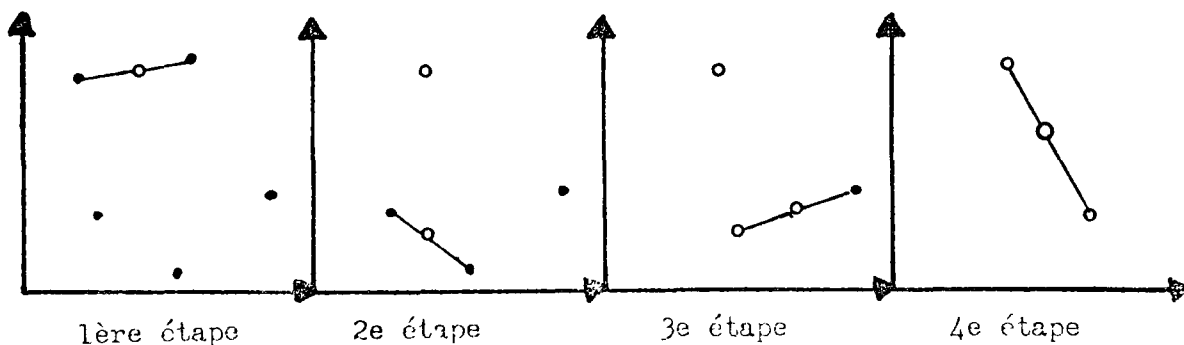
Pour obtenir la distance entre ce groupe et les autres observations ou autres groupes déjà formés, il suffit de prendre la distance moyenne.

Matériellement ceci combine les deux rangées et les deux colonnes des observations respectives en une rangée et une colonne. Ainsi la dimension de la matrice est réduite à chaque itération jusqu'à ce que finalement un seul groupe contienne toutes les observations.

FIGURE 21

REPRESENTATION GRAPHIQUE DE L'ALGORITHME DE CLASSIFICATION

CONGROUPE



Voici graphiquement (figure 21) l'évolution du processus avec un exemple à deux dimensions.

La première étape consiste à remplacer les deux points les plus rapprochés par une seule valeur moyenne.

La deuxième étape fait de même pour les deux autres plus rapprochés.

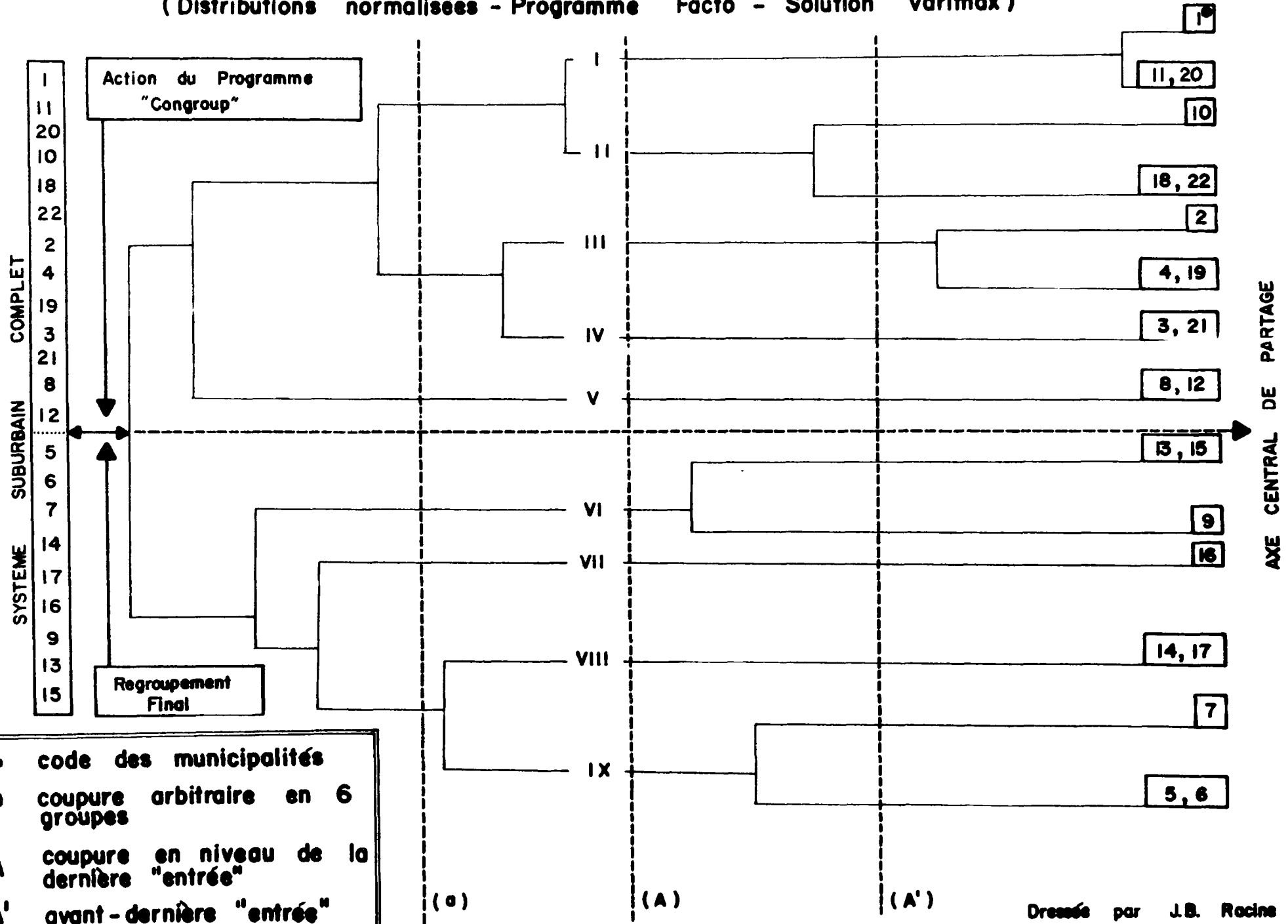
La troisième et la quatrième étapes continuent le processus jusqu'à l'obtention d'un seul point représentatif de l'ensemble.

Il en ressort un dendogramme (figure 22) qui permet de constater de quelle façon les vingt-deux municipalités se sont groupées. Malheureusement il n'existe pas de solution analytique qui pourrait déterminer le nombre de groupes qui optimiserait la classification, bien qu'il soit possible de déterminer à quelle étape une nouvelle réduction du nombre de groupes augmente de façon minimale la somme des carrés des distances à l'intérieur d'un groupe (King, 1969, p. 199). Selon J.B. Racine (1972) une solution provisoire serait d'établir la coupure au niveau de l'entrée dans le système de la dernière observation, c'est-à-dire au moment où l'itération en cours, au lieu de créer un groupe se contente de compléter le système en lui adjoignant la dernière observation, restée jusque là résiduelle.

Le programme congroup possède une option de contiguïté que peu de géographes ont utilisée par le passé. Sans cette

FIGURE 22

**SYSTEME ET SOUS-SYSTEMES DE L'UTILISATION DU SOL URBAIN
DANS LES MUNICIPALITES DE LA RIVE-SUD METROPOLITAINE (MONTREAL)**
(Distributions normalisées - Programme Facto - Solution Varimax)



- code des municipalités
- o coupure arbitraire en 6 groupes
- A coupure en niveau de la dernière "entrée"
- A' avant-dernière "entrée"
- I, II... groupes choisis

option le programme groupe parfois des unités d'espace non contigues ce qui n'est pas toujours intéressant lorsqu'on travaille à une étude de régionalisation ou que l'on soit en face de tout autre besoin qui nécessite que les unités d'observation forment un continuum. La matrice de contiguïté (tableau 21) entre les observations devient le médium de contrainte à la classification des unités spatiales non contigues.

La technique du congroup pose cependant au géographe un certain nombre de problèmes spécifiques, ne serait-ce que parce qu'elle donne le même poids à chacune des variables (les poids locaux des facteurs) qui entrent dans l'analyse, même si l'un d'eux ne compte que pour 5% de la variance totale (Racine, 1972). Ceci biaise certains groupes en incluant des observations dont les caractéristiques ne sont pas en accord avec celles des autres éléments du groupe.

Tel quel cependant, le programme congroup permet de classifier les observations selon six méthodes différentes que nous n'avons pas eu le temps d'étudier ou d'expérimenter, étant donné l'extrême complexité de la structure interne du programme qui exigerait pour fins de compréhension adéquate une thèse de maîtrise pour lui seul.

Mais pour avoir travaillé de nombreuses fois avec le programme nous pouvons affirmer que malgré sa complexité il contient un très grand potentiel que peu de géographes semblent connaître, peut-être à cause d'un manque d'information

ou d'un manque de familiarité avec le programme. Il faut noter que le congrup est particulièrement efficace pour déterminer des hiérarchies plutôt que des classifications proprement dites. Pour la détermination des typologies, le linkage analysis et l'analyse discriminante s'avèrent plus adéquates.

2.3 b) Le "linkage analysis"

Développé par McQuitty (1957) le linkage analysis présente une façon originale de traiter avec la même méthode la réduction des attributs et la réduction des observations. Les avantages du linkage analysis sont la rapidité et l'objectivité. Une matrice d'une vingtaine de variables peut être interprétée en moins de dix minutes.

Cette méthode de groupement est basée sur le plus grand indice d'association qui puisse exister entre une variable et toutes les autres. Ceci élimine donc le problème de savoir à quelle limite nous devons inclure une variable dans l'analyse. Dans le cas du linkage analysis la limite inférieure est déterminée exclusivement par les données.

Au lieu de chercher une structure simple à l'aide de facteurs comme dans le cas de l'analyse factorielle, le linkage analysis détermine une structure typique dans laquelle chaque membre d'un type est plus semblable aux membres de son type qu'aux membres des autres types (McQuitty, 1957, p. 219).

Les caractéristiques que possèdent les membres d'un même type forment un prototype qui se définit opérationnellement comme la centroïde des attributs, commune à chaque membre du TYPE. Il est alors possible de calculer les correspondances typiques (typal relevancies) au moyen de l'analyse factorielle, à cette différence près que celle-ci est effectuée à partir d'une sous-matrice des corrélations, c'est-à-dire quelques variables seulement, au lieu de l'ensemble de la matrice.

La matrice des corrélations est le point de départ pour l'analyse des liens de dépendance. Si une variable i a son coefficient de corrélation le plus élevé avec une autre variable j et de même cette variable j a son coefficient de corrélation le plus élevé avec la variable i , elles forment une paire de variables réciproques. Cependant le phénomène n'est pas toujours le même. Ainsi une variable peut être en forte corrélation avec une autre mais celle-ci ne l'est pas nécessairement avec la première et peut tout aussi bien avoir une plus grande relation avec une troisième. A partir d'une matrice des corrélations on déduit autant de groupes que de paires réciproques. Après quoi il suffit d'y rattacher les variables résiduelles, qui ont une forte corrélation avec l'une ou l'autre des variables formant une paire réciproque.

Voici un exemple à partir de la matrice des corrélations des douze variables utilisées auparavant pour l'analyse factorielle.

1. Souligner dans chacune des colonnes la corrélation positive la plus élevée
2. Etablir le tableau des fortes corrélations

1 → 3: .56
 2 → 4: .36
 3 → 1: .56
 4 → 2: .36
 5 → 9: .76
 6 → 5: .70
 7 → 12: .41
 8 → 4: .19
 9 → 5: .76
 10 → 5: .64
 11 → 6: .46
 12 → 7: .41

3. Choisir la liaison qui a la plus forte relation

5 ↔ 9

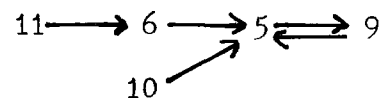
Les variables 5 et 9 sont réciproques, considérées comme soeurs et sont identifiées par un trait double

4. Vérifier si d'autres membres s'y rattachent. En effet 10 se rattache à 5 ainsi qu'à 6. Cependant rien ne s'accroche à 9 excepté 5 que nous avons déjà choisi.

6 → 5 ↔ 9
 10 ↗

Les variables 6 et 10 deviennent cousines du premier degré.

5. Déterminer s'il y a lieu des variables cousines du second degré, c'est-à-dire celles liées aux variables 6 et 10 dans l'exemple en cours. La variable 11 a un lien avec la 6e tandis que la variable 10 n'a pas de lien du second degré et rien ne se rattache à la variable 11. Ainsi se termine le premier groupe.



6. Eliminer les paires de variables qui ont été choisies pour le premier groupe. Les paires restantes sont:

$$\begin{array}{l}
 1 \longrightarrow 3: .56 \\
 2 \longrightarrow 4: .36 \\
 3 \longrightarrow 1: .56 \\
 4 \longrightarrow 2: .36 \\
 7 \longrightarrow 12: .41 \\
 8 \longrightarrow 4: .19 \\
 12 \longrightarrow 7: .41
 \end{array}$$

7. Choisir la plus forte corrélation entre variables qui n'ont pas encore été éliminées.

$$1 \longleftrightarrow 3$$

Aucune autre variable s'y ajoute donc on en conclue qu'elles forment le groupe II.

8. Refaire le procédé jusqu'à ce que toutes les variables aient été choisies. On obtient avec l'exemple en cours deux autres groupes dont l'un composé des variables 7 et 12 et le

quatrième des variables $2 \rightleftarrows 4 \leftarrow 8$.

Le lecteur sera sans doute étonné que l'on rattache dans le quatrième groupe la variable 8 (parcs) à la 4 qui représente le résidentiel unifamilial avec un coefficient de corrélation aussi faible que .19. Ceci est le défaut d'une analyse qui se veut trop objective, basée uniquement sur le chiffre. Car ce .19 représente une indépendance plutôt qu'une relation et devrait constituer une entité. Il serait plus juste de lier la variable 8 à la 2 (importance du terrain vacant) dont la relation est de -.50. Ce principe de choisir uniquement les valeurs positives des coefficients de corrélation constitue la principale faiblesse du linkage analysis, ce qui n'est pas le cas pour l'analyse factorielle qui considère les oppositions. Pour une matrice de corrélation donnée une analyse factorielle qui n'aurait que des facteurs unipolaires positifs donnerait les mêmes résultats que ceux du linkage analysis.

A cause de l'incapacité de saisir simultanément les relations positives et négatives, ce que fait admirablement bien l'analyse factorielle cette méthode n'est presque plus employée pour réduire les attributs. Toutefois elle s'est avérée efficace pour grouper et classifier des unités d'observations selon certains types (Racine, 1972).

Le procédé est essentiellement le même excepté que la matrice de corrélation est calculée à partir de la trans-

posée de la matrice originale. Les variables deviennent des observations et les observations deviennent des variables. De ce fait la matrice des corrélations traduit les relations qui existent entre les observations, au lieu d'entre les variables. Dans notre exemple les municipalités qui se ressemblent auront un coefficient de corrélation élevé et inversement.

Comme on l'a expliqué dans la partie théorique il est possible d'établir une sous-matrice des corrélations avec chaque groupe. Une analyse factorielle de cette matrice donne des facteurs dont les saturations, appelées correspondances typiques, indiquent le poids que tient une variable ou une observation dans la formation du groupe.

Ceci peut se faire à la main. Prenons en exemple le premier groupe obtenu dans le premier exemple.

1. Etablir la sous-matrice des corrélations qui est la suivante:

	5	6	9	10	11
5	(.76)	.70	.76	.64	.30
6	.70	(.70)	.57	.53	.46
9	.76	.57	(.76)	.55	.09
10	.64	.53	.55	(.64)	.38
11	.30	.46	.04	.38	(.46)
Σ	3.16	2.96	2.03	2.74	1.69

2. Estimer la communauté dans la diagonale en choisissant

la plus haute corrélation de chaque colonne.

3. Calculer la somme de chaque colonne.

$$3.16 \quad 2.96 \quad 2.03 \quad 2.74 \quad 1.69$$

4. Additionner les sommes de chaque colonne.

$$3.16 + 2.96 + 2.03 + 2.74 + 1.69 = 12.58$$

5. Extraire l'inverse de la racine carrée de cette valeur de façon à obtenir un coefficient e

$$e = \frac{1}{\sqrt{\Sigma M}} = \frac{1}{\sqrt{12.58}} = .278$$

6. Multiplier chacune des sommes des colonnes par le coefficient e

$$5: 3.16 \times .278 \quad 0.87$$

$$6: 2.96 \times .278 \quad 0.82$$

$$9: 2.03 \times .278 \quad 0.56$$

$$10: 2.74 \times .278 \quad 0.76$$

$$11: 1.69 \times .278 \quad 0.46$$

7. Les résultats de 6. sont les correspondances typiques.

8. Faire le tableau des correspondances typiques en présupposant que les variables qui ne sont pas incluses dans le groupe ont une valeur de zéro. (Tableau 22).

9. Monter au carré les correspondances typiques, faire la somme des carrés et diviser par le nombre de variables pour obtenir la proportion déterminée par le facteur. (Tableau 22, 2e colonne).

TABLEAU 22
CORRESPONDANCES TYPIQUES

1	0.0					
2	0.0					
3	0.0					
4	0.0			} $\frac{1.98}{12} = 16.5\%$	Correspondances typiques du premier groupe avec 16.5% d'explication.	
5	.9	.81				
6	.6	.36				
7	0.0					
8	0.0					
9	.6	.36				
10	.6	.36				
11	.3	.09				
12	0.0		$\Sigma = 1.98$			

Le congrup et le linkage analysis témoignent cependant d'une certaine difficulté à saisir adéquatement toutes les qualités d'un groupement maximum. L'analyse discriminante à laquelle nous avons consacré un article en collaboration avec le professeur Racine dans la revue française L'Espace Géographique (Racine-Lemay, 1972), détermine un optimum de classification qui mérite qu'on s'y attarde quelque peu.

2.3 c) La fonction discriminatoire

A partir de l'analyse des différents facteurs de différenciation spatiale, le géographe est conduit à définir, lorsqu'il passe des variables aux observations, des types de combinaisons spatiales. Le groupement des aires élémentaires $y_1, y_2, \dots, y_i, \dots, y_n$ qui découpent à l'origine l'ensemble spatial E soumis à l'étude, doit aboutir à découvrir des sous-ensembles homogènes Y_1, Y_2, \dots, Y_n , n étant inférieur à n . Le problème de la classification en géographie est en effet toujours le même: établir dans l'ensemble E des sous-ensembles E'_1, E'_2, \dots, E'_i , tels que les attributs (les poids locaux factoriels p_1, p_2, \dots, p_i si la classification prend comme point de départ la matrice semi réduite par l'analyse, la matrice F), varient moins à l'intérieur de E'_i qu'ils ne varient lorsque l'on passe de E'_i à un quelconque des E' . Il s'agit donc, en d'autres termes, de déterminer des groupes, (classes, types, régions) d'observation de façon telle que la variance intra-groupe soit minimale et la variance inter-groupe soit maximale. Tel est justement le rôle de l'analyse discriminatoire. Le lecteur trouvera dans les travaux de Casetti et de King une présentation plus complète que la nôtre de la fonction discriminatoire linéaire, simple ou multiple. Il pourra également se reporter à l'indispensable ouvrage de Cooley et Lohnes (1966). Le programme que nous avons utilisé a été adapté en langage FORTRAN par Casetti et nous a été fourni comme tous les autres par notre collègue torontois Bryn Greer-Wooten qui l'a mis au point pour le système actuel l'I.B.M.

360-65. C'est en suivant les articulations de ce programme que nous présenterons au lecteur la logique interne de l'analyse discriminatoire.

i) Définitions

Au départ, le principe général de l'analyse discriminatoire est relativement simple: il s'agit, à partir de la découverte d'une série de fonctions linéaires mutuellement orthogonales séparant de la façon la plus adéquate possible un certain nombre de classes ou de groupes, d'assigner tel ou tel individu (ici des observations) à tel ou tel groupe avec une probabilité d'erreur qui soit minimale; autrement dit, estimer la position de chacun des individus (ou des observations) sur les droites qui séparent les groupes ou les classes de façon optimale. En termes de classification le modèle général peut donc s'écrire;

$$P(H_j/X_i), \quad i=1,2,\dots,N \quad \text{et} \quad j=1,2,\dots,g$$

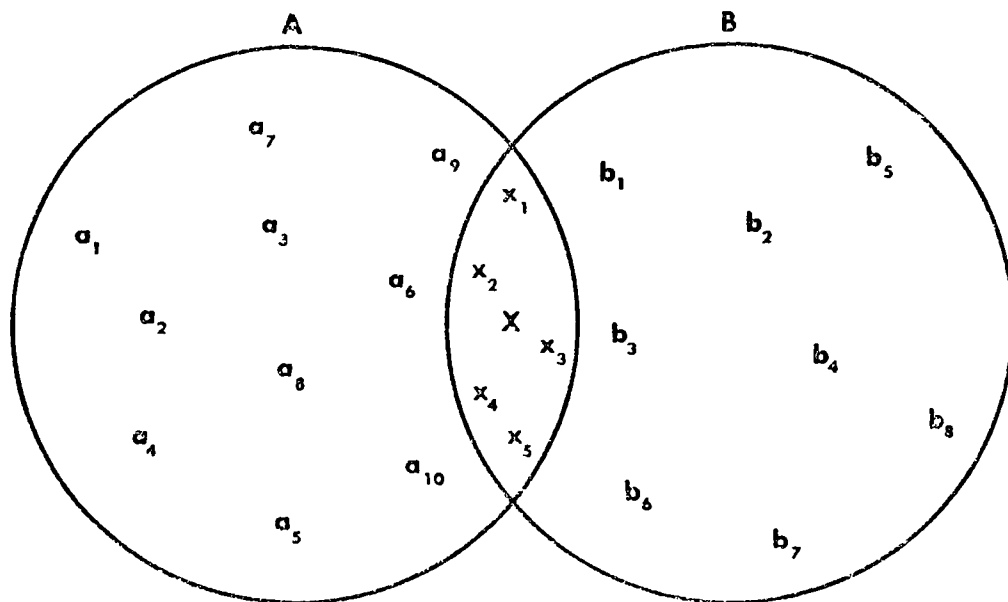
Le modèle se lit ainsi: "probabilité de l'hypothèse j, étant donné les attributs spatiaux (ici les poids locaux des facteurs) de l'individu i". L'hypothèse j est la suivante: l'individu i est un membre du groupe j. Il y a autant d'hypothèses possibles que de groupes g pour chacun des individus. La méthode consiste donc à choisir l'hypothèse la plus probable en comparant le profil des attributs spatiaux d'une observation des différents groupes.

Il nous faut souligner cependant dès le départ deux contraintes fondamentales de l'analyse: les attributs spatiaux peuvent

être communs à plusieurs observations; en revanche les observations ne peuvent appartenir à plus d'un seul groupe. Mais une des richesses de l'analyse discriminatoire consiste justement à permettre au géographe, dans une étude régionale ou dans une étude sur les zones d'influence, de trancher en face de cas marginaux, sur l'appartenance desquels il ne peut se décider avec objectivité. Supposons en effet que l'on ait devant soi deux groupes d'observations, les groupes A et B. Chacun de ces groupes rassemble donc un certain nombre d'éléments caractérisés par un certain nombre de mesures ou variables, en géographie, les "attributs spatiaux".

FIGURE 23

REPRESENTATION GRAPHIQUE DE LA FONCTION DISCRIMINATOIRE



Si l'on trace deux cercles à partir du centre du groupe A et du centre du groupe B on découvre (cf. figure 23) que certains éléments (ici les éléments x_1, x_2, x_3, x_4, x_5) sont communs aux deux ensembles. Le rôle de la fonction discriminatoire est d'affecter les éléments de ce sous-ensemble jointif X à l'un ou l'autre des deux groupes, le plus fidèlement possible, en maximisant l'hétérogénéité entre les deux groupes et en minimisant la probabilité d'erreur d'attribution. Il est évident qu'un tel procédé suppose à priori une certaine forme de classification initiale, modifiable, et que chacun des éléments pris en considération soit classifiable dans l'un ou l'autre des deux groupes considérés. Le rôle des itérations discriminatoires, telles que définies par CASETTI est cependant fort différent même s'il s'appuie sur celui de la fonction discriminatoire. La première, s'inscrit dans ce que nous appellerons la procédure discriminatoire, qui peut se définir comme un ensemble de règles permettant de situer un nouvel élément dans l'un ou l'autre des différents groupes (ou classes, types ou régions) d'une classification ou d'une régionalisation. C'est à cette fin que la procédure discriminatoire a été utilisée par Leslie KING dans le premier exemple, qu'il lui consacre dans son livre Statistical Analysis in Geography, page 206.

Soulignons au passage l'intérêt d'une telle formulation en géographie régionale. Soient deux régions nodales A et B et une aire géographique intermédiaire, dont les attributs spatiaux ont des caractéristiques telles qu'à priori il semblerait qu'on ait pu

les tirer aussi bien de la région A que de la région B. Etant donné la complexité des relations pouvant exister entre les attributs spatiaux, facteurs de la différenciation régionale, comment différencier ces deux régions nodales de façon telle qu'on puisse assigner l'aire intermédiaire X à l'une plutôt qu'à l'autre? Pour ce faire il faut évidemment chercher, en utilisant les valeurs des attributs spatiaux qualifiant les observations, une fonction telle que la dispersion à l'intérieur des groupes soit minimale tandis que la dispersion entre les groupes soit maximale. On retombe ainsi sur le problème classique de l'analyse de variance.

L'idée fondamentale de CASETTI fut alors d'appliquer la procédure à des groupes déjà constitués dont les différents membres seraient caractérisés quantitativement par les poids locaux des facteurs obtenus par la décomposition de la matrice de corrélation en ses principales composantes. Ses itérations discriminatoires (I.D.) ne font qu'appliquer, en la réitérant de façon successive, la procédure discriminatoire à un ensemble d'éléments préalablement groupés, soit à partir d'un choix subjectif, soit à partir de l'un ou l'autre des différents algorithmes de classement dont on dispose aujourd'hui. Ces I.D. ont pour effet de transformer une classification initiale T_0 en une classification finale T_i optimale au moyen de i procédures discriminatoires.

ii) De la matrice originale à la matrice des groupes

La matrice d'information originale qui va être soumise au traitement informatique comporte deux entrées. Les colonnes accueillent les attributs spatiaux ou variables aléatoires qui vont permettre aux éléments (ou observations) d'être classés de façon optimale après que la matrice E' ait subi successivement les effets de l'analyse factorielle et des premiers algorithmes de classification (Passage de la matrice E' aux matrices P et T dans le tableau No.²³). Les tableaux No.^{24,25,26}, voir aux pages 172-4 progressivement dégagés de l'analyse des différentes caractéristiques de l'utilisation du sol urbain dans les municipalités de la zone métropolitaine de Montréal au Sud du Saint-Laurent permettent au lecteur à l'aide d'un exemple concret de suivre le processus que nous décrivons.

Sur notre tableau nous avons défini une matrice typologique d'ordre $n' \times p$. Mais selon la terminologie habituelle des traités décrivant le processus de l'analyse discriminatoire nous utiliserons plutôt le code n_g (groupes) que le code n' , pour définir la matrice au sein de laquelle les différents éléments (observations) du système ont déjà été rassemblés en un certain nombre n_g de groupes. Chaque élément de chacun des groupes est défini (ou caractérisé) par une série d'attributs spatiaux qui ne sont autres que les poids locaux (scores) que l'élément en question a obtenu sur les facteurs de différenciation du système auquel il appartenait. Avant que le test discriminatoire de l'optimalité de groupement considéré

n'ait été effectué, la matrice T d'ordre $n_g \times p$ se présente donc une classification initiale T_0 dans laquelle par exemple l'élément $a_{hi,j}$ est la $i^{\text{ème}}$ observation possédant la $j^{\text{ème}}$ caractéristique et appartenant au $h^{\text{ème}}$ groupe.

CLASSIFICATION INITIALE T_0

		1	2	3	4.....j.....p	<u>attributs spatiaux</u>
<u>Groupes</u>						
	1					
I	2					
	3					
	4					
II	5					
	6					
III	7					
	8					
	.					
.	.					
.	.					
h	i				$a_{hi,j}$	
.	.					
.	.					
.	.					
.	.					
n_g	n <u>observations</u>					$(n_g \ n)$

La règle générale à laquelle obéissent toutes les procédures discriminatoires est simple: il s'agit d'ajouter un nouvel

élément au groupe dont le point représentatif (moyenne ou centroïde) est le plus proche (l'utilisation des poids locaux factoriels indépendants les uns des autres permet tant d'effectuer directement des mesures de distance en suivant la théorème de Pythagore) du point qualifiant l'élément considéré dans l'espace multidimensionnel. Les procédures discriminatoires peuvent cependant se prêter à quelques variantes selon les buts du chercheur. Quelques unes de ces modifications possibles méritent d'être soulignées au passage en vue de recherches ultérieures. Elles portent sur trois points principaux:

- a) La transformation qui génère l'espace discriminatoire à partir de l'espace original (la nôtre, partant de l'espace factoriel sera linéaire et orthogonale).
- b) L'ensemble des règles arithmétiques ou algébriques qui permettent de déterminer les points centraux représentatifs des groupes (nous utiliserons quant à nous la moyenne à titre de "centroïde" du groupe dans l'espace discriminatoire, suivant en cela le programme de CASETTI).
- c) La fonction qui calcule la distance entre les points les plus rapprochés dans l'espace discriminatoire (nous en resterons évidemment à l'espace euclidien multidimensionnel).

iii) Les itérations de l'optimisation typologique

Comme nous l'avons déjà indiqué le processus de l'itération consiste dans l'application successive de la procédure discri-

minatoire à la classification initiale telle que mise en ordre dans la matrice T_0 . Cette matrice possède n éléments composant n_g groupes différents. Ainsi x_{hi} est le $i^{\text{ème}}$ élément du $h^{\text{ème}}$ groupe. Supposons maintenant, au plan théorique, que l'on incorpore un nouvel élément au système qui n'appartienne à aucun des groupes définis. La matrice possède alors $n + 1$ élément et un nouveau groupe X_0 qui n'est constitué que par un seul élément. Ce nouvel élément, que l'on peut identifier par les coordonnées $X_{0,n+1}$ doit obligatoirement entrer, selon les postulats de départ de l'analyse discriminatoire, dans l'un de n_g groupes déjà définis. Appliquée à cet élément et aux valeurs de ses attributs spatiaux la procédure discriminatoire P.D. va se traduire par l'assignation de $x_{0,n+1}$ à l'un ou l'autre des n_g groupes prédéfinis. On obtiendra donc la relation suivante:

$$\text{P.D. } (x_{0,n+1}) = x_{h,n+1} \quad (\text{I})$$

En réalité ce même procédé est appliqué à chacun des éléments du système ordonné dans la matrice T_0 . Dès que l'un de ses éléments constituants se révèle être plus proche de la centroïde d'un autre groupe, le groupe h par exemple, que de la centroïde de son groupe initial il est transféré dans le groupe h . Dans le cas contraire, il demeurera dans son groupe original. On comprend que l'application de la procédure discriminatoire à tous les éléments de la classification initiale T_0 peut à l'extrême se traduire par une réorganisation complète de la distribution des éléments dans chacun des groupes considérés. A l'inverse cependant les deux clas-

sifications seront rigoureusement identiques si après la première itération il n'y a eu aucune permutation, c'est-à-dire si les éléments des groupes de la matrice T_1 sont organisés selon la même structure que celle qui est propre à la matrice T_0 . En d'autres termes:

$$T_0 = T_1 \text{ si } t_0(h,i) = T_1(h,i) \tag{II}$$

dans laquelle $h = 1 \text{ à } n_g$ et $i = 1 \text{ à } n$

Si le programme a dû effectuer une ou plusieurs permutations, il s'en suit évidemment que ce déplacement des unités d'observation d'un groupe à l'autre change les valeurs des centroides des groupes qui, soit ont été amputés d'un ou plusieurs éléments, soit au contraire ont fait de nouvelles acquisitions. L'idée de CASSETTI fut donc d'appliquer la procédure discriminatoire autant de fois qu'il serait nécessaire pour obtenir finalement une classification qui ne changerait pas à la suite d'une nouvelle itération. Après avoir réalisé les permutations le programme recommence donc tous les calculs à partir de ce nouveau découpage et réalise autant d'itérations discriminatoires qu'il est nécessaire pour obtenir enfin une structure stable, c'est-à-dire que deux itérations donnent le même résultat. Cette structure stable exprime la division optimale.

En d'autres termes nous pouvons écrire en utilisant une série d'exposants successifs et croissants:

$$P.D. (T_0) = T_1 \tag{III}$$

$$\text{P.D.}(T_1) = \text{P.D.} \text{ P.D.}(T_0) = \text{P.D.}^2(T_0) \quad (\text{IV})$$

$$\text{P.D.}(T_i) = \text{P.D.} \text{ P.D.}(T_{i-1}) = \text{P.D.}^i(T_0) = T_i \quad (\text{V})$$

dans lesquelles:

P.D. = procédure discriminatoire appliquée à

T_0 = classification originale

T_i = classification finale

i = nombre d'itérations discriminatoires

(T pouvant varier de T_0 à T_i)

La limite de l'analyse est en effet atteinte quand la dernière classification est identique à l'avant dernière. C'est-à-dire lorsque l'on obtient:

$$\text{P.D.}^{i-1}(T_0) = \text{P.D.}^i(T_0) \quad \text{ou encore} \quad T_{i-1} = T_i \quad (\text{VI})$$

CASETTI a démontré qu'une classification initiale ne peut avoir plus d'une limite et que la variabilité existant à l'intérieur des groupes d'une classification générée par un algorithme discriminatoire était plus faible que la variabilité interne des groupes définis au départ dans la classification initiale. L'inverse est vrai pour la variabilité existant entre les groupes. La fonction discriminatoire possède évidemment comme toute équation algébrique e discriminants, racines latentes ou "eigenvalues"* (tous ces termes

* Nous conserverons dans cet ouvrage le terme utilisé jusqu'à ce jour par les seuls utilisateurs du programme.

sont équivalents) qui servent à sa résolution et qui mesurent en valeur absolue la puissance discriminatoire de la fonction. Le rapport de la valeur de l'eigenvalue obtenue par chacun des groupes à la somme des valeurs des différentes eigenvalues donne la proportion de la variance inter-groupes exprimée par la fonction.

On démontre également que l'on peut juger de la différence entre la classification initiale et la classification limite, ou encore de la différence entre n'importe quelle classification à n'importe quel niveau d'itération en utilisant les deux rapports suivants (cf. KING, 1969):

$$R^0 = \frac{\text{variabilité intra-groupe}}{\text{variabilité totale}} = 1 / (1 + \sum_{i=1}^e \lambda_i); \quad 0 < R^0 < 1;$$

$$R = \frac{\text{variabilité inter-groupes}}{\text{variabilité intra-groupe}} = \sum_{i=1}^e \lambda_i; \quad 0 < R < \infty$$

Si R^0 est petit, la variabilité (ou la dispersion) à l'intérieur des groupes est relativement faible et la différence entre les groupes doit donc être élevée. La classification est donc d'autant meilleure que R^0 est faible. A l'inverse de faibles valeurs de R correspondent à une classification médiocre tandis que de fortes valeurs de R indiquent qu'une classification est bonne. Il existe d'ailleurs bien d'autres points de repère pour tester la validité d'une classification par rapport à une autre ou d'un procédé de groupement par rapport à un autre; le nombre des itérations nécessaires pour aboutir à une structure stable par exemple, le nombre

de permutations que le programme a dû effectuer, des tests de variances dans les écarts. Leslie KING en suivant CASETTI ne semble malheureusement pas s'être préoccupé de savoir si tous ces tests concordent et surtout si les itérations discriminatoires, appliquées à des matrices typologiques issues de différents algorithmes de classement - qui comme l'a rappelé Brian J.L. BERRY ont tous des qualités et leurs défauts - donnaient des résultats analogues. Ce n'est qu'après avoir livré aux presses la deuxième partie de son article sur les "Modèles graphiques et mathématiques en géographie humaine" que J.B. RACINE a découvert qu'il pouvait y avoir contradiction entre les différentes mesures de l'optimalité d'une classification et qu'en outre, selon que les distributions statistiques originales avaient été ou non normalisées, les résultats finaux pouvaient varier sensiblement. De la même manière nul ne s'est encore prononcé sur l'optimalité du nombre de groupes à prendre en considération au départ. Autant de questions qui ont fait l'objet sous la direction de J.B. RACINE de recherches récentes à l'Université d'Ottawa, recherches et résultats dont il rendra compte, avec l'un ou l'autre de ses assistants de recherche dans de prochaines publications. Le problème n'est pas très grave en soi tant que le chercheur reste au niveau théorique et de la recherche fondamentale. Lorsqu'en revanche la géographie doit déboucher sur des résultats qui seront utilisés par ceux qui ont la charge de prendre des décisions en matière d'aménagement du territoire, de telles ambiguïtés dans le processus de calcul pourraient se révéler catastro-

phiques.

Quoiqu'il en soit, dans le cadre de la problématique définie dans cet ouvrage, le problème ne se pose pas puisqu'il s'agit d'utiliser les itérations discriminatoires non pas pour tester l'optimalité d'une classification mais pour établir des corrélations ou comme on le verra plus loin, des "correspondances" entre matrices typologiques issues de l'analyse des différentes trames visibles ou invisibles dont la superposition définit l'espace géographique. Nous croyons cependant qu'auparavant le lecteur appréciera de pouvoir suivre avec un exemple le cheminement du calcul qui permet aux différentes itérations discriminatoires d'aboutir aux résultats définis plus haut.

iv) La structure de l'algorithme itératif dans le programme DISCRIM*

Il a été mathématiquement démontré, bien avant les travaux des géographes CASETTI et KING d'ailleurs, que la fonction dis-

* L'exemple choisi ci-après est toujours celui de l'analyse des structures de l'utilisation du sol urbain dans les municipalités de la rive-sud métropolitaine de Montréal. Nous avons cependant enlevé, à cette étape de l'analyse notre ancienne observation numéro 4 (la réserve indienne de Caughnawaga) qui pour être englobée dans la zone métropolitaine n'offre malheureusement pas des données statistiques comparables à celles obtenues pour les autres municipalités ne "bénéficiant pas" d'un statut particulier. Le problème se pose, surtout lorsqu'il s'agit de passer de la trame de l'utilisation du sol, évidemment exacte pour cette municipalité comme pour les autres, aux différentes trames de l'espace socio-économique et démographique que nous allons comparer entre elles plus loin. Quant à la classification initiale soumise au test discriminatoire, elle est issue de l'algorithme de McQUITTY ("linkage analysis") appliqué à une matrice des corrélations de rang $n \times n$ (observations \times observations).

criminatoire pouvait s'écrire de la façon suivante en langage matriciel**;

$$\left| G - \lambda I \right| = 0 \quad (I = \text{matrice unitaire}) \quad (\text{VII})$$

dans laquelle:

$$G - (V' V)^{-1} D'D \quad \text{et} \quad \lambda = \frac{v' D'Dv}{v' V'V_v} \quad (\text{VIII})$$

Ce sont donc les valeurs de ces différents termes que le programme mis au point par CASSETTI va chercher à découvrir.

- La première étape consiste à calculer les moyennes de chaque groupe pour chaque variable. La première matrice prise en considération est donc une matrice de rang $n_g \times p$.

** Nous avons réussi à reprendre toute la démonstration mathématique sur laquelle reposent les notions de fonction et de procédure discriminatoire. Cette démonstration est cependant hors de la portée de la très grande majorité des géographes qui doivent bien - c'est la condition du progrès - s'en remettre quelques fois aux spécialistes et leur faire confiance. De toutes manières cette démonstration ne pouvait avoir sa place dans ce travail.

- L'étape suivante consiste dans le calcul de la moyenne obtenue sur l'ensemble des attributs spatiaux (poids locaux des facteurs) par les membres de chacun des groupes.

MOYENNES

Groupe 1	-0.0563	0.7965	-0.0808	0.0223
Groupe 2	-0.6365	0.2775	-0.0475	1.8110
Groupe 3	0.2764	-0.1926	-0.1160	-0.2476
Groupe 4	0.3855	-0.3885	-0.6277	-0.3805
Groupe 5	-1.3120	-1.6205	1.5950	-0.2250
Ensemble	-0.0361	-0.0385	-0.0340	0.0024

- L'ordinateur calcule alors, sans toutefois l'écrire sur la sortie du programme, la matrice D qui définit la différence entre la moyenne de chaque groupe et la moyenne totale, de telle sorte que puisse être utilisée la matrice de rang $n_g \times p$:

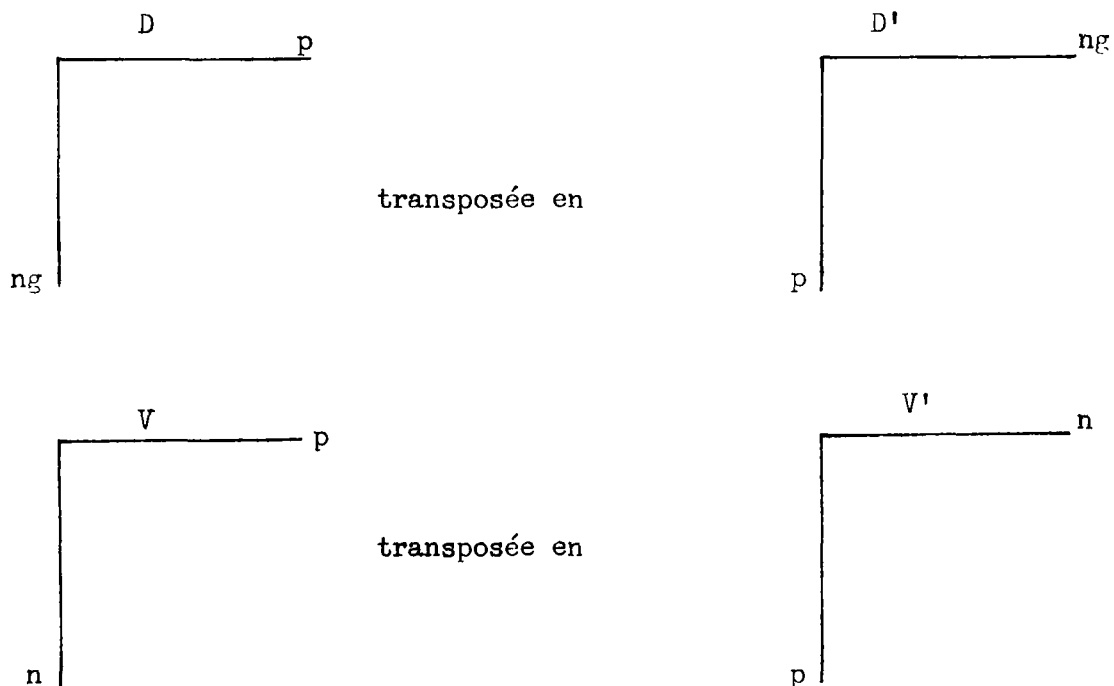
$$D = \bar{X} \text{ (groupes)} - \bar{X} \text{ (ensemble)}$$

- De la même manière l'ordinateur découvre aussi les valeurs de la matrice V (de rang $n \times p$) qui expriment la différence entre les valeurs de la matrice originale T_0 et la moyenne de chacun des groupes:

$$V = t - \bar{X} \text{ (groupe)}$$

- Ces deux dernières matrices sont alors transposées de façon à donner les matrices D' et V'. Le procédé est simple: il s'a-

git de mettre les colonnes à la place des rangées et inversement de la manière suivante:



- Les deux matrices D' et V' sont alors respectivement post-multipliées par les matrices D et V de telle sorte que l'on obtienne:

$$D' \times D = D'D \quad \text{et} \quad V' \times V = V'V$$

MATRICE D'D

2.2642	1.6161	-2.3453	-1.0357
1.6161	3.4460	-2.3999	1.1205
-2.3 53	-2.3999	3.0152	-0.1480
-1.0357	1.1205	-0.1480	3.5322

MATRICE V'V

17.1664	-1.9157	5.4718	3.4888
-1.9157	10.2631	3.3156	-1.9166
5.4 18	3.3156	12.2816	0.8114
3.4888	-1.9166	0.8114	12.9109

- Si l'on se reporte à la relation fondamentale (VIII) on comprendra que cette nouvelle étape ait pour but de calculer l'inverse de la matrice V'V et de définir ainsi les valeurs de la matrice $V'V^{-1}$, de laquelle il faudra dégager les racines caractéristiques, les eigenvalues. Ces dernières sont proportionnelles au poids de chaque variable dans la fonction discriminatoire. Soit e le nombre total d'eigenvalues. On peut écrire que la proportion de la variance originale de l'ensemble considéré est déterminée par le rapport:

$$\frac{\lambda_j}{\sum_{i=1}^e \lambda_i} \times 100 \quad (j = \text{une quelconque des } \underline{\text{eigenvalues}}) \quad (\text{IX})$$

La somme des eigenvalues ($R = \sum_{i=1}^e \lambda_i$) est égale au rapport initial entre la variance intra-groupe et la variance inter-groupe. c'est pourquoi si R est petit la classification doit être considérée comme pauvre. On remarquera cependant que R peut avoir une valeur faible pour deux raisons, comme toujours lorsque l'on a af-

faire à un rapport: le dénominateur peut être trop grand comme le numérateur peut être trop petit. C'est pourquoi on utilise également l'indice R^0 qui exprime le rapport de la variabilité interne à la variabilité totale de l'ensemble. Un R^0 de faible valeur montre que la différence entre les groupes est bien établie et qu'ils sont nettement séparés.

EIGENVALUE	1	1.2886362
EIGENVALUE	2	0.4109840
EIGENVALUE	3	0.0497269
EIGENVALUE	4	0.0016194
Somme des EIGENVALUES	$(\sum_{i=1}^e \lambda_i)$	= 1.7509661

- La matrice suivante est appelée matrice résultante.

Elle contient, comme dans une matrice factorielle, les saturations des variables initiales (dans notre cas les attributs spatiaux définis en termes de poids locaux des facteurs) dans chacune des fonctions discriminatoires.

MATRICE RESULTANTE

0.4205	0.3965	-0.2005	-0.7662
0.6359	-0.3117	0.7358	0.0384
-0.6448	-0.1167	0.5386	-0.5845
0.0555	-0.8556	-0.3583	-0.2644

Les quatre rangées de cette matrice contiennent les coefficients de quatre équations d'un type bien connu, du moins des praticiens de l'analyse de régression multiple et qui n'est en fait qu'une extension de la fonction bi-variée standardisée de type $y=ax$. On peut écrire en effet quatre fonctions discriminatoires Z_1, Z_2, Z_3, Z_4 :

$$Z_1 = 0.4204 p_1 + 0.3965 p_2 - 0.2005 p_3 - 0.7662 p_4$$

$$Z_2 = 0.6359 p_1 - 0.3117 p_2 + 0.7358 p_3 + 0.0384 p_4$$

$$Z_3 = 0.6448 p_1 - 0.1167 p_2 + 0.5386 p_3 - 0.5845 p_4$$

$$Z_4 = 0.0555 p_1 - 0.8556 p_2 - 0.3583 p_3 - 0.644 p_4$$

- A partir de la matrice dite "résultante", il est donc possible de calculer le poids réel de chaque élément (observation) par rapport à chacune des fonctions discriminatoires, de façon analogue au procédé que l'on utilise pour calculer avec les saturations et les valeurs standardisées originales les poids locaux de facteurs. Il suffit de multiplier la matrice originale (entrée) par la matrice résultante. On obtient ainsi la matrice des scores discriminatoires, la matrice XR de rang $n \times p$:

MATRICE XRGROUPE 1

1	BOUCHERVILLE	0.4958	-0.4682	1.1221	-0.2128
2	LAFLECHE	-0.1935	-0.4350	0.4127	-0.3847
3	CHAT-HEIGHT	0.5619	-0.6207	1.2660	-0.1924
4	CHAT-CENTRE	0.5922	0.0562	1.2389	-0.2625
5	DELSON	1.5417	-0.5185	-0.9658	1.9594
6	ST-LAMBERT	0.2190	0.3136	0.2009	-0.2164

GROUPE 2

7	BROSSARD	0.1118	-0.5457	-0.7098	0.1861
8	ST-HUBERT	-0.0320	-3.2199	-1.3951	-0.0915

GROUPE 3

9	CANDIAC	0.7298	0.2835	0.7314	1.2542
10	CHATEAUGUAY	-0.2536	-0.3146	0.3567	-0.6499
11	GREENFIELD PARK	0.8290	0.6162	0.6094	-0.614
12	ST-BRUNO	-0.0021	0.9521	0.1661	0.6409
13	STE-CATHERINE	-0.7550	0.0834	-1.6799	1.1058
14	LEMOYNE	0.0632	1.6649	-1.5794	-2.1509
15	ST-CONSTANT	-0.2275	-0.5206	0.1995	-0.1913

GROUPE 4

16	PREVILLE	-0.5555	1.1690	0.3694	2.1073
17	LA PRAIRIE	0.9668	1.1162	-1.4671	0.9798
18	JACOUES-CARTIER	0.4276	-0.6622	-0.0068	-0.9681
19	LONGUEUIL	0.3558	1.0679	-1.1551	-1.4901

GROUPE 5

20	LERY	-3.2732	0.3160	-0.216	0.5766
21	NOTRE-DAME	-1.9729	-0.3335	0.2822	-0.4362

La même fonction discriminatoire permet également de trouver pour chacun des groupes concernés le point moyen représentatif du groupe tout entier (la centroïde) dans l'espace discrimina-

toire. Il convient à cette fin de multiplier la matrice des moyennes par la matrice des saturations ou matrice résultante. Les moyennes du départ sont donc pondérées par les poids respectifs des éléments (poids exprimés par les saturations) dans les fonctions discriminatoires. On obtient ainsi la matrice XAVR de rang $n_g \times p$:

MATRICE XAVR

GROUPE	1	0.5362	-0.2803	0.5458	0.1151
GROUPE	2	0.0399	-1.8828	-0.3427	0.0473
GROUPE	3	0.0548	0.3950	-0.1709	-0.0859
GROUPE	4	0.2987	0.6727	-0.5649	0.1572
GROUPE	5	-2.6231	-0.0088	0.0103	0.0702

- Il devient enfin possible de trouver, dans l'espace discriminatoire, les distances séparant les groupes les uns des autres. Il suffit d'additionner la différence des carrés de la distance existant entre les groupes pris deux à deux. Notre programme ne le fait pas, mais c'est cette technique que Leslie KING a utilisée avec des préoccupations complètement différentes des nôtres pour comparer l'évolution respective des systèmes urbains du Québec et de l'Ontario (1967). Dans notre programme la dernière matrice que l'on obtient pour terminer la première itération définit plutôt la distance de chacun des éléments (observations) du système complet à chacune des centroïdes de chacun des sous-systèmes considérés au départ (les groupes). Cette distance se calcule aisément (formule X) en utilisant d'une part la valeur de la centroïde de chacun des grou-

pes et d'autre part la valeur des scores discriminatoires obtenus par chacun des éléments tels qu'ordonnés dans la matrice XR,

$$D = \sum_{i=1}^n \sqrt{(XR_i, j) - XAVR_i, j)^2} \quad (X)$$

(j = 1 à p)

Les valeurs D obtenues permettent à l'ordinateur de décider immédiatement si un élément placé au départ dans tel ou tel groupe a mieux sa place dans tel ou tel autre pris comme un tout. Le programme effectue donc les permutations qui s'imposent et recommence tous les calculs (itération 2) à partir de ce nouveau découpage et réalise autant "itérations discriminatoires" qu'il est nécessaire pour arriver à une matrice finale T_i qui soit identique à la matrice T_{i-1} . Dans notre exemple on constate que les municipalités de Delson et de St-Lambert originellement "entrées" dans le groupe I ont mieux leur place la première dans le groupe 4, la seconde dans le groupe 3. A l'inverse Brossard, d'abord placée dans le groupe 2 avec St-Hubert a mieux sa place dans le groupe 1, St-Hubert demeurant un cas spécifique qui ne s'associe à rien. Un fort pourcentage du territoire de cette municipalité, qui à bien des égards ressemble en effet à sa voisine Brossard est consacré à un aéroport militaire. On comprend pourquoi la fonction discriminatoire en ait fait un cas particulier. On pourrait ainsi commenter toutes les permutations de la matrice des distances à la centroïde des groupes. Qu'il nous suffise de dire ici qu'il a fallu deux

itérations pour obtenir une structure optimale et que c'est la troisième itération qui, ne déformant pas la classification mise en place par la deuxième, a démontré que la structure de T_2 était optimale. La somme des eigenvalues, qui n'était que de 1.70 au premier test, atteignait alors une valeur bien plus élevée, témoignant de l'amélioration de la classification: 4.33. Quand on sait d'expérience qu'en général il faut de 3 à 6 itérations pour stabiliser une classification, on conçoit tout ce que ce merveilleux système qu'est l'ordinateur a apporté à la science géographique. Un tel algorithme est en effet inconcevable sans le secours de l'informatique.

Les matrices qui suivent montrent les résultats de la première itération et de la troisième (celle qui confirme la classification optimale établie par la seconde) ainsi que le bilan final de toutes les permutations auxquelles l'ordinateur a dû se livrer.

TABLEAU 23

ANALYSE DISCRIMINATOIRE DE LA TYPOLOGIE ISSUE DU PROGRAMME LINYAUtilisation du sol non normalisé-varimax

GROUPE 1

		<u>Poids locaux des facteurs</u>			
1	Boucherville	-0.0070	1.2230	0.4360	0.0300
2	Laflèche	-0.0340	0.2700	0.5800	0.3120
3	Chat-Height	-0.0820	1.4090	0.4750	0.1200
4	Chat-Centre	0.2460	1.2270	0.4130	-0.4550
5	Delson	-0.8100	0.4490	-2.4420	0.4000
6	St-Lambert	0.3490	0.1920	0.0530	-0.2820

GROUPE 2

7	Brossard	-0.4480	0.7200	0.2560	0.1330
8	St-Hubert	-0.8250	-0.1650	-0.3510	3.4800

GROUPE 3

9	Candiac	-0.7050	0.9230	-0.7580	-0.8910
10	Chateauguay	0.2050	0.1650	0.7160	0.3050
11	Greenfield Park	0.9700	0.7740	0.0620	-0.5610
12	St-Bruno	-0.1980	-0.1210	-0.3410	-1.1140
13	Ste-Catherine	-0.8500	-1.6610	-1.0000	0.2500
14	Lemoyne	2.7030	-1.5610	0.1130	-0.1400
15	St-Constant	-0.1900	0.1330	0.3960	0.4180

GROUPE 4

16	Préville	-1.5950	-0.3490	-0.6850	-1.8850
17	La Prairie	0.4050	-0.7130	-1.0800	-0.5800
18	Jacques-Cartier	0.7340	0.4160	0.3050	0.0210
19	Longueuil	1.9980	-0.9080	-0.1420	0.0280

GROUPE 5

20	Léry	-1.8480	-2.2930	1.5580	-0.6020
21	Notre-Dame	-0.7760	-0.9480	1.0220	0.1520

MATRICE DES DISTANCES
DISTANCE A LA CENTROIDE DES GROUPES
(ITERATION 1)

GROUPE 1		1	2	3	4	5	ASSIGNATION
1	Boucherville	<u>0.6904</u>	2.1029	1.6209	2.0703	3.3548	1
2	Laflèche	<u>0.9077</u>	1.7052	1.0865	1.6489	2.5403	1
3	Chat-Height	<u>0.8579</u>	2.1185	1.8393	2.2891	3.4802	1
4	Chat-Centre	<u>0.8599</u>	2.5811	1.5564	1.0730	3.4586	1
5	Delson	2.5989	2.8568	2.8036	<u>2.5244</u>	4.7030	4
6	St-Lambert	0.8259	2.2850	<u>0.4345</u>	0.9281	2.8800	3
GROUPE 2							
7	Brossard	<u>0.5315</u>	1.7088	1.3182	1.7735	2.8750	1
8	St-Hubert	3.5741	<u>1.7088</u>	3.8175	4.0016	4.3619	2
GROUPE 3							
9	Candiac	<u>1.2990</u>	2.7891	1.7545	1.7948	3.6400	1
10	Chateauguay	1.1162	1.8763	<u>1.0931</u>	1.6675	2.5102	3
11	Greenfield Park	<u>1.1860</u>	2.8626	1.2340	1.4063	3.6217	1
12	St-Bruno	1.4930	2.9410	0.9775	<u>0.9679</u>	2.8535	4
13	Ste-Catherine	2.7812	2.7215	2.1096	<u>1.8976</u>	2.7252	4
14	Lemoyne	3.7013	4.3576	2.8104	<u>2.7273</u>	4.1854	4
15	St-Constant	<u>0.9245</u>	1.5092	1.0326	1.5513	2.4708	1
GROUPE 4							
16	Préville	2.7004	3.7972	2.4645	<u>2.3773</u>	3.1529	4
17	La Prairie	2.6334	3.4622	2.0415	<u>1.4607</u>	4.1408	4
18	Jacques-Cartier	<u>1.2792</u>	1.6686	1.4359	1.8375	3.0881	1
19	Longueuil	2.7055	3.4395	1.8665	<u>1.7948</u>	3.7182	4
GROUPE 5							
20	Léry	3.9663	4.0122	3.3955	3.6268	<u>0.9266</u>	4
21	Notre-Dame	2.5830	2.6601	2.2205	2.6012	<u>0.9266</u>	4

TABLEAU 25
MATRICE DES DISTANCES
A LA CENTROIDE DES GROUPES
 (ITERATION 3)

GROUPE 1		1	2	3	4	5	ASSIGNATION
1	Boucherville	<u>0.4659</u>	3.4579	1.0182	2.6371	3.4368	1
2	Chat-Heigt	<u>0.6198</u>	3.4468	1.2370	2.8590	3.5819	1
3	Chat-Centre	<u>0.5454</u>	3.9304	1.0757	2.4680	3.5447	1
4	Brossard	<u>0.5694</u>	3.1490	0.8918	2.2535	2.9186	1
5	Candiac	<u>1.3614</u>	4.2885	2.0209	1.9760	3.7170	1
GROUPE 2							
6	St-Hubert	3.5922	<u>0.0</u>	3.2639	4.1610	4.0123	2
GROUPE 3							
7	Laflèche	1.1002	3.0323	<u>0.4660</u>	2.1759	2.5606	2
8	Greenfield Park	1.2870	4.2191	<u>1.1135</u>	1.9676	3.6881	3
9	St-Constant	1.1825	2.8847	<u>0.6002</u>	2.0399	2.4707	2
10	Jacques-Cartier	1.5413	2.8680	<u>0.7794</u>	2.4234	3.2773	2
11	St-Lambert	1.1391	3.6915	<u>0.6219</u>	1.4248	2.9157	3
12	Chateauguay	1.3098	3.1256	<u>0.4571</u>	2.2020	2.5501	3
GROUPE 4							
13	Delson	2.7148	3.9953	3.0124	<u>2.2817</u>	4.5561	4
14	St-Bruno	1.7723	4.3992	1.7230	<u>1.0221</u>	2.9420	4
15	Ste-Catherine	3.1827	3.7794	2.7556	<u>1.5652</u>	2.5865	4
16	Lemoyne	4.0706	5.2733	3.1776	<u>2.9314</u>	4.1016	4
17	Préville	2.8879	5.1924	3.1431	<u>2.1901</u>	3.7997	4
18	La Prairie	3.1151	4.8373	2.9128	<u>1.1320</u>	4.1657	4
19	Longueuil	3.0785	4.4915	2.2125	<u>1.9874</u>	3.7058	4
GROUPE 5							
20	Léry	4.1407	4.7014	3.6901	3.5796	<u>0.9656</u>	5
21	Notre-Dame	2.7028	3.4580	2.1274	2.9561	<u>0.9656</u>	5

BILAN DES PERMUTATIONS
APRES TROIS ITERATIONS

GROUPE 1

1	Boucherville	1	1	1
2	Laflèche	1	3	3
3	Chat-Height	1	1	1
4	Chat-Centre	1	1	1
5	Delson	4	4	4
6	St-Lambert	3	3	3

GROUPE 2

7	Brossard	1	1	1
8	St-Hubert	2	2	2

GROUPE 3

9	Candiac	1	1	1
10	Chateauguay	3	3	3
11	Greenfield Park	1	3	3
12	St-Bruno	4	4	4
13	Ste-Catherine	4	4	4
14	Lemoyne	4	4	4
15	St-Constant	1	3	3

GROUPE 4

16	Préville	4	4	4
17	La Prairie	4	4	4
18	Jacques-Cartier	1	3	3
19	Longueuil	4	4	4

GROUPE 5

20	Léry	5	5	5
21	Notre-Dame	5	5	5
		↑	↑	↑

Un tel bilan est évidemment trop riche de signification pour ne servir qu'à tester l'optimalité d'une classification. Dès 1964 CASETTI suggérait que l'on pourrait l'utiliser pour identifier les véritables noyaux d'une classification ou d'une régionalisation. Ces noyaux "centraux" sont formés par les observations qui ne changent pas de groupe entre la classification T_0 et la classification terminale T_1 . Dans le cas d'une étude de régionalisation, il faudrait en outre que ces observations soient géographiquement contigues. C'est là une extension fort intéressante du programme et elle a été utilisée dans une étude sur la structure spatiale des Etats-Unis que J.B. RACINE publiera bientôt en collaboration avec une de ses étudiantes de maîtrise, assistante de recherche, Mlle Micheline MEUNIER. CASETTI suggérait également que l'on utilise ses itérations discriminatoires pour tester la validité d'une classification en considérant comme variables d'entrée deux ensembles caractérisant différents lieux dans l'espace ou dans le temps. Par exemple un modèle de classification des villes réalisé aux Etats-Unis pourrait être comparé à un modèle analogue dérivé de l'information recueillie sur les villes européennes. C'est ainsi que J.B. RACINE a trouvé au départ la justification théorique d'un nouveau modèle de recherche typologique qui lui était d'ailleurs imposé tout à la fois par les contraintes de l'information qu'il détenait et par la nécessité de rendre opérationnel son modèle global d'une recherche géographique informatique utilisant simultanément les démarches inductive et déductive et tenant compte des exigences tant idiographiques que nomothétiques de la géographie.

2.4) LES ETAPES DE LA MODELISATION DES SYSTEMES SPATIAUX

Après avoir déterminé des structures, découvert des typologies à l'aide des procédés factoriels et des algorithmes de classification, le géographe est en droit de se demander quelles sont les variables et les observations qui ont un rôle prépondérant sur le faciès spatial et de plus quel serait la répercussion sur le système spatial si on modifiait l'un ou l'autre des paramètres. Pour répondre à ces questions des outils s'offrent aux géographes pour développer des modèles qui vont lui permettre de simuler partiellement les résultats obtenus au stade de la réduction des attributs et des observations.

Les divers types de régression ont apporté à la géographie des éléments nouveaux surtout avec l'apport de la cartographie des résidus et méritent qu'on s'y attarde un peu.

2.4 a) La régression

Le lecteur a sans doute remarqué que le fondement de l'analyse multivariée en géographie repose au départ sur la notion de corrélation prise au sens statistique du mot. Bien que la corrélation statistique soit employée depuis peu de temps les géographes ont toujours essayé d'expliquer et de chercher des relations fonctionnelles entre les phénomènes en vue de la prédiction. De plus en plus on s'aperçoit que les relations spatiales sont complexes et qu'il devient difficile d'estimer des relations et des fonctions lorsqu'il y a plusieurs variables en jeu. C'est pourquoi la notion de régression permet de découvrir les liens de dépendance entre les variables, et de passer d'un modèle théorique à la réalité afin de vérifier statistiquement l'hypothèse initiale du modèle.

Afin d'éviter le reproche que ce travail soit trop mathématique au détriment de la géographie nous n'entrerons pas dans les détails de la régression qui en fait est relativement bien expliquée dans tout livre de statistique. Toutefois nous croyons qu'il est bon de passer en revue les différentes mesures qui apparaissent à la sortie de l'ordinateur afin que l'utilisateur soit en mesure de saisir la portée et la signification des résultats.

En géographie, les liaisons entre plusieurs phénomènes sont plus ou moins directement proportionnels les uns

aux autres contrairement aux phénomènes physiques où dans la plupart des cas des relations exactes se traduisent sur un graphique par une droite qui pour toute valeur de x , a une valeur correspondante y rigoureusement déterminée. Pour faciliter l'étude du processus d'interdépendance la connaissance de l'une des variables ne doit pas nécessairement déterminer l'autre exactement, mais la viser, avec le plus de précision possible, de façon à fournir l'estimation la plus probable. (Racine, 1972, à paraître). La réalisation de cette estimation sera directement proportionnelle à la concentration d'un "nuage" de points pris d'une courbe du premier degré $Y = a + bx$. C'est donc dire que la corrélation sera d'autant plus grande dans la mesure où les points s'alignent le plus près de la droite. Il va sans dire que la ligne de régression est extrêmement liée au coefficient de corrélation. En effet, un coefficient de corrélation positif près de 1 exprime une liaison linéaire positive (à chaque accroissement de la valeur x équivaut un accroissement proportionnel de y) et inversement pour une corrélation négative près de -1. Dans le cas d'une corrélation non significative il existe une possibilité infinie de droites susceptibles de donner la meilleure approximation de l'ensemble des points ce qui indique qu'il n'y a pas de relation linéaire précise entre x et y .

Le géographe qui tente une modification partielle à l'aide de la régression doit:

1. Déterminer la ligne de régression qui traduit le mieux la répartition de l'ensemble des points des variables x et y .
2. Vérifier la linéarité de la droite pour savoir si elle est représentative des observations.
3. Analyser la liaison trouvée afin de connaître le degré de signification de la liaison.

i) Détermination de la ligne de régression

La meilleure représentation d'un ensemble de points est fournie par la droite dont la somme des carrés des distances verticales aux divers points est minimale. Cette droite ($y = ax + b$) passe obligatoirement par le point de rencontre des moyennes des variables. Avec des données standardisées elle passe par l'origine puisque les moyennes sont égales à zéro; "l'intercepte" de la droite disparaît et l'équation prend la forme $y = ax$. La pente de la droite correspond au coefficient d'estimation appelé aussi coefficient de régression et équivalent à la tangente de l'angle formée par la droite et l'axe des x .

Nos travaux d'assistant avec J.B. Racine nous ont conduit à traiter les données de la relation densité-distance au centre-ville, pour la rive sud de la région métropolitaine de Montréal. Nous ne reprendrons pas ici toute la théorie qui a largement été expliquée ultérieurement par Clark (1951, 1958).

Berry, Simons et Tennant (1963), Newling (1966), Mercer (1968).

Le but de cette étude était de déterminer l'équation de la droite qui permettait de situer par rapport à son axe, l'axe où se situeraient toutes les observations si la densité dépendait de façon rigide de la distance au centre dans chacune des municipalités de la zone métropolitaine de Montréal (Racine). La variable dépendante du modèle était la densité au mille carré (moyenne pour la rive sud de 5,24) et comme variable indépendante la distance au centre (distance à vol d'oiseau) "à condition que l'oiseau vole en ligne droite" entre la centroïde de chacune des municipalités et le coeur du C.P.D. La relation trouvée est inverse comme nous nous y attendions mais son coefficient de corrélation linéaire est relativement faible $r = -0,44$, c'est-à-dire que moins de 20% de la variation de la densité était attribuable à la distance brute. En fonction de l'équation, la liaison est la suivante:

$$\text{Densité} = 9,53 + 0,03961 \times \text{la distance brute}$$

Etant donné que la relation n'était pas linéaire il a fallu transformer l'équation en logarithme pour assurer la linéarité et la normalité. Il va de soi que ceci n'a aucun effet sur le coefficient de corrélation qui est demeuré à $-0,44$ mais l'équation est devenue beaucoup plus significative.

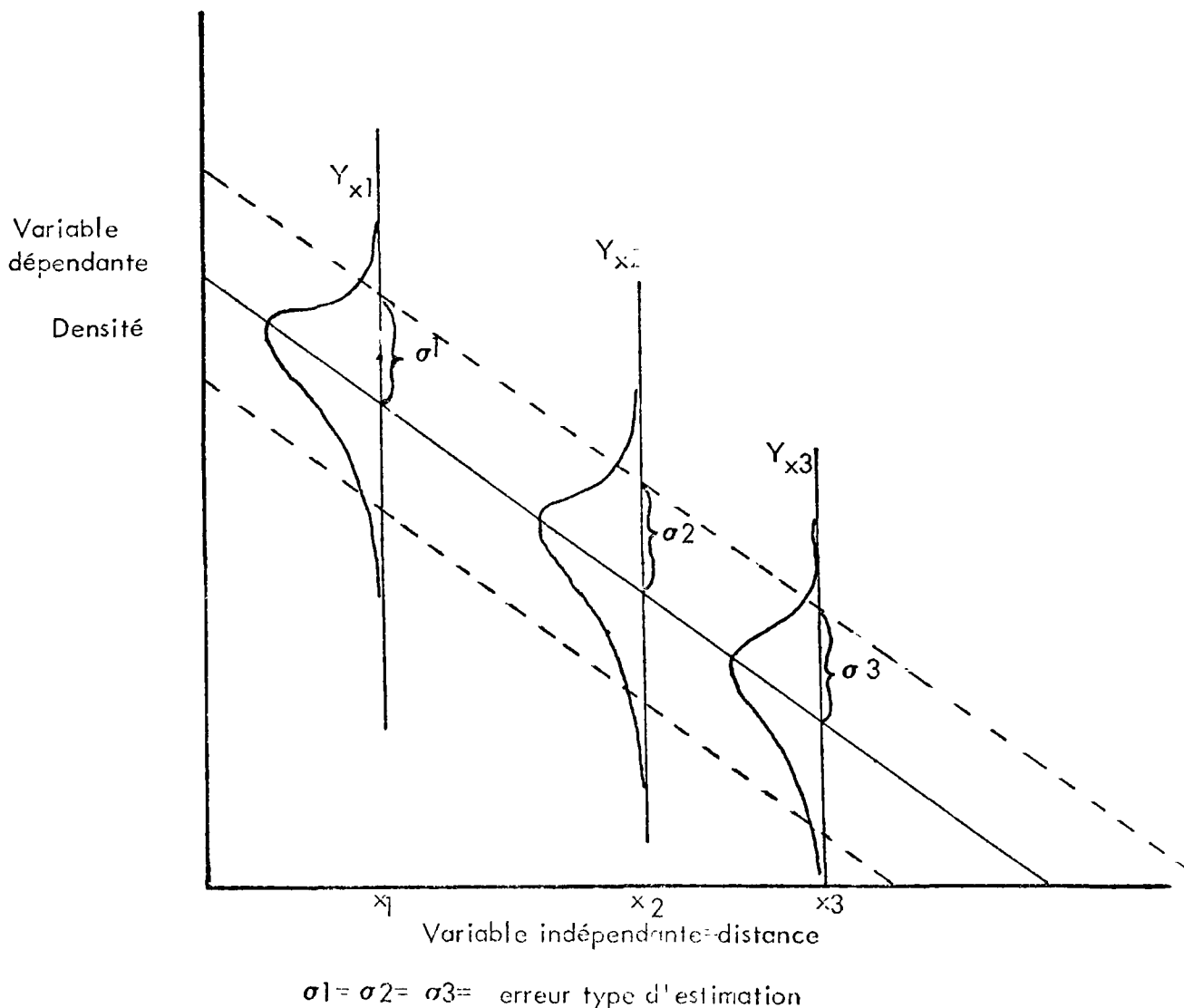
$$\text{Log } y = 9,96 + 0,94 \log x + e = 1,24$$

$$t = 3,70$$

La valeur de e représente l'erreur-type d'estimation correspondant à l'écart-type des colonnes en grandeur et en signification. Il est postulé que l'écart-type est constant d'une colonne à l'autre et que le postulat de l'homoscédasticité vu auparavant est également respecté. Pour chaque valeur de x (distance) il y a une distribution normale de y (densité). C'est précisément l'écart-type de cette distribution qui correspond à l'erreur-type d'estimation. La figure (24) illustre à propos cette notion.

FIGURE 24

REPRESENTATION GRAPHIQUE DE L'ERREUR-TYPE D'ESTIMATION



L'erreur-type d'estimation sert à fixer les limites entre lesquelles la valeur estimée de y a une probabilité de .68. Ces limites seront égales à la valeur estimée de y ou - l'écart-type. Avec l'équation suivante,

$$\text{LOG } Y = -0.03961 x + 9.53 = 5.04 \quad \text{si } x = 2$$

il y a 68 chances sur 100 que le logarithme de la densité soit entre 4.44 et 14.52. On en conclut que la précision de la régression sera d'autant meilleure dans la mesure où l'erreur standard d'estimation sera relativement petite. Pour ce qui est de l'épreuve de linéarité elle est faite avec le concours du test de t qui doit rejeter l'hypothèse nulle suivante: il n'y a aucune régression de y sur x c'est-à-dire que dans l'équation suivante

$$Y = ax + b$$

la valeur de a est égale à 0. Une acceptation de l'hypothèse nulle revient à dire qu'il n'y a pas de régression significative de Y sur X . Un test de signification peut aussi être fait à l'aide de l'analyse de variance basée elle aussi sur l'hypothèse nulle qu'il n'y a aucune relation entre les variables.

ii) L'analyse de variance comme test de signification pour la régression

La ligne de régression étant la meilleure droite possible qui représente l'ensemble des points il s'ensuit qu'il y a une variation dû au fait que tous les points ne sont

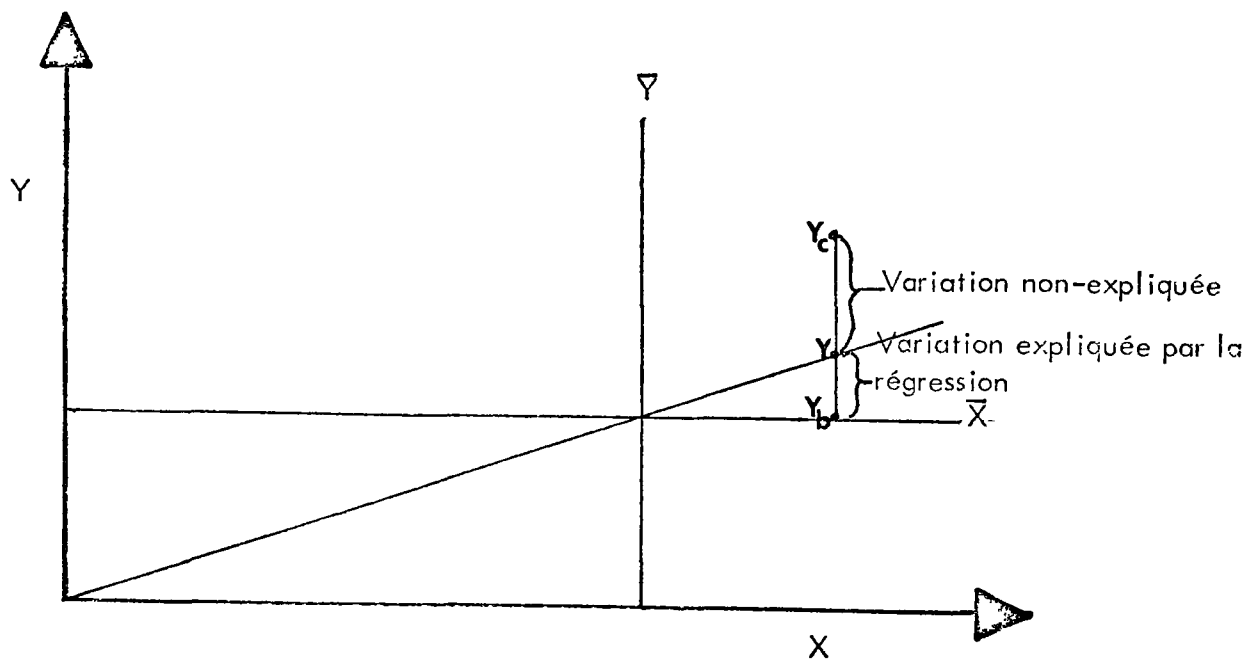
pas situés sur la ligne. De cette variation totale, une partie peut être expliquée par la régression tandis que l'autre demeure non expliquée.

Variation totale = variation expliquée + variation non expliquée

La figure (25) visualise la notion de variation.

FIGURE 25

REPRESENTATION GRAPHIQUE DE LA VARIATION DANS LA REGRESSION



La variation totale pour une valeur donnée est la distance d'un point à la moyenne . On appelle Y_c le point situé sur la droite de la régression. $Y - Y_c$ est égale à la variance non expliquée et $Y - \bar{Y}_b$ égale la variation expliquée par la régression.

Le rapport de la variance expliquée sur la variance non expliquée fournit un indice F qui permet à l'aide de tables de savoir si le rapport est significatif.

$$F = \frac{\text{variance expliquée}}{\text{variance non expliquée}}$$

n = nombre d'observations

m = nombre de variables

$m - 1$ degrés de liberté de la variance expliquée par la régression

$n - m$ degrés de liberté de la variance non expliquée

Les degrés de liberté correspondent au nombre de données indépendantes diminuées du nombre de restrictions.

Si la variation expliquée par la ligne de régression n'est pas différente significativement de la variance non expliquée le rapport F sera de deux niveaux.

1. Plus petit que 1 ou près de 0, la variance expliquée (numérateur) est plus petite que la variance non expliquée (dénominateur) et on calcule qu'il n'y a aucune relation entre les deux variables et que la modélisation n'est pas adéquate.

2. Par contre si F est plus grand que 1 la régression réussit par le fait même à expliquer la plus grande partie de la variation totale. Cependant il faut avoir recours aux tables pour savoir exactement quel est le niveau d'acceptation. En résumé si F est très grand on rejette l'hypothèse nulle qu'il n'y a aucune corrélation, et si F n'est pas significativement grand il est possible qu'il y ait une corrélation modérée et si F est très petit ou près de 0 la régression n'est pas valable.

iii) L'analyse des résidus et son intérêt géographique

Finalement la régression produit les valeurs résiduelles. Le principe des résidus est fort simple et très utile en géographie. Lorsque l'équation de régression est obtenue les valeurs de Y seront estimées en introduisant chaque valeur de la variable indépendante dans l'équation. Les résidus sont équivalents à la différence entre les valeurs initiales et estimées de Y . Ainsi moins une observation est conforme au modèle plus son résidu s'éloigne de 0. Un résidu positif indique que la valeur de Y a été sous-estimée par rapport à la régression et un résidu négatif signifie une surestimation de Y . Une régression est d'autant plus précise si les valeurs résiduelles tendent vers 0. La cartographie des résidus permet de fournir une image spatiale de la modélisation et la vérification de l'hypothèse de recherche à la base du modèle.

Avec J.B. Racine nous avons repris l'analyse pour la seule rive sud métropolitaine en utilisant la densité résidentielle et différentes approximations de la distance.

- distance brute
- distance-temps par automobile
- distance-temps par autobus et métro
- distance moyenne aux quatre ponts (Champlain, Victoria, Jacques-Cartier, Mercier)
- accessibilité relative du C.B.D. par rapport aux municipalités de la rive sud.

Cette dernière variable comprend les poids locaux de chacune des municipalités de la rive sud sur la principale et unique composante de l'analyse factorielle de toutes les autres approximations de la distance.

Les coefficients de corrélation obtenus sont négatifs et assez élevés.

- 0.70 pour la distance brute
- 0.51 pour la distance-temps par automobile au pont le plus proche
- 0.68 pour la distance-temps autobus jusqu'au métro de Longueuil
- 0.58 pour la distance moyenne aux quatre ponts
- 0.66 pour l'accessibilité relative

La distance brute de centroïde à centroïde donne la meilleure approximation de la distance ce qui est conforme au

modèle jusqu'à date. Quoique satisfaisante, une distribution des points sur papier graphique révélait que la relation n'était pas linéaire. Pour pallier à ceci il suffisait de transformer la densité en logarithme pour assurer la linéarité.

Les résultats obtenus ont été beaucoup plus élevés et significatifs.

- 0.80 pour la distance brute
- 0.64 pour la distance-temps automobile au pont le plus proche
- 0.80 pour la distance-temps autobus jusqu'au métro de Longueuil
- 0.71 pour l'accessibilité relative

Le test de t indique avec une valeur de 5.80 que distance brute est nettement meilleur modèle.

Le tableau (27) illustre les résultats obtenus avec chacune des régressions.

On remarque que malgré l'importance de la valeur de l'intercepte dans la relation qui unit la densité à la distance au centre, plusieurs municipalités s'écartent encore significativement de la droite de régression.

La régression est utile pour le géographe qui veut arriver à déterminer avec la plus grande précision possible les inter-relations entre les variables constituantes de l'espace qu'il étudie. Il est en mesure, par ce principe de simulation primaire, d'arriver à prédire l'effet par une variable de la

modification d'une autre inter-reliée avec elle. Il est certain que les processus stochastiques raffinent beaucoup les méthodes de prédiction mais il n'en demeure pas moins que la régression reste un outil difficile à manier, certes, mais très important pour l'analyse spatiale.

La régression multiple est une autre facette de l'immense domaine de la régression. Contrairement à la précédente on l'utilise pour obtenir le meilleur ajustement d'une série de mesures de variables indépendantes et dépendantes, à une équation de la forme,

$$Y = b + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

Y = variable dépendante

b, a₁, a₂, a₃ = coefficients à déterminer

x₁, x₂, x₃ = variables indépendantes

Le passage de la régression simple à la régression multiple oblige différentes distributions à être mises en corrélation sur plusieurs axes perpendiculaires. Une liaison rigide ne se traduit donc plus par une droite mais plutôt par un plan ou par une figure géométrique ayant autant de dimensions qu'il y a de variables indépendantes. Pour P_i points formant un nuage dans l'espace il s'agit donc de trouver un plan d'estimation passant aussi près que possible de l'ensemble de ces points (Liorzou, 1966, p. 188).

ANALYSE DE LA DISTRIBUTION DES DENSITES RESIDENTIELLES DANS LA
ZONE METROPOLITAINE DE MONTREAL AU SUD DU SAINT-LAURENT - (1966)

A - Les variables

- Log. y = Logarithme naturel de la densité de population à l'acre résidentielle.
 x_1 = Distance au centre (1/10 de mille) de chacune des centroides municipales au C.B.D.
 x_2 = Temps de parcours automobile jusqu'au côté NW du pont le plus proche de la centroide des différentes municipalités.
 x_3 = Temps de parcours en autobus jusqu'au métro de Longueuil.
 x_4 = Distance moyenne en milles du principal carrefour des municipalités aux quatre têtes de pont (compte non tenu du pont L.H. Lafontaine, postérieur à 1966).

B - Les modèles

1. Log. y = - 0.00725 x_1 + 3.88397 (r = - 0.79951 ; t = 5.80 ; e = 0.34782)
 2. Log. y = - 0.06699 x_2 + 3.81449 (r = - 0.64404 ; t = 3.67 ; e = 0.44299)
 3. Log. y = - 0.02717 x_3 + 3.63820 (r = - 0.79806 ; t = 5.77 ; e = 0.34894)
 4. Log. y = - 0.09371 x_4 + 3.8286 (r = - 0.71132 ; t = 4.41 ; e = 0.40702)

C - Valeurs résiduelles

Municipalités	Valeurs observées Log. y	Différence entre les valeurs observées et les valeurs estimées par les modèles				
		avec x_1	avec x_2	avec x_3	avec x_4	
Boucherville	1	3.0445	0.2799	0.1678	0.0313	0.1623
Brossard	2	2.6391	-0.3677	-0.9075	-0.1839	-0.5805
Candiac	3	2.5649	-0.2881	-0.6467	-0.3395	-0.5234
Chateauguay	4	2.3026	-0.1148	-0.1722	0.0231	-0.0173
Chateauguay-Centre	5	2.7080	0.4864	0.2333	0.4286	0.4256
Chateauguay-Height	6	2.5649	0.0751	0.0902	0.2855	0.2638
Delson	7	2.9444	0.0754	-0.1332	0.2030	-0.1533
Greenfield Park	8	3.2581	-0.1496	0.2474	-0.1083	-0.1113
Jacques-Cartier	9	3.6376	0.2314	0.4940	0.2711	0.4649
LaFlèche	10	3.4340	0.0662	0.4233	0.1219	0.1301
La Prairie	11	3.4340	0.6092	0.0884	0.4480	0.3925
Lemoyne	12	4.2195	0.7234	1.0079	0.8531	0.8782
Lery	13	1.9459	-0.0836	-0.1939	-0.3336	-0.1023
Longueuil	14	3.7842	0.3055	0.5056	0.2003	0.5740
Notre-Dame	15	3.1780	-0.0832	-0.1675	-0.2971	-0.0883
Préville	16	3.1780	-0.3260	-0.1675	-0.2427	-0.1351
Saint-Bruno	17	2.3979	-0.2232	-0.2108	-0.5609	-0.3718
Sainte-Catherine	18	2.4849	-0.5849	-0.4588	-0.1750	-0.5003
Saint-Constant	19	2.3979	-0.3414	-0.4118	-0.1261	-0.4936
Saint-Hubert	20	2.8332	-0.1105	-0.0435	-0.3974	-0.2551
Saint-Lambert	21	3.4012	-0.1790	0.2566	-0.1011	0.0411

$$\bar{y} = 2.96919 \quad \bar{x}_1 = 126.186 \quad \bar{x}_2 = 12.619 \quad \bar{x}_3 = 24.619 \quad \bar{x}_4 = 9.171$$

La méthode par étapes (step-wise) utilise les résultats intermédiaires pour obtenir de précieuses informations statistiques. Car avec l'équation de la régression multiple finale on obtient un grand nombre d'équations de régression partielles. Celles-ci s'obtiennent par l'addition à chaque étape d'une variable supplémentaire. Ainsi la première équation est une équation de régression simple à une variable

$$Y = b + a_1 x_1$$

et la seconde est en trois dimensions

$$Y = b + a_1 x_1 + a_2 x_2$$

et ainsi de suite. La variable que l'on ajoute à chaque étape est celle qui apporte la plus grande amélioration à l'estimation du plan. Cette technique de "pas à pas" permet d'introduire des variables qui peuvent sembler significatives à la première étape et être éclipsées par la suite en fonction de la rentrée des autres variables dans le système, qui donnent un meilleur ajustement.

2.4 b) La covariance multiple: un compromis entre la régression et les contraintes spatiales

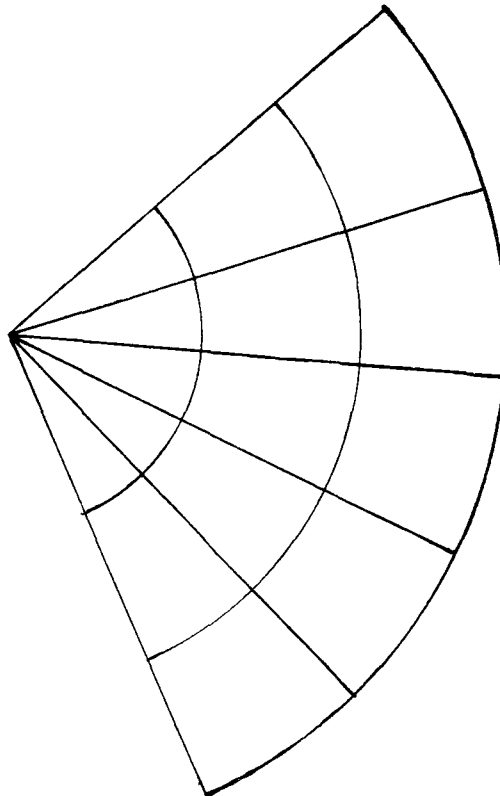
Pour le géographe, la covariance multiple est une façon de mesurer la variabilité d'un modèle dans l'espace. Lorsqu'on a étudié la relation distance-densité nous avons appliqué le modèle sans aucune discrimination sur l'ensemble de

la rive sud. Cependant il est peut-être possible qu'il y ait des différences spatiales qui font que le modèle ne soit pas significatif pour l'ensemble du territoire. Par exemple une autoroute modifie grandement une distance-temps. Pour contrebalancer les contraintes spatiales le géographe est intéressé à savoir s'il y a un effet de groupe (ou effets sectoriels) dans un modèle.

Ainsi l'équation de régression qui était $y = f(x) + e$ devient $y = f(x) + \text{effet sectoriel } e$. Il va de soi que si l'on veut déterminer l'effet de groupe ou l'effet sectoriel, il faut diviser sa région en groupes ou en secteurs. A titre d'exemple la rive sud serait divisible en secteurs dont le point de rencontre des rayons serait le C.R.D. (figure 26).

FIGURE 26

SCHEMA D'UNE CONFIGURATION SECTORIELLE



Lorsque les groupes sont déterminés la variable vide (dummy variable) est introduite spécifiquement à chaque groupe. La valeur constituante de cette variable est un (1) dont la représentation matricielle est la suivante:

$$Y = \text{Variables indépendantes} + \text{Effet sectoriel}$$

$$Y = ax_{11} + ax_{12} \dots \dots \dots + ax_{i,j} \quad X_{K1} \quad X_{K2} \quad \dots \dots$$

	$y_1 = ax_{11} + ax_{12}$		1	0
Groupe 1	$y_2 = ax_{21} \quad ax_{22}$		1	0
$K = 1$	$y_3 = ax_{31} \quad ax_{32}$		1	0

	$y_{nk} = ax_{nk1} \quad ax_{nk2}$		1	0

$i = 1 \text{ à } nK$
 $j = 1 \text{ à } m$
 $m = \text{nombre de variables}$
 $n_k = \text{nombre d'observations dans le groupe } K$

	$y_1 = ax_{11} \quad ax_{12}$		0	1
Groupe 2	$y_2 = ax_{21} \quad ax_{22}$		0	1
$K = 2$	$y_3 = ax_{31} \quad ax_{32}$		0	1
	$y_4 = ax_{41} \quad ax_{42}$		0	1

	$y_{nk} = ax_{nk1} \quad ax_{nk2}$		0	1

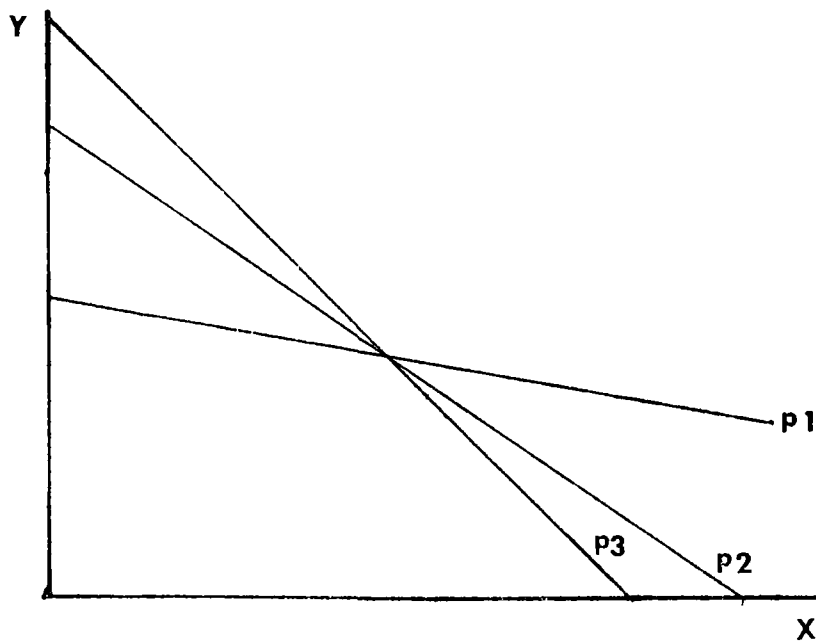
Avec la matrice de régression sectorielle il est important de savoir s'il y a hétérogénéité de la variance entre les groupes. L'hypothèse nulle repose donc sur l'homoscédasticité.

Pour pouvoir continuer l'analyse il faut pouvoir, grâce au test du X^2 , accepter cette hypothèse car s'il y avait une différence dans la variance entre les divers groupes, aucune comparaison ne saurait être statistiquement réalisable à cause des différentes relations dans les groupes.

Après vérification de l'égalité des variances entre les groupes, un deuxième test vérifie si le coefficient de régression est identique pour chaque groupe c'est-à-dire que les lignes de régression ont la même pente (figure 27).

FIGURE 27

LIGNES DE RÉGRESSION A PENTES VARIABLES



A l'aide du test de F la vérification de l'hypothèse nulle suivante s'impose.

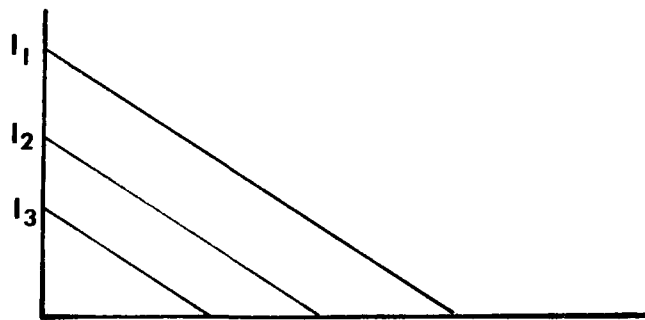
Mais là encore, il faut accepter H_0 pour poursuivre l'analyse. Un rejet de H_0 à ce niveau indique qu'il existe différentes relations fonctionnelles entre les groupes.

Finalement le troisième test considère les variations de l'intercepte de chaque ligne de régression. L'hypothèse nulle est la suivante: il n'y a aucune différence significative entre les moyennes.

Contrairement aux deux premiers tests, pour qu'il y ait effet de groupe il est impératif de pouvoir rejeter cette hypothèse nulle.

FIGURE 28

LIGNES DE REGRESSION A INTERCEPTES VARIABLES

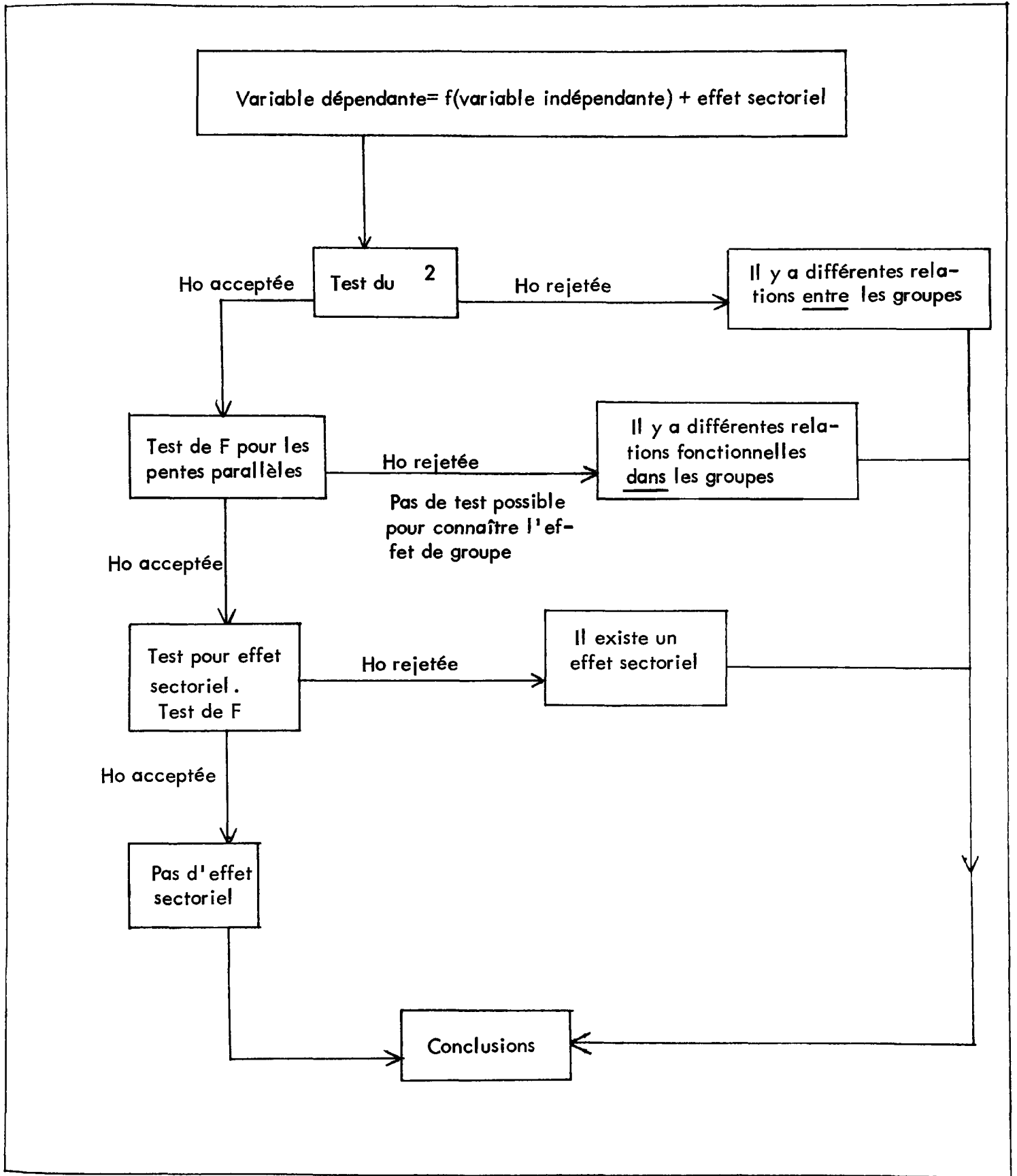


L'intercepte de la figure (28) étant différent signifie que le modèle n'est pas constant pour toute la région et qu'il existe donc des différences spatiales.

L'organigramme (figure 29) mesure les différentes étapes de la multivariance spatiale.

FIGURE 29

ORGANIGRAMME DE LA COVARIANCE MULTIPLE



Nous n'avons malheureusement pas d'exemple pour démontrer avec plus de clarté cette méthode importante qui en ce qui nous concerne n'en est qu'au stade expérimental actuellement.

Nous n'exposons ici que la théorie, qui nous l'espérons suscitera de l'intérêt auprès des adeptes de la modélisation spatiale, tous ceux qui, en particulier, travaillent sur les effets de distorsions de l'espace liés aux modalités tertiaires, aux effets d'agglomération, aux effets sectoriels, que ceux-ci dépendent du passage d'axes routiers ou des modalités de l'expansion au sein des aires métropolitaines des différents secteurs socio-économiques identifiés dès 1939 par l'économiste Homer Hoyt.

2.4 c) L'analyse canonique

Après avoir découvert à l'aide de l'analyse factorielle, les relations structurelles qui sous-tendent l'ensemble des concomitances observées dans la matrice des corrélations, après avoir établi des typologies à l'aide des algorithmes de classification, le géographe sera sans doute intéressé à savoir s'il existe des liens ou des relations entre les ensembles de variables ou entre des structures déterminées à l'aide de l'analyse multivariée. La corrélation canonique est l'outil tout indiqué pour trouver la corrélation maximum entre les fonctions linéaires de deux ensembles de variables. Le problème est de trouver deux ensembles de poids, en quelque

sorte des facteurs a et b, qui puissent maximiser la corrélation entre les équations de régression de chaque ensemble. A titre d'exemple prenons le cas où nous voulons savoir s'il y a corrélation entre la structure factorielle de l'utilisation du sol et celle de la structure familiale pour la rive sud de Montréal. (Caughnawaga n'est pas inclus ici dû à la difficulté d'avoir des données adéquates pour la deuxième série d'attributs). Les deux groupes X et Y de facteurs scores suivent et sont présentés algébriquement par les équations suivantes:

$$\begin{aligned}
 Y_1 &= a_1 y_{11} + a_2 y_{12} + a_3 y_{13} + a_4 y_{14} & X_1 &= a_1 x_{11} + a_2 x_{12} + a_3 x_{13} + a_4 x_{14} \\
 Y_2 &= a_1 y_{21} + a_2 y_{22} + a_3 y_{23} + a_4 y_{24} & X_2 &= a_1 x_{21} + a_2 x_{22} + a_3 x_{23} + a_4 x_{24} \\
 &\cdot & & \cdot \\
 &\cdot & & \cdot \\
 &\cdot & & \cdot \\
 &\cdot & & \cdot \\
 Y_{21} &= a_1 y_{21,1} + a_2 y_{21,2} + a_3 y_{21,3} + a_4 y_{21,4} & X_{21} &= a_1 x_{21,1} + a_2 x_{21,2} + a_3 x_{21,3} + a_4 x_{21,4}
 \end{aligned}$$

Nous retrouvons donc les équations de régression multiple que nous avons déjà expliquées. Géométriquement la corrélation canonique vérifie dans quelle mesure une observation occupe la même position relative dans l'espace à z dimensions, (z=4 pour l'utilisation du sol) et dans l'espace à

w dimensions ($w = 4$ pour la structure familiale).

Le point de départ est la matrice de corrélation de toutes les variables entre elles (z et w). Cette matrice a la propriété d'être divisible en quatre sections de la façon suivante:

$$R = \begin{array}{c|c} R_{11} & R_{12} \\ \hline R_{21} & R_{22} \end{array}$$

La matrice R_{11} contient les corrélations entre les variables z (poids locaux des facteurs de l'utilisation du sol).

La matrice R_{22} contient les corrélations entre les variables w (poids locaux des facteurs du statut familial).

La matrice R_{12} contient les corrélations entre les variables z et les variables w.

La matrice R_{21} est la transposée de la matrice R_{12} .

De ces quatre matrices R_{11} , R_{12} , R_{21} , R_{22} , on peut tirer une équation (Anderson, 1958) dont la minimisation fournit des racines caractéristiques (eigenvalues), terme déjà expliqué au nombre de variables du plus grand groupe. L'extraction de la racine carrée de la plus grande eigenvalue donne le coefficient de corrélation maximum entre les combinaisons linéaires des deux ensembles.

Voici les quatre eigenvalues avec les coefficients de corrélation canonique respectifs pour l'exemple en cours.

<u>Eigenvalue</u>	<u>Corrélation canonique</u>	
.86917	.86917	0.93229
.34601	.3466	0.58822
.20560	.20560	.45343
.02131	.02131	.14599

Mais ici revient l'éternel problème du degré de signification. Bartlett (1941, 1947) est arrivé à déterminer un indice Lambda (Λ) dont la distribution s'apparentait avec celle de χ^2 avec l'hypothèse nulle que les variables du premier ensemble sont non corrélées au second. Lambda peut se calculer par la formule suivante:

$$\Lambda = \prod_{i=w}^1 (1 - \lambda_i)$$

$$\Lambda_4 = (1 - 0.021) = .97$$

$$\Lambda_3 = .97 (1 - 0.12) = .84$$

$$\Lambda_2 = .77 (1 - 0.50) = .42$$

$$\Lambda_1 = .50 (1 - 0.85) = .06$$

Le χ^2 est obtenu par la formule suivante:

$$\chi^2 = - [N - .5 (w + z + 1)] \log e \Lambda$$

N nombre d'observations

w nombre de variables dans le groupe w

z nombre de variables dans le groupe z

Avec le premier lambda on obtient:

$$\chi^2 = - [21 - .5 (4 + 4 + 1)] \log e \quad 0.97$$

$$\chi^2 = 16.5 \times 2.813 = 46.46$$

Les degrés de liberté sont $(w - t) (z - t)$ où $t = i - 1$. Ainsi à la première eigenvalue $i = 1$ et $t = 0$, à la deuxième $i = 2$, $T = 1$, etc.

Donc le premier facteur canonique $\chi^2 = 46.46$ avec $(w - t) (z - t) = (4 - 0) (4 - 0) = 16$ degrés de liberté. La valeur du χ^2 est de 39.252 à .001 de probabilité. Avec un tel résultat on rejette l'hypothèse nulle puisqu'on a obtenu 46.46 et on conclut qu'il y a une relation entre la structure de l'utilisation du sol et le statut familial pour les 21 municipalités. Habituellement on reconsidère que le premier facteur, les autres sont rarement significatifs.

Cependant cette déduction statistique ne nous instruit pas sur la structure de la relation, il est impossible de dire quelles sont les variables qui, de l'utilisation du sol et du statut familial, ont le plus contribué à la relation. Pour répondre à cette question on se sert de saturations tout comme dans l'analyse factorielle. Dans notre exemple on a obtenu pour la première eigenvalue les saturations suivantes:

<u>Utilisation du sol</u>		<u>Famille</u>	
0.10388	F ₄	0.44579	F ₄
-0.13009	F ₃	0.59734	F ₂
-0.24498	F ₂	0.07180	F ₃
-0.85850	F ₁	-0.66280	F ₁

On remarque que la relation qui existe entre l'utilisation du sol et la famille est en partie due à la relation qui existe entre les deux premiers facteurs de chaque structure. Le premier facteur de l'utilisation correspond au niveau d'urbanisation défini par le résidentiel bifamilial et multifamilial, le terrain institutionnel en opposition au résidentiel unifamilial et au terrain vacant. D'autre part le premier facteur du statut familial se définit par l'habitat et le type d'occupant dont les variables composantes sont le pourcentage des maisons multifamiliales et d'appartements duplexes, le nombre de locataires, la proportion des chefs de famille de 20 à 24 ans et le nombre des familles qui n'ont pas d'enfants en opposition au pourcentage des maisons unifamiliales et du nombre de propriétaires.

Cette méthode de corrélation inter-structurelle est une combinatoire astucieuse de la corrélation et de l'analyse factorielle. On peut obtenir des poids locaux qui indiquent dans quelle mesure une observation spatiale reflète la relation entre les structures. Actuellement la façon de calculer

les scores de l'analyse canonique est la même que l'analyse factorielle. Nous croyons toutefois que cette méthode est biaisée (ceci serait à vérifier dans des recherches subséquentes) étant donné qu'elle ne tient pas compte du poids relatif de chaque structure. Cette question demeure une préoccupation pour les géographes dont le souci constant est la juste interprétation de la structure spatiale.

Malheureusement la corrélation canonique ne peut pas établir simultanément la corrélation entre plus de deux structures. Avec plusieurs trames dont chacune contient un aspect précis de l'espace, le procédé est ardu et long à cause de toutes les possibilités de corrélation entre les structures. Par exemple pour connaître la corrélation entre cinq trames on doit faire dix analyses canoniques. Une question se pose: comment procéder pour connaître simultanément les relations entre les structures découvertes à partir de plusieurs types d'analyse? La réponse à cette question dépasse les limites de l'analyse multivariée pour rejoindre le domaine de la théorie générale des systèmes avec laquelle nous nous familiariserons au niveau du doctorat.

2.5) CONCLUSION

A titre de conclusion nous sommes en droit de nous demander quel apport supplémentaire offrent les méthodes quantitatives précédentes à celles de la géographie qualitative.

Les descriptions qualitatives ne renseignent le chercheur ni sur la nature des combinaisons qu'il découvre ni sur leur intensité et ne disent rien des dimensions latentes qui sous-tendent la structure spatiale analysée. Cependant l'analyse factorielle permet de découvrir des relations indépendantes les unes des autres, de replacer ces dimensions dans l'espace et de maximiser la typologie obtenue à l'aide des différents algorithmes de classification. Elle donne des facteurs susceptibles de fournir des éléments de modélisation. L'analyse factorielle apparaît comme une nécessité dans la recherche des structures et dans la démarche scientifique. Les algorithmes de classification déterminent les sous-systèmes les plus homogènes, les plus précis et finalement ceux qui sont le plus fidèles à la réalité. Les techniques de régression offrent la possibilité de faire intervenir spécifiquement certaines variables susceptibles de fournir des éléments-clé d'interprétation et de connaître le rôle de certaines variables dans l'élaboration de la structure. Tous ces programmes offrent l'avantage de travailler sur une multitude de variables et d'unités d'observation pour en déterminer les

trames indépendantes.

Ces divers programmes peuvent être montés en batterie pour former un système informatique et analytique global, celui là même que nous allons présenter dans la dernière partie. Il s'agit d'un système parmi d'autres. Mais il nous a paru particulièrement conforme à la double logique idiographique et nomothétique de la géographie.

001734/2

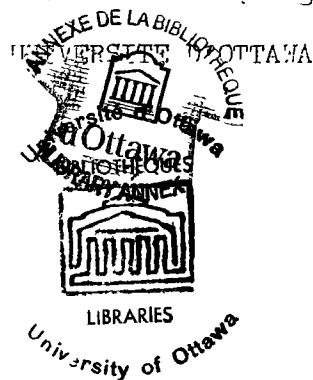
INFORMATION ET ESPACE GÉOGRAPHIQUE

LE MODELE S.E.T.C.E.G.

par

Guy Lenay

Baccalauréat ès Arts (Géographie)



Thèse présentée au Département de Géographie de

l'Université d'Ottawa

en vue de l'obtention de la

Maîtrise ès Arts

OTTAWA - SEPTEMBRE 1972

CHAPITRE III

UN SYSTEME INFORMATIQUE DE L'ANALYSE MULTIVARIEE
EN GEOGRAPHIE

LE MODELE S.E.I.G.E.G.

3.0) INTRODUCTION

Notre travail d'assistant de trois ans auprès de M. J.B. Racine à l'Université d'Ottawa nous a amenés à nous familiariser avec la technique informatique et les méthodes quantitatives. Les difficultés rencontrées pour adapter la programmation à la géographie, à établir les liens entre différents programmes, ont poussé notre intérêt à développer un système analytique qui permettra à d'autres, nous l'espérons du moins, de mettre à profit l'expérience que nous avons acquise en ce domaine.

Dans le cadre de notre effort pour rendre notre système opérationnel, nous avons voulu satisfaire deux exigences. La première, d'ordre pédagogique, était de fournir un outil facilement utilisable à une population de géographes et pouvant aussi servir à l'enseignement de la méthodologie géographique; la deuxième, d'ordre scientifique, était de faire en sorte que ce système soit relativement puissant afin de contribuer à la recherche et de développer de nouvelles techniques d'approche pour la connaissance de notre milieu. Tel est l'esprit avec lequel nous avons entrepris cette recherche que nous vous livrons dans cette dernière partie.

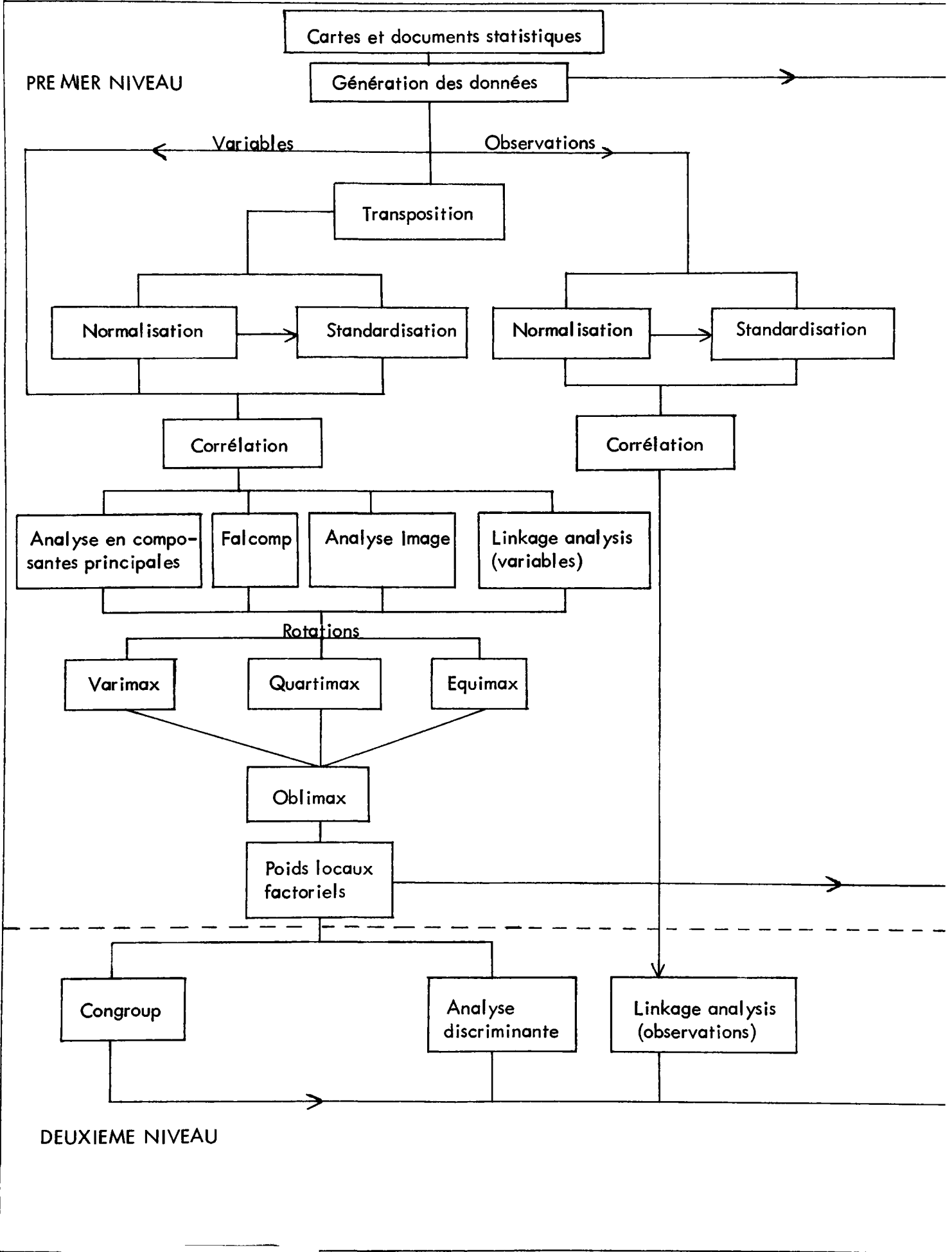
3.1) CARACTERISTIQUES GENERALES DU SYSTEME

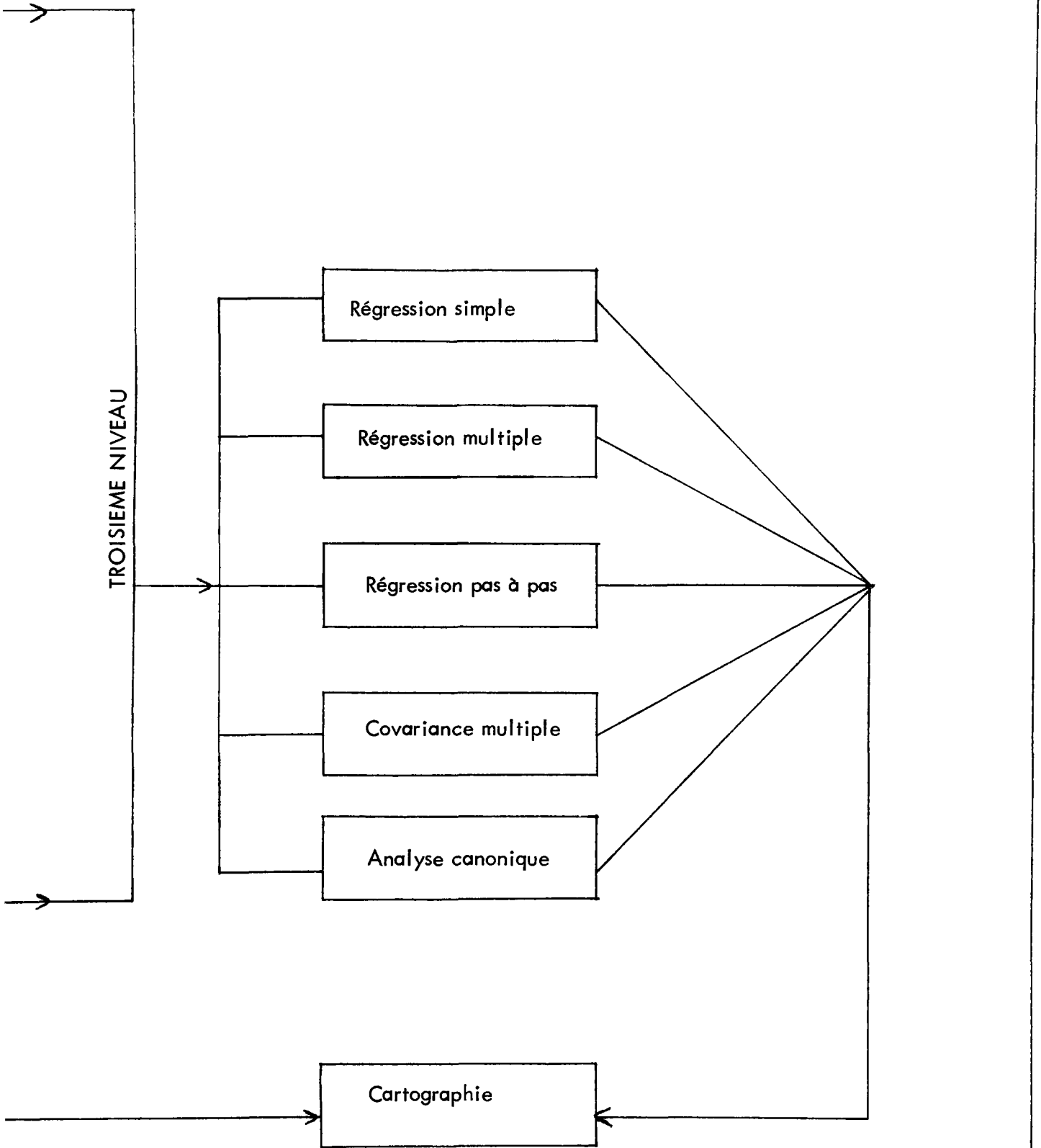
Le système au complet est composé de treize programmes assistés de cinquante-sept sous-programmes (voir annexe 2) écrits en langage Fortran sur environ 10,000 cartes. Il est aussi disponible sur bande magnétique à la librairie mécanographique du Centre de Calcul de l'Université d'Ottawa.

L'ordinogramme (figure 30) du système comprend trois parties principales.

1. Les programmes auxiliaires à la préparation des données à être introduites dans le système et ceux du traitement de l'information en vue de la réduction des attributs spatiaux.
2. Les programmes dont le but est de réduire les observations et d'exécuter les algorithmes de classification.
3. Les programmes nécessaires à la modélisation.

Lorsque le géographe possède sa matrice d'information spatiale il a le choix, selon ses buts, de travailler soit au niveau de la réduction des attributs, soit au niveau de la typologie. En fait il est préférable d'utiliser les résultats de la première réduction, les poids locaux des facteurs, comme matrice d'entrée aux programmes de classification.





3.1 a) Niveau de la préparation des données et de la réduction des attributs spatiaux

Lorsque la matrice d'information spatiale est recueillie, la première option qui s'offre à l'utilisateur est la transformation ou la génération de ses données. Le programme de génération est à la disposition de l'utilisateur qui désire manipuler ou combiner des variables pour en former des nouvelles. A titre d'exemple supposons que sa matrice d'entrée contienne des variables telles que le nombre d'habitants pour certaines villes et la superficie en milles carrés pour les unités d'espace correspondantes; il peut alors générer la variable densité en divisant la première par la deuxième. Ce programme peut aussi transformer des données en pourcentage, additionner, diviser, soustraire, mettre à la puissance n'importe quelle variable de la matrice d'entrée.

La manière d'utiliser ce programme sera expliquée ultérieurement. A la suite de ce programme un second choix s'offre à l'utilisateur, celui de la transformation. En passant par la transposition le géographe a l'intention de travailler au niveau des observations et sera conduit au "linkage analysis". Par contre en ne transposant pas sa matrice d'entrée il s'oriente vers la réduction des attributs spatiaux. Qu'il ait exécuté ou non la transposition, l'étape suivante est la normalisation et/ou la standardisation qui conduit à

la corrélation soit entre les attributs soit entre les observations, selon l'option que l'utilisateur a choisi au niveau de la transposition. La corrélation termine pour ainsi dire l'étape dite de la préparation des données. Après quoi s'offre une vaste sélection de programmes dont l'analyse en composantes principales, l'adaptation combinée FALCOMP que nous avons proposée, l'analyse-image et le linkage analysis. Ce dernier a la caractéristique d'être ambivalent c'est-à-dire capable de réduire soit des variables, soit des observations et c'est pourquoi il a une position mitoyenne entre le niveau de la réduction des variables et celui de la diminution des observations. Le cheminement logique du système conduit vers les différents types de rotation tels que Varimax, Quartimax, Equimax et Oblimax. Les trois premières sont optionnelles tandis que l'Oblimax est exécutée à chaque fois comme test d'orthogonalité entre les facteurs. Ce dernier programme clôt la première partie du système.

La section suivante a pour objectif la classification et le groupement des unités spatiales.

3.1 b) Réduction des observations

Après les analyses structurales les poids locaux des facteurs sont calculés automatiquement par la méthode classique (voir deuxième partie, section du calcul des scores) et celle de Kaiser. La cartographie des facteurs scores en

autant de cartes que de facteurs est possible en les divisant en classes tel que:

- supérieur à 1'écart-type
- entre 0,5 et 1 écart-type
- entre 0 et -0,5 écart-type
- entre -0,5 et -1 écart-type
- inférieur à -1 écart-type

Notre système malheureusement ne fait pas cette classification automatiquement. Il ordonne toutefois les facteurs en ordre décroissant ce qui facilite énormément la tâche. La superposition des cartes offre au chercheur l'opportunité de tirer des conclusions sur la conjoncture structurale étudiée.

Après les scores, le congroup, le discrim et le linka sont disponibles pour développer une typologie ou étudier la répartition de ses facteurs dans le milieu étudié. Ces trois programmes aboutissent à la cartographie manuelle ou automatique par le Système Symap. Le temps manquant il a été impossible pour nous de relier notre système à celui du Symap. Cependant la clarté et la simplicité d'utilisation du système Symap font que le chercheur ne devrait pas avoir de difficulté majeure à faire cartographier automatiquement les résultats obtenus à l'aide du S.E.I.G.E.G.

Pour celui qui serait en mal de savoir quelle méthode de classification utiliser, nous lui conseillons d'exécuter les trois (congroup, discrim, linka), de comparer les résultats et d'opter pour ceux qui correspondent le plus soit à son hypothèse de départ, soit aux résultats qui font le meilleur écho à la réalité empirique observée. L'hypothèse de départ est quelquefois traduite sous forme de modèle, elle exige alors le recours à la simulation partielle.

3.1 c) La simulation partielle et les différents types de régression

Cette dernière partie du système utilise dans la plupart des cas, les données de la matrice d'information spatiale et les "facteurs scores" qui expriment en fait le poids synthétique de plusieurs variables. Les programmes de cette section sont:

- i la régression simple et multiple
- ii la régression par étape
- iii la multivariance
- iv l'analyse canonique

Nous ne reprendrons pas la démonstration de ces méthodes mais soulignerons seulement la raison, d'ordre fonctionnel autant que méthodologique, de leur position à la fin du système. En effet, le géographe qui veut étudier une structure spatiale et en arriver à détecter les mécanismes de son

organisation doit au départ connaître les différentes composantes indépendantes de l'espace étudié afin de pouvoir modifier l'un ou l'autre des paramètres tout en laissant les autres invariables. Cela nécessite des résultats factoriels adéquats.

Cette condition offre l'opportunité de connaître avec précision l'effet de la variation d'une structure sur différentes caractéristiques spatiales. Ce moyen d'employer la régression est d'après nous nettement supérieur aux simples régressions entre deux ou trois variables qui fournissent à peine plus d'information qu'une simple corrélation. En inter-reliant des structures et des variables on s'approche de la technique de la simulation complète et de la théorie générale des systèmes.

Avec la modélisation partielle les tables de résidus prennent une signification particulière pour les géographes. Ceci est d'autant plus vrai que les autres systèmes dédiés aux sociologues, économistes ou autres tel que S.P.S.S. (Scientific Package for Social Sciences) ne donnent pas les résidus ni les poids locaux des facteurs. Tout comme ces derniers les résidus sont cartographiables et par le fait même offrent l'avantage d'être comparés aux résultats obtenus à la première et à la deuxième partie du système par l'entremise d'une dimension commune: l'espace.

La figure 30 visualise l'organisation du système S.E.T.G.E.G. et indique les cheminements possibles offerts aux géographes. On remarquera les trois différents niveaux du système et leur convergence vers la cartographie, qui montre bien la préoccupation avec laquelle nous avons élaboré l'agencement des programmes.

Ceci nous conduit finalement à l'aspect technique du système. Le lecteur trouvera ci-après les divers informations et renseignements qui lui permettront d'utiliser avec le maximum d'efficacité le système S.E.I.G.E.G.

3.2) CARACTERISTIQUES TECHNIQUES DU S.E.T.G.E.G. ET MODE D'EMPLOI

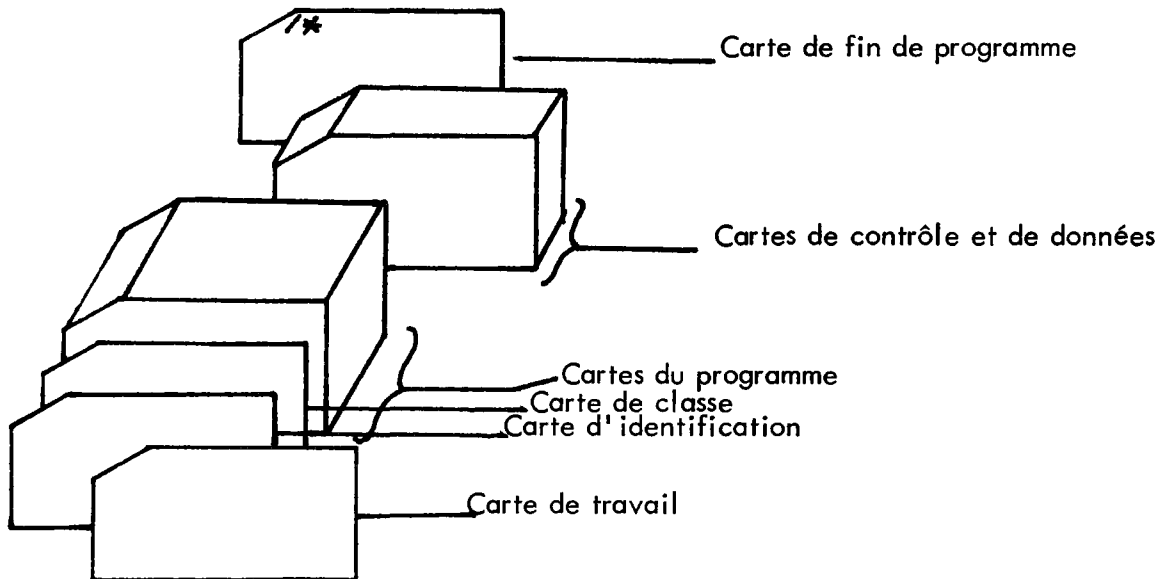
Chaque fois qu'un usager passe un programme tel que le S.E.I.G.E.G., sa soumission doit avoir la configuration suivante:

- i carte de travail
- ii carte d'identification
- iii carte de classe
- iv cartes du programme
- v cartes de contrôle et les données
- vi carte de fin de travail

Les cartes sont représentées schématiquement à la figure (31).

FIGURE 31

SCHEMA GENERAL DE LA SOUMISSION D'UN PROGRAMME A L'ORDINATEUR



* Les cartes en retrait à droite sont-elles à modifier pour les différents programmes

FIGURE 32

CARTES MASTRESSES DU PROGRAMME

CLASSE=B

GUY LENAY MSGLEVEL=02,00

7700004921 JOB

UNIVERSITÉ OTTAWA
CENTRE D'INFORMATIQUE

UNIVERSITY OF OTTAWA
COMPUTING CENTRE

0000
1111
2222
3333
4444
5555
6666
7777
8888
9999

CATEGORIE - CLAYS		IMPRIMER - PRINT		LIGES LINES		TRACEUR - PLOTTER		CARTES - CARDS		SPECIAL INSTRUCTIONS	
NO DE SOUS-SESSION - JOB NO		CHOIX - CORE		MIN		A PERFORER - PUNCH		NO		SPECIAL INSTRUCTIONS	
NO DE SOUS-SESSION - JOB NO		CHOIX - CORE		MIN		A PERFORER - PUNCH		NO		SPECIAL INSTRUCTIONS	

REMARQUES

- La carte de travail doit être changée à chaque soumission. D'ailleurs après la lecture de son programme on ne reçoit que la section gauche de la ligne pointillée. (figure 32)

- La carte d'identification doit avoir le nom de l'utilisateur dans les colonnes 19 à 33.

- La carte de classe est habituellement incorporée au système par l'opérateur lui-même. Elle indique la priorité du travail. Le système IBM 360-65 possède quatre classes principales:

Classe A : Programme utilisant moins de 100 K de mémoire, et moins de deux minutes de temps-machine.

Classe B : Utilise entre 100 et 200 K de mémoire et moins de deux minutes de temps-machine.

Classe C : Utilise entre 200 et 300 K de mémoire avec un temps de calcul de deux à cinq minutes.

Classe D : Plus de 200 K et plus de cinq minutes. Emploie des rubans magnétiques ou des disques.

Le S.E.I.G.E.G. sur carte passe en classe C avec 225 K et deux minutes approximativement de temps-machine. Lorsqu'il est en librairie sur bande magnétique il passe en classe D.

3.3) CARTES DE CONTROLE DE LA PREMIERE PARTIE DU SYSTEME S.E.I.G.E.G.

Le rôle de ces cartes est de fournir les renseignements nécessaires au programme afin qu'il exécute l'analyse multivariée voulue.

3.3 a) Cartes générales de contrôle.

i) Carte de spécificité de la matrice d'information spatiale.

Colonne	Contenu	Explication	Sigle de la constante*	Exemple
1-3	Nombre d'observations	maximum 150	NOB	022
4-6	Nombre de variables	maximum 100	NVAR	012
7	1 ou 0	si l'utilisateur a l'intention d'identifier ses variables - cas contraire	IPOV	1
8	1 ou 0	- s'il y a un titre à l'analyse - pas de titre	ITIT	1

* Nous avons cru bon de déterminer le sigle de la constante pour but de renseigner celui qui voudrait se référer au programme. Il n'est nullement nécessaire pour faire fonctionner le système.

- iii) Carte d'identification des variables obligatoire si
IPOV = 1 (première carte)

Colonne	Contenu
1 - 20	nom de la variable

Il doit y avoir un nombre de cartes égal au nombre de variables. Sinon une erreur se produit. Exemple figure (33).

- iv) Cartes de format: 5 obligatoires.

L'utilisateur dispose de cinq cartes pour écrire son format de lecture de données et les cinq cartes doivent être incluses même s'il ne les utilise pas au complet.

Colonne	Contenu
1 -80	(20A1/9F8.2/3F8.2)
	N.B. commencer à la colonne 1.

Le format* ci-haut indique à l'ordinateur de lire trois cartes pour chaque observation. Sur la première carte lire 20 caractères alphanumériques (20A1) qui sont le nom de l'unité d'observation. Le symbole (/) signifie de passer à la carte suivante sur laquelle l'ordinateur devra trouver 9 nombres de cinq chiffres, un point décimal, deux chiffres (exemple: 34678.25). Le second / demande à l'ordinateur de

* Si la notion de format n'est pas connue il serait souhaitable de consulter un livre d'introduction sur la programmation en Fortran.

lire la carte suivante qui devra comprendre 3 nombres de 5 chiffres au centième de précision.

Quatre cartes blanches suivent. La raison pour laquelle on réserve cinq cartes pour le format de lecture est dû à certains formats qui sont parfois très longs à cause d'une très grande quantité de variables à lire.

Voici l'exemple du format de lecture accompagné des cartes de donnée de la première observation qu'il doit lire. (figure 34).

N.B. Si IPOV = 1 (cf. première carte) le nom des observations doit être lu absolument. Nous conseillons d'identifier les variables et les observations car cela facilite l'interprétation des résultats. L'esthétique des tableaux de sortie a été développée en fonction du format de lecture F8.2, et nous conseillons fortement cette base pour les formats mais ceci n'empêche pas cependant d'en utiliser d'autres. Exemple, F15.0.

v) Cartes de données

Les données (matrice d'information spatiale) suivent les cartes de format. On peut utiliser plus d'une carte par observation cependant les limites du système sont de 100 variables et de 150 observations. Les données doivent être préparées avec un point décimal car le système n'accepte pas de nombre entier. A titre d'exemple le nombre 10 pour être admis

par le système doit être accompagné d'un point donc 10, ou 10.0.

Les valeurs maximum et minimum du système sont 99,999,999.99,999 et -99,999,999,999,999. Si l'occasion se présente où des valeurs excèdent les minimum et les maximum, il suffit de diviser par 10, 100 ou 1,000 toutes les valeurs de la variable. Ce procédé rend la variable acceptable par le système sans en changer la représentativité relative.

vi) Cartes de sélection

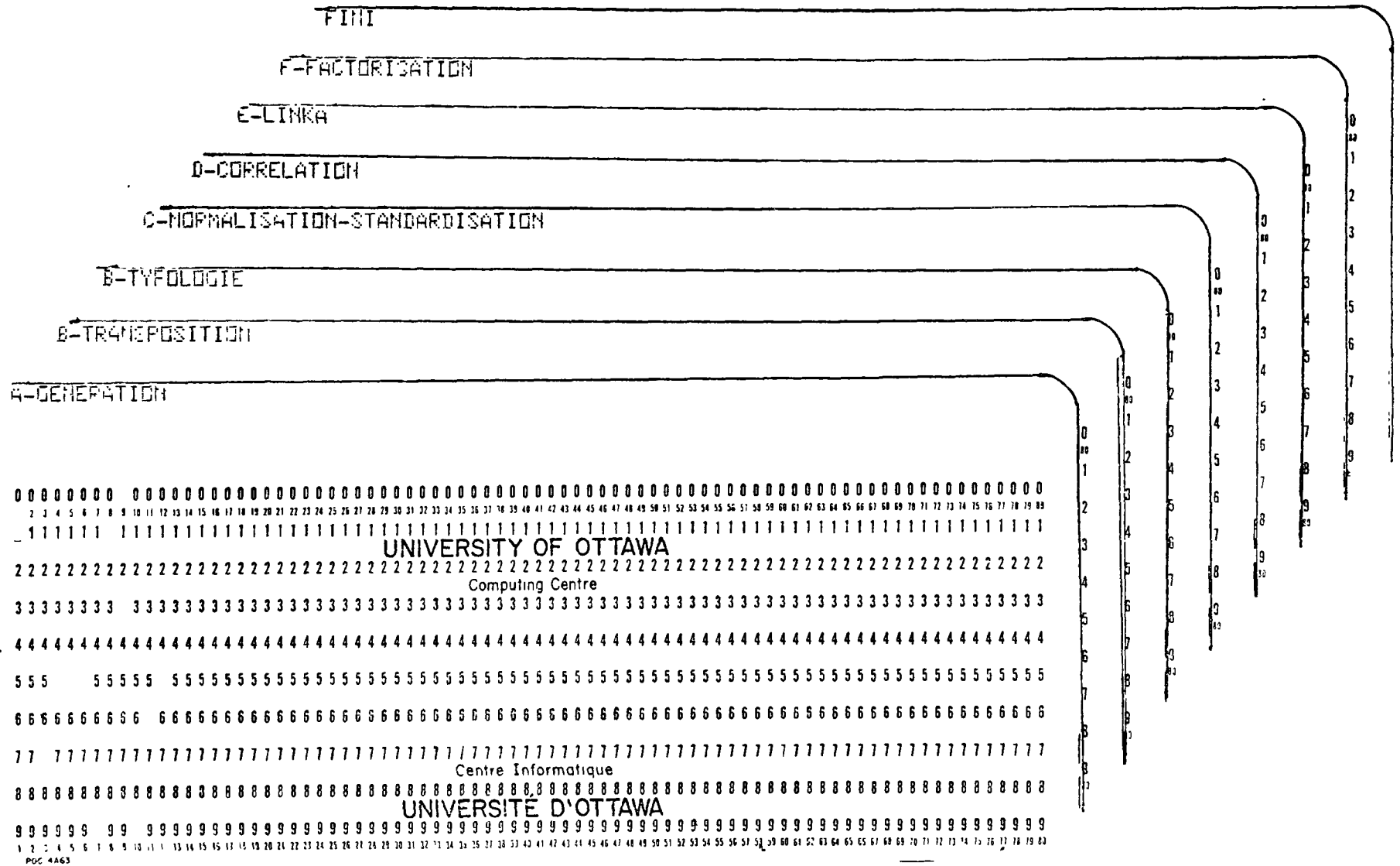
Le principe sur lequel nous nous sommes basés pour établir les cartes de sélection est identique à celui développé au laboratoire d'analyse spatiale de l'Université de Harvard. Chacune des sélections est précédée d'une lettre alphabétique. L'ordre séquentiel dans lequel doivent être mises les cartes de sélection doit correspondre à l'ordre alphabétique. A titre d'exemple une sélection avec la lettre E ne doit pas être mise avant celles de A, B, C, ou D. Cette contrainte oblige à respecter l'organigramme du système sans provoquer de complications. A la figure 35 on retrouve les cartes de sélection de la première partie du système.

3.3 b) Carte: A - Génération

Cette option sert à générer et à modifier les variables de la matrice d'information spatiale. La configuration

FIGURE 35

CARTES DE SELECTION DE LA PREMIERE PARTIE DU SYSTEME



générale des cartes de contrôle qui suivent la carte A- Génération est la suivante:

Colonne	Contenu	Explication
1-3	Un choix parmi les abréviations suivantes	
	ADD	Addition
	SOU	Soustraction
	MUL	Multiplication
	DIV	Division
	PUI	Puissance
	SQR	Racine carrée
	EXP	Exponentielle
	LOG	Logarithme naturel
	LCO	Logarithme commun
	ABS	Valeur absolue
	PER	Pourcentage
	FIN	Fin
4	(Indique le début de la description des opérations
5-10	VAR XXX	XXX (toujours trois chiffres) correspond au numéro de la nouvelle variable générée. Par exemple si on crée la variable 15 on retrouvera dans les colonnes 5-10 VAR 015

Colonne	Contenu	Explication
11	,	Virgule de séparation des paramètres
12-17	VAR XXX	Numéro de la variable qui sert à la génération. ex. VAR 003
18)	Cette parenthèse doit apparaître dans la colonne 18 si le code d'opération (col 1-3) est un des suivants SQR, EXP, LOG, LCO, ABS, PER
	ou	
	,	Cette virgule de séparation de paramètres si le code d'opération est un des suivants, ADD, SOU, MUL, DIV, PUT
19-24	VAR XXX	Variable ou constante qui sert à la génération s'il y a une virgule dans la colonne 18
25)	Parenthèse de la fin
40-60		Nom de la nouvelle variable générée.

Remarques: il doit y avoir une carte de contrôle pour chaque génération. Sur la dernière carte écrire le mot FIN dans les trois premières colonnes.

Voici quelques exemples susceptibles d'écclairer ce qui précède.

Supposons que la variable 5 de la matrice d'information spatiale corresponde au nombre d'habitants par mille, et

que la variable 10 soit la superficie en milles carrés pour chacune des villes. A l'aide de ces deux variables on peut créer la variable densité no. 13 en divisant la variable 5 par la variable 10. La carte de contrôle sera:

	Colonne 40-60
DIV(VAR011,VAR005,VAR010)	DENSITE

Exemple de transformation en pourcentage de la variable 14 pour créer la variable 15.

	Colonne 40-60
PER(VAR015,VAR014)	VARIABLE 14 en %

Exemple d'addition d'une constante 10.25 à la variable 1 pour créer la variable 2

	Colonne 40-60
ADD(VAR002,VAR001,10.25)	Addition de 10.25 à la variable 1.

Exemple d'extraction de la racine carrée de la variable 3 pour remplacer la variable 3.

	Colonne 40-60
SQR(VAR003,VAR003)	Racine carrée de la variable 3.

Le programme accepte n'importe quelle sorte de constante. En voici quelques unes:

10
 10.
 0.2
 1/2
 1./2
 3.4/5.6
 1/.003

Il est impératif que les numéros des variables soient continus. C'est-à-dire que si au départ par exemple la matrice d'information spatiale contient 12 variables la nouvelle variable générée devra porter le numéro 13 et ainsi de suite ou de 1 à 12 si elle en remplace une autre.

Le programme accepte plusieurs cartes de contrôle. La dernière carte de la section A-Génération le mot FIN doit être perforé dans les colonnes 1-3.

Voici un exemple des cartes de contrôle de la section A.Génération, figure 36 .

3.3 c) Carte: B - Transposition

Avec cette carte l'ordinateur transpose automatiquement la matrice d'information spatiale.

3.3 d) Carte: B - Typologie

Le lecteur notera qu'il a deux B (B - transposition et B - typologie). Ceci est dû à la position identique qu'ils occupent dans le schème opérationnel de système, l'utilisateur doit employer l'un ou l'autre ou aucun des deux.

Cette carte de sélection ordonne à l'ordinateur de transposer la matrice d'information spatiale, d'y effectuer la corrélation entre les observations et d'exécuter le linkage analysis.

Il n'y a pas de carte de contrôle pour B - typologie sauf que la carte C - normalisation et ses cartes de contrôle sont obligatoires.

3.3 e) Carte: C - Normalisation - Standardisation (obligatoire pour
B - typologie)

Demande de la normalisation et/ou de la standardisation de la matrice d'entrée.

Deux cartes de contrôle successives.

- Carte 1.

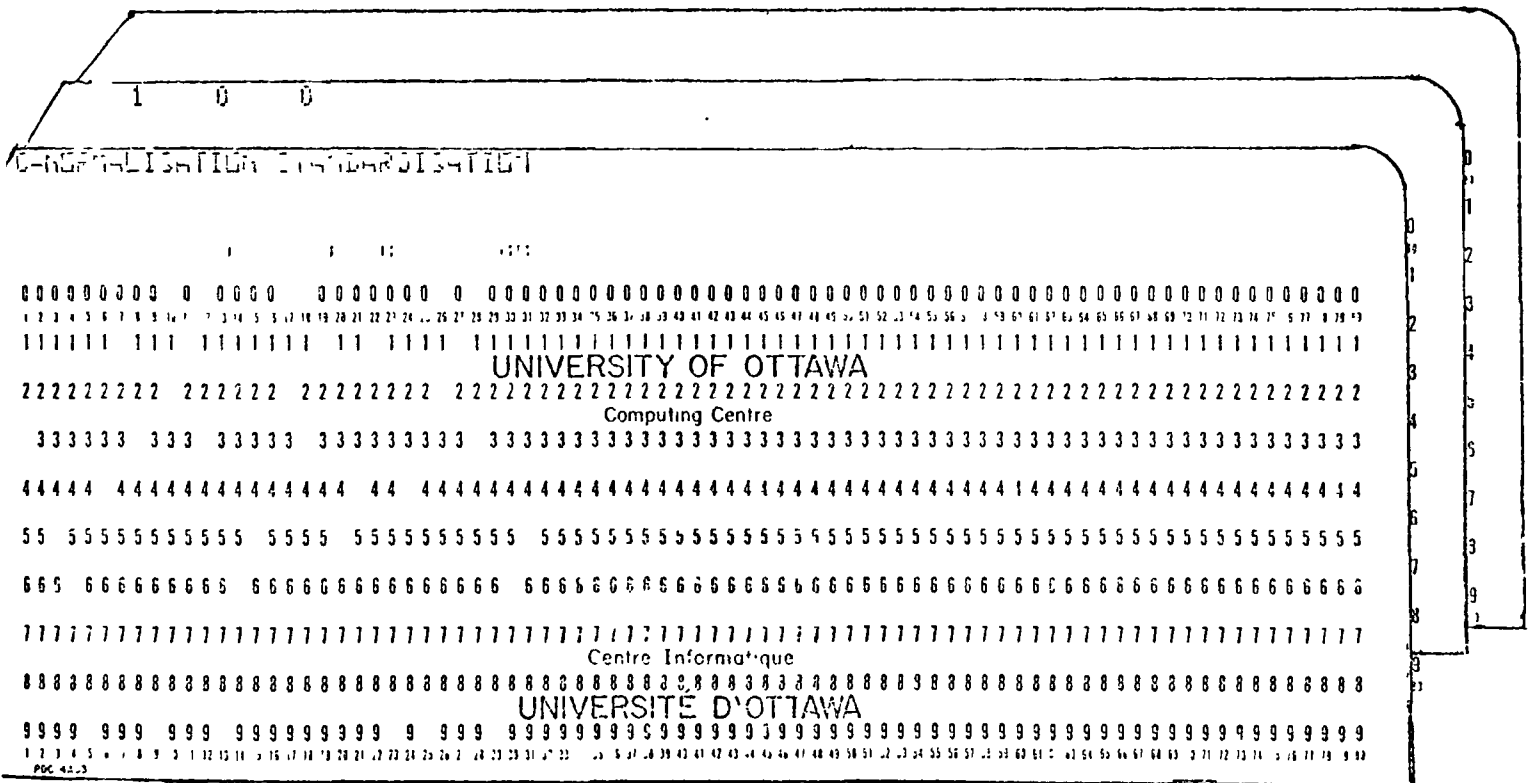
Colonne	Contenu	Explication	Sigle	Exemple
5	1	Perforation de la matrice normalisée	NPXS	1
	0	Pas de perforation		
10	1	Perforation de la matrice standardisée	NPSS	0
	0	Pas de perforation		
15	2	Ne pas normaliser ni standardiser	1MISS	0
	1	Standardiser seulement		
	0	Normaliser et standardiser		

- Carte 2.

L'addition de cette carte offre l'opportunité d'empêcher que certaines variables soient normalisées ou standardisées. Le code 1 perforé dans la colonne correspondante du numéro de la variable établit la contrainte. Par exemple si nous ne voulons pas que la variable 10 soit normalisée on perfore 1 dans la colonne 10. Pour plus de 80 variables il faut ajouter une troisième carte, parce que chaque carte contient 80 colonnes. Si la normalisation doit être effectuée

sur toutes les variables on met une carte blanche. Il est important que cette carte y soit.

Exemple:



vations. Aucune carte de contrôle succède cette carte.

3.3 h) Carte: F - Factorisation

Carte de sélection qui donne accès aux différentes analyses factorielles.

- Cartes de contrôle:

Colonnes	Contenu	Explication	Sigle	Exemple
1-2	01	Principale composante	FACTO	02
	02	Image-ana- lysis		
	03	FALCOMP		
3-4	-1	Rotation quartimax	IROT	01
	00	Rotation equimax		
	01	Rotation varimax		
5-10	000.10	Valeur mini- mum (cut-off) pour laquelle les eigenvalues sont considérées. Ainsi lorsque ICON 1.0 l'or- dinateur ne con- sidère que les facteurs avec des eigenvalues supérieures à 1. Si ICON 0.0 on obtient autant de facteurs que de variables.	ICON	0001.0

Colonnes	Contenu	Explication	Sigle	Exemple
11-12	01	Les satura- tions rotation- nées ainsi que les facteurs`	NPUF	00
	ou	scores sont per- forés sur carte		
	00	Pour ne pas per- forer les satu- rations et les facteurs scores		
13-14	01	Pour imprimer les résultats intermédiaires de l'image ana- lysis	NPIR	01
	ou			
	00	Pour ne pas im- primer les ré- sultats inter- médiaires		
15-16	01	Pour perforer des eigenvecteurs par colonne	NPUEV	00
	ou			
	00	Pour ne pas per- forer les eigen- vecteurs		
17-18	00	Pour imprimer les facteurs rotationnés normalisés	NP	00
	ou			
	01	Pour ne pas les imprimer		
19-20	00	Pour normali- sation des facteurs rota- tionnés	NURI	00
	ou			
	01	Pour ne pas nor- maliser les fac- teurs		

3.4) LES CARTES DE CONTROLE DES ALGORITHMES DE L'ANALYSE TYPO-
LOGIQUE - LE CONGROUPE ET LE DISCRIM

Le congroupe et le discrim étant des programmes qui demandent réflexion sur les données avant de les passer à l'analyse typologique, c'est volontairement que nous ne les avons pas intégrés au système qui aurait utilisé les données automatiquement et sans discernement. Le géographe qui a l'intention d'utiliser ces programmes doit s'en servir séparément.

3.4 a) Caractéristiques et organisation du programme de groupement géographique

But du programme: Le programme groupe des observations en employant le critère de la proximité dans l'espace euclidien. La matrice d'entrée est la matrice d'information spatiale ou ce qui semble beaucoup mieux la matrice des poids locaux des facteurs, obtenue de l'analyse factorielle. A cette matrice le programme peut automatiquement imposer une matrice de contiguïté sur le groupement.

Ce programme composé originalement par Neely et Mazukelli et adapté pour le département de géographie de l'Université d'Ottawa peut prendre une matrice d'entrée de 300 observations, par 150 variables. Au départ le programme possède n groupes dont chacun contient un élément. L'algorithme

examine la matrice des distances au carré et joint les deux groupes séparés par la distance **minimum**. Chaque étape de la procédure de groupement décroît de un le nombre de groupes jusqu'à l'obtention d'un seul groupe comprenant toutes les observations.

Si l'option de contiguité est spécifiée, seules les distances entre des groupes contigus seront prises en considération.

L'algorithme de classification offre l'avantage de grouper de six façons différentes.

1. A partir de la matrice de distance, l'ordinateur additionne les sommes des carrés à chaque étape et minimise par la suite l'augmentation avec chaque addition d'un nouveau membre.

2. La deuxième façon de grouper est inverse de la première et groupe ce qui est plus éloigné et converge vers les groupes les plus proches.

3. La troisième manière est identique à la deuxième mais avec pondération par les moyennes des distances.

4. Groupe à partir de la distance minimum et pondérée à chaque étape par les moyennes des distances.

5. Groupe en se basant sur la moyenne des carrés minimum pour chaque groupe.

6. Le critère de groupement est l'attraction maximum à partir de la formule de la gravité.

$$A = \frac{M_1 \times M_2}{D^2}$$

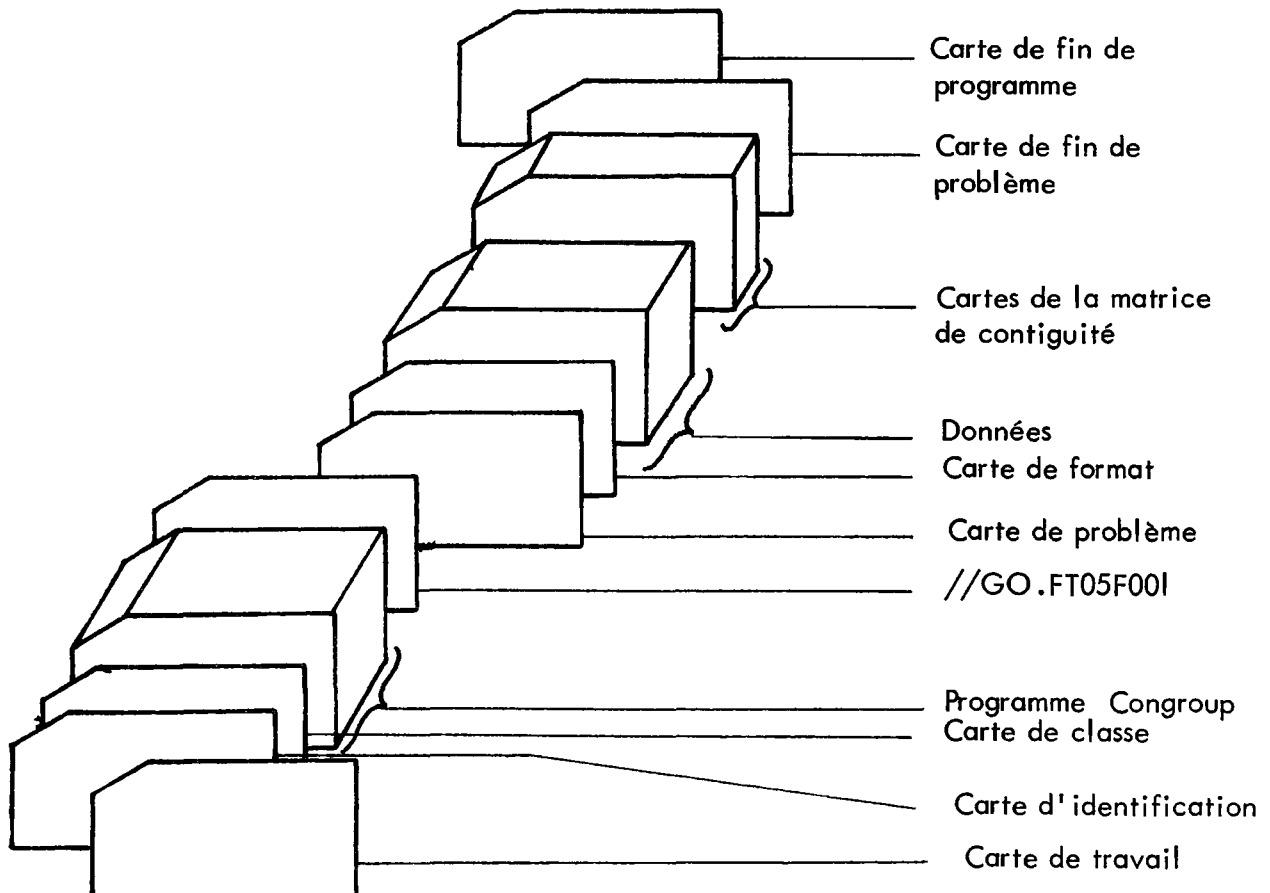
M_1 et M_2 représentent la densité de chaque membre.

Nous avons essayé les six possibilités en utilisant les facteurs scores de l'utilisation du sol de la rive sud de Montréal, et les résultats obtenus étaient identiques pour la première, quatrième et cinquième méthodes. La plupart des géographes, avec Tobler et King, emploient uniquement la première méthode. Toutefois les expériences restent à faire.

Organisation des cartes du programme (figure 37)

FIGURE 37

SCHEMA DE L'ORGANISATION DU PROGRAMME CONGROIP



Les 5 premières cartes ont été expliquées à la section 3.2. Nous allons porter notre attention sur les suivantes:

Carte de problème:

Colonnes	Contenu	Explication	Sigle	Exemple
1-4	PROB	Indique à l'ordinateur un nouveau problème	Celui de la constante	
5-6	05	Unité de lecture des cartes	NTIN	05
16-18	XXX*	Nombre de variables (maximum 150)	NVAR	004
19	1	Lire les données sur carte	INCO	1
	2	Lire les données sur bande magnétique		
20	X	Nombre de cartes du format utilisé (habituellement une)	NFMT	1
26	1	Groupe selon le minimum du carré des distances	KODE	1
	2	Réciproque à la proximité		

* Les X signifient un nombre quelconque que l'utilisateur doit perforer. Cette notice s'applique aux autres tableaux subséquents.

Colonnes	Contenu	Explication	Sigle	Exemple
	3	Proximité pondérée		
	4	Distance minimum pondérée		
	5	Moyenne des carrés minimum		
	6	Attraction maximum		
39	2	Contrainte de contiguité: lire une matrice de contiguité	NCLUS	0
	0	Ne pas la lire		

Si INCLUS:2 l'utilisateur doit fournir une matrice de contiguité de $n \times n$ observations. La diagonale de cette matrice contient des zéros. Un 1 dénote deux observations contigues et un 0 l'inverse.

La contiguité entre les observations à partir d'une matrice d'information spatiale se définit comme étant l'unité d'espace qui a un côté commun avec une autre. (cf. partie théorique du congroup au chapitre II)

Colonnes	Contenu	Explication	Sigle	Exemple
46-48		Nombre d'observations	NOOB	022
49-80		Titre de l'analyse	TITRE	Titre de l'analyse

2e carte: carte de format

Cette carte contient le format de lecture des données en incluant les parenthèses initiale et finale, ex. (4F8.3).

Il doit y avoir un nombre identique à la valeur attribuée à NFMT sur la carte de problème.

4e : les cartes de données.

Sur ces cartes on retrouve habituellement la matrice d'information spatiale ou les facteurs scores.

5e : La matrice de contiguité si NCLUS=2

Cette matrice est perforée dans le format (8011) et doit contenir autant de cartes que le nombre d'observations.

6e : Carte de fin de problème

Colonne 1-4 FINS

7e : Carte de fin de programme /* Colonnes 1 et 2.

Le lecteur trouvera un exemple à la figure (3⁸).

3.4 b) Caractéristiques et organisation du programme discrim

Nous ne reprendrons pas la structure complète du programme puisque le début est identique à celui du congroup et à l'ensemble des autres programmes. Nous n'allons expliquer que les cartes de contrôle. A la figure (3^o) on retrouve le schéma qui illustre l'organisation du programme.

Colonnes	Contenu	Explication	Sigle	Exemple
1ère carte de contrôle				
1-80	TITRE			
2e carte de contrôle				
1-2	XXX	Nombre de variables (minimum 2 et maximum 10)	NVAR	04
3		Toujours égale à 1. Permet d'écrire les résultats intermédiaires	PROPT	1
4-5		Correspond au maximum d'itérations voulues (habituellement 6)	MAXIT	06

3e : Carte de format; même procédé que pour le congroup.

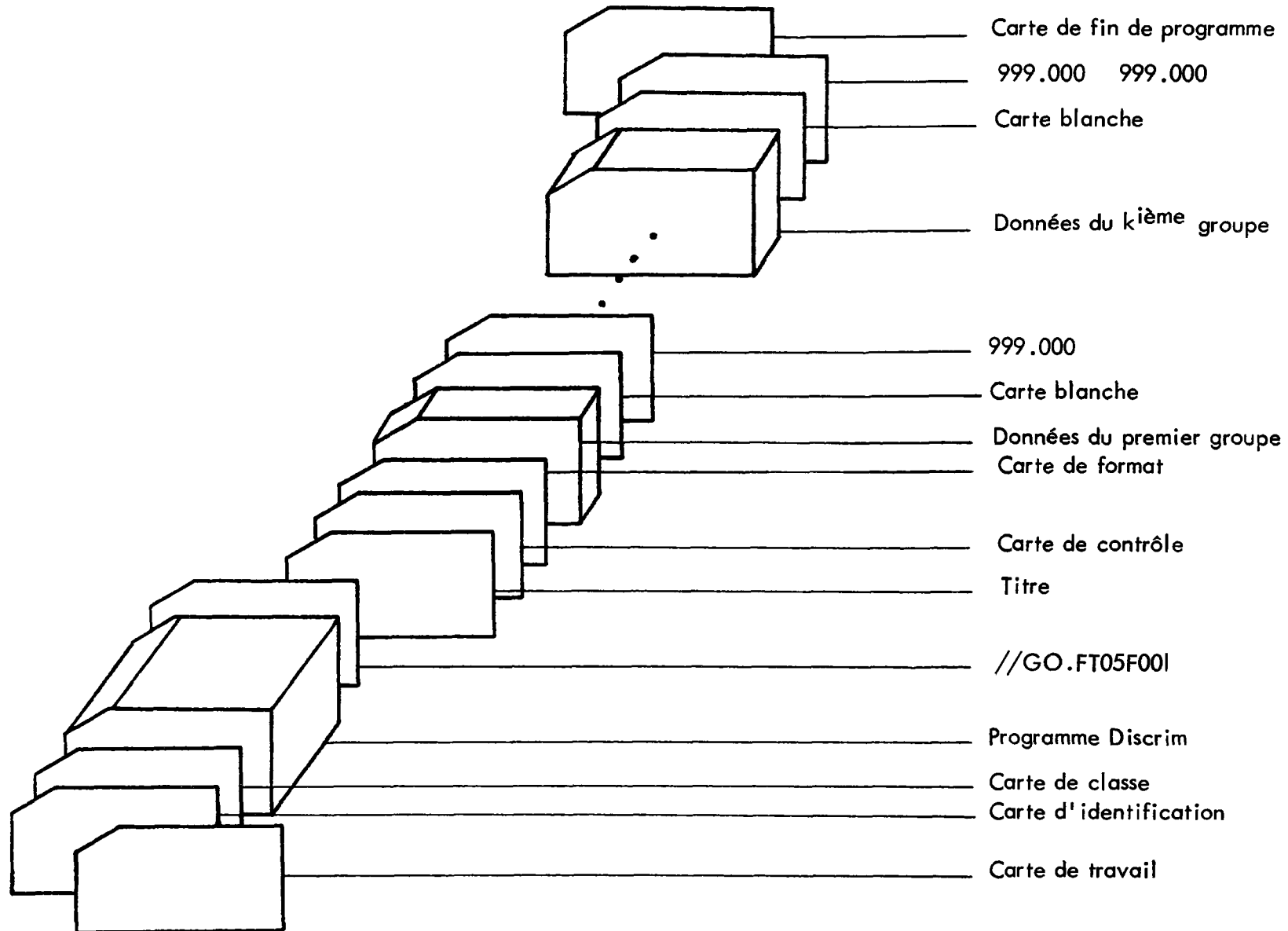
4 : Les données

Les données sont organisées de la façon suivante.

Une première carte avec le nom des observations dans les co-

FIGURE 39

SCHEMA DE L'ORGANISATION DU PROGRAMME DISCRIM



lonnes 1 à 12.

Une dernière carte avec les données conformes au format de lecture. Ceci est répété pour chaque observation d'un groupe.

A la fin d'un groupe on inclut une carte blanche suivie d'une autre sur laquelle le nombre 999,000 est perforé en concordance avec la première donnée.

A la fin complètement de tous les groupes, au lieu de perforer un seul 999 il y en a deux, correspondant aux deux premières données. Cf. figure (40)

Les données du discrim sont habituellement composées de poids locaux factoriels, quoiqu'il soit possible de mettre une matrice d'information spatiale. La matrice d'entrée maximum acceptée par le programme est de 150 observations et 10 facteurs scores ou variables groupés en 15 groupes.

La dernière carte est toujours /* dans la première et deuxième colonnes.

3.4 c) Caractéristiques et cartes de contrôle du programme mulcov

Originellement programmé par Stanley Kneiff d'Towa et légué par Bryn Greer-Wooten ce programme a une capacité de 80 observations et de 12 variables.

La figure (41) représente la structure du programme.

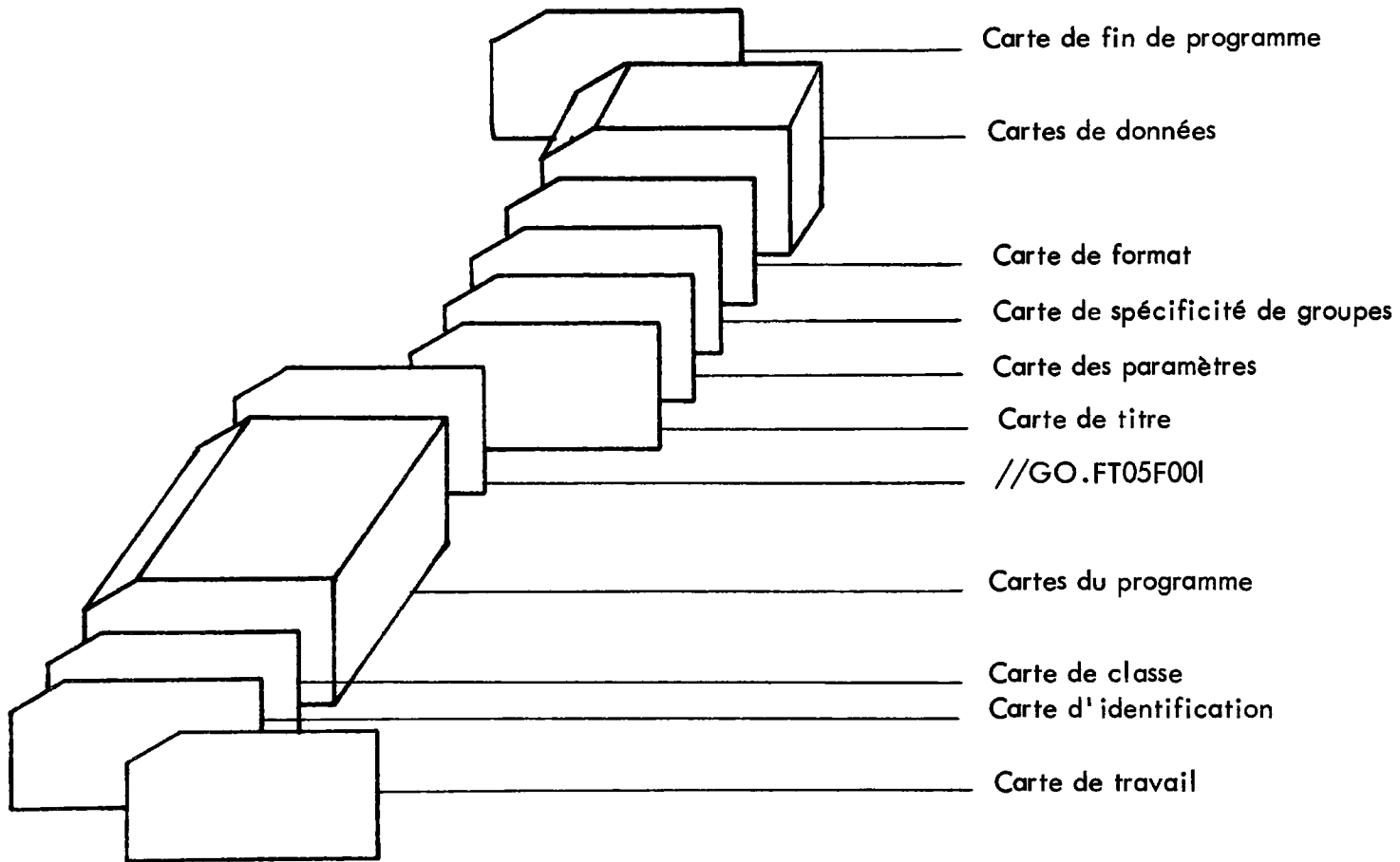
Quatre cartes de contrôle sont nécessaires à son exécution.

Colonnes	Contenu	Explication	Sigle	Exemple
1ère carte. Titre de l'analyse. 80 colonnes				
2e carte des paramètres.				
1-2	XX	Nombre de variables (maximum 12)	N	08
3-4	XX	Nombre de groupes (maximum 10)	NS	06
5-6	00	Option du test de l'homoscédasticité. L'exécution est terminée si les variables ne sont pas homoscédastiques	KKK	
	ou			
	01	L'exécution continue		01
7-8		Option pour les résidus de la régression	NREST	01
	00	Les résidus ne sont pas imprimés		
	ou			
	01	Les résidus sont imprimés		
9-10	XX	Numéro séquentiel de la variable dépendante	DEF	03

3e carte : Nombre d'observations dans chaque groupe.

FIGURE 41

SCHEMA DE L'ORGANISATION DU PROGRAMME MULCOV



Colonnes	Contenu	Explication	Sigle	Exemple
4-5	XX	Nombre d'observations du premier groupe	1GROUP(1)	20
9-10	XX	Nombre d'observations du deuxième groupe	1GROUP(2)	09
14-15	XX	Nombre d'observations du troisième groupe	1GROUP(3)	14
etc.	etc.	etc.	etc.	etc.

4e : Carte de format ex. (4F8.3)

Format de lecture des données d'entrée.

5e : Les données d'entrée

6e : Carte de fin de programme

1-2 / *

Voici un exemple à la figure (42).

3.5) TROISIEME PARTIE DU MODELE S.F.I.G.E.G.: SOUS-SYSTEME DE LA MODELISATION PARTIELLE

Les trois programmes constituant de cette section proviennent de la série des programmes S.S.P. (Scientific Subroutine Package) écrits par la compagnie I.B.M. Nous avons toutefois adapté ces programmes pour des géographes qui sont souvent obligés de faire plusieurs régressions. Avec ce système que nous avons développé plusieurs simulations linéaires

peuvent être effectuées d'un seul coup à l'aide de différentes cartes de contrôle. Figure (43).

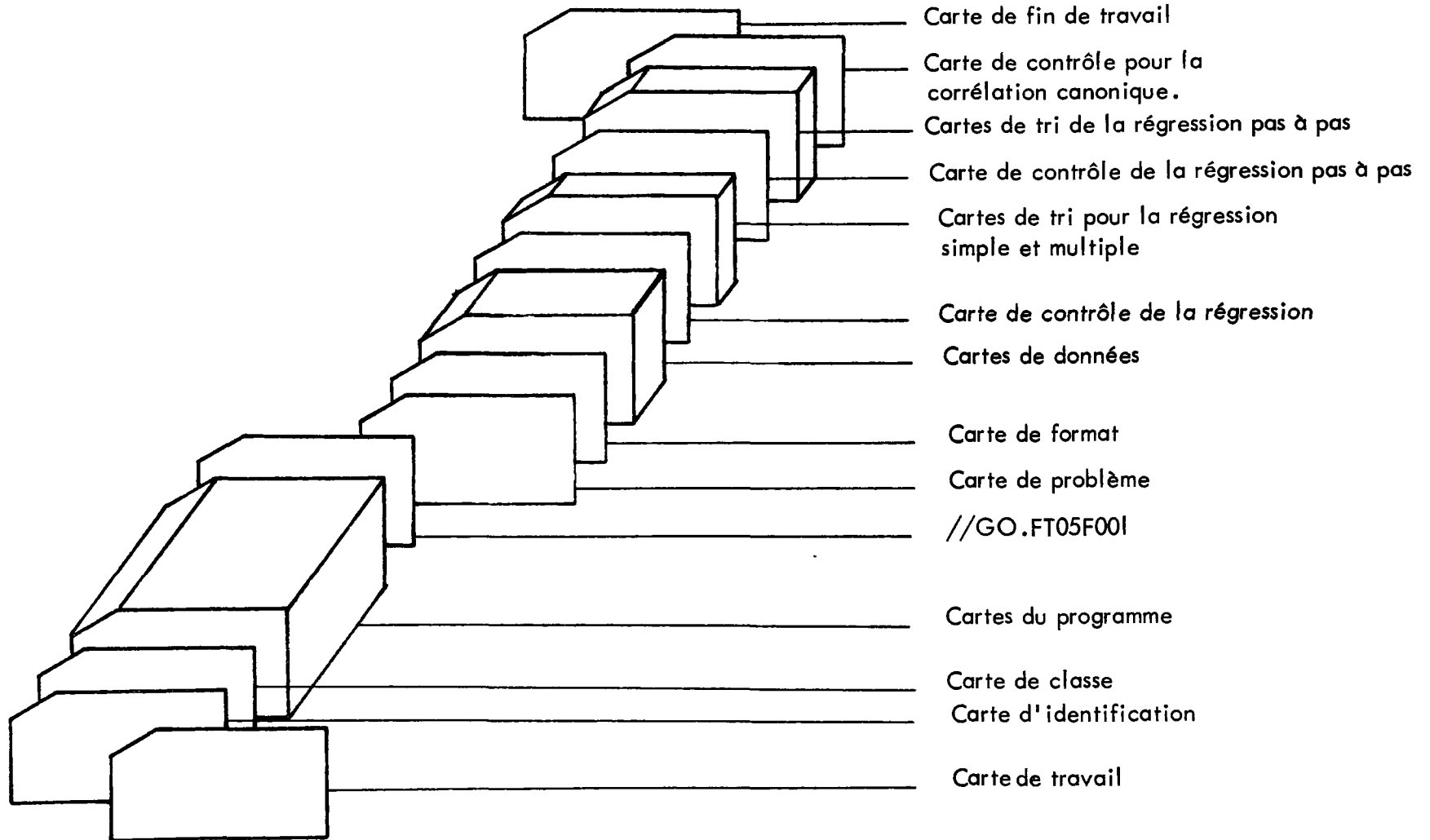
La régression simple, la régression multiple, la régression pas à pas et l'analyse canonique représentent les diverses options de cette troisième partie.

Voici la structure du programme.

Colonnes	Contenu	Explication	Sigle	Exemple
1ère : Carte de problème				
1-3	XXX	Nombre d'observations (maximum 150)	N	022
3-6	XXX	Nombre de variables	M	009
7-8	01	Pour exécuter la régression simple et multiple	IRI	01
	ou 00	Pour ne pas exécuter la régression simple ou multiple		
9-10	01	Pour exécuter la régression pas à pas	IR2	01
	ou 00	Pour ne pas exécuter la régression pas à pas.		
11-12	01	Pour exécuter la régression canonique	IR3	01
	ou 00	Pour ne pas exécuter la régression canonique		

FIGURE 13

SCHEMA DE L'ORGANISATION DES PROGRAMMES DE LA MODELISATION PARTIELLE



2e : Carte de format

Ecrire dans les colonnes 1-80 le format de lecture des données. ex. (9F8,2)

3e : Cartes de données

Les données sont perforées selon le format écrit sur la carte précédente.

4e : Cartes de contrôle obligatoire si TR1 = 01

- Carte de spécificité de la régression simple et de la régression multiple.

Colonnes	Contenu	Explication	Sigle	Exemple
1-6	XXXXXX	Nom du problème. Les caractères peuvent être alphanumériques	PR,PR1	GEOG72
7-11	XXXXX	Nombre d'observations	N	00022
12-13	XX	Nombre de variables	M	009
14-15	XX	Nombre de cartes de tri pour les régressions simple et multiple (voir ci-dessous)	NS	03

- Cartes de tri des régressions simple et multiple.

Les cartes de tri servent à spécifier la variable dépendante et l'ensemble des variables indépendantes dans une

régression linéaire multiple. N'importe quelle variable peut être choisie comme variable dépendante et de même pour les variables indépendantes. Plusieurs tris peuvent être tentés sur le même ensemble de variables.

Colonnes	Contenu	Explication	Sigle	Exemple
1-2		Code d'option pour la table des résidus	NRESI	01
	01 ou 00	Imprimer Ne pas imprimer		
3-4	XX	Numéro de la variable dépendante	NDEP	03
5-6	XX	Nombre de variables indépendantes	K	03
7-8	XX	Numéro de la première variable indépendante	ISAVE(1)	01
9-10	XX	Numéro de la deuxième variable indépendante	ISAVE(2)	02
11-12	XX	Numéro de la troisième variable indépendante	ISAVE(3)	08
etc.	etc.	etc.	etc.	etc.

Pour chaque régression voulue une carte de tri doit spécifier les paramètres nécessaires. Ainsi trois régressions nécessitent trois cartes de tri.

5e : Cartes de contrôle obligatoires si IR2 = 01

- Carte de spécificité de la régression pas à pas.

Colonnes	Contenu	Explication	Sigle	Exemple
1-6	XXXXXX	Numéro du problème	PR PRI	PROR10
7-11	XXXXX	Nombre d'observations (maximum 150)	N	00022
12-13	XX	Nombre de variables (maximum 100)	M	09
14-15	XX	Nombre de cartes de tri pour la régression pas à pas	NS	02
16-21		Cette constante est la préparation de la somme des carrés qui sera employée comme limite pour savoir si une variable doit entrer dans la régression. Si la préparation de la somme des carrés est supérieure à PCT la variable ne sera pas incluse dans le modèle. La constante doit avoir un point décimal, ex. 0.0. Le choix de cette constante dépend du but de l'analyse. Habituellement on emploie 0.01 (1%). Si toutes les variables doivent être incluses dans l'analyse employez 0.0	PCT	000.01

Colonnes	Contenu	Explication	Siège	Exemple
22		Code d'option pour la table des résidus	NRESI	01
	01	Imprimer		
	00	Ne pas imprimer		

- Cartes de tri de la régression pas à pas

Le procédé pour la régression pas à pas est différent de celui de la régression simple et multiple.

Le système employé est celui des codes.

Code 0: Variable indépendante choisie par le programmeur.

Code 1: Variable indépendante forcée à entrer dans la régression.

Code 2: Variable exclue de la régression.

Code 3: Code de la variable dépendante.

Supposons que l'on ait 12 variables et que l'on fasse une régression en prenant la variable 4 comme dépendante et les variables 3, 6, 7, 8, 9, 10 comme indépendantes. On aurait la carte de tri suivante:

Colonnes	Contenu
1	2
2	2
3	0

Colonnes	Contenu
4	3
5	2
6	0
7	0
8	0
9	0

Si on considère la variable 9 comme la plus importante dans le modèle la carte de tri sera la suivante:

Colonnes	Contenu
1	2
2	2
3	0
4	3
5	2
6	0
7	0
8	0
9	1

Une carte de tri par régression est obligatoire.

6e: Cartes de contrôle obligatoires si $IR3 = 1$

Nous rappelons que l'analyse canonique cherche la corrélation maximale entre deux ensembles de variables w et z . Le nombre de variables dans le premier ensemble doit être

égal ou plus grand que le nombre de variables dans le second ensemble.

Les limites sont de 150 observations et 20 variables ou poids locaux factoriels.

Carte de spécificité de l'analyse canonique.

Colonnes	Contenu	Explication	Sigle de la constante	Exemple
1-6	XXXXXX	Numéro du problème ou titre	PR,PRI	RTVSUD
7-11	XXXXX	Nombre d'observations (maximum 150)	N	00022
12-13	XX	Nombre de variables dans le premier ensemble	MP	04
14-15	XY	Nombre de variables dans le second ensemble	MQ	03

Nous rappelons que si l'une des constantes suivantes IR1, TR2, IR3 est égale à 0 il ne faut absolument pas inclure les cartes de contrôle respectives. A titre d'exemple si on ne désire pas la régression pas à pas TR2 = 0 alors on n'inclut pas la carte de spécificité de la régression pas à pas ni les cartes de tri de ce genre de régression.

A la fin du programme ne pas oublier d'inclure la carte de fin du programme /*. La figure (44) illustre les

cartes de contrôle que nous venons d'expliquer.

Ceci termine l'aspect proprement technique et opérationnel du modèle S.E.I.G.E.G. Il reste cependant que l'élaboration d'un tel système ne se fait pas sans comporter des problèmes d'intégration et d'utilisation. En fait, le principal problème que nous avons rencontré en édifiant ce système a été celui de développer une méthode d'utilisation accessible à un public géographe (ou en voie de le devenir) qui n'a pratiquement pas d'expérience en informatique. Ceci nécessite donc de prévoir toutes les possibilités d'utilisation tout en laissant une certaine flexibilité. De plus, le fait de vouloir satisfaire à des exigences pédagogiques et scientifiques oblige à certains compromis entre la structure interne du système et la disposition des résultats. La capacité de mémoire de l'ordinateur ne pouvait accepter simultanément le grand nombre de programmes (13) et de sous-programmes (57) **et** nous a obligé à réorganiser la structure interne de notre système pour permettre à l'ordinateur de compiler chaque programme au moment précis où il s'en sert et de l'effacer par la suite. Nous ne décrivons pas la méthode utilisée parce qu'elle n'est pas de connotation géographique et qu'elle est relativement compliquée. Nous avons rencontré le problème de la conservation des données intermédiaires entre les différents programmes. A cet effet nous avons dû utiliser 5 unités de disques qui jouent le rôle de banque de données tempo-

raires. Et finalement nous avons essayé de minimiser le temps-machine et le nombre de mémoires nécessaires pour exécuter un problème.

3.6) CONCLUSION

Avec les programmes du S.E.I.G.E.G. que nous avons expliqués, le géographe découvre des structures et des typologies; cependant aucune technique ne lui permet de connaître l'évolution des structures et des typologies obtenues puisque tout cela est purement statique. La seule façon d'étudier le dynamisme de l'espace consiste dans le recours à l'analyse stochastique de la matrice chrono-spatiale et à l'analyse spectrale. Nous nous proposons d'utiliser ces méthodes sur un problème géographique en France au cours de nos études de doctorat. Quoique pleinement conscient que ce système est loin d'être parfait et qu'il y a place pour de nombreuses améliorations nous croyons qu'il sera utile autant aux débutants qu'à ceux qui sont familiarisés avec l'analyse multivariée en géographie.

Cependant nous n'avons pas "vidé" la question et plusieurs problèmes restent en suspens tels celui de l'auto-corrélation, celui d'une modélisation globale, celui de la recherche d'une matrice typologique déduite, celui de la correspondance typologique entre les types de combinaison. De plus nous croyons que l'analyse sociométrique des questionnaires demeure l'un des problèmes sur lesquels il faudra se pencher tôt ou tard. Il nous était évidemment impossible d'aborder tous ces problèmes dans le cadre d'une maîtrise. Il reste que le progrès de la science géographique réclame que quelqu'un s'y

attache de toute urgence et pour les raisons pédagogiques que nous avons déjà exposées, fasse dans un cadre académique ou dans un autre une présentation d'ensemble sensiblement analogue à la nôtre.

CONCLUSION

Au terme de ce travail un bilan s'impose. Si nous avons présenté ces techniques, c'est qu'elles étaient nécessaires à l'élaboration d'une méthode qui puisse saisir simultanément aussi bien les limites spatiales que les attributs qui les caractérisent. Si le recours à l'informatique s'impose, c'est qu'il répond à des besoins et sous-tend une réalité auxquels le géographe aura tôt ou tard à faire face. En effet la multitude des données, les nombreux rapports et l'information sans cesse grandissante font que l'ordinateur n'est plus un luxe comme le prétendent plusieurs, mais une nécessité. C'est en réponse à cette nécessité que nous avons développé ce système informatique. Si on ramène le problème sur le plan de la logique ce système repose sur une logique technique organisée en fonction des besoins opérationnels du géographe car c'est à celui-ci, en fonction des critères géographiques, qu'il revient de décider lequel parmi tous ces outils est le plus conforme à la nature du problème qu'il étudie et lesquels, parmi les différents résultats obtenus à partir de divers algorithmes, sont les plus représentatifs de la situation géographique analysée. (Racine, 1972).

De plus ce travail avait pour but d'assurer une transition entre une géographie descriptive et une géographie quantitative. Nous insistons ici sur le terme "transition" car il ne s'agit pas de remplacer l'une par l'autre. De plus nous avons tenté d'assurer cette transition socialement en rendant disponible une série de méthodes et de techniques à une population non initiée, en particulier l'école française et en démontrant à l'école américaine et anglo-saxon-

ne qu'il pouvait exister des liens logiques entre les différentes méthodes, ce qu'elles avaient souvent négligé de faire jusqu'à présent.

Nous avons tenté d'incorporer cette transition dans un circuit analytique qui soit à la fois géographique, pédagogique et heuristique. Géographique en ce sens qu'il est orienté en fonction de l'espace et que l'ensemble du circuit converge vers la représentation cartographique des structures déterminées, par sa capacité à fournir des éléments d'explication pour divers phénomènes dont la base demeure principalement spatiale et finalement par la possibilité de saisir simultanément plusieurs attributs de l'environnement.

Le circuit analytique est aussi pédagogique car il renferme une méthodologie qui, nous en sommes certains, va servir aux étudiants. La simplicité du système informatique fournit les moyens à l'étudiant d'expérimenter lui-même sur l'espace qu'il choisit, des théories ou des modèles. (En somme le S.F.I.G.E.G. devient un laboratoire de géographie quantitative mais peut tout aussi bien servir aux chercheurs qui ont un problème précis). Le fait de délimiter différents niveaux tels que celui de la préparation des données, celui de la réduction des attributs, celui des observations facilite l'accessibilité de ces méthodes.

Finalement ce circuit analytique est aussi heuristique car il offre au chercheur un pouvoir de traitement de son information en concordance avec la démarche propre à la géographie quantitative. Avec ce système il est en mesure de découvrir des éléments

nouveaux qui feront progresser la connaissance géographique jusqu'au point où elle réussira à s'implanter comme science.

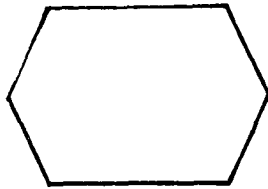
Il va sans dire qu'un tel circuit s'encadre entre les deux extrêmes de la logique géographique: la logique de la mise en ordre des variables et la logique de l'estimation des variables. La logique existant entre les variables se réfère au fait que chaque colonne de la matrice d'information spatiale peut être représentée par une carte et si la superposition des cartes laisse entrevoir des relations on peut en conclure qu'il existe un lien entre les variables et ce lien deviendra logique par la suite lorsqu'il contribuera à définir les structures internes du système. L'autre extrême fait appel à la logique qui pousse à choisir une variable plutôt qu'une autre. Le problème du choix des variables n'est pas simple à résoudre et il est impératif que toute analyse quantitative subisse au préalable une réflexion théorique sur le choix des variables à utiliser et en ce sens il n'existe aucun travail purement quantitatif.

Enfin une dernière question: qu'est-ce que le S.E.T.G.E.G. nous offre de plus? Trois éléments de réponse se détachent. Premièrement qu'il existe des méthodes complémentaires à la géographie qualitative susceptibles de renforcer et d'éclairer des faits qui étaient mal définis à la suite d'études descriptives. Deuxièmement il nous a appris que ces méthodes pouvaient être montées en batterie pour former un système opérationnel correspondant à une démarche logique de la pensée géographique. Troisièmement il montre que la technique est encadrée dans la géographie et qu'il n'y a point de

travail vraiment géographique sans réflexion théorique au préalable.

Le S.E.I.G.F.G. est beaucoup plus qu'un outil; il sous-tend une forme de pensée géographique et contribue de façon incommensurable à ouvrir des perspectives vers les processus stochastiques et la théorie générale des systèmes.

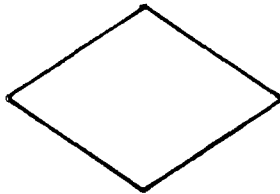
APPENDICE A



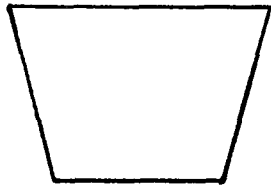
BOUCLE



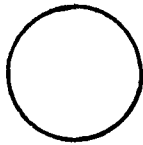
CALCUL



DECISION



ENTREE



RACCORD



SORTIE



DEBUT ET FIN

APPENDICE B

NOM DES SOUS-PROGRAMMES	FONCTION	SOURCE
Programme principal du S.E.I.G.E.G.		
MOUT	Tableaux de sortie	Guy Lemay
TRANS	Transposition de matrice	S.S.P.(Scientific Subroutine Package)
JTRANS	Transposition des tableaux de sortie	Guy Lemay
TRANST	Génération des données	Guy Lemay
EXTRAC	Décodage de la variable à générer	Guy Lemay et André Leclerc
GETNUM	Identification des constantes	Idem
VARCHK	Vérification des variables	Idem
MESSGE	Message et erreur	Idem
DIMAX	Sécurité de la limite du système	Idem
TSTNUM	Vérification numérique	Idem
BLK-DATA	Mémoire temporaire commune	Idem
TYPO	Prépare les données pour la typologie	Guy Lemay
NORM	Prépare les données pour la normalisation	Guy Lemay
Normalisation		
RPRINT	Imprime une matrice	Greer-Wootten S.S.P.

NOM DES SOUS-PROGRAMMES	FONCTION	SOURCE
Corrélation		
DATA	Garde les données en mémoire	S.S.P.
CORRE	Exécute la corrélation	S.S.P.
ORDRE	Mise en ordre décroissant des coefficients de corrélation	Guy Lemay
CORCOM	Corrélation avec test de T	Greer-Wootten
Linka		
PHI	Calcul du coefficient de corrélation PHI	
LINANA	Exécute le "linkage analysis"	
FACTER	Calcule les facteurs	
SORTR	Met en ordre décroissant les facteurs	
EVALUE	Calcule les valeurs propres	U.B. Manley
Facto		
A	Programme principal de l'analyse en composantes principales	Guy Lemay
INNORM	Standardisation	Greer-Wootten
EIGEN	Calcul des valeurs propres	S.S.P.
TRACE	Pourcentage cumulatif des valeurs propres	S.S.P.

NOM DES SOUS-PROGRAMMES	FONCTION	SOURCE
LOAD	Calcul des saturations	S.S.P.
B	Programme principal de Falcomp et de l'analyse image	Chuck Crawford
INVERT	Inversion de matrice	S.S.P.
C	Rotation des facteurs	Chuck Crawford et Guy Lemay
VARMAX	Rotation Varimax	S.S.P.
GRAPH	Graphique des facteurs	Guy Lemay
OBLI	Préparation des facteurs pour la rotation oblique	Greer-Wootten
ROTAT 3	Rotation oblique des facteurs	Idem
ROOT	Calcul des racines des facteurs rotationnés	Idem
Programme des facteurs scores	Calcul des poids locaux	Chuck Crawford
ORS	Mise en ordre des facteurs scores	Guy Lemay
Congroup	Algorithme de classification	Neely et Mazukelli
ERROR	Détection des erreurs	Idem
CNTGTY	Préparation de la matrice de contiguité	Idem
GRPING	Groupeement des observations	Idem
SLH	Calcul des distances	Idem

NOM DES SOUS-PROGRAMMES	FONCTION	SOURCE
SRH	Calcul des distances	Idem
UPKMAT	Matrice mise à date	Idem
XIPHF	Fonction de calcul	Idem
XILHF	Fonction de calcul	Idem
Discrim	Analyse discriminante	Cassetti
DIRNM	Préparation des matrices de distance	Idem
HDTAG	Calcul des valeurs des diagonales (valeurs propres)	Idem
Mulcov	Multicovariance	Greer-Wootten
MINV	Inversion de matrice	S.S.P.
MULTR	Régression multiple	S.S.P.
Modélisation partielle	Prépare les données pour les régressions	Guy Lemay
R1	Régression simple	S.S.P.
R2	Régression pas à pas	S.S.P.
MSTR	Conversion des données et mise en mémoire	S.S.P.
LOC	Location comprimée des matrices en mémoire	S.S.P.
STPRG	Programme principal de la régression multiple	S.S.P.
STOUT	Programme de sortie pour la régression pas à pas	S.S.P.

NOM DES SOUS-PROGRAMMES	FONCTION	SOURCE
CANOR	Calcul de la corrélation canonique	S.S.P.
NROOT	Calcul des valeurs propres et des vecteurs propres (eigenvalues, eigenvectors)	S.S.P.

APPENDICE C

```

//STEP1 EXEC FORTGCLG,REGIGN.GO=208K,TIME.GO=2
//FORT.SYSIN DD *
COMMON/GU/NUBS(150,20),NATR(100,20),DD(150),      FMT(100),ITITR(20)
CCMCN ENME(150,100)
CCMCN A(2000),T(2000)
DIMENSION DON(150,100)
DIMENSION IND(2),KODE(16)
EQUIVALENCE(ENME,DON)
DATA KODE/'A-GE','B-TR','C-NO','D-CO','E-LI','F-FA','G-FA','H-IM',
+ 'I-SI','J-RE','K-CA','L-CC','M-DI','9999','FINI','B-TY'/
DATA BLANC/'  '/
INTEGER R/1/,W/3/,P/2/
INTEGER ERRFL1

```

C
C
C

```

WRITE(W,1)
1 FORMAT('1',44X,41('*')/45X,41('*')/45X,2('*'),37X,2('*')/45X,2('*')
+),37X,2('*')/
+ 45X,2('*'),1X,'INFORMATIQUE ET ESPACE GEOGRAPHIQUE',1X,2('*')/45
+X,2('*'),37X,2('*')/45X,2('*'),8X,'LE MODELE S.E.I.G.E.G.',7X,2('*')
+ )/45X,2('*'),37X,2('*')/45X,2('*')
+ ,10X,'THESE DE MAITRISE',10X,2('*')/45X,2('*'),37X,2('*')/45X,2(
+ '*'),12X,'PAR GUY LEMAY',12X,2('*')/45X,2('*'),37X,2('*')/45X,2('*')
+ ),6X,'DEPARTEMENT DE GEOGRAPHIE',6X,2('*')/45X,2('*'),37X,2('*')/
+ 45X,2('*'),9X,'UNIVERSITE D''OTTAWA',9X,2('*')/45X,2('*'),37X,2('*')
+ )/45X,2('*'),16X,'1972',17X,2('*')/45X,2('*'),37X,2('*')/2(45X,41
+ ('*')/),10(/)

```

C
C
C

```

READ(R,2) NOB,NVAR,IPCV,ITIT,I TRANS,INOR
2 FCRMAT(2I3,74I1)
KK=20
IF(ITIT.LT.1) GO TO 4
READ(R,3) (ITITR(JJ),JJ=1,20)
3 FORMAT(20A4)
WRITE(W,9)
9 FCRMAT(51X,'T I T R E   D E   L '' A N A L Y S E ',//)
WRITE(W,5) (ITITR(JJ),JJ=1,20)
5 FORMAT(2(27X,8C('I')/),27X,20A4/2(27X,8C('I')/),5(/)
4 WRITE(W,6) NOB,NVAR
6 FORMAT(131('X')//5X,'I L Y A ',I3,' OBSERVATIONS ET ',I3,' VARIABLE
+S DANS CETTE ANALYSE'//131('X')//)

```

C
C
C

```

N=NOB
M=NVAR
IF(IPCV.LT.1) GO TO 20
WRITE(W,10)
10 FCRMAT(5X,'LISTE DES VARIABLES:'//5X,20('- '))
DO 7 I=1,M
READ(R,8) (NATR(I,J),J=1,KK)
8 FCRMAT(20A1)
7 WRITE(W,11) (NATR(I,J),J=1,KK)
11 FCRMAT(/26X,'I',3X,20A1,3X,'I')
READ(R,12) (FMT(JJ),JJ=1,100)
12 FORMAT(4(20A4/),20A4)
WRITE(W,13)

```

```

13 FORMAT(5(/),5X,'LISTE DES OBSERVATIONS: '/5X,20(' - '))
C
C
C
DO 14 I=1,150
DC 14 J=1,100
ENME(I,J)=0.0
DON(I,J)=0.0
14 CCNTINUE
DO 15 I=1,N
READ(R,FMT) (NCBS(I,K),K=1,KK),(DD(J),J=1,M)
WRITE(W,16) (NCBS(I,K),K=1,KK),(DD(J),J=1,M)
C
16 FORMAT(1X,20A1/,5F8.2/5F8.2/)
DC 15 J=1,M
ENME(I,J)=DD(J)
DON(I,J)=DD(J)
15 CONTINUE
WRITE(W,100) ((ENME(I,J),J=1,M),I=1,N)
100 FORMAT(6F8.2)
DC 18 I=1,N
18 WRITE(W,11) (NCBS(I,J),J=1,KK)
GO TO 30
C
C
C
20 DO 21 I=1,M
DC 21 J=1,20
21 NATR(I,J)=BLANC
DO 22 I=1,N
DC 22 J=1,20
22 NCBS(I,J)=BLANC
DO 26 I=1,N
READ(R,FMT) (DD(J),J=1,M)
DO 26 J=1,M
ENME(I,J)=DD(J)
26 CCNTINUE
C
C
C
30 WRITE(W,31)
31 FORMAT('1',51X,'MATRICE D'INFORMATION SPATIALE'/52X,3J('=')//)
CALL MOUT(N,M,KK)
C
C
C
C
270 IGE=C
ITR=C
INC=0
ICC=0
ILI=0
INCT=0
INCS=C
ITY=0
ICOMP=0
WRITE(8) KK,N,M,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+I=1,N),((ENME(I,J),J=1,M),I=1,N)
C
89 READ(R,90) IND
WRITE(W,93) INC

```

```
93 FORMAT(1X,15('*')/1X,A4,A2/1X,15('*'))
90 FORMAT(A4,A2)
  DC 91 I=1,16
  IF(IND(1).EQ.KCODE(I)) GO TO 95
91 CONTINUE
```

C
C
C

```
  WRITE(W,92) INC
92 FCRMAT(A4,A2,5X,'EST UN ORDRE INVALIDE')
  GO TO 89
```

C
C

```
95 GO TO (110,120,130,140,150,160,170,180,190,200,210,220,230,240,
+250,260), I
```

C
C
C

```
110 CALL TRANSF(DCN,N,M,ERRFL1, IDIMAX, IPOV, IGE)
  WRITE(W,111)((DCN(I,J),J=1,M),I=1,N)
111 FORMAT(10F8.3)
  REWIND 8
  DO 112 I=1,N
  DO 112 J=1,M
  ENME(I,J)=DCN(I,J)
112 CONTINUE
  WRITE(8) KK,N,M,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+I=1,N),((ENME(I,J),J=1,M),I=1,N)
  IF(ERRFL1.NE.1.AND.IDIMAX.NE.1) GO TO 89
  WRITE(3,61)
61 FCRMAT(1X,'EXECUTION TERMINEE DU A LA DETECTION D'ERREURS AU NIVE
+AU DU MODULE DE LA TRANSFORMATION')
  GO TO 89
```

C
C
C

```
120 CALL JTRANS(N,M, ITR)
  READ(R,121) IPTR
121 FORMAT(I1)
  IF(IPTR.NE.1) GO TO 89
  DC 124 I=1,N
  WRITE(2,125) (ENME(I,J),J=1,M)
125 FCRMAT(9F8.3)
124 CONTINUE
  GO TO 89
```

C
C
C

```
130 CALL NORM(N,M, INO, ITY, INOT, INOS)
  GO TO 89
```

C
C
C
C

```
140 ICO =1
  GO TO 89
150 ILI=1
  GO TO 89
160 ICOMP=1
  READ(R,161) IFACTC, IRCT, CGN, NPUF, NPIR, NPUEV, NP, NORI, IBUS
```

```

161 FORMAT(2I2,F6.1,8I2)
    WRITE(12) (ITITR(JJ),JJ=1,20),IFACTO,IROT,CON,NPUF,NPIR,NPUEV,
+NP,NORI,IBUG,ICOMP
    GO TO 89
170 CONTINUE
180 CONTINUE
190 CONTINUE
200 CONTINUE
210 CONTINUE
220 CONTINUE
230 CONTINUE
240 CONTINUE
    GO TO 89

```

```

C
260 CALL TYPO (N,M,ITR,ITY)
    GO TO 89
250 WRITE(8) IGE,ITR,ITY,INC,INOT,INOS,ICO,ILI,ICOMP
    RETURN
    END
    SLBRCUTINE MOLT(N,M,KK)
    COMMON/GU/NOBS(150,20),NATR(100,20),DD(150),    FMT(100),ITITR(20)
    COMMON ENME(150,100)
    COMMON A(2000),T(2000)
    DIMENSION NNATR(20,100)
    DATA EQU,LIT,GE,RV,IG,VA,AB/'=', 'I', 'OBSE', 'RVAT', 'IONS', 'VARI',
+ 'ABLE' /
    INTEGER R/1/,W/3/,P/2/
    INTEGER DIGIT(6)/'1','2','3','4','5','6'/
    INTEGER F1(8)/'(1X','29A1',' ',' ',' ',' ','(A1','15A1',' '),2A',
+ '1) '/
    INTEGER F2(9)/'(1X','2A1','26X','A1',' ',' ','(A1','15X)',
+ ',2A1',') '/
    INTEGER F3(14)/'(1X','2A1','7X,3','A4,7','X,A1',' ',' ',' ',
+ '(A1','1X,2','A4,1','X',' ','5X)',',2A1',') '/
    WRITE(W,1) (EQL,I=1,131)
1  FCRMAT(1X,131A1)
    L=0
    DO 2 I=1,M
    DO 2 J=1,20
    L=L+1
2  A(L)=NATR(I,J)
    CALL TRANS (A,T,20,M)
    L=0
    DO 3 I=1,20
    DO 3 J=1,M
    L=L+1
3  NNATR(I,J)=T(L)
    J1=1
    NMO=C
7  J2=J1+5
    NMC=M-J2
    IF(NMO) 5,10,15
5  NMOO=NMO+6
    NC=NMC
    J2=J1+(NO-1)
    N1=2
    GO TO 20
10 NO=C
    N1=2
    GO TO 20

```

```

15 NO=6
   N1=1
20 F1(4)=DIGIT(NO)
   F2(5)=DIGIT(NO)
   F3(7)=DIGIT(NC)
   WRITE(W,F1) LIT,LIT,(EQU,J=1,26),LIT,(LIT,(EQU,J=1,15),K=1,NO),
+ (LIT,J=1,N1)
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
   DO 30 I=1,20
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
30 WRITE(W,18) (NATR(I,J),J=J1,J2)
18 FORMAT(1H+,30X,6(7X,A1,8X))
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
   WRITE(W,F1) LIT,LIT,(EQU,J=1,26),LIT,(LIT,(EQU,J=1,15),K=1,NO),
+ (LIT,J=1,N1)
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
   WRITE(W,F3) LIT,LIT,OB,RV,IC,LIT,(LIT,VA,AB,J=1,NO),(LIT,J=1,N1)
   WRITE(W,19) (J,J=J1,J2)
19 FORMAT(1H+,30X,6(11X,I3,2X))
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
   WRITE(W,F1) LIT,LIT,(EQU,J=1,26),LIT,(LIT,(EQU,J=1,15),K=1,NO),
+ (LIT,J=1,N1)
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
   DC 21 I=1,N
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
21 WRITE(W,22) (NCBS(I,J),J=1,KK),I,(ENME(I,JJ),JJ=J1,J2)
22 FORMAT(1H+,3X,20A1,1X,I3,3X,6(2X,F8.3,6X))
   WRITE(W,F2) LIT,LIT,LIT,(LIT,J=1,NO),(LIT,J=1,N1)
   WRITE(W,F1) LIT,LIT,(EQU,J=1,26),LIT,(LIT,(EQU,J=1,15),K=1,NO),
+ (LIT,J=1,N1)
   IF(NMO) 100,25,25
25 J1=J2+1
   IF(J1.GT.M)GOTC 100
   WRITE(W,35)
35 FORMAT('1',5(/))
   GC TG 7
100 WRITE(W,1) (EQU,I=1,131)
   RETURN
   END
   SUBROUTINE TRANS(A,T,N,M)
   DIMENSION A(1),T(1)
   IR=0
   DO 10 I=1,N
   IJ=I-N
   DO 10 J=1,M
   IJ=IJ+N
   IR=IR+1
10 T(IR)=A(IJ)
   RETURN
   END
   SUBROUTINE JTRANS(N,M,ITR)
   COMMON/GU/NOBS(150,20),NATR(100,20),DD(150), FMT(100),ITITR(20)
   COMMON ENME(150,100)
   COMMON A(2000),T(2000)
   DIMENSION INTER(30,30)
   INTEGER R/1/,W/3/,P/2/
   KK=20
   ITR=1
   DO 04 I=1,M
   DO 64 J=1,20

```

```

64 INTER(I,J)=NATR(I,J)
   DO 65 I=1,N
   DO 65 J=1,20
65 NATR(I,J)=NOBS(I,J)
   DO 66 I=1,M
   DO 66 J=1,20
66 NCBS(I,J)=INTER(I,J)
   L=0
   DO 61 I=1,N
   DO 61 J=1,M
   L=L+1
   A(L)=ENME(I,J)
61 CONTINUE
   CALL TRANS(A,T,M,N)
   L=0
   DO 62 I=1,M
   DO 62 J=1,N
   L=L+1
   ENME(I,J)=T(L)
62 CONTINUE
   WRITE(W,103) ((ENME(I,J),J=1,N),I=1,M)
103 FCRMAT(10F8.2)
   WRITE(W,63)
63 FCRMAT('1',56X,'MATRICE TRANSPONSEE'/57X,18('=')//)
   CALL MOUT(M,N,KK)
   ITEMP=N
   N=M
   M=ITEMP
   RETURN
   END
SUBROUTINE TRANS(VARI,MO,NV,ERRFL1,IDIMAX,ITIT,IGE)
IMPLICIT INTEGER*4(A-Y)
REAL*4 VARI(150,100)
C LA SCUS -ROUTINE A TROIS ARGUMENTS
C VAR EST LA MATRICE SUR LAQUELLE ON EFFECTUE LA TRANSFORMATION
C NV EST LE NOMBRE DE VARIABLES
C MO EST LE NOMBRE D'OBSERVATIONS
C TOUTES LES VARIABLES SONT ENTIERES EXCEPTE CELLES DEBUTANT PAR UN Z
C CETTE CARTE DOIT ETRE INCLUSE DANS TOUTES LES SOUSROUTINES
C LA SCSRROUTINE TRANSFCR UTILISE LES SOUSROUTINE SUIVANTES
C 1) VARCHK
C 2) GETNUM
C 3) EXTRACT
COMMON/GU/NOBS(150,20),NATR(100,20),DD(150), FMT(100),ITITR(20)
COMMON/DIG/DIGIT(10),DCT
COMMON/ALPHA/A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y
COMMON/CHAR/BLANK,COMMA,SLASH,LBRAK,RBRAK
DIMENSION CARD(80)
DIMENSION NATRF(20)
C TOUTES LES CARTES DE DIMENSION(EXCEPTE VARI),DE DATA ET DE COMMON
C DOIVENT ETRE INCLUSES DANS TOUTES LES SOUS ROUTINES
C ICI DEBUTE LE PROGRAMME
WRITE(3,1205)
1205 FORMAT('1',51X,'DIAGNOSTIQUE DE LA TRANSFORMATION'/52X,33('=')//)
   IVAR1=NV
   IVAR2=NV
   ERRFL1=0
   IGE=1
555 IF(ERRFLG.EQ.1) ERRFL1=1
   ERRFLG=0

```

```

ARG3=0
KDIV=C
JLOG=C
KLCC=0
IF(ITIT.LT.1) GO TO 11
READ(1,5) (CARD(II),II=1,40),(NATRF(II),II=1,20)
5  FORMAT(80A1)
WRITE(3,6) (CARD(II),II=1,40),(NATRF(II),II=1,20)
6  FORMAT(1H1,'CARD=',80A1)
GO TO 12
11 READ(1,7) (CARD(II),II=1,40)
7  FORMAT(40A1)
WRITE(3,8) (CARD(II),II=1,40)
8  FORMAT(1H1,'CARTE=',40A1)
C ICI ON SAUTE LES BLANCS
12 DC 10 II=1,40
P=II
IF(CARD(II).NE.BLANK) GO TO 20
10 CONTINUE
CALL MESSGE(1,P)
GO TO 555
C A CHAQUE ERREUR ON LIT UNE AUTRE CARTE ET ON CONTINUE DE VERIFIER
20 IOP=0
C ICI ON VERIFIE QUELLE TRANSFORMATION ON VEUT FAIRE
IF(CARD(P).EQ.F.AND.CARD(P+1).EQ.I.AND.CARD(P+2).EQ.N) GO TO 1201
IF(CARD(P).EQ.A.AND.CARD(P+1).EQ.D.AND.CARD(P+2).EQ.D) IOP=1
IF(CARD(P).EQ.S.AND.CARD(P+1).EQ.O.AND.CARD(P+2).EQ.U) IOP=2
IF(CARD(P).EQ.M.AND.CARD(P+1).EQ.U.AND.CARD(P+2).EQ.L) IOP=3
IF(CARD(P).EQ.C.AND.CARD(P+1).EQ.I.AND.CARD(P+2).EQ.V) IOP=4
IF(CARD(P).EQ.LP.AND.CARD(P+1).EQ.U.AND.CARD(P+2).EQ.I) IOP=5
IF(CARD(P).EQ.S.AND.CARD(P+1).EQ.Q.AND.CARD(P+2).EQ.R) IOP=6
IF(CARD(P).EQ.E.AND.CARD(P+1).EQ.X.AND.CARD(P+2).EQ.LP) IOP=7
IF(CARD(P).EQ.L.AND.CARD(P+1).EQ.O.AND.CARD(P+2).EQ.G) IOP=8
IF(CARD(P).EQ.L.AND.CARD(P+1).EQ.C.AND.CARD(P+2).EQ.O) IOP=9
IF(CARD(P).EQ.A.AND.CARD(P+1).EQ.B.AND.CARD(P+2).EQ.S) IOP=10
IF(CARD(P).EQ.LP.AND.CARD(P+1).EQ.E.AND.CARD(P+2).EQ.R) IOP=11
WRITE(3,1000) IOP,CARD(P),CARD(P+1),CARD(P+2)
1000 FORMAT(1H0,I3,3A1)
IF(IOP.NE.0) GO TO 25
CALL MESSGE(2,P)
CALL MESSGE(9,F)
ERRFLG=1
GO TO 555
C ICI ON SAIT QUE LE CODE D'OPERATION EST CORRECT, DONC ON AUGMENTE P DE 3
25 P=P+3
IF(CARD(P).EQ.LBRAK) GO TO 30
CALL MESSGE(3,P)
CALL MESSGE(9,P)
ERRFLG=1
GO TO 555
30 P=P+1
C ICI ON VEUT RECONNAITRE LA PREMIERE VARIABLE
C ON APPELE LA SOUS ROUTINE VARCHK
CALL VARCHK(CARD,P,ERR,VAR1)
WRITE(3,1001) CARD(P),P,ERR,VAR1
IF(VAR1.GT.IVAR1) IVAR1=VAR1
IF(VAR1.GT.NV) IVAR2=IVAR2+1
IF(ITIT.LT.1) GO TO 41
DO 40 II=1,20
NATR(VAR1,II)=NATRF(II)

```

```

40 CONTINUE
GO TO 43
41 DO 42 II=1,20
NATR(VAR1,II)=BLANK
42 CONTINUE
1001 FCKMAT(1,II,A2,314)
43 IF(ERR.EQ.0) GO TO 50
CALL MESSGE(4,P)
CALL MESSGE(9,F)
ERRFLG=1
GO TO 555
C ICI IL DEVRAIT Y AVOIR UNE VIRGULE
50 IF(CARD(P).EQ.CCMA)GO TO 55
WRITE(3,1001) CARD(P),P
CALL MESSGE(5,F)
CALL MESSGE(9,P)
ERRFLG=1
GO TO 555
C ICI ON VERIFIE POUR LA DEUXIEME VARIABLE
55 P=P+1
C CN AUGMENTE TOUJOUR LE POINTEUR DE LA CARTE POUR SAVOIR OU ON EST
CALL VARCHK(CARD,P,ERR2,VAR2)
WRITE(3,1001) CARD(P),P,ERR2,VAR2
IF(ERR2.EQ.0) GO TO 60
CALL MESSGE(4,P)
CALL MESSGE(9,F)
ERRFLG=1
GO TO 555
60 IF (CARD(P).EQ.RBRAK)GO TO 5000
C SI ON A UNE PARENTHESE CN A SEULEMENT DEUX ARGUMENTS ET ON EXECUTE
C
IF(CARD(P).EQ.CCMA)GO TO 61
CALL MESSGE(5,P)
CALL MESSGE(9,P)
GO TO 555
61 P=P+1
WRITE(3,1001) CARD(P),P
IF (CARD(P).NE.V)GO TO 70
C VERIFIE SI ON A UNE VARIABLE DANS LE TROISIEME ARGUMENT
CALL VARCHK(CARD,P,ERR3,VAR3)
WRITE(3,1001) CARD(P),P,ERR3,VAR3
IF(ERR3.EQ.0) GO TO 65
CALL MESSGE(6,P)
CALL MESSGE(9,F)
ERRFLG=1
GO TO 555
65 ARG3=1
C ARG3 NOUS INDIQUE SI LE TROISIEME ARGUMENT EST UNE VARIABLE
WRITE(3,1001) CARD(P),P
IF(CARD(P).EQ.RBRAK) GO TO 5000
CALL MESSGE(7,P)
CALL MESSGE(9,F)
ERRFLG=1
GO TO 555
70 DO 1 II=P,80
JJ=II
C VERIFIE SI LE TROISIEME ARGUMENT EST UNE DIVISION
IF(CARD(II).EQ.SLASH) GO TO 75
1 CONTINUE
DO 2 II=P,80

```

```

      IF(CARD(II).EQ.RBRAK) FOINT=II-1
2 CONTINUE
C ICI CN EXTRAIT UN NOMBRE
CALLTSTNUM(CARD,P,POINT,ERR5)
WRITE(3,1001) CARD(P),P,POINT,ERR5
IF(ERR5.EQ.0)GOTO 71
CALL MESSGE(8,P)
CALL MESSGE(9,P)
ERRFLG=1
GO TO 555
71 CALL GETNUM (CARD,P,POINT,ZVAR3,ERRN)
WRITE(3,1002) CARD(P),P,POINT,ZVAR3,ERRN
1002 FORMAT(50X,A2,2I4,F8.2,I4)
ARG3=2
C ARG3 VEUT DIRE QUE LE TROISIEME ARGUMENT EST UN NOMBRE
GO TO 5000
75 P5=JJ-1
P6=JJ+1
C P5 INDIQUE LA LOCATION PRECEDANT LE /
C P6 INDIQUE LA LOCATION SUIVANT LE /
DO 3 KK=P,80
P7=KK-1
C P7 INDIQUE LA GAUCHE DE LA PARENTHESE DROITE
IF(CARD(KK).EQ.RBRAK) GO TO 80
3 CONTINUE
CALL MESSGE(7,P)
CALL MESSGE(9,P)
ERRFLG=1
GO TO 555
80 CALL TSTNUM(CARD,P,P5,ERR6)
WRITE(3,1001) CARD(P),P,P5,ZNUMB1,ERR6
IF(ERR6.EQ.0)GO TO 81
CALL MESSGE(8,P)
CALL MESSGE(9,P)
ERRFLG=1
GO TO 555
81 CALLGETNUM(CARD,P,P5,ZNUMB1,ERR7)
WRITE(3,1003) CARD(P),P,P5,ZNUMB1,ERR7
1003 FORMAT(1H0,25(' '),A2,2I4,F8.3,I4)
CALL TSTNUM(CARD,P6,P7,ERR7)
WRITE(3,1001) CARD(P),P6,P7,ERR7
IF (ERR7.EQ.0) GO TO 82
CALL MESSGE(8,P)
CALL MESSGE(9,P)
ERRFLG=1
GO TO 555
82 CALL GETNUM(CARD,P6,P7,ZNUMB2,ERR2)
WRITE(3,1003) CARD(P),P,P5,ZNUMB2,ERR2
ZVAR3=ZNUMB1/ZNUMB2
WRITE(3,1004) ZVAR3
1004 FORMAT(1H0,'ZVAR3=',F10.5)
ARG3=2
C ICI TERMINE LA SEPARATION DES ELEMENTS ET LA VERIFICATION DE LA SYNTAXE
5000 WRITE(3,5003)
5003 FORMAT(1X,'JE SUIS ICI')
IF(ERRFLG.EQ.1) GO TO 555
5001 GO TO (100,200,300,400,500,600,700,800,900,1050,1100), IOP
C
C ADDITION
C

```

```

100 IF(ARG3.EQ.1) GO TO 101
    DC 110 II=1,MC
    VARI(II,VAR1)=VARI(II,VAR2)+ZVAR3
    IF(VARI(II,VAR1)-999999999.99999) 104,104,105
104 WRITE(3,112) VARI(II,VAR1)
112 FCRMAT(5X,'VARI=',F10.3)
110 CONTINUE
    WRITE(3,102) VARI,VAR2,ZVAR3
102 FCRMAT(1X,'LA VARIABLE',I3,' EST LR RESULTAT DE LA SOMME DE LA VAR
+IABLE',I3,' ET DE LA CONSTANCE',F15.5)
    CALL MESSAGE(8,F)
    GO TO 555
105 CALL DIMAX(II,VAR1,IDIMAX)
    GO TO 110
101 DO 111 II=1,MO
    VARI(II,VAR1)=VARI(II,VAR2)+VARI(II,VAR3)
    IF(VARI(II,VAR1)-999999999.99999) 113,113,116
113 WRITE(3,114) VARI(II,VAR1)
114 FCRMAT(15X,'VARI=',F10.3)
111 CONTINUE
    WRITE(3,115) VARI,VAR2,VAR3
115 FCRMAT(1X,'LA VARIABLE',I3,' EST LE RESULTAT DE LA SOMME DE LA VAR
+IABLE',I3,' ET DE LA VARIABLE',I3)
    GO TO 555
116 CALL DIMAX(II,VAR1,IDIMAX)
    GO TO 111

```

C
C
C

```

200 IF(ARG3.EQ.1) GO TO 201
    DO 210 II=1,MC
    VARI(II,VAR1)=VARI(II,VAR2)-ZVAR3
    IF(VARI(II,VAR1)-999999999.99999) 204,204,206
204 WRITE(3,212) VARI(II,VAR1)
212 FCRMAT(5X,'VARI--',F15.3)
210 CONTINUE
    WRITE(3,205) VARI,VAR2,ZVAR3
205 FCRMAT(1X,'LA VARIABLE',I3,' EST LE RESULTAT DE LA DIFFERENCE ENTR
+E LA VARIABLE',I3,' ET LA CONSTANCE',F15.5)
    GO TO 555
206 CALL DIMAX(II,VAR1,IDIMAX)
    GO TO 210
201 DO 211 II=1,MO
    VARI(II,VAR1)=VARI(II,VAR2)-VARI(II,VAR3)
    IF(VARI(II,VAR1)-999999999.99999) 216,216,218
216 WRITE(3,214) VARI(II,VAR1)
214 FCRMAT(15X,'VARI--',F15.3)
211 CONTINUE
    WRITE(3,215) VARI,VAR2,VAR3
215 FCRMAT(1X,' LA VARIABLE',I3,' EST LE RESULTAT DE LA DIFFERENCE ENT
+RE LA VARIABLE',I3,' ET LA VARIABLE',I3)
    GO TO 555
218 CALL DIMAX(II,VAR1,IDIMAX)
    GO TO 211

```

C
C
C

```

300 IF(ARG3.EQ.1) GO TO 301
    DC 310 II=1,MC
    VARI(II,VAR1)=VARI(II,VAR2)*ZVAR3

```

```

      IF(VARI(II,VAR1)-999999999.99999) 303,303,304
303 WRITE(3,312) VARI(II,VAR1)
312 FCFMAT(5X,'VARI*=',F15.3)
310 CONTINUE
      WRITE(3,305) VAR1,VAR2,ZVAR3
305 FCFMAT(1X,'LA VARIABLE',I3,' EST LE RESULTAT DU PRODUIT ENTRE LA V
+ARIABLE',I3,' ET LA CCNSTANTE',F15.5)
      GC TC 555
304 CALL DIMAX(II,VAR1,IDIMAX)
      GO TO 310
301 DC 311 II=1,MO
      VARI(II,VAR1)=VARI(II,VAR2)*VARI(II,VAR3)
      IF(VARI(II,VAR1)-999999999.99999) 316,316,317
316 WRITE(3,314) VARI(II,VAR1)
314 FCFMAT(15X,'VARI**=',F15.3)
311 CONTINUE
      WRITE(3,315) VAR1,VAR2,VAR3
315 FCFMAT(1X,'6A VARIABLE',I3,' EST LA MULTIPLICATION DES VARIABLES',
+I3,' ET',I3)
      GC TC 555
317 CALL DIMAX(II,VAR1,IDIMAX)
      GC TC 311

```

```

C
C DIVISION
C
400 IF(ARG3.EQ.1) GO TO 401
      IF(ZVAR3.EQ.0.C) GO TC 409
      DO 410 II=1,MO
      VARI(II,VAR1)=VARI(II,VAR2)/(ZVAR3)
      IF(VARI(II,VAR1)-999999999.99999) 406,406,407
406 WRITE(3,412) VARI(II,VAR1)
412 FCFMAT(5X,'VARI/=',F15.3)
410 CONTINUE
      WRITE(3,405) VAR1,VAR2,ZVAR3
405 FCFMAT(1X,'LA VARIABLE',I3,' EST LA RESULTAT DE LA DIVISION DE LA
+VARIABLE',I3,' PAR LA CONSTANCE',F15.5)
      GC TO 555
407 CALL DIMAX(II,VAR1,IDIMAX)
      GO TO 410
401 DC 404 II=1,MC
      IF(VARI(II,VAR3).NE.0.0) GC TO 404
      WRITE(3,403) II,VAR3
403 FCFMAT(1X,'L'CBSEVATION',I3,' DE LA VARIABLE',I3,' A UNE VALEUR D
+E ZERC: IMPOSSIBILITE D'EFFECTUER LA DIVISION')
      KDIV=1
404 CCNTINUE
      IF(KDIV.EQ.1) GO TO 555
      DC 411 II=1,MO
      VARI(II,VAR1)=VARI(II,VAR2)/VARI(II,VAR3)
      IF(VARI(II,VAR1)-999999999.99999) 416,416,417
416 WRITE(3,414) VARI(II,VAR1)
414 FCFMAT(15X,'VARI//=',F15.3)
411 CONTINUE
      WRITE(3,415) VAR1,VAR2,VAR3
415 FCFMAT(1X,'LA VARIABLE',I3,' EST LE RESULTAT DE LA DIVISION DE LA
+VARIABLE',I3,' PAR LA VARIABLE',I3)
      GC TO 555
409 WRITE(3,402)
402 FCFMAT(1X,'LE DIVISEUR EST EGALE A ZERU : IMPOSSIBLE D'EFFECTUER
+LA DIVISION')

```

```
GC TC 555
417 CALL DIMAX(II,VAR1,IDIMAX)
GO TO 411
```

```
C
C METTRE A LA PUISSANCE
C
```

```
500 IF(ZVAR3.NE.0.0)GO TO 505
WRITE(3,501)
501 FORMAT(5X,' LA VALEUR DE L''EXPOSANT EST EGALE A ZERO: DONC ENONCE
+INVALIDE PUISQUE TOUTE VALEUR EXPOSANT ZERO EST EGALE A UN')
GC TC 555
505 DO 510 II=1,MO
VARI(II,VAR1)=VARI(II,VAR2)**(ZVAR3)
IF(VARI(II,VAR1)-999999999.99999) 503,503,504
503 WRITE(3,512) VARI(II,VAR1)
512 FCRMAT(5X,' PUISSANCE VARI=',F15.3)
510 CONTINUE
WRITE(3,506) VARI,VAR2,ZVAR3
506 FCRMAT(1X,' LA VARIABLE',I3,' EST LE RESULTAT DE LA VARIABLE',I3,'
+A LA PUISSANCE',F15.5)
GC TC 555
504 CALL DIMAX(II,VAR1,IDIMAX)
GO TO 510
```

```
C
C EXTRACTION DE LA RACINE CARRE
C
```

```
600 DO 610 II=1,MO
IF(VARI(II,VAR1)-999999999.99999) 603,603,604
603 VARI(II,VAR1)=SQRT(VARI(II,VAR2))
WRITE(3,612) VARI(II,VAR1)
612 FCRMAT(5X,' SQRT VARI=',F15.3)
610 CONTINUE
WRITE(3,605) VARI,VAR2
605 FCRMAT(1X,' LA VARIABLE',I3,' EST L''EXTRACTION DE LA RACINE CARREE
+ DE LA VARIABLE',I3)
GC TC 555
604 CALL DIMAX(II,VAR1,IDIMAX)
GO TO 610
```

```
C
C EXPONENTIELLE
C
```

```
700 DO 710 II=1,MO
VARI(II,VAR1)=EXP(VARI(II,VAR2))
IF(VARI(II,VAR1)-999999999.99999) 720,720,725
720 WRITE(3,712) VARI(II,VAR1)
712 FORMAT(5X,' EXP=VARI=',F15.3)
710 CONTINUE
WRITE(3,705) VARI,VAR2
705 FORMAT(1X,' LA VARIABLE',I3,' EST LA TRANSFORMATION EXPONENTIELLE D
+E LA VARIABLE',I3)
GC TC 555
725 CALL DIMAX(II,VAR1,IDIMAX)
GO TO 710
```

```
C
C LOGARITHME NATUREL
C
```

```
800 DO 805 II=1,MO
IF(VARI(II,VAR2).NE.0.0) GO TO 805
WRITE(3,806) II,VAR2
806 FORMAT(1X,' L''OBSERVATION',I3,' DE LA VARIABLE',I3,' A UNE VALEUR
```

+PLUS PETITE OU EGALE A ZERO :'/25X,'DONC IMPOSSIBILITE D''EN EXTRA
+IRE LE LOGARITHME NATUREL')

JLOG=1

805 CCNTINUE

IF(JLOG.EQ.1) GC TO 555

DO 810 II=1,MU

VARI(II,VAR1)=ALOG(VARI(II,VAR2))

IF(VARI(II,VAR1)-999999999.99999) 803,803,804

803 WRITE(3,811) VARI(II,VAR1)

811 FCRMAT(25X,'LOG VARI=',F15.5)

810 CCNTINUE

WRITE(3,815)VAR1,VAR2

815 FCRMAT(1X,'LA VARIABLE',I3,' EST LA TRANSFORMATION EN LOGARITHME N
+ATUREL DE LA VARIABLE',I3)

GO TO 555

804 CALL DIMAX(II,VAR1,IDIMAX)

GC TO 810

C

C LOGARITHME DANS LA BASE 10

C

900 DO 905 II=1,MO

IF(VARI(II,VAR2).NE.0.0.AND.VARI(II,VAR2).GT.0.0) GO TO 905

WRITE(3,906) II,VAR2

906 FCRMAT(1X,'L''OBSERVATION',I3,' DE LA VARIABLE',I3,' A UNE VALEUR
+PLUS PETITE OU EGALE A ZERO :'/25X,'DONC IMPOSSIBILITE D''EN EXTRA
+IRE LE LOGARITHME')

KLOG=1

905 CCNTINUE

IF(KLOG.EQ.1) GO TO 555

DC 910 II=1,MC

VARI(II,VAR1)=ALOG10(VARI(II,VAR2))

IF(VARI(II,VAR1)-999999999.99999) 903,903,904

903 WRITE(3,911) VARI(II,VAR1)

911 FCRMAT(1X,'LOGARITHME DE VARI=',F15.5)

910 CCNTINUE

WRITE(3,915) VAR1,VAR2

915 FCRMAT(1X,'LA VARIABLE',I3,' EST LE LOGARITHME DANS LA BASE 10 DE
+LA VARIABLE',I3)

GC TO 555

904 CALL DIMAX(II,VAR1,IDIMAX)

GC TO 910

C

C VALEURS ABSOLUES

C

1050 DC 1060 II=1,MC

VARI(II,VAR1)=ABS(VARI(II,VAR2))

IF(VARI(II,VAR1)-999999999.99999) 1051,1051,1052

1051 WRITE(3,1061) VARI(II,VAR1)

1061 FCRMAT(5X,'VALEUR ABSCLUE DE VARI=',F15.3)

1060 CCNTINUE

WRITE(3,1055) VAR1,VAR2

1055 FCRMAT(1X,'LA VARIABLE',I3,' EST LA CONVERSION DE LA VARIABLE',I3,
+' EN VALEURS ABSCLUES')

GO TO 555

1052 CALL DIMAX(II,VAR1,IDIMAX)

GC TO 1060

C

C POURCENTAGE

C

1100 SPER=0.0

```

      DC 1110 II=1,MC
      SPER=SPER+VARI(II,VAR2)
1110 CONTINUE
      WRITE(3,1111) VAR2
1111 FORMAT(1X,' PCLRCENTAGES DE LA VARIABLE',I3)
      DC 1120 II=1,MC
      VARI(II,VAR1)=(VARI(II,VAR2)*100)/SPER
      IF(VARI(II,VAR1)-99999999.99999) 1101,1101,1102
1101 WRITE(3,1121) VARI(II,VAR1)
1121 FCRMAT(25X,F15.5,'(')
1120 CONTINUE
      WRITE(3,1115) VAR1,VAR2
1115 FCRMAT(1X,'LA VARIABLE',I3,' EST LA TRANSFORMATION EN POURCENTAGE
+DE LA VARIABLE',I3)
      GO TO 555
1102 CALL DIMAX(II,VAR1,IDIMAX)
      GO TO 1120
1201 WRITE(3,1200) IVAR1
1200 FCRMAT(5X,'IVARI-IVARI-IVARI-IVARI-IVARI=',I3)
      IF(IVAR1.EQ.IVAR2) GO TO 1202
      WRITE(3,1208)
1208 FCRMAT(1H1,'LISTE DES VARIABLES TRANSFORMEES')
      DO 1220 VAR1=1,IVAR1
      WRITE(3,1207)(MATR(VAR1,II),II=1,20)
1207 FORMAT(25X,20A1)
1220 CONTINUE
      ERRFL1=1
      WRITE(3,1209)
1209 FCRMAT(1X,'LES NOUVELLES VARIABLES AJOUTEES NE SONT PAS CONTIGUES
+A CELLES DEJA EXISTANTES')
      RETURN
1202 WRITE(3,1210)
1210 FCRMAT('1',56X,'MATRICE TRANSFORMEE'/57X,19('='))//)
      DO 1204 II=1,MO
      WRITE(3,1206) (VARI(II,VAR1),VAR1=1,IVAR1)
1206 FORMAT(8F15.5)
1204 CONTINUE
      NV=IVAR1
      RETURN
      END
      SUBROUTINE EXTRAC (CARD,P1,P2,ZNUMB,TEST)
      IMPLICIT INTEGER*4(A-Y)
      COMMON/GU/NOBS(150,20),MATR(100,20),DD(150),      FMT(100),ITITR(20)
      COMMON/DIG/DIGIT(10),DCT
      COMMON/ALPHA/A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,LP,Q,R,S,T,U,V,W,X,Y
      COMMON/CHAR/BLANK,COMMA,SLASH,LBRAK,RBRAK
      DIMENSION CARD(80),      ZNUMBE(10)
      NDIGIT=P2-P1+1
      ZNUMB=0.0
      KJ=0
      DO 10 KK=P1,P2
      DO 10 LL=1,10
      IF(CARD(KK).NE.DIGIT(LL)) GO TO 10
      KJ=KJ+1
      ZNUMBE(KJ)=LL-1
10 CONTINUE
      IF(TEST.EQ.1)GO TO 15
      DO 20 II=1,NDIGIT
      ZNLMB=(ZNUMBE(II))*(10**((NDIGIT-II)))+ZNUMB
      WRITE(3,40) ZNLMB

```

```

40 FORMAT(1H0,' EXTRAC=' ,F10.5)
RETURN
15 CC 30 II=1,NDICIT
30 ZNLMB=(ZNLMBE(II))*(1./(10**(II))))+ZNLMB
WRITE(3,41) ZNLMB
41 FCRMAT(1H0,' EXTRAC30=' ,F10.3)
RETURN
END
SUBROUTINE GETNUM(CARD,P,PP,ZNUM,ERR)
IMPLICIT INTEGER*4(A-Y)
COMMON/GU/NOBS(150,20),NATR(100,20),DD(150), FMT(100),ITITR(20)
COMMON/DIG/DIGIT(10),DOT
COMMON/ALPHA/A,E,C,D,E,F,G,H,I,J,K,L,M,N,O,LP,Q,R,S,T,U,V, W,X,Y
COMMON/CHAR/BLANK,COMMA,SLASH,LBRAK,RBRAK
DIMENSION CARD(80)
ZNUM=0.0
PPP=PP
CC 10 II=P,PPP
IDOT=II
IF(CARD(II).EQ.DOT) GO TO 50
10 CONTINUE
CALL EXTRAC (CARD,P,PPP,ZNUM,0)
WRITE(3,70) CARD(P),P,PPP,ZNUM
70 FORMAT(1H0,' GETNUM=' ,A2,2I4,F8.3)
RETURN
50 P3=IDOT-1
P4=IDOT+1
CALL EXTRAC (CARD,P,P3,ZNUM1,0)
IF(P4.GT.PPP) GO TO 80
CALL EXTRAC(CARD,P4,PPP,ZNUM2,1)
GO TO 90
80 ZNUM2=0.0
90 ZNUM=ZNUM1+ZNUM2
WRITE(3,60) ZNUM
60 FCRMAT(1H0,' ZNUM=' ,F10.3)
RETURN
END
SUBROUTINE VARCHK(CARD,P,ERR,VAR)
IMPLICIT INTEGER*4(A-Y)
COMMON/GU/NOBS(150,20),NATR(100,20),DD(150), FMT(100),ITITR(20)
COMMON/DIG/DIGIT(10),DOT
COMMON/ALPHA/A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,LP,Q,R,S,T,U,V, W,X,Y
COMMON/CHAR/BLANK,COMMA,SLASH,LBRAK,RBRAK
DIMENSION CARD(80), NUM(3)
ERR=0
WRITE(3,500) CARD(P),CARD(P+1),CARD(P+2)
500 FORMAT(1H1,' ENTERED VARCHK' ,3A1)
IF(CARD(P).EQ.V.AND.CARD(P+1).EQ.A.AND.CARD(P+2).EQ.R)GO TO 40
ERR=1
WRITE(3,101) ERR,P
101 FCRMAT(1H0,' ERR=' ,2I5)
RETURN
40 P=P+3
TST1=0
TST2=0
TST3=0
CC 10 II=1,10
IF(CARD(P).EQ.DIGIT(II)) TST1=1
IF(CARD(P+1).EQ.DIGIT(II)) TST2=1
IF(CARD(P+2).EQ.DIGIT(II)) TST3=1

```

```

10 CONTINUE
  STST=TST1+TST2+TST3
  WRITE(3,102) TST1,TST2,TST3,CARD(P),CARD(P+1),CARD(P+2),STST
102 FORMAT(1H0,3I5,2X,3A1,I5)
  IF(STST.EQ.3) GO TO 41
  ERR=1
  RETURN
41 PP=P+2
  DO 100 II=P,PP
  DO 100 JJ=1,10
  IF(CARD(II).EQ.DIGIT(JJ)) NUM(II-P+1)=JJ-1
  WRITE(3,104) II,P,JJ
104 FORMAT(2GX,3I5)
100 CONTINUE
  VAR=100*NUM(1)+10*NUM(2)+NUM(3)
  WRITE(3,103) VAR,(NUM(II),II=1,3)
103 FCRMAT(1H0,4I5)
  P=P+3
  RETURN
  END
  SUBROUTINE MESSGE(N,P)
  WRITE(3,100) N,M
100 FCRMAT(1H0,10('*'),5X,I4,2X,I4,5X,10('*'))
  GO TO (10,20,30,40,50,60,70,80,90),N
  10 WRITE(3,11)
  11 FCRMAT(1X,'IL Y A UNE CARTE NON PERFOREE PARMIS LES ORDRE DE LA TR
  +NSFORMATION')
  RETURN
  20 WRITE(3,21) P
  21 FCRMAT(1X,'IL Y UNE ERREUR A LA COLONNE',I2,' AU NIVEAU DU CODE D''
  +OPERATION ')
  RETURN
  30 WRITE(3,31) P
  31 FCRMAT(1X,'IL DEVRAIT Y AVOIR UNE PARENTHÈSE GAUCHE A LA COLONNE
  +',I2)
  RETURN
  40 WRITE(3,41)P
  41 FCRMAT(1X,'IL Y A UNE ERREUR A LA COLONNE',I2)
  RETURN
  50 WRITE(3,51) P
  51 FCRMAT(1X,'IL MANQUE UNE VIRGULE A LA ',I2,' IÈME COLONNE')
  RETURN
  60 WRITE(3,61) P
  61 FCRMAT(1X,'IL Y A UNE ERREUR A LA ',I2,' IÈME COLONNE')
  RETURN
  70 WRITE(3,71) P
  71 FCRMAT(1X,'IL MANQUE UNE PARENTHÈSE DROITE A LA ',I2,' IÈME COLONNE')
  RETURN
  80 WRITE(3,81) P
  81 FCRMAT(1X,'IL Y A UNE ERREUR A LA ',I2,' IÈME COLONNE AU NIVEAU DE
  +LA CONSTANTÉ')
  RETURN
  90 WRITE(3,91)
  91 FCRMAT(1X,' DU A LA DETECTION D''ERREUR IL EST IMPOSSIBLE D''EFFEC
  +TUER LA TRANSFORMATION')
  RETURN
  END
  SUBROUTINE DIMAX(II,VARI)
  WRITE(3,1) II,VARI
  1 FCRMAT(1X,'L''OBSERVATION',I3,' DE LA VARIABLE',I3,' A UNE VALEUR

```

```

+PLUS GRANDE QUE 999 999 999.99999:CE QUI DEPASSE LA LA LIMITE'/'MA
+XIMALE DU SYSTEME S.E.I.G.E.G.')
  ICIMAX=1
  RETURN
  END
  SLBRCTINE TSTNUM (CARD,P,PP,ERR)
  IMPLICIT INTEGER*4(A-Y)
  COMMON/GU/NOBS(150,20),NATR(100,20),DD(150),      FMT(100),ITITR(20)
  COMMON/DIG/DIGIT(10),DOT
  COMMON/ALPHA/A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,LP,Q,R,S,T,U,V,W,X,Y
  COMMON/CHAR/BLANK,CCMMA,SLASH,LBRAK,RBRAK
  DIMENSION CARD(80)
  ERR=0
  DC 10 KK=P,PP
  TST=0
  DC 11 JJ=1,10
  IF(CARD(KK).EQ.DIGIT(JJ).OR.CARD(KK).EQ.DOT) TST=1
11 CONTINUE
  IF(TST.EQ.0) GO TO 50
10 CCNTINUE
  RETURN
50 ERR=1
  RETURN
  END
  BLCK DATA
  IMPLICIT INTEGER*4(A-Y)
  COMMON/GU/NOBS(150,20),NATR(100,20),DD(150),      FMT(100),ITITR(20)
  COMMON/DIG/DIGIT(10),DOT
  COMMON/ALPHA/A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,LP,Q,R,S,T,U,V,W,X,Y
  COMMON/CHAR/BLANK,COMMA,SLASH,LBRAK,RBRAK
  DATA A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,LP,Q,R,S,T,U,V,W,X,Y/'A','B','C
+', 'D','E','F','G','H','I','J','K','L','M','N','O','P','Q','R','S',
+', 'T','U','V','W','X','Y'/
  DATA DIGIT /'0','1','2','3','4','5','6','7','8','9'/
  DATA BLANK,COMMA,SLASH,LBRAK,RBRAK/' ','/','(',')','.',DOT/'.'/
  END
  SLBRCTINE TYPE(N,M,ITR,ITY)
  COMMON/GU/NOBS(150,20),NATR(100,20),DD(150),      FMT(100),ITITR(20)
  COMMON ENME(150,100)
  ICTY=0
  ICSTY=0
  ITY=1
  CALL JTRANS(N,M,ITR)
  KK=20
  WRITE(3,1)
1 FORMAT(1X,'JE SLIS DANS LA TYPO')
  REWIND 10
  WRITE(10)KK,N,M,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+',I=1,N),((ENME(I,J),J=1,M),I=1,N),ICTY,ICSTY,ITY
  RETURN
  END
  SLBRCTINE NORP(N,M,INC,ITY,INOT,INOS)
  COMMON/GU/NOBS(150,20),NATR(100,20),DD(150),      FMT(100),ITITR(20)
  COMMON ENME(150,100)
  DIMENSION JXK(100)
  INC=1
  NP=N
  NV=M
  READ(1,1) NPXS,NPSS,ICOR,IMISS,IXC,ISTDC,IPC
1 FORMAT(7I5)

```


	17.80	25.80	62.40	96.50	3.00	0.00	0.30	1.00	2.40	06
	4.90	3.50	0.00	726.00	14.00	.85	.72			06
CHATAUGUAY-HEIGHTS										
	58.20	37.40	56.60	97.10	3.90	0.00	0.00	0.50	3.30	07
	0.50	0.00	1.60	109.00	12.00	.87	.76			07
DELSCN										
	29.30	26.00	18.50	99.00	0.00	1.00	0.00	0.00	0.80	08
	1.50	48.70	4.60	123.00	18.00	.02	.00			08
GREENFIELD-PARK										
	99.20	21.20	59.00	85.40	14.40	0.50	0.00	2.20	4.20	09
	12.10	1.30	0.00	383.00	25.00	.88	.77			09
JACQUES-CARTIER										
	29.90	24.10	49.30	74.00	15.90	0.40	9.80	0.40	7.70	10
	4.60	6.90	6.60	1262.00	37.00	.84	.71			10
LAFLECHE										
	75.90	32.30	52.50	68.60	12.00	0.00	19.40	1.70	3.80	11
	4.30	3.50	1.70	407.00	30.00	-.76	.58			11
LA PRAIRIES										
	12.40	12.40	13.90	60.40	32.20	1.30	6.70	19.80	3.90	12
	2.50	45.50	2.10	252.00	30.00	-.06	.00			12
LEMOYNE										
	99.20	7.00	67.50	31.90	49.60	8.10	10.40	1.00	6.50	13
	9.00	6.00	2.50	140.00	67.00	.96	.92			13
LERY										
	20.40	12.50	53.10	34.20	1.10	0.00	65.10	30.30	0.60	14
	0.60	0.00	3.00	354.00	6.00	.76	.58			14
LONGUEUIL										
	86.90	6.20	61.20	49.30	42.10	2.60	6.00	2.70	12.40	15
	5.60	11.40	0.30	598.00	43.00	.98	.96			15
NOTRE-DANE										
	96.90	15.60	78.10	59.20	0.80	0.00	40.00	0.60	1.90	16
	0.60	0.00	3.10	125.00	23.00	.93	.86			16
PREVILLE										
	46.70	14.50	40.60	99.90	0.00	0.00	0.00	43.00	1.40	17
	0.50	0.00	0.00	88.00	23.00	.52	.27			17
ST-BRLENC										
	8.70	7.70	61.20	93.10	6.50	0.60	0.00	15.30	3.20	18
	4.30	8.20	0.20	840.00	10.00	.92	.85			18
STE-CATHERINE										
	13.80	15.70	48.20	99.70	0.30	0.00	0.00	25.00	1.60	19
	1.20	7.30	1.00	186.00	11.00	.74	.55			19
ST-CONSTANT										
	2.40	32.60	55.40	69.10	12.10	0.00	18.80	0.00	4.00	20
	4.40	2.00	1.70	513.00	10.00	.77	.59			20
ST-HUBERT										
	14.10	24.70	37.10	86.30	6.00	0.00	7.70	2.80	4.40	21
	1.50	1.10	28.40	930.00	16.00	.52	.27			21
ST-LAMBERT										
	79.20	10.90	59.20	88.60	8.30	0.60	2.30	8.00	6.20	22
	3.50	8.80	3.40	523.00	29.00	.95	.90			22

B-TYPE
C-NORMALISATION

D-CO
F-FA
03-1 1.0000100000001
FINI

//STEP2 EXEC FCRTGCLG,PARM.FORT='ECD',REGION.GO=200K,TIME.GO=3

```

//FCRT.SYSIN DD *
C PROGRAMME DE NORMALISATION DES DONNEES D APRES SNEDECOR MODIFIE PAR
C GLY LEMAY DEPARTEMENT DE GEOGRAPHIE UNIVERSITE D OTTAWA
C CETTE VERSION N'EST PAS FINALE
C LES TRANSFORMATIONS POSSIBLES SONT : LOG10,RACINE CARREE,RACINE CU-
C BIQUE,LA MISE AU CARRE ET AU CUBE
C
C CE PROGRAMME EXECUTE LA STANDARDISATION SUR DEMANDE
C
C
C NP NOMBRE DE PLACES
C NV NOMBRE DE VARIABLES
C
C TEST DE LA NORMALITE DES DISTRIBUTIONS D'APRES SNEDECOR
C
C JXK=1 POUR LA SELECTION DES VARIABLES DONT ON NE VEUT PAS LA
C TRANSFORMATION
C JXK EST PERFORE SUR UNE CARTE SEPAREMENT
C
C IMISS=1 POUR LA STAND.SEULEMENT,GT 1 POUR NE PAS NORM. NI STAND,
C =0 POUR EXECUTION DES DEUX.(NORM+STAND.)
C
C
C
C DIMENSION A(30,210),STDS(30,210),Q(210,9),XBAR(210),VARB(210),
+VARC(210),VARD(210),GCNE(210),GTWO(210),TEED(210),ESSC(210),
+ESST(210),SSD(210),SST(210),FMT(20),IHEAD(20),JXK(210),B(30,210),
+AA(210),TEET(210)
DIMENSION NATR(100,20),NOBS(100,20)
EQUIVALENCE(B,STDS)
C
996 READ(8) KK,N,M,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+I=1,N),((A(I,J),J=1,M),I=1,N)
READ(8) IGE,ITR,ITY,INO,INOT,INGS,ICO,ILI,ICOMP
WRITE(6,9070) IGE,ITR,ITY,INO,INOT,INGS,ICO,ILI,ICOMP
9070 FCRMAT(1X,9I3)
ITT=C
TEMP=N
N=M
M=TEMP
NACO=C
ICTY=C
ICSTY=0
IF(ITY.EQ.0.AND.INOT.EQ.0.AND.INO.EQ.0) GO TO 9001
IF(ITY.EQ.1.AND.INOT.EQ.1.AND.INO.EQ.1) GO TO 998
IF(ITY.EQ.1.AND.INOT.EQ.0.AND.INO.EQ.0) GO TO 9002
605 IF(INO.EQ.1.AND.INGS.EQ.1) GO TO 620
GO TO 9004
C 605 READ(5,101,END=9001) NP,NV,NPXS,NPSS,(IHEAD(JJ),JJ=1,10),ICOR,IMISS
998 READ(10,END=999) NP,NV,NPXS,NPSS,ICOR,IMISS,IXC,ISTDC,IPC,
+(IHEAD(JJ),JJ=1,20),((A(I,J),J=1,NV),I=1,NP),(JXK(J),J=1,NV)
ITT=1
GO TO 999
620 READ(11,END=997) NP,NV,NPXS,NPSS,ICOR,IMISS,IXC,ISTDC,IPC,
+(IHEAD(JJ),JJ=1,20),((A(I,J),J=1,NV),I=1,NP),(JXK(J),J=1,NV)
997 ITY=0
999 CONTINUE
PRINT101, NP,NV,NPXS,NPSS,(IHEAD(JJ),JJ=1,20),ICOR,IMISS,IXC,ISTDC
*,IPC
101 FORMAT(4I3,20A4,23X,5I1)

```

```

C   READ (5,1) (FMT(JJ),JJ=1,20)
C   1 FORMAT (20A4)
      M=NP
      N=NV
C   READ 1784,(JXK(J),J=1,N)
C1784 FORMAT(80I1)
      DC 61 I=1,M
C   61 READ(5,63) (A(I,J),J=1,N)
      61 CONTINUE
      63 FCRMAT(4F8.3)
C   61 READ(5,FMT) (A(I,J),J=1,N)
      PRINT 60,(IHEAD(JJ),JJ=1,20)
      60 FCRMAT('1',2X,'MATRICE DE DONNES ORIGINALES',3X, 20A4)
      DO 1001 I=1,NP
      WRITE(6,1002) (A(I,J),J=1,NV)
1002 FCRMAT(1X,8F10.2)
1001 CONTINUE
      L=0
      CALL RPFINT(A,N,P,L)
      XM=M

C
      DC 7 I=1,M
      DC 7 J=1,N
      7 B(I,J)=A(I,J)
      IF(IMISS.GT.1) GO TO 9000
      DO2 J=1,N
      IF(ISS2.EQ.1) GO TO 2
      DC 717 JN=1,10
717 AA(JN)=0.
      JN=0
      ISKT=0
      ISI=0
      ISKIP=0
100 XXX=0.
      DO3 I=1,M
      3 XXX=XXX+A(I,J)
      XBAR(J)=XXX/XM
      DO 4001 K=1,4
4001 C(J,K) =0.0
      DC 4000 I=1,M
      Q(J,2) = Q(J,2)+((A(I,J)-XBAR(J))**2)
      Q(J,3) = Q(J,3)+((A(I,J)-XBAR(J))**3)
4000 Q(J,4) = Q(J,4)+((A(I,J)-XBAR(J))**4)
      VARB(J) = Q(J,2)/(XM-1.0)
      VARC(J) = (XM* Q(J,3))/( (XM-1.)*(XM-2.))
4003 VARD(J)=(XM*((XM+1.)*C(J,4))-((3.*(XM-1.))*((Q(J,2)**2)/XM)))/
      *((XM-1.)*(XM-2.)*(XM-3.))
      GCNE(J)=(VARC(J))/(VARB(J)*SQRT(VARB(J)))
      GTWO(J)=VARD(J)/(VARB(J)**2)
      ESSO(J)=((6.*XM)*(XM-1.))/( (XM-2.)*(XM+1.)*(XM+3.))
      ESST(J)=((24.*XM)*((XM-1.)**2))/( (XM-3.)*(XM-2.)*(XM+3.)*(XM+5.))
      SSU(J)=SQRT(ESSO(J))
      SST(J)=SQRT(ESST(J))
      TEEU(J)=GCNE(J)/SSU(J)
      TEET(J)=GTWO(J)/SST(J)
      IF(IMISS.EQ.1) GO TO 2
      JN=JN+1
      XS=TEEU(J)
      AA(JN)=XS
      XK=TEET(J)

```

```

      IF(ISKIP.GT.0) GO TO 44
      IF(XS)73,20,102
73  IF(XS.GT.(-1.96)) GO TO 20
      GC TC 98
102 IF(XS.LT.1.96) GG TO 20
98  IF(JN.EQ.1) PRINT 99,J,XS,XK
99  FORMAT(///IX,'VARIABLE',I3,2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST
+DE KURTOSI',F10.5)
      IF(JXK(J).GT.0) GO TO 2
      IF(ISWT.EQ.0) GC TO 583
      IF(ISWT.EQ.10) GO TO 25
      IF(ISWT.EQ.(-1)) GO TO 27
      IF(ISWT.EQ.1) GC TO 29
      IF(ISWT.EQ.(-2)) GO TO 31
      IF(ISWT.EQ.2) GC TO 35
583 DO 584 I=1,M
      A(I,J)=B(I,J)
      IF(A(I,J).LE.0.) GO TO 585
584 A(I,J)=ALOG10(A(I,J))
      ISWT=10
      GC TC 100
25  DO 26 I=1,M
      A(I,J)=B(I,J)
      IF(A(I,J).LT.0.) GO TO 33
26  A(I,J)=SQRT(A(I,J))
      ISWT=-1
      GC TC 100
27  DO 28 I=1,M
      A(I,J)=B(I,J)
28  A(I,J)=A(I,J)**2
      ISWT=1
      GO TO 100
29  DO 70 I=1,M
      A(I,J)=B(I,J)
      IF(A(I,J).LT.0.) GO TO 34
      IF(A(I,J).EQ.0.) GO TO 70
      A(I,J)=EXP((ALOG(A(I,J)))/3.)
70  CONTINUE
      ISWT=-2
      GO TO 100
31  DO 32 I=1,M
      A(I,J)=B(I,J)
32  A(I,J)=A(I,J)**3
      ISWT=2
      GC TC 100
33  JN=JN+1
      AA(JN)=99.
      PRINT 19
19  FORMAT(IX,'IMPOSSIBILITE DE TRANSFORMER PAR LA RACINE CARREE PARCE
+ QUE LA VARIABLE CONTIENT DES VALEURS PLUS PETITES QUE 0')
      GC TC 27
34  JN=JN+1
      AA(JN)=99.
      PRINT 18
18  FORMAT(IX,'IMPOSSIBILITE DE TRANSFORMER PAR LA RACINE CUBIQUE PARC
+E QUE LA VARIAELE CNTIENT DES VALEURS PLUS PETITES QUE 0')
      GC TO 31
585 JN=JN+1
      AA(JN)=99.
      PRINT 588,J

```

```

588 FORMAT(1X,'LA TRANSFORMATION EN LOG EST IMPOSSIBLE POUR LA VARIABLE',I4)
      GO TO 25
35  AA(2)=SQRT((AA(2)**2))
      AA(3)=SQRT((AA(3)**2))
      AA(4)=SQRT((AA(4)**2))
      AA(5)=SQRT((AA(5)**2))
      AA(6)=SQRT((AA(6)**2))
      ISKIP =5
      IF(AA(2).LT.AA(3).AND.AA(2).LT.AA(4).AND.AA(2).LT.AA(5).AND.AA(2).
*LT.AA(6)) IS1=5
      IF(AA(3).LT.AA(4).AND.AA(3).LT.AA(5).AND.AA(3).LT.AA(6)) IS1=1
      IF(AA(4).LT.AA(5).AND.AA(4).LT.AA(6)) IS1=2
      IF(AA(5).LT.AA(6)) IS1=3
      IF(AA(6).LT.AA(2).AND.AA(6).LT.AA(3).AND.AA(6).LT.AA(4).AND.AA(6).
*LT.AA(5)) IS1=4
      IF(IS1.EQ.5) GO TO 583
      IF(IS1.EQ.1) GO TO 25
      IF(IS1.EQ.2) GO TO 27
      IF(IS1.EQ.3) GO TO 29
      IF(IS1.EQ.4) GO TO 31
      PRINT 23
23  FORMAT(1X,'ERREUR AU NIVEAU DE LA VERIFICATION')
      GO TO 2
44  IF(IS1.EQ.1) GO TO 10
      IF(IS1.EQ.5) GO TO 589
      IF(IS1.EQ.2) GO TO 15
      IF(IS1.EQ.3) GO TO 76
      IF(IS1.EQ.4) GO TO 45
      PRINT 24
24  FORMAT(1X,'ERREUR LORS DE LA SECONDE VERIFICATION')
      GO TO 2
589 PRINT 591,J,XS,XK
591 FORMAT(1X,/1X,' LA TRANSFORMATION EN LOG EST LA MEILLEURE POUR LA V
+ARIEBLE',I4/1X,'TEST DE SYMMETRIE=',F10.5,2X,'TEST DE KURTOSE=',
+F10.5)
      GO TO 2
10 PRINT 12,J,XS,XK
12 FORMAT(/1X,' LA RACINE CARRE EST LA MEILLEURE TRANSFORMATION POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
      GO TO 2
15 PRINT 77,J,XS,XK
77 FORMAT(/1X,'LA MISE AU CARRE EST LA MEILLEURE TRANSFORMATION POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
      GO TO 2
76 PRINT 74,J,XS,XK
74 FORMAT(/1X,'LA RACINE CUBIQUE EST LA MEILLEURE TRANSFORMATION POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
      GO TO 2
45 PRINT 119,J,XS,XK
119 FORMAT(/1X,'LA MISE AU CUBE EST LA MEILLEURE TRANSFORMATION POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
      GO TO 2

```

```

20 IF(ISWT.EQ.0) PRINT 230,J,XS,XK
230 FORMAT(/1X,'PAS DE TRANSFORMATION NECESSAIRE POUR NORMALISER LA V
+ARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',F10.
+5)
IF(ISWT.EQ.10) PRINT 555,J,XS,XK
555 FORMAT(/1X,'LA TRANSFORMATION EN LOG EST NECESSAIRE POUR TRANSFORM
+ER LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOS
+E',F10.5)
IF(ISWT.EQ.(-1)) PRINT 57,J,XS,XK
57 FORMAT(/1X,'LA RACINE CARREE EST LA TRANSFORMATION NECESSAIRE POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
IF(ISWT.EQ.1) PRINT 58,J,XS,XK
58 FORMAT(/1X,'LA MISE AU CARRE EST LA TRANSFORMATION NECESSAIRE POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
IF(ISWT.EQ.(-2)) PRINT 52,J,XS,XK
52 FORMAT(/1X,'LA RACINE CUBIQUE EST LA TRANSFORMATION NECESSAIRE POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
IF(ISWT.EQ.2) PRINT 53,J,XS,XK
53 FORMAT(/1X,'LA MISE AU CUBE EST LA MEILLEURE TRANSFORMATION POUR
+LA VARIABLE',I4/2X,'TEST DE SYMMETRIE',F10.5,2X,'TEST DE KURTOSE',
+F10.5)
2 CONTINUE
IF(ISS2.EQ.1) GO TO 605
IF(IMISS.EQ.1) GO TO 709
DC 564 J=1,N
DC 564 I=1,M
564 B(I,J)=0.
PRINT 501,(IHEAD(JJ),JJ=1,20)
501 FORMAT('1','MATRICE NORMALISEE',5X,20A4)
L=0
CALL RPRINT(A,N,M,L)
IF(IMISS.EQ.-1) GO TO 209
C
C STANDARDISATION DE LA MATRICE NORMALISEE
C
709 DO 41 J=1,N
DC 41 I=1,M
ZXX=SQRT(Q(J,2)/XM)
41 STDS(I,J)=(A(I,J)-XBAR(J))/ZXX
PRINT 62
62 FORMAT('1','MATRICE STANDARDISEE')
L=1
CALL RPRINT(STDS,N,M,L)
DO 800 I=1,M
DC 800 J=1,N
A(I,J)=STDS(I,J)
800 CONTINUE
209 PRINT 210
210 FORMAT('1',50X,'MEYENNES')
PRINT 199,(XBAR(I),I=1,N)
199 FORMAT(1X,5F20.3)
PRINT 300
300 FORMAT(1HK,50X,4HVARB)
PRINT 299,(VARE(I),I=1,N)
299 FORMAT(1X,5F20.3)
PRINT 4109
4109 FORMAT(1HK,50X,4HVARC)

```

```

PRINT 399, (VARC(I), I=1, N)
399 FCRMAT(1X, 5F20.3)
PRINT 510
510 FCRMAT(1HK, 50X, 4FVARU)
PRINT 499, (VARC(I), I=1, N)
499 FCRMAT(1X, 5F20.3)
IF(NPXS.GT.0) GC TC 602
GO TO 603
602 PUNCH 705, (IHEAD(JJ), JJ=1, 20)
705 FCRMAT(20A4)
DC 608 I=1, M
608 PUNCH 601, (A(I, J), J=1, N)
601 FCRMAT(5F15.3)
603 IF(NPSS.GT.0) GO TO 604
GO TO 9000
604 PUNCH 705, (IHEAD(JJ), JJ=1, 20)
DU 606 I=1, M
606 PUNCH 607, (STDS(I, J), J=1, N)
607 FCRMAT(10F8.4)
9000 CONTINUE
IF(ITY.EQ.1.AND.INOT.EQ.1.AND.INO.EQ.1) GO TO 9002
IF(INO.EQ.1.AND.INOS.EQ.1) JJ=3
GO TO 9005
9002 JJ=2
GO TO 9005
9004 JJ=3
GC TC 9005
9001 JJ=1
9005 GO TO (9010, 9015, 9020), JJ
9010 IF(ICC-1) 9030, 9035, 9030
9035 ICSTY=1
GO TO 9015
9037 NACC=1
9030 REWIND 8
IF(ITT.EQ.1) ITY=1
WRITE(8) KK, M, N, ((NATR(I, J), J=1, KK), I=1, N), ((NOBS(I, J), J=1, KK),
+I=1, M), ((A(I, J), J=1, N), I=1, M), ITY, NACC, ICO, ILI, ICCMP
WRITE(6, 9050) ((A(I, J), J=1, N), I=1, M)
9050 FCRMAT(1X, 10F8.3)
WRITE(6, 9051) NACO, ICO, ILI, ICCMP
9051 FCRMAT(1X, 'UNITE8', 413)
GC TC 9040
9015 ICTY=1
REWIND 10
WRITE(10) KK, M, N, ((NATR(I, J), J=1, KK), I=1, N), ((NOBS(I, J), J=1, KK),
+I=1, M), ((A(I, J), J=1, N), I=1, M), ICTY, ICSTY, ITY
WRITE(6, 9060) ((A(I, J), J=1, N), I=1, M)
9060 FCRMAT(1X, 10F8.3)
WRITE(6, 9061) ICTY, ICST, ITY
9061 FCRMAT(1X, 'UNITE 10', 313)
GC TC 605
9020 IF(ICC-1) 9030, 9037, 9030
9040 RETURN
END

```

```

SUBROUTINE RPRINT(R, M, N, L)
C PRINTS OUT RECTANGULAR MATRICES
C ALL ROWS AND COLUMNS ARE APPROPRIATELY LABELLED
C R(I, J) = MATRIX TO BE PRINTED OUT
C M=WIDTH OF R. MAX. AS DIMENSIONED. NO. OF VARIABLES
C N=LENGTH OF R. MAX. AS DIMENSIONED. NO. OF PLACES

```

```

C      L=1 FOR FIG.4 OUTPUT
C      L=0 FOR FIG.2 OUTPUT
C      L.GT.1 FOR FIG.0 OUTPUT
C
C      DIMENSION J(30),R(30,30)
C      DIMENSION J(210),R(30,210)
C      DIMENSION J(210),R(30,210)
      NPAGE = 0
      MI=7
      M1 = 0
      M2 = 0
      JSEC = 0
      DO 8 I = 1,M
6      J(I) = I
9      M1 = M2 + 1
      M2 = M1 + MI
      IF(M2.LE.M) GO TO 13
      M2 = M
13     JSEC = JSEC + 1
      NI = 49
      N1 = 0
      N2 = 0
2      N1 = N2 + 1
      N2 = N1 + NI
      IF(N2.LE.N) GO TO 3
      N2 = N
3      NPAGE = NPAGE + 1
      WRITE(6,17) JSEC,NPAGE
17     FORMAT(1H0 2UX 9H SECTION      I3, 7H PAGE      I3/ )
      WRITE(6,22) (J(I), I = M1,M2)
22     FORMAT(' ', 'RANGEE', 2X, I3, 7I10)
      IF(L.GT.1) GO TO 301
      IF(L.LT.1) GO TO 201
      DC 23 I = N1,N2
23     WRITE(6,24) I, (R(I,K), K = M1,M2 )
24     FORMAT(1H ,I2, 8F10.3)
      GO TO 31
201    DC 29 I = N1,N2
29     WRITE(6,30) I, (R(I,K), K = M1,M2 )
30     FORMAT(1H ,I2, 8F10.3)
      GO TO 31
301    DO 302 I=N1,N2
302    WRITE(6,304) I, (R(I,K),K=M1,M2)
304    FORMAT(1HJ,I2, 8F10.3)
31     WRITE(6,1)
1      FORMAT(1H1)
      IF(N2.LT.N) GO TO 2
      IF(M2.LT.M) GO TO 9
      RETURN
      END

```

```

//LKED.SYSLMOD DD DSN=+MYLIB(MAIN),DISP=(NEW,PASS),UNIT=SYSDA,
//      SPACE=(1600,(160,50,1))
//GO.FT07F001 DD SYSCUT=B
//GC.FT06F001 DD SYSCUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=665)
//GO.FT08F001 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT08F001,DISP=(SHR,PASS)
//GC.FT10F001 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT10F001,DISP=(SHR,PASS)
//GC.FT11F001 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT11F001,DISP=(SHR,PASS)
//GC.FT05F001 DD *

```

```

//STEP3 EXEC FORTGCLC,PARM.FCRT='BCD',PARM.LKED='OVLY,LIST,XREF',
//          REGION.GO=225K,TIME.GO=3
//FCRT.SYSIN DD *                               SOURCE DECK NEXT
DIMENSION A(100,50)
C DIMENSION XBAR(250),STD(250),B(250),D(250),T(250)
C DIMENSION RX(250,250),R(40000)
C DO 160 I=1,250
  DIMENSION XBAR(74),STD(74),B(74),D(74),T(74)
  DIMENSION RX(74,74),R(3000)
  DIMENSION S(74,74),IVBLE(74)
  DIMENSION NATR(100,20),NOBS(100,20)
  DIMENSION X(500)
  DO 160 I=1,74
160 T(I)=0.000
  READ(8) KK,N,M,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+I=1,N),((A(I,J),J=1,M),I=1,N),ITY,NACC,IC0,ILI,ICOMP
510 REWIND 8
  WRITE(8) KK,N,M,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+I=1,N),((A(I,J),J=1,M),I=1,N),ICOMP
  IF(ICC.EQ.1.OR.NACC.EQ.1) GO TO 500
520 READ(10)KK,N,M,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+I=1,N),((A(I,J),J=1,M),I=1,N),ICTY,ICSTY,ITY
  WRITE(6,163) ((A(I,J),J=1,M),I=1,N)
  WRITE(6,164) ICTY,ICSTY,ITY
163 FCRMAT(10F12.3)
164 FCRMAT(3I4)
  ILI=1
  ITY=2
  GC TO 500
  1 FORMAT(2I4)
  2 FORMAT(1X,'VARIABLE          MOYENNE      ECART-TYPE ')
  3 FORMAT(4X,I2,F14.5,F12.5)
  5 FORMAT(1X,'RANGEE',I3/10F12.5)
500 L=0
  IPC=0
  CALL CORCOM (IPC,M,N,A)
  DO 161 J=1,M
  DO 161 I=1,N
  L=L+1
  X(L)=A(I,J)
161 CONTINUE
  CALL CORRE (N,M,1,X,XBAR,STD,RX,R,B,D,T)
  DO 200 I=1,M
  DO 210 J=1,M
  IF(I-J)202,205,205
202 JJ=I+(J-J)/2
  GO TO 210
205 JJ=J+(I-I)/2
210 T(J)=R(JJ)
200 WRITE(2,204) (T(J),J=1,M),I
204 FCRMAT(9F8.2/6F8.2,24X,I8)
  WRITE(6,2)
  DO 100 I=1,M
100 WRITE(6,3) I,XBAR(I),STD(I)
  WRITE(6,9)
  9 FCRMAT('1',57X,'MATRICE DE CORRELATION'/58X,22('='))//
  I1=1
  30 I2=I1+9
  WRITE(6,10)
  10 FCRMAT('1',1X,I30('='))

```

```

WRITE(6,11)
11 FCRMAT(1X,'I',8X,'I',10(11X,'I'))
WRITE(6,12)
12 FCRMAT(1X,'I RANGE E I',10(' COLONNE I'))
WRITE(6,13) (J,J=11,12)
13 FORMAT(1X,'I',8X,'I',10(4X,I3,4X,'I'))
WRITE(6,11)
WRITE(6,8)
8 FORMAT(1X,130('='))
WRITE(6,11)
DC 150 I=1,M
DO 140 J=1,M
IF(I-J) 105,130,130
105 K=1+(J-J-I)/2
GC TC 140
130 K=J+(I-I-1)/2
140 T(J)=R(K)
WRITE(6,21) I,(T(J),J=11,12)
21 FCRMAT(1X,'I ',I3,' I',10(2X,F6.3,' I'))
150 WRITE(6,11)
WRITE(6,8)
I1=I2+1
IF(I1.GT.M)GOTO 50
WRITE(6,25)
25 FCRMAT(////)
GO TO 30
50 DC 320 I=1,74
DO 320 J=1,74
320 S(I,J)=0.0
REWIND 11
JPHI=1
WRITE(11) M,M,JPHI
DC 300 I=1,M
DO 311 J=1,M
IF(I-J)305,330,330
305 K=I+(J-J-I)/2
GO TO 310
330 K=J+(I-I-1)/2
310 S(I,J)=R(K)
WRITE(13) R(K)
311 CONTINUE
IF(ITY.EQ.2) WRITE(11) I,(S(I,J),J=1,M)
300 CONTINUE
CALL CRDRE (M,M,S)
IF(ITY.EQ.1) GC TC 520
IF(ITY.EQ.2) RETURN
400 RETURN
END
SUBROUTINE DATA (M,D,K)
DIMENSION D(74)
DIMENSION D(250)
RETURN
END
SUBROUTINE CURRE (N,M,IC,X,XBAR,STD,RX,R,B,D,T)
DIMENSION X(1),XBAR(1),STD(1),KX(1),R(1),B(1),D(1),T(1)
JJM=0
DO 100 J=1,M
E(J)=0.0
100 T(J)=0.0
K=(M*M+M)/2

```

```

00003210
00003220
00003240
00003250
00003260
00003270
00003280

```

DC 102 I=1,K	00003290
102 R(I)=0.0	00003300
FN=N	00003310
L=0	00003320
IF(10) 105, 127, 105	00003330
105 DO 108 J=1,M	00003340
DC 107 I=1,N	00003350
L=L+1	00003360
107 T(J)=T(J)+X(L)	00003370
XBAR(J)=T(J)	00003380
108 T(J)=T(J)/FN	00003390
DC 115 I=1,N	00003400
JK=0	00003410
L=I-N	00003420
DC 110 J=1,M	00003430
L=L+N	00003440
C(J)=X(L)-T(J)	00003450
110 E(J)=E(J)+C(J)	00003460
DO 115 J=1,M	00003470
DO 115 K=1,J	00003480
JK=JK+1	00003490
115 R(JK)=R(JK)+D(J)*C(K)	00003500
GO TO 205	00003510
127 IF(N-M) 130, 130, 135	00003520
130 KK=N	00003530
GC TG 137	00003540
135 KK=M	00003550
137 DO 140 I=1,KK	00003560
CALL DATA(M,D,JJM)	00003570
DC 140 J=1,M	00003580
T(J)=T(J)+D(J)	00003590
L=L+1	00003600
140 RX(L)=D(J)	00003610
FKK=KK	00003620
DO 150 J=1,M	00003630
XBAR(J)=T(J)	00003640
150 T(J)=T(J)/FKK	00003650
L=0	00003660
DC 180 I=1,KK	00003670
JK=0	00003680
DC 170 J=1,M	00003690
L=L+1	00003700
170 D(J)=RX(L)-T(J)	00003710
DC 180 J=1,M	00003720
B(J)=B(J)+D(J)	00003730
DO 180 K=1,J	00003740
JK=JK+1	00003750
180 R(JK)=R(JK)+D(J)*C(K)	00003760
IF(N-KK) 205, 205, 185	00003770
185 KK=N-KK	00003780
DO 200 I=1,KK	00003790
JK=0	00003800
CALL DATA(M,D,JJM)	00003810
DO 190 J=1,M	00003820
XBAR(J)=XBAR(J)+D(J)	00003830
D(J)=D(J)-T(J)	00003840
190 B(J)=B(J)+D(J)	00003850
DC 200 J=1,M	00003860
DC 200 K=1,J	00003870
JK=JK+1	00003880

200	R(JK)=R(JK)+D(J)*E(K)	00003890
205	JK=C	00003900
	DC 210 J=1,M	00003910
	XBAR(J)=XBAR(J)/FN	00003920
	DC 210 K=1,J	00003930
	JK=JK+1	00003940
210	R(JK)=R(JK)-B(J)*B(K)/FN	00003950
	JK=C	00003960
	DC 220 J=1,M	00003980
	JK=JK+J	00003990
220	STD(J)=SQRT(AES(R(JK)))	00004000
	DC 230 J=1,M	00004010
	DC 230 K=J,M	00004020
	JK=J+(K*K-K)/2	00004030
	L=M*(J-1)+K	00004040
	RX(L)=R(JK)	00004050
	L=M*(K-1)+J	00004060
	RX(L)=R(JK)	00004070
	IF(STD(J)*STD(K)) 225, 222, 225	00004080
222	R(JK)=0.0	00004090
	GC TO 230	00004100
225	R(JK)=R(JK)/(STD(J)*STD(K))	00004110
230	CONTINUE	00004120
	FN=SQRT(FN-1.0)	00004130
	DC 240 J=1,M	00004140
240	STD(J)=STD(J)/FN	00004150
	L=-M	00004160
	DO 250 I=1,M	00004170
	L=L+M+1	00004180
250	B(I)=RX(L)	00004190
	RETURN	00004200
	END	
	SLBRoutine ORDRE(M,N,A)	
	DIMENSION IVBL(74,74)	
	DIMENSION B(74,74)	
	DIMENSION A(74,74),IVBLE(74)	
	DO 700 I=1,74	
	DC 700 J=1,74	
	IVBL(I,J)=0	
700	B(I,J)=0.0	
	DC 850 KK=1,M	
	DC 900 II=1,N	
900	IVBLE(II)=II	
	DC 840 I=1,N	
	J=I+1	
	IF(J.GT.N) GO TO 840	
	DC 801 L=J,N	
	IF(A(I,KK).GE.A(L,KK)) GOTO 801	
	TEMP=A(I,KK)	
	A(I,KK)=A(L,KK)	
	A(L,KK)=TEMP	
	ITEMP=IVBLE(I)	
	IVBLE(I)=IVBLE(L)	
	IVBLE(L)=ITEMP	
801	IVBL(I,KK)=IVBLE(I)	
840	B(I,KK)=A(I,KK)	
	IVBL(I,KK)=IVBLE(I)	
850	CONTINUE	
	I1=1	
31	I2=I1+4	


```

DU 200 I=1,N1
IK=I+1
SX1=S(I)
SX2=S2(I)
CC 200 J=IK,N
SXY=SX(K)
SY=S(J)
SY2=S2(J)
RM=M
R=(RM*SXY-SX1*SY)/SQRT ((RM*SX2-SX1**2)*(RM*SY2-SY**2))
DATASIGL/2H.1,3H.05,3H.02,3F.01,4H.001,4HNTSG,4HNOJF/
IF(RM-2.) 161,161,162
161 SIG=SIGL(7)
T=0.
GO TO 500
162 IF(R.EQ.1.0) GO TO 163
IF(R.EQ.(-1.0)) GO TO 164
T=R*SQRT ((RM-2.)/(1.-R**2))
TMCD=ABS(T)
GO TO 165
163 T=99999.9999
SIG=SIGL(5)
GO TO 500
164 T=-99999.9999
SIG=SIGL(5)
GO TO 500
165 NDF= RM-2.
CATAV(1,1),V(1,2),V(1,3),V(1,4),V(1,5)/6.314,12.706,31.821,63.657,
1 636.619/
CATAV(2,1),V(2,2),V(2,3),V(2,4),V(2,5)/2.920,4.303,6.965,9.925,
1 31.598/
CATAV(3,1),V(3,2),V(3,3),V(3,4),V(3,5)/2.353,3.182,4.541,5.814,
1 12.941/
CATAV(4,1),V(4,2),V(4,3),V(4,4),V(4,5)/2.132,2.776,3.747,4.604,
1 8.610/
CATAV(5,1),V(5,2),V(5,3),V(5,4),V(5,5)/2.015,2.571,3.365,4.032,
1 6.859/
CATAV(6,1),V(6,2),V(6,3),V(6,4),V(6,5)/1.943,2.447,3.143,3.707,
1 5.959/
CATAV(7,1),V(7,2),V(7,3),V(7,4),V(7,5)/1.895,2.365,2.998,3.499,
1 5.405/
CATAV(8,1),V(8,2),V(8,3),V(8,4),V(8,5)/1.860,2.306,2.896,3.355,
1 5.041/
CATAV(9,1),V(9,2),V(9,3),V(9,4),V(9,5)/1.833,2.262,2.821,3.250,
1 4.781/
CATAV(10,1),V(10,2),V(10,3),V(10,4),V(10,5)/1.812,2.228,2.764,
1 3.169,4.587/
CATAV(11,1),V(11,2),V(11,3),V(11,4),V(11,5)/1.796,2.201,2.718,
1 3.106,4.437/
CATAV(12,1),V(12,2),V(12,3),V(12,4),V(12,5)/1.782,2.179,2.681,
1 3.055,4.318/
CATAV(13,1),V(13,2),V(13,3),V(13,4),V(13,5)/1.771,2.160,2.650,
1 3.012,4.221/
CATAV(14,1),V(14,2),V(14,3),V(14,4),V(14,5)/1.761,2.145,2.624,
1 2.977,4.140/
CATAV(15,1),V(15,2),V(15,3),V(15,4),V(15,5)/1.753,2.131,2.602,
1 2.947,4.073/
CATAV(16,1),V(16,2),V(16,3),V(16,4),V(16,5)/1.746,2.120,2.583,
1 2.921,4.015/
CATAV(17,1),V(17,2),V(17,3),V(17,4),V(17,5)/1.740,2.110,2.567,

```

```

1      2.898,3.963/
  CATAV(18,1),V(18,2),V(18,3),V(18,4),V(18,5)/1.734,2.101,2.552,
1      2.878,3.922/
  CATAV(19,1),V(19,2),V(19,3),V(19,4),V(19,5)/1.729,2.093,2.539,
1      2.861,3.883/
  CATAV(20,1),V(20,2),V(20,3),V(20,4),V(20,5)/1.725,2.086,2.528,
1      2.845,3.850/
  CATAV(21,1),V(21,2),V(21,3),V(21,4),V(21,5)/1.721,2.080,2.518,
1      2.831,3.815/
  CATAV(22,1),V(22,2),V(22,3),V(22,4),V(22,5)/1.717,2.074,2.508,
1      2.819,3.792/
  CATAV(23,1),V(23,2),V(23,3),V(23,4),V(23,5)/1.714,2.069,2.500,
1      2.807,3.767/
  CATAV(24,1),V(24,2),V(24,3),V(24,4),V(24,5)/1.711,2.064,2.492,
1      2.797,3.745/
  CATAV(25,1),V(25,2),V(25,3),V(25,4),V(25,5)/1.708,2.060,2.485,
1      2.787,3.725/
  CATAV(26,1),V(26,2),V(26,3),V(26,4),V(26,5)/1.706,2.056,2.479,
1      2.779,3.707/
  CATAV(27,1),V(27,2),V(27,3),V(27,4),V(27,5)/1.703,2.052,2.473,
1      2.771,3.690/
  CATAV(28,1),V(28,2),V(28,3),V(28,4),V(28,5)/1.701,2.048,2.467,
1      2.763,3.674/
  CATAV(29,1),V(29,2),V(29,3),V(29,4),V(29,5)/1.699,2.045,2.462,
1      2.756,3.659/
  CATAV(30,1),V(30,2),V(30,3),V(30,4),V(30,5)/1.697,2.042,2.457,
1      2.750,3.646/
  CATAV(40,1),V(40,2),V(40,3),V(40,4),V(40,5)/1.684,2.021,2.423,
1      2.704,3.551/
  CATAV(60,1),V(60,2),V(60,3),V(60,4),V(60,5)/1.671,2.000,2.390,
1      2.660,3.400/
  CATAV(120,1),V(120,2),V(120,3),V(120,4),V(120,5)/1.658,1.980,2.358
1      ,20.617,3.373/
  CATAV(200,1),V(200,2),V(200,3),V(200,4),V(200,5)/1.645,1.960,2.326
1      ,2.576,3.29/
  IF(NDF=30) 301,301,302
301 GC TC 400
302 IF(NDF=200)303,304,305
305 NDF=200
304 GC TC 400
303 IF(NDF=40)306,307,308
308 NDF=30
307 GC TC 400
306 IF(NDF=60)309,310,311
311 NDF=40
310 GC TC 400
309 IF(NDF=120)312,400,313
312 NDF=60
  GC TC 400
313 NDF=120
400 IF(V(NDF,1)-TMCD)404,405,406
406 SIG=SIGL(6)
  GO TO 500
405 SIG=SIGL(1)
  GC TC 500
404 IF(V(NDF,5)-TMCD)407,407,408
407 SIG=SIGL(5)
  GO TC 500
408 IL=2
409 IF(V(NDF,IL)-TMCD)412,411,410

```

```

410 SIG=SIGL(IL-1)
GC TC 500
411 SIG=SIGL(IL)
GO TO 500
412 IL=IL+1
GC TC 409
500 IF(KN-48)30,71,71
71 WRITE(6,14)
14 FORMAT('1',40X,' CORRELATIONS AVEC TEST DE T'/113X,' NIVEAU'/
+1X,' VARIABLES',3X,' CORRELATION',7X,' SOMME X',4X,' SOMME X**2',
+7X,' SOMME Y',4X,' SOMME Y**2',5X,' SOMME X*Y',9X,' TEST DE T',1X,
+' DE SIG',3X,' OBSER.')
```

```

CORRCOMP
CORRCOMP
CORRCOMP
```

C

```

200 K=K+1
30 KN=KN+1
17 WRITE(6,13)I,J,R,S(I),S2(I),S(J),S2(J),SX(K),T,SIG,RM
13 FORMAT(1H,2I5,F14.6,5F14.0,3X,F13.5,2X,A6,F10.0)
JNK=JNK+1
RR(JNK)=R
```

```

CORRCOMP
```

```

IF(IPC.LE.0) GO TO 2010
XXX=1.00000
NK1=1
NN=N-1
DO 2007 J=1,NN
NK2=NK1+N-J-1
WRITE(2,2008) XXX,(RR(I),I=NK1,NK2)
2008 FORMAT(8F10.5)
NK1=NK2+1
2007 CONTINUE
PLNCH 2009,XXX
2009 FORMAT(F10.5)
2010 RETURN
```

```

END
SUBROUTINE RPRINT(R,M,N,L)
C PRINTS OUT RECTANGULAR MATRICES
C ALL ROWS AND COLUMNS ARE APPROPRIATELY LABELLED
C R(I,J) = MATRIX TO BE PRINTED OUT
C M=WIDTH OF R. MAX. AS DIMENSIONED. NO. OF VARIABLES
C N=LENGTH OF R. MAX. AS DIMENSIONED. NO. OF PLACES
C L=1 FOR F10.4 OUTPUT
C L=0 FOR F10.2 OUTPUT
C L.GT.1 FOR F10.0 OUTPUT
```

```

DIMENSION J(50), R(100,50)
NPAGE = 0
MI = 9
M1 = 0
M2 = 0
JSEC = 0
DO 8 I = 1,M
8 J(I) = I
9 M1 = M2 + 1
M2 = M1 + MI
IF(M2.LE.M) GO TO 13
M2 = M
13 JSEC = JSEC + 1
NI = 49
```

```

N2 = 0
2 N1 = N2 + 1
N2 = N1 + N1
IF(N2.LE.N) GO TO 3
N2 = N
3 NPAGE = NPAGE + 1
WRITE(6,17) JSEC,NPAGE
17 FCRMAT(1H0 20X 9F SECTION I3, 7H PAGE I3/ )
WRITE(6,22) (J(I), I = M1,M2)
22 FORMAT(1X,'RANGEE',2X,10I10)
IF(L.GT.1) GO TO 301
IF(L.LT.1) GO TO 201
DO 23 I = N1,N2
23 WRITE(6,24) I, (R(I,K), K = M1,M2 )
24 FCRMAT(1H I5, 5X 10F10.4 )
GO TO 31
201 DO 29 I = N1,N2
29 WRITE(6,30) I, (R(I,K), K = M1,M2 )
30 FCRMAT(1H I5, 4X 10F10.2 )
GO TO 31
301 DO 302 I=N1,N2
302 WRITE(6,304) I,(R(I,K),K=M1,M2)
304 FCRMAT(1HJ,I5,5X,10F10.0)
31 WRITE(6,1)
1 FCRMAT(1H1)
IF(N2.LT.N) GO TO 2
IF(M2.LT.M) GO TO 9
RETURN
END

```

```

//LKED.SYSLMOD DD DSNNAME=+GUYLE(MAIN),DISP=(NEW,PASS),UNIT=SYSDA,
// SPACE=(1600,(160,50,1))

```

```

//LKED.SYSIN DD *
INSERT MAIN
OVERLAY 1
INSERT CORCOM,RPRINT
OVERLAY 1
INSERT CORRE,DATA
OVERLAY 1
INSERT CRDRE

```

```

//GO.FT06F001 DD SYSCUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=1330)
//GC.FT13F001 DD UNIT=SYSDA,DSN=+TEMP5,DISP=(NEW,PASS),
// SPACE=(TRK,(30,1)),DCB=(RECFM=VBS,LRECL=524,BLKSIZE=3148)
//GC.FT08F001 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT08F001,DISP=(SHR,PASS)
//GC.FT10F001 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT10F001,DISP=(SHR,PASS)
//GC.FT11F001 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT11F001,DISP=(SHR,PASS)
//GC.FT02F001 DD DUMMY,DCB=(BLKSIZE=80,BUFNO=1)
//GC.SYSIN DD *

```

DATA NEXT

```

//STEP4 EXEC FORTGCLG,PARM.FORT='BCD',REGION.GO=200K,TIME.GO=3
//FCRT.SYSIN DD *
SOURCE DECK NEXT

```

```

C
C PHI COEFFICIENT OF ASSOCIATION BETWEEN BINARY DATA VECTORS
C ORIGINALLY SET UP FOR CONNECTIVITY MATRICES (SQUARE)
C
C
C PRINT OF FINAL PHI COEFFICIENT MATRIX
C ALSO PUNCHES OUT THE PHI MATRIX
C IF PUNCH OF PHI MATRIX IS DESIRED LET IPUNCH EQUAL 0

```

```

C      DIMENSION PHI(74,74),C(74,74),JAREA(74)
C      CCMCN/CURD/CO(74,2),C(74,74),DIST(74,74),AA(74),IKONT,RINDEX(74)
      READ(11) N,M,JPHI
101  FORMAT(3I4)
      IF(JPHI.GE.1) GO TO 105
      DC 5 I=1,74
1    FORMAT(74F1.0,1X,12)
      READ(9,1)(C(I,J),J=1,74),JAREA(I)
      5 CONTINUE
C
      DC 9 I=1,N
      DO 9 J=1,M
9    PHI(I,J)=0.0
      DC 7 J=1,M
      DO 7 L=J,M
      ID=0
      IC=0
      IA=0
      IB=0
      DC 6 I=1,N
      IF(C(I,J).EQ.0..AND.C(I,L).EQ.0.) ID=ID+1
      IF(C(I,J).EQ.0..AND.C(I,L).EQ.1.) IC=IC+1
      IF(C(I,J).EQ.1..AND.C(I,L).EQ.0.) IB=IB+1
      IF(C(I,J).EQ.1..AND.C(I,L).EQ.1.) IA=IA+1
6    CONTINUE
      A=FLCAT(IA)
      B=FLOAT(IB)
      E=FLJAT(IC)
      D=FLOAT(ID)
      ANUM=(A*D)-(B*E)
      DENOM=SQRT(((A+B)*(E+D)*(A+E)*(B+D)))
      P=ANUM/DENOM
      IF(P.EQ.0) P=.0001
      PHI(J,L)=P
      PHI(L,J)=PHI(J,L)
      PHI(J,J)=1.00000
      7 CONTINUE
C      CALL RPRINT (PHI,M,N,1)
100  CONTINUE
      GO TO 120
105  DC 110 I=1,N
      READ(11) JAREA(I),(PHI(I,J),J=1,M)
      WRITE(6,111) JAREA(I),(PHI(I,J),J=1,M)
111  FORMAT(1X,'VARIABLE',13/(10F12.5))
110  CONTINUE
120  CALL LINANA(N,NTIME,PHI)
      STCP
      END
      SUBROUTINE RPRINT(R,M,N,L)
C      PRINTS CUT RECTANGULAR MATRICES
C      ALL ROWS AND COLUMNS ARE APPROPRIATELY LABELLED
C      R(I,J) = MATRIX TO BE PRINTED OUT
C      M=WIDTH OF R. MAX. AS DIMENSIONED. NO. OF VARIABLES
C      N=LENGTH OF R. MAX. AS DIMENSIONED. NO. OF PLACES
C      L=1 FOR F10.4 OUTPUT
C      L=0 FOR F10.2 OUTPUT
C      L.GT.1 FOR F10.0 OUTPUT
C
      DIMENSION J(74),K(74,74)

```

```

NPAGE = 0
MI = 9
M1 = 0
M2 = 0
JSEC = 0
DO 8 I = 1,M
8 J(I) = I
9 M1 = M2 + 1
M2 = M1 + MI
IF(M2.LE.M) GO TO 13
M2 = M
13 JSEC = JSEC + 1
N1 = 49
N1 = 0
N2 = 0
2 N1 = N2 + 1
N2 = N1 + NI
IF(N2.LE.N) GO TO 3
N2 = N
3 NPAGE = NPAGE + 1
WRITE(6,17) JSEC,NPAGE
17 FCRMAT(1H0 20X 9F SECTION I3, 7H PAGE I3/ )
WRITE(6,22) (J(I), I = M1,M2)
22 FCRMAT(6H0 ROW 3X 10I10 )
IF(L.GT.1) GO TO 301
IF(L.LT.1) GO TO 201
DC 23 I = N1,N2
23 WRITE(6,24) I, (R(I,K), K = M1,M2 )
24 FCRMAT(1H 15, 5X 10F10.4 )
GC TO 31
201 DC 29 I = N1,N2
29 WRITE(6,30) I, (R(I,K), K = M1,M2 )
30 FCRMAT(1H 15, 4X 10F10.2 )
GC TO 31
301 DO 302 I=N1,N2
302 WRITE(6,304) I, (R(I,K),K=M1,M2)
304 FCRMAT(1HJ,15,5X,10F10.0)
31 WRITE(6,1)
1 FCRMAT(1HL)
IF(N2.LT.N) GC TO 2
IF(M2.LT.M)GO TO 9
RETURN
END
SUBROUTINE LINANA(N,NTIME,CCR)
C
C
C DIMENSION CDR(74,74),HICOL(74,3),TYPE(74,74),LIST(74),TFCT(74)
C DIMENSION MEME(74),SCOR(74,74),B(74),MCQUIT(74)
C COMMON S(74,74)
C REAL MCQUIT
C
C
C LINKAGE ANALYSIS PERFORMED ON THE PHI MATRIX. THE METHOD USED IS
C DERIVED FROM L. L. MCQUITTY, ' EDUCATIONAL AND PSYCHOLOGICAL
C MEASUREMENT ', VOL 17, 1957.
C
C
C NCTY=0
C
C FINDING HIGHEST CCLUMN VALUE

```

```

C
DO 3 J=1,N
FICCL(J,1)=0.
DC 3 I=1,N
IF(I.EQ.J) GO TO 3
IF(CUR(I,J).LE.FICOL(J,1)) GO TO 3
5 HICCL(J,1)=COR(I,J)
FICOL(J,2)=FLOAT(I)
FICCL(J,3)=FLOAT(J)
3 CONTINUE

C
C FINDING LARGEST VALUE OF MATRIX
C
110 DO 6 I=1,N
IF(I.EQ.1) GO TO 7
IF(HICOL(I,1).LE.RN) GO TO 6
7 RN=FICOL(I,1)
RNCI=FICOL(I,2)
RNCJ=FICOL(I,3)
6 CONTINUE
DC 8 I=1,N
IF(HICOL(I,1).EQ.0.0) GO TO 8
GO TO 9
8 CONTINUE
GO TO 600

C
C FINDING MEMBERS FOR EACH TYPE
C
9 ITEST=1
NUM=2

C
C DETERMINE TYPES
C
ACTY=NOTY+1
K=NCTY
TYPE(K,1)=RNCI
TYPE(K,2)=RNCJ
NOI=RNCI
NOJ=RNCJ

C
C ZERO OUT HICOL
C
DC 106 J=1,3
HICCL(NCI,J)=0.
HICCL(NOJ,J)=0.
106 CONTINUE

C
C FINDING FIRST ORDER NEIGHBOURS
C
DC 102 I=1,N
IF(HICOL(I,2).EQ.RNCI.OR.HICOL(I,2).EQ.RNOJ) GO TO 51
GO TO 102
51 ITEST=0
NUM=NUM+1
TYPE(K,NUM)=0.
TYPE(K,NUM)=HICCL(I,3)
DC 105 J=1,3
FICCL(I,J)=0.
105 CONTINUE
102 CONTINUE

```

```

      IF(ITEST.EQ.1) GC TC 99
      ITEST=1
C
C   FINDING SECOND AND HIGHER ORDER NEIGHBOURS
C
      DC 103 IJ=3,NUM
      RNCK=TYPE(K,IJ)
      DO 104 I=1,N
      IF(HICOL(I,2).EQ.RNUK) GO TO 52
      GC TC 104
52  ITEST=0
      NUM=NUM+1
      TYPE(K,NUM)=HICCL(I,3)
      DO 107 J=1,3
      HICCL(I,J)=0.
107  CCNTINUE
104  CONTINUE
103  CCNTINUE
      99 MEMB(K)=NUM
      GO TO 110
C
C   FORMING SUB MATRIX AND FINDING TYPAL RELEVANCIES
C
600  K=0
      DO 200 I=1,NOTY
      MEMBR=MEMB(I)
      JJ=0
      DO 201 J=1,MEMBR
      IS=TYPE(I,J)
      LL=0
      JJ=JJ+1
      DO 220 L=1,MEMBR
      IL=TYPE(I,L)
      LL=LL+1
      SCOR(JJ,LL)=0.
      SCOR(LL,JJ)=0.
      SCOR(JJ,LL)=COR(IS,IL)
220  SCOR(LL,JJ)=COR(IL,IS)
201  CCNTINUE
      NPUF=0
      NPIR=0
      NPUEV=0
      NDIAG=1
      NSQRE=0
      WRITE(6,21)I
21  FORMAT(1H1,10X,'CORRELATION SUBMATRIX FOR GROUP',I5)
      CALL RPRINT(SCOR,LL,JJ,1)
      CALL FACTER (MEMBR,NPUF,NPIR,NPUEV,NDIAG,NSQRE,SCOR)
      WRITE(6,22)
22  FORMAT(1HJ,10X,'TYPAL RELEVANCIES')
      CALL RPRINT(S,LL,JJ,1)
C
C   CALCULATE MCQUIT ARRAY
C
      DO 202 J=1,MEMBR
      E(J)=0.
      B(J)=TYPE(I,J)+(FLOAT(I)*100.)
      TFCT(J)=S(J,1)
202  CCNTINUE
      NBB=MEMBR

```

```

IAS=-1
NB=1
CALL SORTR(MEMBR,TFCT,B,NBB,NB,IAS,LIST)
WRITE(6,302)
302 FCRMAT(1HK)
WRITE(6,25)
25 FCRMAT(1HJ,10X,'GROUP MEMBERS')
DO 203 J=1,MEMBR
K=K+1
IF(TFCT(J)) 210,211,211
211 MCQUIT(K)=TFCT(J)+B(J)
GO TO 204
210 MCQUIT(K)=TFCT(J)-B(J)
204 WRITE(6,301)MCQUIT(K)
301 FCRMAT(1HK,F15.5)
203 CONTINUE
200 CONTINUE
RETURN
END
SLBROUTINE FACTER (N,NPUF,NFIR,NPUEV,NDIAG,NSQRE,R)
C
C PRINCIPAL COMPONENTS ANALYSIS
C PROGRAMMED ORIGINALLY IN MCGILL UNIVERSITY PSYCHOLOGY DEPARTMENT
C AUTHOR CHUCK CRAWFORD FOR PH.D THESIS ON IMAGE ANALYSIS (1966)
C RESULTS ARE PRINTED AND CAN BE PUNCHED IF NPUF > 0
C BY EITHER ROWS OR COLUMNS ON FORMAT 8F9.4.
C N=NUMBER OF VARIABLES
C DEFINITION OF ARGUMENTS
C R(I,J) = CORRELATION MATRIX
C S(I,J)= OUTPUT OF FACTOR LOADINGS
C RA=ACCURACY KRUN=1(DESIGNATES EIGENVECTORS)
C
C CHANGED BY P. STROBACH (1968)
C
DIMENSION R(74,74),G(74,74),S2(74),SSX(74)
COMMON S(74,74)
EQUIVALENCE (S(1,1),G(1,1))
C
130 FCRMAT(8F9.4)
DC 21 I = 1, N
DC 21 J = 1, N
S(I,J) = 0.0
G(I,J) = 0.0
21 S2(I) = 0.0
L = 1
KRUN=1
RA = .0001
DO 23 I = 1, N
DC 23 J = 1, N
23 G(I,J) = R(I,J)
45 CALL EVALUE(R,74,74,S,74,74,N,RA,KRUN)
DC 29 I = 1, N
DO 29 J = 1, N
29 IF(I.EQ.J) S2(I) = R(I,J)
DC 2 J = 1, N
NS = J + 1
IF(NS.GT.N) GO TO 2
DC 2 K = NS, N
IF(ABS(S2(J)).GT.ABS(S2(K)))GO TO 2
CON13880
CON13890
CON13900
CON13920
CON13930
CON13940
CON13960
CON13970
CON13980
CON14000
CON14070
CON14030
CON14040
CON14080
CON14090
CON14100
CON14110
CON14120
CON14150
CON14160
CON14170
CON14180
CON14190
CON14200
CON14230
CON14240
CON14250
CON14260
CON14270
CON14280
CON14290

```

TEMP = S2(K)	CON14300
S2(K) = S2(J)	CON14310
S2(J) = TEMP	CON14320
DC 3 I = 1, N	CON14330
TEMP = S(I,K)	CON14340
S(I,K) = S(I,J)	CON14350
3 S(I,J) = TEMP	CON14360
2 CONTINUE	CON14370
IF(NPUEV.EQ.0) GO TO 44	CON14410
DO 47 J = 1, N	CON14420
47 WRITE(7,130) (S(I,J),I = 1, N)	CON14430
44 CONTINUE	CON14440
DC 31 J = 1, N	CON14450
A = SGRT(ABS(S2(J)))	CON14460
DC 31 I = 1, N	CON14470
XX = S(I,J)*A	CON14480
31 S(I,J) = XX	CON14490
DC 4840 J = 1, N	CON14530
4840 SSX(J) = 0.0	CON14540
DC 5 J = 1, N	CON14550
DO 5 I = 1, N	CON14560
5 SSX(J) = SSX(J) + S(I,J)**2	CON14570
IF(NPUF) 900, 915, 901	CON14610
901 DO 912 J = 1, N	CON14620
912 WRITE(7,130) (S(I,J),I = 1, N)	CON14630
IF(NPUF.NE.(-1)) GO TO 915	CON14640
900 DO 916 I = 1, N	CON14650
916 WRITE(7,130) (S(I,J),J = 1, N)	CON14660
915 RETURN	CON14670
END	CON14680
SUBROUTINE SORTR (N,A,B,NBB,NB,IAS,LIST)	00000010
C	00000020
C THIS SUBROUTINE SORTS NUMERICAL DATA IN EITHER ASCENDING	
C CR DESCENDING ORDER	
C	
C PROGRAMMED BY G.RATZER (MAY.1966)	
C REVISED - P.G. LAMB (JUNE.1967)	
C MCGILL COMPUTING CENTER	
C	00000270
C DIMENSION A(1),E(1),LIST(1)	00000280
C REAL LIST	
C	00000260
C M=N	00000300
DO 1 K=1,N	00000310
1 LIST(K)=A(K)	00000320
IF(IAS)98,88,88	00000330
98 M=(M+2)/3	00000340
MP1=M+1	00000350
DC 99 I=MP1,N	00000360
IM=I-M	00000370
IP1=I+1	00000380
DO 97 L=1,IM,M	00000390
J=IP1-L	00000400
J1=J-M	00000410
IF(LIST(J1)-LIST(J))70,99,99	00000420
70 SAVE=LIST(J1)	00000430
LIST(J1)=LIST(J)	00000440
LIST(J)=SAVE	00000450
97 CONTINUE	00000460
99 CONTINUE	00000470

	IF(M-1)101,101,96	00000480
88	M=(M+2)/3	00000490
	MP1=M+1	00000500
	DO 89 I=MP1,N	00000510
	IM=I-M	00000520
	IP1=I+1	00000530
	DO 87 L=1,IM,M	00000540
	J=IP1-L	00000550
	J1=J-M	00000560
	IF(LIST(J1)-LIST(J))89,89,71	00000570
71	SAVE=LIST(J1)	00000580
	LIST(J1)=LIST(J)	00000590
	LIST(J)=SAVE	00000600
87	CONTINUE	00000610
89	CONTINUE	00000620
	IF(M-1)101,101,88	00000630
101	DO 11 I=1,N	00000640
	DO 10 J=I,N	00000650
	IF(LIST(I)-A(J))10,12,10	00000660
10	CONTINUE	00000670
12	DO 86 K=1,NB	00000680
	INDEX1=(K-1)*NBB+I	00000690
	INDEX2=(K-1)*NBB+J	00000700
	SAVE=B(INDEX1)	00000710
	B(INDEX1)=B(INDEX2)	00000720
86	B(INDEX2)=SAVE	00000730
	SAVE=A(I)	00000740
	A(I)=A(J)	00000750
	A(J)=SAVE	00000760
11	CONTINUE	00000770
	RETURN	00000780
	END	00000790
	SLBRoutine EVALUE(A,IL,IJ,S,IM,IN,N,RA,KRUN)	CON11980
C		CON11990
C	CALCULATES EIGEN VALUES AND EIGEN VECTORS	CON12000
C	FOR A REAL SYMMETRIC MATRIX	CON12010
C		CON12020
C	SOURCE- 'MATHEMATICAL METHODS FOR DIGITAL COMPUTERS'	CON12030
C	A.RALSTON + F.WILF, P.84	CON12040
C		CON12050
C	PROGRAMMED BY MRS. U.B.MANLEY (JUNE.1964)	CON12060
C	MCGILL COMPUTING CENTER	CON12070
C		CON12080
	DIMENSION A(1),S(1)	CON12090
	DC 5 I=1,N	CON12100
	DC 5 J=1,N	CON12110
	M=(J-1)*IM+I	CON12120
5	S(M)=0.	CON12130
	DO 6 I=1,N	CON12140
	M=(I-1)*IM+I	CON12150
6	S(M)=1.	CON12160
	IND=0	CON12170
	VI=0.	CON12180
	FN=N	CON12190
	DO 10 I=1,N	CON12200
	M=I-1	CON12210
	IF(M)10,90,11	CON12220
11	DO 8 K=1,M	CON12230
	II=(K-1)*IL+I	CON12240
8	VI=VI+A(II)**2	CON12250

90	M=I+1	CON12260
	IF(M-N)4,4,10	CON12270
4	DO 5 K=M,N	CON12280
	II=(K-1)*IL+I	CON12290
5	VI=VI+A(II)**2	CON12300
10	CONTINUE	CON12310
	VI=SQRT(VI)	CON12320
	IF(VI)1,145,1	CON12330
1	VE=RA*VI/FN	CON12340
106	VI=VI/FN	CON12350
107	I=2	CON12360
108	J=1	CON12370
109	II=(I-1)*IL+J	CON12380
	IF(ABS(A(II))-VI)122,110,110	CON12390
110	IND=1	CON12400
	II=(I-1)*IL+I	CON12410
	JJ=(J-1)*IL+J	CON12420
	SMY=(A(JJ)-A(II))/2.	CON12430
	JJ=(I-1)*IL+J	CON12440
	SLAMB=-A(JJ)	CON12450
	CMEGA=SLAMB/SQRT(SLAMB**2+SMY**2)	CON12460
	IF(SMY)111,1111,1111	CON12470
111	OMEGA=-OMEGA	CON12480
1111	SI=CMEGA/SQRT(2.*(1.+SQRT(1.-OMEGA**2)))	CON12490
	CG=SQRT(1.-SI**2)	CON12500
	DO 114 L=1,N	CON12510
	LL=L-I	CON12520
	IF(LL)1113,1140,1113	CON12530
1113	LJ=L-J	CON12540
	IF(LJ)1133,1140,1133	CON12550
1133	II=(J-1)*IL+L	CON12560
	JJ=(I-1)*IL+L	CON12570
	III=(L-1)*IL+I	CON12580
	JJJ=(L-1)*IL+J	CON12590
	B1=A(II)	CON12600
	B2=A(JJ)	CON12610
	A(II)=B1*CO-B2*SI	CON12620
	A(JJ)=B1*SI+B2*CO	CON12630
	A(III)=A(JJ)	CON12640
	A(JJJ)=A(II)	CON12650
1140	IF(KRUN)2222,114,2222	CON12660
2222	II=(J-1)*IM+L	CON12670
	JJ=(I-1)*IM+L	CON12680
	S1=S(II)	CON12690
	S2=S(JJ)	CON12700
	S(II)=S1*CO-S2*SI	CON12710
	S(JJ)=S1*SI+S2*CO	CON12720
114	CONTINUE	CON12730
	II=(I-1)*IL+I	CON12740
	JJ=(J-1)*IL+J	CON12750
	JI=(I-1)*IL+J	CON12760
	IIJ=(J-1)*IL+I	CON12770
	B1=A(JJ)	CON12780
	B2=A(II)	CON12790
	B3=A(JI)	CON12800
	A(JJ)=B1*CO**2+B2*SI**2-2.*B3*SI*CO	CON12810
	A(II)=B1*SI**2+B2*CO**2+2.*SI*CO*B3	CON12820
	A(JI)=(B1-B2)*SI*CO+B3*(CO*CO-SI*SI)	CON12830
	A(IIJ)=A(JI)	CON12840
122	J2=J-(I-1)	CON12850

```

      IF(J2)1124,124,1124
1124 J=J+1
      GC TO 109
      124 N1=N-I
      IF(N1)125,120,125
      125 I=I+1
      GC TO 108
      126 IF(IND)127,128,127
      127 INL=0
      GC TO 107
      128 IF(VI-VE)130,130,106
      130 RETURN
      145 WRITE(6,140)
      146 FORMAT(1H0,21HINPUT MATRIX DIOGONAL)
      RETURN
      END

```

```

CON12860
CON12870
CON12880
CON12890
CON12900
CON12910
CON12920
CON12930
CON12940
CON12950
CON12960
CON12970
CON12980
CON12990
CON13000
CON13010

```

```

//LKED.SYSLMOD DD DSNAME=+RACIN(MAIN),DISP=(NEW,PASS),UNIT=SYSDA,
//                SPACE=(1600,(160,50,1))

```

```

//GC.FT06F001 DD SYSCUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=1330)
//GC.FT11F001 DD UNIT=SYSDA,DSN=*.STEP1.GU.FT11F001,DISP=(OLD,DELETE)
//GC.SYSIN DD *

```

```

//STEP5 EXEC FORTGCLG,PARM.FORT='BCD',PARM.LKED='OVLY,LIST,XREF',
//                REGION.GO=300K,TIME.GO=3

```

```

//FCRT.SYSIN DD *
COMMON SD(40,40)
DIMENSION XNAME(20)
DIMENSION SUMF(40)
DIMENSION D(40)
DIMENSION NATR(100,20),NOBS(150,20),AA(40,40)
  READ(8) KK,N,P,((NATR(I,J),J=1,KK),I=1,M),((NOBS(I,J),J=1,KK),
+I=1,N),((AA(I,J),J=1,M),I=1,N),ICOMP
  WRITE(6,22)((AA(I,J),J=1,M),I=1,N)
22 FORMAT(1X,10F8.3)
  CALL INNURM(N,M,AA)
  IF(ICOMP.EQ.0) GC TO 10
  READ(12)(XNAME(JJ),JJ=1,20),IFACTO,IROT,CON,NPUF,NPIR,NPUEV,
+NP,NCRM,IBUG,ICOMP
  NSQRE=1
  NSQ=0
  WRITE(6,1) IROT, N,M,CON, NPUF,NPIR,NPUEV,IFACTO,NSQRE,NP,NORM
1 FORMAT(3I2,F6.1,7I2)
  PR=0.0
  PRI=0.0
  CALL A(PR,PRI,N,M,CCN,IFACTO,K,IROT,R,SUMF)
  NF=K
  CALL B(M,NPUF,NPIR,NPUEV,IFACTO,NSQRE,NF,XNAME,CON,SUMF)
  REWIND 12
  WRITE(12)(XNAME(JJ),JJ=1,20),IFACTO,IROT,CON,NPUF,NPIR,NPUEV,
+NSQRE,NP,NORM,NSQ,IBUG,ICOMP,N,M,NF
  WRITE(6,2)(XNAME(JJ),JJ=1,20),IFACTO,IROT,CON,NPUF,NPIR,NPUEV,
+NSQRE,NP,NORM,NSQ,IBUG,ICOMP,N,M,NF
2 FORMAT(1X,20A4,2I2,F6.1,14I2)
  CALL C(NF,M,NF,NF,NP,NCRM,IROT,NPUF,XNAME)
10 RETURN
  END
  SUBROUTINE A(PR,PRI,N,M,CCN,IFACTO,K,IROT,R,SUMF)
C PROGRAMME DE L'ANALYSE EN COMPOSANTE PRINCIPALE, SOLUTION VARIMAX

```

```

C  SOURCE PRINCIPALE: IBM. SSP
C  ALAPTATION PAR GUY LEMAY UNIVERSITE D'OTTAWA
C  VERSICK NON FINALE ----- UNE SECONDE EN PREPARATION
C  PCCR RENSEIGNEMENTS : GUY LEMAY
C                          DEPARTEMENT DE GEOGRAPHIE
C                          UNIVERSITE D'OTTAWA
C                          OTTAWA, LNT.
C
C  R
C  PREMIERE CARTE
C      CCL 1-6 IDENTIFICATION DU PROBLEME
C      CCL 7-11 N=NCMBRE D'OBSERVATIONS
C      CCL 12-13 M= NOMBRE DE VARIABLES
C      CCL 14-19 CCN= "CUT-OFF" DES EIGENVALUES (F6.0)
C  DEUXIEME CARTE FURMAT DE LECTURE
C  LES DONNEES PAR LA SUITE
      CCMCN SD(40,40)
      DIMENSION STORE(10,10)
      DIMENSION B(40),C(40),S(40),T(40),XBAR(40)
      DIMENSION V(1600)
      DIMENSION FMT(20)
      DIMENSION F(40,40)
      DIMENSION R(820)
      DIMENSION SUMF(40)
C  INTEGER BLANC,ETCIL,BAR
      1 FCRMAT(21H)FACTOR ANALYSIS.....A4,A2//3X,12HNO. OF CASES,4X,16/3X,
      116HNC. OF VARIABLES,16/)
      2 FCRMAT(6H)MEANS/(8F15.5)
      3 FCRMAT(20H)STANDARD DEVIATIONS/(8F15.5)
      4 FCRMAT(25H)CORRELATION COEFFICIENTS)
      5 FCRMAT(1X,'VARIABLE',13/(12F10.4))
      6 FCRMAT(//2X,'EIGENVALUES'/(12F10.5))
      7 FCRMAT(1X,'POURCENTAGE CUMULATIF DES EIGENVALUES'/(12F10.5))
      8 FCRMAT(//2X,'EIGENVECTEURS')
      9 FCRMAT(1X,'VECTEUR',13/12F10.5/12F10.5)
     10 FCRMAT(//2X,'MATRICE DES SATURATIONS (' ,13, ' FACTEURS  )')
     11 FCRMAT(1X,'VARIABLE',13/12F10.5/12F10.5)
     12 FCRMAT(1H0/10H ITERATION,7X,9H)VARIANCES/8H CYCLE)
     13 FCRMAT(16,F20.6)
     14 FCRMAT(1H0/24H ROTATED FACTOR MATRIX (13,9H FACTORS))
     15 FCRMAT(9H)VARIABLEI3/(10F12.5))
     16 FCRMAT(1H0/23H CHECK ON COMMUNALITIES//9H VARIABLE,7X,8H)ORIGINAL,
     112X,5HFINAL,10X,10H)DIFFERENCE)
     17 FCRMAT(16,3F18.5)
     19 FCRMAT(5H)ONLY,12,30H FACTOR RETAINED. NO ROTATION)
    100 DO 120 I=1,M
          DC 110 J=1,M
          IF(I-J) 102, 104, 104
    102 L=I+(J*J-J)/2
          GC TC 109
    104 L=J+(I*I-I)/2
    109 READ(13) R(L)
          IF(I.NE.J.AND.R(L).EQ.1.0) R(L)=0.999
    110 D(J)=R(L)
          WRITE(4) (C(J),J=1,M)
          WRITE(2,22) (C(J),J=1,M)
     22 FORMAT(9F8.3/4F8.3)
    120 WRITE (6,5) I,(L(J),J=1,M)
    20 FCRMAT(10F12.5)

```

	K=0	
	IF(IFACTO.EQ.3) GO TO 1000	
	IF(IFACTO.GT.1) GO TO 188	
1000	MV=0	00000580
	CALL EIGEN (R,V,M,MV)	00000590
	CALL TRACE (M,R,CCN,K,D)	00000600
	DO 130 I=1,K	00000610
	L=I+(I*I-I)/2	00000620
130	S(I)=R(L)	00000630
	WRITE (6,6) (S(J),J=1,K)	00000640
	WRITE (6,7) (D(J),J=1,K)	00000650
	WRITE (6,8)	00000660
	L=0	00000670
	DO 150 J=1,K	00000680
	DO 140 I=1,M	00000690
	L=L+1	00000700
140	D(I)=V(L)	00000710
150	WRITE (6,9) J,(D(I),I=1,M)	00000720
	CALL LOAD (M,K,R,V)	00000730
	WRITE (6,10) K	00000740
	DO 180 I=1,M	00000750
	SUMF(I)=0.0	
	DO 170 J=1,K	00000760
	L=M*(J-1)+1	00000770
	D(J)=V(L)	00000780
	SUMF(I)=SUMF(I)+D(J)*D(J)	
170	F(I,J)=D(J)	00000800
	WRITE(4) (D(J),J=1,K)	
180	WRITE (6,11) I,(D(J),J=1,K)	00000830
	WRITE(6,181)	
181	FORMAT(/1X,'COMMUNAUTES')	
	WRITE(6,182)(SUMF(I),I=1,M)	
182	FORMAT(10F12.5/10F12.5)	
	WRITE(6,6000)	00000840
6000	FORMAT('1',T45,'MATRICE DES FACTEURS NON ROTATIONNES')	00000850
	WRITE(6,6001) ((F(I,J),J=1,K),I=1,M)	00000860
6001	FORMAT(10(1X,F10.5))	00000870
	CALL GRAPH(K,M,F)	
	IF(K-1) 185, 185, 188	00001480
185	WRITE (6,19) K	00001490
	IRCT=2	
188	RETURN	
	END	
C	SUBROUTINE INNGRM (N,M,A)	
	SUBROUTINE INNGRM(N,M)	00002370
	DIMENSION STORE(40,40)	
	DIMENSION D(40),AMEAN(40),STDEV(40),SUM(40),SUMSQ(40)	00002390
	DIMENSION A(40,40),IVBLE(131)	
	DIMENSION FMT(20)	
C	READ(5,21) (FMT(J),J=1,20)	
21	FORMAT(20A4)	
	DO 8 J=1,M	00002410
	SUM(J)=0.0	00002420
8	SUMSQ(J)=0.0	00002430
	WRITE(3,309)	00002440
309	FORMAT(1H1,'INFLI MATRIX')	00002450
	DO 15 J=1,N	00002460
C	READ(1,51) (D(I),I=1,M),TEMP	00002470
C	READ(5,FMT) (D(I),I=1,M)	
C	WRITE(6,501) (D(I),I=1,M),TEMP	00002500

C	WRITE(6,501) (C(I),I=1,M)	00002520
501	FORMAT(10F10.2)	
	DC 16 I=1,M	00002530
	SUM(I)=SUM(I)+A(J,I)	
	SUMSQ(I)=SUMSQ(I)+A(J,I)*A(J,I)	
16	STORE(J,I)=A(J,I)	
15	CONTINUE	00002580
	WRITE(6,100)	
100	FORMAT('1',30X,'MISE DES VARIABLES EN ORDRE DECROISSANT')	
	DO 850 KK=1,M	00002590
	WRITE(6,101) KK	
101	FORMAT(' ',10X,'VARIABLE',I3)	
	DO 900 II=1,N	00002600
900	IVBLE(II)=II	00002610
	DC 840 I=1,N	00002620
	J=I+1	00002630
	IF(J.GT.N) GO TO 840	00002640
	DC 801 L=J,N	00002650
	IF(A(I,KK).GE.A(L,KK)) GOTO 801	00002660
	TEMP=A(I,KK)	00002670
	A(I,KK)=A(L,KK)	00002680
	A(L,KK)=TEMP	00002690
	ITEMP=IVBLE(I)	00002700
	IVBLE(I)=IVBLE(L)	00002710
	IVBLE(L)=ITEMP	000 275
801	CONTINUE	00002730
840	WRITE(6,810) IVBLE(I),A(I,KK)	00002740
C	810 FORMAT(15,10X,F12.5)	00002750
	810 FORMAT(15,10X,F15.5)	00002760
	WRITE(6,9002)	00002770
C	DATA ELANC,ETOIL,BAR/' ','*', 'I'/	00000150
5002	FORMAT('1',110('**'))	00002780
850	CONTINUE	00002790
	FNS=N	00002800
	DC 10 J=1,M	00002810
	AMEAN(J)=SUM(J)/FNS	00002820
	T1=SUMSQ(J)/FNS	00002830
	T2=(SUM(J)**2)/(FNS**2)	00002840
	T3=T1-T2	00002850
10	STDEV(J)=SQRT(T3)	00002860
	WRITE(6,61)	00002870
61	FORMAT(1H1,5X,'INPUT MATRIX COLUMN MEANS')	00002880
	WRITE(6,62) (AMEAN(J),J=1,M)	00002890
C	62 FORMAT('0',12F10.4)	00002900
	62 FORMAT('0',5F15.4)	00002910
	WRITE(6,63)	00002920
63	FORMAT('-',5X,'INPUT MATRIX COLUMN STANDARD DEVS.')	00002930
	WRITE(6,62) (STDEV(J),J=1,M)	00002940
	WRITE(6,64)	00002950
64	FORMAT('1',5X,'MATRICE STANDARDISEE')	
	DC 12 I=1,N	00002970
	DU 13 J=1,M	00002980
13	STORE(I,J)=(STORE(I,J)-AMEAN(J))/STDEV(J)	00002990
C	WRITE(6,65) I	
65	FORMAT(' ',3X,'OBSERVATION',I5)	
	WRITE(6,66) (STORE(I,J),J=1,M)	00003020
	WRITE(7,20) (STORE(I,J),J=1,M)	
20	FORMAT(12F10.4)	
C66	FORMAT(' ',12F10.4)	00003030
66	FORMAT(' ',12F10.4)	

12	CONTINUE	00003050
99	RETURN	00003060
	END	00003070
	SUBROUTINE DATA(M,D,K)	00003080
	DIMENSION STORE(10,10)	
	DIMENSION D(40)	00003100
	K=K+1	00003110
	DC 15 J=1,M	00003120
15	D(J)=STORE(K,J)	00003130
	WRITE(7,61) (D(I),I=1,M)	00003140
61	FORMAT(12F10.1)	00003170
	61 FORMAT(12F10.4)	00003180
	RETURN	00003190
	END	00003200
	SUBROUTINE CORRE (N,M,IG,X,XBAR,STD,RX,R,B,D,T)	00003210
	DIMENSION X(1),XBAR(1),STD(1),RX(1),R(1),B(1),D(1),T(1)	00003220
	JJM=0	00003240
	DO 100 J=1,M	00003250
	E(J)=0.0	00003260
100	T(J)=0.0	00003270
	K=(M*M+M)/2	00003280
	DC 102 I=1,K	00003290
102	K(I)=0.0	00003300
	FN=N	00003310
	L=0	00003320
	IF(IG) 105, 127, 105	00003330
105	DO 108 J=1,M	00003340
	DC 107 I=1,N	00003350
	L=L+1	00003360
107	T(J)=T(J)+X(L)	00003370
	XBAR(J)=T(J)	00003380
108	T(J)=T(J)/FN	00003390
	DC 115 I=1,N	00003400
	JK=0	00003410
	L=L-N	00003420
	DC 110 J=1,M	00003430
	L=L+N	00003440
	D(J)=X(L)-T(J)	00003450
110	E(J)=B(J)+D(J)	00003460
	DO 115 J=1,M	00003470
	DO 115 K=1,J	00003480
	JK=JK+1	00003490
115	R(JK)=R(JK)+D(J)*D(K)	00003500
	GO TO 205	00003510
127	IF(N-M) 130, 130, 135	00003520
130	KK=N	00003530
	GO TO 137	00003540
135	KK=M	00003550
137	DO 140 I=1,KK	00003560
	CALL DATA(M,D,JJM)	00003570
	DC 140 J=1,M	00003580
	T(J)=T(J)+D(J)	00003590
	L=L+1	00003600
140	RX(L)=D(J)	00003610
	FKK=KK	00003620
	DO 150 J=1,M	00003630
	XBAR(J)=T(J)	00003640
150	T(J)=T(J)/FKK	00003650
	L=L	00003660
	DC 180 I=1,KK	00003670

JK=0	00003680
DC 170 J=1,M	00003690
L=L+1	00003700
170 D(J)=RX(L)-T(J)	00003710
DC 180 J=1,M	00003720
B(J)=B(J)+D(J)	00003730
DO 180 K=1,J	00003740
JK=JK+1	00003750
180 R(JK)=R(JK)+D(J)*D(K)	00003760
IF(N-KK) 205, 205, 1E5	00003770
1E5 KK=N-KK	00003780
DO 200 I=1, KK	00003790
JK=0	00003800
CALL DATA(M,D, JJM)	00003810
DO 190 J=1,M	00003820
XBAR(J)=XBAR(J)+E(J)	00003830
D(J)=D(J)-T(J)	00003840
190 B(J)=B(J)+D(J)	00003850
DC 200 J=1,M	00003860
DC 200 K=1,J	00003870
JK=JK+1	00003880
200 R(JK)=R(JK)+D(J)*D(K)	00003890
205 JK=C	00003900
DO 210 J=1,M	00003910
XBAR(J)=XBAR(J)/FN	00003920
DO 210 K=1,J	00003930
JK=JK+1	00003940
210 R(JK)=R(JK)-B(J)*B(K)/FN	00003950
JK=0	00003960
DO 220 J=1,M	00003970
JK=JK+J	00003980
220 STD(J)=SQRT(ABS(R(JK)))	00003990
DO 230 J=1,M	00004000
DC 230 K=J,M	00004010
JK=J+(K*K-K)/2	00004020
L=M*(J-1)+K	00004030
RX(L)=R(JK)	00004040
L=M*(K-1)+J	00004050
RX(L)=R(JK)	00004060
IF(STD(J)*STD(K)) 225, 222, 225	00004070
222 R(JK)=0.0	00004080
GC TC 230	00004090
225 R(JK)=R(JK)/(STD(J)*STD(K))	00004100
230 CONTINUE	00004110
FN=SQRT(FN-1.0)	00004120
DO 240 J=1,M	00004130
240 STD(J)=STD(J)/FN	00004140
L=-M	00004150
DC 250 I=1,M	00004160
L=L+M+1	00004170
250 E(I)=RX(L)	00004180
RETURN	00004190
END	00004200
SUBROUTINE EIGEN(A,R,N,MV)	00004210
DIMENSION A(1),R(1)	00004220
5 RANGE=1.0E-6	00004230
IF(MV-1) 10,25,10	00004240
10 IQ=-N	00004250
DO 20 J=1,N	00004260
IQ=IQ+N	00004270

00	20	I=1,N	00004280
		IJ=IC+I	00004290
		R(IJ)=0.0	00004300
		IF(I-J) 20,15,20	00004310
15		R(IJ)=1.0	00004320
20		CONTINUE	00004330
25		ANORM=0.0	00004340
		DC 35 I=1,N	00004350
		DO 35 J=I,N	00004360
		IF(I-J) 30,35,30	00004370
30		IA=I+(J*J-J)/2	00004380
		ANORM=ANORM+A(IA)*A(IA)	00004390
35		CONTINUE	00004400
		IF(ANORM) 165,165,40	00004410
40		ANORM=1.414*SQRT(ANORM)	00004420
		ANORMX=ANORM*RANGE/FLOAT(N)	00004430
		IND=0	00004440
		THR=ANORM	00004450
45		THR=THR/FLOAT(N)	00004460
50		L=1	00004470
55		M=L+1	00004480
60		MW=(M*M-M)/2	00004490
		LQ=(L*L-L)/2	00004500
		LM=L+MQ	00004510
62		IF(ABS(A(LM))-THR) 130,65,65	00004520
65		IND=1	00004530
		LL=L+LQ	00004540
		MM=M+MQ	00004550
		X=0.5*(A(LL)-A(MM))	00004560
68		Y=-A(LM)/SQRT(A(LM)*A(LM)+X*X)	00004570
		IF(X) 70,75,75	00004580
70		Y=-Y	00004590
75		SINX=Y/SQRT(2.0*(1.0+(SQRT(1.0-Y*Y))))	00004600
		SINX2=SINX*SINX	00004610
78		CCSX=SQRT(1.0-SINX2)	00004620
		COSX2=COSX*COSX	00004630
		SINCS=SINX*COSX	00004640
		ILQ=N*(L-1)	00004650
		IMQ=N*(M-1)	00004660
		DC 125 I=1,N	00004670
		IC=(I*I-1)/2	00004680
		IF(I-L) 80,115,80	00004690
80		IF(I-M) 85,115,90	00004700
85		IM=I+MQ	00004710
		GO TO 95	00004720
90		IM=M+IQ	00004730
95		IF(I-L) 100,105,105	00004740
100		IL=I+LQ	00004750
		GO TO 110	00004760
105		IL=L+IQ	00004770
110		X=A(IL)*COSX-A(IM)*SINX	00004780
		A(IM)=A(IL)*SINX+A(IM)*COSX	00004790
		A(IL)=X	00004800
115		IF(MV-1) 120,125,120	00004810
120		ILR=ILQ+I	00004820
		IMR=IMQ+I	00004830
		X=R(ILR)*COSX-R(IMR)*SINX	00004840
		R(IMR)=R(ILR)*SINX+R(IMR)*CCSX	00004850
		R(ILR)=X	00004860
125		CONTINUE	00004870

X=2.0*A(LM)*SINCS	00004880
Y=A(LL)*COSX2+A(MM)*SINX2-X	00004890
X=A(LL)*SINX2+A(MM)*COSX2+X	00004900
A(LM)=(A(LL)-A(MM))*SINCS+A(LM)*(COSX2-SINX2)	00004910
A(LL)=Y	00004920
A(MM)=X	00004930
130 IF(M-N) 135,140,135	00004940
135 M=M+1	00004950
GO TO 60	00004960
140 IF(L-(N-1)) 145,150,145	00004970
145 L=L+1	00004980
GO TO 55	00004990
150 IF(IND-1) 160,155,160	00005000
155 IND=0	00005010
GO TO 50	00005020
160 IF(THR-ANRMX) 165,165,45	00005030
165 IQ=-N	00005040
DO 185 I=1,N	00005050
IQ=IQ+N	00005060
LL=I+(I*I-I)/2	00005070
JQ=N*(I-2)	00005080
DO 185 J=I,N	00005090
JQ=JQ+N	00005100
MM=J+(J*J-J)/2	00005110
IF(A(LL)-A(MM)) 170,185,185	00005120
170 X=A(LL)	00005130
A(LL)=A(MM)	00005140
A(MM)=X	00005150
IF(MV-1) 175,185,175	00005160
175 DO 180 K=1,N	00005170
ILR=IQ+K	00005180
IMR=JQ+K	00005190
X=R(ILR)	00005200
R(ILR)=R(IMR)	00005210
180 R(IMR)=X	00005220
185 CCNTINUE	00005230
RETURN	00005240
END	00005250
SUBROUTINE TRACE (M,R,CON,K,D)	00005260
DIMENSION R(1),C(1)	00005270
FM=M	00005280
L=0	00005290
DO 100 I=1,M	00005300
L=L+I	00005310
100 D(I)=R(L)	00005320
K=0	00005330
DO 110 I=1,M	00005340
IF(D(I)-CON) 120,105,105	00005350
105 K=K+1	00005360
110 D(I)=D(I)/FM	00005370
120 DO 130 I=2,K	00005380
130 C(I)=D(I)+D(I-1)	00005390
RETURN	00005400
END	00005410
SUBROUTINE LOAD (M,K,R,V)	00005420
DIMENSION R(1),V(1)	00005430
L=0	00005440
JJ=0	00005450
DO 160 J=1,K	00005460
JJ=JJ+J	00005470

```
150 SQ= SQRT(R(JJ))
DO 160 I=1,M
L=L+1
```

```
00005480-
00005490
00005500
00005510
00005520
00005530
```

```
160 V(L)=SQ*V(L)
RETURN
END
```

```
SUBROUTINE B (N,NPUF,NPIR,NPUEV,NDIAG,NSQRE,NF,XNAME,CON,SJMF)
```

```
FACTCR ANALYSIS
```

```
PROGRAMMED ORIGINALLY IN MCGILL UNIVERSITY PSYCHOLOGY DEPARTMENT
```

```
AUTHOR CHUCK CRAWFORD FOR PH D THESIS ON IMAGE ANALYSIS (1966)
```

```
THIS PROGRAM CALCULATES IMAGE FACTORS, PRINCIPLE FACTORS OR
COMMUNALITY FACTORS. THE RESULTS ARE PRINTED AND CAN BE PUNCHED
BY EITHER ROWS OR COLUMNS ON FORMAT 8F9.4.
```

```
CONTROL CARD PUNCHED CN FORMAT6I2
```

```
N = NUMBER OF VARIABLES
```

```
NPUF = 00 FACTORS NOT PUNCHED
```

```
      = 01 FACTORS PUNCHED BY COLUMNS
```

```
      = -1 FACTORS PUNCHED BY ROWS
```

```
NPIR = 01 FOR PRINTING OF INTERMEDIATE RESULTS OF IMAGE ANALYSIS
```

```
NPUEV = 01 EIGENVECTORS PUNCHED BY COLUMNS
```

```
NDIAG = 00 FOR IMAGE ANALYSIS
```

```
      = 01 FOR PRINCIPLE FACTORS
```

```
      = 03 FOR COMMUNALITIES
```

```
NSQRE = 01 BOTH UPPER AND LOWER TRIANGLES PUNCHED
```

```
      = 00 LOWER TRIANGLE PUNCHED
```

```
SECOND CARD VARIABLE FORMAT
```

```
THIRD CARD TITLES, ETC.
```

```
DEFINITION OF ARGUMENTS
```

```
FR(JJ) = READ FORMAT
```

```
R(I,J) = CORRELATION MATRIX
```

```
S(I,J) = OUTPUT, EIGENVECTORS AND/OR FACTORS
```

```
RA = ACCURACY
```

```
KRLN = 1, FOR EIGENVECTORS
```

```
LIMENSION R(40,40)
```

```
DIMENSION S(40,40),XNAME(20),SSX(40) ,FR(20),S2(40)
```

```
DIMENSION G(40,40)
```

```
DIMENSION SJMF(40)
```

```
EQUIVALENCE(S,G)
```

```
10 FORMAT(6I2/20A4/20A4)
```

```
20 FCRMAT(1HK)
```

```
30 FORMAT(1H1)
```

```
40 FCRMAT(1HJ,20A4)
```

```
50 FCRMAT(1HJ, 7HINVERSE///)
```

```
60 FCRMAT(1HJ, 4HDET=E20.8, 4HEPS=E20.8)
```

```
70 FCRMAT(1HJ,23HIMAGE COVARIANCE MATRIX///)
```

```
80 FCRMAT(1HJ, 12HEIGENVECTORS///)
```

```
90 FORMAT(1HJ, 11HEIGENVALUES///)
```

```
100 FCRMAT(1HJ,13HIMAGE FACTORS///)
```

```
110 FORMAT(1HJ, 18HCORRELATION MATRIX///)
```

```
120 FORMAT(1HJ, 2HS2///)
```

```
130 FCRMAT(8F9.4)
```

```
140 FORMAT(1HJ,8F15.8)
```

```
150 FCRMAT(1HJ, 28HANTI IMAGE COVARIANCE MATRIX///)
```

```
160 FCRMAT(1HJ,46HRCCT MEAN SQUARE ANTI IMAGE COVARIANCE MATRIX=F20.9)
```

```
170 FORMAT(20A4)
```

```
180 FCRMAT(1HJ, 26HFACTORS, ONES IN DIAGONALS///)
```

```
190 FCRMAT(1HJ, 15HSUMS OF SQUARES///)
```

```

      IF(NDIAG.EQ.1) GO TO 914
      DC 21 I = 1, N
      DO 21 J = 1, N
      R(I,J) = 0.0
      S(I,J) = 0.0
      G(I,J) = 0.0
21  S2(I) = 0.0
      REWIND 4
      IF(NSQRE) 15,15,16
16  DC 18 J = 1, N
18  READ(4) (R(I,J),I = 1, N)
      IF(NSQRE.GT.0) GO TO 14
15  DC 17 J = 1, N
17  READ(4) (R(I,J),I = J,N)
      DO 19 I=1,N
      NI=I+1
      DO 19 J=NI,N
19  R(I,J) = R(J,I)
14  CONTINUE
      WRITE(6,40) (XNAME(JJ),JJ=1,20)
      WRITE (6,110)
      L = 1
      KRLN = 1
      RA = .0001
      IF(NDIAG.GE.3) GC TO 316
      DO 315 I = 1, N
      DC 315 J = 1, N
315 IF(I.EQ.J) R(I,J) = 1.0
      GO TO 318
316 DC 317 I=1,N
      DC 317 J=1,N
317 IF(I.EQ.J) R(I,J)=SUMF(I)
      CALL RPRINT(R,N,N,L)
      IF(NDIAG.GE.3) GC TO 45
318 DO 23 I = 1, N
      DC 23 J = 1, N
      23 G(I,J) = R(I,J)
      CALL RPRINT(R,N,N,L)
      CALL INVERT(R,N,N,DELT,EPS)
      IF(NPIR.GT.0) WRITE(6,50)
      IF(NPIR.GT.0) CALL RPRINT(R,N,N,L)
      DC 25 I=1,N
      DO 25 J=1,N
      25 IF(I.EQ.J) S2(I)=1.0/R(I,J)
320 WRITE(6,120)
      WRITE(6,140) (S2(I),I=1,N)
C  ANTI IMAGE COVARIANCE MATRIX
      DO 27 I=1,N
      DO 27 J=1,N
      XX=S2(I)*R(I,J)*S2(J)
      27 R(I,J)=XX
      IF(NPIR.GT.0) WRITE(6,150)
      IF(NPIR.GT.0) CALL RPRINT(R,N,N,L)
      SUMQ=0.0
      DC 35 I=1,N
      DC 35 J=1,N
      IF(I.EQ.J) GO TO 35
      SUMQ=SUMQ+R(I,J)**2
35  CONTINUE
      AN=N

```

```

DN=SQRT(SUMQ/(AN*(AN-1.0)))
WRITE(6,160) DN
C IMAGE COVARIANCE MATRIX
DC 37 I=1,N
DO 37 J=1,N
XX=G(I,J)+R(I,J)
IF(I.EQ.J) XX=XX-2.0*S2(I)
37 R(I,J)=XX
G(I,J)=0.0
IF(NPIR.GT.0) WRITE(6,70)
IF(NPIR.GT.0) CALL RPRINT(R,N,N,L)
45 CALL EVALUE(R,40,40,S,40,40,N,RA,KRUN)
WRITE(6,80)
CALL RPRINT(S,N,N,L)
DC 29 I = 1, N
DO 29 J = 1, N
29 IF(I.EQ.J) S2(I) = R(I,J)
DC 2 J = 1, N
NS = J + 1
IF(NS.GT.N) GO TO 2
DO 2 K = NS, N
IF(ABS(S2(J)).GT.ABS(S2(K)))GO TO 2
TEMP = S2(K)
S2(K) = S2(J)
S2(J) = TEMP
DC 3 I = 1, N
TEMP = S(I,K)
S(I,K) = S(I,J)
3 S(I,J) = TEMP
2 CONTINUE
NF=0
WRITE(6,300) (S2(I),I=1,N),CON
300 FURMAT(12F8.2,10X,F6.2)
DO 200 I=1,N
IF(ABS(S2(I)).LT.CON) GO TO 43
NF=NF+1
200 CONTINUE
43 IF(NPUEV.EQ.0) GO TO 44
WRITE(2,170) (XNAME(JJ),JJ=1,20)
DC 47 J = 1, NF
47 WRITE(2,130) (S(I,J),I = 1, N)
44 CONTINUE
WRITE(6,90)
WRITE(6,140) (S2(I),I = 1, NF)
DO 31 J = 1, NF
A = SQRT(ABS(S2(J)))
DO 31 I = 1, N
XX = S(I,J)*A
31 S(I,J) = XX
DO 32 I=1,N
WRITE(4) (S(I,J),J=1,NF)
32 CONTINUE
WRITE(6,30)
IF(NDIAG.GT.2) WRITE(6,180)
IF(NDIAG.EQ.2) WRITE(6,100)
CALL RPRINT (S,NF,N,L)
CALL GRAPH(NF,N,S)
DC 4840 J = 1, NF
4840 SSX(J) = 0.0
LC 5 J = 1, NF

```

```

DO 5 I = 1, N
5 SSX(J) = SSX(J) + S(I,J)**2
WRITE(6,190)
WRITE(6,140) (SSX(J), J = 1, NF)
WRITE(6,30)
IF(NPUF.NE.0) WRITE(2,170)(XNAME(JJ),JJ=1,20)
IF(NPUF) 900, 914, 901
901 DC 912 J = 1, NF
912 WRITE(2,130) (S(I,J), I = 1, N)
IF(NPUF.NE.(-1)) GO TC 914
900 DC 916 I = 1, N
916 WRITE(2,130) (S(I,J), J = 1, NF)
914 RETURN

```

END

SUBROUTINE INVERT(A,N,M,DELTA,EPS)

C THIS SUBROUTINE WILL INVERT OR SOLVE SYSTEMS OF SIMULTANEOUS
C EQUATIONS
C A IS THE MATRIX OF COEFFICIENTS AND MUST BE DIMENSIONED THE SAME
C IN BOTH THIS SUBROUTINE AND IN THE PROGRAM CALLING IT
C N = NUMBER OF ROWS
C M = NUMBER OF COLUMNS = N + NUMBER OF RT-HAND SIDES
C N AND M MUST BE LESS THAN OR EQUAL TO 100
C DELTA = DETERMINANT OF MATRIX FORMED BY COEFFS OF LEFT HAND SIDES
C DELTA SET TO 0. IF MATRIX SINGULAR OR EQUATIONS LIN DEPENDENT
C EPS = SMALLEST PIVOT USED IN THE ELIMINATION TECHNIQUE

DIMENSION A(40,40),C(40),D(40),IZ(40)

DELTA=1.

DC 2 J=1,N

2 IZ(J)=J

DC 778 I=1,N

K=I

Y=A(I,I)

IP=I+1

IF(I.GE.N)GOTC301

DO 30J=IP,N

W=A(I,J)

IF(ABS(W).LE.ABS(Y))GOTO30

K=J

Y=W

30 CONTINUE

301 DELTA=DELTA*Y

IF(I.EQ.1)EPS=ABS(Y)

EPS=AMIN1(EPS,ABS(Y))

IF(EPS.EQ.0.)RETRN

Y=1./Y

DC 6 J=1,N

C(J)=A(J,K)

A(J,K)=A(J,I)

6 A(J,I)=-C(J)*Y

DO 66 J=1,M

A(I,J)=A(I,J)*Y

66 D(J)=A(I,J)

A(I,I)=Y

J=IZ(I)

IZ(I)=IZ(K)

IZ(K)=J

DC 778 K=1,N

IF(K.EQ.I)GOTC778

7 DO 777 J=1,M

IF(J.NE.I)A(K,J)=A(K,J)-D(J)*C(K)

012

013

014

015

016

017

019

021

024

025

026

029

030

031

032

035

036

037

038

039

```

777 CONTINUE                                044
778 CCNTINUE                                046
    DO 121 I=1,N                            047
122 K=IZ(I)
    IF(K.EQ.1)GOTO121
    5 DO 10 J=1,M
      W=A(I,J)                                050
      A(I,J)=A(K,J)                          051
10  A(K,J)=W                                  052
      IP=IZ(I)                                053
      IZ(I)=IZ(K)                            054
      IZ(K)=IP                               055
12  DELTA=-DELTA                             056
    GOTO122
121 CONTINUE                                057
    RETURN
    END                                       069
    SUBROUTINE RPRINT(R,M,N,L)
    PRINTS OUT RECTANGULAR MATRICES
    ALL ROWS AND COLUMNS ARE APPROPRIATELY LABELLED
    R(I,J) = MATRIX TO BE PRINTED OUT
    M=WIDTH OF R. MAX. AS DIMENSIONED. NO. OF VARIABLES
    N=LENGTH OF R. MAX. AS DIMENSIONED. NO. OF PLACES
    L=1 FOR F10.4 OUTPUT
    L=0 FOR F10.2 OUTPUT
    L.GT.1 FOR F10.0 OUTPUT

    DIMENSION R(40,40)
    DIMENSION J(40)
    NPAGE = 0
    MI = 9
    M1 = 0
    M2 = 0
    JSEC = 0
    DO 8 I = 1,M
8    J(I) = I
    9 M1 = M2 + 1
      M2 = M1 + MI
      IF(M2.LE.M) GO TO 13
      M2 = M
13  JSEC = JSEC + 1
      NI = 49
      N1 = 0
      N2 = 0
    2 N1 = N2 + 1
      N2 = N1 + NI
      IF(N2.LE.N) GO TO 3
      N2 = N
    3 NPAGE = NPAGE + 1
      WRITE(6,17) JSEC,NPAGE
17  FORMAT(1H0 20X 9H SECTION      13, 7H PAGE      13/ )
      WRITE(6,22) (J(I), I = M1,M2)
22  FORMAT(' ',3HRCW,7X,11,9I13)
      IF(L.GT.1) GO TO 301
      IF(L.LT.1) GO TO 201
      DC 23 I = N1,N2
    23 WRITE(6,24) I, (R(I,K), K = M1,M2 )
24  FORMAT(1H ,I2,10F13.3)
      GO TO 31
201 DC 29 I = N1,N2

```

```

29 WRITE(6,30) I, (R(I,K), K = M1,M2 )
30 FORMAT(1H ,I2,10F13.3)
   GO TO 21
301 DC 302 I=N1,N2
302 WRITE(6,304) I,(R(I,K),K=M1,M2)
304 FORMAT(1HJ,I2,10F13.3)
31 WRITE(6,1)
   1 FORMAT(1H1)
   IF(N2.LT.N) GO TO 2
   IF(M2.LT.M) GO TO 9
   RETURN
   END

```

```

SUBROUTINE EVALUE(A,IL,IJ,S,IM,IN,N,RA,KRUN)

```

```

CALCULATES EIGEN VALUES AND EIGEN VECTORS
FOR A REAL SYMMETRIC MATRIX

```

```

SOURCE- 'MATHEMATICAL METHODS FOR DIGITAL COMPUTERS'
A.RALSTON + H.WILF, P.84

```

```

PROGRAMMED BY MRS. U.B.MANLEY (JUNE.1964)
MCGILL COMPUTING CENTER

```

```

DIMENSION A(1),S(1)

```

```

DO 5 I=1,N
DC 5 J=1,N
M=(J-1)*IM+I
5 S(M)=0.
DC 6 I=1,N
M=(I-1)*IM+I
6 S(M)=1.
IND=0
VI=0.
FN=N
DC 10 I=1,N
M=I-1
IF(M)10,90,11
11 DO 8 K=1,M
II=(K-1)*IL+I
8 VI=VI+A(II)**2
9C M=I+1
IF(M-N)4,4,10
4 DC 9 K=M,N
II=(K-1)*IL+I
9 VI=VI+A(II)**2
10 CONTINUE
VI=SQRT(VI)
IF(VI)1,145,1
1 VE=RA*VI/FN
106 VI=VI/FN
107 I=2
108 J=1
109 II=(I-1)*IL+J
IF(ABS(A(II))-VI)122,110,110
110 IND=1
II=(I-1)*IL+I
JJ=(J-1)*IL+J
SMY=(A(JJ)-A(II))/2.

```

```

CON11980
CON11990
CON12000
CON12010
CON12020
CON12030
CON12040
CON12050
CON12060
CON12070
CON12080
CON12090
CON12100
CON12110
CON12120
CON12130
CON12140
CON12150
CON12160
CON12170
CON12180
CON12190
CON12200
CON12210
CON12220
CON12230
CON12240
CON12250
CON12260
CON12270
CON12280
CON12290
CON12300
CON12310
CON12320
CON12330
CON12340
CON12350
CON12360
CON12370
CON12380
CON12390
CON12400
CON12410
CON12420
CON12430

```

JJ=(I-1)*IL+J	CON12440
SLAMB=-A(JJ)	CON12450
OMEGA=SLAMB/SQRT(SLAMB**2+SMY**2)	CON12460
IF(SMY)111,1111,1111	CON12470
111 OMEGA=-OMEGA	CON12480
1111 SI=OMEGA/SQRT(2.*(1.+SQRT(1.-OMEGA**2)))	CON12490
CC=SQRT(1.-SI**2)	CON12500
DO 114 L=1,N	CON12510
LL=L-I	CON12520
IF(LL)1113,1140,1113	CON12530
1113 LJ=L-J	CON12540
IF(LJ)1133,1140,1133	CON12550
1133 II=(J-1)*IL+L	CON12560
JJ=(I-1)*IL+L	CON12570
III=(L-1)*IL+I	CON12580
JJJ=(L-1)*IL+J	CON12590
B1=A(II)	CON12600
B2=A(JJ)	CON12610
A(II)=B1*CO-B2*SI	CON12620
A(JJ)=B1*SI+B2*CO	CON12630
A(III)=A(JJ)	CON12640
A(JJJ)=A(II)	CON12650
1140 IF(KRUN)2222,114,2222	CON12660
2222 II=(J-1)*IM+L	CON12670
JJ=(I-1)*IM+L	CON12680
S1=S(II)	CON12690
S2=S(JJ)	CON12700
S(II)=S1*CO-S2*SI	CON12710
S(JJ)=S1*SI+S2*CO	CON12720
114 CCNTINUE	CON12730
II=(I-1)*IL+I	CON12740
JJ=(J-1)*IL+J	CON12750
JI=(I-1)*IL+J	CON12760
IIJ=(J-1)*IL+I	CON12770
B1=A(JJ)	CON12780
B2=A(II)	CON12790
B3=A(JI)	CON12800
A(JJ)=B1*CO**2+B2*SI**2-2.*B3*SI*CO	CON12810
A(II)=B1*SI**2+B2*CO**2+2.*SI*CO*B3	CON12820
A(JI)=(B1-B2)*SI*CO+B3*(CO*CO-SI*SI)	CON12830
A(IIJ)=A(JI)	CON12840
122 J2=J-(I-1)	CON12850
IF(J2)1124,124,1124	CON12860
1124 J=J+1	CON12870
GO TO 109	CON12880
124 N1=N-I	CON12890
IF(N1)125,126,125	CON12900
125 I=I+1	CON12910
GO TO 108	CON12920
126 IF(IND)127,128,127	CON12930
127 IND=0	CON12940
GO TO 107	CON12950
128 IF(VI-VE)130,130,106	CON12960
130 RETURN	CON12970
145 WRITE(6,146)	CON12980
146 FORMAT(1H0,21HINPLT MATRIX DIOGNAL)	CON12990
RETURN	CON13000
END	CON13010
SUBROUTINE C (M,N,MINF,MAXF,NP,NORM,NVQ,NPUF,TITLE)	
CRTHGNAL ROTATION	

```

C PROGRAMMED ORIGINALLY IN MCGILL UNIVERSITY PSYCHOLOGY DEPARTMENT
C AUTHOR CHUCK CRAWFORD FOR PH D THESIS ON IMAGE ANALYSIS (1966)
C
C DEFINITION DES PARAMETRES
C M= NOMBRE DE FACTEURS
C MINF NOMBRE MINIMUM DE FACTEURS PREVUS
C MAXF NOMBRE MAXIMUM DE FACTEURS PREVUS
C NP= 0 PCUR IMPRIMER LES FACTEURS NORMALISES
C NCRM=0 PCUR NORMALISATION DES FACTEURS
C NVQ=-1 PCUR ROTATION QARTIMAX
C NVQ=00 PCUR RLATION EQUIMAX
C NVQ=01 PCUR ROTATION VARIMAX
C DEFINITION OF ARGUMENTS
C
C FIRST CARD, FCFMAT(8I2)
C
C M = NUMBER OF FACTORS
C N = NUMBER OF TESTS
C MINF = MINIMUM NUMBER OF FACTORS EXPECTED
C MAXF = MAXIMUM NUMBER OF FACTORS EXPECTED
C NP = 0 FOR PRINTING NORMALIZED FACTORS
C NCRM = 0 FOR NORMALIZATION
C NVQ = -1 FOR QUARTIMAX
C NVQ = 00 FOREQUIMAX
C NVQ = 01 FOR VARIMAX
C NPUF = 01 FOR PUNCH BY ROWS OF VARIMAX ROTATION
C
C SECCNL CARD, TITLES, ETC. FORMAT(16A5)
C
C THIRD CARD -- VARIABLE FORMAT FOR DATA READ-IN (16A5)
C
C XF(I,J) = FACTOR MATRIX PUNCHED BY ROWS
C H2(I) = COMMUNALITIES
C T = NUMERATOR OF VARIMAX
C B = DENCMINATOR OF VARIMAX
C
C INTEGER BLANC, ETOIL, BAK, TIR, PLUS
C INTEGER XP(2)
C DATA BLANC, ETOIL, BAR, TIR, PLUS/ ' ', '* ', 'I ', '- ', '+ ' /
C DIMENSION MPL(100,100)
C DIMENSION X(40,40), XF(40,40), VAR(40), SUMS(40), TITLE(80), K(40), H2(7
C *5), TV(100), SX(40), SSX(40), SUMF(40), SUM4X(40)
C DIMENSION PERCNT(20)
10 FORMAT(8I2/16A4/16A4)
20 FCFMAT(8F9.4)
30 FCFMAT(1HK, 10FVARIMAX ROTATION///)
40 FCFMAT(1HK, 21FTRANSFORMATION MATRIX///)
50 FCFMAT(1H1)
60 FCFMAT(1HK, 17FURNROTATED FACTORS///)
70 FCFMAT(1HK, 28FRGTATED DENORMALIZED FACTORS///)
80 FCFMAT(1HK, 26FRCTATED NCRMALIZED FACTORS///)
90 FCFMAT(1HK, 15FSUMS OF SQUARES///)
100 FCFMAT(1HK, 17HVARIMAX CRITERION///)
110 FCFMAT(1HJ, 10F12.4)
120 FCFMAT(1HJ, 8HSTOP ONE///)
130 FCFMAT(1HJ, 8FSTOP TWO///)
140 FCFMAT(1HJ, 10FSTOP THREE///)
150 FCFMAT(1HJ, 28FVARIANCE OF SQUARED LOADINGS///)
1600 FCFMAT(1FJ, 7HCYCLE =14, 4X, 4HTV =F12.7, 4X, 3HP =F12.7, 4X, 7HINTER =F1
12.7, 4X, 7HINTRA =F12.7)

```

```

170 FORMAT(1HJ, 2I2, 2F15.10)
180 FORMAT(1HJ, 14)
200 FORMAT(1HJ, 13HFCLRTH POWERS//)
201 FORMAT(1HJ, 18FQUARTIMAX ROTATION)
202 FORMAT(1HJ, 16HECLIMAX ROTATION)
203 FORMAT(1HJ, 16FVARIMAX ROTATION)
210 FORMAT(1HJ,20A4)
220 FORMAT(1HJ, 21HINTRA FACTOR PRODUCTS)
    IF(INVQ.NE.2) GO TO 12
    WRITE(6,11)
    11 FORMAT(1X,' UN SEUL FACTEUR: PAS DE ROTATION')
    GO TO 1010
    12 CONTINUE
    REWIND 4
    DO 1 I=1,N
    DC 1 J = 1, M
    1 XF(I,J) = 0.0
    DC 526 I=1,N
    READ(4) (X(I,J),J=1,N)
    WRITE(6,110) (X(I,J),J=1,N)
526 CCNTINUE
    DC 527 I=1,N
    READ(4) (XF(I,J),J=1,M)
    WRITE(6,110) (XF(I,J),J=1,M)
527 CCNTINUE
    DO 400 I=1,100
    DC 400 J=1,100
    MPL(I,J)=BLANC
400 CONTINUE
    DC 401 I=1,100
    MPL(I,50)=ETOIL
401 CONTINUE
    DC 402 J=1,100
    MPL(50,J)=ETOIL
402 CONTINUE
    DC 411 I=1,N
    DO 410 J=1,2
    XP(J)=50+(XF(I,J)*100)/2
    WRITE(6,409) XP(J)
409 FORMAT(5X,I10)
410 CCNTINUE
    MPL(XP(1),XP(2))=PLUS
411 CONTINUE
    DC 420 I=1,100
    WRITE(6,419) (MPL(I,J),J=1,100)
419 FORMAT(100A1)
420 CCNTINUE
C   ORDER BY SUMS CF SQUARES
    DO 5 J = 1, M
    5 SSX(J) = 0.0
    DC 6 J = 1, M
    DO 6 I = 1, N
    6 SSX(J) = SSX(J) + XF(I,J)**2
    DO 7 J = 1, M
    NS = J + 1
    IF(NS.GT.N)GO TO 7
    DO 777 KK = NS, N
    IF(SSX(J).GE.SSX(KK)) GO TO 7
    TEMP = SSX(KK)
    SSX(KK) = SSX(J)

```

```

      SSX(J) = TEMP
      DO 8 I = 1, N
      TEMP = XF(I, KK)
      XF(I, KK) = XF(I, J)
8     XF(I, J) = TEMP
777  CLCONTINUE
7     CONTINUE
      WRITE(6, 50)
      WRITE(6, 60)
      L = 1
      WRITE(6, 210) (TITLE(JJ), JJ = 1, 20)
      CALL PRINT(XF, M, N, L)
      WRITE(6, 90)
      WRITE(6, 110) (SSX(J), J = 1, M)
      DC 99 JNCTF = MINF, MAXF
      XJN = JNCTF
      IF(NVQ) 301, 302, 303
301  VEQ = 0.0
      GO TO 304
302  VEQ = XJN/2.0
      GO TO 304
303  VEQ = 1.0
304  CONTINUE
975  DC 528 I = 1, N
      DO 528 J = 1, JNCTF
528  X(I, J) = 0.0
      DC 529 I = 1, N
      DO 529 J = 1, JNCTF
529  X(I, J) = XF(I, J)
      DC 308 NV = 1, 100
308  TV(NV) = 0.0
      DC 3 I=1, N
3     H2(I) = 0.0
C     NORMALIZE
      IF(NORM.GT.0) GO TO 33
      DO 4 I = 1, N
      DO 4 J = 1, JNCTF
4     H2(I) = H2(I) + X(I, J)**2
      DO 37 I = 1, N
      DO 37 J = 1, JNCTF
      XX = X(I, J)/SQRT(H2(I))
37    X(I, J) = XX
33    CONTINUE
      CCNS = SQRT(2.0)/2.0
      MM = JNCTF*(JNCTF - 1)/2
      E = .00116
      AN = N
      NV=0
22    CONTINUE
      NCC = 0
      IF(NOC.GE.MM) WRITE(6, 140)
      IF(NOC.EQ.MM) GO TO 13
      NV = NV + 1
      DO 2 J = 1, JNCTF
      SUMF(J) = 0.0
      SX(J) = 0.0
      SSX(J) = 0.0
      SUM4X(J) = 0.0
      VAR(J) = 0.0
2     SUMS(J) = 0.0

```

```

IF(NV.GT.50) WRITE(6,120)
IF(NV.GT.50) GC TC 13
S4X = 0.0
SIJ = 0.0
SV = 0.0
DC 47 J = 1, JNOTF
DC 123 I = 1, N
IXY = I + 1
IF(IXY.GT.N)GO TO 124
DC 123 II = IXY,N
123 SUM4X(J) = SUM4X(J) + 2.*((X(I,J)**2)*(X(II,J)**2))
124 DC 46 I = 1, N
X2 = X(I,J)**2
SX(J) = SX(J) + X2
46 SSX(J) = SSX(J) + X2**2
47 VAR(J) = (AN*SSX(J) - VEQ*SX(J)**2)/AN**2
DC 48 J = 1, JNOTF
S4X = S4X + SSX(J)
SIJ = SIJ + SUM4X(J)
48 SV = SV + VAR(J)
TV(NV) = SV
GO TO 703
974 WRITE(6,100)
WRITE(6,110) (VAR(J),J = 1, JNOTF)
WRITE(6,200)
WRITE(6,110) (SSX(J),J = 1, JNOTF)
WRITE(6,220)
WRITE(6,110) (SUM4X(J),J = 1, JNOTF)
703 CCNTINUE
XINTER = (AN - VEQ)*S4X/AN**2
XINTRA = VEQ*SIJ/AN**2
AM = JNOTF
P = 0.0
IF(NVC.EQ.(-1)) P=(AM*S4X-AN)/(AN*AM-AN)
IF(NVC.EQ.00) F= (2.0*AM*TV(NV)-2.0+AM)/(2.0*AM-2.0)
IF(NVQ.EQ.01) P=(AM*TV(NV))/(AM-1.0)
WRITE(6,160) NV,TV(NV),P,XINTER,XINTRA
MI = JNOTF - 1
DO 500 IP = 1, MI
MJ = IP + 1
IF(MJ.GT.JNOTF)GC TO 22
DO 500 JP = MJ, JNOTF
T = 0.0
B = 0.0
TAN4T = 0.0
SIN4T = 0.0
COS4T = 0.0
CCT4T = 0.0
CALL VARMAX(X,IP,JP,N,T,B,VEQ)
IF((T - B).GT.(0.0)) GO TO 1042
IF((T - B).LT.(0.0)) GO TO 1041
IF((T - B).EQ.(0.0))GO TC 1433
1433 IF((T + B - E).LT.(0.0)) NOC = NOC + 1
IF((T + B - E).LT.(0.0)) GO TO 500
C SECTION 4D3
1043 COS4T = CCNS
SIN4T = CCNS
GO TO 500
C SECTION 4D1
1041 TAN4T = ABS(T)/ABS(B)

```

```

      IF(TAN4T.LT.E) GO TO 8000
      CCS4T = 1.0/SQRT(1.0 + TAN4T**2)
      SIN4T = TAN4T*CCS4T
      GO TO 5000
C     SECTION 8
8000 IF(B.GE.0.0) NOC = NOC + 1
      IF(B.GE.0.0) GO TO 500
      CCSP = CONS
      SINP = CONS
      GO TO 1000
C     SECTION 402
1042 COT4T = ABS(T)/ABS(B)
      IF(COT4T.LT.E) GO TO 9000
      SIN4T = 1.0/SQRT(1.0 + COT4T**2)
      CCS4T = COT4T*SIN4T
      GO TO 5000
C     SECTION 9
9000 SIN4T = 1.0
      CCS4T = 0.0
      GO TO 5000
C     SECTION 5
5000 CCS2T = SQRT((1.0 + CCS4T)/2.0)
      SIN2T = SIN4T/(2.0*CCS2T)
      CCST = SQRT((1.0 + CCS2T)/2.0)
      SINT = SIN2T/(2.0*CCST)
      GO TO 1300
C     SECTION 6
1300 IF(B.LE.0.0) GO TO 1250
      COSP = COST
      SINP = SINT
      GO TO 7000
1250 COSP = CONS*COST + CONS*SINT
      SINP = ABS(CONS*CCST - CONS*SINT)
      GO TO 7000
C     SECTION 7
7000 IF(T.GT.0.0) GO TO 1000
      SINP = -SINP
      GO TO 1000
C     SECTION 10
1000 DO 21 I = 1, N
      XX = X(I,IP)*COSP + X(I,JP)*SINP
      YY = -X(I,IP)*SINP + X(I,JP)*COSP
      X(I,IP) = XX
      21 X(I,JP) = YY
      500 CONTINUE
      GO TO 22
      13 CONTINUE
      WRITE(6,50)
      IF(NP.EQ.0) WRITE(6,80)
      IF(NP.EQ.0) CALL PRINT(X,JNOTF,N,L)
C     DENORMALIZE
      IF(NORM.GT.0) GO TO 34
      DO 38 I = 1, N
      DO 38 J=1,JNOTF
      XX = X(I,J)*SQRT(H2(I))
      38 X(I,J) = XX
      34 WRITE(6,70)
C     ORDER FACTORS BY SUMS OF SQUARES
      DO 9 J = 1, JNOTF
      PERCNT(J)=0.0

```

```

9  SUMS(J) = 0.0
   DO 49 J = 1, JNCTF
   DC 49 I = 1, N
49  SUMS(J) = SUMS(J) + X(I,J)**2
   REALJZ=N
   DC 533 J=1,JNOTF
   PERCNT(J)=SUMS(J)/REALJZ
533  CONTINUE
   WRITE (6,430)
430  FORMAT(//'  PERCENT EXPLANATION FOR EACH ROTATED FACTOR'/)
   DO 440 J=1,JNOTF
   WRITE (6,204) J,PERCNT(J)
204  FORMAT(I5,F20.7)
440  CONTINUE
   DC 530 J = 1, JNCTF
   NS = J + 1
   IF(NS.GT.JNOTF)GO TO 532
   DC 530 KK = NS, JNOTF
   IF(SUMS(J) .GE. SUMS(KK)) GO TO 530
   TEMP = SUMS(KK)
   SUMS(KK) = SUMS(J)
   SUMS(J) = TEMP
   DC 531 I = 1, N
   TEMP = X(I, KK)
   X(I, KK) = X(I, J)
531  X(I, J) = TEMP
530  CONTINUE
532  IF(NVQ) 2000,2001,2002
2000  WRITE(6,201)
      GO TO 2003
2001  WRITE(6,202)
      GO TO 2003
2002  WRITE(6,203)
2003  WRITE(6,210) (TITLE(JJ),JJ = 1, 20)
      DC 51 I=1,N
      WRITE(4) (X(I,J),J=1,JNOTF)
51  CONTINUE
   CALL PRINT(X,JNCTF,N,L)
   DO 600 I=1,N
   DC 600 J=1,JNOTF
   XF(I,J) = X(I,J)
600  CONTINUE
   CALL GRAPH(JNOTF,N,X)
   DEL=0.0
   EPS=.00001
   MAXI=50
   IF=1
   KU=1
   DINC=0.0
   DEND=0.0
   NOSTOP=0
   CALL CBLI (N,JNCTF,DEL,EPS,MAXI,IF,KU,DINC,DEND,NOSTOP,XF)
   IF(NPLF.EQ.0) GO TO 57
   DC 55 I=1,N
55  PUNCH 56,(X(I,J),J=1,JNOTF)
56  FORMAT(10F8.4)
57  WRITE(6,90)
   WRITE(6,110) (SUMS(J), J = 1, JNOTF)
59  CONTINUE
   WRITE(6,50)

```

00002280

00002290

00002310

00002320

00002330

```

1010 CALL EXIT
      END
      SLBRoutine PRINT(R,M,N,L)
      DIMENSION J(40),R(40,40)
C     PRINT OUT RECTANGULAR MATRICES
C
C     R I,J=MATRIX TO BE PRINTED CUT
C     M = WIDTH OF R. MAX. AS DIMENSIONED. NO. OF VARIABLES
C     N = LENGTH OF R. MAX. AS DIMENSIONED. NO. OF PLACES
C     L = 1 FOR F10.4 OUTPUT
C     L = 0 FOR F10.2 OUTPUT
C     L = 2 FOR F10.0 OUTPUT
C
      NPAGE=0
      MI=9
      M1=0
      M2=0
      JSEC=0
      DO 40 I=1,M
40    J(I)=I
50    M1=M2+1
      M2=M1+MI
      IF(M2-M)70,70,60
60    M2=M
70    JSEC=JSEC+1
      NI=49
      N1=0
      N2=0
80    N1=N2+1
      N2=N1+NI
      IF(N2-N)100,100,90
90    N2=N
100   NPAGE=NPAGE+1
      WRITE(6,110) JSEC,NPAGE
110   FCRMAT(1H0,25X,'SECTION',15,'PAGE',15/)
      WRITE(6,120) (J(I),I=M1,M2)
120   FCRMAT(6H0 ROW 3X 10I 10)
      IF(L-1)160,130,190
130   DO 140 I=N1,N2
140   WRITE(6,150)I,(R(I,K),K=M1,M2)
150   FCRMAT(1H 15, 5X 10F10.4)
      GO TO 220
160   DO 170 I=N1,N2
170   WRITE(6,180)I,(R(I,K),K=M1,M2)
180   FCRMAT(1H 15, 4X, 10F10.2)
      GO TO 220
190   DO 200 I=N1,N2
200   WRITE(6,210)I,(R(I,K),K=M1,M2)
210   FCRMAT(1HJ,15,5X,10F10.0)
220   WRITE(6,230)
230   FCRMAT(1H1)
      IF(N2-N)80,240,240
240   IF(M2-M)50,250,250
250   RETURN
      END
      SUBROUTINE VARMAX(X,IP,JP,N,T,B,VEQ)
      DIMENSION X(40,40)
      AA = 0.0
      BB = 0.0
      CC = 0.0

```

```

CC = 0.0
DC 21 I = 1, N
XX = X(I,IP)
YY = X(I,JP)
U = (XX - YY)*(XX + YY)
V = 2.0*XX*YY
CC = CC + (U + V)*(U - V)
DD = DD + 2.0*U*V
AA = AA + U
21 BB = BB + V
FN = N
T = DD - VEQ*(2.0*AA*BB/FN)
E = CC - VEQ*(AA**2 - BB**2)/FN
RETURN
END
SUBROUTINE GRAPH(K,M,F)
INTEGER BLANC,ETOIL,BAR,TIR
DATA BLANC,ETOIL,BAR,TIR /' ','*','I','-' /
DIMENSION MAT(40,100)
DIMENSION F(40,40),IFACT(40)
DO 650 KK=1,K
DC 700 II=1,M
700 IFACT(II)=II
DC640 I=1,M
J=I+1
IF (J.GT.M) GO TO 640
DC 601 L=J,M
IF(F(I,KK).GE.F(L,KK)) GO TO 601
TEMP=F(I,KK)
F(I,KK)=F(L,KK)
F(L,KK)=TEMP
ITEMP=IFACT(I)
IFACT(I)=IFACT(L)
IFACT(L)=ITEMP
601 CONTINUE
640 CONTINUE
WRITE(6,611)
611 FORMAT('0',130(' - '))
WRITE(6,612)
612 FORMAT(22X,'-1.0 -0.9 -0.8 -0.7 -0.6 -0.5 -0.4 -0.3 -0.2 -0.1
+ +.1 +.2 +.3 +.4 +.5 +.6 +.7 +.8 +.9 +1.0')
WRITE(6,611)
DO 502 I=1,M
DC 502 J=1,100
502 MAT(I,J)= BLANC
DO 503 I=1,M
503 MAT(I,50)=BAR
DO 506 I=1,M
AF=(100*F(I,KK))/2
IX=AF
C BF(M,KK)=IX(M,KK)
C IF((AF(M,KK)-BF(M,KK)).GE.0.5) IX(M,KK)=IX(M,KK)+1
504 IX=50+IX
MM=IX
NN=MM-1
IF(NN.GE.50) GO TO 510
NN=MM+1
DC 509 J=NN,49
509 MAT(I,J)=TIR
MAT(I,MM)=ETOIL

```

```

00000160
00000170
00000880
00000890
00000900
00000910
00000920
00000930
00000940
00000950
00000960
00000970
00000980
00000990
00001000
00001010
00001020
00001050
00001060
00001070
00001080
00001090
00001100
00001110
00001120
00001130
00001140
00001150
00001160
00001190
00001200
00001230
00001240
00001250
00001260
00001270
00001280

```

```

GC TG 506
510 DO 508 J=51,NN
508 MAT(I,J) =TIR
505 MAT(I,MM)=ETOIL
506 WRITE(6,507)IFACT(I),F(I,KK),(MAT(I,J),J=1,100)
507 FCRMAT(15,5X,F12.5,3X,10A1)
WRITE(6,611)
WRITE(6,612)
WRITE(6,611)
650 CONTINUE
RETURN
END
SUBROUTINE OBLI (N,M,DEL,EPS,MAXI,IF,K,DINC,DEND,NOSTOP,A)
DIMENSION A(40,40),T(40,40),F(20),S(40,40),B(40,40)
3 FCRMAT(2X,13,15,F5.1,F10.6,3I5,2F5.1,15)

C
C N=NOMBKE DE VARIABLES
C M=NOMBRE DE FACTEURS
C DEL=DELTA
C EPS= HABITUELLEMENT .00001
C MAXI=NOMBRE MAXIMUM D'ITERATIONS
C IF=PATTERN INITIAL DES FACTEURS
C =1 ORTHOGONAL;CAS HABITUEL
C =0 CBLIQUE
C =0 OBLIQUE
C K==NORMALISATION DE KAISER
C =1 EMPLOYEE
C =0 NON EMPLOYEE
C DINC=DELTA INCREMENT, DOIT ETRE POSITIF
C DEND= LA PLUS GRANDE VALEUR DE DELTA
C NOSTOP=1 S'IL YA DES DENNES ADDITIONNEE
C NCSTOP=1 S'IL YA DES DONNES ADDITIONNELLES
C =0 SI NON
C LES TROIS DERNIER CHAMPS DOIVENT DEUMEURER BLANC POUR UN SIMPL
C PROBLEME AVEC UN SEUL ENSEMBLE DE DONNEES
C
C
C
C 501 READ(5,3) N,M,DEL,EPS,MAXI,IF,K,DINC,DEND,NOSTOP
C READ(5,4) F
4 FORMAT(20A4)
C DC 5 J=1,N
C 5 READ(5,F) (A(J,I),I=1,M)
C WRITE(6,4) F
501 WRITE(6,25) N,M,MAXI,EPS
25 FORMAT(1X,' NOMBRE DE VARIABLES',I6/1X,' NOMBRE DE FACTEURS',I8/
+1X,' LA LIMITE DES ITERATIUNS',I10/1X,'EPSILON',8X,F10.6//)
WRITE(6,1)
1 FORMAT(1X,'PATTERN ORIGINAL DES FACTEURS'//)
CC 7J=1,N
7 WRITE(6,8) J,(A(J,I),I=1,M)
C MISE EN MEMOIRE DE LA STRUCTURE FACTORIELLE POUR EMPLOIE SUBSEQUENT
CC 50 I=1,N
DO 50 J=1,M
50 B(I,J)=A(I,J)
502 WRITE(6,510) DEL
510 FORMAT(1X,'DELTA',F8.2)
CALL ROTAT3(A,T,N,M,DEL,EPS,MAXI,IF,K,S)
IF(K.EQ.0) WRITE(6,6)
6 FCRMAT(1X,' RCTATION SANS LA NCRMALISATION D''APRES KAISER'//)
IF(K.EQ.1) WRITE(6,9)

```

```

00001290
00001300
00001310
00001320
00001330
00001340
00001350
00001360
00001370
00001380

```

```

9 FORMAT(1X,'APRES LA ROTATION AVEC LA NORMALISATION DE KAISER'//)
WRITE(6,19)
19 FORMAT(1X,'STRUCTURE FACTORIELLE'//)
DO 10 J=1,N
10 WRITE(6,8) J,(A(J,I),I=1,M)
8 FORMAT(1X,I4,5F12.5)
WRITE(6,11)
11 FORMAT(1X,'CORRELATIONS FACTORIELLES'//)
DO 12 I=1,M
12 WRITE(6,8) I,(T(I,J),J=1,M)
WRITE(6,15)
15 FORMAT(1X,'FACTOR STRUCTURE'//)
DO 20 I=1,N
20 WRITE(6,8) I,(S(I,J),J=1,M)

C
C BOUCLE POUR LES DIFFERENTES VALEURS DE DELTA
C
IF(DINC) 40,40,35
35 DEL=DEL+DINC

C
C REMEMORISER LA STRUCTURE FACTORIELLE ORIGINALE POUR UTILISATION
C AVEC LE PROCHAIN DELTA
DO 60 I=1,N
DO 60 J=1,M
60 A(I,J)=B(I,J)
IF(DEL-DEND)502,502,40

C
C BOUCLE POUR DES DONNEES ADDITIONNELLES
C
40 IF(NCSTOP) 45,45,501
45 RETURN
END
SUBROUTINE ROTAT3 (A,TT,LV,LR,GG,ACC,MR,INI,NDR,SS)
DIMENSION A(40,40),TT(40,40),C(40),S(40),V(40),TL(40),FL(40)
+,XL(40),YL(40),SS(40,40)
WRITE(6,882)
882 FORMAT(1X,'ROTATION FOR DIRECT OBLIMIN LOADINGS'
+/1X,' ITERATION CIROLB'/13X,' CRITERION ')
NV=LV
NR=LR
GA=GG/FLOAT(NV)
DO 30 I=1,NV
S(I)=0.0
DO 4 J=1,NR
4 S(I)=S(I)+A(I,J)*A(I,J)
IF(NDR) 30,30,32
32 FL(I)=SQRT(S(I))
S(I)=1.0
DO 2 J=1,NR
2 A(I,J)=A(I,J)/FL(I)
30 CONTINUE
DO 33 I=1,NR
IF(INI) 31,31,3
3 DO 6 J=1,NR
6 TT(I,J)=0.0
TT(I,I)=1.0
31 C(I)=0.0
V(I)=0.0
DO 33 J=1,NV
AA =A(J,I)*A(J,I)

```

```

      C(I)=C(I)+AA
33  V(I)=V(I)+AA*AA
      G=0.0
      D=0.0
      DC 34 J=1,NR
      D=D+C(J)
      V(J)=V(J)-GA*C(J)*C(J)
34  G=G+V(J)
      H=0.0
      DC 5 I=1,NV
      H=H+S(I)*S(I)
      H=H-GA*D*D
      FC =F-G
      FCC=FC*ACU
      L=0
      WRITE(6,52) L,FC
52  FCRMAT(I6,F16.6)
70  DD 80 L=1,MR
      DC 81 IP=1,NR
      DC 81 IQ=1,NR
      IF(IP-IQ)82,81,82
82  D=D-C(IP)-C(IQ)
      G=G-V(IP)-V(IQ)
      P=0.0
      R=0.0
      T=0.0
      U=0.0
      Y=0.0
      Z=0.0
      DD 83 I=1,NV
      A1=A(I,IP)
      A2=A(I,IQ)
      AA=A1*A1
      BB=A2*A2
      AB=A1*A2
      S(I)=S(I)-AA-BB
      Z=Z+AB
      R=R+AA*BB
      P=P+AB*AA
      T=T+AA*S(I)
E3  O=O+AB*S(I)
      X=TT(IP,IQ)
      GAC=GA*C(IP)
      R=R-GAC*C(IQ)
      P=P-GAC*Z
      C=C-GA*Z*D
      U=U-GA*C(IQ)*C
      T=T-GAC*D
100 P1=1.5*(X-P/V(IP))
      Q1=.5*(V(IP)-4.0*X*P+R+2.0*T)/V(IP)
      R1=.5*(X*(T+R)-P-O)/V(IP)
      CALL ROCT(P1,Q1,R1,A2)
      A22=A2*A2
      P=1.0+2.0*X*A2+A22
      IF(P)89,90,90
E5  WRITE(6,91) P
91  FCRMAT(2X,F20.7)
      P=-P
90  A1=SQRT(P)
      A3=A2/A1

```

```

A11=P*P
C(IP)=P*C(IP)
V(IP)=A11*V(IP)
Z=0.0
Y=L.0
DC 84 I=1,NV
A(I,IQ)=A(I,IQ)-A2*A(I,IP)
A(I,IP)=A1*A(I,IP)
BB=A(I,IQ)*A(I,IQ)
Z=Z+BB*BB
Y=Y+BB
84 S(I)=S(I)+A(I,IP)*A(I,IP)+BB
V(IQ)=Z-GA*Y*Y
C(IQ)=Y
D=D+Y+C(IP)
G=G+V(IP)+V(IQ)
A1=1.0/A1
DC 85 I=1,NR
TT(I,IP)=A1*TT(I,IP)+A3*TT(I,IQ)
85 TT(IP,I)=TT(I,IP)
TT(IP,IP)=1.0
81 CONTINUE
F=0.0
DC 86 I=1,NV
86 H=H+S(I)*S(I)
F=F-GA*D*D-G
WRITE(6,52) L,F
157 IF(ABS(F-F0)-FCC)38,38,80
80 FC = F
38 CONTINUE
IF(NUK) 88,88,71
71 DC 36 I=1,NR
DO 36 J=1,NV
36 A(J,I)=A(J,I)*FL(J)
88 DC 190 I=1,NV
DO 190 J=1,NR
190 SS(I,J)=0.0
DC 95 K=1,NV
DO 95 I=1,NR
DO 95 J=1,NR
95 SS(K,I)=SS(K,I)+A(K,J)*TT(J,I)
99 RETURN
END
SUBROUTINE RCOI(P,Q,R,X)
H=(ABS(P)+ABS(Q)+ABS(R))*1.E-5
F2=2.0*P
A=P*P-3.0*Q
IF(A) 1,1,2
1 X=0.0
GO TO 5
2 A=SQRT(A)
X=(A-F)/3.0
X1=-(P+A)/3.0
IF(X*(Q+X*(P+X))+X1*(Q+X1*(P+X1))+R+R)3,4,4
3 X=X+1.0
GO TO 5
4 X=X1-1.0
5 DC 7 I=1,50
F=R+X*(Q+X*(P+X))
FP=Q+X*(P2+3.0*X)

```

```

DX=F/FP
X=X-CX
IF (ABS(F)-H)6,6,7
7 CONTINUE
WRITE(6,8) P,Q,R,X
8 FCRMAT(2H *4E2C.8)
6 RETURN
END

```

```

//LKED.SYSLMUD DD DSN=+MICHE(MAIN),DISP=(NEW,PASS),UNIT=SYSDA,
//          SPACE=(1600,(160,50,1))

```

```

//LKED.SYSIN DD *
INSERT MAIN,GRAPH
CVERLAY FIRST
INSERT A
OVERLAY 1
INSERT INNGRM,CORRE,DATA
CVERLAY 1
INSERT EIGEN,TRACE,LOAD
CVERLAY FIRST
INSERT B
OVERLAY SECOND
INSERT RPRINT,INVERT
CVERLAY SECOND
INSERT EVALUE
CVERLAY FIRST
INSERT C
OVERLAY 3
INSERT PRINT
OVERLAY 3
INSERT VARMAX
CVERLAY 3
INSERT CBLI,ROTAT3,RECT

```

```

//GC.FT06FU01 DD SYSOUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=1330)          00006730
//GC.FT07FU01 DD UNIT=SYSDA,DSN=+TEMPO,DISP=(NEW,PASS),                  00006740
//          SPACE=(TRK,(5,1)),DCB=(RECFM=FB,LRECL=120,BLKSIZE=3600)      00006750
//GC.FT04FU01 DD UNIT=SYSDA,DSN=+TEMP6,DISP=(NEW,PASS),
//          SPACE=(TRK,(30,1)),DCB=(RECFM=VBS,LRECL=524,BLKSIZE=3148)
//GC.FT12FU01 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT12FU01,DISP=(SHR,PASS)
//GC.FT13FU01 DD UNIT=SYSDA,DSN=*.STEP3.GO.FT13FU01,DISP=(SHR,PASS)
//GC.FT08FU01 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT08FU01,DISP=(SHR,PASS)
//GC.FT02FU01 DD DUMMY,DCB=(BLKSIZE=80,BUFNO=1)
//GC.FT05FU01 DD DDNAME=INPUT
//GC.INPUT DD *

```

```

//STEP6 EXEC FORTGCLG,PARM.FORT='BCD',PARM.LKED='OVLY,LIST,XREF',
//          REGICN.GC=220K,TIME.GO=3

```

```

//FCRT.SYSIN DD *
C FACTOR SCORES PROGRAM          00006820
C                                00006830
C                                00006840
C PROGRAMMED ORIGINALLY IN MCGILL UNIVERSITY PSYCHOLOGY DEPARTMENT      00006850
C AUTHCK CHUCK CRAWFORD FOR PH D THESIS ON IMAGE ANALYSIS (1966)        00006860
C ADAPTATION BY GUY LEMAY UNIVERSITY OF OTTAWA
C THIS IS NOT A FINAL VERSION
C LV=THE NUMBER OF VARIABLES          00006870
C NV=THE NUMBER OF SIGNIFICANT FACTORS 00006880
C NS=THE NUMBER OF PLACES FOR WHICH FACTOR SCORES TO BE OBTAINED        00006890
C NPUF=1 IF FACTOR SCORES TO BE PUNCHED OUT 00006900
C NSQ=.GT. 0 IF ONLY THE LOWER HALF OF THE CORRELATION MATRIX READ      00006910

```

C		00006920
C	PROGRAM SETUP	00006930
C	PROGRAM	00006940
C	HEADER CARD	00006950
C	FORMAT FOR CORRELATION MATRIX	00006960
C	FORMAT FOR FACTOR LOADING MATRIX	00006970
C	FORMAT FOR STANDARD SCORES	00006980
C	CORRELATION MATRIX (LV X LV)	00006990
C	FACTOR LOADING MATRIX (LV X NV)	00007000
C	STANDARD SCORES OF RAW DATA	00007010
C		00007020
C	DO NOT FORGET TO DIMENSION THE TWO SUBROUTINES	00007030
C		00007040
C		00007050
C		00007060
C		00007070
C		00007080
	COMMON R(131,40),S(131,40),S2(40),XNAME(80),KKK(40),C(40),D(40),IZ	00007090
	X(40),SSX(40),FR(20)	00007100
	COMMON/LE/BANK(40,40)	
	DIMENSION VEE(131),FMT(20),FFT(20),IHEAD(20)	00007110
	DIMENSION ZSCOR(20)	
	DIMENSION HFSCSL(150,20)	
	DIMENSION B(150,20),IPLAC(150)	
	DIMENSION RKEEP(131,40)	00007140
	DOUBLE PRECISION SCORE(131)	00007150
	READ(12) (XNAME(JJ),JJ=1,20),IFACTO,IROT,CON,NPUF,NPIR,NPUEV,	
	+NSQRE,NP,NORM,NSQ,IBUG,ICOMP,NS,LV,NV	
	WRITE(6,998) (XNAME(JJ),JJ=1,20),IFACTO,IROT,CON,NPUF,NPIR,NPUEV,	
	+NSQRE,NP,NORM,NSQ,IBUG,ICOMP,NS,LV,NV	
558	FORMAT(1X,20A4,2I2,F6.1,14I2)	
	IBUG=1	
	ICOMP=1	
	IF(ICOMP.EQ.0) GO TO 999	
839	FORMAT(20A4)	00007170
	WRITE(6,838) (IHEAD(JJ),JJ=1,20)	00007180
838	FORMAT(1H1,25X,20A4)	00007190
1	FORMAT(5I5,50X,15)	00007210
	READ(1,7) (FMT(JJ),JJ=1,20)	00007220
	READ(1,7) (FFT(JJ),JJ=1,20)	00007230
	READ(1,7) (FR(JJ),JJ=1,20)	00007240
7	FORMAT(20A4)	00007250
	IF(NSQ.GT.0) GO TO 405	00007260
	DC 2 I=1,LV	00007270
2	READ(4) (R(I,J),J=1,LV)	
	GO TO 406	00007290
405	DC 407 J=1,LV	00007300
407	READ(4) (R(I,J),J=1,LV)	
	DC 408 I=1,LV	00007320
	NI=I+1	00007330
	DO 408 J=NI,LV	00007340
408	R(I,J)=R(J,I)	00007350
	DC 409 I=1,LV	00007360
409	R(I,I)=1.0	00007370
	IF(IBUG.LE.0) GO TO 329	00007380
406	WRITE(6,305)	00007390
305	FORMAT(1HK,35HCORRELATION MATRIX	00007400
	L=1	00007410
	CALL RPRINT(R,LV,LV,L)	00007420
	DO 400 I=1,LV	00007430

	DC 400 J=1, LV	00007440
400	RKEEP(I, J)=R(I, J)	00007450
	DC 3001 KFOIS=1, 2	00007460
325	DO 4 I=1, LV	00007470
	READ(4) (S(I, J), J=1, NV)	
4	WRITE(6, FFT)(S(I, J), J=1, NV)	00007490
	IF(IBM.G.LE.0) GC TO 353	00007500
	WRITE(6, 307)	00007510
307	FORMAT(1HK, 35HFACTOR LOADING MATRIX)	00007520
	CALL RPRINT(S, NV, LV, L)	00007530
	DC 630 I=1, NS	
	DC 630 K=1, NV	
	HFSCSU(I, K)=0.0	
630	CONTINUE	
	DC 632 I=1, NS	
	READ(7, FR) (ZSCOR(J), J=1, LV)	
	WRITE(6, FR) (ZSCOR(J), J=1, LV)	
	DC 632 K=1, NV	
	DO 632 J=1, LV	
	HFSCSU(I, K)=HFSCSU(I, K)+ZSCOR(J)*S(J, K)	
632	CONTINUE	
	REWIND 7	
	IF(KFCIS-2) 634, 633, 634	
633	WRITE(6, 635)	
635	FORMAT(1X, 'FACTEURS SCORES----METHODE CLASSIQUE ROTATION')	
	GC TO 637	
634	WRITE(6, 636)	
636	FORMAT(1X, 'FACTEURS SCORES---METHODE CLASSIQUE NON ROTATIONES')	
637	DC 638 I=1, NS	
	WRITE(6, 639) I, (HFSCSU(I, K), K=1, NV)	
639	FORMAT(1X, I3, 13F8.3)	
638	CONTINUE	
	CALL ORS(NV, NS, HFSCSU)	
353	CALL INVERT(R, LV, LV, DELTA, EPS)	00007540
	IF(IBM.G.LE.0) GC TO 354	00007550
	WRITE(6, 306)	00007560
306	FORMAT(1HK, 35HINVERSE CORRELATION MATRIX)	00007570
	CALL RPRINT(R, LV, LV, L)	00007580
354	DO 6 I=1, LV	00007590
6	S2(I)=1.0/R(I, I)	00007600
	REWIND 9	00007610
	DC 101 I=1, LV	00007620
	DC 101 J=1, LV	00007630
	XX= S2(I)*R(I, J)	00007640
101	R(I, J) = XX	00007650
	DC 102 I=1, LV	00007660
	DO 102 J=1, LV	00007670
	IF (I.EQ.J) R(I, J) = 1.-R(I, J)	00007680
102	R(I, J) = 0.0 - R(I, J)	00007690
C	HAVE COMPUTED W AT THIS STAGE	00007700
	DO 103 I=1, NV	00007710
	DO 103 J=1, LV	00007720
	XX=0.	00007730
	DC 104 II=1, LV	00007740
	XX=XX+S(II, I)*R(II, J)	00007750
104	CONTINUE	00007760
103	WRITE(9) XX	00007770
C	HAVE COMPUTED A*W AT THIS STAGE	00007780
	REWIND 9	00007790
	DC 105 I=1, NV	00007800

DO 105 J=1,NV	00007810
R(I,J) =0.	00007820
DO 105 II=1,LV	00007830
105 R(I,J) = R(I,J) + S(II,I) * S(II,J)	00007840
IF(IBUG.LE.0) GO TO 534	00007850
WRITE(6,309)	00007860
309 FORMAT(1HK,35F)A MATRIX BEFORE INVERSION	00007870
CALL RPRINT(R,NV,NV,L)	00007880
C HAVE COMPUTED A*A AT THIS STAGE	00007890
534 CALL INVERT (R,NV,NV,DELTA,EPS)	00007900
IF(IBUG.LE.0) GO TO 531	00007910
WRITE(6,308)	00007920
308 FORMAT(1HK,35F)INVERTED A*A MATRIX	00007930
CALL RPRINT(R,NV,NV,L)	00007940
C HAVE COMPUTED A*A TO THE -1 AT THIS STAGE	00007950
531 DO 110 I=1,NV	00007960
DO 110 J=1,LV	00007970
110 READ(9) S(I,J)	00007980
REWIND 9	00007990
DO 968 J=1,NV	00008000
DC 968 I=1,NV	00008010
968 WRITE(9) R(I,J)	00008020
REWIND 9	00008030
DC 945 I=1,LV	00008040
DC 945 J=1,LV	00008050
945 R(I,J)=0.	00008060
DC 106 J=1,NV	00008070
DC 106 I=1,NV	00008080
READ(9) XX	00008090
DC 107 II=1,LV	00008100
107 R(J,II)=R(J,II)+XX*S(I,II)	00008110
106 CONTINUE	00008120
C HAVE COMPUTED A*A-1A*W AT THIS STAGE	00008130
C COMPLETE FACTOR SCORES	00008140
WRITE(6,287) (IHEAD(JJ),JJ=1,20)	00008150
287 FORMAT(1H1,10X,18H)FACTOR SCORES FOR ,20A4)	00008160
REWIND 12	
DC 202 K=1,NS	00008170
READ(7,FR) (VEE(I),I=1,LV)	00008180
DO 201 I =1,NV	00008190
SCORE(I) =0.	00008200
B(K,I)=0	00008210
DC 201 J=1,LV	00008220
SCORE(I) = SCORE (I) + R(I,J) * VEE(J)	00008230
201 B(K,I)=SCORE(I)	00008240
NK1=1	00008250
NK2=10	00008260
JJ=1	00008270
205 IF(NPUF.NE. 1) GO TO 207	00008280
IF(NK2.GT.NV) NK2=NV	00008290
C GC 800 I=NK1,NK2	00008300
C WRITE(2,204) SCORE(I),K,JJ	00008310
WRITE(2,204) (SCORE(I),I=NK1,NK2),K,JJ	00008320
C 800 JJ=JJ+1	00008330
204 FORMAT(7F8.3,18X,2I3)	00008340
207 CONTINUE	00008350
IF(NK2.EQ.NV) GO TO 202	00008360
NK1 = NK1+9	00008370
NK2 = NK2 + 9	00008380
JJ = JJ + 1	00008390

IF(NK2.GT.NV) NK2 = NV	00008400
GL TC 205	00008410
202 WRITE(6,203) K,(SCORE(I),I=1,NV)	00008420
REWIND 7	00008430
203 FLKMAT(1H0,I5,2X,(10F12.3/(7X,10F12.3)))	00008440
4000 WRITE(6,2000)	00008450
2000 FCRMAT('1',T45,'INTERPRETATION DES FACTEURS SCORES')	00008460
WRITE(6,3010) ((E(K,I),I=1,NV),K=1,NS)	00008470
3010 FORMAT(10(1X,F10.5))	00008480
CALL CRS(NV,NS,B)	
DO 500 I=1,LV	00008690
DU 500 J=1,LV	00008700
500 R(I,J)=RKEEP(I,J)	00008710
3001 CONTINUE	00008720
999 RETURN	00008740
END	00008750
SUBROUTINE ORS(NV,NS,B)	
DIMENSION B(150,20),IPLAC(150)	
WRITE(6,1) NV,NS	
1 FORMAT(2I4)	
DO 2 II=1,NS	
WRITE(6,3) (B(II,KK),KK=1,NV)	
3 FORMAT(1X,13F8.4)	
2 CONTINUE	
DO 3000 KK=1,NV	00008490
DO 3500 II=1,NS	00008500
3500 IPLAC(II)=II	00008510
DO 3600 K=1,NS	00008520
I=K+1	00008530
IF(I.GT.NS) GO TO 3600	00008540
DO 3700 L=I,NS	00008550
IF(B(K,KK).GE.B(L,KK)) GO TO 3700	00008560
TEMP=B(K,KK)	00008570
B(K,KK)=B(L,KK)	00008580
B(L,KK)=TEMP	00008590
ITEMP=IPLAC(K)	00008600
IPLAC(K)=IPLAC(L)	00008610
IPLAC(L)=ITEMP	00008620
3700 CONTINUE	00008630
3600 WRITE(6,3601) IPLAC(K),B(K,KK)	00008640
3601 FCRMAT(15,10X,F12.3)	00008650
WRITE(6,3602)	00008660
3602 FCRMAT('0',110(' - '))	00008670
3000 CONTINUE	00008680
RETURN	
END	
SUBROUTINE INVERT(A,N,M,DELTA,EPS)	00008760
C THIS SUBROUTINE WILL INVERT OR SOLVE SYSTEMS OF SIMULTANEOUS	00008770
C EQUATIONS	00008780
C A IS THE MATRIX OF COEFFICIENTS AND MUST BE DIMENSIONED THE SAME	00008790
C IN BOTH THIS SUBROUTINE AND IN THE PROGRAM CALLING IT	00008800
C N = NUMBER OF ROWS	00008810
C M = NUMBER OF COLUMNS = N + NUMBER OF RT-HAND SIDES	00008820
C M AND N MUST BE LESS THAN OR EQUAL TO 90	00008830
C DELTA = DETERMINANT OF MATRIX FORMED BY COEFFS OF LEFT HAND SIDES	00008840
C DELTA SET TO 0. IF MATRIX SINGULAR OR EQUATIONS LIN DEPENDENT	00008850
C EPS = SMALLEST IVOT USED IN THE ELIMINATION TECHNIQUE	00008860
C COMMON R(131,40),S(131,40),S2(40),XNAME(80),KKK(40),C(40),D(40),I	00008870
C X(40),SSX(40),FF(20)	00008880
C DIMENSION A(131,40),G(131,40)	00008890

EQUIVALENCE (S(1,1),C(1,1))	00008900
DELTA=1.	00008910
DO 2 J=1,N	00008920
2 IZ(J)=J	00008930
DO 778 I=1,N	00008940
K=I	00008950
Y=A(I,I)	00008960
IP=I+1	00008970
IF(I.GE.N)GOTC301	00008980
DO 30J=IP,N	00008990
W=A(I,J)	00009000
IF(ABS(W).LE.AES(Y))GOTO30	00009010
K=J	00009020
Y=W	00009030
30 CCNTINUE	00009040
301 DELTA=DELTA*Y	00009050
IF(I.EQ.1)EPS=ABS(Y)	00009060
EPS=AMIN1(EPS,AES(Y))	00009070
IF(EPS.EQ.0.)RETURN	00009080
Y=1./Y	00009090
DO 6 J=1,N	00009100
C(J)=A(J,K)	00009110
A(J,K)=A(J,I)	00009120
6 A(J,I)=-C(J)*Y	00009130
DO 66 J=1,M	00009140
A(I,J)=A(I,J)*Y	00009150
66 D(J)=A(I,J)	00009160
A(I,I)=Y	00009170
J=IZ(I)	00009180
IZ(I)=IZ(K)	00009190
IZ(K)=J	00009200
DO 778 K=1,N	00009210
IF(K.EQ.1)GOTC778	00009220
7 DO 777 J=1,M	00009230
IF(J.NE.1)A(K,J)=A(K,J)-D(J)*C(K)	00009240
777 CCNTINUE	00009250
778 CONTINUE	00009260
DO 121 I=1,N	00009270
122 K=IZ(I)	00009280
IF(K.EQ.1)GOTO121	00009290
9 DO 10 J=1,M	00009300
W=A(I,J)	00009310
A(I,J)=A(K,J)	00009320
10 A(K,J)=W	00009330
IP=IZ(I)	00009340
IZ(I)=IZ(K)	00009350
IZ(K)=IP	00009360
12 DELTA=-DELTA	00009370
GOTO122	00009380
121 CCNTINUE	00009390
RETURN	00009400
END	00009410
SUBROUTINE RPRINT(R,M,N,L)	00009420
C PRINTS CUT RECTANGULAR MATRICES	00009430
C ALL ROWS AND COLUMNS ARE APPROPRIATELY LABELLED	00009440
C R(I,J) = MATRIX TO BE PRINTED OUT	00009450
C M=WIDTH OF R. MAX. AS DIMENSIONED. NO. OF VARIABLES	00009460
C N=LENGTH OF R. MAX. AS DIMENSIONED. NO. OF PLACES	00009470
C L=1 FOR F10.4 OUTPUT	00009480
C L=0 FOR F10.2 CLTPUT	00009490

```

C      L.GT.1 FOR F10.0 OUTPUT
C
      DIMENSION J(40),R(131,40)
      NPAGE = 0
      M1 = 9
      M1 = 0
      M2 = 0
      JSEC = 0
      DO 8 I = 1,M
8      J(I) = 1
9      M1 = M2 + 1
      M2 = M1 + M1
      IF(M2.LE.M) GO TO 13
      M2 = M
13     JSEC = JSEC + 1
      N1 = 49
      N1 = 0
      N2 = 0
2      N1 = N2 + 1
      N2 = N1 + N1
      IF(N2.LE.N) GO TO 3
      N2 = N
3      NPAGE = NPAGE + 1
      WRITE(6,17) JSEC,NPAGE
17     FCRMAT(1H,20X,9H SECTION      13, 7H PAGE      13/ )
      WRITE(6,22) (J(I), I = M1,M2)
22     FCRMAT(' ',3HROW,7X,I1,9I13)
      IF(L.GT.1) GO TO 301
      IF(L.LT.1) GO TO 201
      DC 23 I = N1,N2
23     WRITE(6,24) I, (R(I,K), K = M1,M2 )
24     FCRMAT(1H ,I2,10F13.3)
      GC TO 31
201    DC 29 I = N1,N2
29     WRITE(6,30) I, (R(I,K), K = M1,M2 )
C30    FCRMAT(1H ,I2,10F13.3)
30     FCRMAT(1H ,I2,8E14.7)
      GC TO 31
301    DC 302 I=N1,N2
302    WRITE(6,304) I,(R(I,K),K=M1,M2)
C304   FCRMAT(1HJ,I2,10F13.3)
304    FCRMAT(1HJ,I2,8E14.7)
31     WRITE(6,1)
      1 FCRMAT(1H1)
      IF(N2.LT.N) GC TO 2
      IF(M2.LT.M) GO TO 9
      RETURN
      END

```

```

00009500
00009510
00009520
00009530
00009540
00009550
00009560
00009570
00009580
00009590
00009600
00009610
00009620
00009630
00009640
00009650
00009660
00009670
00009680
00009690
00009700
00009710
00009720
00009730
00009740
00009750
00009760
00009770
00009780
00009790
00009800
00009810
00009820
00009830
00009840
00009850
00009860
00009870
00009880
00009890
00009900
00009910
00009920
00009930
00009940
00009950
00009960
00009970

```

```

//LKED.SYSLMOD DD DSNAME=+GUY02(MAIN),DISP=(NEW,PASS),UNIT=SYSDA,
//          SPACE=(1600,(160,50,1))
//LKED.SYSIN DD *
  INSERT MAIN,KPRINT
  OVERLAY 1
  INSERT INVERT
  CVERLAY 1
  INSERT ORS

```

```

//GC.FT06F001 DD SYSCUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=1330) 00010010
//GO.FT07F001 DD UNIT=SYSDA,DSN=*.STEP5.GO.FT07F001,DISP=(OLD,DELETE) 0001003

```

```

//GC.FT04FU01 DD UNIT=SYSDA,DSN=*.STEP5.GO.FT04FU01,DISP=(OLD,DELETE)      0001002
//GC.FT12FU01 DD UNIT=SYSDA,DSN=*.STEP1.GO.FT12FU01,DISP=(SHR,PASS)
//GC.FT09FU01 DD UNIT=SYSDA,SPACE=(TRK,(8,1)),DISP=(NEW,DELETE),          00010040
//          DSN=+TEMP7,DCB=(RECFM=VBS,LRECL=8,BLKSIZE=84)
//GC.FT02FU01 DD DUMMY,DCB=(BLKSIZE=80,BUFNO=1)
//GC.SYSIN DD *                      DATA NEXT 00010070
(10F12.5)                             00010100
(10F12.5)                             00010110
(12F10.1)                             00010120

```

```

//          ,,,,1100),*GLY LEMAY          ',MSGLEVEL=(2,0),
//A EXEC ASMFCOMP
//ASM.SYSIN DD *
OR      CSECT
        ENTRY AND
        L      15,0(1)
        L      0,0(15)
        L      1,4(1)
        C      0,0(1)
        BR     14
AND     L      15,0(1)
        L      0,0(15)
        L      1,4(1)
        N      0,0(1)
        BR     14
        END

```

```

//B EXEC FCRTCOMP,PARM=FCRT='BCD,LOAD,NODECK,LINECNT=45'
//FCRT.SYSIN DD *                      SOURCE DECK NEXT
        WRITE (6,3222)
3222  FORMAT (///'U',T41,'THE UNIVERSITY OF IOWA'/T41,
* 'DEPARTMENT OF GEOGRAPHY'/T45,'PROGRAM LIBRARY'/T41,
* 'MODIFIED FROM A PROGRAM'/T40,'WRITTEN BY PETER M. NEELY')

```

```

C
C      ***** CONGROUP *****
C      ***** SET FOR MAXIMUM OF 250 OBSERVATIONS *****
C
        DIMENSION A(45000),NA(300),FMT(180),TITLE(8),IFMT(180)
        DATA PROB,FINS/'PROB','FINS'/
        INTEGER X,Y,AND,CR
        COMMON A, NA, FMT, DUM(140)
        NSYSIN = 5
        NT = 6
        NPAGE = 1
        NSYZ = 45000
100   READ(5,1001,END=147)WORD,NTIN,MVAR,INCO,NFMT,KODE,NCLUS,NOOB,TITLE
        IF(WORD.EQ.PROB)GO TO 19
        IF(WORD.EQ.FINS)GO TO 13
12   CALL ERROR('DSTNCE',12,NT,WORD)
13   WRITE(6,2005)WORD
        CALL EXIT
19   IF(NTIN)1,1,2
1   NTIN = NSYSIN
        GO TO 4
2   IF(NVAR*NOOB)9,9,3
3   IF(((NOOB+10)*NCOB)/2-NSYZ)4,4,9
9   CALL ERROR('DSTNCE',9,NT,'N*P')
91  CALL ERROR('DSTNCE',91,NT,'INCO')
4   X = 16**6-1+7*16**6
        Y = 16**6-1+3*16**6

```

```

Y = CR(TITLE(1),Y)
NNFMT = NFMT * 20
IF(X.EQ.Y)GO TC 6
6 WRITE(6,2007)TITLE,NPAGE
NPAGE = NPAGE + 1
READ(5,2000) (IFMT(I),I=1,NNFMT)
WRITE(NT,2008)(IFMT(I),I=1,NNFMT)
I1=NSYZ-NOOB * MVAR + 1 - MVAR
I2 = I1 + MVAR-1
DO 10 I = 1,NCCB
I1 = I1 + MVAR
I2 = I2 + MVAR
IF(INCO - 2)8,7,91
7 READ(NTIN)(A(J),J = I1,I2)
GC TC 10
8 READ(NTIN,IFMT) (A(J),J=I1,I2)
WRITE(6,IFMT) (A(J),J=I1,I2)
10 CONTINUE
9999 WRITE(NT,2001)MVAR,NCCB
K=0
I1=NSYZ-NOOB * MVAR + 1
DC 24 I = 1,NCCB
J1=I1
DC 22 J=I,NOOB
K=K+1
A(K)= J.
I2=I1
J2=J1
DC 20 L = 1,MVAR
A(K) = A(K) + (A(I2) - A(J2))**2
I2=I1 + 1
20 J2=J1+1
22 J1=J1+MVAR
24 I1=I1+MVAR
IF(NCLUS)29,29,28
28 CALL CNTGTY(A,NCCB,NSYZ,NSYSIN)
29 IF(NCLUS - 2)30,32,60
30 NCLUS = 1
32 DC 35 I=1,NOOB
LL=NCOB + 1 - I
35 NA(LL) = I
NN=((NOOB - 1) / 16 ) * 16 + 1
CC 48 I = 1,NN,16
LL = MINO(I+15,NCCB)
MM=NOOB - I + 1
CC 46 M = 1,MM
IF(MOD(M-1,50))39,38,39
38 WRITE(6,2009)TITLE,NPAGE
NPAGE = NPAGE + 1
WRITE(NT,2002)(NA(K),K=I,LL)
WRITE(NT,2003)
39 LL = MINO(LL,NCCB - M + 1)
L = NA(I) + ((2*NOOB-M) * (M-1))/2
CC 44 K = I,LL
FMT(K) = A(L)
44 L = L - 1
46 WRITE(NT,2004)M,(FMT(K),K = I,LL)
48 CONTINUE
60 CALL GRPING(NCCB,MVAR,KODE,NCLUS,NT,NPAGE,TITLE)
GC TC 100

```

```

147 CALL EXIT
C
1001 FORMAT(A4,I2,9X,I3,2I1,5X,I1,12X,I1,4X,I5,8A4)
2000 FORMAT('1')
2001 FORMAT('J NUMBER OF VARIABLES = ',I9,70X,'SUBROUTINE DSTNCE ',/,
*'C NUMBER OF OBSERVATIONS = ',I6,70X,'DATED 1971'///)
2002 FORMAT('NOBS NO. ',I6I7)
2003 FORMAT('1')
2004 FORMAT(1X,I4,3X,I6F7.2)
2005 FORMAT(1HC/1H0,20X,A4)
2006 FORMAT(20A4)
2007 FORMAT('1',8A4,T56,' START OF PROBLEM - GROUPING ',T120,' PAGE',I3)
2008 FORMAT('JFORMAT FOR INPUT DATA IS-',( ' ',T33,20A4))
2009 FORMAT('1',8A4,T56,'DISTANCE SQUARED MATRIX',T120,' PAGE ',I3)
END
SUBROUTINE GRPING(NOOB,MVAR,KODE,NCLUS,NT,NPAGE,TITLE)
INTEGER XI LHF,XIRHF,AND,OR,BLANK,DATA
DATA BLAN2,JX,JY,JZ / ' ','+', 'I', '-' /
DIMENSION A(45000),NA(45000),JTEM(300),NATA(300),DATA(300),
*MINX(300),SUMX(300),TGI2(300),SXSQ(300),ITEM(300),TITLE(8)
COMMON A,NATA,MINX,SUMX,TGI2,SXSQ,ITEM
COMMON AVOG,BCT,JJ,K,NZ,NOIN,VMAX,B1,DEFR,J,LL,N4,NDUT,VMIN,B2,I2,
IKEY,L,N5,N,XX,BLANK,II,KK,MM,NCYC,RMAX,YY,BNCT,I,KOUT,M,NN,TRACE
EQUIVALENCE (A,NA),(NATA,DATA)
C
BLANK=BLAN2
NCYC=0
KCLNT = NOOB * MVAR
AVOG = 6.623 * 10. ** 23
VMAX = 0.
TRACE = 0.
DO 2 I = 1,NOOB
L = (I-1) * NOOB + I - ((I-1)*I)/2
TRACE = TRACE + A(L)
DO 13 I = 1,NOCB
DATA(I) = I+1
MINX(I) = 0
ITEM(I) = 0
13 SUMX(I) = 1.
NATA(NOOB) = J
C
NM = NOCB
IF(KODE)16,16,14
14 IF(KODE-6)17,17,15
15 CALL ERKOR('GRUP',15,NT,'KODE')
16 KODE=1
17 WRITE(NT,2000)NOOB,KODE
WRITE(NT,2016)TITLE,NPAGE
NPAGE = NPAGE + 1
WRITE(NT,2006)
20 I = 1
VMAX=AVOG
VMIN=AVOG
GO TO (24,22,22,24,24,22),KODE
22 VMAX=-VMAX
VMIN=-VMIN
C
J = I
24 IF(NATA(J))34,31,34
30
C

```

SET UP TO KICK OUT

BEGIN SEARCH

KICK OUT AS SET UP

```

31 IF(NATA(I))33,50,33
33 I=NATA(I)
   GU TO 24
34 J=NATA(J)
   C
   C M IS THE INDEX TO FIND C(I,J)
   C L IS THE INDEX TO FIND C(I,I)
   C K IS THE INDEX TO FIND C(J,J)
   M=(I-1)*NCOB+J-((I-1)*I)/2
   L=(I-1)*NCOB+I-((I-1)*I)/2
   K=(J-1)*NCOB+J-((J-1)*J)/2
   IF(A(M))30,35,35
35 BOT=A(M)-A(L)-A(K)
   DEFR=SUMX(I)+SUMX(J)-1.
   BNOT=(A(M)*(SUMX(I)+SUMX(J))-A(L)*SUMX(I)-A(K)*SUMX(J))/((SUMX(I)
1+SUMX(I))*SUMX(J))
   C
   C CODE 1 (WARD) GIVEN DISTANCE SQUARES AS INPUT - COMPUTE SUMS OF
   C SQUARES AT EACH STEP (FOR CURRENT GROUPS) - AS A LINEAR FUNCTION
   C OF SQUARED DISTANCES THEN MINIMIZE INCREMENT WITH EACH ADDITION
   C CODE 2 DOES CONVERSE FOR PROXIMITY
   C CODE 3 GROUPS ON MAXIMUM PROXIMITY - UPDATING ON BASIS OF AVERAGES
   C CODE 4 GROUPS ON MINIMUM DISTANCE - UPDATING ON BASIS OF AVERAGES
   C CODE 5 GROUPS ON MINIMUM MEAN SQUARE FOR THAT GROUP BEING FORMED
   C CODE 6 GROUPS ON MAXIMUM ATTRACTION (DENSITY = M), A = M*M / D**2
   C ATTRACTION IS UPDATED AS MAXIMUM TO ANY MEMBER IN GROUP
   C
   GC TO (40,39,41,42,36,37),KODE
36 BOT = A(M) / DEFR
   GU TO 40
37 IF(A(M))372,371,372
371 VMIN = 0.
   II=I
   JJ=J
   GO TO 50
372 IF(A(M)-VMIN)30,30,38
38 VMIN=A(M)
   GO TO 44
39 IF(BCT - VMIN)30,30,43
40 IF(BCT - VMIN)43,30,30
41 IF(VMAX - BNOT)43,30,30
42 IF(VMAX - BNOT)30,30,43
43 VMAX = BNOT
   VMIN = BOT
44 II = I
   JJ = J
   GU TO 30
   C
   C NATA(I) CONTAINS INDEX OF NEXT ITEM TO BE TESTED - HENCE
   C ENABLING ONE TO SKIP ITEMS WHICH ARE ALREADY GROUPED
   C
50 NCYC = NCYC + 1
   I=II
51 IF(NATA(I) - JJ)52,53,52
52 I = NATA(I)
   GO TO 51
53 NATA(I) = NATA(JJ)
   I = II
   NOIN = I
   C THE LEFT HALF OF MINX(I) CONTAINS INDEX OF NEXT MEMBER IN GROUP
60 I2 = XILHF(MINX(I))

```

```

        IF(I2)61,62,61
61      I = I2
        NCIN = NOIN + 1
        GO TO 60
62      CALL SLH(JJ,MINX(I))
C       THE RIGHT HALF OF MINX(J) GETS CYCLE NUMBER WHEN GROUPED
        CALL SRH(NCYC,MINX(I))
        I = JJ
        NDUT = 1
65      I2 = XLHF(MINX(I))
        IF(I2)66,70,66
66      I=I2
        NCUT = NDUT + 1
        GO TO 65
70      TRACE = TRACE + VMIN
        TGI2(NCYC) = TRACE
        IF(KODE - 3)76,75,75
75      TGI2(NCYC) = VMIN
C76     CONTINUE
C       THE LEFT HALF OF MINX(I) = 0 FOR LAST MEMBER IN GROUP
C       HENCE FIND ZERO IN LEFT HALF OF WORD, THEN UPDATE
76      WRITE(NT,2001)NCYC,TRACE,VMIN,II,JJ,NOIN,NDUT,VMAX
        I = 1
        M = (II - 1) * NOOB + JJ - ((II - 1) * II) / 2
        L = (II - 1) * NCCB + II - ((II - 1) * II) / 2
        K = (JJ - 1) * NOOB + JJ - ((JJ - 1) * JJ) / 2
        XX = A(M)
        B1 = A(M) * (SUMX(II) + SUMX(JJ))
        B2 = A(L) * SUMX(II) + A(K) * SUMX(JJ)
        A(L) = XX
80      IF(I-II)81,82,86
81      M = (I-1)*NOOB + II - ((I - 1) * I) / 2
        GO TO 87
82      IF(NATA(I))83,96,83
83      I = NATA(I)
        GO TO 80
86      M = (II - 1) * NCCB + I - ((II - 1) * II) / 2
87      IF(I - JJ)88,90,90
88      MM = (I - 1) * NCCB + JJ - ((I - 1) * I) / 2
        GO TO 92
90      MM = (JJ - 1) * NOOB + I - ((JJ - 1) * JJ) / 2
92      L=(I-1)*NOOB+I-((I-1)*I)/2
        XX = SIGN(1.,AMAX1(A(M),A(MM)))
        A(M) = ABS(A(M))
        A(MM) = ABS(A(MM))
        GO TO (94,94,94,94,94,93),KODE
93      A(M) = AMAX1(A(M),A(MM))
        GO TO 95
94      A(M) = (A(MM)* (SUMX(I) + SUMX(JJ)) + A(M)*(SUMX(I)+SUMX(II)) +
1B1 - B2 - A(L) * SUMX(I)) / (SUMX(I) + SUMX(JJ) + SUMX(II))
95      A(M) = SIGN(A(M),XX)
        A(MM) = A(M)
        GO TO 82
96      SUMX(II) = SUMX(II) + SUMX(JJ)
C       IF(SXSQ(JJ).GE.(0.14455217E-78)) GO TO 1196
C       WRITE(6,1197) SXSQ(II),SXSQ(JJ),(II),(JJ)
C1197 FORMAT('0',130('-'))/'0',2E15.8,214/'0',130('-'))
1196   SXSQ(II) = SXSQ(II) + SXSQ(JJ)
        I = 1
        NN = NCYC + 2

```

```

KEY = MM - 1
C          ITEM(I) GETS INDICES OF GROUP MEMBERS FOR PRINTING
C          ITEM(I) HAS ORDERED LIST AT CONCLUSION OF GROUPING
C          AND ALSO HAS BOTH RIGHT AND LEFT HALVES OF MINX(I)
100      J = I
        KK = 0
102      KK = KK + 1
        ITEM(KK) = J
        J = XILHF(J)
        IF(XILHF(MINX(J)))106,108,106
106      J = MINX(J)
        GO TO 102
108      M = (I - 1) * NOOB + I - ((I - 1) * I) / 2
        IF(XILHF(MINX(I)))110,109,110
109      KEY = KEY + 1
        ITEM(KEY) = I
        GO TO 118
110      DEFR = KK - 1
        RMAX = A(M)
        IF(NCLLS - 2)112,112,118
112      GO TO (111,117,117,111,111,117),KODE
111      RMAX = RMAX / DEFR
117      CALL UPKMAT(JTEM,ITEM,KK)
        IF(I1 - ITEM(1))116,115,116
116      WRITE(NT,2002)KK,RMAX,(JTEM(K),K=1,KK)
        GO TO 118
115      WRITE(NT,2004)NCYC,KK,RMAX,(JTEM(K),K = 1,KK)
118      IF(NATA(I))119,120,119
119      I = NATA(I)
        GO TO 100
120      IF(KEY - MM)124,122,122
122      KK = KEY - MM + 1
        WRITE(NT,2003)KK,(ITEM(K),K=MM,KEY)
124      WRITE(NT,2006)
        IF(NOOB - 1 - NCYC)130,130,20
C
C          END OF GROUPING LOOP
130      IF(NCLLS - 2)149,129,149
129      DO 131 I=1,NOOB
        LL = NOOB + 1 - I
131      MINX(LL) = ITEM (I)
        NN = ((NOOB - 1) / 16) * 16 + 1
        CALL UPKMAT(JTEM,MINX,LL)
        DO 148 I = 1,NN,16
        LL=MIN0(I+15,NOOB)
        MM = NOOB - I + 1
        DO 146 M = 1,MM
        IF(MOD(M-1,50))139,138,139
138      WRITE(NT,2017)TITLE,NPAGE
        NPAGE = NPAGE + 1
        WRITE(NT,2005)(JTEM(K),K = I,LL)
        WRITE(NT,2006)
139      J=AND(ITEM(M),(2**16-1))
        LL=MIN0(LL,NOOB-M+1)
        DO 144 K=1,LL
        L=AND(MINX(K),(2**16-1))
        IF(J-L)140,142,142
140      N=L+((2*NOOB-J)*(J-1))/2
        GO TO 144
142      N=J+((2*NOOB-L)*(L-1))/2
144      SUMX(K)=A(N)

```

```

146 WRITE (NT,2007) J, (SUMX (K), K=1, LL)
148 CONTINUE
149 I5=0
160 I1=I5+1
      I2=I5+2
      I3=I5+3
      I4=I5+4
      I5=MINO (I5+100, NCOB)
      WRITE (NT, 2018) TITLE, NPAGE
      NPAGE=NPAGE+1
      DO 150 I=I1, I5
150  DATA (I)=BLANK
      CALL LPKMAT (JTEM, ITEM, 195)
      WRITE (NT, 2008) (JTEM (I), I=I1, I5, 4)
      WRITE (NT, 2012) (JTEM (I), I=I2, I5, 4)
      WRITE (NT, 2015) (JTEM (I), I=I3, I5, 4)
      WRITE (NT, 2014) (JTEM (I), I=I4, I5, 4)
      DC 153 I=1, NCOB
      DC 151 L=1, NCOB
      J=L+1
      JJ=XIFHF (ITEM (J))
      IF (I-JJ) 151, 152, 151
151  CONTINUE
152  K=J-1
      DATA (K)=JX
      II=XILHF (ITEM (K))
      JJ=XILHF (ITEM (J))
      WRITE (NT, 2009) I, II, JJ, TG12 (I), (DATA (M), M=I1, I5), JY
153  CCNTINUE
      JY=JZ
      N2=I5+2
      WRITE (NT, 2010) (JY, I=I1, N2)
      WRITE (NT, 2011) (JTEM (I), I=I1, I5, 4)
      WRITE (NT, 2012) (JTEM (I), I=I2, I5, 4)
      WRITE (NT, 2013) (JTEM (I), I=I3, I5, 4)
      WRITE (NT, 2014) (JTEM (I), I=I4, I5, 4)
      IF (NUOB-15) 161, 161, 160
161  RETURN
C
2000  FORMAT (23H1NUMBER OF OBSERVATIONS 18,50X,39H SUBROUTINE GRPING
* DATED 20 DEC 67 // 23H0GROUPING CODE (KODE) 18,53X, 34H OBSER
* VATIONS ARE GROUPED ACCORDING /'0',83X,'TO SPECIFICATION OF KODE'
*, / '0')
2001  FORMAT (5H STEP 14,6X, 5H SUM = E15.6,13H MIN DIST = E15.6,
* 11H FOR ITEMS 15,5H AND 15,6H WITH 215,9H MEMBERS E18.6)
2002  FORMAT (9X,13,2X,F13.4,7H ITEMS 24I4 / (34X,24I4))
2003  FORMAT (14X,13,10X,7H SINGLES 24I4 / (34X,24I4))
2004  FORMAT (5H STEP 13,14,2X,F13.4,7H ITEMS 24I4 / (34X,24I4))
2005  FORMAT (8H OBS NO 16I7)
2006  FORMAT (1H0)
2007  FORMAT (1X,14,3X,16F7.2)
2008  FORMAT (19H0 ITEMS GROUPES 9X,25I4)
2009  FORMAT (1X,14,215,E14.5,2H I,10I1)
2010  FORMAT (30X,10Z A1)
2011  FORMAT (28X,25I4)
2012  FORMAT (29X,25I4)
2013  FORMAT (30X,25I4)
2014  FORMAT (31X,25I4)
2015  FORMAT (25H ETAPE I J VALEUR 5X,25I4)
2016  FORMAT ('1',8A4,T56,'GROUPING SUMMARY ',T120,'PAGE ',I3)

```

```

2017 FORMAT('1',8A4,T56,'DISTANCES IN ORDER GROUPED',T120,'PAGE ',I3)
2018 FORMAT('1',8A4,T56,'STEP GRAPH OF GROUPING',T120,'PAGE ',I3)
END
SUBROUTINE CNTGTY(A,NCCB,NSYZ,NSYSIN)
DIMENSION A(1)
K=0
I1=NSYZ+1
I2=NSYZ+NOOB
DC 20 I=1,NOOB
READ(NSYSIN,1001)(A(J),J = I1,I2)
WRITE(6,1001)(A(J),J = I1,I2)
DO 5 J = I1,I2
5 IF(A(J).EQ.0.) A(J)=-1
DC 10 J=I,NOOB
L=NSYZ+J
K=K+1
10 A(K)=SIGN(A(K),A(L))
20 CONTINUE
C
1001 FORMAT(72F1.0)
RETURN
END
SUBROUTINE ERRCR(WORD,NUMBER,NOUT,WD)
INTEGER WORD(2),WD
WRITE(NOUT,2001)WORD,NUMBER,WD
CALL EXIT
C
2001 FORMAT('0*** ERRCR STCP *** - ROUTINE ',A4,A2,' AT STATEMENT ',
*16,18X,10HTROUBLE IS,10X,A4)
END
SUBROUTINE SLH(I1,I2)
INTEGER AND,OR
I2=AND(I2,(-(2**16)))
I2=CR(I2,I1)
I2=AND(I2,(2**16-1))
2020 FORMAT(' ',I1 = ',18,'I2 = ',18)
RETURN
END
SUBROUTINE SRH(I1,I2)
INTEGER OR
I3=I1*2**16
I2=CR(I3,I2)
RETURN
END
SUBROUTINE UPKMAT(JTEM,ITEM,KK)
INTEGER AND
DIMENSION JTEM(KK),ITEM(KK)
DC 1 IJK=1,KK
1 JTEM(IJK) = AND(ITEM(IJK),(2**16-1))
RETURN
END
INTEGER FUNCTION XIRHF(I1)
XIRHF=I1/2**16
RETURN
END
INTEGER FUNCTION XILHF(I1)
INTEGER AND
XILHF=AND(I1,(2**16-1))
RETURN
END

```

```
//C EXEC LKCOMPGR, PARM=LKED='XREF,LIST',
// CCND=(4,LT,A.ASM),(4,LT,B.FORT)), REGION.GO=250K, TIME.GO=70
//LKED.SYSLIB DD DSN=SYS1.FORTLIB,DISP=SHR
//LKED.SYSIN DD *
ENTRY MAIN
```

```
//GC.FTC6F001 DD SYSOUT=A,DCB=(LRECL=133,RECFM=FBA,BLKSIZE=917)
//GC.FTC5F001 DD *
```

```
PRGB          411      1          2          22  UTILI. DU SOL VAR 1A12 NO ROTA
(4F8.3)
-0.423 -0.849  0.836 -0.152  0.0      0.0      0.0
-0.356 -0.963  0.831  0.230  0.0      0.0      0.0
-1.767  0.747  0.014 -0.285  0.0      0.0      0.0
-0.463 -0.003 -0.207 -1.355  0.0      0.0      0.0
 0.827 -0.891  0.418 -0.082  0.0      0.0      0.0
-0.260 -0.731  0.825 -0.657  0.0      0.0      0.0
-1.093 -1.066  0.844  0.003  0.0      0.0      0.0
-0.586  1.750  1.054  0.705  0.0      0.0      0.0
 0.760 -0.604  1.300 -1.019  0.0      0.0      0.0
 0.464  0.055  0.542  0.900  0.0      0.0      0.0
 0.270 -0.798  0.300  0.414  0.0      0.0      0.0
-0.031  2.888 -0.632  0.507  0.0      0.0      0.0
 2.182  0.835 -0.185 -0.134  0.0      0.0      0.0
-0.137 -0.852 -3.425  0.770  0.0      0.0      0.0
 2.156  0.910 -0.009 -0.073  0.0      0.0      0.0
 0.394 -1.352 -1.453  0.492  0.0      0.0      0.0
-1.118  0.344 -0.981 -1.641  0.0      0.0      0.0
-0.137 -0.028 -0.164 -1.371  0.0      0.0      0.0
-0.464 -0.002 -0.208 -1.350  0.0      0.0      0.0
 0.133 -0.910  0.328  0.535  0.0      0.0      0.0
-1.247  0.211  0.185  3.403  0.0      0.0      0.0
 0.495  0.094  0.243 -0.461  0.0      0.0      0.0
```

```

  1      1      1
  1      1      11     1
  1      1      1
111      11
11
```

```

      11
  1 1      1      1
  1 1 1      11
  1      1
```

```

      1
      1
      1
      1
```

FINS

```
// EXEC FORTGCLG, PARM='BCD', REGION.GO=200K, TIME.GO=3
//FORT.SYSIN DD * SOURCE DECK NEXT
C *** DISCRIMINANT ITERATIONS PROGRAM
```



```

C   *** READ AND PRINT DATA
    WRITE (6,103)
103  FORMAT (/40X,4FDATA/40X4H----)
    K=1
    N=1
    3  WRITE (6,104) N
104  FORMAT (1X6HSUESET, I3)
    KPREV=K
    4  READ(5,FMT)(NAME(K,J),J=1,3),(X(K,J),J=1,NVAR)
    IF (X (K,1) -999.) 6,7,6
    6  WRITE(6,105) K, (NAME(K,J),J=1,3),(X(K,J),J=1,NVAR)
105  FORMAT(3X,I3,2X,3A4,1X,10F11.4/20X,F11.4)
    K=K+1
    GO TO 4
C   *** END OF DATA SUBSET
    7  WRITE(6,99)(NAME(K,J),J=1,3),(X(K,J),J=1,NVAR)
99  FORMAT(8X,3A4,1X,10F11.4/20X,5F11.4)
    NROW(N)= K-KPREV
    IF (X(K,2)-999.) 8,9,8
    8  N=N+1
    GO TO 3
C
C   *** END OF DATA
    9  MAXRC = K-1
    DO 10 K=1,MAXRC
    10  IL(K)=K
    DC 235 M=1,N
235  NROW(M)=NROW(M)
C
C   *** (REENTRY POINT SUCCESSIVE ITERATIONS)
C   *** COMPUTE SUMS AND AVERAGES
    5  DC 13 J=1,NVAR
    SUMTOT(J)=0.0
    K=0
    DC 12 M=1,N
    NR=NROW(M)
    SUMSET(M,J)=0.0
    IF (NR) 241,241,240
240  DO 11 I=1,NR
    K=K+1
    11  SUMSET(M,J) = SUMSET(M,J) + X(K,J)
241  XBAR(M,J)=SUMSET(M,J)/FLOAT(NR)
    12  SUMTOT(J) = SUMTOT(J) + SUMSET(M,J)
    13  XBARCV(J)=SUMTOT(J)/FLOAT(MAXRC)
C
C   **** PRINT AVERAGES
    PRINT 1000
    WRITE (6,106)
106  FORMAT (/40X,8FAVERAGES/40X,8H-----//)
    DC 20 M=1,N
    20  WRITE (6,107) M,(XBAR(M,J),J=1,NVAR)
107  FORMAT(8X,'SUBSET',I3,4X,10F11.4/21X,5F11.4)
    WRITE (6,108) (XBARCV(J),J=1,NVAR)
108  FORMAT(8X,'OVERALL',6X,10F11.4/21X,5F11.4)
C
C   *** DEVELOP D MATRIX
    DO 14 M=1,N
    DC 14 J=1,NVAR
    14  D(M,J) = XBAR(M,J) -XBAROV(J)

```

```

C      ***** DEVELOP V MATRIX
      K=L
      DO 45 M=1,N
      NR=NRGW(M)
      IF (NK) 45,45,44
44     DO 16 I=1,NR
      K=K+1
      DO 16 J=1,NVAR
16     V(K,J) = X(K,J) - XBAR(M,J)
45     CCNTINUE
C      *** TRANSPOSE D AND V MATRICES
      DC 15 I=1,N
      DC 15 J=1,NVAR
15     DT(J,I) = D(I,J)
      DC 17 I=1,MAXRC
      DC 17 J=1,NVAR
17     VT(J,I) = V(I,J)
C
C      ***** DEVELOP DD, VV MATRICES, AND PRINT OUT IF
      DO 19 I=1,NVAR
      DC 19 J=1,NVAR
      DD(I,J) =0.0
      VV(I,J) =0.0
      DC 18 K=1,N
18     DD(I,J) = DD(I,J) + DT(I,K) * D(K,J)
      DO 19 K=1,MAXRC
19     VV(I,J) = VV(I,J) + VT(I,K) * V(K,J)
      IF (PROPT) 201,201,200
200    WRITE (6,109)
109    FCRMAT (/ /40X10HMATRIX DTD//)
      DO 21 I=1,NVAR
21     WRITE (6,110) (DD(I,J),J=1,NVAR)
110    FCRMAT (10X,10F12.4)
      WRITE (6,111)
111    FCRMAT (/ /40X,10HMATRIX VTV//)
      DO 22 I=1,NVAR
22     WRITE (6,110) (VV(I,J),J=1,NVAR)
C
201    CALL DIRNM(DD,NVAR,VV,RES,RCCTR)
C      ***** PRINT ROOTR
      PRINT 1000
      DO 67 I=1,NVAR
67     WRITE (6,117) I,ROOTR(I)
117    FCRMAT (5X15,F20.8)
C      PRINT RESULT MATRIX
      WRITE (6,143)
143    FCRMAT (/ /40X13HRESULT MATRIX//)
      DO 65 I=1,NVAR
65     WRITE (6,110) (RES(I,J),J=1,NVAR)
C
C      *** COMPUTE XR
      DC 206 K=1,MAXRC
      DC 206 J=1,NVAR
      XR(K,J)=0.0
      DC 206 L=1,NVAR
206    XR(K,J) = XR(K,J) + X(K,L) * RES(L,J)
C
C      *** PRINT XR IF
      PRINT 1000
      IF (PROPT) 208,208,207

```

```

207 WRITE (6,100) TITL
WRITE (6,127) IT
WRITE (6,144)
144 FORMAT (//40X9FMATRIX XR//)
K=C
DC 85 M=1,N
WRITE (6,104) M
NR=NRCW(M)
IF (NR) 85,85,84
84 DC 81 I=1,NR
K=K+1
81 WRITE(6,105) K,(NAME(K,J),J=1,3),(XR(K,J),J=1,NVAR)
85 CONTINUE
C
C *** COMPUTE XAVR
208 DC 82 M=1,N
DO 82 J=1,NVAR
XAVR(M,J)=0.0
DC 82 L=1,NVAR
82 XAVR(M,J) = XAVR(M,J) + XBAR(M,L) * RES(L,J)
C
C *** PRINT XAVR IF
PRINT 1000
IF (PROPT) 210,210,209
209 WRITE (6,145)
145 FORMAT (//40X11HMATRIX XAVR//)
DC 83 M=1,N
83 WRITE(6,107) M,(XAVR(M,J),J=1,NVAR)
C
C *** SET FORMAT CONTROL FOR THIS WIDTH
210 MM=N
IF (MM-4) 213,213,214
213 IFMT=1
GO TO 217
214 IF (MM-7) 215,215,216
215 IFMT=2
GO TO 217
216 IFMT=3
C
C *** COMPUTE DISTANCE MATRIX
217 DO 211 K=1,MAXRO
DC 211 MM=1,N
DIST(K,MM) =0.0
DO 212 L=1,NVAR
212 DIST(K,MM) = DIST(K,MM) + (XR(K,L)-XAVR(MM,L))**2
211 DIST(K,MM)=SQRT(DIST(K,MM))
C
C *** COMPUTE RANKS (WHICH COL OF GIVEN ROW HAS LEAST VALUE)
C *** STORE RANKS IN TEMP VECTOR
DO 224 K=1,MAXRO
DISTLC=DIST(K,1)
IRTEMP(K)=1
DC 224 MM=2,N
IF (DIST(K,MM) -DISTLC) 223,224,224
223 DISTLO=DIST(K,MM)
IRTEMP(K)=MM
224 CONTINUE
C
C *** PRINT DISTANCE MATRIX, AND CURRENT RANKS
C *** DO TITLES

```

```

PRINT 1000
WRITE (6,100) TITL
WRITE (6,127) IT
WRITE (6,121) (MM,MM=1,N)
121 FCRMAT (/ /28X37H-DISTANCE FROM AVERAGES OF SUBSETS.... /
1 1HJ,20X,10(18,2X))
GO TO (218,219,220),1FMT
218 WRITE (6,151)
GO TO 221
219 WRITE (6,152)
GO TO 221
220 WRITE (6,153)
151 FORMAT (67X,4HRANK)
152 FCRMAT (92X,4HRANK)
153 FORMAT (122X,4HRANK)
C *** CC BODY
221 K=0
DO 229 M=1,N
WRITE (6,104) M
NR=NRCW(M)
IF (NR) 229,229,222
222 CC 228 I=1,NR
K=K+1
WRITE(6,154)K,(NAME(K,J),J=1,3),(DIST(K,MM),MM=1,N)
154 FCRMAT(3HJ ,13,2X,3A4,1X,10F10.4/20X,5F10.4)
GO TO (225,226,227),1FMT
225 WRITE (6,161) IRTEMP(K)
GO TO 228
226 WRITE (6,162) IRTEMP(K)
GO TO 228
227 WRITE (6,163) IRTEMP(K)
161 FORMAT (68XI2)
162 FCRMAT (93XI2)
163 FORMAT (123XI2)
228 CONTINUE
229 CCNTINUE
C
C *** UNSCRAMBLE TEMP RANKS, STORE PERM FOR THIS ITER
CC 50 K=1,MAXRC
L=ID(K)
50 IRANK(L,IT) = IRTEMP(K)
C
IF (IT-1) 24,24,25
25 IF (IT-MAXIT) 26,70,70
C
C *** ARE RESULTS THIS ITER SAME AS PREVIOUS
26 DO 27 K=1,MAXRC
IF (IRANK(K,IT) - IRANK(K,IT-1)) 24,27,24
27 CONTINUE
GO TO 70
C
24 KNU=C
DO 234 M=1,N
NRCW(M)=0
CC 234 K=1,MAXRC
IF (IRTEMP(K)-M) 234,231,234
231 NRCW(M)=NROW(M)+1
KNU=KNU+1
DO 232 J=1,NVAR
232 XALT(KNU,J) = X(K,J)

```

```

DO 233 J=1,3
233 NAMALT(KNU,J) = NAME(K,J)
    IDALT(KNU) = IC(K)
234 CONTINUE
C
C   *** COPY REARRANGED DATA BACK INTO WORKING LOCATIONS
DO 35 K=1,MAXRC
DO 34 J=1,NVAR
34 X(K,J) = XALT(K,J)
    ID(K)=IDALT(K)
DO 35 J=1,3
35 NAME(K,J) = NAMALT(K,J)
C
C   *** PROCESS NEW DATA FROM SCRATCH
IT=IT+1
PRINT 1000
WRITE (6,100) TITL
WRITE (6,127) IT
GO TO 5
C
C
C
C   *** INTERRUPT ITERS (20 ITERS OR CONVERGING)
70 WRITE (6,100) TITL
71 WRITE (6,123) IT
123 FORMAT (/5X10HRESULTS OF ,I3,11H ITERATIONS/5X,24(1H-))
WRITE (6,124) (J,J=1, IT)
124 FORMAT (42X,20I3)
C
C   *** SORT NAMES BACK INTO ORIGINAL ROWS
DO 36 K=1,MAXRC
L=ID(K)
DO 36 J=1,3
36 NAMALT(L,J) = NAME(K,J)
C
C   *** PRINT TABLE OF RANKS FOR ALL ITERATIONS
PRINT 1000
K=0
DO 73 M=1,N
WRITE (6,164) M
164 FORMAT (/20X,6F5.2,13)
NR=NCRIG(M)
DO 73 I=1,NR
K=K+1
73 WRITE(6,125)K, (NAMALT(K,J),J=1,3), (IRANK(K,J),J=1,IT)
125 FORMAT(22X,I3,3X,3A4,2X,20I3)
GO TO 2
999 STOP
END
C
SUBROUTINE DIRNM.
C
SUBROUTINE DIRNM, DIAGONALIZATION OF A REAL NONSYMMETRIC MATRIX
OF THE FORM B-INVERSE * A.
C
A AND B ARE M BY M INPUT MATRICES. UPON RETURN VECTOR XL CONTAINS
THE EIGENVALUES OF B-1*A, AND MATRIX X CONTAINS THE EIGENVECTORS
IN ITS COLUMNS, NORMALIZED. SUBROUTINE HDIAG IS REQUIRED.
C
A, M, B, X, AND XL ARE DUMMY NAMES AND MAY BE CHANGED IN THE
CALLING STATEMENT.
C
SUBROUTINE DIRNM (A, M, B, X, XL)

```

```

C   DIMENSION A(10,10), B(10,10), X(10,10), XL(10)
C   DIMENSION A(15,15), B(15,15), X(15,15), XL(15)
C   CALL HDIAG (B, M, U, X, NR)
C   DO 1 I = 1, M
C   1 XL(I)=1.0/SQRT(AES(B(I,I)))
C   DO 2 I = 1, M
C   DO 2 J = 1, M
C   2 B(I,J) = X(I,J) * XL(J)
C   DO 3 I = 1, M
C   DO 3 J = 1, M
C   X(I,J) = 0.0
C   DO 3 K = 1, M
C   3 X(I,J) = X(I,J) + B(K,I) * A(K,J)
C   DO 4 I = 1, M
C   DO 4 J = 1, M
C   A(I,J) = 0.0
C   DO 4 K = 1, M
C   4 A(I,J) = A(I,J) + X(I,K) * E(K,J)
C   A NOW CONTAINS B-1/2PRIME * A * B-1/2 OF THE NOTES.
C   TRACE = 0.0
C   DO 10 I = 1, M
C   10 TRACE = TRACE + A(I,I)
C   WRITE (6,11) TRACE
C   11 FORMAT (33HTRACE OF B-1/2PRIME *A* B-1/2 = F14.7)
C   CALL HDIAG (A, M, U, X, NR)
C   DO 5 I = 1, M
C   5 XL(I) = A(I,I)
C   SUMR = 0.0
C   DO 12 I = 1, M
C   SUMR = SUMR + XL(I)
C   12 WRITE (6,13) I,XL(I)
C   13 FORMAT (12H0EIGENVALUE I2,F14.7)
C   WRITE (6,14) SUMR
C   14 FORMAT (22H0SUM OF EIGENVALUES = F14.7)
C   DO 6 I = 1, M
C   DO 6 J = 1, M
C   A(I,J) = 0.0
C   DO 6 K = 1, M
C   6 A(I,J) = A(I,J) + B(I,K) * X(K,J)
C   DO 9 J = 1, M
C   SUMV = 0.0
C   DO 7 I = 1, M
C   7 SUMV = SUMV + (A(I,J)**2)
C   DEN=SQRT(SUMV)
C   DO 8 I = 1, M
C   8 X(I,J) = A(I,J) / DEN
C   9 CONTINUE
C   COLUMNS OF X(I,J) ARE NOW NORMALIZED.
C   RETURN
C   END
C   SUBROUTINE HDIAG.

```

```

C   PROGRAMMED BY F. J. CCRBATO AND M. MERWIN OF THE M. I. T.
C   COMPUTATION CENTER.

```

```

C   THIS SUBROUTINE COMPUTES THE EIGENVALUES AND EIGENVECTORS
C   OF A REAL SYMMETRIC MATRIX, H, OF ORDER N ( WHERE N MUST BE LESS
C   THEN 51), AND PLACES THE EIGENVALUES IN THE DIAGONAL ELEMENTS OF
C   THE MATRIX H, AND PLACES THE EIGENVECTORS (NORMALIZED) IN THE
C   COLUMNS OF THE MATRIX U. IEGEN IS SET AS 1 IF ONLY EIGENVALUES

```

```

C ARE DESIRED, AND IS SET TO 0 WHEN VECTORS ARE REQUIRED. NR CON-
C TAINS THE NUMBER OF ROTATIONS DONE.
C
C H, N, IEGEN, U, AND NR OF THE ARGUMENT LIST ARE DUMMY VARIABLES
C AND MAY BE NAMED DIFFERENTLY IN THE CALLING OF THE SUBROUTINE.
C
C SUBROUTINE PLACES COMPUTER IN THE FLOATING TRAP MODE
C THE SUBROUTINE OPERATES ONLY ON THE ELEMENTS OF H THAT ARE TO THE
C RIGHT OF THE MAIN DIAGONAL. THUS, ONLY A TRIANGULAR
C SECTION NEED BE STORED IN THE ARRAY H.
C SUBROUTINE HCDIAG (H,N,IEGEN,U,NR)
C DIMENSION H(10,10), U(10,10), X(10), IQ(10)
C DIMENSION H(15,15),U(15,15),X(15),IQ(15)
C IF (IEGEN) 15,10,15
10 DO 14 I=1,N
    DO 14 J=1,N
    IF(I-J)12,11,12
11 U(I,J)=1.0
    GO TO 14
12 U(I,J)=0.
14 CONTINUE
15 NR = 0
    IF (N-1) 1000,1000,17
C SCAN FOR LARGEST OFF-DIAGONAL ELEMENT IN EACH ROW
C X(I) CONTAINS LARGEST ELEMENT IN ITH ROW
C IQ(I) HOLDS SECCND SUBSCRIPT DEFINING POSITION OF ELEMENT
17 NMII=N-1
    DO 30 I=1,NMII
    X(I) = 0.
    IPL1=I+1
    DO 30 J=IPL1,N
    IF(X(I)-ABS(H(I,J)))20,20,30
20 X(I)=ABS(H(I,J))
    IQ(I)=J
30 CONTINUE
C SET INDICATOR FOR SHUT-OFF.RAP=2**-27,NR=NO. OF ROTATIONS
RAP=7.450580596E-9
HDTEST=1.0E38
C FIND MAXIMUM OF X(I) S FOR PIVOT ELEMENT AND
C TEST FOR END OF PROBLEM
40 DO 70 I=1,NMII
    IF (I-1) 60,60,45
45 IF ( XMAX- X(I)) 60,70,70
60 XMAX=X(I)
    IPIV=I
    JPIV=IQ(I)
70 CONTINUE
C IS MAX. X(I) EQUAL TO ZERO, IF LESS THAN HDTEST, REVISE HDTEST
IF ( XMAX) 1000,1000,80
80 IF (HDTEST) 90,90,85
85 IF (XMAX - HDTEST) 90,90,148
90 HDIMIN=ABS(H(1,1))
    DO 110 I= 2,N
    IF(HDIMIN-ABS(H(I,I)))110,110,100
100 HDIMIN=ABS(H(I,I))
110 CONTINUE
    HDTEST=HDIMIN*RAP
C RETURN IF MAX.H(I,J) LESS THAN(2**-27)ABS(H(K,K)-MIN)
IF (HDTEST- XMAX) 148,1000,1000
148 NR = NR+1

```

```

C      COMPUTE TANGENT, SINE AND COSINE,H(I,I),H(J,J)
150 TANG=SIGN(2.0,(H(IPIV,IPIV)-H(JPIV,JPIV)))*H(IPIV,JPIV)/(ABS(H(IPI
  *V,IPIV)-H(JPIV,JPIV))+SQRT((H(IPIV,IPIV)-H(JPIV,JPIV))**2+4.0*H(IP
  *IV,JPIV)**2))
  CCSINE=1.0/SQRT(1.0+TANG**2)
  SINE=TANG*CCSINE
  HII=H(IPIV,IPIV)
  H(IPIV,IPIV)=CCSINE**2*(HII+TANG*(2.*H(IPIV,JPIV)+TANG*H(JPIV,JPIV
  1)))
  H(JPIV,JPIV)=CCSINE**2*(H(JPIV,JPIV)-TANG*(2.*H(IPIV,JPIV)-TANG*H
  11))
  H(IPIV,JPIV)=0.
C      PSEUDOC RANK THE EIGENVALUES
C      ADJUST SINE AND CCS FOR COMPUTATION OF H(IK) AND U(IK)
  IF (H(IPIV,IPIV) - H(JPIV,JPIV)) 152,153,153
152 HTEMP = H(IPIV,IPIV)
  H(IPIV,IPIV) = H(JPIV,JPIV)
  H(JPIV,JPIV) = HTEMP
C      RECCMPUTE SINE AND COS
  HTEMP=SIGN(1.0,-SINE)*CCSINE
  CCSINE=ABS(SINE)
  SINE = HTEMP
153 CONTINUE
C      INSPECT THE IQS BETWEEN I+1 AND N-1 TO DETERMINE
C      WHETHER A NEW MAXIMUM VALUE SHOULD BE COMPUTED SINCE
C      THE PRESENT MAXIMUM IS IN THE I OR J ROW.
  DC 350 I=1,NM1
  IF(I-IPIV)210,350,200
200 IF(I-JPIV)210,350,210
210 IF(IQ(I)-IPIV)230,240,230
230 IF(IQ(I)-JPIV)350,240,350
240 K=IQ(I)
250 HTEMP=H(I,K)
  H(I,K)=0.
  IPL1=I+1
  X(I) =0.
C      SEARCH IN DEPLETED ROW FOR NEW MAXIMUM
  DC 320 J=IPL1,N
  IF(X(I)-ABS(H(I,J)))300,300,320
300 X(I)=ABS(H(I,J))
  IQ(I)=J
320 CONTINUE
  H(I,K)=HTEMP
350 CCNTINUE
  X(IPIV) =0.
  X(JPIV) =0.
C      CHANGE THE OTHER ELEMENTS OF H
  DC 530 I=1,N
  IF(I-IPIV)370,530,420
370 HTEMP = H(I,IPIV)
  H(I,IPIV) = CCSINE*HTEMP + SINE*H(I,JPIV)
  IF(X(I)-ABS(H(I,IPIV)))380,390,390
380 X(I)=ABS(H(I,IPIV))
  IQ(I) = IPIV
390 H(I,JPIV) = -SINE*HTEMP + COSINE*H(I,JPIV)
  IF(X(I)-ABS(H(I,JPIV)))400,530,530
400 X(I)=ABS(H(I,JPIV))
  IQ(I) = JPIV
  GC TC 530
420 IF(I-JPIV)430,530,480

```

```

430 HTEMP = H(IPIV,I)
H(IPIV,I) = CCSINE*HTEMP + SINE*H(I,JPIV)
IF(X(IPIV)-ABS(H(IPIV,I)))440,450,450
440 X(IPIV)=ABS(H(IPIV,I))
IQ(IPIV) = I
450 H(I,JPIV) = -SINE*HTEMP + COSINE*H(I,JPIV)
IF(X(I)-ABS(H(I,JPIV)))400,530,530
480 HTEMP = H(IPIV,I)
H(IPIV,I) = COSINE*HTEMP + SINE*H(JPIV,I)
IF(X(IPIV)-ABS(H(IPIV,I)))490,500,500
490 X(IPIV)=ABS(H(IPIV,I))
IQ(IPIV) = I
500 H(JPIV,I) = -SINE*HTEMP + CCSINE*H(JPIV,I)
IF(X(JPIV)-ABS(H(JPIV,I)))510,530,530
510 X(JPIV)=ABS(H(JPIV,I))
IQ(JPIV) = I
530 CONTINUE
C TEST FOR COMPUTATION OF EIGENVECTORS
IF(IEGEN)40,540,40
540 DC 550 I=1,N
HTEMP=U(I,IPIV)
U(I,IPIV)=COSINE*HTEMP+SINE*U(I,JPIV)
550 U(I,JPIV)=-SINE*HTEMP+COSINE*U(I,JPIV)
GC TC 40
1000 RETURN
END

```

```

//LKED.SYSLMOD DD DSNNAME=+DISCR(MAIN),DISP=(NEW,PASS),UNIT=SYS DA,
// SPACE=(1600,(160,50,1))
//GC.FTU6FU01 DD SYSCUT=A,DCB=(LRECL=131,RECFM=FBA,BLKSIZE=917)
//GC.FTU5FU01 DD *

```

VAR 1A 12

```

04106
(3A4/4F8.3)
STE-CATHERI.
-0.855 0.414 -0.926 0.925 0.0 0.0 0.0 0 0
CAUGHNAWAGA
-0.855 0.432 -0.941 0.925 0.0 0.0 0.0 0 0
PREVILLE
-1.457 -0.250 -1.336 1.112 0.0 0.0 0.0 0 0
CANCIAC
-0.915 0.997 -0.215 0.932 0.0 0.0 0.0 0 0
ST-ERUND
0.450 0.047 -0.877 1.140 0.0 0.0 0.0 0 0

999.000
999.000
GREENFIELD P
0.966 -0.908 -0.014 1.442 0.0 0.0 0.0 J 0

999.000
LONGUEIL
1.891 -0.714 -1.132 -0.255 0.0 0.0 0.0 0 0
LEMZYNE
2.091 -0.945 -1.179 -0.439 0.0 0.0 0.0 0 0
ST-LAMBERT
0.562 0.155 -0.077 0.147 0.0 0.0 0.0 0 0

```

```

999.000
LAFLECHE

```

U.050	-0.156	0.974	-0.824	0.0	0.0	0.0	0	0
ST-CONSTANT								
0.172	-0.219	0.986	-0.721	0.0	0.0	0.0	0	0
JACQUES-CART								
0.822	0.443	0.637	-0.700	0.0	0.0	0.0	0	0

999.000								
CHATEAUGLAY								
0.110	-0.780	0.501	-0.477	0.0	0.0	0.0	0	0
NCTRE-DANE								
-0.970	-2.000	-0.072	-1.359	0.0	0.0	0.0	0	0

999.000								
LERY								
-1.694	-0.391	-1.499	-2.205	0.0	0.0	0.0	0	0
ST-FILBERT								
-0.214	0.963	0.429	-0.888	0.0	0.0	0.0	0	0
DELSON								
-0.175	1.850	0.572	0.353	0.0	0.0	0.0	0	0
LA PRAIRIES								
1.023	1.801	-1.054	-0.940	0.0	0.0	0.0	0	0

999.000								
BOUCHERVILLE								
-0.116	-0.133	1.298	0.390	0.0	0.0	0.0	0	0
CHAT-HEIGHT								
-0.571	-0.786	1.056	0.728	0.0	0.0	0.0	0	0
CHAT-CENTRE								
0.017	-0.447	0.816	1.197	0.0	0.0	0.0	0	0
BROSSARD								
-0.405	0.224	1.566	-0.227	0.0	0.0	0.0	0	0

999.000 999.000

```
//S360 EXEC FORTGCLG,PARM='BCD',REGION.GD=100K,TIME.GD=3
//FORT.SYSIN DD *
```

```
C
C PROGRAM FOR THE GENERALIZED ANALYSIS OF COVARIANCE
```

```
C*****
```

```
C ORIGINALLY PROGRAMMED IN UNIVERSITY OF IOWA DEPT. OF GEOGRAPHY
C BY STANLEY KNEIPP
```

```
C **** THERE ARE FOUR CONTROL CARDS ****
C CARD ONE -- TITLE CARD CAN USE ALL 80 COLUMNS
C CARD TWO -- PARAMETER CARD WITH OPTIONS
```

```
C          COLS 1 - 2 NUMBER OF VARIABLES
C          COLS 3 - 4 NUMBER OF GROUPS
C          COLS 5 - 6 OPTION FOR TEST OF HOMOSCEDASTICITY
C                      00 = DISCONTINUE EXECUTION
C                      01 = CONTINUE EXECUTION
C          COLS 7 - 8 OPTION FOR RESIDUALS FROM REGRESSION
C                      00 = NOT PRINTED
C                      01 = RESIDUALS PRINTED
C          COLS 9 - 10 SEQUENCE NUMBER OF THE DEPENDENT VARIABLE
```

```
C NOTE HERE THAT THIS PROGRAM IS GENERALIZED SO THAT IT CAN HANDLE
C MORE THAN ONE INDEPENDENT VARIABLE
```

```
C
```



```

590 JJ=NDEP-1
   DO 182 I=1,NNS
182 BART(I,1)=BEARS(I)
   NDEP=1
180 CONTINUE
   DC 141 I=1,NNS
141 WRITE (3,17) (BART(I,J),J=1,M)
   REWIND 3

```

```

C
C CHECK ON CORRECT TAPE READ
C

```

```

   DC 301 I=1,NNS
   READ (3,17) (CHECK(I,J),J=1,M)
301 PRINT 302, (CHECK(I,J), J=1,M)
302 FORMAT (12F10.3)
   REWIND 3
   JJ=1
   DC 351 L=1,NS
   N=IGRCUP(L)
   DO 351 I=1,N
   BART(I,L)=BART(JJ,1)
351 JJ=JJ+1
   DO 500 J=1,NS
   SUM=0.0
   N=IGRCUP(J)
   DC 501 I=1,N
501 SUM=SUM+BART(I,J)
   X=N
500 XBAR(J)=SUM/X
   ZSUM=0.0
   DO 30 J=1,NS
   ESUM(J)=0.0
   N=IGRCUP(J)
   X=N
   DC 502 I=1,N
502 ESUM(J)=ESUM(J)+((BART(I,J)-XBAR(J))**2)
   ZSUM=ZSUM+ESUM(J)
   VARI(J)=ESUM(J)/(X-1.0)
30 STD(J)=SQRT(VARI(J))
   Z=NNS-NS
   XMNSQ=ZSUM/Z
   BLCM=Z*(ALOG10(XMNSQ))
   WALK=0.0
   DC 580 J=1,NS
   RR=IGROUP(J)-1
580 WALK =WALK+(RR*(ALOG10(VARI(J))))
   TCP =2.3026*(BLCM-WALK)
   BB=NS-1
   C=1.0/(3.0*BB)
   AZCUE=0.0
   DO 503 J=1,NS
   RR=IGROUP(J)-1
503 AZCUE=AZCUE+(1.0/RR)
   AZCUE=AZCUE-(1.0/Z)
   TRESH=(C*AZCUE)+1.0
   CHI =TOP/TRESH
   IDF=NS-1
   WRITE(6,6) CHI
   WRITE(6,7) IDF
   IF(SIG(IDF)-CHI) 330,330,331

```

```

330 IF(NKK) 332,333,332
333 WRITE (6,8)
GO TO 300
332 WRITE(6,9)
GO TO 336
331 WRITE (6,10)
336 CCNTINUE
ASUM=C.0
XXX=0.0
DC 600 I=1,NNS
600 READ (3,17) (BART(I,J), J=1,M)
REWIND 3
MMM=M*M
DO 510 I=1,M
510 ESUM(I)=0.0
DC 511 I=1,MMM
511 DSUM(I)=0.0
DC 143 IJK=1,NS
N=IGRCUP(IJK)
WRITE (6,11) IJK
307 REWIND 4
CALL CORRE (N,M,U,X,XBAR,STD,RX,R,D,B,T)
REWIND 4
IF(XXX) 309,309,310
309 DO 196 I=1,MMM
196 DSUM(I)=DSUM(I)+RX(I)
DC 197 I=1,M
197 ESUM(I)=ESUM(I)+RX(I)
310 CONTINUE
IF(NS) 108,108,109
108 WRITE (6,12)
GO TO 300
109 CCNTINUE
CALL CRDR (M,R,NDEP,K,ISAVE,RX,RY)
WRITE(6,40) M,R(IJK),NDEP,K,ISAVE(IJK),RX(IJK),RY(IJK)
40 FCRMAT(15,F10.5,315,2F10.5)
CALL MINV(RX,K,DET,B,T)
IF (DET) 112,110,112
110 WRITE (6,13)
GO TO 300
112 CALL MULTR (N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)
MM=K+1
WRITE (6,14)
DC 115 J=1,K
L=ISAVE(J)
115 WRITE(6,15)L,XBAR(L),STD(L),RY(J),B(J),SB(J),T(J)
WRITE (6,16)
L=ISAVE(MM)
WRITE(6,15)L,XBAR(L),STD(L)
WRITE(6,18)ANS(1),ANS(2),ANS(3)
WRITE(6,19)
L=ANS(8)
WRITE(6,20)K,ANS(4),ANS(6),ANS(10),L,ANS(7),ANS(9)
L=N-1
SUM=ANS(4)+ANS(7)
WRITE (6,21)L,SUM
IF(NRESI) 700,700,120
120 WRITE (6,22)
MM=ISAVE(K+1)
DO 140 II=1,N

```

```

      READ(4) (W(J), J=1, M)
      SUM=ANS(1)
      DO 130 J=1, K
      L=ISAVE(J)
130  SLM=SUM+W(L)*B(J)
      RESI=W(MM)-SUM
140  WRITE (6,23) II, W(MM), SUM, RESI
700  REWIND 4
      IF(XXX) 380, 380, 308
380  ASUM=ASLM+ANS(7)
143  CONTINUE
      REWIND 3
      DO 610 I=1, NNS
610  WRITE (3, 17) (EART(I, J), J=1, M)
      WRITE (6, 24)
      XXX=1.0
      N=NNS
      REWIND 3
      GO TO 307
308  KK=K*K
      MN=M+M
      J=M+2
      DO 195 I=1, KK
      RX(I)=0.0
      RX(I)=DSUM(J)
      IF(J-MN) 200, 201, 195
200  J=J+1
      GO TO 195
201  MN=MN+M
      J=J+2
195  CONTINUE
      CALL MINV(RX, K, DET, B, T)
      DO 202 I=1, K
202  B(I)=C.0
      J=1
      DO 206 II=1, K
      FSUM=0.0
      DO 205 I=1, K
      FSLM=FSUM+(RX(J)*ESUM(I+1))
205  J=J+1
206  E(II)=FSUM
      FSLM=C.0
      DO 207 I=1, K
207  FSLM=FSUM+(B(I)*ESUM(I+1))
      SSW=ASUM
      SSCR=ESUM(1)-FSUM
      SST=ANS(7)
      SSRC=SSCR-SSW
      SSAAM=SST-SSCR
      IFW=NNS-(NS*M)
      DFW=IFW
      IFRC=(NS*M)-M-NS+1
      DFRC=IFRC
      IFCR=NNS-M-NS+1
      DFCR=IFCR
      IFAAM=NS-1
      DFAAM=IFAAM
      IFT=NNS-M
      DFT=IFT
      XSW=SSW/DFW

```

```

XSRC=SSRC/DFRC
XSCR=SSCR/DFCR
XSAAM=SSAAM/DFAAM
XST=SST/DFT
F1=XSRC/XSW
F2=XSAAM/XSCR
WRITE (6,25)
WRITE (6,26) IFW,SSW,XSW,IFRC,SSRC,XSRC,IFCR,SSCR,XSCR,IFAAM,SSAA
1M,XSAAM,IFT,SST,XST
WRITE (6,27)
WRITE (6,28) F1,IFRC,IFW,F2,IFAAM,IFCR
1 FCRMAT(20A4)
2 FCRMAT(1H1,20A4)
3 FCRMAT (5I2)
4 FCRMAT (10I5)
5 FCRMAT(20A4)
6 FCRMAT (1H0, 25HCHI-SQUARE TEST STATISTIC, F15.4)
7 FCRMAT(1H0,43HDEGREES OF FREEDOM FOR CHI-SQUARE STATISTIC,I6)
8 FCRMAT (1H0,53HDEPENDENT VARIABLES NOT HOMOSCEDASTIC JOB TERMINATE
1D.)
9 FCRMAT (1H0, 37HDEPENDENT VARIABLES NOT HOMOSCEDASTIC)
10 FCRMAT (1H0, 37HDEPENDENT VARIABLES ARE HOMOSCEDASTIC)
11 FCRMAT (1H1,50X, 27HREGRESSION ANALYSIS - GROUP, 15)
12 FCRMAT(48HNUMBER OF GROUPS NOT SPECIFIED. JOB TERMINATED.)
13 FCRMAT (41H0THE MATRIX IS SINGULAR. JOB TERMINATED.)
14 FCRMAT (9H0VARIABLE,5X,4HMEAN,6X,8HSTANDARD,6X,11HCORRELATION,4X,1
10HREGRESSION,4X,10HSTD. ERROR,5X,8HCOMPUTED/6H NO.,18X,9HDEVIATI
20N,7X,6HX VS Y,7X,11HCOEFFICIENT,3X,13HOF REG. COEFF.,3X,7HT VALUE)
15 FCRMAT (1H ,14,6F14.5)
16 FCRMAT (10H DEPENDENT)
18 FCRMAT (1H0/10H INTERCEPT,13X,F13.5//23H MULTIPLE CORRELATION ,F1
13.5//23H STD. ERROR OF ESTIMATE,F13.5//)
19 FCRMAT (1H0,21X,39HANALYSIS OF VARIANCE FOR THE REGRESSION//5X,19H
15SOURCE OF VARIATION,7X,7HDEGREES,7X,6HSUM OF,10X,4HMEAN,12X,7HF VA
22LUE/30X,10HOF FREEDOM,4X,7HSQUARES,9X,7HSQUARES)
20 FCRMAT(30H ATTRIBUTABLE TO REGRESSION ,16,3F16.5/30H DEVIATION F
1RGM REGRESSION ,16,2F16.5)
21 FCRMAT (1H ,5X,5HTOTAL,19X,16,F16.5)
22 FCRMAT (1H ,15X,18HTABLE OF RESIDUALS//9H CASE NO.,5X,7HY VALUE,5X
1,10HY ESTIMATE,6X,8HRESIDUAL)
23 FCRMAT (1H ,16,F15.5,2F14.5)
24 FCRMAT (1H1,50X,29HREGRESSION ANALYSIS FOR TOTAL)
25 FCRMAT(1H1,15X,36HANALYSIS OF COVARIANCE SUMMARY TABLE//,8X,7HSOUR
10CES,12X,7HDEGREES,8X,6HSUM OF,11X,4HMEAN/26X,10HOF FREEDOM,5X,7HSQ
20UARES,10X,7HSQUARES)
26 FCRMAT(25H WITHIN AREAS ,16,2F17.4/25H REGRESSION COEFF
11ICIENTS ,16,2F17.4/25H COMMON REGRESSION LINE ,16,2F17.4/25H ADJU
22STED MEANS ,16,2F17.4/25H TOTAL ,16,2F1
37.4)
27 FCRMAT (1H0,7X,7HF-TESTS,24X,4HTEST,9X,7HDEGREES/37X,9HSTATISTIC,5
1X,10HCF FREEDOM)
28 FCRMAT(21H PARALLEL REGRESSIONS,10X,F15.4,I9,1H,, 15/32H SIGNIFICA
15NCE OF AREAL GROUPINGS,F14.4,I9,1H,,15)
300 CONTINUE
GO TO 999
END
SUBROUTINE CURRE(N,M,L,X,XBAR,STD,RX,R,B,D,T)
DIMENSION XBAR (1),STD(1),RX(1),R(1),B(1),D(1),T(1)
DC 100 J=1,M

```

```

100 T(J)=0.0
    K=(M*M+M)/2
    DC 102 I=1,K
102 R(I)=0.0
    FN=N
    L=0
    IF(N-M) 130,130,135
130 KK=N
    GC TC 137
135 KK=M
137 DO 140 I=1,KK
    CALL DATA (M,D)
    DC 140 J=1,M
    T(J)=T(J)+D(J)
    L=L+1
140 RX(L)=D(J)
    FKK=KK
    DO 150 J=1,M
    XBAR(J)=T(J)
150 T(J)=T(J)/FKK
    L=0
    DC 180 I=1,KK
    JK=0
    DO 170 J=1,M
    L=L+1
170 D(J)=RX(L)-T(J)
    CC 180 J=1,M
    B(J)=B(J)+D(J)
    DO 180 K=1,J
    JK=JK+1
180 R(JK)=R(JK)+D(J)*D(K)
    IF (N-KK) 205,205,185
185 KK=N-KK
    DC 200 I=1,KK
    JK=0
    CALL DATA (M,D)
    DC 190 J=1,M
    XBAR(J)=XBAR(J)+D(J)
    C(J)=D(J)-T(J)
190 B(J)=B(J)+D(J)
    DO 200 J=1,M
    DO 200 K=1,J
    JK=JK+1
200 R(JK)=R(JK)+D(J)*D(K)
205 JK=0
    DO 210 J=1,M
    XBAR(J)=XBAR(J)/FN
    DC 210 K=1,J
    JK=JK+1
210 R(JK)=R(JK)-B(J)*B(K)/FN
    JK=0
    DO 220 J=1,M
    JK=JK+J
220 STD(J)=SQRT(AES(R(JK)))
    DO 230 J=1,M
    DO 230 K=J,M
    JK=J+(K*K-K)/2
    L=M*(J-1)+K
    RX(L)=R(JK)
    L=M*(K-1)+J

```

```

      RX(L)=R(JK)
      IF(STD(J)*STD(K)) 225,222,225
222 R(JK)=C.O
      GO TO 230
225 R(JK)=R(JK)/(STD(J)*STD(K))
230 CONTINUE
      FN=SQRT(FN-1.O)
      DC 240 J=1,M
240 STD(J)=STD(J)/FN
      L=-M
      DC 250 I=1,M
      L=L+M+1
250 B(I)=RX(L)
      RETURN
      END
      SUBROUTINE DATA (M,D)

```

DATA 029
 DATA 001
 DATA 002
 DATA 003
 DATA 004
 DATA 005
 DATA 006
 DATA 007
 DATA 008
 DATA 009
 DATA 010
 DATA 011
 DATA 012
 DATA 013
 DATA 014
 DATA 015
 DATA 016
 DATA 002
 DATA 018
 DATA 019
 DATA 020
 DATA 021
 DATA 022
 DATA 023
 DATA 024
 DATA 025
 DATA 026
 DATA 002
 DATA 028
 DATA 030

.....

SAMPLE INPLT SUBROUTINE - DATA

PURPOSE

READ AN OBSERVATION (M DATA VALUES) FROM INPLT DEVICE.
 THIS SUBROUTINE IS CALLED BY THE SUBROUTINE CORRE AND MUST
 BE PROVIDED BY THE USER. IF SIZE AND LOCATION OF DATA
 FIELDS ARE DIFFERENT FROM PROBLEM TO PROBLEM, THIS SUB-
 ROUTINE MUST BE RECOMPILED WITH A PROPER FORMAT STATEMENT.

USAGE

CALL DATA (M,D)

DESCRIPTION OF PARAMETERS

M - THE NUMBER OF VARIABLES IN AN OBSERVATION.
 D - OUTPLT VECTOR OF LENGTH M CONTAINING THE OBSERVATION
 DATA.

REMARKS

THE TYPE OF CONVERSION SPECIFIED IN THE FORMAT MUST BE
 EITHER F OR E.

SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED

NONE

.....

DIMENSION D(12)

THE FOLLOWING FORMAT MUST BE IDENTICAL TO FORMAT 17 IN MAIN PROGRAM

1 FCRMAT(9F10.3)

READ AN OBSERVATION FROM INPUT DEVICE.

READ (3,1) (D(I),I=1,M)

INPUT DATA ARE WRITTEN ON LOGICAL TAPE 04 FOR THE RESIDUAL ANALY-
 SIS PERFORMED IN THE SAMPLE MULTIPLE REGRESSION PROGRAM.

WRITE (4) (D(I),I=1,M)

DATA 034
 DATA 035
 DATA 036
 DATA 038
 DATA 039
 DATA 040
 DATA 041

```
RETURN
END
SUBROUTINE ORDER (M,R,NDEP,K,ISAVE,RX,RY)
DIMENSION R(1),ISAVE(1),RX(1),RY(1)
MM=0
DO 130 J=1,K
L2=ISAVE(J)
IF(NDEP-L2) 122,123,123
122 L=NDEP+(L2*L2-L2)/2
GO TO 125
123 L=L2+(NDEP*NDEP-NDEP)/2
125 RY(J)=R(L)
DO 130 I=1,K
L1=ISAVE(I)
IF(L1-L2) 127,128,128
127 L=L1+(L2*L2-L2)/2
GO TO 129
128 L=L2+(L1*L1-L1)/2
129 MM=MM+1
130 RX(MM)=R(L)
ISAVE(K+1)=NDEP
RETURN
END
SUBROUTINE MINV(A,N,D,L,M)
DIMENSION A(1),L(1),M(1)
D=1.0
NK=-N
DO 80 K=1,N
NK=NK+N
L(K)=K
M(K)=K
KK=NK+K
BIGA=A(KK)
DO 20 J=K,N
IZ=N*(J-1)
DO 20 I=K,N
IJ=IZ+I
10 IF(ABS(BIGA)-AES(A(IJ))) 15,20,20
15 BIGA=A(IJ)
L(K)=I
M(K)=J
20 CONTINUE
J=L(K)
IF(J-K) 35,35,25
25 KI=K-N
DO 30 I=1,N
KI=KI+N
HOLD=-A(KI)
JI=KI-K+J
A(KI)=A(JI)
30 A(JI)=HOLD
35 I=M(K)
IF(I-K) 45,45,38
38 JP=N*(I-1)
DO 40 J=1,N
JK=NK+J
JI=JP+J
HOLD=-A(JK)
A(JK)=A(JI)
40 A(JI)=HOLD
```

```

45 IF (BIGA) 48,46,48
46 D=0.0
   RETURN
48 DO 55 I=1,N
   IF(I-K) 50,55,50
50 IK=NK+I
   A(IK)=A(IK)/(-BIGA)
55 CONTINUE
   DO 65 I=1,N
   IK=NK+I
   IJ=I-N
   DO 65 J=1,N
   IJ=IJ+N
   IF(I-K) 60,65,60
60 IF(J-K) 62,65,62
62 KJ=IJ-I+K
   A(IJ)=A(IK)*A(KJ)+A(IJ)
65 CONTINUE
   KJ=K-N
   DO 75 J=1,N
   KJ=KJ+N
   IF (J-K) 70,75,70
70 A(KJ)=A(KJ)/BIGA
75 CONTINUE
   D=D*BIGA
   A(KK)=1.0/BIGA
80 CONTINUE
   K=N
100 K=(K-1)
   IF(K) 150,150,105
105 I=L(K)
   IF(I-K) 120,120,108
108 JQ=N*(K-1)
   JF=N*(I-1)
   DO 110 J=1,N
   JK=JQ+J
   HCLD=A(JK)
   JI=JR+J
   A(JK)=-A(JI)
110 A(JI)=HOLD
120 J=M(K)
   IF(J-K) 100,100,125
125 KI=K-N
   DO 130 I=1,N
   KI=KI+N
   HCLD=A(KI)
   JI=KI-K+J
   A(KI)=-A(JI)
130 A(JI)=HOLD
   GO TO 100
150 RETURN
   END
SUBROUTINE MULTR (N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)
DIMENSION XBAR(1),STD(1),D(1),RX(1),RY(1),ISAVE(1),B(1),SB(1),
1 T(1),ANS(1)
   MM=K+1
   DO 100 J=1,K
100 B(J)=0.0
   DO 110 J=1,K
   LI=K*(J-1)

```

```

      DC 110 I=1,K
      L=L1+I
110  B(J)=B(J)+RY(I)*RX(L)
      RM=0.0
      BC=0.0
      L1=ISAVE(MM)
      DC 120 I=1,K
      RM=RM+B(I)*RY(I)
      L=ISAVE(I)
      E(I)=E(I)*(STD(L1)/STD(L))
120  BU=BO+B(I)*XBAR(L)
      BU=XBAR(L1)-BO
      SSAR=RM*D(L1)
122  RM=SQRT(ABS(RM))
      SSDR=D(L1)-SSAR
      FN=N-K-1
      SY=SSDR/FN
      DO 130 J=1,K
      L1=K*(J-1)+J
      L=ISAVE(J)
125  SB(J)=SQRT(ABS((RX(L1)/D(L))*SY))
130  T(J)=B(J)/SB(J)
135  SY=SQRT(ABS(SY))
      FK=K
      SSARM=SSAR/FK
      SSDRM=SSDR/FN
      F=SSARM/SSDRM
      ANS(1)=BO
      ANS(2)=RM
      ANS(3)=SY
      ANS(4)=SSAR
      ANS(5)=FK
      ANS(6)=SSARM
      ANS(7)=SSDR
      ANS(8)=FN
      ANS(9)=SSDRM
      ANS(10)=F
      RETURN
      END

```

```

//LKED.SYSLMOD DD DSN=+MULCO(MAIN),DISP=(NEW,PASS),UNIT=SYSDA,
//          SPACE=(1600,(160,50,1))
//GC.FT03F001 DD UNIT=SYSDA,DSN=+TEMP1,DISP=(NEW,PASS),
//          SPACE=(TRK,(30,1)),DCB=(RECFM=VBS,LRECL=524,BLKSIZE=3148)
//GC.FT04F001 DD UNIT=SYSDA,DSN=+TEMP1,DISP=(NEW,PASS),
//          SPACE=(TRK,(30,1)),DCB=(RECFM=VBS,LRECL=524,BLKSIZE=3148)
//GC.FT06F001 DD SYSCLT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=1330)
//GC.FT05F001 DD *
//          ),* GUY LEMAY *,MSGLEVEL=(2,0),
//STP2 EXEC FORTGCLG,PARM.FORT='BCD',PARM.LKED='OVLY,LIST,XREF',
//          REGION.GO=200K,TIME.GO=3
//FCRT.SYSIN DD *
COMMON/LE/BANK(150,100)
DIMENSION D(100),FMT(20)
READ(5,5) N,M
5 FORMAT(2I5)
READ(5,2) IR1,IR2,IR3
2 FORMAT(3I5)
READ(5,6) (FMT(JJ),JJ=1,20)
6 FORMAT(20A4)

```

```

DU 1 I=1,N
READ(5,FMT) (C(J),J=1,M)
WRITE(6,FMT) (C(J),J=1,M)
DU 4 J=1,M
4 BANK(I,J)=C(J)
1 CONTINUE
IF(IR1.EQ.1) CALL R1
IF(IR2.EQ.1) CALL R2
IF(IR3.EQ.1) CALL R3
RETURN
END
PROGRAM REGRE

```

.....	REGR	20
	REGR	30
SAMPLE MAIN PROGRAM FOR MULTIPLE REGRESSION - REGRE	REGR	40
	REGR	50
PURPOSE	REGR	60
(1) READ THE PROBLEM PARAMETER CARD FOR A MULTIPLE REGRES-	REGR	70
SIGN, (2) READ SUBSET SELECTION CARDS, (3) CALL THE SUB-	REGR	80
ROUTINES TO CALCULATE MEANS, STANDARD DEVIATIONS, SIMPLE	REGR	90
AND MULTIPLE CORRELATION COEFFICIENTS, REGRESSION COEFFI-	REGR	100
CIENTS, T-VALUES, AND ANALYSIS OF VARIANCE FOR MULTIPLE	REGR	110
REGRESSION, AND (4) PRINT THE RESULTS.	REGR	120
	REGR	130
REMARKS	REGR	140
THE NUMBER OF OBSERVATIONS, N, MUST BE GREATER THAN M+1,	REGR	150
WHERE M IS THE NUMBER OF VARIABLES. IF SUBSET SELECTION	REGR	160
CARDS ARE NOT PRESENT, THE PROGRAM CAN NOT PERFORM MULTIPLE	REGR	170
REGRESSION.	REGR	180
AFTER RETURNING FROM SUBROUTINE MINV, THE VALUE OF DETER-	REGR	190
MINANT (DET) IS TESTED TO CHECK WHETHER THE CORRELATION	REGR	200
MATRIX IS SINGULAR. IF DET IS COMPARED AGAINST A SMALL	REGR	210
CONSTANT, THIS TEST MAY ALSO BE USED TO CHECK NEAR-	REGR	220
SINGULARITY.	REGR	230
	REGR	240
SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	REGR	250
CORRE (WHICH, IN TURN, CALLS THE SUBROUTINE NAMED DATA)	REGR	260
ORDER	REGR	270
MINV	REGR	280
MULTR	REGR	290
	REGR	300
METHOD	REGR	310
REFER TO B. OSTLE, 'STATISTICS IN RESEARCH', THE IOWA STATE	REGR	320
COLLEGE PRESS', 1954, CHAPTER 8.	REGR	330
	REGR	340
.....	REGR	350
	REGR	360
THE FOLLOWING DIMENSIONS MUST BE GREATER THAN OR EQUAL TO THE	REGR	370
NUMBER OF VARIABLES, M..	REGR	380
	REGR	390
SUBROUTINE R1		
INTEGER E,F,G		
DATA E,F,G/1,6,5/		
DIMENSION XBAR(80),STD(80), D(80), RY(80), ISAVE(80), B(80),		
1 SB(80),T(80), W(80)		
	REGR	420
DIMENSION COE(11),P(30,50)		
THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	REGR	430
PRODUCT OF M*M..	REGR	440
	REGR	450

C	DIMENSION RX(6400)	REGR 470
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO	REGR 480
C	(M+1)*M/2..	REGR 490
C		REGR 500
C	LIMENSION R(3240)	REGR 520
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO 10..	REGR 530
C		REGR 540
C	DIMENSION ANS(10)	REGR 550
C		REGR 560
C	REGR 570
C		REGR 580
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	REGR 590
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	REGR 600
C	STATEMENT WHICH FOLLOWS.	REGR 610
C		REGR 620
C	DOUBLE PRECISION XBAR,STD,RX,R,D,B,T,RY,DET,SB,ANS,SUM	REGR 630
C		REGR 640
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	REGR 650
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	REGR 660
C	ROUTINE.	REGR 670
C		REGR 680
C	REGR 690
C		REGR 700
C	1 FORMAT(A4,A2,I5,2I2)	REGR 710
C	2 FORMAT(25H1MULTIPLE REGRESSION.....A4,A2//6X,14HSELECTION.....I2//	REGR 720
C	1)	REGR 730
C	3 FORMAT(9HVARIABLE,5X,4HMEAN,6X,8HSTANDARD,6X,11HCORRELATION,4X,	REGR 740
C	11HREGRESSION,4X,10HSTD. ERROR,5X,8HCOMPUTED/6H NO.,18X,9HDEVIATREGR	REGR 750
C	2ICN,7X,6HX VS Y,7X,11HCOEFFICIENT,3X,12HOF REG.COEF.,3X,7HT VALUE)REGR	REGR 760
C	4 FORMAT(1H ,I4,6F14.5)	REGR 770
C	5 FORMAT(10H DEPENDENT)	REGR 780
C	6 FORMAT(10H/10H INTERCEPT,10X,F16.5//23H MULTIPLE CORRELATION ,F13REGR	REGR 790
C	1.5//23H STD. ERROR OF ESTIMATE,F13.5//)	REGR 800
C	7 FORMAT(10H,21X,39HANALYSIS OF VARIANCE FOR THE REGRESSION//5X,19HSREGR	REGR 810
C	10LRCE OF VARIATION,7X,7HDEGREES,7X,6HSUM OF,10X,4HMEAN,12X,7HF VALREGR	REGR 820
C	2LE/30X,10HOF FREEDOM,4X,7HSQUARES,9X,7HSQUARES)	REGR 830
C	8 FORMAT(30H ATTRIBUTABLE TO REGRESSION ,I6,3F16.5/30H DEVIATION FREGR	REGR 840
C	1RCM REGRESSION ,I6,2F16.5)	REGR 850
C	9 FORMAT(1H ,5X,5HTOTAL,19X,I6,F16.5)	REGR 860
C	10 FORMAT(36I2)	REGR 870
C	11 FORMAT(1H ,15X,18HTABLE OF RESIDUALS//9H CASE NO.,5X,7HY VALUE,5X,REGR	REGR 880
C	110HY ESTIMATE,6X,8HRESIDUAL)	REGR 890
C	12 FORMAT(1H ,I6,F15.5,2F14.5)	REGR 900
C	13 FORMAT(53H1NUMBER OF SELECTIONS NOT SPECIFIED. JOB TERMINATED.)	REGR 910
C	14 FORMAT(52H0THE MATRIX IS SINGULAR. THIS SELECTION IS SKIPPED.)	REGR 920
C		REGR 930
C	REGR 940
C		REGR 950
C	READ PROBLEM PARAMETER CARD	REGR 960
C		REGR 970
C	100 READ(5,1) PR,PR1,N,M,NS	
C	PR.....PROBLEM NUMBER (MAY BE ALPHAMERIC)	REGR 990
C	PR1.....PROBLEM NUMBER (CONTINUED)	REGR1000
C	N.....NUMBER OF OBSERVATIONS	REGR1010
C	M.....NUMBER OF VARIABLES	REGR1020
C	NS.....NUMBER OF SELECTIONS	REGR1030
C		REGR1040
C	LOGICAL TAPE 13 IS USED AS INTERMEDIATE STORAGE TO HOLD INPUT	REGR1050

C	LATA. THE INPUT DATA ARE WRITTEN ON LOGICAL TAPE 13 BY THE	REGR1060
C	SPECIAL INPUT SUBROUTINE NAMED DATA. THE STORED DATA MAY BE USED	REGR1070
C	FOR RESIDUAL ANALYSIS.	REGR1080
C		REGR1090
C	REWIND 13	REGR1100
C		REGR1110
C	IC=0	REGR1120
C	X=0.0	REGR1130
C		REGR1140
C	CALL CCORR (N,M,IC,X,XBAR,STD,RX,R,D,B,T)	REGR1150
C		REGR1160
C	REWIND 13	REGR1170
C		REGR1180
C	TEST NUMBER OF SELECTIONS	REGR1190
C		REGR1200
C	IF(NS) 108, 108, 109	REGR1210
C	108 WRITE (6,13)	REGR1220
C	GC TO 300	REGR1230
C		REGR1240
C	109 DO 200 I=1,NS	REGR1250
C	WRITE (6,2) PR,PR1,I	REGR1260
C	READ SUBSET SELECTION CARD	REGR1280
C		REGR1290
C	READ(5,10) NRESI,NDEP,K,(ISAVE(J),J=1,K)	
C	NRESI.....OPTION CODE FOR TABLE OF RESIDUALS	REGR1310
C	0 IF IT IS NOT DESIRED.	REGR1320
C	1 IF IT IS DESIRED.	REGR1330
C	NDEP.....DEPENDENT VARIABLE	REGR1340
C	K.....NUMBER OF INDEPENDENT VARIABLES INCLUDED	REGR1350
C	ISAVE.....A VECTOR CONTAINING THE INDEPENDENT VARIABLES	REGR1360
C	INCLUDED	REGR1370
C		REGR1380
C	CALL ORDER (M,R,NDEP,K,ISAVE,RX,RY)	REGR1390
C		REGR1400
C	CALL MINV (RX,K,DET,B,T)	REGR1410
C		REGR1420
C	TEST SINGULARITY OF THE MATRIX INVERTED	REGR1430
C		REGR1440
C	IF(DET) 112, 110, 112	REGR1450
C	110 WRITE (6,14)	REGR1460
C	GC TO 200	REGR1470
C		REGR1480
C	112 CALL MULTR (N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)	REGR1490
C		REGR1500
C	PRINT MEANS, STANDARD DEVIATIONS, INTERCORRELATIONS BETWEEN	REGR1510
C	X AND Y, REGRESSION COEFFICIENTS, STANDARD DEVIATIONS OF	REGR1520
C	REGRESSION COEFFICIENTS, AND COMPUTED T-VALUES	REGR1530
C		REGR1540
C	MM=K+1	REGR1550
C	WRITE (6,3)	REGR1560
C	DO 115 J=1,K	REGR1570
C	L=ISAVE(J)	REGR1580
C	115 WRITE (6,4) L,XBAR(L),STD(L),RY(J),B(J),SB(J),T(J)	REGR1590
C	WRITE (6,5)	REGR1600
C	L=ISAVE(MM)	REGR1610
C	WRITE (6,4) L,XBAR(L),STD(L)	REGR1620
C		REGR1630
C	PRINT INTERCEPT, MULTIPLE CORRELATION COEFFICIENT, AND STANDARD	REGR1640
C	ERROR OF ESTIMATE	REGR1650
C		REGR1660

```

C      WRITE (6,6) ANS(1),ANS(2),ANS(3)
C      PRINT ANALYSIS OF VARIANCE FOR THE REGRESSION
C
C      WRITE (6,7)
C      L=ANS(8)
C      WRITE (6,8) K,ANS(4),ANS(6),ANS(10),L,ANS(7),ANS(9)
C      L=N-1
C      SUM=ANS(4)+ANS(7)
C      WRITE (6,9) L,SUM
C      IF(NRES1) 200, 200, 120
C
C      PRINT TABLE OF RESIDUALS
C
C      120 WRITE (6,2) PR,PR1,I
C      WRITE (6,11)
C      MM=ISAVE(K+1)
C      DO 140 II=1,N
C      READ (13) (W(J),J=1,M)
C      SUM = ANS(1)
C      DO 130 J=1,K
C      L=ISAVE(J)
C      130 SUM=SUM+W(L)*B(J)
C      RESI=W(MM)-SUM
C      140 WRITE (6,12) II,W(MM),SUM,RESI
C      REWIND 13
C      200 CONTINUE
C      GC TO 100
C      300 CONTINUE
C      RETURN
C      END

```

```

REGR1670
REGR1680
REGR1690
REGR1700
REGR1710
REGR1720
REGR1730
REGR1740
REGR1750
REGR1760
REGR1770
REGR1780
REGR1790
REGR1800
REGR1810
REGR1820
REGR1830
REGR1840
REGR1850
REGR1860
REGR1870
REGR1880
REGR1890
REGR1900
REGR1910
REGR1920
REGR1930
REGR1940
REGR1950

```

```

C
C      .....STEP 10
C      .....STEP 20
C      .....STEP 30
C      SAMPLE MAIN PROGRAM FOR STEP-WISE MULTIPLE REGRESSION - STEPR STEP 40
C      .....STEP 50
C      PURPOSE STEP 60
C      (1) READ THE PROBLEM PARAMETER CARD FOR A STEP-WISE MULTIPLE STEP 70
C      REGRESSION, (2) READ SUBSET SELECTION CARDS, (3) CALL THE STEP 80
C      SUBROUTINE TO CALCULATE MEANS, STANDARD DEVIATIONS, SIMPLE STEP 90
C      CORRELATION COEFFICIENTS, AND (4) CALL THE SUBROUTINE TO STEP 100
C      PERFORM EACH STEP OF REGRESSION ANALYSIS. STEP 110
C      .....STEP 120
C      REMARKS STEP 130
C      THE NUMBER OF OBSERVATIONS, N, MUST BE GREATER THAN M+2, STEP 140
C      WHERE M IS THE NUMBER OF VARIABLES. IF SELECTION CARDS ARE STEP 150
C      NOT PRESENT, THIS PROGRAM CAN NOT PERFORM STEP-WISE MULTIPLE STEP 160
C      REGRESSION. STEP 170
C      .....STEP 180
C      SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED STEP 190
C      CORRE (WHICH, IN TURN, CALLS THE SUBROUTINE DATA) STEP 200
C      MSTR (WHICH, IN TURN, CALLS THE SUBROUTINE LOC) STEP 210
C      STPRG (WHICH, IN TURN, CALLS THE SUBROUTINE STOUT) STEP 220
C      .....STEP 230
C      METHOD STEP 240
C      REFER TO C. A. BENNETT AND N. L. FRANKLIN, "STATISTICAL STEP 250
C      ANALYSIS IN CHEMISTRY AND THE CHEMICAL INDUSTRY", JOHN WILEY STEP 260
C      AND SONS, 1954, APPENCIX 6A. STEP 270
C      .....STEP 280
C      .....STEP 290

```

C		STEP 300
C	THE FOLLOWING DIMENSICNS MUST BE GREATER THAN OR EQUAL TO THE	STEP 310
C	NUMBER OF VARIABLES, M..	STEP 320
C		STEP 330
	SUBROUTINE R2	
	DIMENSION XBAR(35),STD(35),D(35),B(35),T(35),IDX(35),L(35)	STEP 340
C		STEP 350
C	THE FLLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	STEP 360
C	PRODUCT OF M*M..	STEP 370
C		STEP 380
	DIMENSION RX(1225)	STEP 390
C		STEP 400
C	THE FCLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	STEP 410
C	(M+1)*M/2..	STEP 420
C		STEP 430
	DIMENSION R(630)	STEP 440
C		STEP 450
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO 5..	STEP 460
C		STEP 470
	DIMENSION NSTEP(5)	STEP 480
C		STEP 490
C	THE FCLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO 11..	STEP 500
C		STEP 510
	DIMENSION ANS(11)	STEP 520
C		STEP 530
C	STEP 540
C		STEP 550
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	STEP 560
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	STEP 570
C	STATEMENT WHICH FOLLOWS.	STEP 580
C		STEP 590
C	DOUBLE PRECISION XBAR,STD,RX,R,B,T,ANS,YES	STEP 600
C		STEP 610
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	STEP 620
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	STEP 630
C	ROUTINE.	STEP 640
C		STEP 650
C	STEP 660
C		STEP 670
	1 FCRMAT(A4,A2,I5,2I2,F6.0,2I2)	
	2 FCRMAT(53HNUMBER OF SELECTIONS NOT SPECIFIED. JOB TERMINATED.)	STEP 690
	3 FCRMAT(35H1STEP-WISE MULTIPLE REGRESSION.....A4, A2)	STEP 700
	4 FORMAT(31H0VARIABLE MEAN STANDARD/4X,3HNO.16X,9HDEVIATION)	STEP 710
	5 FCRMAT(4X, I2, F14.5, F12.5)	STEP 720
	6 FORMAT(19H1CORRELATION MATRIX)	STEP 730
	7 FORMAT(4HOROWI3/(10F12.5))	STEP 740
	8 FCRMAT(72I1)	STEP 750
	9 FCRMAT(23H0NUMBER OF CBSERVATIONS15)	STEP 760
	10 FORMAT(20H NUMEER OF VARIABLES3X, I5)	STEP 770
	11 FCRMAT(21H NUMBER OF SELECTIONS2X, I5)	STEP 780
	12 FORMAT(28H0CONSTANT TO LIMIT VARIABLESF9.5)	STEP 790
	13 FORMAT(/15H1SELECTION.....I2)	STEP 800
	14 FORMAT(16X,18HTABLE OF RESIDUALS//9H CASE NO.5X,7HY VALUE5X,10HY E	STEP 810
	1STIMATE6X,8HRESIDUAL)	STEP 820
	15 FCRMAT(I7, F15.5, 2F14.5)	STEP 830
	16 FCRMAT(1H)	STEP 840
	17 FCRMAT(1H1)	STEP 850
	18 FCRMAT(1H0, '***CCLUMN', I4, ' OF SELECTION CARD', I5, ' IS IN ERROR.	STEP 860
	1 IT IS POSSIBLE THAT COLUMNS SUCCEEDING THAT COLUMN ARE ALSO'	STEP 870
	2/' INCORRECT. THE SELECTION IS IGNORED.***')	STEP 880

```

19 FORMAT(1H0,'****SELECTION CARD',I5,' DOES NOT NAME ONE AND ONLY ONSTEP 890
   1E DEPENDENT VARIABLE. SELECTION IGNORED.****') STEP 900
20 FCRMAT(1H0,'****EITHER THE MATRIX IS SINGULAR, OR THE RESIDUAL SUMSTEP 910
   1 OF SQUARES IS NEGATIVE IMPLYING EXTREME ILL CONDITION.',/, ' SELECSTEP 920
   2TION IGNORED.****') STEP 930
21 FCRMAT(1H0,'****',I6,' OBSERVATIONS ARE TOO FEW TO ALLOW PARAMETERSTEP 940
   1 ESTIMATION FOR',I5,' VARIABLES. JOB TERMINATED.****') STEP 950
C STEP 960
C READ PROBLEM PARAMETER CARD STEP 970
C STEP 980
NPR=0
100 READ (5,1) PR1,PR2,N,M,NS,PCT,NR,NSET STEP 990
NPR=NPR+1
C PR1.....PRCELEM CODE (MAY BE ALPHAMERIC) STEP1000
C PR2.....PROBLEM CODE (CONTINUED) STEP1010
C N .....NUMBER OF OBSERVATIONS STEP1020
C M .....NUMEER OF VARIABLES STEP1030
C NS.....NUMBER OF SELECTIONS STEP1040
C PCT.....A CONSTANT VALLE OF PROPORTION OF SUM OF SQUARES THAT STEP1050
C WILL BE USED TO LIMIT VARIABLES ENTERING IN THE REGRES- STEP1060
C SION STEP1070
C NR.....OPTION CODE FOR TABLE OF RESIDUALS STEP1080
C 0 - IF IT IS NCT DESIRED STEP1090
C 1 - IF IT IS DESIRED STEP1100
C STEP1110
WRITE (6,3) PR1,PR2 STEP1120
WRITE (6,9) N STEP1130
WRITE (6,10) M STEP1140
IF(N-M-2) 101,101,102 STEP1150
101 WRITE(6,21) N,M STEP1160
STOP STEP1170
102 WRITE (6,11) NS STEP1180
WRITE (6,12) PCT STEP1190
C STEP1200
C LOGICAL TAPE 13 IS USED AS INTERMEDIATE STORAGE TO HOLD INPUT STEP1210
C DATA. THE INPUT DATA ARE WRITTEN ON LOGICAL TAPE 13 BY THE STEP1220
C SPECIAL INPUT SUBROUTINE NAMED DATA. THE STORED DATA MAY BE USED STEP1230
C FOR RESIDUAL ANALYSIS. STEP1240
C STEP1250
REWIND 13 STEP1260
C STEP1270
IC=0 STEP1280
X=0.0 STEP1290
C STEP1300
CALL CCRRE (N,M,IC,X,XBAR,STD,RX,R,B,D,T) STEP1310
C STEP1320
REWIND 13 STEP1330
C STEP1340
C PRINT MEANS AND STANDARD DEVIATION STEP1350
C STEP1360
WRITE (6,4) STEP1370
DO 105 I=1,M STEP1380
105 WRITE (6,5) I,XEAR(I),STD(I) STEP1390
C STEP1400
C PRINT CORRELATION MATRIX STEP1410
C STEP1420
WRITE (6,6) STEP1430
DO 130 I=1,M STEP1440
DO 125 J=1,M STEP1450
IF(I-J) 110, 120, 120 STEP1460

```

110	K=I+(J*J-J)/2	STEP1470
	GO TO 125	STEP1480
120	K=J+(I*I-I)/2	STEP1490
125	T(J)=R(K)	STEP1500
130	WRITE (6,7) I,(T(J),J=1,M)	STEP1510
C		STEP1520
C	TEST NUMBER OF SELECTIONS	STEP1530
C		STEP1540
	IF(NS) 135, 135, 140	STEP1550
135	WRITE (6,2)	STEP1560
	GO TO 200	STEP1570
C		STEP1580
C	SAVE THE MATRIX OF SUMS OF CROSS-PRODUCTS OF DEVIATIONS	STEP1590
C		STEP1600
140	CALL MSTK (RX,R,M,U,1)	STEP1610
C		STEP1620
	NSEL=1	STEP1630
	GO TO 150	STEP1640
C		STEP1650
C	COPY THE MATRIX OF SUMS OF CROSS-PRODUCTS OF DEVIATIONS	STEP1660
C		STEP1670
145	CALL MSTR (R,RX,M,1,0)	STEP1680
C		STEP1690
C	READ A SELECTION CARD	STEP1700
C		STEP1710
150	WRITE (6,13) NSEL	STEP1720
	READ (5,8) (IDX(J),J=1,M)	STEP1730
C		STEP1740
C	IN EACH POSITION OF IDX, ONE OF THE FOLLOWING CODES MUST BE	STEP1750
C	SPECIFIED..	STEP1760
C	0 OR BLANK - INDEPENDENT VARIABLE AVAILABLE FOR SELECTION	STEP1770
C	1 - INDEPENDENT VARIABLE TO BE FORCED IN REGRESSION	STEP1780
C	2 - VARIABLE TO BE DELETED	STEP1790
C	3 - DEPENDENT VARIABLE	STEP1800
C		STEP1810
	N35=0	STEP1820
	DO 155 K=1,M	STEP1830
	IF (IDX(K)) 152,153,153	STEP1840
152	WRITE (6,18) K,NSEL	STEP1850
	GO TO 185	STEP1860
153	IF (IDX(K)-3) 155,154,152	STEP1870
154	N35=N35+1	STEP1880
155	CONTINUE	STEP1890
	IF (N35-1) 156,157,156	STEP1900
156	WRITE (6,19) NSEL	STEP1910
	GO TO 185	STEP1920
C	CALL THE SUBROUTINE TO PERFORM A STEP-WISE REGRESSION ANALYSIS	STEP1930
C		STEP1940
157	CALL STPRG (M,N,RX,XBAR,IDX,PCT,NSTEP,ANS,L,B,STD,T,D,IER)	STEP1950
	IF (IER) 158,159,158	STEP1960
158	WRITE (6,20)	STEP1970
	GO TO 185	STEP1980
C		STEP1990
C	FIND WHETHER TO PRINT THE TABLE OF RESIDUALS	STEP2000
C		STEP2010
159	IF(NR) 185, 185, 160	STEP2020
C		STEP2030
C	PRINT THE TABLE OF RESIDUALS	STEP2040
C		STEP2050
C		STEP2060

160	WRITE (6,13) NSEL	STEP2070
	WRITE (6,16)	STEP2080
	WRITE (6,14)	STEP2090
	MM=NSTEP(1)	STEP2100
	DO 180 I=1,N	STEP2110
	READ (13) (D(J),J=1,M)	STEP2120
	YEST=ANS(9)	STEP2130
	K=NSTEP(4)	STEP2140
	DO 170 J=1,K	STEP2150
	KK=L(J)	STEP2160
170	YEST=YEST+B(J)*D(KK)	STEP2170
	RESI=C(MM)-YEST	STEP2180
180	WRITE (6,15) I,D(MM),YEST,RESI	STEP2190
	REWIND 13	STEP2200
C		STEP2210
C	TEST TO SEE WHETHER ALL SELECTIONS ARE COMPLETED	STEP2220
C		STEP2230
185	IF(NSEL-NS) 190, 300, 300	STEP2250
190	NSEL=NSEL+1	STEP2260
	WRITE (6,17)	STEP2270
C	GO TO 145	STEP2280
C		
300	IF(NPR-NSET) 100, 301, 301	
200	CONTINUE	STEP2290
301	RETURN	
	END	STEP2300
C		MCAN 10
C	MCAN 20
C		MCAN 30
C	SAMPLE MAIN PROGRAM FOR CANCNICAL CORRELATION - MCAND	MCAN 40
C		MCAN 50
C	PUPPOSE	MCAN 60
C	(1) READ THE PROBLEM PARAMETER CARD FOR A CANONICAL	MCAN 70
C	CORRELATION, (2) CALL TWO SUBROUTINES TO CALCULATE SIMPLE	MCAN 80
C	CORRELATIONS, CANONICAL CORRELATIONS, CHI-SQUARES, DEGREES	MCAN 90
C	OF FREEDOM FOR CHI-SQUARES, AND COEFFICIENTS FOR LEFT AND	MCAN 100
C	RIGHT HAND VARIABLES, NAMELY CANONICAL VARIATES, AND (3)	MCAN 110
C	PRINT THE RESULTS.	MCAN 120
C		MCAN 130
C	REMARKS	MCAN 140
C	THE NUMBER OF LEFT HAND VARIABLES MUST BE GREATER THAN	MCAN 150
C	OR EQUAL TO THE NUMBER OF RIGHT HAND VARIABLES.	MCAN 160
C		MCAN 170
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	MCAN 180
C	CORRE (WHICH, IN TURN, CALLS THE INPUT SUBROUTINE NAMED	MCAN 190
C	CATA.)	MCAN 200
C	CANOR (WHICH, IN TURN, CALLS THE SUBROUTINES MINV AND	MCAN 210
C	NROOT. NROOT, IN TURN, CALLS THE SUBROUTINE EIGEN.)	MCAN 220
C		MCAN 230
C	METHOD	MCAN 240
C	REFER TO W. W. COOLEY AND P. R. LOHNES, 'MULTIVARIATE PRO-	MCAN 250
C	CEDURES FOR THE BEHAVIORAL SCIENCES', JOHN WILEY AND SONS,	MCAN 260
C	1962, CHAPTER 3.	MCAN 270
C		MCAN 280
C	MCAN 290
C		MCAN 300
C	THE FOLLOWING DIMENSIONS MUST BE GREATER THAN OR EQUAL TO THE	MCAN 310
C	TOTAL NUMBER OF VARIABLES M (M=MP+MQ, WHERE MP IS THE NUMBER OF	MCAN 320
C	LEFT HAND VARIABLES, AND MQ IS THE NUMBER OF RIGHT HAND VARI-	MCAN 330
C	ABLES)..	MCAN 340

C		MCAN 350
	SUBROUTINE R3	
	DIMENSION XBAR(20),STD(20),CANR(20),CHISQ(20),VDF(20)	MCAN 360
C		MCAN 370
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	MCAN 380
C	PRODUCT OF M*M..	MCAN 390
C		MCAN 400
	DIMENSION RX(400)	MCAN 410
C		MCAN 420
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO	MCAN 430
C	(M+1)*M/2..	MCAN 440
C		MCAN 450
	DIMENSION R(210)	MCAN 460
C		MCAN 470
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	MCAN 480
C	PRODUCT OF MP*MQ..	MCAN 490
C		MCAN 500
	DIMENSION CCEFL(400)	MCAN 510
C		MCAN 520
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	MCAN 530
C	PRODUCT OF MQ*MQ..	MCAN 540
C		MCAN 550
	DIMENSION CCEFR(400)	MCAN 560
C		MCAN 570
C	MCAN 580
C		MCAN 590
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	MCAN 600
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	MCAN 610
C	STATEMENT WHICH FOLLOWS.	MCAN 620
C		MCAN 630
C	DOUBLE PRECISION XBAR,STD,RX,R,CANR,CHISQ,CCEFL,CCEFR	MCAN 640
C		MCAN 650
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	MCAN 660
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	MCAN 670
C	ROUTINE.	MCAN 680
C		MCAN 690
C	MCAN 700
C		MCAN 710
	1 FORMAT(A4,A2,I5,2I2)	MCAN 720
	2 FORMAT(27H1CANONICAL CORRELATION.....,A4,A2//22H NO. OF OBSERVAT	MCAN 730
	IONS,8X,I4/29H NO. OF LEFT HAND VARIABLES,I5/30H NO. OF RIGHT	MCAN 740
	HAND VARIABLES,14/)	MCAN 750
	3 FORMAT(6HMEANS/(8F15.5))	MCAN 760
	4 FORMAT(26HSTANDARD DEVIATIONS/(8F15.5))	MCAN 770
	5 FORMAT(25HOCORRELATION COEFFICIENTS)	MCAN 780
	6 FORMAT(4HROW,I3/(10F12.5))	MCAN 790
	7 FORMAT(1HU//12H NUMBER OF, 7X,7HLARGEST,7X,13HCORRESPONDING,3IX,	MCAN 800
	17HDEGREES/13H EIGENVALUES,5X,10HEIGENVALUE,7X,9HCANONICAL,7X,	MCAN 810
	26HLAMBDA,5X,10HCHI-SQUARE,7X,2HU/4X,7HREMOVED,7X,9HREMAINING,7X,	MCAN 820
	31H-CORRELATION,32X,7HFREEDOM/)	MCAN 830
	8 FORMAT(1H ,17,F19.5,F16.5,2F14.5,5X,I5)	MCAN 840
	9 FORMAT(1HU/22H CANONICAL CORRELATION,F12.5)	MCAN 850
	10 FORMAT(39HU COEFFICIENTS FOR LEFT HAND VARIABLES/(8F15.5))	MCAN 860
	11 FORMAT(40HU COEFFICIENTS FOR RIGHT HAND VARIABLES/(8F15.5))	MCAN 870
C		MCAN 880
C	MCAN 890
C		MCAN 900
C	READ PROBLEM PARAMETER CARD	MCAN 910
C		MCAN 920
	100 READ (5,1) PR,FRI,N,MP,MQ	MCAN 930

C	PK.....PROBLEM NUMBER (MAY BE ALPHAMERIC)	MCAN 940
C	PR1.....PROBLEM NUMBER (CONTINUED)	MCAN 950
C	N.....NUMBER OF OBSERVATIONS	MCAN 960
C	MP.....NUMBER OF LEFT HAND VARIABLES	MCAN 970
C	MQ.....NUMBER OF RIGHT HAND VARIABLES	MCAN 980
C		MCAN 990
C	WRITE (6,2) PR,PR1,N,MP,MQ	MCAN1000
C		MCAN1010
C	M=MP+MQ	MCAN1020
C	IC=0	MCAN1030
C	X=0.0	MCAN1040
C		MCAN1050
C	CALL CORRE (N,M,IC,X,XBAR,STD,RX,R,CANR,CHISQ,COEFL)	MCAN1060
C		MCAN1070
C	PRINT MEANS, STANDARD DEVIATIONS, AND CORRELATION	MCAN1080
C	COEFFICIENTS OF ALL VARIABLES	MCAN1090
C		MCAN1100
C	WRITE (6,3) (XBAR(I),I=1,M)	MCAN1110
C	WRITE (6,4) (STD(I),I=1,M)	MCAN1120
C	WRITE (6,5)	MCAN1130
C	DO 160 I=1,M	MCAN1140
C	DO 150 J=1,M	MCAN1150
C	IF(I-J) 120, 130, 130	MCAN1160
C	120 L=I+(J*J-J)/2	MCAN1170
C	GO TO 140	MCAN1180
C	130 L=J+(I*I-I)/2	MCAN1190
C	140 CANR(J)=R(L)	MCAN1200
C	150 CONTINUE	MCAN1210
C	160 WRITE (6,6) I,(CANR(J),J=1,M)	MCAN1220
C		MCAN1230
C	CALL CANOR (N,MP,MQ,R,XBAR,STD,CANR,CHISQ,NDF,COEFR,CCEFL,RX)	MCAN1240
C		MCAN1250
C	PRINT EIGENVALUES, CANONICAL CORRELATIONS, LAMBDA, CHI-SQUARES,	MCAN1260
C	DEGREES OF FREEDOMS	MCAN1270
C		MCAN1280
C	WRITE (6,7)	MCAN1290
C	DO 170 I=1,MQ	MCAN1300
C	N1=I-1	MCAN1310
C		MCAN1320
C	TEST WHETHER EIGENVALUE IS GREATER THAN ZERO	MCAN1330
C		MCAN1340
C	IF(XBAR(I)) 165, 165, 170	MCAN1350
C	165 MM=N1	MCAN1360
C	GO TO 175	MCAN1370
C	170 WRITE (6,8) N1,XBAR(I),CANR(I),STD(I),CHISQ(I),NDF(I)	MCAN1380
C	MM=MQ	MCAN1390
C		MCAN1400
C	PRINT CANONICAL COEFFICIENTS	MCAN1410
C		MCAN1420
C	175 N1=0	MCAN1430
C	N2=0	MCAN1440
C	DO 200 I=1,MM	MCAN1450
C	WRITE (6,9) CANR(I)	MCAN1460
C	DO 180 J=1,MP	MCAN1470
C	N1=N1+1	MCAN1480
C	180 XBAR(J)=COEFL(N1)	MCAN1490
C	WRITE (6,10) (XBAR(J),J=1,MP)	MCAN1500
C	DO 190 J=1,MQ	MCAN1510
C	N2=N2+1	MCAN1520
C	190 XBAR(J)=COEFR(N2)	MCAN1530

	WRITE (6,11) (XBAR(J),J=1,M)	MCAN1540
200	CONTINUE	MCAN1550
C	GL TO 100	MCAN1560
	RETURN	
	END	MCAN1570

C		PLRG 10
C	PLRG 20
C		PLRG 30
C	SAMPLE MAIN PROGRAM FOR POLYNOMIAL REGRESSION - POLRG	PLRG 40
C		PLRG 50
C	PURPOSE	PLRG 60
C	(1) READ THE PROBLEM PARAMETER CARD FOR A POLYNOMIAL REGRES-	PLRG 70
C	PRINT THE REGRESSION COEFFICIENTS AND ANALYSIS OF VARIANCE	PLRG 90
C	TABLE FOR POLYNOMIALS OF SUCCESSIVELY INCREASING DEGREES,	PLRG 100
C	AND (4) OPTIONALLY PRINT THE TABLE OF RESIDUALS AND A PLOT	PLRG 110
C	OF Y VALLES AND Y ESTIMATES.	PLRG 120
C		PLRG 130
C	REMARKS	PLRG 140
C	THE NUMBER OF OBSERVATIONS, N, MUST BE GREATER THAN M+1,	PLRG 150
C	WHERE M IS THE HIGHEST DEGREE POLYNOMIAL SPECIFIED.	PLRG 160
C	IF THERE IS NO REDUCTION IN THE RESIDUAL SUM OF SQUARES	PLRG 170
C	BETWEEN TWO SUCCESSIVE DEGREES OF THE POLYNOMIALS, THE	PLRG 180
C	PROGRAM TERMINATES THE PROBLEM BEFORE COMPLETING THE ANALY-	PLRG 190
C	SIS FOR THE HIGHEST DEGREE POLYNOMIAL SPECIFIED.	PLRG 200
C		PLRG 210
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	PLRG 220
C	GDATA	PLRG 230
C	ORDER	PLRG 240
C	MINV	PLRG 250
C	MULTR	PLRG 260
C	PLOT (A SPECIAL PLOT SUBROUTINE PROVIDED FOR THE SAMPLE	PLRG 270
C	PROGRAM.)	PLRG 280
C		PLRG 290
C	METHOD	PLRG 300
C	REFER TO B. OSTLE, 'STATISTICS IN RESEARCH', THE IOWA STATE	PLRG 310
C	COLLEGE PRESS', 1954, CHAPTER 6.	PLRG 320
C		PLRG 330
C	PLRG 340
C	SUBROUTINE R4	
C		PLRG 350
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	PLRG 360
C	PRODUCT OF N*(M+1), WHERE N IS THE NUMBER OF OBSERVATIONS AND M	PLRG 370
C	IS THE HIGHEST DEGREE POLYNOMIAL SPECIFIED..	PLRG 380
C		PLRG 390
C	DIMENSION X(1100)	PLRG 400
C		PLRG 410
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO THE	PLRG 420
C	PRODUCT OF M*M..	PLRG 430
C		PLRG 440
C	DIMENSION DI(100)	PLRG 450
C		PLRG 460
C	THE FOLLOWING DIMENSION MUST BE GREATER THAN OR EQUAL TO	PLRG 470
C	(M+2)*(M+1)/2..	PLRG 480
C		PLRG 490
C	DIMENSION C(66)	PLRG 500
C		PLRG 510
C	THE FOLLOWING DIMENSIONS MUST BE GREATER THAN OR EQUAL TO M..	PLRG 520
C		PLRG 530

C	DIMENSION B(10),E(10),SB(10),T(10)	PLRG 540
C		PLRG 550
C	THE FOLLOWING DIMENSICNS MUST BE GREATER THAN OR EQUAL TO (M+1)..	PLRG 560
C		PLRG 570
C	DIMENSION XBAR(11),STD(11),COE(11),SUMSQ(11),ISAVE(11)	PLRG 580
C		PLRG 590
C	THE FOLLOING DIMENSION MUST BE GREATER THAN OR EQUAL TO 10..	PLRG 600
C		PLRG 610
C	COMMON BANK(40,40)	
C	DIMENSION ANS(10)	PLRG 620
C		PLRG 630
C	THE FOLLOWING DIMENSION WILL BE USED IF THE PLOT OF OBSERVED DATA	PLRG 640
C	AND ESTIMATES IS DESIRED. THE SIZE OF THE DIMENSION, IN THIS	PLRG 650
C	CASE, MUST BE GREATER THAN OR EQUAL TO N*3. OTHERWISE, THE SIZE	PLRG 660
C	OF DIMENSION MAY BE SET TO 1.	PLRG 670
C		PLRG 680
C	DIMENSION P(300)	PLRG 690
C		PLRG 700
C	PLRG 710
C		PLRG 720
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	PLRG 730
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	PLRG 740
C	STATEMENT WHICH FOLLOWS.	PLRG 750
C		PLRG 760
C	DOUBLE PRECISION X,XBAR,STD,D,SUMSQ,DI,E,B,SB,T,ANS,DET,COE	PLRG 770
C		PLRG 780
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	PLRG 790
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	PLRG 800
C	ROUTINE.	PLRG 810
C		PLRG 820
C	PLRG 830
C		PLRG 840
C	1 FORMAT(A4,A2,I5,I2,I1)	PLRG 850
C	2 FORMAT(2F6.0)	PLRG 860
C	3 FORMAT(27H1POLYNOMIAL REGRESSION.....,A4,A2/)	PLRG 870
C	4 FORMAT(23HNUMBER OF CBSERVATIONS,I6//)	PLRG 880
C	5 FORMAT(32H0POLYNOMIAL REGRESSION OF DEGREE,I3)	PLRG 890
C	6 FORMAT(12H0 INTERCEPT,E20.7)	PLRG 900
C	7 FORMAT(26H0 REGRESSION COEFFICIENTS/(6E20.7))	PLRG 910
C	8 FORMAT(1H0/24X,24HANALYSIS OF VARIANCE FOR,I4,19H DEGREE POLYNOMI	PLRG 920
C	1AL/)	PLRG 930
C	9 FORMAT(1H0,5X,19HSOURCE OF VARIATION,7X,9HDEGREE OF,7X,6HSUM OF,9XPLRG	PLRG 940
C	1,4HMEAN,10X,1FF,5X,20HIMPROVEMENT IN TERMS/33X,7HFREEDOM,8X,7HSQUAPLRG	PLRG 950
C	2RES,7X,6HSQUARE,7X,5HVALUE,8X,17HOF SUM OF SQUARES)	PLRG 960
C	10 FURMAT(20H0 DUE TO REGRESSION,I2X,I6,F17.5,F14.5,F13.5,F20.5)	PLRG 970
C	11 FORMAT(32H DEVIATION ABOUT REGRESSION ,I6,F17.5,F14.5)	PLRG 980
C	12 FORMAT(8X,5HTCTAL,19X,I6,F17.5///)	PLRG 990
C	13 FORMAT(17H0 NO IMPROVEMENT)	PLRG1000
C	14 FCRMAT(1H0//27X,18HTABLE OF RESIDUALS//16H OBSERVATION NO.,5X,7HX	PLRG1010
C	1VALUE,7X,7HY VALLE,7X,10HY ESTIMATE,7X,8HRESIDUAL/)	PLRG1020
C	15 FOKMAT(11H0,3X,I6,F18.5,F14.5,F17.5,F15.5)	PLRG1030
C		PLRG1040
C	PLRG1050
C		PLRG1060
C	READ PROBLEM PARAMETER CARD	PLRG1070
C		PLRG1080
C	100 READ (5,1) PR,FRI,N,M,NPLDT	PLRG1090
C		PLRG1100
C	PR....PROBLEM NUMBER (MAY BE ALPHAMERIC)	PLRG1110
C	PR1....PROBLEM NUMBER (CONTINUED)	PLRG1120

C	N.....NUMBER OF OBSERVATIONS	PLRG1130
C	M.....HIGHEST DEGREE POLYNOMIAL SPECIFIED	PLRG1140
C	NPLOT.OPTION CLDE FOR PLOTTING	PLRG1150
C	0 IF PLOT IS NOT DESIRED.	PLRG1160
C	1 IF PLOT IS DESIRED.	PLRG1170
C		PLRG1180
C	PRINT PROBLEM NUMBER AND N.	PLRG1190
C		PLRG1200
C	WRITE (6,3) PR,PRI	PLRG1210
C	WRITE (6,4) N	PLRG1220
C		PLRG1230
C	READ INPUT DATA	PLRG1240
C		PLRG1250
C	IV1=4	
C	IV2=8	
C	L=N*M	PLRG1260
C	DC 110 I=1,N	PLRG1270
C	J=L+I	PLRG1280
C		PLRG1290
C	X(I) IS THE INDEPENDENT VARIABLE, AND X(J) IS THE DEPENDENT	PLRG1300
C	VARIABLE.	PLRG1310
C		PLRG1320
C	X(I)=BANK(I,IV1)	
C	X(J)=BANK(I,IV2)	
C	110 CCNTINUE	
C		PLRG1340
C	CALL GDATA (N,M,X,XBAR,STD,D,SUMSQ)	PLRG1350
C		PLRG1360
C	MM=M+1	PLRG1370
C	SUM=0.0	PLRG1380
C	NT=N-1	PLRG1390
C		PLRG1400
C	DC 200 I=1,M	PLRG1410
C	ISAVE(I)=I	PLRG1420
C		PLRG1430
C	FORM SUBSET OF CORRELATION COEFFICIENT MATRIX	PLRG1440
C		PLRG1450
C	CALL ORDER (MM,D,MM,I,ISAVE,DI,E)	PLRG1460
C		PLRG1470
C	INVERT THE SUBMATRIX OF CORRELATION COEFFICIENTS	PLRG1480
C		PLRG1490
C	CALL MINV (DI,I,DET,B,T)	PLRG1500
C		PLRG1510
C	CALL MULTR (N,I,XBAR,STD,SUMSQ,DI,E,ISAVE,B,SB,T,ANS)	PLRG1520
C		PLRG1530
C	PRINT THE RESULT OF CALCULATION	PLRG1540
C		PLRG1550
C	WRITE (6,5) I	PLRG1560
C	IF(ANS(7)) 140,130,130	PLRG1570
C	130 SUMIP=ANS(4)-SUM	PLRG1580
C	IF(SUMIP) 140, 140, 150	PLRG1590
C	140 WRITE (6,13)	PLRG1600
C	GC TO 210	PLRG1610
C	150 WRITE (6,6) ANS(1)	PLRG1620
C	WRITE (6,7) (B(J),J=1,I)	PLRG1630
C	WRITE (6,8) I	PLRG1640
C	WRITE (6,9)	PLRG1650
C	SUM=ANS(4)	PLRG1660
C	WRITE (6,10) I,ANS(4),ANS(6),ANS(10),SUMIP	PLRG1670
C	NI=ANS(8)	PLRG1680

WRITE (6,11) N1,ANS(7),ANS(9)	PLRG1690
WRITE (6,12) N1,SLMSQ(MM)	PLRG1700
C	PLRG1710
C SAVE COEFFICIENTS FOR CALCULATION OF Y ESTIMATES	PLRG1720
C	PLRG1730
COE(1)=ANS(1)	PLRG1740
DC 160 J=1,I	PLRG1750
160 COE(J+1)=B(J)	PLRG1760
LA=I	PLRG1770
C	PLRG 10
200 CONTINUE	PLRG1780
C	PLRG1790
C TEST WHETHER PLOT IS DESIRED	PLRG1800
C	PLRG1810
210 IF(NPLOT) 100, 100, 220	PLRG1820
C	PLRG1830
C CALCULATE ESTIMATES	PLRG1840
C	PLRG1850
220 NP3=N+N	PLRG1860
DO 230 I=1,N	PLRG1870
NP3=NP3+1	PLRG1880
P(NP3)=COE(1)	PLRG1890
L=I	PLRG1900
DC 230 J=1,LA	PLRG1910
P(NP3)=P(NP3)+X(L)*COE(J+1)	PLRG1920
230 L=L+N	PLRG1930
C	PLRG1940
C COPY OBSERVED DATA	PLRG1950
C	PLRG1960
N2=N	PLRG1970
L=N*M	PLRG1980
DC 240 I=1,N	PLRG1990
P(I)=X(I)	PLRG2000
N2=N2+1	PLRG2010
L=L+1	PLRG2020
240 P(N2)=X(L)	PLRG2030
C	PLRG2040
C PRINT TABLE OF RESIDUALS	PLRG2050
C	PLRG2060
WRITE (6,3) PR,PR1	PLRG2070
WRITE (6,5) LA	PLRG2080
WRITE (6,14)	PLRG2090
NP2=N	PLRG2100
NP3=N+N	PLRG2110
DO 250 I=1,N	PLRG2120
NP2=NP2+1	PLRG2130
NP3=NP3+1	PLRG2140
RESID=P(NP2)-P(NP3)	PLRG2150
250 WRITE (6,15) I,P(I),P(NP2),P(NP3),RESID	PLRG2160
C	PLRG2170
C CALL PLOT (LA,P,N,3,22,1)	PLRG2180
C	PLRG2190
GO TO 100	PLRG2200
END	PLRG2210
CCORR 20
C	CORR 30
C SUBROUTINE CORRE	CORR 40
C	CORR 50
C PURPOSE	CORR 60
C COMPUTE MEANS, STANDARD DEVIATIONS, SUMS OF CROSS-PRODUCTS	CORR 70

```

C                                     CORR 90
C USAGE                               CORR 100
C   CALL CORRE (N,M,IO,X,XBAR,STD,RX,R,B,D,T)  CORR 110
C                                     CORR 120
C DESCRIPTION OF PARAMETERS           CORR 130
C   N   - NUMBER OF OBSERVATIONS. N MUST BE > OR = TO 2.  CORR 140
C   M   - NUMBER OF VARIABLES. M MUST BE > OR = TO 1.    CORR 150
C   IO  - OPTION CODE FOR INPUT DATA  CORR 160
C         0 IF DATA ARE TO BE READ IN FROM INPUT DEVICE IN THE CORR 170
C           SPECIAL SUBROUTINE NAMED DATA. (SEE SUBROUTINES CORR 180
C           USED BY THIS SUBROUTINE BELOW.)  CORR 190
C         1 IF ALL DATA ARE ALREADY IN CORE.  CORR 200
C   X   - IF IO=0, THE VALUE OF X IS 0.0.  CORR 210
C         IF IO=1, X IS THE INPUT MATRIX (N BY M) CONTAINING CORR 220
C           DATA.  CORR 230
C   XBAR - OUTPUT VECTOR OF LENGTH M CONTAINING MEANS.  CORR 240
C   STD  - OUTPUT VECTOR OF LENGTH M CONTAINING STANDARD CORR 250
C           DEVIATIONS.  CORR 260
C   RX   - OUTPUT MATRIX (M X M) CONTAINING SUMS OF CROSS- CORR 270
C           PRODUCTS OF DEVIATIONS FROM MEANS.  CORR 280
C   R    - OUTPUT MATRIX (ONLY UPPER TRIANGULAR PORTION OF THE CORR 290
C           SYMMETRIC MATRIX OF M BY M) CONTAINING CORRELATION CORR 300
C           COEFFICIENTS. (STORAGE MODE OF 1)  CORR 310
C   B    - OUTPUT VECTOR OF LENGTH M CONTAINING THE DIAGONAL CORR 320
C           OF THE MATRIX OF SUMS OF CROSS-PRODUCTS OF CORR 330
C           DEVIATIONS FROM MEANS.  CORR 340
C   D    - WORKING VECTOR OF LENGTH M.  CORR 350
C   T    - WORKING VECTOR OF LENGTH M.  CORR 360
C                                     CORR 370
C REMARKS                               CORR 380
C   CORRE WILL NOT ACCEPT A CONSTANT VECTOR.  CORR 390
C                                     CORR 400
C SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED  CORR 410
C   DATA(M,D) - THIS SUBROUTINE MUST BE PROVIDED BY THE USER.  CORR 420
C   (1) IF IO=0, THIS SUBROUTINE IS EXPECTED TO  CORR 430
C       FURNISH AN OBSERVATION IN VECTOR D FROM AN  CORR 440
C       EXTERNAL INPUT DEVICE.  CORR 450
C   (2) IF IO=1, THIS SUBROUTINE IS NOT USED BY  CORR 460
C       CORRE BUT MUST EXIST IN JOB DECK. IF USER  CORR 470
C       HAS NOT SUPPLIED A SUBROUTINE NAMED DATA,  CORR 480
C       THE FOLLOWING IS SUGGESTED.  CORR 490
C           SUBROUTINE DATA  CORR 500
C           RETURN  CORR 510
C           END  CORR 520
C                                     CORR 530
C METHOD  CORR 540
C   PRODUCT-MOMENT CORRELATION COEFFICIENTS ARE COMPUTED.  CORR 550
C                                     CORR 560
C .....  CORR 570
C                                     CORR 580
C SUBROUTINE CORRE (N,M,IO,X,XBAR,STD,RX,R,B,D,T)  CORR 590
C DIMENSION X(1),XBAR(1),STD(1),RX(1),R(1),B(1),D(1),T(1)  CORR 600
C                                     CORR 610
C .....  CORR 620
C                                     CORR 630
C IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE  CORR 640
C C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION  CORR 650
C STATEMENT WHICH FOLLOWS.  CORR 660
C                                     CORR 670
C DOUBLE PRECISION XBAR,STD,RX,R,B,T  CORR 680

```

C		CORR 690
C	THE C MUST ALSO BE REMOVED FROM DCUBLE PRECISION STATEMENTS	CORR 700
C	APPEARING IN CTRER ROUTINES USED IN CONJUNCTION WITH THIS	CORR 710
C	ROUTINE.	CORR 720
C		CORR 730
C	THE DCUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO	CORR 740
C	CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. SQRT AND ABS IN	CORR 750
C	STATEMENT 220 MUST BE CHANGED TO DSQRT AND DABS.	CORR 760
C		CORR 770
C	CORR 780
C		CORR 790
C	INITIALIZATION	CORR 800
C		CORR 810
	JJM=0	
	DC 100 J=1,M	CORR 820
	B(J)=0.0	CORR 830
100	T(J)=0.0	CORR 840
	K=(M*M+M)/2	CORR 850
	DO 102 I=1,K	CORR 860
102	R(I)=0.0	CORR 870
	FN=N	CORR 880
	L=0	CORR 890
C		CORR 900
	IF(IC) 105, 127, 105	CORR 910
C		CORR 920
C	DATA ARE ALREADY IN CORE	CORR 930
C		CORR 940
105	DO 108 J=1,M	CORR 950
	DC 107 I=1,N	CORR 960
	L=L+1	CORR 970
107	T(J)=T(J)+X(L)	CORR 980
	XBAR(J)=T(J)	CORR 990
108	T(J)=T(J)/FN	CORR1000
C		CORR1010
	DC 115 I=1,N	CORR1020
	JK=0	CORR1030
	L=I-N	CORR1040
	DC 110 J=1,M	CORR1050
	L=L+N	CORR1060
	D(J)=X(L)-T(J)	CORR1070
110	B(J)=B(J)+D(J)	CORR1080
	DO 115 J=1,M	CORR1090
	DC 115 K=1,J	CORR1100
	JK=JK+1	CORR1110
115	R(JK)=R(JK)+D(J)*D(K)	CORR1120
	GC TO 205	CORR1130
C		CORR1140
C	READ OBSERVATIONS AND CALCULATE TEMPORARY	CORR1150
C	MEANS FROM THESE DATA IN T(J)	CORR1160
C		CORR1170
127	IF(N-M) 130, 130, 135	CORR1180
130	KK=N	CORR1190
	GO TO 137	CORR1200
135	KK=M	CORR1210
137	DO 140 I=1,KK	CORR1220
	CALL DATA (M,D,JJM)	
	DC 140 J=1,M	CORR1240
	T(J)=T(J)+D(J)	CORR1250
	L=L+1	CORR1260
140	RX(L)=D(J)	CORR1270

FKK=KK	CORR1280
DC 150 J=1,M	CORR1290
XBAR(J)=T(J)	CORR1300
150 T(J)=T(J)/FKK	CORR1310
C	CORR1320
C CALCULATE SUMS OF CROSS-PRODUCTS OF DEVIATIONS	CORR1330
C FROM TEMPORARY MEANS FOR M OBSERVATIONS	CORR1340
C	CORR1350
L=0	CORR1360
DC 160 I=1, KK	CORR1370
JK=0	CORR1380
DC 170 J=1, M	CORR1390
L=L+1	CORR1400
170 D(J)=RX(L)-T(J)	CORR1410
DC 180 J=1, M	CORR1420
E(J)=E(J)+D(J)	CORR1430
DC 180 K=1, J	CORR1440
JK=JK+1	CORR1450
180 R(JK)=R(JK)+D(J)*D(K)	CORR1460
C	CORR1470
IF(N-KK) 205, 205, 185	CORR1480
C	CORR1490
C READ THE REST OF OBSERVATIONS ONE AT A TIME, SUM	CORR1500
C THE OBSERVATION, AND CALCULATE SUMS OF CROSS-	CORR1510
C PRODUCTS OF DEVIATIONS FROM TEMPORARY MEANS	CORR1520
C	CORR1530
185 KK=N-KK	CORR1540
DC 200 I=1, KK	CORR1550
JK=0	CORR1560
CALL DATA (M, D, JJM)	
DC 190 J=1, M	CORR1580
XBAR(J)=XBAR(J)+D(J)	CORR1590
D(J)=D(J)-T(J)	CORR1600
190 B(J)=B(J)+D(J)	CORR1610
DC 200 J=1, M	CORR1620
DC 200 K=1, J	CORR1630
JK=JK+1	CORR1640
200 R(JK)=R(JK)+D(J)*D(K)	CORR1650
C	CORR1660
C CALCULATE MEANS	CORR1670
C	CORR1680
205 JK=0	CORR1690
DC 210 J=1, M	CORR1700
XBAR(J)=XBAR(J)/FN	CORR1710
C	CORR1720
C ADJUST SUMS OF CROSS-PRODUCTS OF DEVIATIONS	CORR1730
C FROM TEMPORARY MEANS	CORR1740
C	CORR1750
DC 210 K=1, J	CORR1760
JK=JK+1	CORR1770
210 K(JK)=R(JK)-B(J)*B(K)/FN	CORR1780
C	CORR1790
C CALCULATE CORRELATION COEFFICIENTS	CORR1800
C	CORR1810
JK=0	CORR1820
DC 220 J=1, M	CORR1830
JK=JK+J	CORR1840
220 STD(J)=SQRT(ABS(R(JK)))	CORR1850
DC 230 J=1, M	CORR1860
DC 230 K=J, M	CORR1870

```

JK=J+(K*K-K)/2
L=M*(J-1)+K
RX(L)=R(JK)
L=M*(K-1)+J
RX(L)=R(JK)
IF(STD(J)*STD(K)) 225, 222, 225
222 R(JK)=0.0
GC TC 230
225 R(JK)=R(JK)/(STD(J)*STD(K))
230 CONTINUE

C
C   CALCULATE STANDARD DEVIATIONS
C
   FN=SQRT(FN-1.0)
DU 240 J=1,M
240 STD(J)=STD(J)/FN

C
C   COPY THE DIAGONAL OF THE MATRIX OF SUMS OF CROSS-PRODUCTS OF
C   DEVIATIONS FROM MEANS.
C
   L=-M
   DC 250 I=1,M
   L=L+M+1
250 B(I)=RX(L)
   RETURN
   END
SUBROUTINE DATA(M,D,K)
CCMCN/LE/BANK(150,100)
DIMENSION(100)
K=K+1
DC 15 J=1,M
15 D(J)=BANK(K,J)
WRITE(6,1) J, D(J)
1 FORMAT(1X,I4,20X,F12.5)
WRITE(13) (D(I),I=1,M)
RETURN
END

C
C .....
C
C   SUBROUTINE CANOR
C
C   PURPOSE
C   COMPUTE THE CANONICAL CORRELATIONS BETWEEN TWO SETS OF
C   VARIABLES. CANOR IS NORMALLY PRECEDED BY A CALL TO SUBROU-
C   TIME CORRE.
C
C   USAGE
C   CALL CANOR (N,MP,MQ,RR,ROOTS,WLAM,CANR,CHISQ,NDF,COEFR,
C   COEFL,R)
C
C   DESCRIPTION OF PARAMETERS
C   N - NUMBER OF OBSERVATIONS
C   MP - NUMBER OF LEFT HAND VARIABLES
C   MQ - NUMBER OF RIGHT HAND VARIABLES
C   RR - INPUT MATRIX (ONLY UPPER TRIANGULAR PORTION OF THE
C   SYMMETRIC MATRIX OF M X M, WHERE M = MP + MQ)
C   CONTAINING CORRELATION COEFFICIENTS. (STORAGE MODE
C   CF 1)
C   ROOTS - CLPUT VECTOR OF LENGTH MQ CONTAINING EIGENVALUES

```

```

CORR1880
CORR1890
CORR1900
CORR1910
CORR1920
CORR1930
CORR1940
CORR1950
CORR1960
CORR1970
CORR1980
CORR1990
CORR2000
CORR2010
CORR2020
CORR2030
CORR2040
CORR2050
CORR2060
CORR2070
CORR2080
CORR2090
CORR2100
CORR2110
CORR2120
CORR2130

CANO 10
CANO 20
CANO 30
CANO 40
CANO 50
CANO 60
CANO 70
CANO 80
CANO 90
CANO 100
CANO 110
CANO 120
CANO 130
CANO 140
CANO 150
CANO 160
CANO 170
CANO 180
CANO 190
CANO 200
CANO 210
CANO 220
CANO 230

```

C		COMPUTED IN THE NROOT SUBROUTINE.	CANO 240
C	WLAM	- OUTPUT VECTOR OF LENGTH MQ CONTAINING LAMBDA.	CANO 250
C	CANR	- OUTPUT VECTOR OF LENGTH MQ CONTAINING CANONICAL	CANO 260
C		CORRELATIONS.	CANO 270
C	CHISQ	- OUTPUT VECTOR OF LENGTH MQ CONTAINING THE	CANO 280
C		VALUES OF CHI-SQUARES.	CANO 290
C	NDF	- OUTPUT VECTOR OF LENGTH MQ CONTAINING THE DEGREES	CANO 300
C		OF FREEDOM ASSOCIATED WITH CHI-SQUARES.	CANO 310
C	COEFR	- OUTPUT MATRIX (MQ X MQ) CONTAINING MQ SETS OF	CANO 320
C		RIGHT HAND COEFFICIENTS COLUMNWISE.	CANO 330
C	COEFL	- OUTPUT MATRIX (MP X MQ) CONTAINING MQ SETS OF	CANO 340
C		LEFT HAND COEFFICIENTS COLUMNWISE.	CANO 350
C	R	- WORK MATRIX (M X M)	CANO 360
C			CANO 370
C	REMARKS		CANO 380
C		THE NUMBER OF LEFT HAND VARIABLES (MP) SHOULD BE GREATER	CANO 390
C		THAN OR EQUAL TO THE NUMBER OF RIGHT HAND VARIABLES (MQ).	CANO 400
C		THE VALUES OF CANONICAL CORRELATION, LAMBDA, CHI-SQUARE,	CANO 410
C		DEGREES OF FREEDOM, AND CANONICAL COEFFICIENTS ARE COMPUTED	CANO 420
C		ONLY FOR THOSE EIGENVALUES IN ROOTS WHICH ARE GREATER THAN	CANO 430
C		ZERO.	CANO 440
C			CANO 450
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED		CANO 460
C	MINV		CANO 470
C	NROOT	(WHICH, IN TURN, CALLS THE SUBROUTINE EIGEN.)	CANO 480
C			CANO 490
C	METHOD		CANO 500
C		REFER TO W. W. COOLEY AND P. R. LOHNES, 'MULTIVARIATE PRO-	CANO 510
C		CEDURES FOR THE BEHAVIORAL SCIENCES', JOHN WILEY AND SONS,	CANO 520
C		1962, CHAPTER 3.	CANO 530
C			CANO 540
C		CANO 550
C			CANO 560
C	SUBROUTINE CANGR (N,MP,MQ,RR,ROOTS,WLAM,CANR,CHI SQ,NDF,COEFR,		CANO 570
C	1	COEFL,R)	CANO 580
C		DIMENSION RR(1),ROOTS(1),WLAM(1),CANR(1),CHISQ(1),NDF(1),COEFR(1),	CANO 590
C	1	COEFL(1),R(1)	CANO 600
C			CANO 610
C		CANO 620
C			CANO 630
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE		CANO 640
C	C	IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	CANO 650
C		STATEMENT WHICH FOLLOWS.	CANO 660
C			CANO 670
C	DOUBLE PRECISION RR,ROOTS,WLAM,CANR,CHISQ,COEFR,COEFL,R,DET,SUM		CANO 680
C			CANO 690
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS		CANO 700
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS		CANO 710
C	ROUTINE.		CANO 720
C			CANO 730
C	THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO		CANO 740
C	CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. SQRT IN STATEMENT		CANO 750
C	165 MUST BE CHANGED TO DSQRT. ALOG IN STATEMENT 175 MUST BE		CANO 760
C	CHANGED TO DLOG.		CANO 770
C			CANO 780
C		CANO 790
C			CANO 800
C	PARTITION INTERCORRELATIONS AMONG LEFT HAND VARIABLES, BETWEEN		CANO 810
C	LEFT AND RIGHT HAND VARIABLES, AND AMONG RIGHT HAND VARIABLES.		CANO 820
C			CANO 830

M=MP+MQ	CAN0 840
N1=0	CAN0 850
DC 105 I=1,M	CAN0 860
DU 105 J=1,M	CAN0 870
IF(I-J) 102, 103, 103	CAN0 880
102 L=1+(J*J-J)/2	CAN0 890
GO TU 104	CAN0 900
103 L=J+(I*I-I)/2	CAN0 910
104 N1=N1+1	CAN0 920
105 R(N1)=RR(L)	CAN0 930
L=MP	CAN0 940
DO 108 J=2,MP	CAN0 950
N1=M*(J-1)	CAN0 960
DC 108 I=1,MP	CAN0 970
L=L+1	CAN0 980
N1=N1+1	CAN0 990
108 R(L)=R(N1)	CAN01000
N2=MP+1	CAN01010
L=0	CAN01020
DO 110 J=N2,M	CAN01030
N1=M*(J-1)	CAN01040
DC 110 I=1,MP	CAN01050
L=L+1	CAN01060
N1=N1+1	CAN01070
110 CCEFL(L)=R(N1)	CAN01080
L=0	CAN01090
DC 120 J=N2,M	CAN01100
N1=M*(J-1)+MP	CAN01110
DO 120 I=N2,M	CAN01120
L=L+1	CAN01130
N1=N1+1	CAN01140
120 COEFR(L)=R(N1)	CAN01150
C	CAN01160
C SOLVE THE CANONICAL EQUATION	CAN01170
C	CAN01180
L=MP*MP+1	CAN01190
K=L+MP	CAN01200
CALL MINV (R,MP,DET,R(L),R(K))	CAN01210
C	CAN01220
C CALCULATE T = INVERSE OF R11 * R12	CAN01230
C	CAN01240
DC 140 I=1,MP	CAN01250
N2=0	CAN01260
DC 130 J=1,MP	CAN01270
N1=I-MP	CAN01280
ROOTS(J)=0.0	CAN01290
DC 130 K=1,MP	CAN01300
N1=N1+MP	CAN01310
N2=N2+1	CAN01320
130 ROOTS(J)=ROOTS(J)+R(N1)*COEFL(N2)	CAN01330
L=I-MP	CAN01340
DO 140 J=1,MP	CAN01350
L=L+MP	CAN01360
140 R(L)=ROOTS(J)	CAN01370
C	CAN01380
C CALCULATE A = R21 * T	CAN01390
C	CAN01400
L=MP*MP	CAN01410
N3=L+1	CAN01420
DO 160 J=1,MP	CAN01430

N1=0	CAN01440
DO 160 I=1,MQ	CAN01450
N2=MP*(J-1)	CAN01460
SLM=0.0	CAN01470
DO 150 K=1,MP	CAN01480
N1=N1+1	CAN01490
N2=N2+1	CAN01500
150 SLM=SLM+COEFL(N1)*R(N2)	CAN01510
L=L+1	CAN01520
160 R(L)=SLM	CAN01530
C	CAN01540
C	CAN01550
C	CAN01560
C	CAN01570
L=L+1	CAN01580
CALL NROOT (MQ,R(N3),COEFR,ROOTS,R(L))	CAN01590
C	CAN01600
C	CAN01610
C	CAN01620
C	CAN01630
DO 210 I=1,MQ	CAN01640
C	CAN01650
C	CAN01660
C	CAN01670
TEST WHETHER EIGENVALUE IS GREATER THAN ZERO	CAN01680
C	CAN01690
C	CAN01700
IF(ROOTS(I)) >= 0, 220, 165	CAN01710
C	CAN01720
C	CAN01730
C	CAN01740
C	CAN01750
C	CAN01760
C	CAN01770
C	CAN01780
C	CAN01790
C	CAN01800
C	CAN01810
C	CAN01820
C	CAN01830
C	CAN01840
C	CAN01850
C	CAN01860
C	CAN01870
C	CAN01880
C	CAN01890
C	CAN01900
C	CAN01910
C	CAN01920
C	CAN01930
C	CAN01940
C	CAN01950
C	CAN01960
C	CAN01970
C	CAN01980
C	CAN01990
C	CAN02000
C	CAN02010
C	CAN02020
C	CAN02030

```

COEFL(K)=0.0
DC 190 JJ=1,MC
N1=N1+MP
N2=N2+1
190 CCEFL(K)=COEFL(K)+R(N1)*COEFR(N2)
200 CCEFL(K)=CCEFL(K)/CANR(I)
210 CONTINUE
220 RETURN
END

```

CAN02040
CAN02050
CAN02060
CAN02070
CAN02080
CAN02090
CAN02100
CAN02110
CAN02120

```

C
C .....
C
C SUBROUTINE EIGEN
C
C PURPOSE
C   COMPUTE EIGENVALUES AND EIGENVECTORS OF A REAL SYMMETRIC
C   MATRIX
C
C USAGE
C   CALL EIGEN(A,R,N,MV)
C
C DESCRIPTION OF PARAMETERS
C   A - ORIGINAL MATRIX (SYMMETRIC), DESTROYED IN COMPUTATION.
C       RESULTANT EIGENVALUES ARE DEVELOPED IN DIAGONAL OF
C       MATRIX A IN DESCENDING ORDER.
C   R - RESULTANT MATRIX OF EIGENVECTORS (STORED COLUMNWISE,
C       IN SAME SEQUENCE AS EIGENVALUES)
C   N - ORDER OF MATRICES A AND R
C   MV- INPUT CODE
C       0  COMPUTE EIGENVALUES AND EIGENVECTORS
C       1  COMPUTE EIGENVALUES ONLY (R NEED NOT BE
C         DIMENSIONED BUT MUST STILL APPEAR IN CALLING
C         SEQUENCE)
C
C REMARKS
C   ORIGINAL MATRIX A MUST BE REAL SYMMETRIC (STORAGE MODE=1)
C   MATRIX A CANNOT BE IN THE SAME LOCATION AS MATRIX R
C
C SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C   NONE
C
C METHOD
C   DIAGONALIZATION METHOD ORIGINATED BY JACOBI AND ADAPTED
C   BY VON NEUMANN FOR LARGE COMPUTERS AS FOUND IN "MATHEMATICAL
C   METHODS FOR DIGITAL COMPUTERS", EDITED BY A. RALSTON AND
C   H.S. WILF, JOHN WILEY AND SONS, NEW YORK, 1962, CHAPTER 7
C
C .....
C
C SUBROUTINE EIGEN(A,R,N,MV)
C DIMENSION A(1),R(1)
C
C .....
C
C IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE
C C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION
C STATEMENT WHICH FOLLOWS.
C
C DOUBLE PRECISION A,R,ANORM,ANRMX,THR,X,Y,SINX,SINX2,COSX,
C 1 COSX2,SINCS,RANGE

```

EIGE 10
EIGE 20
EIGE 30
EIGE 40
EIGE 50
EIGE 60
EIGE 70
EIGE 80
EIGE 90
EIGE 100
EIGE 110
EIGE 120
EIGE 130
EIGE 140
EIGE 150
EIGE 160
EIGE 170
EIGE 180
EIGE 190
EIGE 200
EIGE 210
EIGE 220
EIGE 230
EIGE 240
EIGE 250
EIGE 260
EIGE 270
EIGE 280
EIGE 290
EIGE 300
EIGE 310
EIGE 320
EIGE 330
EIGE 340
EIGE 350
EIGE 360
EIGE 370
EIGE 380
EIGE 390
EIGE 400
EIGE 410
EIGE 420
EIGE 430
EIGE 440
EIGE 450
EIGE 460
EIGE 470
EIGE 480
EIGE 490
EIGE 500
EIGE 510

C		EIGE 520
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	EIGE 530
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	EIGE 540
C	ROUTINE.	EIGE 550
C		EIGE 560
C	THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO	EIGE 570
C	CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. SQRT IN STATEMENTS	EIGE 580
C	40, 68, 75, AND 78 MUST BE CHANGED TO DSQRT. ABS IN STATEMENT	EIGE 590
C	62 MUST BE CHANGED TO DABS. THE CONSTANT IN STATEMENT 5 SHOULD	EIGE 600
C	BE CHANGED TO 1.0D-12.	EIGE 610
C		EIGE 620
C	EIGE 630
C		EIGE 640
C	GENERATE IDENTITY MATRIX	EIGE 650
C		EIGE 660
C	5 RANGE=1.0E-6	EIGE 670
C	IF(MV-1) 10,25,10	EIGE 680
C	10 IC=-N	EIGE 690
C	DO 20 J=1,N	EIGE 700
C	IQ=IQ+N	EIGE 710
C	DC 20 I=1,N	EIGE 720
C	IJ=IQ+I	EIGE 730
C	R(IJ)=0.0	EIGE 740
C	IF(I-J) 20,15,20	EIGE 750
C	15 R(IJ)=1.0	EIGE 760
C	20 CONTINUE	EIGE 770
C		EIGE 780
C	COMPUTE INITIAL AND FINAL NORMS (ANORM AND ANCRMX)	EIGE 790
C		EIGE 800
C	25 ANORM=0.0	EIGE 810
C	DO 35 I=1,N	EIGE 820
C	DC 35 J=I,N	EIGE 830
C	IF(I-J) 30,35,30	EIGE 840
C	30 IA=I+(J-J-J)/2	EIGE 850
C	ANCRM=ANORM+A(IA)*A(IA)	EIGE 860
C	35 CONTINUE	EIGE 870
C	IF(ANORM) 165,165,40	EIGE 880
C	40 ANCRM=1.414*SQRT(ANORM)	EIGE 890
C	ANRMX=ANORM*RANGE/FLOAT(N)	EIGE 900
C		EIGE 910
C	INITIALIZE INDICATORS AND COMPUTE THRESHOLD, THR	EIGE 920
C		EIGE 930
C	IND=0	EIGE 940
C	THR=ANORM	EIGE 950
C	45 THR=THR/FLOAT(N)	EIGE 960
C	50 L=1	EIGE 970
C	55 M=L+1	EIGE 980
C		EIGE 990
C	COMPUTE SIN AND COS	EIGE1000
C		EIGE1010
C	60 MQ=(M*M-M)/2	EIGE1020
C	LG=(L*L-L)/2	EIGE1030
C	LM=L+MQ	EIGE1040
C	62 IF(ABS(A(LM))-THR) 130,65,65	EIGE1050
C	65 IND=1	EIGE1060
C	LL=L+LQ	EIGE1070
C	MM=M+MQ	EIGE1080
C	X=0.5*(A(LL)-A(MM))	EIGE1090
C	68 Y=-A(LM)/SQRT(A(LM)*A(LM)+X*X)	EIGE1100
C	IF(X) 70,75,75	EIGE1110

70	Y=-Y	EIGE1120
75	SINX=Y/ SQRT(2.0*(1.0+(SQRT(1.0-Y*Y))))	EIGE1130
	SINX2=SINX*SINX	EIGE1140
78	CCSX= SQRT(1.0-SINX2)	EIGE1150
	CCSX2=CUSX*CUSX	EIGE1160
	SINCS =SINX*CCSX	EIGE1170
C		EIGE1180
C	ROTATE L AND M COLUMNS	EIGE1190
C		EIGE1200
	ILQ=N*(L-1)	EIGE1210
	IMQ=N*(M-1)	EIGE1220
	DC 125 I=1,N	EIGE1230
	IQ=(1*I-I)/2	EIGE1240
	IF(I-L) 80,115,80	EIGE1250
80	IF(I-M) 85,115,90	EIGE1260
85	IM=I+MQ	EIGE1270
	GC TC 95	EIGE1280
90	IM=M+IQ	EIGE1290
95	IF(I-L) 100,105,105	EIGE1300
100	IL=I+LQ	EIGE1310
	GU TO 110	EIGE1320
105	IL=L+IQ	EIGE1330
110	X=A(IL)*CCSX-A(IM)*SINX	EIGE1340
	A(IM)=A(IL)*SINX+A(IM)*CCSX	EIGE1350
	A(IL)=X	EIGE1360
115	IF(MV-1) 120,125,120	EIGE1370
120	ILR=ILQ+I	EIGE1380
	IMR=IMQ+I	EIGE1390
	X=R(ILR)*CCSX-R(IMR)*SINX	EIGE1400
	R(IMR)=R(ILR)*SINX+R(IMR)*CCSX	EIGE1410
	R(ILR)=X	EIGE1420
125	CONTINUE	EIGE1430
	X=2.0*A(LM)*SINCS	EIGE1440
	Y=A(LL)*CCSX2+A(MM)*SINX2-X	EIGE1450
	X=A(LL)*SINX2+A(MM)*CCSX2+X	EIGE1460
	A(LM)=(A(LL)-A(MM))*SINCS+A(LM)*(CCSX2-SINX2)	EIGE1470
	A(LL)=Y	EIGE1480
	A(MM)=X	EIGE1490
C		EIGE1500
C	TESTS FOR COMPLETION	EIGE1510
C		EIGE1520
C	TEST FOR M = LAST COLUMN	EIGE1530
C		EIGE1540
130	IF(M-N) 135,140,135	EIGE1550
135	M=M+1	EIGE1560
	GO TO 60	EIGE1570
C		EIGE1580
C	TEST FOR L = SECOND FROM LAST COLUMN	EIGE1590
C		EIGE1600
140	IF(L-(N-1)) 145,150,145	EIGE1610
145	L=L+1	EIGE1620
	GU TO 55	EIGE1630
150	IF(IND-1) 160,155,160	EIGE1640
155	IND=0	EIGE1650
	GU TO 50	EIGE1660
C		EIGE1670
C	COMPARE THRESHOLD WITH FINAL NORM	EIGE1680
C		EIGE1690
160	IF(THR-ANRMX) 165,165,45	EIGE1700
C		EIGE1710

C	SCRT EIGENVALUES AND EIGENVECTORS	EIGE1720
C		EIGE1730
165	IQ=-N	EIGE1740
	DC 185 I=1,N	EIGE1750
	IQ=IQ+N	EIGE1760
	LL=I+(I*I-I)/2	EIGE1770
	JG=N*(I-2)	EIGE1780
	DC 185 J=I,N	EIGE1790
	JQ=JG+N	EIGE1800
	MP=J+(J*J-J)/2	EIGE1810
	IF(A(LL)-A(MM)) 170,185,185	EIGE1820
170	X=A(LL)	EIGE1830
	A(LL)=A(MM)	EIGE1840
	A(MM)=X	EIGE1850
	IF(MV-1) 175,185,175	EIGE1860
175	DC 180 K=1,N	EIGE1870
	ILR=IQ+K	EIGE1880
	IMR=JG+K	EIGE1890
	X=R(ILR)	EIGE1900
	R(ILR)=K(IMR)	EIGE1910
180	R(IMR)=X	EIGE1920
185	CUNTINL	EIGE1930
	RETURN	EIGE1940
	END	EIGE1950
C		LOC 10
C	LOC 20
C		LOC 30
C	SUBROUTINE LOC	LOC 40
C		LOC 50
C	PURPCSE	LOC 60
C	COMPUTE A VECTOR SUBSCRIPT FOR AN ELEMENT IN A MATRIX OF	LOC 70
C	SPECIFIED STORAGE MODE	LOC 80
C		LOC 90
C	USAGE	LOC 100
C	CALL LOC (I,J,IR,N,M,MS)	LOC 110
C		LOC 120
C	DESCRIPTION OF PARAMETERS	LOC 130
C	I - ROW NUMBER OF ELEMENT	LOC 140
C	J - CCLUMN NUMBER OF ELEMENT	LOC 150
C	IR - RESULTANT VECTOR SUBSCRIPT	LOC 160
C	N - NUMBER OF ROWS IN MATRIX	LOC 170
C	M - NUMBER OF COLUMNS IN MATRIX	LOC 180
C	MS - ONE DIGIT NUMBER FOR STORAGE MODE OF MATRIX	LOC 190
C	0 - GENERAL	LOC 200
C	1 - SYMMETRIC	LOC 210
C	2 - DIAGONAL	LOC 220
C		LOC 230
C	REMARKS	LOC 240
C	NONE	LOC 250
C		LOC 260
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	LOC 270
C	NONE	LOC 280
C		LOC 290
C	METHOD	LOC 300
C	MS=0 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N*M ELEMENTS	LOC 310
C	IN STORAGE (GENERAL MATRIX)	LOC 320
C	MS=1 SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N*(N+1)/2 IN	LOC 330
C	STORAGE (UPPER TRIANGLE OF SYMMETRIC MATRIX). IF	LOC 340
C	ELEMENT IS IN LOWER TRIANGULAR PORTION, SUBSCRIPT IS	LOC 350
C	CORRESPONDING ELEMENT IN UPPER TRIANGLE.	LOC 360

C	MS=2	SUBSCRIPT IS COMPUTED FOR A MATRIX WITH N ELEMENTS	LOC	370
C		IN STORAGE (DIAGONAL ELEMENTS OF DIAGONAL MATRIX).	LOC	380
C		IF ELEMENT IS NOT ON DIAGONAL (AND THEREFORE NOT IN	LOC	390
C		STORAGE), IR IS SET TO ZERO.	LOC	400
C			LOC	410
C		LOC	420
C			LOC	430
C		SUBROUTINE LOC(I,J,IR,N,M,MS)	LOC	440
C			LOC	450
		IX=I	LOC	460
		JX=J	LOC	470
		IF(MS-1) 10,20,30	LOC	480
10		IRX=N*(JX-1)+IX	LOC	490
		GC TO 36	LOC	500
20		IF(IX-JX) 22,24,24	LOC	510
22		IRX=IX+(JX*JX-JX)/2	LOC	520
		GC TO 36	LOC	530
24		IRX=JX+(IX*IX-IX)/2	LOC	540
		GC TO 36	LOC	550
30		IRX=0	LOC	560
		IF(IX-JX) 36,32,36	LOC	570
32		IRX=IX	LOC	580
36		IR=IRX	LOC	590
		RETURN	LOC	600
		END	LOC	610
C			MINV	10
C		MINV	20
C			MINV	30
C		SUBROUTINE MINV	MINV	40
C			MINV	50
C		PURPOSE	MINV	60
C		INVERT A MATRIX	MINV	70
C			MINV	80
C		USAGE	MINV	90
C		CALL MINV(A,N,D,L,M)	MINV	100
C			MINV	110
C		DESCRIPTION OF PARAMETERS	MINV	120
C		A - INPUT MATRIX, DESTROYED IN COMPUTATION AND REPLACED BY	MINV	130
C		RESULTANT INVERSE.	MINV	140
C		N - ORDER OF MATRIX A	MINV	150
C		D - RESULTANT DETERMINANT	MINV	160
C		L - WORK VECTOR OF LENGTH N	MINV	170
C		M - WORK VECTOR OF LENGTH N	MINV	180
C			MINV	190
C		REMARKS	MINV	200
C		MATRIX A MUST BE A GENERAL MATRIX	MINV	210
C			MINV	220
C		SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	MINV	230
C		NONE	MINV	240
C			MINV	250
C		METHOD	MINV	260
C		THE STANDARD GAUSS-JORDAN METHOD IS USED. THE DETERMINANT	MINV	270
C		IS ALSO CALCULATED. A DETERMINANT OF ZERO INDICATES THAT	MINV	280
C		THE MATRIX IS SINGULAR.	MINV	290
C			MINV	300
C		MINV	310
C			MINV	320
C		SUBROUTINE MINV(A,N,D,L,M)	MINV	330
C		DIMENSION A(1),L(1),M(1)	MINV	340
C			MINV	350

L	MINV	360
C		MINV	370
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	MINV	380
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	MINV	390
C	STATEMENT WHICH FOLLOWS.	MINV	400
C		MINV	410
C	DOUBLE PRECISION A,D,BIGA,HOLD	MINV	420
C		MINV	430
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	MINV	440
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	MINV	450
C	ROUTINE.	MINV	460
C		MINV	470
C	THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO	MINV	480
C	CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. ABS IN STATEMENT	MINV	490
C	10 MUST BE CHANGED TO DABS.	MINV	500
C		MINV	510
C	MINV	520
C		MINV	530
C	SEARCH FOR LARGEST ELEMENT	MINV	540
C		MINV	550
	D=1.0	MINV	560
	NK=-N	MINV	570
	DO 80 K=1,N	MINV	580
	NK=NK+N	MINV	590
	L(K)=K	MINV	600
	M(K)=K	MINV	610
	KK=NK+K	MINV	620
	BIGA=A(KK)	MINV	630
	DO 20 J=K,N	MINV	640
	IZ=N*(J-1)	MINV	650
	DO 20 I=K,N	MINV	660
	IJ=IZ+I	MINV	670
10	IF(ABS(BIGA)- ABS(A(IJ))) 15,20,20	MINV	680
15	BIGA=A(IJ)	MINV	690
	L(K)=I	MINV	700
	M(K)=J	MINV	710
20	CONTINUE	MINV	720
C		MINV	730
C	INTERCHANGE ROWS	MINV	740
C		MINV	750
	J=L(K)	MINV	760
	IF(J-K) 35,35,25	MINV	770
25	KI=K-N	MINV	780
	DO 30 I=1,N	MINV	790
	KI=KI+N	MINV	800
	HOLD=-A(KI)	MINV	810
	JI=KI-K+J	MINV	820
	A(KI)=A(JI)	MINV	830
30	A(JI)=HOLD	MINV	840
C		MINV	850
C	INTERCHANGE COLUMNS	MINV	860
C		MINV	870
35	I=M(K)	MINV	880
	IF(I-K) 45,45,38	MINV	890
38	JP=N*(I-1)	MINV	900
	DO 40 J=1,N	MINV	910
	JK=NK+J	MINV	920
	JJ=JP+J	MINV	930
	HOLD=-A(JK)	MINV	940
	A(JK)=A(JJ)	MINV	950

40	A(JI) =HOLD	MINV 960
C		MINV 970
C	DIVIDE COLUMN BY MINUS PIVOT (VALUE OF PIVOT ELEMENT IS	MINV 980
C	CONTAINED IN BIGA)	MINV 990
C		MINV1000
45	IF(BIGA) 48,46,48	MINV1010
46	D=0.0	MINV1020
	RETURN	MINV1030
48	DO 55 I=1,N	MINV1040
	IF(I-K) 50,55,50	MINV1050
50	IK=NK+I	MINV1060
	A(IK)=A(IK)/(-BIGA)	MINV1070
55	CONTINUE	MINV1080
C		MINV1090
C	REDUCE MATRIX	MINV1100
C		MINV1110
	DO 65 I=1,N	MINV1120
	IK=NK+I	MINV1130
	HOLD=A(IK)	MINV1140
	IJ=I-N	MINV1150
	DO 65 J=1,N	MINV1160
	IJ=IJ+N	MINV1170
	IF(I-K) 60,65,60	MINV1180
60	IF(J-K) 62,65,62	MINV1190
62	KJ=IJ-I+K	MINV1200
	A(IJ)=HOLD*A(KJ)+A(IJ)	MINV1210
65	CONTINUE	MINV1220
C		MINV1230
C	DIVIDE ROW BY PIVOT	MINV1240
C		MINV1250
	KJ=K-N	MINV1260
	DO 75 J=1,N	MINV1270
	KJ=KJ+N	MINV1280
	IF(J-K) 70,75,70	MINV1290
70	A(KJ)=A(KJ)/BIGA	MINV1300
75	CONTINUE	MINV1310
C		MINV1320
C	PRODUCT OF PIVOTS	MINV1330
C		MINV1340
	D=D*BIGA	MINV1350
C		MINV1360
C	REPLACE PIVOT BY RECIPROCAL	MINV1370
C		MINV1380
	A(KK)=1.0/BIGA	MINV1390
80	CONTINUE	MINV1400
C		MINV1410
C	FINAL ROW AND COLUMN INTERCHANGE	MINV1420
C		MINV1430
	K=N	MINV1440
100	K=(K-1)	MINV1450
	IF(K) 150,150,105	MINV1460
105	I=L(K)	MINV1470
	IF(I-K) 120,120,108	MINV1480
108	JQ=N*(K-1)	MINV1490
	JK=N*(I-1)	MINV1500
	DO 110 J=1,N	MINV1510
	JK=JQ+J	MINV1520
	HOLD=A(JK)	MINV1530
	JI=JR+J	MINV1540
	A(JK)=-A(JI)	MINV1550

```

110 A(JI) =HOLD
120 J=M(K)
    IF(J-K) 100,100,125
125 KI=K-N
    DO 130 I=1,N
    KI=KI+N
    HCLD=A(KI)
    JI=KI-K+J
    A(KI)=-A(JI)
130 A(JI) =HCLD
    GO TO 100
150 RETURN
    END

```

```

MINV1560
MINV1570
MINV1580
MINV1590
MINV1600
MINV1610
MINV1620
MINV1630
MINV1640
MINV1650
MINV1660
MINV1670
MINV1680

```

```

C
C
C .....
C
C     SUBROUTINE MSTR
C
C     PURPOSE
C     CHANGE STORAGE MODE OF A MATRIX
C
C     USAGE
C     CALL MSTR(A,R,N,MSA,MSR)
C
C     DESCRIPTION OF PARAMETERS
C     A - NAME OF INPLT MATRIX
C     R - NAME OF OUTPUT MATRIX
C     N - NUMBER OF RCWS AND COLUMNS IN A AND R
C     MSA - ONE DIGIT NUMBER FOR STORAGE MODE OF MATRIX A
C           0 - GENERAL
C           1 - SYMMETRIC
C           2 - DIAGONAL
C     MSR - SAME AS MSA EXCEPT FOR MATRIX R
C
C     REMARKS
C     MATRIX R CANNOT BE IN THE SAME LOCATION AS MATRIX A
C     MATRIX A MUST BE A SQUARE MATRIX
C
C     SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
C     LJC
C
C     METHOD
C     MATRIX A IS RESTRUCTURED TO FORM MATRIX R.
C     MSA MSR
C     0 0 MATRIX A IS MOVED TO MATRIX R
C     0 1 THE UPPER TRIANGLE ELEMENTS OF A GENERAL MATRIX
C         ARE USED TO FORM A SYMMETRIC MATRIX
C     0 2 THE DIAGONAL ELEMENTS OF A GENERAL MATRIX ARE USED
C         TO FORM A DIAGONAL MATRIX
C     1 0 A SYMMETRIC MATRIX IS EXPANDED TO FORM A GENERAL
C         MATRIX
C     1 1 MATRIX A IS MOVED TO MATRIX R
C     1 2 THE DIAGONAL ELEMENTS OF A SYMMETRIC MATRIX ARE
C         USED TO FORM A DIAGONAL MATRIX
C     2 0 A DIAGONAL MATRIX IS EXPANDED BY INSERTING MISSING
C         ZERO ELEMENTS TO FORM A GENERAL MATRIX
C     2 1 A DIAGONAL MATRIX IS EXPANDED BY INSERTING MISSING
C         ZERO ELEMENTS TO FORM A SYMMETRIC MATRIX
C     2 2 MATRIX A IS MOVED TO MATRIX R

```

```

MSTR 10
MSTR 20
MSTR 30
MSTR 40
MSTR 50
MSTR 60
MSTR 70
MSTR 80
MSTR 90
MSTR 100
MSTR 110
MSTR 120
MSTR 130
MSTR 140
MSTR 150
MSTR 160
MSTR 170
MSTR 180
MSTR 190
MSTR 200
MSTR 210
MSTR 220
MSTR 230
MSTR 240
MSTR 250
MSTR 260
MSTR 270
MSTR 280
MSTR 290
MSTR 300
MSTR 310
MSTR 320
MSTR 330
MSTR 340
MSTR 350
MSTR 360
MSTR 370
MSTR 380
MSTR 390
MSTR 400
MSTR 410
MSTR 420
MSTR 430
MSTR 440
MSTR 450
MSTR 460
MSTR 470

```

```

C .....MSTR 480
C .....MSTR 490
SUBROUTINE MSTR(A,R,N,MSA,MSR)MSTR 500
DIMENSION A(1),R(1)MSTR 510
C .....MSTR 520
LD 20 I=1,NMSTR 530
DC 20 J=1,NMSTR 540
C .....MSTR 550
C IF K IS GENERAL, FORM ELEMENTMSTR 560
C .....MSTR 570
IF(MSR) 5,10,5MSTR 580
C .....MSTR 590
C IF IN LOWER TRIANGLE OF SYMMETRIC OR DIAGONAL R, BYPASSMSTR 600
C .....MSTR 610
5 IF(I-J) 10,10,20MSTR 620
10 CALL LOC(I,J,IR,N,N,MSR)MSTR 630
C .....MSTR 640
C IF IN UPPER AND OFF DIAGONAL OF DIAGONAL R, BYPASSMSTR 650
C .....MSTR 660
IF(IR) 20,20,15MSTR 670
C .....MSTR 680
C OTHERWISE, FORM R(I,J)MSTR 690
C .....MSTR 700
15 R(IR)=0.0MSTR 710
CALL LOC(I,J,IA,N,N,MSA)MSTR 720
C .....MSTR 730
C IF THERE IS NO A(I,J), LEAVE R(I,J) AT 0.0MSTR 740
C .....MSTR 750
IF(IA) 20,20,18MSTR 760
18 R(IR)=A(IA)MSTR 770
20 CONTINUEMSTR 780
RETURNMSTR 790
ENDMSTR 800
C .....MULT 10
C .....MULT 20
SUBROUTINE MLTRMULT 30
C .....MULT 40
PURPOSEMULT 50
C .....MULT 60
PERFORM A MULTIPLE LINEAR REGRESSION ANALYSIS FOR A MULT 70
DEPENDENT VARIABLE AND A SET OF INDEPENDENT VARIABLES. THISMULT 80
SUBROUTINE IS NORMALLY USED IN THE PERFORMANCE OF MULTIPLE MULT 90
AND POLYNOMIAL REGRESSION ANALYSES.MULT 100
C .....MULT 110
USAGEMULT 120
CALL MLTR (N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)MULT 130
C .....MULT 140
DESCRIPTION OF PARAMETERSMULT 150
C .....MULT 160
N - NUMBER OF OBSERVATIONS.MULT 170
C .....MULT 180
K - NUMBER OF INDEPENDENT VARIABLES IN THIS REGRESSION.MULT 180
XBAR - INPUT VECTOR OF LENGTH M CONTAINING MEANS OF ALL MULT 190
VARIABLES. M IS NUMBER OF VARIABLES IN OBSERVATIONS.MULT 190
C .....MULT 200
STD - INPUT VECTOR OF LENGTH M CONTAINING STANDARD DEVI- MULT 200
ATIONS OF ALL VARIABLES.MULT 210
C .....MULT 220
D - INPUT VECTOR OF LENGTH M CONTAINING THE DIAGONAL OF MULT 220
THE MATRIX OF SUMS OF CROSS-PRODUCTS OF DEVIATIONS MULT 230
FROM MEANS FOR ALL VARIABLES.MULT 240
C .....MULT 250
RX - INPUT MATRIX (K X K) CONTAINING THE INVERSE OF MULT 250
INTERCORRELATIONS AMONG INDEPENDENT VARIABLES.MULT 260
C .....MULT 270
RY - INPUT VECTOR OF LENGTH K CONTAINING INTERCORRELA- MULT 270

```

C	TIONS OF INDEPENDENT VARIABLES WITH DEPENDENT	MULT 280
C	VARIABLE.	MULT 290
C	ISAVE - INPUT VECTOR OF LENGTH K+1 CONTAINING SUBSCRIPTS OF	MULT 300
C	INDEPENDENT VARIABLES IN ASCENDING ORDER. THE	MULT 310
C	SUBSCRIPT OF THE DEPENDENT VARIABLE IS STORED IN	MULT 320
C	THE LAST, K+1, POSITION.	MULT 330
C	B - OUTPUT VECTOR OF LENGTH K CONTAINING REGRESSION	MULT 340
C	COEFFICIENTS.	MULT 350
C	SB - OUTPUT VECTOR OF LENGTH K CONTAINING STANDARD	MULT 360
C	DEVIATIONS OF REGRESSION COEFFICIENTS.	MULT 370
C	T - OUTPUT VECTOR OF LENGTH K CONTAINING T-VALUES.	MULT 380
C	ANS - OUTPUT VECTOR OF LENGTH 10 CONTAINING THE FOLLOWING	MULT 390
C	INFORMATION..	MULT 400
C	ANS(1) INTERCEPT	MULT 410
C	ANS(2) MULTIPLE CORRELATION COEFFICIENT	MULT 420
C	ANS(3) STANDARD ERROR OF ESTIMATE	MULT 430
C	ANS(4) SUM OF SQUARES ATTRIBUTABLE TO REGRES-	MULT 440
C	SION (SSAR)	MULT 450
C	ANS(5) DEGREES OF FREEDOM ASSOCIATED WITH SSAR	MULT 460
C	ANS(6) MEAN SQUARE OF SSAR	MULT 470
C	ANS(7) SUM OF SQUARES OF DEVIATIONS FROM REGRES-	MULT 480
C	SION (SSDR)	MULT 490
C	ANS(8) DEGREES OF FREEDOM ASSOCIATED WITH SSDR	MULT 500
C	ANS(9) MEAN SQUARE OF SSDR	MULT 510
C	ANS(10) F-VALUE	MULT 520
C		MULT 530
C	REMARKS	MULT 540
C	N MUST BE GREATER THAN K+1.	MULT 550
C		MULT 560
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	MULT 570
C	NCNE	MULT 580
C		MULT 590
C	METHOD	MULT 600
C	THE GAUSS-JORDAN METHOD IS USED IN THE SOLUTION OF THE	MULT 610
C	NORMAL EQUATIONS. REFER TO W. W. COOLEY AND P. R. LOHNES,	MULT 620
C	'MULTIVARIATE PROCEDURES FOR THE BEHAVIORAL SCIENCES',	MULT 630
C	JOHN WILEY AND SONS, 1962, CHAPTER 3, AND B. OSTLE,	MULT 640
C	'STATISTICS IN RESEARCH', THE IOWA STATE COLLEGE PRESS,	MULT 650
C	1954, CHAPTER 8.	MULT 660
C		MULT 670
C	MULT 680
C		MULT 690
C	SUBROUTINE MULTR (N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)	MULT 700
C	DIMENSION XBAR(1),STD(1),D(1),RX(1),RY(1),ISAVE(1),B(1),SB(1),	MULT 710
C	1 T(1),ANS(1)	MULT 720
C		MULT 730
C	MULT 740
C		MULT 750
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	MULT 760
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	MULT 770
C	STATEMENT WHICH FOLLOWS.	MULT 780
C		MULT 790
C	DOUBLE PRECISION XBAR,STD,D,RX,RY,B,SB,T,ANS,RM,BO,SSAR,SSDR,SY,	MULT 800
C	1 FN,FK,SSARM,SSDRM,F	MULT 810
C		MULT 820
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	MULT 830
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	MULT 840
C	ROUTINE.	MULT 850
C		MULT 860
C	THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO	MULT 870

C		MULT 880
C	CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. SQRT AND ABS IN	MULT 890
C	STATEMENTS 122, 125, AND 135 MUST BE CHANGED TO DSQRT AND DABS.	MULT 900
C	MULT 910
C		MULT 920
C	MM=K+1	MULT 930
C		MULT 940
C	BETA WEIGHTS	MULT 950
C		MULT 960
C	DO 100 J=1,K	MULT 970
C	100 B(J)=0.0	MULT 980
C	DO 110 J=1,K	MULT 990
C	L1=K*(J-1)	MULT1000
C	DC 110 I=1,K	MULT1010
C	L=L1+I	MULT1020
C	110 E(J)=B(J)+RY(I)*RX(L)	MULT1030
C	RM=0.0	MULT1040
C	BO=0.0	MULT1050
C	L1=ISAVE(MM)	MULT1060
C		MULT1070
C	Coefficient of Determination	MULT1080
C		MULT1090
C	DO 120 I=1,K	MULT1100
C	RM=RM+B(I)*RY(I)	MULT1110
C		MULT1120
C	REGRESSION COEFFICIENTS	MULT1130
C		MULT1140
C	L=ISAVE(I)	MULT1150
C	B(I)=B(I)*(STD(L1)/STD(L))	MULT1160
C		MULT1170
C	INTERCEPT	MULT1180
C		MULT1190
C	120 BC=BC+B(I)*XBAR(L)	MULT1200
C	BC=XBAR(L1)-BC	MULT1210
C		MULT1220
C	SUM OF SQUARES ATTRIBUTABLE TO REGRESSION	MULT1230
C		MULT1240
C	SSAR=RM*D(L1)	MULT1250
C		MULT1260
C	MULTIPLE CORRELATION COEFFICIENT	MULT1270
C		MULT1280
C	122 RM= SQRT(ABS(RM))	MULT1290
C		MULT1300
C	SUM OF SQUARES OF DEVIATIONS FROM REGRESSION	MULT1310
C		MULT1320
C	SSDR=D(L1)-SSAR	MULT1330
C		MULT1340
C	VARIANCE OF ESTIMATE	MULT1350
C		MULT1360
C	FN=N-K-1	MULT1370
C	SY=SSDR/FN	MULT1380
C		MULT1390
C	STANDARD DEVIATIONS OF REGRESSION COEFFICIENTS	MULT1400
C		MULT1410
C	DC 130 J=1,K	MULT1420
C	L1=K*(J-1)+J	MULT1430
C	L=ISAVE(J)	MULT1440
C	125 SB(J)= SQRT(ABS((RX(L1)/D(L))*SY))	MULT1450
C		MULT1460
C	COMPUTED T-VALUES	MULT1470

C		MULT1480
C	130 T(J)=B(J)/SB(J)	MULT1490
C		MULT1500
C	STANDARD ERROR OF ESTIMATE	MULT1510
C		MULT1520
C	135 SY= SQRT(ABS(SY))	MULT1530
C		MULT1540
C	F VALUE	MULT1550
C		MULT1560
C	FK=K	MULT1570
C	SSARM=SSAR/FK	MULT1580
C	SSURM=SSDR/FN	MULT1590
C	F=SSARM/SSDRM	MULT1600
C		MULT1610
C	ANS(1)=BO	MULT1620
C	ANS(2)=RM	MULT1630
C	ANS(3)=SY	MULT1640
C	ANS(4)=SSAR	MULT1650
C	ANS(5)=FK	MULT1660
C	ANS(6)=SSARM	MULT1670
C	ANS(7)=SSDR	MULT1680
C	ANS(8)=FN	MULT1690
C	ANS(9)=SSDRM	MULT1700
C	ANS(10)=F	MULT1710
C	RETURN	MULT1720
C	END	MULT1730
C		NROO 10
C	NROO 20
C		NROO 30
C	SUBROUTINE NRCCCT	NROO 40
C		NROO 50
C	PURPOSE	NROO 60
C	COMPUTE EIGENVALUES AND EIGENVECTORS OF A REAL NONSYMMETRIC	NROO 70
C	MATRIX OF THE FORM B-INVERSE TIMES A. THIS SUBROUTINE IS	NROO 80
C	NORMALLY CALLED BY SUBROUTINE CANOR IN PERFORMING A	NROO 90
C	CANONICAL CORRELATION ANALYSIS.	NROO 100
C		NROO 110
C	USAGE	NROO 120
C	CALL NRCCCT (M,A,B,XL,X)	NROO 130
C		NROO 140
C	DESCRIPTION OF PARAMETERS	NROO 150
C	M - ORDER OF SQUARE MATRICES A, B, AND X.	NROO 160
C	A - INPUT MATRIX (M X M).	NROO 170
C	B - INPUT MATRIX (M X M).	NROO 180
C	XL - OUTPUT VECTOR OF LENGTH M CONTAINING EIGENVALUES OF	NROO 190
C	B-INVERSE TIMES A.	NROO 200
C	X - OUTPUT MATRIX (M X M) CONTAINING EIGENVECTORS COLUMN-	NROO 210
C	WISE.	NROO 220
C		NROO 230
C	REMARKS	NROO 240
C	NONE	NROO 250
C		NROO 260
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	NROO 270
C	EIGEN	NROO 280
C		NROO 290
C	METHOD	NROO 300
C	REFER TO W. W. COOLEY AND P. R. LOHNES, 'MULTIVARIATE PRO-	NROO 310
C	CEDURES FOR THE BEHAVIORAL SCIENCES', JOHN WILEY AND SONS,	NROO 320
C	1962, CHAPTER 3.	NROO 330
C		NROO 340

C	NR00	350
C		NR00	360
	SUBROUTINE NRCCT (M,A,B,XL,X)	NR00	370
	DIMENSION A(1),B(1),XL(1),X(1)	NR00	380
C		NR00	390
C	NR00	400
C		NR00	410
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	NR00	420
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	NR00	430
C	STATEMENT WHICH FOLLOWS.	NR00	440
C		NR00	450
C	DOUBLE PRECISION A,B,XL,X,SUMV	NR00	460
C		NR00	470
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	NR00	480
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	NR00	490
C	ROUTINE.	NR00	500
C		NR00	510
C	THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO	NR00	520
C	CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. SQRT IN STATEMENTS	NR00	530
C	110 AND 175 MUST BE CHANGED TO DSQRT. ABS IN STATEMENT 110	NR00	540
C	MUST BE CHANGED TO DABS.	NR00	550
C		NR00	560
C	NR00	570
C		NR00	580
C	CCOMPUTE EIGENVALUES AND EIGENVECTORS OF B	NR00	590
C		NR00	600
	K=1	NR00	610
	CC 100 J=2,M	NR00	620
	L=M*(J-1)	NR00	630
	DO 100 I=1,J	NR00	640
	L=L+1	NR00	650
	K=K+1	NR00	660
100	B(K)=B(L)	NR00	670
C		NR00	680
C	THE MATRIX B IS A REAL SYMMETRIC MATRIX.	NR00	690
C		NR00	700
	MV=0	NR00	710
	CALL EIGEN (B,X,M,MV)	NR00	720
C		NR00	730
C	FORM RECIPROCAL OF SQUARE ROOT OF EIGENVALUES. THE RESULTS	NR00	740
C	ARE PREMULTIPLIED BY THE ASSOCIATED EIGENVECTORS.	NR00	750
C		NR00	760
	L=0	NR00	770
	DO 110 J=1,M	NR00	780
	L=L+J	NR00	790
110	XL(J)=1.0/ SQRT(ABS(B(L)))	NR00	800
	K=0	NR00	810
	DO 115 J=1,M	NR00	820
	CC 115 I=1,M	NR00	830
	K=K+1	NR00	840
115	B(K)=X(K)*XL(J)	NR00	850
C		NR00	860
C	FORM (B**(-1/2))PRIME * A * (B**(-1/2))	NR00	870
C		NR00	880
	CC 120 I=1,M	NR00	890
	N2=0	NR00	900
	CC 120 J=1,M	NR00	910
	N1=M*(I-1)	NR00	920
	L=M*(J-1)+I	NR00	930
	X(L)=0.0	NR00	940

DO 120 K=1,M	NR00 950	
N1=N1+1	NR00 960	
N2=N2+1	NR00 970	
120 X(L)=X(L)+B(N1)*A(N2)	NR00 980	
L=L+1	NR00 990	
DO 130 J=1,M	NR001000	
DO 130 I=1,J	NR001010	
N1=I-M	NR001020	
N2=M*(J-1)	NR001030	
L=L+1	NR001040	
A(L)=0.0	NR001050	
DO 130 K=1,M	NR001060	
N1=N1+M	NR001070	
N2=N2+1	NR001080	
130 A(L)=A(L)+X(N1)*B(N2)	NR001090	
C	NR001100	
C	NR001110	
C	NR001120	
CALL EIGEN (A,X,M,MV)	NR001130	
L=L+1	NR001140	
DO 140 I=1,M	NR001150	
L=L+I	NR001160	
140 XL(I)=A(L)	NR001170	
C	NR001180	
C	NR001190	
C	NR001200	
DO 150 I=1,M	NR001210	
N2=0	NR001220	
DO 150 J=1,M	NR001230	
N1=I-M	NR001240	
L=M*(J-1)+I	NR001250	
A(L)=0.0	NR001260	
DO 150 K=1,M	NR001270	
N1=N1+M	NR001280	
N2=N2+1	NR001290	
150 A(L)=A(L)+B(N1)*X(N2)	NR001300	
L=L+1	NR001310	
K=0	NR001320	
DO 180 J=1,M	NR001330	
SUMV=0.0	NR001340	
DO 170 I=1,M	NR001350	
L=L+1	NR001360	
170 SUMV=SUMV+A(L)*A(L)	NR001370	
175 SUMV= SQRT(SUMV)	NR001380	
DO 180 I=1,M	NR001390	
K=K+1	NR001400	
180 X(K)=A(K)/SUMV	NR001410	
RETURN	NR001420	
END	NR001430	
C	ORDE 10	
CORDE 20	
C	ORDE 30	
C	SUBROUTINE ORDER	ORDE 40
C	ORDE 50	
C	PURPOSE	ORDE 60
C	CONSTRUCT FROM A LARGER MATRIX OF CORRELATION COEFFICIENTS	ORDE 70
C	A SUBSET MATRIX OF INTERCORRELATIONS AMONG INDEPENDENT	ORDE 80
C	VARIABLES AND A VECTOR OF INTERCORRELATIONS OF INDEPENDENT	ORDE 90
C	VARIABLES WITH DEPENDENT VARIABLE. THIS SUBROUTINE IS	ORDE 100
C	NORMALLY USED IN THE PERFORMANCE OF MULTIPLE AND POLYNOMIAL	ORDE 110

C	REGRESSION ANALYSES.	ORDE 120
C		ORDE 130
C	USAGE	ORDE 140
C	CALL ORDER (M,K,NDEP,K,ISAVE,RX,RY)	ORDE 150
C		ORDE 160
C	DESCRIPTION OF PARAMETERS	ORDE 170
C	M - NUMBER OF VARIABLES AND ORDER OF MATRIX R.	ORDE 180
C	R - INPUT MATRIX CONTAINING CORRELATION COEFFICIENTS.	ORDE 190
C	THIS SUBROUTINE EXPECTS ONLY UPPER TRIANGULAR	ORDE 200
C	PORTION OF THE SYMMETRIC MATRIX TO BE STORED (BY	ORDE 210
C	COLUMN) IN R. (STORAGE MODE OF 1)	ORDE 220
C	NDEP - THE SUBSCRIPT NUMBER OF THE DEPENDENT VARIABLE.	ORDE 230
C	K - NUMBER OF INDEPENDENT VARIABLES TO BE INCLUDED	ORDE 240
C	IN THE FORTHCOMING REGRESSION. K MUST BE GREATER	ORDE 250
C	THAN OR EQUAL TO 1.	ORDE 251
C	ISAVE - INPUT VECTOR OF LENGTH K+1 CONTAINING, IN ASCENDING	ORDE 260
C	ORDER, THE SUBSCRIPT NUMBERS OF K INDEPENDENT	ORDE 270
C	VARIABLES TO BE INCLUDED IN THE FORTHCOMING REGRES-	ORDE 280
C	SIGN.	ORDE 290
C	UPON RETURNING TO THE CALLING ROUTINE, THIS VECTOR	ORDE 300
C	CONTAINS, IN ADDITION, THE SUBSCRIPT NUMBER OF	ORDE 310
C	THE DEPENDENT VARIABLE IN K+1 POSITION.	ORDE 320
C	RX - OUTPUT MATRIX (K X K) CONTAINING INTERCORRELATIONS	ORDE 330
C	AMONG INDEPENDENT VARIABLES TO BE USED IN FORTH-	ORDE 340
C	COMING REGRESSION.	ORDE 350
C	RY - OUTPUT VECTOR OF LENGTH K CONTAINING INTERCORRELA-	ORDE 360
C	TIONS OF INDEPENDENT VARIABLES WITH DEPENDENT	ORDE 370
C	VARIABLES.	ORDE 380
C		ORDE 390
C	REMARKS	ORDE 400
C	NONE	ORDE 410
C		ORDE 420
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	ORDE 430
C	NONE	ORDE 440
C		ORDE 450
C	METHOD	ORDE 460
C	FROM THE SUBSCRIPT NUMBERS OF THE VARIABLES TO BE INCLUDED	ORDE 470
C	IN THE FORTHCOMING REGRESSION, THE SUBROUTINE CONSTRUCTS THE	ORDE 480
C	MATRIX RX AND THE VECTOR RY.	ORDE 490
C		ORDE 500
C	ORDE 510
C		ORDE 520
C	SUBROUTINE ORDER (M,R,NDEP,K,ISAVE,RX,RY)	ORDE 530
C	DIMENSION R(1),ISAVE(1),RX(1),RY(1)	ORDE 540
C		ORDE 550
C	ORDE 560
C		ORDE 570
C	IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE	ORDE 580
C	C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION	ORDE 590
C	STATEMENT WHICH FOLLOWS.	ORDE 600
C		ORDE 610
C	DOUBLE PRECISION R,RX,RY	ORDE 620
C		ORDE 630
C	THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS	ORDE 640
C	APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS	ORDE 650
C	ROUTINE.	ORDE 660
C		ORDE 670
C	ORDE 680
C		ORDE 690
C	COPY INTERCORRELATIONS OF INDEPENDENT VARIABLES	ORDE 700

C	WITH DEPENDENT VARIABLE	ORDE 710
C		ORDE 720
	MM=0	ORDE 730
	CC 130 J=1,K	ORDE 740
	L2=ISAVE(J)	ORDE 750
	IF(NDEP-L2) 122, 123, 123	ORDE 760
122	L=NDEP+(L2*L2-L2)/2	ORDE 770
	GC TC 125	ORDE 780
123	L=L2+(NDEP*NDEP-NDEP)/2	ORDE 790
125	RY(J)=R(L)	ORDE 800
C		ORDE 810
C	COPY A SUBSET MATRIX OF INTERCORRELATIONS AMONG	ORDE 820
C	INDEPENDENT VARIABLES	ORDE 830
C		ORDE 840
	DU 130 I=1,K	ORDE 850
	L1=ISAVE(I)	ORDE 860
	IF(L1-L2) 127, 128, 128	ORDE 870
127	L=L1+(L2*L2-L2)/2	ORDE 880
	GC TC 129	ORDE 890
128	L=L2+(L1*L1-L1)/2	ORDE 900
129	MM=MM+1	ORDE 910
130	RX(MM)=R(L)	ORDE 920
C		ORDE 930
C	PLACE THE SUBSCRIPT NUMBER OF THE DEPENDENT	ORDE 940
C	VARIABLE IN ISAVE(K+1)	ORDE 950
C		ORDE 960
	ISAVE(K+1)=NDEP	ORDE 970
	RETURN	ORDE 980
	END	ORDE 990
C		STOU 10
C	STOU 20
C		STOU 30
C	SAMPLE OUTPUT SUBROUTINE STOUT	STOU 40
C		STOU 50
C	PURPOSE	STOU 60
C	PRINT THE RESULT OF A STEP-WISE MULTIPLE REGRESSION. THIS	STOU 70
C	SUBROUTINE IS CALLED BY THE SUBROUTINE STPRG.	STOU 80
C		STOU 90
C	USAGE	STOU 100
C	CALL STOUT (NSTEP,ANS,L,B,S,T,NSTOP)	STOU 110
C		STOU 120
C	DESCRIPTION OF PARAMETERS	STOU 130
C	NSTEP - INPUT VECTOR OF LENGTH 5 CONTAINING THE FOLLOWING	STOU 140
C	INFORMATION..	STOU 150
C	NSTEP(1) DEPENDENT VARIABLE	STOU 160
C	NSTEP(2) NUMBER OF VARIABLES FORCED TO ENTER	STOU 170
C	IN THE REGRESSION	STOU 180
C	NSTEP(3) NUMBER OF VARIABLES DELETED	STOU 190
C	NSTEP(4) THE LAST STEP NUMBER	STOU 200
C	NSTEP(5) THE LAST VARIABLE ENTERED	STOU 210
C	ANS - INPUT VECTOR OF LENGTH 11 CONTAINING THE FOLLOWING	STOU 220
C	INFORMATION FOR THE LAST STEP..	STOU 230
C	ANS(1) SUM OF SQUARES REDUCED	STOU 240
C	ANS(2) PROPORTION REDUCED	STOU 250
C	ANS(3) CUMULATIVE SUM OF SQUARES REDUCED	STOU 260
C	ANS(4) CUMULATIVE PROPORTION REDUCED	STOU 270
C	ANS(5) SUM OF SQUARES OF THE DEPENDENT VARIABLE	STOU 280
C	ANS(6) MULTIPLE CORRELATION COEFFICIENT	STOU 290
C	ANS(7) F-VALUE FOR ANALYSIS VARIANCE (FOR THE	STOU 300
C	REGRESSION)	STOU 310

```

C          ANS(8) STANDARD ERROR OF ESTIMATE          STOU 320
C          ANS(9) INTERCEPT                          STOU 330
C          ANS(10) ADJUSTED MULTIPLE R                 STOU 340
C          ANS(11) ADJUSTED STANDARD ERROR OF ESTIMATE STOU 350
C      L      - INPUT VECTOR OF LENGTH K (K=M-NSTEP(3)-1) CONTAIN- STOU 360
C              ING VARIABLES ENTERED IN THE REGRESSION. L(1)=FIRSTSTOU 370
C              VARIABLE ENTERED, L(2)=SECOND VARIABLE ENTERED, ETC. STOU 380
C      B      - INPUT VECTOR OF LENGTH K (K=M-NSTEP(3)-1) CONTAIN- STOU 390
C              ING REGRESSION COEFFICIENTS CORRESPONDING TO THE STOU 400
C              VARIABLES IN VECTOR L                   STOU 410
C      S      - INPUT VECTOR OF LENGTH K (K=M-NSTEP(3)-1) CONTAIN- STOU 420
C              ING STANDARD ERRORS OF REGRESSION COEFFICIENTS STOU 430
C              CORRESPONDING TO THE VARIABLES IN VECTOR L STOU 440
C      T      - INPUT VECTOR OF LENGTH K (K=M-NSTEP(3)-1) CONTAIN- STOU 450
C              ING COMPUTED T-VALUES CORRESPONDING TO THE VARIABLESSTOU 460
C              IN VECTOR L                              STOU 470
C      NSTOP - OUTPUT OPTION CODE TO STOP THE STEP-WISE REGRESSION STOU 480
C              1 - IF THE STEP-WISE REGRESSION IS TO BE TERMI- STOU 490
C                  NATED BY SOME CRITERIA OTHER THAN PROPORTION STOU 500
C                  OF SUM OF SQUARES, SUCH AS F-TEST AND SO ON, STOU 510
C                  THIS SUBROUTINE MAY BE MODIFIED TO PERFORM STOU 520
C                  DESIRED TESTS. WHEN IT BECOMES NO LONGER STOU 530
C                  NECESSARY TO CONTINUE THE STEP-WISE REGRES- STOU 540
C                  SION, SET NSTOP EQUAL TO 1.           STOU 550
C              0 - IF THE STEP-WISE REGRESSION IS TO BE CONTINUEDSTOU 560
C                  STOU 570
C      REMARKS                                          STOU 580
C          THE CONTENTS OF THE VECTORS NSTEP, ANS, L ARE REQUIRED IN STOU 590
C          SUBSEQUENT STEPS AND MUST NOT BE DESTROYED. STOU 600
C                  STOU 610
C      SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED STOU 620
C          NONE                                          STOU 630
C                  STOU 640
C                  STOU 650
C      .....STOU 660
C                  STOU 670
C      SUBROUTINE STOUT (NSTEP,ANS,L,B,S,T,NSTOP) STOU 680
C                  STOU 690
C      DIMENSION NSTEP(1),ANS(1),L(1),B(1),S(1),T(1) STOU 700
C                  STOU 710
C      .....STOU 720
C                  STOU 730
C          IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE STOU 740
C          C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION STOU 750
C          STATEMENT WHICH FOLLOWS. STOU 760
C                  STOU 770
C      DOUBLE PRECISION ANS,B,S,T STOU 780
C                  STOU 790
C          THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS STOU 800
C          APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS STOU 810
C          ROUTINE. STOU 820
C                  STOU 830
C      .....STOU 840
C                  STOU 850
C      1 FORMAT(/5H1STEP13) STOU 860
C      2 FORMAT(22H0VARIABLE ENTERED.....I2) STOU 870
C      3 FORMAT(40H0SUM OF SQUARES REDUCED IN THIS STEP.....F13.3) STOU 880
C      4 FORMAT(40H0PROPORTION REDUCED IN THIS STEP.....F13.3) STOU 890
C      5 FORMAT(40H0CUMMLLATIVE SUM OF SQUARES REDUCED.....F13.3) STOU 900
C      6 FORMAT(40H0CUMMLATIVE PROPORTION REDUCED.....F13.3,4H OFF13. STOU 910

```

	13)		STOU 920
	7	FORMAT(4H0FOR13,18H VARIABLES ENTERED)	STOU 930
	8	FORMAT(38H MULTIPLE CORRELATION COEFFICIENT...F9.3)	STOU 940
	9	FORMAT(38H F-VALUE FOR ANALYSIS OF VARIANCE...F9.3)	STOU 950
	10	FORMAT(38H STANDARD ERROR OF ESTIMATE.....F9.3)	STOU 960
	11	FORMAT(757H VARIABLE REGRESSION STD. ERROR OF COMPUT	STOU 970
		1ED/56H NUMBER COEFFICIENT REG. COEFF. T-VALUE)	STOU 980
	12	FORMAT(5X,I3,F16.5,F16.5,F14.3)	STOU 990
	13	FORMAT(12H INTERCEPTF14.5)	STOU1000
	14	FORMAT(31H0DEPENDENT VARIABLE.....I2)	STOU1010
	15	FORMAT(31H NUMBER OF VARIABLES FORCED....I2)	STOU1020
	16	FORMAT(31H NUMBER OF VARIABLES DELETED...I2)	STOU1030
	17	FORMAT(20H (FORCED VARIABLE))	STOU1040
	18	FORMAT(38H (ADJUSTED FOR D.F.).....F9.3)	STOU1050
C			STOU1060
C		TEST WHETHER THIS IS THE FIRST STEP	STOU1070
C			STOU1080
		IF(NSTEP(4)-1) 30, 30, 35	STOU1090
	30	WRITE (6,14) NSTEP(1)	STOU1100
		WRITE (6,15) NSTEP(2)	STOU1110
		WRITE (6,16) NSTEP(3)	STOU1120
C			STOU1130
C		PRINT THE RESULT OF A STEP	STOU1140
C			STOU1150
	35	WRITE (6,1) NSTEP(4)	STOU1160
		WRITE (6,2) NSTEP(5)	STOU1170
		IF(NSTEP(4)-NSTEP(2)) 37, 37, 38	STOU1180
	37	WRITE (6,17)	STOU1190
	38	WRITE (6,3) ANS(1)	STOU1200
		WRITE (6,4) ANS(2)	STOU1210
		WRITE (6,5) ANS(3)	STOU1220
		WRITE (6,6) ANS(4), ANS(5)	STOU1230
		WRITE (6,7) NSTEP(4)	STOU1240
		WRITE (6,8) ANS(6)	STOU1250
		WRITE(6,18)ANS(10)	STOU1260
		WRITE (6,9) ANS(7)	STOU1270
		WRITE (6,10) ANS(8)	STOU1280
		WRITE(6,18)ANS(11)	STOU1290
		WRITE (6,11)	STOU1300
		N=NSTEP(4)	STOU1310
		DC 40 I=1,N	STOU1320
	40	WRITE (6,12) L(I),B(I),S(I),T(I)	STOU1330
		WRITE (6,13) ANS(9)	STOU1340
C			STOU1350
		NSTEP=0	STOU1360
		RETURN	STOU1370
		END	STOU1380
C			STPR 10
C		STPR 20
C			STPR 30
C		SUBROUTINE STPRG	STPR 40
C			STPR 50
C		PURPOSE	STPR 60
C		TO PERFORM A STEPWISE MULTIPLE REGRESSION ANALYSIS FOR A	STPR 70
C		DEPENDENT VARIABLE AND A SET OF INDEPENDENT VARIABLES. AT	STPR 80
C		EACH STEP, THE VARIABLE ENTERED INTO THE REGRESSION EQUATION	STPR 90
C		IS THAT WHICH EXPLAINS THE GREATEST AMOUNT OF VARIANCE	STPR 100
C		BETWEEN IT AND THE DEPENDENT VARIABLE (I.E. THE VARIABLE	STPR 110
C		WITH THE HIGHEST PARTIAL CORRELATION WITH THE DEPENDENT	STPR 120
C		VARIABLE). ANY VARIABLE CAN BE DESIGNATED AS THE DEPENDENT	STPR 130

VARIABLE. ANY INDEPENDENT VARIABLE CAN BE FORCED INTO OR DELETED FROM THE REGRESSION EQUATION, IRRESPECTIVE OF ITS CONTRIBUTION TO THE EQUATION.

STPR 140
STPR 150
STPR 160
STPR 170
STPR 180
STPR 190
STPR 200
STPR 210
STPR 220
STPR 230
STPR 240
STPR 250
STPR 260
STPK 270
STPR 280
STPR 290
STPR 300
STPR 310
STPR 320
STPR 330
STPR 340
STPR 350
STPR 360
STPR 370
STPR 380
STPR 390
STPR 400
STPR 410
STPR 420
STPR 430
STPR 440
STPR 450
STPR 460
STPR 470
STPR 480
STPR 490
STPR 500
STPR 510
STPR 520
STPR 530
STPR 540
STPR 550
STPR 560
STPR 570
STPR 580
STPR 590
STPR 600
STPR 610
STPR 620
STPK 630
STPR 640
STPR 650
STPR 660
STPR 670
STPR 680
STPR 690
STPR 700
STPR 710
STPR 720
STPR 730

USAGE

CALL STPRG (M,N,D,XBAR,IDX,PCT,NSTEP,ANS,L,B,S,T,LL,IER)

DESCRIPTION OF PARAMETERS

- M - TOTAL NUMBER OF VARIABLES IN DATA MATRIX
- N - NUMBER OF OBSERVATIONS
- D - INPUT MATRIX (M X M) OF SUMS OF CROSS-PRODUCTS OF DEVIATIONS FROM MEAN. THIS MATRIX WILL BE DESTROYED.
- XBAR - INPUT VECTOR OF LENGTH M OF MEANS
- IDX - INPUT VECTOR OF LENGTH M HAVING ONE OF THE FOLLOWING CODES FOR EACH VARIABLE.
 - 0 - INDEPENDENT VARIABLE AVAILABLE FOR SELECTION
 - 1 - INDEPENDENT VARIABLE TO BE FORCED INTO THE REGRESSION EQUATION
 - 2 - VARIABLE NOT TO BE CONSIDERED IN THE EQUATION
 - 3 - DEPENDENT VARIABLETHIS VECTOR WILL BE DESTROYED
- PCT - A CONSTANT VALUE INDICATING THE PROPORTION OF THE TOTAL VARIANCE TO BE EXPLAINED BY ANY INDEPENDENT VARIABLE. THOSE INDEPENDENT VARIABLES WHICH FALL BELOW THIS PROPORTION WILL NOT ENTER THE REGRESSION EQUATION. TO ENSURE THAT ALL VARIABLES ENTER THE EQUATION, SET PCT = 0.0.
- NSTEP- OUTPUT VECTOR OF LENGTH 5 CONTAINING THE FOLLOWING INFORMATION
 - NSTEP(1)- THE NUMBER OF THE DEPENDENT VARIABLE
 - NSTEP(2)- NUMBER OF VARIABLES FORCED INTO THE REGRESSION EQUATION
 - NSTEP(3)- NUMBER OF VARIABLE DELETED FROM THE EQUATION
 - NSTEP(4)- THE NUMBER OF THE LAST STEP
 - NSTEP(5)- THE NUMBER OF THE LAST VARIABLE ENTERED
- ANS - OUTPUT VECTOR OF LENGTH 11 CONTAINING THE FOLLOWING INFORMATION FOR THE LAST STEP
 - ANS(1)- SUM OF SQUARES REDUCED BY THIS STEP
 - ANS(2)- PROPORTION OF TOTAL SUM OF SQUARES REDUCED
 - ANS(3)- CUMULATIVE SUM OF SQUARES REDUCED UP TO THIS STEP
 - ANS(4)- CUMULATIVE PROPORTION OF TOTAL SUM OF SQUARES REDUCED
 - ANS(5)- SUM OF SQUARES OF THE DEPENDENT VARIABLE
 - ANS(6)- MULTIPLE CORRELATION COEFFICIENT
 - ANS(7)- F RATIO FOR SUM OF SQUARES DUE TO REGRESSION
 - ANS(8)- STANDARD ERROR OF THE ESTIMATE (RESIDUAL MEAN SQUARE)
 - ANS(9)- INTERCEPT CONSTANT
 - ANS(10)-MULTIPLE CORRELATION COEFFICIENT ADJUSTED FOR DEGREES OF FREEDOM.
 - ANS(11)-STANDARD ERROR OF THE ESTIMATE ADJUSTED FOR DEGREES OF FREEDOM.
- L - OUTPUT VECTOR OF LENGTH K, WHERE K IS THE NUMBER OF INDEPENDENT VARIABLES IN THE REGRESSION EQUATION. THIS VECTOR CONTAINS THE NUMBERS OF THE INDEPENDENT VARIABLES IN THE EQUATION.
- B - OUTPUT VECTOR OF LENGTH K, CONTAINING THE PARTIAL

```

C          REGRESSION COEFFICIENTS CORRESPONDING TO THE          STPR 740
C          VARIABLES IN VECTOR L.                                STPR 750
C          S - OUTPUT VECTOR OF LENGTH K, CONTAINING THE STANDARD STPR 760
C          ERRORS OF THE PARTIAL REGRESSION COEFFICIENTS,      STPR 770
C          CORRESPONDING TO THE VARIABLES IN VECTOR L.         STPR 780
C          T - OUTPUT VECTOR OF LENGTH K, CONTAINING THE COMPUTED STPR 790
C          T-VALUES CORRESPONDING TO THE VARIABLES IN VECTOR L. STPR 800
C          LL - WORKING VECTOR OF LENGTH M                      STPR 810
C          IER - 0, IF THERE IS NO ERROR.                       STPR 820
C          1, IF RESIDUAL SUM OF SQUARES IS NEGATIVE OR IF THE STPR 830
C          PIVOTAL ELEMENT IN THE STEPWISE INVERSION PROCESS IS STPR 840
C          ZERO. IN THIS CASE, THE VARIABLE WHICH CAUSES THIS STPR 850
C          ERROR IS NOT ENTERED IN THE REGRESSION, THE RESULT STPR 860
C          PRIOR TO THIS STEP IS RETAINED, AND THE CURRENT     STPR 870
C          SELECTION IS TERMINATED.                             STPR 880
C                                                                STPR 890
C                                                                STPR 900
REMARKS                                                         STPR 910
C          THE NUMBER OF DATA POINTS MUST BE AT LEAST GREATER STPR 920
C          NUMBER OF INDEPENDENT VARIABLES PLUS ONE. FORCED VARIABLES STPR 930
C          ARE ENTERED INTO THE REGRESSION EQUATION BEFORE ALL OTHER STPR 940
C          INDEPENDENT VARIABLES. WITHIN THE SET OF FORCED VARIABLES, STPR 950
C          THE ONE TO BE CHOSEN FIRST WILL BE THAT ONE WHICH EXPLAINS STPR 960
C          THE GREATEST AMOUNT OF VARIANCE.                     STPR 970
C          INSTEAD OF USING, AS A STOPPING CRITERION, A PROPORTION OF STPR 980
C          THE TOTAL VARIANCE, SOME OTHER CRITERION MAY BE ADDED TO STPR 990
C          SUBROUTINE STOUT.                                     STPR1000
C                                                                STPR1010
SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED                     STPR1020
C          STOUT(NSTEP,ANS,L,B,S,T,NSTOP)                       STPR1030
C          THIS SUBROUTINE MUST BE PROVIDED BY THE USER. IT IS AN STPR1040
C          OUTPUT ROUTINE WHICH WILL PRINT THE RESULTS OF EACH STEP OF STPR1050
C          THE REGRESSION ANALYSIS. NSTOP IS AN OPTION CODE WHICH IS STPR1060
C          ONE IF THE STEPWISE REGRESSION IS TO BE TERMINATED, AND IS STPR1070
C          ZERO IF IT IS TO CONTINUE. THE USER MUST CONSIDER THIS IF STPR1080
C          SOME OTHER STOPPING CRITERION THAN VARIANCE PROPORTION IS TO STPR1090
C          BE USED.                                             STPR1100
C                                                                STPR1110
METHOD                                                           STPR1120
C          THE ABBREVIATED COOLITTLE METHOD IS USED TO (1) DECIDE VARI- STPR1130
C          ABLES ENTERING IN THE REGRESSION AND (2) COMPUTE REGRESSION STPR1140
C          COEFFICIENTS. REFER TO C. A. BENNETT AND N. L. FRANKLIN, STPR1150
C          'STATISTICAL ANALYSIS IN CHEMISTRY AND THE CHEMICAL INDUS- STPR1160
C          TRY', JOHN WILEY AND SONS, 1954, APPENDIX 6A.       STPR1170
C                                                                STPR1180
C          ..... STPR1190
C          SUBROUTINE STPRG (M,N,C,XBAR,IDX,PCT,NSTEP,ANS,L,B,S,T,LL,IER) STPR1200
C                                                                STPR1210
C          DIMENSION D(1),XBAR(1),IDX(1),NSTEP(1),ANS(1),L(1),B(1),S(1),T(1), STPR1220
C          ILL(1)                                               STPR1230
C                                                                STPR1240
C          ..... STPR1250
C                                                                STPR1260
C          IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE STPR1270
C          C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION STPR1280
C          STATEMENT WHICH FOLLOWS.                             STPR1290
C                                                                STPR1300
C          DOUBLE PRECISION C,XBAR,ANS,B,S,T,RD,RE             STPR1310
C                                                                STPR1320
C          THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS STPR1330

```

NEW=K	STPR1940
50 CONTINUE	STPR1950
GO TO 75	STPR1960
C	STPR1970
C	STPR1980
C	STPR1990
55 DL 70 I=1,M	STPR2000
IF(I-MY) 60, 70, 60	STPR2010
60 IF(LL(I)) 70, 70, 62	STPR2020
62 LYP=LY+I	STPR2030
IP=M*(I-1)+I	STPR2040
RE=D(LYP)*D(LYP)/D(IP)	STPR2050
IF(RD-RE) 64, 70, 70	STPR2060
64 RD=RE	STPR2070
NEW=I	STPR2080
70 CONTINUE	STPR2090
C	STPR2100
C	STPR2110
C	STPR2120
C	STPR2130
C	STPR2140
75 IF(RD) 77,77,76	STPR2150
76 IF(ANS(5)-(ANS(3)+RD))77,77,78	STPR2160
77 IER=1	STPR2170
GO TO 150	STPR2180
78 RE=RD/ANS(5)	STPR2190
IF(RE-PCT) 150, 80, 80	STPR2200
C	STPR2210
C	STPR2220
C	STPR2230
80 LL(NEW)=0	STPR2240
L(NL)=NEW	STPR2250
ANS(1)=RD	STPR2260
ANS(2)=RE	STPR2270
ANS(3)=ANS(3)+RD	STPR2280
ANS(4)=ANS(4)+RE	STPR2290
NSTEP(4)=NL	STPR2300
NSTEP(5)=NEW	STPR2310
C	STPR2320
C	STPR2330
C	STPR2340
C	STPR2350
85 ANS(6)= SQRT(ANS(4))	STPR2360
RD=NL	STPR2370
RE=CNM-RD	STPR2380
RE=(ANS(5)-ANS(3))/RE	STPR2390
ANS(7)=(ANS(3)/RD)/RE	STPR2400
90 ANS(8)= SQRT(RE)	STPR2410
C	STPR2420
C	STPR2430
C	STPR2440
IF=M*(NEW-1)+NEW	STPR2450
RD=D(IP)	STPR2460
LYP=NEW-M	STPR2470
DO 100 J=1,M	STPR2480
LYP=LYP+M	STPR2490
IF(LL(J)) 100, 94, 97	STPR2500
94 IF(J-NEW) 96, 98, 96	STPR2510
96 IJ=M*(J-1)+J	STPR2520
L(IJ)=D(IJ)+D(LYP)*D(LYP)/RD	STPR2530

57	D(LYP)=D(LYP)/RD	STPR2540
	GO TO 100	STPR2550
98	L(IP)=1.0/RD	STPR2560
100	CONTINUE	STPR2570
C		STPR2580
C	CCOMPUTE REGRESSION COEFFICIENTS	STPR2590
C		STPR2600
	LYP=LY+NEW	STPR2610
	B(NL)=D(LYP)	STPR2620
	IF(NL-1) 112, 112, 105	STPR2630
105	ID=NL-1	STPR2640
	DO 110 J=1, ID	STPR2650
	IJ=NL-J	STPR2660
	KK=L(IJ)	STPR2670
	LYF=LY+KK	STPR2680
	B(IJ)=D(LYP)	STPR2690
	DO 110 K=1, J	STPR2700
	IK=NL-K+1	STPR2710
	MK=L(IK)	STPR2720
	LYF=M*(MK-1)+KK	STPR2730
110	B(IJ)=B(IJ)-D(LYP)*B(IK)	STPR2740
C		STPR2750
C	CCOMPUTE INTERCEPT	STPR2760
C		STPR2770
112	ANS(9)=XBAR(MY)	STPR2780
	DO 115 I=1, NL	STPR2790
	KK=L(I)	STPR2800
	ANS(9)=ANS(9)-B(I)*XBAR(KK)	STPR2810
	IJ=M*(KK-1)+KK	STPR2820
114	S(I)=ANS(8)*SQRT(D(IJ))	STPR2830
115	T(I)=B(I)/S(I)	STPR2840
C		STPR2850
C	PERFORM A REDUCTION TO ELIMINATE THE LAST VARIABLE ENTERED	STPR2860
C		STPR2870
	IF=M*(NEW-1)	STPR2880
	DO 130 I=1, M	STPR2890
	IJ=I-M	STPR2900
	IK=NEW-M	STPR2910
	IP=IP+1	STPR2920
	IF(LL(I)) 130, 130, 120	STPR2930
120	DO 126 J=1, M	STPR2940
	IJ=IJ+M	STPR2950
	IK=IK+M	STPR2960
	IF(LL(J)) 126, 122, 122	STPR2970
122	IF(J-NEW) 124, 126, 124	STPR2980
124	D(IJ)=D(IJ)-D(IP)*D(IK)	STPR2990
126	CONTINUE	STPR3000
	L(IP)=D(IP)/(-RD)	STPR3010
130	CONTINUE	STPR3020
C		STPR3030
C	ADJUST STANARD ERROR OF THE ESTIMATE AND MULTIPLE CORRELATION	STPR3040
C	CCEFFICIENT	STPR3050
C		STPR3060
	RD=N-NSTEP(4)	STPR3070
	RD=CNM/RD	STPR3080
132	ANS(10)=SQRT(1.0-(1.0-ANS(6)*ANS(6))*RD)	STPR3090
134	ANS(11)=ANS(8)*SQRT(RD)	STPR3100
C		STPR3110
C	CALL THE OUTPUT SUBROUTINE	STPR3120
	CALL STOUT (NSTEP,ANS,L,B,S,T,NSTOP)	STPR3130

C		STPR3140
C	TEST WHETHER THE STEP-WISE REGRESSION WAS TERMINATED IN	STPR3150
C	SUBROUTINE STCLT	STPR3160
C		STPR3170
C	IF(NSTOP) 140, 140, 150	STPR3180
C		STPR3190
C	140 CONTINUE	STPR3200
C		STPR3210
C	150 RETURN	STPR3220
C	END	STPR3230
C		PLOT 10
C	PLOT 20
C		PLOT 30
C	SUBROUTINE FLCT	PLOT 40
C		PLOT 50
C	PURPOSE	PLOT 60
C	PLOT SEVERAL CROSS-VARIABLES VERSUS A BASE VARIABLE	PLOT 70
C		PLOT 80
C	USAGE	PLOT 90
C	CALL PLOT (NU,A,N,M,NL,NS)	PLOT 100
C		PLOT 110
C	DESCRIPTION OF PARAMETERS	PLOT 120
C	NU - CHART NUMBER (3 DIGITS MAXIMUM)	PLOT 130
C	A - MATRIX OF DATA TO BE PLOTTED. FIRST COLUMN REPRESENTS	PLOT 140
C	BASE VARIABLE AND SUCCESSIVE COLUMNS ARE THE CROSS-	PLOT 150
C	VARIABLES (MAXIMUM IS 9).	PLOT 160
C	N - NUMBER OF ROWS IN MATRIX A	PLOT 170
C	M - NUMBER OF COLUMNS IN MATRIX A (EQUAL TO THE TOTAL	PLOT 180
C	NUMBER OF VARIABLES). MAXIMUM IS 10.	PLOT 190
C	NL - NUMBER OF LINES IN THE PLOT. IF 0 IS SPECIFIED, 50	PLOT 200
C	LINES ARE USED.	PLOT 210
C	NS - CODE FOR SORTING THE BASE VARIABLE DATA IN ASCENDING	PLOT 220
C	ORDER	PLOT 230
C	0 SORTING IS NOT NECESSARY (ALREADY IN ASCENDING	PLOT 240
C	ORDER).	PLOT 250
C	1 SORTING IS NECESSARY.	PLOT 260
C		PLOT 270
C	REMARKS	PLOT 280
C	NONE	PLOT 290
C		PLOT 300
C	SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED	PLOT 310
C	NONE	PLOT 320
C		PLOT 330
C	PLOT 340
C		PLOT 350
C	SUBROUTINE PLCT(NU,A,N,M,NL,NS)	PLOT 360
C	DIMENSION OUT(101),YPR(11),ANG(9),A(1)	PLOT 370
C		PLOT 380
C	1 FORMAT(1H1,60X,7H CHART ,I3,/))	PLOT 390
C	2 FCRMAT(1H ,F11.4,5X,101A1)	PLOT 400
C	3 FCRMAT(1H)	PLOT 410
C	4 FORMAT(10H 123456789)	PLOT 420
C	5 FCRMAT(10A1)	PLOT 430
C	7 FORMAT(1H ,16X,101H.)	PLOT 440
C	1)	PLOT 450
C	8 FCRMAT(1H0,9X,11F10.4)	PLOT 460
C		PLOT 470
C	PLOT 480
C		PLOT 490
C	NLL=NL	PLOT 500

C		PLOT 510
	IF(NS) 16, 16, 10	PLOT 520
C		PLOT 530
C	SORT BASE VARIABLE DATA IN ASCENDING ORDER	PLOT 540
C		PLOT 550
10	DO 15 I=1,N	PLOT 560
	CC 14 J=I,N	PLOT 570
	IF(A(I)-A(J)) 14, 14, 11	PLOT 580
11	L=I-N	PLOT 590
	LL=J-N	PLOT 600
	DO 12 K=1,M	PLOT 610
	L=L+N	PLOT 620
	LL=LL+N	PLOT 630
	F=A(L)	PLOT 640
	A(L)=A(LL)	PLOT 650
12	A(LL)=F	PLOT 660
14	CONTINUE	PLOT 670
15	CONTINUE	PLOT 680
C		PLOT 690
C	TEST NLL	PLOT 700
C		PLOT 710
16	IF(NLL) 20, 18, 20	PLOT 720
18	NLL=50	PLOT 730
C		PLOT 740
C	PRINT TITLE	PLOT 750
C		PLOT 760
20	WRITE(6,1)NO	PLOT 770
C		PLOT 780
C	DEVELOP BLANK AND DIGITS FOR PRINTING	PLOT 790
C		PLOT 800
	REWIND 13	PLOT 810
	WRITE (13,4)	PLOT 820
	REWIND 13	PLOT 830
	READ (13,5) BLANK,(ANG(I),I=1,9)	PLOT 840
	REWIND 13	PLOT 850
C		PLOT 860
C	FIND SCALE FOR BASE VARIABLE	PLOT 870
C		PLOT 880
	XSCAL=(A(N)-A(1))/(FLOAT(NLL-1))	PLOT 890
C		PLOT 900
C	FIND SCALE FOR CROSS-VARIABLES	PLOT 910
C		PLOT 920
	M1=N+1	PLOT 930
	YMIN=A(M1)	PLOT 940
	YMAX=YMIN	PLOT 950
	M2=M*N	PLOT 960
	DO 40 J=M1,M2	PLOT 970
	IF(A(J)-YMIN) 28,26,26	PLOT 980
26	IF(A(J)-YMAX) 40,40,30	PLOT 990
28	YMIN=A(J)	PLOT1000
	GC TC 40	PLOT1010
30	YMAX=A(J)	PLOT1020
40	CONTINUE	PLOT1030
	YSCAL=(YMAX-YMIN)/100.0	PLOT1040
C		PLOT1050
C	FIND BASE VARIABLE PRINT POSITION	PLOT1060
C		PLOT1070
	XB=A(1)	PLOT1080
	L=1	PLOT1090
	MY=M-1	PLOT1100

```

I=1
45 F=I-1
XPR=XB+F*XSCAL
IF(A(L)-XPR) 50,50,70
C
C FIND CROSS-VARIABLES
C
50 DO 55 IX=1,101
55 CUT(IX)=BLANK
DO 60 J=1,MY
LL=L+J*N
JP=((A(LL)-YMIN)/YSCAL)+1.0
CUT(JP)=ANG(J)
60 CONTINUE
C
C PRINT LINE AND CLEAR, OR SKIP
C
WRITE(6,2)XPR,(OUT(IZ),IZ=1,101)
L=L+1
GO TO 80
70 WRITE(6,3)
80 I=I+1
IF(I-NLL) 45, 84, 86
84 XPR=A(N)
GO TO 50
C
C PRINT CROSS-VARIABLES NUMBERS
C
86 WRITE(6,7)
YPR(1)=YMIN
DO 90 KN=1,9
90 YPR(KN+1)=YPR(KN)+YSCAL*10.0
YPR(11)=YMAX
WRITE(6,8)(YPR(IP),IP=1,11)
RETURN
END
C
C .....
C
C SUBROUTINE GDATA
C
C PURPOSE
C GENERATE INDEPENDENT VARIABLES UP TO THE M-TH POWER (THE
C HIGHEST DEGREE POLYNOMIAL SPECIFIED) AND COMPUTE MEANS,
C STANDARD DEVIATIONS, AND CORRELATION COEFFICIENTS. THIS
C SUBROUTINE IS NORMALLY CALLED BEFORE SUBROUTINES ORDER,
C MINV AND MLTR IN THE PERFORMANCE OF A POLYNOMIAL
C REGRESSION.
C
C USAGE
C CALL GDATA (N,M,X,XBAR,STD,D,SUMSQ)
C
C DESCRIPTION OF PARAMETERS
C N - NUMBER OF OBSERVATIONS.
C M - THE HIGHEST DEGREE POLYNOMIAL TO BE FITTED.
C X - INPUT MATRIX (N BY M+1) . WHEN THE SUBROUTINE IS
C CALLED, DATA FOR THE INDEPENDENT VARIABLE ARE
C STORED IN THE FIRST COLUMN OF MATRIX X, AND DATA FOR
C THE DEPENDENT VARIABLE ARE STORED IN THE LAST
C COLUMN OF THE MATRIX. UPON RETURNING TO THE

```

```

PLOT1110
PLOT1120
PLOT1130
PLOT1140
PLOT1150
PLOT1160
PLOT1170
PLOT1180
PLOT1190
PLOT1200
PLOT1210
PLOT1220
PLOT1230
PLOT1240
PLOT1250
PLOT1260
PLOT1270
PLOT1280
PLOT1290
PLOT1300
PLOT1310
PLOT1320
PLOT1330
PLOT1340
PLOT1350
PLOT1360
PLOT1370
PLOT1380
PLOT1390
PLOT1400
PLOT1410
PLOT1420
PLOT1430
PLOT1440
PLOT1450
PLOT1460
GDAT 10
GDAT 20
GDAT 30
GDAT 40
GDAT 50
GDAT 60
GDAT 70
GDAT 80
GDAT 90
GDAT 100
GDAT 110
GDAT 120
GDAT 130
GDAT 140
GDAT 150
GDAT 160
GDAT 170
GDAT 180
GDAT 190
GDAT 200
GDAT 210
GDAT 220
GDAT 230
GDAT 240

```

```

C          CALLING ROUTINE, GENERATED POWERS OF THE INDEPENDENT GDAT 250
C          VARIABLE ARE STORED IN COLUMNS 2 THROUGH M.          GDAT 260
C          XBAR - OUTPUT VECTOR OF LENGTH M+1 CONTAINING MEANS OF GDAT 270
C          INDEPENDENT AND DEPENDENT VARIABLES.                 GDAT 280
C          STD  - OUTPUT VECTOR OF LENGTH M+1 CONTAINING STANDARD GDAT 290
C          DEVIATIONS OF INDEPENDENT AND DEPENDENT VARIABLES.  GDAT 300
C          D    - OUTPUT MATRIX (ONLY UPPER TRIANGULAR PORTION OF THE GDAT 310
C          SYMMETRIC MATRIX OF M+1 BY M+1) CONTAINING CORRELA- GDAT 320
C          TION COEFFICIENTS. (STORAGE MODE OF 1)              GDAT 330
C          SUMSQ - OUTPUT VECTOR OF LENGTH M+1 CONTAINING SUMS OF GDAT 340
C          PRODUCTS OF DEVIATIONS FROM MEANS OF INDEPENDENT    GDAT 350
C          AND DEPENDENT VARIABLES.                             GDAT 360
C
C          REMARKS                                             GDAT 370
C          N MUST BE GREATER THAN M+1.                         GDAT 380
C          IF M IS EQUAL TO 5 OR GREATER, SINGLE PRECISION MAY NOT BE GDAT 390
C          SUFFICIENT TO GIVE SATISFACTORY COMPUTATIONAL RESULTS. GDAT 400
C
C          SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED        GDAT 410
C          NONE                                               GDAT 420
C
C          METHOD                                               GDAT 430
C          REFER TO B. OSTLE, 'STATISTICS IN RESEARCH', THE IOWA STATE GDAT 440
C          COLLEGE PRESS, 1954, CHAPTER 6.                   GDAT 450
C
C          ..... GDAT 460
C          ..... GDAT 470
C          ..... GDAT 480
C          ..... GDAT 490
C          ..... GDAT 500
C          ..... GDAT 510
C          SUBROUTINE GDATA (N,M,X,XBAR,STD,D,SUMSQ)          GDAT 520
C          DIMENSION X(1),XBAR(1),STD(1),D(1),SUMSQ(1)      GDAT 530
C          ..... GDAT 540
C          ..... GDAT 550
C          ..... GDAT 560
C          IF A DOUBLE PRECISION VERSION OF THIS ROUTINE IS DESIRED, THE GDAT 570
C          C IN COLUMN 1 SHOULD BE REMOVED FROM THE DOUBLE PRECISION GDAT 580
C          STATEMENT WHICH FOLLOWS.                          GDAT 590
C
C          DOUBLE PRECISION X,XBAR,STD,D,SUMSQ,T1,T2        GDAT 600
C          ..... GDAT 610
C          ..... GDAT 620
C          THE C MUST ALSO BE REMOVED FROM DOUBLE PRECISION STATEMENTS GDAT 630
C          APPEARING IN OTHER ROUTINES USED IN CONJUNCTION WITH THIS GDAT 640
C          ROUTINE.                                          GDAT 650
C          ..... GDAT 660
C          THE DOUBLE PRECISION VERSION OF THIS SUBROUTINE MUST ALSO GDAT 670
C          CONTAIN DOUBLE PRECISION FORTRAN FUNCTIONS. SQRT AND ABS IN GDAT 680
C          STATEMENT 100 MUST BE CHANGED TO DSQRT AND DABS. GDAT 690
C          ..... GDAT 700
C          ..... GDAT 710
C          ..... GDAT 720
C          GENERATE INDEPENDENT VARIABLES                    GDAT 730
C          ..... GDAT 740
C          IF(M-1) 105, 105, 90                             GDAT 750
C          90 L1=0                                           GDAT 760
C          DO 100 I=2,M                                       GDAT 770
C          L1=L1+N                                           GDAT 780
C          DC 100 J=1,N                                       GDAT 790
C          L=L1+J                                           GDAT 800
C          K=L-N                                             GDAT 810
C          100 X(L)=X(K)*X(J)                                  GDAT 820
C          ..... GDAT 830
C          CALCULATE MEANS                                   GDAT 840

```

C		G DAT 850
	105 MM=M+1	G DAT 860
	DF=N	G DAT 870
	L=C	G DAT 880
	DC 115 I=1,MM	G DAT 890
	XBAR(I)=0.0	G DAT 900
	DO 110 J=1,N	G DAT 910
	L=L+1	G DAT 920
	110 XBAR(I)=XBAR(I)+X(L)	G DAT 930
	115 XBAR(I)=XBAR(I)/DF	G DAT 940
C		G DAT 950
	DC 130 I=1,MM	G DAT 960
	130 STD(I)=0.0	G DAT 970
C		G DAT 980
C	CALCULATE SUMS OF CROSS-PRODUCTS OF DEVIATIONS	G DAT 990
C		G DAT 1000
	L=((MM+1)*MM)/2	G DAT 1010
	DO 150 I=1,L	G DAT 1020
	150 D(I)=0.0	G DAT 1030
	DC 170 K=1,N	G DAT 1040
	L=C	G DAT 1050
	DC 170 J=1,MM	G DAT 1060
	L2=N*(J-1)+K	G DAT 1070
	T2=X(L2)-XBAR(J)	G DAT 1080
	STD(J)=STD(J)+T2	G DAT 1090
	DC 170 I=1,J	G DAT 1100
	L1=N*(I-1)+K	G DAT 1110
	T1=X(L1)-XBAR(I)	G DAT 1120
	L=L+1	G DAT 1130
	170 D(L)=D(L)+T1*T2	G DAT 1140
	L=C	G DAT 1150
	DC 175 J=1,MM	G DAT 1160
	DO 175 I=1,J	G DAT 1170
	L=L+1	G DAT 1180
	175 D(L)=D(L)-STD(I)*STD(J)/DF	G DAT 1190
	L=C	G DAT 1200
	DC 180 I=1,MM	G DAT 1210
	L=L+1	G DAT 1220
	SUMSC(I)=D(L)	G DAT 1230
	180 STD(I)= SQRT(ABS(D(L)))	G DAT 1240
C		G DAT 1250
C	CALCULATE CORRELATION COEFFICIENTS	G DAT 1260
C		G DAT 1270
	L=C	G DAT 1280
	DC 190 J=1,MM	G DAT 1290
	DO 190 I=1,J	G DAT 1300
	L=L+1	G DAT 1310
	190 C(L)=D(L)/(STD(I)*STD(J))	G DAT 1320
C		G DAT 1330
C	CALCULATE STANDARD DEVIATIONS	G DAT 1340
C		G DAT 1350
	DF=SQRT(DF-1.0)	G DAT 1360
	DC 200 I=1,MM	G DAT 1370
	200 STD(I)=STD(I)/DF	G DAT 1380
	RETURN	G DAT 1390
	END	G DAT 1400

```
//LKED.SYSIN DD *
INSERT MAIN,CORRE,CATA,MINV
OVERLAY 1
```

```
INSERT R1,ORDER,MULTR,R4,GDATA,PLOT
OVERLAY 1
INSERT R2,MSTR,LOC,STFRG,STCUT
OVERLAY 1
INSERT R3,CANOR,NRCCT,EIGEN
```

```
//GC.FT06FU01 DD SYSOUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=1330)
//GC.FT13FU01 DD UNIT=SYSDA,CCB=(RECFM=VBS,LRECL=76,BLKSIZE=1524),
//  SPACE=(CYL,(5,2)),DSN=+RCB,DISP=(NEW,PASS)
//GC.FT05FU01 DD *
LEMAY 000221201
01010102
LEMAY 221201 0.00101
3
LEMAY 220604
LEMAY 22021
```

0000

BIBLIOGRAPHIE

- BARROWS, H.H., (1923): "Geography as Human Ecology", Annals of the Association of American Geographers, Vol.XIII, no 1, p. 1-14.
- BEAUJEU-GARNIER, Jacqueline (1971): La Géographie: méthodes et perspectives, coll. de Géogr. appli. Masson, 142 p.
- BERRY, Brian (1972): A New Paradigm for Modern Geography (non-publié).
- BERRY, Brian J.L. et HURTON, Frank E. (1970): Geographic Perspectives on Urban Systems with integrated readings, Englewood Cliffs, New Jersey, Prentice-Hall, 564 p.
- BERRY, Brian J.L. et MARBLE Duane W. (1969): Spatial Analysis: a reader in Statistical Geography, New York, Prentice-Hall, 512 p.
- BROWN, Gary Deword (1970): System/360 Job Control Language, John Willey & Sons Inc., Toronto, 292 p.
- BUNGE, William (1966): Theoretical Geography, Lund, Gleerup, "Lund Studies in Geography", sér. C, no 1, 289 p.
- BURT, C. (1952): Tests of Significance in Factor Analysis, British Journal of Psychology, Section 5, p. 109-133.
- CASETTI, E. (1964): "Multiple Discriminant Functions", Computer application in the Earth Science Project, Tech. Rept. no 11, Dept. Geogr., Northwestern University, 63 p.
- CASETTI, E. (1964 a): "Classificatory and Regional analysis by Discriminant Interactions", Computer application in the Earth Science Project, Tech.Rept. no 12, Dept. Geogr., Northwestern University, 95 p.
- COOLEY, William W. et LOHNES, Paul R. (1966): Multivariate procedures for Behavioral Sciences, New York, John Willey and Sons, 211 p.
- CRESS, Paul; DIRKSEN, Paul et GRAHAM, Wesley J. (1968): Fortran IV with Watfor. New Jersey, Prentice-Hall, 390 p.
- FERRIER, J.P. (1972): Introduction à l'analyse spatiale, G.B. vol.2, no. 1, p. 5-9.
- FREY, W. (1945): Sentiment et métabolisme. American Social Review, no.10

- FORESTER, J.W. (1969): Urban Dynamics, Cambridge, The Massachusetts Institute of Urbanology Press, 285 p.
- FRENCH, Hugh et RACINE, Jean-Bernard, éd. (1971): Quantitative and Qualitative Geography: la nécessité d'un dialogue, Ottawa, travaux du département de géographie de l'Université d'Ottawa, no 1, 216 p.
- GIRAULT, Maurice (1964): Calcul des Probabilités en vue des applications, Paris, Dunod, 194 p.
- GEORGE, P. (1969): La Géographie Active, P.U.F., 394 p.
- GEORGE, Pierre (1970): Les Méthodes de la Géographie, Paris, Presses Universitaires de France, "Que sais-je?" 126 p.
- GOODMAN, L.A. et KRUSKAL, W.H. (1954): Measures of Association for Cross Classifications, Journal of the American Statistical Association, vol. 54, p. 732-764.
- GOODMAN, L.A. et KRUSKAL, W.H. (1959): Measures of Association for Cross Classifications II: Further Discussion and References, Journal of the American Statistical Association, vol. 54, p. 123-163.
- GOODMAN, L.A. et KRUSKAL, W.H. (1963): Measures of Association for Cross Classification: III: Approximate Sampling Theory. Journal of the American Statistical Association, juin, p. 310-365.
- GOODMAN, L.A. (1968): The Analysis of Cross Classified Data: Independence, Quasi-independence, and Interactions in Contingency Tables with or Without Missing Entries, Journal of the American Statistical Association, vol. 63, p. 1091-1131.
- GOULD, P.R. (1967): "On the Geographical Interpretation of Eigenvalues", Transactions, Institute of British Geographers, décembre, no 42, p. 53-86, 19 fig. 6 tableaux.
- GOULD, Peter (1970): "Is Statistics Inferens the Geographical Name for a Wild Goose?" Economic Geography, vol. 46, No. 2 (Supplement) June 1970, Proceedings, International Geographical Union Commission on Quantitative Methods, Clark University, p. 439-448.
- GREER-WOOTTEN, Bryn (1971): "Some Reflections on Systems Analysis in Geographic Research", dans FRENCH, H. et RACINE, J.B. op.cit., p. 151-174.

- GUTTMAN, L. (1953): Image Theory for the Structure of Quantitative Variates, Psychometrika 18, p. 277-296.
- HARRIS, C.W. (1962): Some Roo-Guttman Relationship, Psychometrika 27 p. 247-263.
- HARMAN, H.H. (1967): Modern Factor Analysis, 2e ed., Chicago, University of Chicago Press, 474 p.
- HARTSHORNE, Richard (1939): The Nature of Geography, Annals, of the Association of American Geographers. Vol. XXIX, No 3 et 4, Penvsylvania 1948 p.
- HARVEY, David (1969): Explanation in Geography, Londres, Edmond Arnold, 521 p.
- KAYSER, H.F. (1958): The Varimax Criterion for Analytic Rotation of Factor Analysis. Psychometrika, Vol. 23, pp.187-200.
- KAYSER, H.F. (1959): Computer Program for Varimax Rotation in Factor Analysis. Educational and Psychological Measurement, Vol. 19, pp. 205-224.
- KAYSER, H.F. (1963): Image analysis in C.W. Harris, ed., Problems in Measuring Change, Madison, Wis., University of Wisconsin Press, 1963.
- KING, Leslie D. (1969): Statistical Analysis in Geography, Englerwoods Cliffs, Prentice-Hall, 288 p.
- KRUMBEIN, W.C. et GRAYBILL, Franklin A. (1965), An Introduction to Statistical Models in Geology, McGraw-Hill, 475 p.
- KUHN, T.S. (1962): The Structure of Scientific Revolutions, Chicago.
- LARASSE, Jean (1971): "Les problèmes soulevés par la géographie quantitative: quelques réflexions expérimentales et critiques", dans FRENCH, H.M. and/et RACINE, J.R. op. cit., pp. 57-68 et 190-194.
- LIORZOU, A. (1966): Initiation pratique à la statistique, Paris, Gauthier-Villars, 310 p.
- LOHNES, Paul R. et COOLEY, William W. (1968): Introduction to Statistical Procedures: with Computer Exercises, New York, John Willey and Sons, 280 p.

- McQUITTY, I.L. (1957): "Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types Relevancies", Educational and Psychological Measurements, vol. 17, p. 207-222.
- NEUHAUS, Jack O. and WRIGLEY, Charles (1954): The Quartimax Method: An Analytical Approach to Orthogonal Simple Structure, British Journal of Statistical Psychology, 7, p. 81-91.
- OLSSON, G. (1965): Distance and Human Interaction: a Migration Study, Geographic Annals, Series B, 47, 3-43.
- RACINE, J.B. (1972 a): Modèles graphiques et mathématiques en géographie. Revue de Géographie de Montréal. Vol. 26, no 1, pp. 7-34. II: Les algorithmes de l'analyse typologique.
- RACINE, J.B. et LEMAY, G. (1972 b): L'analyse discriminatoire des correspondances typologiques dans l'espace géographique. L'Espace Géographique, Vol.1, no 3, 1972 (à paraître).
- RACINE, J.B. (1972 c): La croissance du Grand Montréal au sud du Saint-Laurent. Thèse de doctorat d'Etat, à paraître.
- RACINE, J.B. (1967 a): La croissance du grand Montréal au sud du Saint-Laurent: le cas de Saint-Bruno de Montarville. Revue de Géographie de Montréal, no 1, pp. 111-147.
- RACINE, J.B. (1967 b): Exurbanisation et métamorphisme péri-urbain: introduction à l'étude de la croissance de la croissance du grand Montréal au sud du Saint-Laurent. Revue de Géographie de Montréal, Vol. XXI, no 2, pp. 313-342.
- RACINE, J.B. (1969): Nouvelle frontière pour la recherche géographique. Cahiers de Géographie de Québec, no 29, pp. 135-168.
- RACINE, J.B. (1970): L'Evolution récente du phénomène périurbain nord américain. I: les observations traditionnelles. Revue de Géographie de Montréal, Vol. 24, no 1, pp. 43-54.
- RACINE, J.B. (1971 a): Le modèle urbain américain: les mots et les choses. Annales de Géographie, no 440, juillet-août, pp. 397-427.

- RACINE, J.B. (1971 b): Modèles graphiques et mathématiques en géographie humaine. I: la transformation des unités statistiques quantitatives en unités géographiques qualitatives. Revue de Géographie de Montréal, Vol. 25, no 4, pp. 323-358.
- RACINE, J.B. et FRENCH, H.M. (1971 c) (Ed.): Quantitative and Qualitative Geography: la nécessité d'un dialogue. Travaux du Département de Géographie de l'Université d'Ottawa, no 1, 1971, Ed. Univ. Ottawa, 216 p.
- REES, H. (1971): Factorial Ecology. Economic Geography, Vol. 47, no 2 (Supplement).
- ROGERS, Andrei (1965): "A Stochastic Analysis of the Spatial Clustering of Retail Establishments", Jour.Amer.Stat.Ass., pp. 1099-1103.
- ROGERS, Andrei (1967): "Theories of Intraurban Spatial Structure: A Disserting View", Land Econ., Vol. XLIII, no 1, pp. 108-112.
- RUMMEL, R.S.: Understanding Factor Analysis, Journal of Resolution, 1967.
- RUTHERFORD, Brent M., (1971): Non-Metric Correlation methods: a sensitivity analysis using Monte Carlo Simulation. Center for Advanced Study of Educational Administration. University of Oregon, 46 p.(non-publié).
- SNEDECOR, G.W. et COCHRAN, W.C. (1967): Statistical Methods, Ames. The Iowa State College Press, 593 p.
- STEVEN, S.S. (1935): The Operational Basis of Psychology, American Journal of Psychology, 47, 323-30.
- THIRODEAU, (1972): Implantation manufacturière dans la région de Montréal, Les Cahiers du C.R.U.R., no 1, 1971, 94 p.
- THOMSON, Godfrey H. (1950): L'analyse factorielle des aptitudes humaines, traduit par Pierre Naville, P.U.F., 413 p.
- THOMAS, F.H. (1962): The Stability of Distance-Population Size Relationships for Iowa Towns from 1900 to 1950, Land Studies in Geography, Series B, vol. 29, p.13-30.
- TOPLER, Waldo (1970): Selected Computer Programs, Dept. Geogr., University of Michigan, 162 p.

- VELDMAN, Donald J. (1967): Fortran Programming for the Behavioral Sciences, New York, Holt, Rinehart and Winston, 406 p.
- WARD, J.H. (1961): Hierarchical Grouping to Maximise Payoff Technical Report Wadd-Tn-61-69, U.S.A.F. Lackland Air Force Base, Texas.
- WARD, J.H. (1963): Hierarchical Grouping to Optimize and Objective Function. Journ.Am.Stat.Ass., Vol. 58, pp. 236-244.