

NOTES DE STATISTIQUE ET D'INFORMATIQUE

94/1

TRAITEMENT DE DONNEES
PAR LE LOGICIEL MINITAB

J.J. CLAUSTRIAUX et A. DELVAUX

Faculté des Sciences
agronomiques

Centre de Recherches
agronomiques

GEMBLoux
(Belgique)

TRAITEMENT DE DONNÉES PAR LE LOGICIEL MINITAB

J.J. CLAUSTRIAUX⁽¹⁾ et A. DELVAUX⁽²⁾

RÉSUMÉ

Le document présente les principales procédures pour la manipulation et le traitement de données au moyen du logiciel statistique Minitab.

SUMMARY

This paper gives main procedures for data management and statistical analysis with the Minitab statistical software.

1. INTRODUCTION

Minitab est un logiciel statistique mis au point, à l'origine, par le Département de Statistique de l'Université de Pennsylvanie (U.S.A.). Il est particulièrement bien adapté pour l'analyse statistique de petits tableaux de données : statistique descriptive, analyse de la variance, méthodes relatives à la corrélation et à la régression simple et multiple, séries chronologiques, tests d'indépendance, méthodes non paramétriques, analyse en composantes principales, analyse discriminante, etc.

Ce logiciel constitue une excellente introduction à l'utilisation de logiciels statistiques plus élaborés, tel que SAS. Il ne nécessite qu'une connaissance élémentaire de l'outil informatique et l'emploi des quelque 500 commandes ou procédures et sous-commandes ou sous-procédures est facilité par la mise à la disposition de l'utilisateur d'une fonction d'aide (HELP).

⁽¹⁾ Professeur à la Faculté des Sciences agronomiques de Gembloux (Unité de Statistique et Informatique).

⁽²⁾ Assistante à la Faculté des Sciences agronomiques de Gembloux (Unité de Statistique et Informatique).

Ce document présente brièvement la version 9 de Minitab. Des informations complémentaires peuvent être trouvées dans le manuel de référence [X, 1992].

Les paragraphes 2, 3 et 4 sont consacrés aux manipulations des données et aux relations entre l'emploi du logiciel Minitab et l'environnement de l'utilisateur de l'ordinateur VAX 4000-100 (système d'exploitation VMS), par l'intermédiaire d'un terminal, type VT240 [X, 1989].

Il convient également de noter que la plupart des commandes présentées dans cette note sont également celles en usage dans les versions du logiciel qui sont implantées sur les micro-ordinateurs de types IBM compatibles et Apple Macintosh. Toutefois, dans ce cas, l'utilisateur peut aussi avoir accès à un éditeur (*data screen*) facilitant la manipulation des données. On peut encore signaler qu'un manuel de référence en français existe pour la version destinée au micro-ordinateur [X, 1993].

Enfin, les paragraphes 5, 6, 7 et 8 décrivent les principales procédures statistiques; leur emploi est illustré par quelques exemples dont les données sont reprises en annexes.

2. LA SÉANCE MINITAB

L'ouverture d'une séance interactive étant effectuée, le logiciel est mis en service en exécutant la fonction :

```
$ MINITAB « RET »(3)
```

A ce moment, apparaissent à l'écran du terminal les caractères :

```
MTB >
```

et une commande Minitab peut être introduite par l'utilisateur.

Des informations générales relatives au logiciel sont obtenues par la commande :

```
MTB > HELP OVERVIEW
```

et des informations relatives aux commandes Minitab elles-mêmes sont affichées quand on frappe :

```
MTB > HELP COMMANDS
```

Une séance Minitab se termine par l'exécution de la commande :

```
MTB > STOP(4)
```

En outre, l'instruction RESTART permet de débiter une nouvelle séance sans faire appel aux instructions de fermeture et d'ouverture.

⁽³⁾ Dorénavant, l'utilisation de la touche « RET » ne sera plus mentionnée.

⁽⁴⁾ Dorénavant, les caractères MTB > ne seront plus indiqués.

3. LECTURE ET TRANSFORMATION DES DONNÉES

3.1. Règles et notations

Chaque commande Minitab et ses qualificatifs sont composés, indifféremment en minuscules ou en majuscules, sur une ligne; seuls les quatre premiers caractères sont nécessaires pour identifier une commande. Celle-ci peut être accompagnée de sous-commandes. Dans ce cas, la commande et, éventuellement, les sous-commandes se terminent par le symbole « ; » et la dernière sous-commande se clôture par le caractère « . ».

La lettre C, suivie d'un numéro d'ordre, identifie une colonne de données, souvent appelée variable selon les conventions habituelles (exemple : C9). La lettre K, accompagnée d'un numéro d'ordre, est réservée aux constantes (exemple : K5), et la lettre M, associée à un numéro d'ordre, représente une matrice (exemple : M7).

3.2. Lecture des données

3.2.1. Données encodées au cours d'une séance Minitab

READ liste des colonnes

A titre d'exemple, la lecture des données reprises à l'annexe 1 [DAGNELIE, 1980] s'effectue de la manière suivante :

```
READ C1 C2 C3
```

et la commande, une fois exécutée, fait apparaître à l'écran les caractères :

```
DATA >
```

qui autorisent l'utilisateur à encoder le premier brelan de données, séparées par un ou plusieurs espaces, de la manière suivante :

```
DATA > 1959 12.2 9.8
```

et ainsi de suite jusqu'au dernier brelan ou à la dernière ligne de données :

```
DATA > 1964 22.3 20.8
```

l'ensemble étant suivi des caractères END.

D'autre part, les données d'une seule colonne, à raison de plusieurs données par ligne d'encodage, peuvent aussi être enregistrées grâce à la commande SET.

Ainsi, la lecture des mêmes données s'effectue aussi de la manière suivante :

```
SET C1
```

```
DATA > 1959 1960 ... 1964
```

```
DATA > SET C2
```

```
DATA > 12.2 14.2 ... 22.3
```

```
DATA > SET C3
```

```
DATA > 9.8 10.8 ... 20.8
```

```
DATA > END(5)
```

⁽⁵⁾ Dorénavant, les caractères DATA > ne seront plus indiqués.

La commande SET permet en outre de générer automatiquement des données. Ainsi, la première colonne de l'exemple peut être créée de la manière suivante :

```
SET C1
1959 : 1964
END
```

D'autre part, une colonne formée des valeurs 1 à 5, répétées trois fois, s'obtient comme suit :

```
SET C4
3(1 : 5)
END
```

tandis qu'une colonne comportant trois valeurs 1, trois valeurs 2, trois valeurs 3, trois valeurs 4 et trois valeurs 5 est automatiquement générée par les commandes :

```
SET C5
(1 : 5)3
END
```

Enfin, à la lecture, chaque ligne de données comporte au plus 80 caractères et l'astérisque est le symbole associé à une donnée manquante.

3.2.2. Données encodées dans un fichier

READ 'nom du fichier' liste des colonnes

Si, grâce à l'éditeur par exemple, les données (annexe 1) ont été préalablement encodées, sans terminer l'ensemble par les caractères END, dans un fichier DONNEES.DAT, la lecture des données s'effectue de la manière suivante (données séparées par un caractère blanc) :

```
READ 'DONNEES' C1 C2 C3
```

ou :

```
READ 'DONNEES' C1-C3
```

3.2.3. Dénomination des colonnes

NAME colonne 1 = 'nom 1', colonne 2 = 'nom 2', ...

A nouveau, toujours pour l'exemple considéré, il est possible d'associer un titre de huit caractères maximum à chaque colonne, en exécutant la commande suivante :

```
NAME C1 = 'ANNEES', C2 = 'ARBRE 1', C3 = 'ARBRE 2'
```

3.3. Correction des données

3.3.1. Correction d'une donnée

LET numéro de la colonne (numéro de la ligne) = nouvelle valeur de la donnée

L'exemple donné ci-dessous permet de substituer la valeur 15,6 à la quatrième observation de la colonne 3, préalablement encodée par erreur 15,8 :

```
LET C3(4) = 15.6
```

3.3.2. Insertion, ajout ou suppression de données

INSERT numéros d'ordre des lignes entre lesquelles il convient d'insérer des données et liste des colonnes

En supposant que le troisième breilan de données de l'exemple n'ait pas été encodé, les instructions suivantes doivent être exécutées pour insérer les valeurs oubliées :

```
INSERT 2 3 C1-C3
1961 16.7 13.3
END
```

L'ajout de données au début du tableau s'effectue en utilisant comme numéros d'ordre des lignes les valeurs 0 et 1. Quant à l'ajout de données à la fin du tableau, il est réalisé en ne spécifiant pas de numéros de lignes.

DELETE numéros d'ordre des lignes à supprimer et liste des colonnes

Pour supprimer les données des lignes 2, 3, 4 et 5 des colonnes 1 et 2 de l'exemple, on introduit la commande :

```
DELETE 2 3 4 5 C1 C2
```

ou :

```
DELETE 2 : 5 C1 C2
```

ERASE liste des colonnes et/ou liste des constantes

Il est possible de supprimer l'ensemble des données figurant dans la colonne 2 de la manière suivante :

```
ERASE C2
```

3.4. Transformation des données

LET constante ou colonne = instruction arithmétique

Toute instruction arithmétique peut faire appel aux symboles + - * / () et à des fonctions de base qui opèrent sur des colonnes ou des constantes, par exemple :

```
LET K1 = 3
LET C4 = C1 * C2 / K1
LET C5 = LOGE ((C1 + C2) * 25)
LET K2 = MEAN (C1)
LET C6 = (C1 - K2)/STDEV(C1)
```

STACK numéros des colonnes à empiler colonne résultante.

Pour créer une colonne de données C4 formée de l'empilement des colonnes C1, C2 et C3, la commande suivante est exécutée :

```
STACK C1 - C3 C4
```

UNSTACK numéro de la colonne à décomposer numéros des colonnes résultantes ;

SUBSCRIPTS numéro de la colonne contenant les indices de décomposition.

Pour décomposer une colonne de données C1 en trois colonnes C3, C4 et C5, en fonction des indices repris dans la colonne C2, on exécute ce qui suit :
UNSTACK C1 C3 - C5 ;
SUBSCRIPTS C2.

4. VISUALISATION, SAUVEGARDE DES DONNÉES, DES COMMANDES ET DES RÉSULTATS, PROGRAMME MINITAB

4.1. Visualisation des données INFORMATION

Cette commande permet d'obtenir des informations générales relatives aux colonnes et aux constantes utilisées.

PRINT liste des colonnes

La commande donnée ci-dessous engendre la visualisation à l'écran des données des colonnes C1 à C3 :

PRINT C1-C3

4.2. Sauvegarde des données

WRITE 'nom du fichier' liste des colonnes

Pour sauver les données de l'exemple et les données résultant des transformations proposées au paragraphe 3.4, dans un fichier .DAT (par exemple SUITE.DAT) du répertoire de l'utilisateur (au maximum : 80 caractères par ligne), la commande suivante doit être introduite :

WRITE 'SUITE' C1-C6

4.3. Sauvegarde des commandes et des résultats

4.3.1. Dans un fichier LIS

OUTFILE 'nom du fichier'

L'exécution de cette commande engendre l'enregistrement, dans un fichier .LIS du répertoire de l'utilisateur, des commandes postérieures à l'exécution de celle-ci et des résultats de leur exécution (au maximum 120 caractères par ligne) ; la commande NOECHO supprime l'enregistrement des commandes dans le fichier .LIS⁽⁶⁾.

4.3.2. Dans un fichier .MTW

Si on souhaite interrompre une séance Minitab, tout en sauvegardant le travail effectué, en vue de le poursuivre ultérieurement, il convient, préalablement à la clôture, d'introduire la commande :

SAVE 'nom du fichier'

qui crée, dans le répertoire, un fichier .MTW, non accessible à l'utilisateur.

Après la réouverture de la séance Minitab, la première commande à exécuter, pour récupérer les données, est la suivante :

RETRIEVE 'nom du fichier'

4.4. Programme Minitab

Lorsqu'on utilise régulièrement les mêmes procédures, on peut les rassembler dans un fichier .MAC et ainsi créer un programme ou macro Minitab. Ce programme peut en outre comporter des répétitions (boucles DO) et des décisions (tests IF). On distingue les macros globales (GMACRO) et les macros locales (MACRO) ; les secondes plus puissantes et plus complexes ne sont pas décrites.

Le fichier d'une macro globale se compose d'une première ligne correspondant au type de programme (GMACRO), d'une seconde ligne avec le nom de la macro et des procédures ; il se termine par l'instruction ENDMACRO.

A titre d'exemple, la macro globale MOMENT (fichier MOMENT.MAC) permettant de calculer le moment centré d'ordre 3, pour les valeurs encodées dans le fichier DONNEES.DAT comporte les instructions suivantes :

```
GMACRO
MOMENT
READ C1;
FILE 'DONNEES'.
PRINT C1
COUNT C1 K1
SUM C1 K2
SSQ C1 K3
LET K4=0
DO K10=1:K1
    LET K4=K4+C1(K10)**3
ENDDO
LET K5=(K4-((3*K2*K3)/K1)+((2*(K2**3))/K1**2))/K1
NOTE Le moment d'ordre 3 vaut:
PRINT K5
ENDMACRO
```

Lors d'une séance Minitab, une procédure est exécutée au moyen de la commande :

% nom du fichier

D'autre part, ce programme peut aussi être exécuté en temps différé sur VAX, à condition de créer un fichier .COM reprenant les instructions suivantes :

⁽⁶⁾ Les commandes PAPER et NOPAPER engendrent le début et la fin de l'impression d'une séance Minitab, lorsqu'une imprimante est reliée au terminal ou au micro-ordinateur.

\$ MINITAB
 % nom du fichier
STOP
Y

Dans ce cas, l'utilisateur lance l'exécution grâce à la commande VMS suivante :

\$ SUBMIT/QUE=BATCHRAP nom du fichier .COM

Les commandes Minitab peuvent également être encodées dans un fichier .MTB, exécuté lors d'une séance Minitab par l'instruction :

EXECUTE 'nom du fichier'

ou exécuté en temps différé sur VAX, à condition de faire précéder les commandes Minitab de l'instruction BATCH et de créer un fichier .COM, reprenant les commandes suivantes :

\$ MINITAB
EXECUTE 'nom du fichier'

Ce type de fichier (.MTB) est toutefois amené à disparaître dans les prochaines versions du logiciel Minitab. En outre, il ne permet pas l'utilisation des instructions DO et IF.

5. STATISTIQUE DESCRIPTIVE

5.1. Commandes

5.1.1. Statistique descriptive à une dimension

Distribution de fréquences, valeurs entières : TABLE C1
 Histogramme : HISTOGRAM C1
 Polygone de fréquences : DOTPLOT C1
 Réduction des données : DESCRIBE C1

5.1.2. Statistique descriptive à deux dimensions

Distribution de fréquences, valeurs entières : TABLE C1 C2
 Diagrammes de dispersion
 C1 = f(C2) : PLOT C1 C2
 C2 = f(C1), C3 = f(C1), etc. : MPLOT C2 C1, C3 C1, etc

5.2. Commentaires et exemples

L'établissement d'une distribution de fréquences à deux dimensions, présent à la fois les fréquences absolues et relatives, nécessite l'exécution de la commande et des sous-commandes suivantes :

TABLE C1 C2 ;
 COUNTS ;
 TOTPERCENTS.

La commande donnée ci-dessous permet de calculer la moyenne et l'écart-type estimé d'une série d'observations (C1) en fonction de deux critères discontinus (C2 et C3) :

TABLE C2 C3 ;
 MEANS C1 ;
 STDEV C1 .

La commande et la sous-commande suivantes permettent de calculer la moyenne, la médiane, la moyenne tronquée de 5 % des données, l'écart-type estimé, l'erreur-standard de la moyenne, le minimum, le maximum, le premier quartile et le troisième quartile d'une série d'observations (C1) en fonction d'un critère discontinu (C2) :

DESCRIBE C1 ;
 BY C2 .

6. DISTRIBUTIONS THÉORIQUES

6.1. Commandes

6.1.1. Variable aléatoire

Fonction de densité de probabilité

$x = C1$ et $f(x) = C2$

ou probabilité

$x = C1$ et $P(x) = C2$: PDF C1 C2 ;
 code de la distribution
 et paramètre(s).

Fonction de répartition

$x = C1$ et $F(x) = C2$: CDF C1 C2 ;

code de la distribution
 et paramètre(s).

Fonction inverse

$F(x) = C1$ et $x = C2$: INVCDF C1 C2 ;
 code de la distribution
 et paramètre(s).

6.1.2. Distributions théoriques (codes des distributions et paramètres)

Distribution binomiale (n, p) : BINOMIAL K1 K2
 Distribution de POISSON (m) : POISSON K1
 Distribution uniforme (a, b) : UNIFORM K1 K2
 Distribution normale (m, σ) : NORMAL K1 K2
 Distribution log-normale (m, σ) : LOGNORMAL K1 K2
 Distribution t de STUDENT (k) : T K1
 Distribution χ^2 de PEARSON (k) : CHISQUARE K1
 Distribution F de SNEDECOR
 (k_1, k_2) : F K1 K2

6.1.3. Génération de nombres pseudo-aléatoires et échantillonnage

Nombres pseudo-aléatoires : RANDOM K C1 ... Cn ;
code de la distribution
et paramètre(s).
Echantillonnage aléatoire : SAMPLE K C1 C2

6.2. Commentaires et exemples

Pour calculer $P(0)$, dans le cas d'une distribution de POISSON de moyenne unitaire, on applique les commandes suivantes :

PDF 0 ;
POISSON 1.

Pour calculer $F(1)$, dans le cas d'une distribution binomiale de paramètres $n = 20$ et $p = 0,05$, les commandes sont :

CDF 1 ;
BINOMIAL 20 0.05.

Pour calculer la valeur x telle que $F(x) = 0,95$, dans le cas d'une distribution de SNEDECOR à 5 et 15 degrés de liberté, les commandes suivantes sont exécutées :

INVCDF 0.95 ;
F 5 15.

Pour des valeurs distribuées selon une loi normale de moyenne nulle et d'écart-type unitaire enregistrées dans la colonne C1, il est possible de calculer les valeurs correspondantes de la fonction de densité de probabilité et de les enregistrer en C2, en exécutant les instructions suivantes :

PDF C1 C2 ;
NORMAL 0 1.

Le graphe de la fonction peut alors être obtenu en faisant appel à la commande suivante :

PLOT C2 C1

D'autre part, la génération de quatre séries de 20 nombres pseudo-aléatoires normaux ($m = 0, \sigma = 1$) est obtenue comme suit :

RANDOM 20 C1 C2 C3 C4 ;
NORMAL 0 1.

7. INFÉRENCE STATISTIQUE CLASSIQUE

7.1. Commandes

7.1.1. Méthodes relatives aux moyennes

Intervalle de confiance :
variance connue : ZINTERVAL σ , C1
variance inconnue : TINTERVAL C1
Test de conformité :
variance connue : ZTEST m_0 σ , C1
variance inconnue : TTEST m_0 , C1
Test de signification :
variance connue : ZTEST σ , C1
variance inconnue : TTEST C1

Test d'égalité de deux moyennes :
échantillons indépendants et
égalité des variances
un échantillon dans C1 et
l'autre dans C2 : TWOSAMPLE C1 C2 ;
POOLED.

les deux échantillons dans C1 et
les numéros des objets
dans C2 : TWOT C1 C2 ;
POOLED.

échantillons indépendants et
inégalité des variances
un échantillon dans C1 et
l'autre dans C2 : TWOSAMPLE C1 C2
les deux échantillons dans C1 et
les numéros des objets
dans C2 : TWOT C1 C2
échantillons associés par paires
: LET C3 = C2-C1
TTEST C3

7.1.2. Test de normalité

Graphique : NSCORES C1 C2
PLOT C1 C2

7.1.3. Test d'indépendance

Test χ^2
table de contingence
dans C1 à Cn : CHISQUARE C1 ... Cn
valeurs dans C1 et C2 : TABLE C1 C2 ;
CHISQUARE 2.

La constante 2 permet d'obtenir à la fois les fréquences observées et attendues.

7.1.4. Analyse de la variance à un critère de classification

Tous les échantillons dans C1 et les numéros des objets dans C2

: ANOVA C1 = C2;
RANDOM C2;
EMS;
RESTRICT;
MEANS C2.

7.1.5. Analyse de la variance à deux critères de classification

Tous les échantillons dans C1 et les numéros des objets dans C2 et C3, modèle croisé, $n = 1$

: ANOVA C1 = C2 C3;
RANDOM C2 C3;
EMS;
RESTRICT;
MEANS C2 C3.

modèle croisé, $n > 1$,

échantillons d'effectifs égaux : ANOVA C1 = C2 C3 C2*C3;
(ou ANOVA C1 = C2 | C3);
RANDOM C2 C3;
EMS;

RESTRICT;
MEANS C2 C3 C2*C3.

modèle croisé fixe, $n > 1$,

échantillons d'effectifs inégaux : GLM C1 = C2 C3 C2*C3
(ou GLM C1 = C2 | C3)

modèle hiérarchisé, $n > 1$,

échantillons d'effectifs égaux : ANOVA C1 = C2 C3(C2);
RANDOM C2 C3;
EMS;
RESTRICT;
MEANS C2 C3.

7.1.6. Analyse de la variance à trois critères de classification

Tous les échantillons dans C1 et les numéros des objets dans C2, C3 et C4, modèle croisé, $n = 1$

: ANOVA C1 = C2 C3 C4 C2*C3
C2*C4 C3*C4;
RANDOM C2 C3 C4;
EMS;
RESTRICT;
MEANS C2 C3 C4 C2*C3
C2*C4 C3*C4.

modèle croisé, $n > 1$,

échantillons d'effectifs égaux : ANOVA C1 = C2 C3 C4
C2*C3 C2*C4 C3*C4 C2*C3*C4;
(ou ANOVA C1 = C2 | C3 | C4);
RANDOM C2 C3 C4;
EMS;
RESTRICT;
MEANS C2 C3 C4 C2*C3
C2*C4 C3*C4 C2*C3*C4.

modèle croisé fixe, $n > 1$,

échantillons d'effectifs inégaux : GLM C1 = C2 C3 C4
C2*C3 C2*C4 C3*C4 C2*C3*C4
(ou GLM C1 = C2 | C3 | C4)

modèle hiérarchisé, $n > 1$,

échantillons d'effectifs égaux : ANOVA C1 = C2 C3(C2) C4(C3);
RANDOM C2 C3 C4;
EMS;
RESTRICT;
MEANS C2 C3 C4.

modèle partiellement hiérarchisé,

$n = 1$ (exemple) : ANOVA C1=C2 C3 C4(C3) C2*C3;
RANDOM C2 C3 C4;
EMS;
RESTRICT;
MEANS C2 C3 C2*C3.

7.1.7. Analyse de la variance à plus de trois critères de classification

Commande ANOVA ou GLM, selon que le modèle est orthogonal ou non.

7.1.8. Analyse de la covariance

Commande ANCOVA ou GLM, et sous-commande COVARIATES, selon que le modèle est orthogonal ou non.

7.1.9. Comparaisons multiples des moyennes

Tous les échantillons dans C1 et les numéros des objets dans C2, AV1 modèle fixe, test de DUNNETT (témoin = k) et ppds

: ONEWAY C1 = C2;
DUNNETT K;
FISHER.

7.1.10. Régression et corrélation simples

Droite de régression $C1 = f(C2)$
avec estimation d'une ou plusieurs valeurs de y
en fonction de valeur(s) de x

: REGRESS C1 1 C2;
PREDICT valeur(s) de x .

Droite de régression $C1 = f(C2)$
avec résidus (C3) et valeurs
estimées (C4)
: REGRESS C1 1 C2 C3 C4
Matrice de covariance
: COVARIANCE C1 ... Cn
Matrice de corrélation
: CORRELATION C1 ... Cn

7.1.11. Régression multiple

Régression multiple $C1 = f(C2, \dots, Cn)$
avec estimation d'une ou
plusieurs valeurs de y
en fonction des valeurs $x_2 \dots x_n$: REGRESS C1 n-1 C2 ... Cn ;
PREDICT $x_2 \dots x_n$.

Régression multiple pas à pas

$C1 = f(C2, \dots, Cn)$
avec modification des valeurs F
théoriques pour l'entrée (F_e) ou
la sortie d'une variable (F_s) : STEPWISE C1 C2 ... Cn ;
FENTER = F_e ;
FREMOVE = F_s .

Meilleurs modèles de régression multiple

$C1 = f(C2, \dots, Cn)$ selon le critère
de valeur maximum de R^2 ,
avec modification du nombre de
modèles (p) et du nombre de
variables explicatives (q à r) : BREG C1 C2 ... Cn ;
BEST p ;
NVAR q r .

7.2. Commentaires et exemples

Les instructions ci-dessous permettent d'obtenir le tableau d'analyse de la variance à un critère de classification pour l'exemple repris à l'annexe 2, lorsque les hauteurs de tous les arbres sont encodées dans C1 et les numéros des types de forêts dans C2 :
ANOVA C1 = C2 ;
MEANS C2 .

Pour obtenir le tableau d'analyse de la variance à deux critères de classification à partir des données reprises à l'annexe 3 (modèle croisé mixte) ainsi que les moyennes par fongicide et par bloc, lorsque les rendements sont encodés dans C1, les numéros des traitements fongicides dans C2 et les numéros des blocs dans C3, il convient d'exécuter les commandes suivantes :

ANOVA C1 = C2 C3 ;
RANDOM C3 ;
MEANS C2 C3 .

Le tableau d'analyse de la variance à deux critères de classification, modèle hiérarchisé mixte (exemple : annexe 4), et le tableau contenant les espérances mathématiques des carrés moyens et les composantes de la variance peuvent être obtenus par les commandes suivantes, lorsque les rendements fourragers sont encodés dans C1, les numéros des types de prairies dans C2 et les numéros des prairies dans C3 :

ANOVA C1 = C2 C3(C2) ;
RANDOM C3 ;
EMS ;
RESTRICT .

Pour réaliser le tableau d'analyse de la variance à trois critères de classification, modèle croisé mixte (exemple : annexe 5) et calculer les moyennes par additif, par dose et par combinaison additif-dose, lorsque les rendements sont encodés dans C1, les numéros des additifs dans C2, les numéros des doses dans C3 et les numéros des blocs dans C4, il convient d'exécuter les commandes suivantes :

ANOVA C1 = C2 C3 C4 C2*C3 C2*C4 C3*C4 ;
RANDOM C4 ;
MEANS C2 C3 C2*C3 .

Pour obtenir le même tableau d'analyse de la variance, avec regroupement des interactions faisant intervenir le facteur blocs, on exécute ce qui suit :

ANOVA C1 = C2 C3 C4 C2*C3 ;
RANDOM C4 .

Les instructions données ci-dessous permettent d'obtenir le tableau d'analyse de la variance à trois critères de classification pour un dispositif en carré latin (exemple : annexe 6), lorsque les rendements sont encodés dans C1, les numéros des traitements dans C2, les numéros des lignes dans C3 et les numéros des colonnes dans C4 :

GLM C1 = C2 C3 C4

Les commandes suivantes permettent d'obtenir l'équation de régression linéaire au sens des moindres carrés, le test de signification du coefficient de régression, le test de signification de l'ordonnée à l'origine et l'estimation d'une valeur moyenne ou d'une valeur individuelle ($x = 1963$) avec leurs intervalles de confiance (exemple : annexe 1), lorsque les années sont encodées dans C1 et les circonférences dans C2 :

REGRESS C2 1 C1 ;
PREDICT 1963 .

8. MÉTHODES COMPLÉMENTAIRES

8.1. Méthodes non paramétriques

8.1.1. Méthodes relatives à un échantillon

Test des séquences : RUNS C1
Test des médianes ($\bar{x} = k$) : STEST K C1
Test de WILCOXON ($\bar{x} = k$) : WTEST K C1

8.1.2. Méthodes relatives à deux échantillons

Test des rangs, échantillons indépendants : un échantillon dans C1 et l'autre dans C2 : MANNWHITNEY C1 C2

8.1.3. Méthodes relatives à plus de deux échantillons

Test des médianes, échantillons indépendants : tous les échantillons dans C1 et les numéros des objets dans C2 : MOOD C1 C2

Test des rangs, échantillons indépendants : tous les échantillons dans C1 et les numéros des objets dans C2 : KRUSKAL-WALLIS C1 C2

Test des rangs, échantillons non indépendants : tous les échantillons dans C1 et les numéros des objets dans C2 et C3 : FRIEDMAN C1 C2 C3

8.2. Méthodes multivariées

8.2.1. Analyse de la variance à plusieurs variables

Commande ANOVA ou GLM, selon que le modèle est orthogonal ou non, en utilisant des sous-commandes : MANOVA, EIGEN, SSCP et PARTIAL.

8.2.2. Analyse discriminante

Les numéros des groupes dans C1 et les variables dans C2 à Cn :
DISCRIMINANT C1 C2 ... Cn ;
PRIORS ;
FITS ;
XVAL ;
PREDICT ...

8.2.3. Analyse des composantes principales

Les variables dans C1 à Cn :
PCA C1 ... Cn ;
COVARIANCE ;
NCOMP = K ;
COEFF C ... ;
SCORES C ...

8.2.4. Analyse factorielle

Les variables dans C1 à Cn :
FACTOR C1 ... Cn ;
COVARIANCE ;

NFAC = K ;
SORT.

D'autres sous-commandes sont disponibles, en particulier pour choisir une méthode de rotation des axes.

8.3. Séries chronologiques et contrôle de la qualité

Le logiciel Minitab permet également d'analyser des séries chronologiques (commandes ARIMA, etc.) et d'appliquer certaines méthodes de contrôle de la qualité (commandes XBARChart, etc.).

Des informations sur les commandes Minitab pour l'analyse des séries chronologiques peuvent être trouvées dans PALM [1987 a et b].

8.4. Calcul matriciel

Lecture d'une matrice (k_1 lignes et k_2 colonnes) : READ K1 K2 M1
Ecriture d'une ou plusieurs matrices : PRINT M1 ... Mn
Transformation de colonnes (vecteurs) en une matrice : COPY C1 ... Cn M1
Transformation d'une matrice en colonnes (vecteurs) : COPY M1 C1 ... Cn
Transposition ($M_2 = M_1'$) : TRANSPOSE M1 M2
Multiplication ($M_3 = M_1 M_2$) : MULTIPLY M1 M2 M3
Addition ($M3 = M1 + M2$) : ADD M1 M2 M3
Soustraction ($M3 = M1 - M2$) : SUBTRACT M2 M1 M3
Extraction de la diagonale : DIAGONAL M1 C1
Inversion ($M_2 = M_1^{-1}$) : INVERT M1 M2
Calcul de valeurs propres : EIGEN M1 C1

9. BIBLIOGRAPHIE

- DAGNELIE P. [1980]. *Théorie et méthodes statistiques : applications agronomiques* (vol. 2). Gembloux, Presses Agronomiques, 463 p.
- PALM R. [1987a]. Etude des séries chronologiques par les méthodes de décomposition. *Notes Stat. Inform.* (Gembloux) 87/1, 25 p.
- PALM R. [1987b]. Etude des séries chronologiques par la méthodologie de Box et JENKINS. *Notes Stat. Inform.* (Gembloux) 87/2, 40 p.
- X [1989]. *VMS User's manual*. Maynard, Digital Equipment, 928 p.
- X [1992]. *Minitab reference manual*. PA State College, Minitab, 504 p.
- X [1993]. *Minitab, manuel de référence*. PA State College, Minitab, 659 p.

10. ANNEXES

Annexe 1. - Circonférences observées au cours de plusieurs années [DAGNELIE, 1980 ; tableau 18.5.1].

Années	Arbre 1	Arbre 2
1959	12,2	9,8
1960	14,2	10,8
1961	16,7	13,3
1962	18,0	15,6
1963	-	-
1964	22,3	20,8

Annexe 2. - Comparaison des hauteurs des arbres de trois types de forêts [DAGNELIE, 1980 ; tableau 14.3.3].

Type 1	Type 2	Type 3
23,4	22,5	18,9
24,4	22,9	21,1
24,6	23,7	21,2
24,9	24,0	22,1
25,0	24,4	22,5
26,2	24,5	23,6
26,3	25,3	24,5
26,8	26,0	24,6
26,8	26,2	26,2
26,9	26,4	26,7
27,0	26,7	-
27,6	26,9	-
27,7	27,4	-
-	28,5	-

Annexe 3. - Comparaison de différents traitements fongicides appliqués à une culture de pomme de terre [DAGNELIE, 1980 ; tableau 15.3.11].

	Bloc 1	Bloc 2	Bloc 3	Bloc 4
Traitement 1	527	604	606	533
Traitement 2	633	600	650	567
Traitement 3	642	708	662	504
Traitement 4	623	550	562	667
Traitement 5	377	408	500	333

Annexe 4. - Comparaison et étude de l'homogénéité des rendements fourragers de deux types de prairies [DAGNELIE, 1980 ; tableau 15.5.2].

Types	1			2		
	1	2	3	1	2	3
Prairies	20,6	15,9	19,2	29,1	15,7	24,3
Parcelle 1	29,9	26,3	18,5	32,7	18,2	21,7
Parcelle 2	19,8	19,8	21,4	34,5	26,9	23,7
Parcelle 3	29,5	22,5	13,3	39,2	32,5	28,9
Parcelle 4	27,0	20,9	18,3	43,4	31,1	22,4

Annexe 5. - Comparaison de différents traitements herbicides appliqués à une culture de pois [DAGNELIE, 1980 ; tableau 16.2.5].

Additifs	Doses	Bloc 1	Bloc 2	Bloc 3	Bloc 4	Bloc 5
1	1	18,0	25,3	22,1	22,1	22,2
1	2	20,6	21,1	20,9	21,0	23,0
1	3	19,1	22,8	20,3	22,2	22,8
2	1	22,7	25,3	22,4	24,3	25,1
2	2	22,8	23,7	23,6	22,0	24,0
2	3	21,0	23,3	21,5	20,5	19,2

Annexe 6. - Comparaison de différents traitements herbicides appliqués à une culture de haricot [DAGNELIE, 1980 ; tableau 16.2.8].

Lignes	Colonnes					
	1	2	3	4	5	6
1	2: 28,7	5: 28,4	4: 25,4	3: 30,7	1: 30,6	6: 30,9
2	6: 31,4	3: 30,1	2: 27,4	5: 26,8	4: 29,8	1: 29,8
3	4: 29,4	6: 29,7	1: 30,4	2: 22,0	5: 24,1	3: 32,9
4	1: 29,6	2: 21,8	5: 22,5	6: 30,0	3: 30,6	4: 28,5
5	3: 25,8	4: 21,9	6: 23,1	1: 24,3	2: 20,7	5: 17,7
6	5: 18,1	1: 23,6	3: 22,5	4: 20,2	6: 23,7	2: 18,9

La collection

NOTES DE STATISTIQUE ET D'INFORMATIQUE

réunit divers travaux (documents didactiques, notes techniques, rapports de recherche, publications, etc.) émanant des services de statistique et d'informatique de la Faculté des Sciences agronomiques et du Centre de Recherches agronomiques de Gembloux (Belgique).

Quelques titres récents:

RAMLOT P. [1986]. Programmation structurée en Cobol, en Fortran et en Basic: aspects méthodologiques. *Notes Stat. Inform.* (Gembloux) 86/7, 33 p.

RAMLOT P., TOUSSAINT A., VANDEVANDEL J.P. [1986]. Programmation structurée en Cobol, en Fortran et en Basic: étude de cas. *Notes Stat. Inform.* (Gembloux) 86/8, 36 p.

PALM R. [1987]. Etude des séries chronologiques par les méthodes de décomposition. *Notes Stat. Inform.* (Gembloux) 87/1, 25 p.

PALM R. [1987]. Etude des séries chronologiques par la méthode de BOX et JENKINS. *Notes Stat. Inform.* (Gembloux) 87/2, 40 p.

PALM R. [1988]. Les critères de validation des équations de régression linéaire. *Notes Stat. Inform.* (Gembloux) 88/1, 27 p.

PALM R., DE BAST A. [1988]. Construction d'un modèle agrométéorologique pour la prévision des productions agricoles dans la Communauté Economique Européenne. *Notes Stat. Inform.* (Gembloux) 88/2, 14 p.

PALM R. [1989]. Quelques éléments de programmation linéaire. *Notes Stat. Inform.* (Gembloux) 89/1, 37 p.

DAGNELIE P. [1989]. Le choix d'une méthode d'analyse statistique et l'examen préliminaire des données. *Notes Stat. Inform.* (Gembloux) 89/2, 17 p.

PALM R. [1990]. La corrélation canonique: principes et application. *Notes Stat. Inform.* (Gembloux) 90/1, 28 p.

CARLETTI G. [1991]. Les micro-ordinateurs: présentation générale. *Notes Stat. Inform.* (Gembloux) 91/1, 27 p.

IEMMA A.F., PALM R. [1992]. Les matrices généralisées et leur utilisation dans le modèle linéaire. *Notes Stat. Inform.* (Gembloux) 92/1, 25 p.

PALM R. [1993]. Les méthodes d'analyse factorielle: principes et applications. *Notes Stat. Inform.* (Gembloux) 93/1, 38 p.