

Introduction à SPSS

Guy Mélard, U.L.B.
Mars 2006
(gmelard@ulb.ac.be)

Objectif de la leçon

Elle a pour but de fournir une introduction à ce logiciel statistique de manière à faciliter son utilisation en complément de SAS, avec l'illustration sur des exemples concrets de statistique appliquée

Exemple: étude clinique en rhumatologie

22 polyarthritiques

- ✓ âge du patient (AGE)
 - ✓ sévérité de l'affection (0 à 4: SEVER)
 - ✓ type de traitement
 - ◆ anti-inflammatoires (TANTINFL) et/ou
 - ◆ stéroïdes (TSTEROI)
 - ✓ dosages de 3 enzymes: 5'NU (FNU), ADA, PNP
 - ✓ pourcentages de lymphographie: LYMPHOST, OKT4, OKT8
- + 3 groupes de contrôle: dosage d'un seul enzyme pour chacun d'eux

Questions:

1. le dosage des enzymes est-il plus bas chez les arthritiques?
2. si oui, y-a-t-il une dépendance entre ce dosage et ...
 - 2.1 l'âge?
 - 2.2 le traitement?
 - 2.3 la sévérité de l'affection?
 - 2.4 les trois pourcentages?

Nature des variables

- ⇒ Classification par nature: qualitative, quantitative
- ⇒ Classification plus fine basée sur l'échelle de mesure
 - ✓ échelle nominale
 - ✓ échelle ordinale
 - ✓ échelle d'intervalle
 - ✓ échelle de rapport

Données manquantes, aberrantes

- ⇒ Données manquantes: pourquoi? Que faire? Effet?
 - ⇒ Données aberrantes: pourquoi? Que faire? Effet?
- N.B. SPSS est le progiciel statistique qui prend le mieux en compte les données manquantes; le problème des données aberrantes nécessite l'emploi de méthodes statistiques robustes.

Introduction à SPSS

- ⇒ Versions de SPSS
 - ⇒ Modules de SPSS
 - ⇒ Fenêtres de SPSS
 - ⇒ Types de menus
 - ⇒ Menus de SPSS
 - ⇒ Graphiques et graphiques interactifs
- N.B. On insiste plus sur la rédaction de programmes pour SPSS que sur l'emploi interactif

Versions de SPSS

- ⇒ Version installée dans la salle Math-Stat : version 11.5 de SPSS pour Windows
- ⇒ La version 12 existe déjà
- ⇒ Texte sur l'utilisation: l'essentiel est déjà dans la version 5.0
- ⇒ Texte introductif: version 5.0

Modules de SPSS

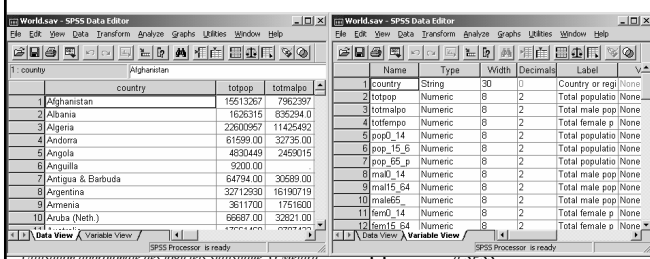
- ⇒ SPSS Base (procédures de base)
- ⇒ SPSS Advanced Statistics (procédures statistiques)
- ⇒ SPSS Professional Statistics (procédures statistiques)
- ⇒ SPSS Tables (préparation de rapports)
- ⇒ SPSS Trends (économétrie et séries chronologiques)
- ⇒ SPSS Categories (analyse des correspondances)

Fenêtres de SPSS

- ⇒ 'SPSS Data Editor' : tableur pour la saisie et la modification des données;
- ⇒ 'SPSS Viewer': la fenêtre de sortie (texte + graphique)
- ⇒ 'SPSS Syntax window': la fenêtre de syntaxe

Fenêtre de saisie

- ⇒ Entrer les données
- ⇒ Introduire et modifier les propriétés des variables



Fenêtre de sortie

- ⇒ Elle reprend les tableaux et les graphiques produits par les méthodes d'analyse de SPSS
- ⇒ On peut sauver son contenu au format SPO (SPSS Output Viewer) qui ne peut être relu que pas SPSS
- ⇒ On peut exporter la sortie au formats
 - ✓ DOC/RTF (figures incorporées)
 - ✓ html (figures séparées)
 - ✓ texte (sans figures)

Fenêtre de syntaxe

- ⇒ coller les commandes choisies dans les menus
- ⇒ modifier la syntaxe (notamment pour réaliser des fonctions qui ne sont pas disponibles par les menus)
- ⇒ exécuter
- ⇒ sauver

Utilisation approfondie des logiciels statistique, G.Mélard

13

Introduction
à SPSS

Programme pour SPSS

```
commandes de définition
commande de description
commandes de transformation ou de sélection
données
commandes de transformation ou de sélection
procédure
commandes de transformation ou de sélection
procédure
etc.
```

Utilisation approfondie des logiciels statistique, G.Mélard

14

Introduction
à SPSS

Sujets

- ⇒ Variable
- ⇒ Cas
- ⇒ Sélection de cas et de variables
- ⇒ Commandes de SPSS
- ⇒ Sous-commande
- ⇒ Conventions
- ⇒ Données

Utilisation approfondie des logiciels statistique, G.Mélard

15

Introduction
à SPSS

Variable

- ⇒ caractérisée par un nom (1 à 8 caractères)
- ⇒ noté var
- ⇒ précédé de #: variable provisoire
- ⇒ précédé de \$: variable système (ex.: \$casenum)

Utilisation approfondie des logiciels statistique, G.Mélard

16

Introduction
à SPSS

Cas

- ⇒ indice de 1 au nombre total de cas
- ⇒ maintenu automatiquement dans une variable \$casenum
- ⇒ données sur 1/plusieurs lignes consécutives
- ⇒ valeurs d'une variable pour différents cas toujours situées de façon identique (même ligne, même colonne)
- ⇒ possibilité d'entrer les données en format libre

Utilisation approfondie des logiciels statistique, G.Mélard

17

Introduction
à SPSS

Sélection de cas et de variables

- ⇒ les cas qui interviennent dans une procédure statistique sont déterminés au moyen des commandes de sélection
- ⇒ les variables qui interviennent dans une procédure statistique mentionnées dans une liste de variables, notées varlist
- ⇒ mot-clé to pour signifier un intervalle
- ⇒ par exemple 4 to 6: équivalent à 4,5,6

Utilisation approfondie des logiciels statistique, G.Mélard

18

Introduction
à SPSS

Commandes de SPSS

- ⇒ Chacune doit se terminer par un point décimal (.)
- ⇒ Commandes sont divisées en deux zones :
 - ✓ le mot-clé (à reproduire tel quel en majuscules ou en minuscules)
 - ✓ les spécifications (qui dépendent de la commande)

Sous-commande

- ⇒ Certaines commandes comportent des sous-commandes séparées par /
- ⇒ Chaque sous-commande commence par un mot-clé: *variable, print, plot, etc.*
- ⇒ Le mot-clé est parfois suivi d'un signe =.
- ⇒ Quelques sous-commandes sont obligatoires (*var* pour spécifier la liste des variables sur lesquelles porte l'analyse)
- ⇒ Des valeurs par défaut sont disponibles pour la plupart des sous-commandes

Conventions

- ⇒ les parties entre [] sont optionnelles
- ⇒ les parties entre { } sont l'objet d'un choix
- ⇒ le premier spécifié est le choix par défaut
- ⇒ les mots en italiques sont à remplacer par des symboles particuliers: *var* = varlist
- ⇒ exemple: varlist doit être remplacé par une liste de noms de variables: *var = aa,ab,ae*
- ⇒ la présentation des commandes est libre
- ⇒ respecter les séparateurs: *, espace / () = etc.*

Données

- ⇒ présentation des données: spécifiée dans la commande *data list*
- ⇒ doivent être précédé de la commande suivante *begin data*
- ⇒ peuvent être suivies de la commande : *end data*
- ⇒ Remarques
 - ✓ *end data* doit commencer en colonne 1
 - ✓ pas abrégée et pas plus d'un espace entre *end* et *data*

Exemple (1)

```
TITLE 'ENZYMES DANS LA POLYARTHRITE' .
DATA LIST FIXED/
    GROUPE 1, AGE 2-3, TANTINFL 4,
    TSTEROI 5, SEVER 6, PCLYMPH 7-8,
    PCOKT4 9-10, PCOKT8 11-12, FNU 13-15,
    ADA 16-19, PNP 20-24 .
VAR LABELS TANTINFL,TRAITEMENT ANTI INFLAMMATOIRE/
    TSTEROI ,TRAITEMENT STEROIDES/ .
MISSING VALUES PCLYMPH,PCOKT4,PCOKT8(0)/
    FNU(999)/ .
N OF CASES 124 .
IF (GROUPE = 1 OR GROUPE = 3) LOGADA = LG10(ADA) .
IF (GROUPE = 1 OR GROUPE = 4) INVPNP = -1000000/PNP .
BEGIN DATA
```

Exemple (2)

```
BEGIN DATA
158114755242166 582 7260
...
4 18300
END DATA .
DESCRIPTIVES PCLYMPH,PCOKT4,PCOKT8 / STATISTICS ALL .
FREQUENCIES SEVER / STATISTICS ALL .
CROSSTABS TABLES = TANTINFL BY TSTEROI /
    SEVER BY TANTINFL, TSTEROI / STATISTICS =
CHISQ .
MEANS TABLES = FNU, ADA, PNP, LOGADA, INVPNP BY GROUPE .
CORRELATIONS AGE, ADA, PNP WITH AGE .
NONPAR CORR AGE, ADA, PNP WITH AGE .
```

Exemple (3)

```
T-TEST      GROUPS = GROUPE(1,2) / VARIABLES = FNU .
T-TEST      GROUPS = GROUPE(1,3) / VARIABLES = ADA, LOGADA .
T-TEST      GROUPS = GROUPE(1,4) / VARIABLES = PNP, INVPNP .
NPAR TESTS  K-S = FNU BY GROUPE(1,2) /
            K-S = ADA BY GROUPE(1,3) /
            K-S = PNP BY GROUPE(1,4) .
NPAR TESTS  M-W=FNU BY GROUPE(1,2) /
            M-W=ADA, LOGADA BY GROUPE(1,3) /
            M-W=PNP, INVPNP BY GROUPE(1,4) / .
T-TEST      PAIRS = PCOKT4 WITH PCOKT8 .
NPAR TESTS  WILCOXON = PCOKT4 WITH PCOKT8 .
FINISH .
```

Les types de menus

- ⇒ Base
- ⇒ Viewer
- ⇒ Graph
- ⇒ Syntax

Les menus de base

- ⇒ File
- ⇒ Edit
- ⇒ View
- ⇒ Data
- ⇒ Transform
- ⇒ Analyze
- ⇒ Graph
- ⇒ Utilities
- ⇒ Window
- ⇒ Help

Menu File

- ⇒ New
- ⇒ Open
- ⇒ Read Text Data données en format fixe ou libre
- ⇒ Close un fichier de données
- ⇒ Save et Save as
- ⇒ Export
- ⇒ Display Data Info
- ⇒ Print
- ⇒ Printer Preview
- ⇒ Stop Processor (Ctrl C)
- ⇒ Exit pour quitter SPSS

New

pour ouvrir un nouveau fichier, de type ...

- ⇒ Data données
- ⇒ Syntax syntaxe
- ⇒ Output sortie

Open

pour ouvrir un fichier existant, de type ...

- ⇒ Data données SPSS, Excel, Lotus 1-2-3, ou dBase
- ⇒ Syntax syntaxe
- ⇒ Output sortie

Syntaxe correspondante

`data list [file=file] [{fixed records=n}] spécif
 {free}
 {list}`

- ⇒ *file* spécifie le nom d'un fichier externe (sinon, les données suivent *begin data*)
- ⇒ *fixed*: format fixe (défaut)
- ⇒ *free* ou *list* indiquent un format libre où les données sont séparées par *espace* ou ,
- ⇒ *n* =nombre de lignes par cas (défaut: 1)
- ⇒ spécifications dépend du nombre de lignes par cas

Utilisation approfondie des logiciels statistique, G.Mélard 31 Introduction à SPSS

Une ligne par cas

On peut combiner les formes suivantes :

- ⇒ var interv
- où var est un nom de variable
- interv est col1-col2
- col1, col2: numéros de colonnes

Exemple : `data list xyz 26-28`

- ⇒ varlist interv

Exemple : `data list ab,ad,ae 36-41`

Utilisation approfondie des logiciels statistique, G.Mélard 32 Introduction à SPSS

Open Database

pour ouvrir un fichier existant, de type base de données...

- ✓ Oracle base de données Oracle
- ✓ SQL Server base de données SQL Server
- ⇒ New Query
- ⇒ Edit Query
- ⇒ Run Query

Utilisation approfondie des logiciels statistique, G.Mélard 33 Introduction à SPSS

Save

- ⇒ Data pour sauvegarder un fichier de données
 - ◆ formats acceptés: SPSS, SPSS portable, ASCII ou texte,
 - ◆ Excel, Lotus 1-2-3, dBase (extension .SAV, .POR, .DAT, XLS, WK1, DBF, respectivement)
- ⇒ Output pour sauvegarder le contenu d'une fenêtre de sortie (extension .SPO)
- ⇒ Syntax pour sauvegarder le contenu d'une fenêtre de syntaxe (extension .SPS)

Utilisation approfondie des logiciels statistique, G.Mélard 34 Introduction à SPSS

Export

- ⇒ On peut exporter les sorties aux formats
 - ✓ DOC/RTF (figures incorporées)
 - ✓ html (figures séparées)
 - ✓ texte (sans figures)
- ⇒ On peut exporter les graphiques aux formats JPG, EPS

Utilisation approfondie des logiciels statistique, G.Mélard 35 Introduction à SPSS

Menu Edit

- ⇒ Undo
- ⇒ Cut
- ⇒ Copy
- ⇒ Copy Object
- ⇒ Paste

Utilisation approfondie des logiciels statistique, G.Mélard 36 Introduction à SPSS

Menu Data

- ⇒ Define Variables
- ⇒ Templates
- ⇒ Insert Variables
- ⇒ Insert Cases
- ⇒ Goto Case
- ⇒ Sort Cases

Utilisation approfondie des logiciels statistique, G.Mélard

37

Introduction
à SPSS

Menu Data (suite)

- ⇒ Transpose
- ⇒ Merge files
- ⇒ Aggregate
- ⇒ Split File
- ⇒ Select Cases
- ⇒ Weight Cases

Utilisation approfondie des logiciels statistique, G.Mélard

38

Introduction
à SPSS

Define Variables

pour définir ou redéfinir une variable (type, description, valeurs manquantes, format)

Utilisation approfondie des logiciels statistique, G.Mélard

39

Introduction
à SPSS

N.B. Valeurs manquantes (missing values)

- ⇒ L'effet est le suivant : les procédures statistiques ne porteront pas sur les cas pour lesquels une des variables utilisées prend comme valeur une des valeurs manquantes
- ⇒ Il y a moyen de contourner cette règle

Utilisation approfondie des logiciels statistique, G.Mélard

40

Introduction
à SPSS

Select Cases

pour sélectionner des cas

- ⇒ Syntaxe: *select if* (logexpr)
- ⇒ Les procédures statistiques et les transformations qui suivent seront effectuées seulement sur les cas pour lesquels logexpr est une expression logique qui est vraie
- ⇒ L'effet d'un *select if* ultérieur est de sélectionner parmi les cas déjà retenus

Utilisation approfondie des logiciels statistique, G.Mélard

41

Introduction
à SPSS

Menu Transform

- ⇒ Compute
- ⇒ Random number seed
- ⇒ Count Occurrences
- ⇒ Recode
- ⇒ Rank cases
- ⇒ Automatic Recode
- ⇒ Run Pending Transform

Utilisation approfondie des logiciels statistique, G.Mélard

42

Introduction
à SPSS

Compute

pour appliquer une expression algébrique

Syntaxe: `compute var = arithmexpr`

⇒ var peut être une nouvelle variable

⇒ arithmexpr expression arithmétique formée à partir

- ✓ constantes et noms de variables existantes
- ✓ opérateurs + - * / ** (exponentiation)
- ✓ parenthèses (et)
- ✓ fonctions, sous la forme f(argument)

Fonctions

- ◆ sqrt racine carrée
- ◆ ln logarithme népérien
- ◆ lg10 logarithme en base 10
- ◆ exp exponentielle en base e
- ◆ rnd arrondi à l'entier le plus proche
- ◆ abs valeur absolue
- ◆ trunc troncation au plus grand entier contenu
- ◆ mod reste de la division par le 2e argument
- ◆ lag valeur pour le cas précédent
- ◆ normal
- ◆ uniform pour générer des données artificielles

Forme conditionnelle

Syntaxe: `if (logexpr) var = arithmexpr`

Pour tout cas pour lequel l'expression logique logexpr est vraie, la variable var reçoit une valeur. Sinon, elle garde sa valeur

Ex. `if (ab > ac or not(ab > ad)) xyz = 1`

Recode

⇒ into Same Variables

- ◆ pour recodifier dans les mêmes variables

⇒ into Different Variables

- ◆ pour recodifier dans de nouvelles variables

Syntaxe de Recode (1)

`recode varlist-1 (subst-1-1)
(subst-1-2) [into varlist]
etc.
[/ varlist-2 (subst-2-1) etc.]`

⇒ Les variables de chaque liste sont affectées par les substitutions

⇒ subst est de la forme:

anciennes valeurs=nouvelle valeur

Syntaxe de Recode (2)

⇒ anciennes valeurs peut être :

- ✓ une liste de valeurs numériques (ex.: 6,7,8,9)
- ✓ un intervalle de valeurs (ex.: 6 thru 9)
- ✓ un intervalle non borné inférieurement ou supérieurement
- ✓ missing
- ✓ `systemis` (valeur manquante système)

Exemples : `lowest thru 5 , 10 thru highest`

Syntaxe de Recode (3)

- ⇒ nouvelle valeur est une valeur numérique unique
 - ⇒ *into varlist*
 - ✓ en l'absence de ceci, les variables sont modifiées
 - ✓ en présence, elles ne sont pas modifiées mais les valeurs sont affectées aux variables de varlist
- Exemple: `recode ab ac (4=0) /`
`ab to ad (lowest thru -1 = -1)`

Menu Analyze (1)

- pour appliquer les procédures statistiques de même nom
- ⇒ Reports
- ⇒ Descriptive statistics
- ⇒ Compare Means
- ⇒ General Linear Models
- ⇒ Correlate
- ⇒ Regression
- ⇒ Loglinear

Menu Analyze (2)

- ⇒ Classify
- ⇒ Data reduction
- ⇒ Scale
- ⇒ Nonparametric tests
- ⇒ Time Series
- ⇒ Survival
- ⇒ Multiple Response

Descriptive Statistics

- ⇒ Frequencies
- ⇒ Descriptives
- ⇒ Explore
- ⇒ Crosstabs
- ⇒ List Cases
- ⇒ Report Summaries in Rows

Explore (1)

Syntaxe:

```
examine variables = varlist [by varlist [, var by var ] ]
[/compare = {groups}][/ {total} ] [/ id = {scasenum}]
{variables} {nototal} {var}
[/ plot = {none} ] [/ statistics = {descriptives} ]
{boxplot} {extremes}
{stemleaf} {all}
{histogram} {none}
{spreadlevel}
{npplot}
{all}
```

Explore (2)

- ⇒ Pour chaque variable dépendante,
- ⇒ Pour chaque valeur des variables de la liste "by varlist"
- ⇒ Pour chaque couple de valeurs des variables de "var by var"
 - ✓ statistiques descriptives
 - ✓ stem-and-leaf display
 - ✓ box-plot

Explore (3)

Sous-commandes

⇒ *nototal* supprime les diagrammes sur l'ensemble des cas.

⇒ *id* sert à identifier les cas

⇒ *plot* donne les diagrammes usuels, plus

✓ *spreadlevel(n)*: puissance n des étendues interquartiles en fonction des puissances n des médianes (n = 0: log en base e),

✓ droite de régression

✓ test d'homogénéité des variances de Levene

Explore (4)

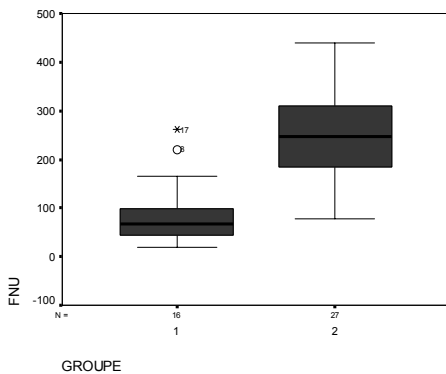
⇒ *npplot* donne

✓ normal probability plot

✓ detrended probability plot

✓ + tests de Shapiro-Wilk et de Kolmogorov-Lilliefors

N.B. *Examine* contient également des sous-commandes *frequencies*, *percentiles*, *measures* (Huber, Andrew, Hampel, Tukey).



Crosstabs

```

crosstabs tables = colvarlist-1 by linvarlist-1
              [ / colvarlist-2 by linvarlist-2 ]
              etc.
[ / statistics = {none} ] [ / cells = {count}
                          {chisq}      {row}
                          {bttau}     {column}
                          {eta}       {total}
                          {corr}      {none}
    
```

Crosstabs (suite)

Tables de contingence produites pour chaque couple de variables (une dans colvarlist et l' autre dans linvarlist). Les valeurs de la variable de colvarlist sont placées dans la première colonne et celles de linvarlist sur la première ligne. Ceci suppose que le nombre de valeurs distinctes est faible.

Crosstabs (suite)

⇒ Principales statistiques:

✓ *statistics = chisq* (chi carré d'indépendance ou d'homogénéité et autres tests (Mantel-Haenszel, Fisher exact 2 x 2, ...))

✓ *statistics = bttau* (coefficient de corrélation de rang de Kendall (tau b))

✓ *statistics = eta* : rapport de corrélation

✓ *statistics = corr* : coefficient de corrélation (+Pearson et Spearman)

✓ N.B. Autres statistiques disponibles: *phi*, *cc*, *lambda*, *uc*, *ctau*, *gamma*, *d*, *kappa*, *risk*, *all*

Crosstabs (fin)

⇒ Les options les plus utiles sont :

- ✓ *cells = row* :
pourcentages de ligne
- ✓ *cells = column* :
pourcentages de colonne
- ✓ *cells = total* :
pourcentages de total
- ✓ *cells = none* :
suppression des tables

Compare Means

- ⇒ Means
- ⇒ One-Sample T Test
- ⇒ Independent-Samples T-Test
- ⇒ Paired-Samples T-Test
- ⇒ One-Way ANOVA

Means

means tables = depvar by indvarlist

[/statistics = {none}]
{anova}
{linearity}
{all}

Effectue une analyse de la variance à un facteur contrôlé (comparaison de plusieurs moyennes) + test de linearité basé sur le rapport de corrélation, sur demande

Independent Samples T-Test

t-test groups = groupspecif /variables = varlist

Test de comparaison de 2 moyennes effectué pour chaque variable de varlist

Les 2 échantillons sont définis comme suit :

- ⇒ si *groupspecif* de la forme *var(val)*,
 - ✓ groupe 1: tous les cas pour lesquels *var* ≥ *val*
 - ✓ groupe 2: tous les autres cas
- ⇒ si *groupspecif* de la forme *var(val-1, val-2)*
 - ✓ groupe 1: tous les cas pour lesquels *var* = *val-1*,
 - ✓ groupe 2: ceux pour lesquels *var* = *val-2*, les autres cas étant ignorés

Paired Samples T-Test

t-test pairs =
varlist-1-1 [with varlist-2-1 [(paired)]]
[/varlist-1-2 [with varlist-2-2 [(paired)]]]
etc.

Comparaison des moyennes des paires de variables, à partir du même échantillon
Chaque variable de *varlist-1* est comparée avec chaque variable de *varlist-2*
Option (*paired*): comparaison des moyennes de la *i*-ème variable de *varlist-1* et de la *i*-ème variable de *varlist-2*

General Linear Models (1)

- ⇒ Univariate analysis of variance (ANOVA)
- ⇒ Multivariate analysis of variance (MANOVA)
- ⇒ Repeated Measures
- ⇒ Variance Component

General Linear Models (2)

- ⇒ Simple Factorial (= comparaison de plusieurs moyennes): plans classiques y compris l'analyse de la covariance
- ⇒ General Factorial: contrastes, comparaisons multiples, plans à mesures répétées (split plot, etc.)
- ⇒ Analyse de variance multivariée (MANOVA) at analyse de mesures répétées

Correlate

- ⇒ Bivariate
- ⇒ Partial (corrélation partielle)
- ⇒ Distance
 - ✓ entre variables, entre cas
 - ✓ similarités, dissimilarités
 - ✓ euclidienne, Manhattan, etc.

Correlate bivariate

```
correlations varlist-1 with varlist-2
  [/ print = {twotail} ] [/ format = {matrix} ]
  {onetail}             {serial}
```

Calcul du coefficient de corrélation totale pour chaque couple de variables obtenu en prenant une variable dans varlist-1 et une autre dans varlist-2 + probabilité de signification (test bilatéral)

Options :

- ⇒ *print = onetail* : test unilatéral
- ⇒ *format = serial* : 1 fois chaque coefficient

Syntaxe de nonpar corr

```
nonpar corr varlist-1 with varlist-2
  [/print={twotail}{spearman}][[/format={matrix}]]
  {onetail}{kendall}          {serial}
  {both}
```

- ⇒ Corrélation de rang de Spearman
- ⇒ Probabilité de signification du test bilatéral
- ⇒ Options :

- ✓ *print = onetail* : test unilatéral
- ✓ *print = kendall* : tau b au lieu de Spearman
- ✓ *print = both* : tous les 2 sont calculés
- ✓ *format = serial* : 1 fois chaque coefficient.

Regression

- ⇒ Linear
- ⇒ Curve Estimation
- ⇒ Binary Logistic
- ⇒ Multinomial Logistic
- ⇒ Ordinal
- ⇒ Probit
- ⇒ Non Linear : régression non linéaire
- ⇒ Weighted Estimation
- ⇒ 2-Stage Least Squares

Regression linéaire

Syntaxe: *regression* liste-de-sous-commandes

Exemple: régression de y en x1, x2 et x3 : *regression variables = y, x1, x2, x3*

/ dependent = y

/ enter

+ graphe des résidus + test de Durbin-Watson (cas de données chronologiques): ajouter

/ residuals = durbin

/ casewise = defaults, all

Les types de sous-commandes

- ⇒ Spécifications de l'analyse décrites par une liste de sous-commandes
- ⇒ Quelques sous-commandes sont obligatoires
- ⇒ Sous-commandes regroupées en 4 classes:
 - classe 1 : les variables et les cas
 - classe 2 : le modèle
 - classe 3 : la méthode d'élaboration du modèle
 - classe 4 : l'analyse des résidus (facultative)

Utilisation approfondie des logiciels statistique, G.Mélard 73 Introduction à SPSS

Classe 1 : les variables et les cas

- 1.1 Les statistiques descriptives désirées
 - ⇒ Défaut : néant
 - ⇒ Exemple : `/descriptives`
- 1.2 Le traitement des valeurs manquantes
 - ⇒ Défaut : les cas avec au moins une valeur manquante sont exclus de l'analyse
- 1.3 La sélection des cas dans l'estimation des paramètres
 - ⇒ Défaut : tous les cas sont utilisés
 - ⇒ Exemple : `/select = $casenum <= 30`

Utilisation approfondie des logiciels statistique, G.Mélard 74 Introduction à SPSS

Classe 1 (suite)

- 1.4 La largeur des sorties
 - ⇒ Défaut : 132 caractères
 - ⇒ Exemple : `/width = 78` (sortie sur 78 col.)
- 1.5 La liste des variables (obligatoire)
 - ⇒ `/variables = varlist`

Utilisation approfondie des logiciels statistique, G.Mélard 75 Introduction à SPSS

Classe 2. Le modèle

- 2.1 Le critère statistique éventuel de la méthode d'élaboration
 - ⇒ Défaut : `pin = 0,05` , `pout = 0,10` , `tolerance = 0,0001`
 - ⇒ Exemple :
`/criteria = pin(0.05), pout(0.20)`

Utilisation approfondie des logiciels statistique, G.Mélard 76 Introduction à SPSS

Classe 2 (suite)

- 2.2 Les statistiques de régression
 - ⇒ Défaut: R^2 , tableau d'analyse de variance, coefficients avec erreurs-types et statistiques t , coefficients réduits ('beta') et coefficients des variables hors équation
 - ⇒ Exemple :
`/statistics = defaults,cha,bcov,zpp,ci,history`

Utilisation approfondie des logiciels statistique, G.Mélard 77 Introduction à SPSS

Classe 2 (suite)

- En plus des statistiques par défaut :
- ✓ changements de R carré d'une étape à l'autre (*cha*)
 - ✓ covariances et corrélations entre coefficients de régression (*bcov*)
 - ✓ coefficients de corrélation totale et partielle (*zpp*)
 - ✓ intervalles de confiance à 95% pour les coefficients de régression (*ci*)
 - ✓ informations après chaque étape (*history*)

Utilisation approfondie des logiciels statistique, G.Mélard 78 Introduction à SPSS

Classe 2 (fin)

2.3 Omission de la constante de régression

⇒ Défaut : régression avec constante

⇒ Exemple : `/origin`

2.4 La variable dépendante (obligatoire)

`/dependent = varlist`

⇒ Noms de variables de varlist doivent appartenir à la liste `variables =`

⇒ Si varlist a plusieurs variables: chacune tour à tour la variable dépendante

Classe 3. Méthode d'élaboration du modèle

⇒ Obligatoire

⇒ On peut choisir une ou plusieurs des méthodes ci-dessous. Si plusieurs méthodes sont appliquées, elles cumulent leurs effets:

`/enter`

(toutes les variables de `variables =` sauf la variable dépendante sont introduites comme variables explicatives)

Classe 3 (suite)

⇒ Autres méthodes :

- `stepwise /stepwise = varlist`

(pas à pas : examen d'entrée ou de sortie)

- `forward /forward = varlist`
(inclusion progressive avec examen d'entrée)

- `backward /backward`
(toutes les variables sont introduites; élimination régressive par examen de sortie)

- `remove /remove = varlist`

(sortie forcée des variables)

- `test /test = (varlist-1)(varlist-2)...`

⇒ varlist omis sauf pour remove et test

Classe 3: Exemples

1) `/enter = d1 to d11/enter temps, temps2`

2) `/stepwise`

3) `/forward`

4) `/enter d1 to d11 /forward`

5) `/test = (cos12,sin12)(cos6,sin6)(cos4,sin4)`
`(cos3,sin3)(cos2p4,sin2p4)(cos2)`

6) `/test = (d1 to d11)(temps, temps2)`

⇒ Problèmes de multicollinéarité évités

Classe 4. L'analyse des résidus

⇒ Facultatif

⇒ On peut créer des variables temporaires :

✓ `pred` (valeurs prédites)

✓ `resid` (résidus)

✓ `zresid` (réduits, / écart-type résiduel)

✓ `sresid` (studentisés, / écart-type propre)

✓ `dresid` (après suppression du cas correspondant dans le modèle),

✓ `sdresid` (idem studentisé),

✓ `cook` (statistiques de Cook), etc.

Classe 4 (suite)

4.1 Spécification des analyses

⇒ Défaut : néant

⇒ Exemple : `/residuals`

(histogramme, graphique en probabilités cumulées normales de `zresid`, 10 valeurs les plus extrêmes de `zresid` et le test de Durbin-Watson)

⇒ Exemple:

`/residuals = durbin, outliers(sresid), id(date)`

Classe 4 (suite)

4.2 Graphes en fonction de l'indice des cas

⇒ Défaut : néant

⇒ Exemple : `/ casewise = defaults, all`
 (pour avoir un graphe de tous les cas)

4.3 Graphes à deux dimensions

⇒ Défaut : néant

⇒ Exemple: `/ scatterplot = (*pred,*resid) (x,y)`
 (produit deux graphes : résidus en fonction des prévisions et y en fonction de x)

Ordre des sous-commandes

⇒ Les 4 classes peuvent être emboîtées les une dans les autres. Exemple :

```

c11 c12 c13 c14 c12 c13 c14 c12 c13 c11 c12 c13
<.....> <.....> <.....> <.....> Classe 1
<.....> <.....> <.....> <.....> Classe 2
<.....> <.....> <.....> <.....> Classe 3
<.....> <.....> <.....> <.....> Classe 4
    
```

Loglinear

⇒ General

⇒ Hierarchical

⇒ Logit

Classify

⇒ Two Step Cluster

⇒ K-means cluster

⇒ Hierarchical cluster

⇒ Discriminant

Data reduction

⇒ Factor analysis

⇒ Correspondance Analysis

⇒ Optimal Scaling

Scale

⇒ Reliability analysis

⇒ Multidimensional scaling

Nonparametric tests

- ⇒ Chi-Square
- ⇒ Binomial
- ⇒ Runs (séquences)
- ⇒ 1-Sample K-S Kolmogorov-Smirnov
- ⇒ 2 Independent Samples
- ⇒ K Independent Samples
- ⇒ 2 Related Samples
- ⇒ K Related Samples

Utilisation approfondie des logiciels statistique, G.Mélard 91 Introduction à SPSS

Nonparametric tests: syntax

- Syntaxe: *npar tests* [nom de test(paramètre)=varlist(paramètre)]
[/nom de test ...]
- ⇒ Le test désiré est indiqué par un mot-clé
 - ⇒ Pour certains tests, un paramètre supplémentaire précise le test à effectuer

Utilisation approfondie des logiciels statistique, G.Mélard 92 Introduction à SPSS

Tests non paramétriques (1)

- ✓ *binomial* test binomial
- ✓ *chisquare* test chi carré d'ajustement (1 échantillon)
- ✓ *k-s* tests de Kolmogorov-Smirnov 1 et 2 échantillons
- ✓ *runs* test des séquences 1 échantillon
- ✓ *mcnemar* test de McNemar de comparaison de proportions
- ✓ *sign* test du signe
- ✓ *wilcoxon* test de Wilcoxon
- ✓ *cochran* test Q de Cochran

Utilisation approfondie des logiciels statistique, G.Mélard 93 Introduction à SPSS

Tests non paramétriques (2)

- ✓ *friedman* test de Friedman (Anova à deux facteurs contrôlés)
- ✓ *kendall* coefficient de concordance W de Kendall
- ✓ *median* test de la médiane
- ✓ *m-w* test de Mann-Whitney
- ✓ *w-w* test de Wald-Wolfowitz
- ✓ *moses* test de Moses des réactions extrêmes
- ✓ *k-w* test de Kruskal-Wallis (Anova à 1 facteur contrôlé)

N.B. : *npar corr* contient les tests de corrélation de rangs de Kendall et de Spearman

Utilisation approfondie des logiciels statistique, G.Mélard 94 Introduction à SPSS

Exemples (1)

- ⇒ *binomial*(p) = varlist (valeur-1 , valeur-2)
 - ⇒ *chisquare* = varlist (min , max) / *expected* = p1,...,pj , où j = max-min+1
 - ⇒ *k-s* (loi,paramètres) = varlist
 - ✓ loi :
 - uniform* paramètres: borne infér, borne supér.
 - normal* moyenne, écart-type
 - poisson* moyenne
- Les paramètres omis sont estimés à partir des données

Utilisation approfondie des logiciels statistique, G.Mélard 95 Introduction à SPSS

Exemples (2)

- ⇒ *runs* (paramètre) = varlist
- paramètre définit la dichotomisation : *mean*, *median*, *mode* ou une valeur numérique spécifiée
- ⇒ *wilcoxon* = varlist-1 with varlist-2
- ⇒ *m-w* = varlist by variable (valeur-1, valeur-2)
- ⇒ *k-s* = varlist by variable (valeur-1 , valeur-2)
- idem pour le test de Kolmogorov-Smirnov 2 échantillons

Utilisation approfondie des logiciels statistique, G.Mélard 96 Introduction à SPSS

Time Series

- ⇒ Exponential Smoothing
- ⇒ Autoregression
- ⇒ ARIMA
- ⇒ Seasonal Decomposition

Utilisation approfondie des logiciels statistique, G.Mélard 97 Introduction à SPSS

Survival

- ⇒ Life tables
- ⇒ Kaplan-Meier
- ⇒ Cox regression
- ⇒ Cox with time-dependent covariates

Utilisation approfondie des logiciels statistique, G.Mélard 98 Introduction à SPSS

Multiple response

- ⇒ Define sets
- ⇒ Frequencies
- ⇒ Crosstabs

Utilisation approfondie des logiciels statistique, G.Mélard 99 Introduction à SPSS

Menu Graphs (1)

- ⇒ Gallery
- ⇒ Interactive
 - Bar, Dot, Line, Ribbon, Drop-Line, Area
 - Pie, Boxplot, Error Bar, Histogram, Scatterplot
- ⇒ Map
 - Range of Values, Graduated Symbol,
 - Dot Density, Individual Values,
 - Bar Chart, Pie Chart, Multiple Themes
- ⇒ graphiques ordinaires

Utilisation approfondie des logiciels statistique, G.Mélard 100 Introduction à SPSS

Menu Graphs (2)

- ⇒ Graphiques ordinaires
 - ✓ Bar
 - ✓ Line
 - ✓ Area
 - ✓ Pie
 - ✓ High-Low
 - ✓ Pareto
 - ✓ Control
 - ✓ Boxplot
 - ✓ Error Bar
 - ✓ Scatter
 - ✓ Histogram
 - ✓ P-P
 - ✓ Q-Q
 - ✓ Sequence
 - ✓ ROC Curve
 - ✓ Time Series
 - ◆ Autocorrelations
 - ◆ Cross-Correlations
 - ◆ Spectral

Utilisation approfondie des logiciels statistique, G.Mélard 101 Introduction à SPSS

Menu Utilities

- ⇒ Variables
- ⇒ File Info
- ⇒ Define Sets
- ⇒ Use Sets
- ⇒ Run Script
- ⇒ Menu Editor

Utilisation approfondie des logiciels statistique, G.Mélard 102 Introduction à SPSS

Menu Window

- ⇒ Minimize All Windows
- ⇒ Tile
- ⇒ Cascade
- ⇒ Icon Bar
- ⇒ Status Bar
- ⇒ (liste des fenêtres disponibles)

Menu Help

- ⇒ Topics table des matières
- ⇒ Tutorial comment employer SPSS
- ⇒ Case Studies
- ⇒ Syntax Guide
 syntaxe des instruction de SPSS (peut être copié et collé dans la
 fenêtre de syntaxe)
- ⇒ Statistics Coach
- ⇒ SPSS Home Page
- ⇒ About

Fin