

STAT-F-413 (ex STAT053) **Utilisation approfondie de logiciels statistiques**

Toufik ZAHAF et Guy M  LARD
Abdelkamel ALJ, assistant

2009-2010

D  tails sur l'organisation du cours STAT-F-413 (ex STAT053)

- ⇒ Organisation et examen
- ⇒ Horaire d  taill  
- ⇒ R  gles

Guy M  lard, Utilisation approfondie des logiciels statistiques

2

Repr  sentations
graphiques

Partie 1 **Repr  sentations graphiques**

Guy M  lard, U.L.B.
2009-2010
(gmelard@ulb.ac.be)

Objectif des deux le  ons

- ⇒ Elle a pour but de fournir une introduction    l'analyse exploratoire des donn  es
- ⇒ ... et en particulier les repr  sentations
 - ✓ par tableau
 - ✓ par graphique
- ⇒ Elle sert aussi de mise    niveau
 - ✓ rappels de statistique descriptive
 - ✓ rappels d'inf  rence statistique

Guy M  lard, Utilisation approfondie des logiciels statistiques

4

Repr  sentations
graphiques

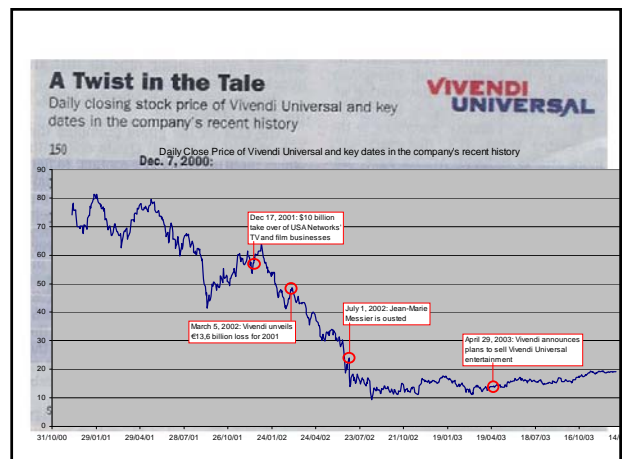
Contenu

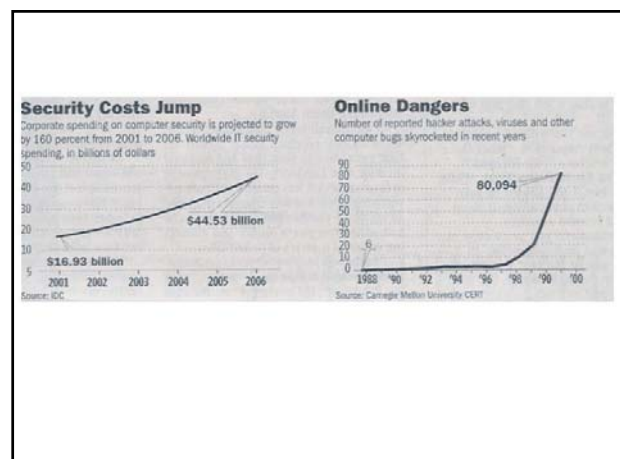
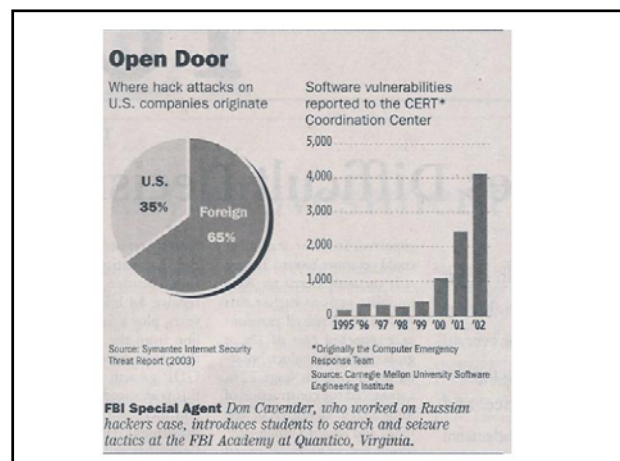
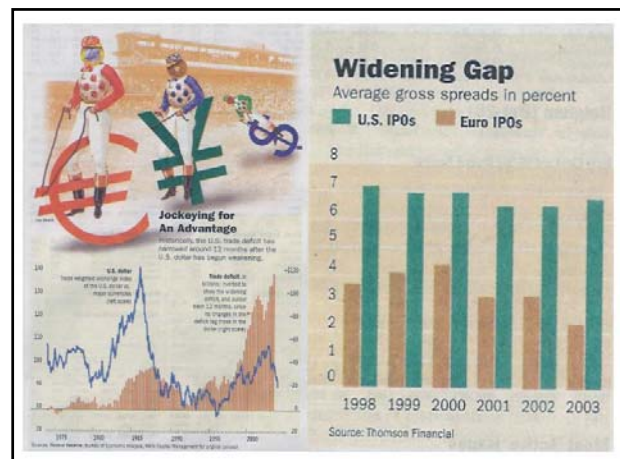
0. Introduction
 1. Statistique descriptive
 2. Tableaux statistiques
 3. Graphiques statistiques
 4. Tableaux et graphiques dynamiques
- Annexe. Quelques exemples de SAS/INSIGHT

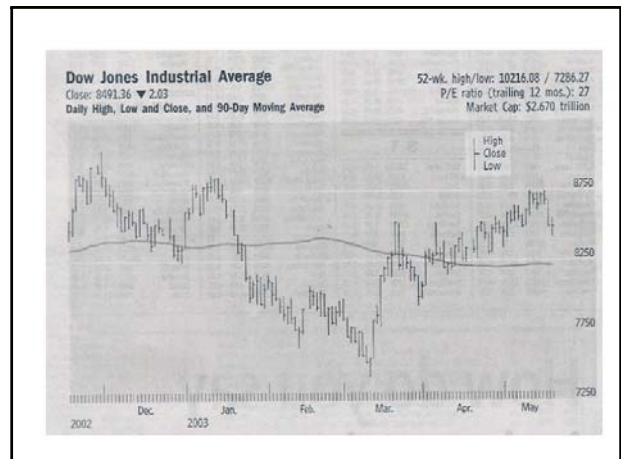
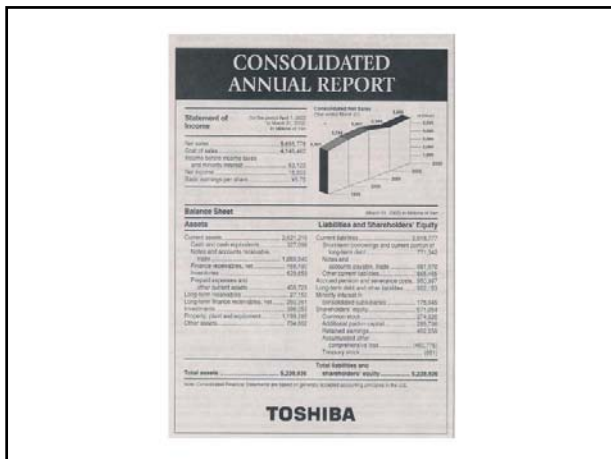
Guy M  lard, Utilisation approfondie des logiciels statistiques

5

Repr  sentations
graphiques







0. Introduction

- 0.1 Emplois de l'informatique en statistique
- 0.2 Acc  s    l'ordinateur
- 0.3 Notions de base
- 0.4 Positionnement de quelques progiciels statistiques

0.1 Emplois de l'informatique en statistique

(bas   sur un article de Pierre DAGNELIE, Biom  trie-Praxim  trie, vol. 15, 1975)

- ⇒ Types d'emploi de l'informatique en statistique
- ⇒ Evolution des mat  riels et logiciels
 - ✓ statistique administrative (recensements, enqu  tes)
 - ✓ statistique de recherche **CE COURS**
 - ✓ enseignement de la statistique
 - ✓ recherche statistique (simulation, ...) cfr Statistique informatique STAT068

Evolution des mat  riels et logiciels

- ⇒ Calculatrices scientifiques    fonctions statistiques
- ⇒ Ordinateurs    fonctions statistiques: disparus
- ⇒ Progiciels (= produits logiciels) statistiques Exemples: SPSS, BMDP, SAS, Minitab, Genstat, Glim, TSP
- ⇒ La plupart des progiciels: d  velopp  s pour grosses machines, transf  r  es +/- sur mini puis sur micro-ordinateurs
- ⇒ Probl  me de choix: Francis (1981)
- ⇒ Logiciels personnels sp  cifiques    une application cfr Statistique informatique STAT068

0.2 Acc  s    l'ordinateur

- ⇒ traitement par lots    distance ("remote batch processing")
- ⇒ traitement interactif
- ⇒ critique

Traitement par lots   distance ("remote batch processing")

- ⇒ pr  paration du travail dans un langage appropri  
- ⇒ introduction du travail   l'aide d'un  diteur
- ⇒ soumission du travail
- ⇒ r  cup  ration des r  sultats (liste d'erreurs):  cran de visualisation ou listage
- ⇒ interpr  tation

Guy M  lard, Utilisation approfondie des logiciels statistiques

19

Repr  sentations
graphiques

Traitement interactif

- ⇒ dialogue avec le progiciel statistique

Guy M  lard, Utilisation approfondie des logiciels statistiques

20

Repr  sentations
graphiques

Critique

- ⇒ Le traitement par lots est plus lent mais ...
permet de r  fl  chir
- ⇒ Le traitement interactif permet des essais (variations de m  thodes, graphiques) mais ...
on peut perdre la d  marche
- ⇒ Les deux modes sont co  teux :
 - ✓ traitement par lots : reprendre une partie du traitement (lecture, transformations, ...)
 - ✓ traitement interactif : co  t plus  lev   sauf PC
- ⇒ Dans les deux cas :
 - ✓ saisie des donn  es une seule fois et v  rifier
 - ✓ il faut conserver les r  sultats en ordre

Guy M  lard, Utilisation approfondie des logiciels statistiques

21

Repr  sentations
graphiques

0.3 Notions de base

- ⇒ Variables :
 - ✓ nombre : 1, 2, plusieurs
 - ✓ nature : qualitative, quantitative
(classification plus fine plus loin)
- ⇒ Observation ou cas : Peut  tre complexe   d  finir
(exemple: questionnaire avec r  ponses multiples)
- ⇒ Ensemble de donn  es :  chantillon al  atoire extrait d'une population (r  elle ou fictive) : une ou plusieurs populations
(dans ce dernier cas, possibilit   de comparaison)
- ⇒ Donn  es manquantes, donn  es aberrantes
- ⇒ Transformation : pour obtenir une relation lin  aire p. ex.

Guy M  lard, Utilisation approfondie des logiciels statistiques

22

Repr  sentations
graphiques

0.4 Positionnement de quelques progiciels statistiques

- ⇒ Tableurs: Lotus 1-2-3, Microsoft Excel: statistique de base uniquement
- ⇒ Gestionnaires de base de donn  es: Microsoft Access: pauvre
- ⇒ Logiciels statistiques (*=disponible ici; †=fut disponible):
 - ✓ SPSS ("Statistical Package for the Social Sciences") (*)
 - ✓ SAS ("Statistical Analysis System") (†)
 - ✓ Statistica (†)
 - ✓ Minitab
 - ✓ STATlab (†)
 - ✓ SPAD (†)
 - ✓ Systat (-)/Mystat (gratuit)

Guy M  lard, Utilisation approfondie des logiciels statistiques

23

Repr  sentations
graphiques

- ⇒ Logiciels  conom  triques: TSP, E-Views, RATS
- ⇒ Biblioth  ques scientifiques:
 - ✓ IMSL (International Math & Stat Library)
 - ✓ NAG (Numerical Algorithm Group)
- ⇒ Environnements interactifs de programmation statistique et  conom  trique:
 - ✓ Gauss
 - ✓ S-Plus /clone: R (gratuit)
 - ✓ MATLAB (*)

Guy M  lard, Utilisation approfondie des logiciels statistiques

24

Repr  sentations
graphiques

1. Statistique descriptive

- 1.1 Exemple :   tude clinique en rhumatologie: 22 polyarthritiques
- 1.2 Les diff  rents types de variables
- 1.3 La fonction de distribution (empirique)
- 1.4 Les principales caract  ristiques de position, de dispersion et de forme
- 1.5 Rappels d'inf  rence statistique
- 1.6 Mod  le statistique

Compl  ment : classification des proc  dures statistiques

Compl  ment : techniques multivari  es

Guy M  lard, Utilisation approfondie des logiciels statistiques

25

Repr  sentations
graphiques

1.1 Exemple:   tude clinique en rhumatologie

22 polyarthritiques

-      ge du patient (AGE)
-    s  v  rit   de l'affection (0    4: SEVER)
-    type de traitement
 -    anti-inflammatoires (TANTINFL) et/ou
 -    st  ro  des (TSTEROI)
-    dosages de 3 enzymes: 5'NU (FNU), ADA, PNP
-    pourcentages de lymphographie: LYMPHOST, OKT4, OKT8

+ 3 groupes de contr  le: dosage d'un seul enzyme pour chacun d'eux

Donn  es (SD-1a)

Guy M  lard, Utilisation approfondie des logiciels statistiques

26

Repr  sentations
graphiques

Questions

- 1. le dosage des enzymes est-il plus bas chez les arthritiques?
- 2. si oui, y-a-t-il une d  pendance entre ce dosage et ...
 - 2.1 l'  ge?
 - 2.2 le traitement?
 - 2.3 la s  v  rit   de l'affection?
 - 2.4 les trois pourcentages?

Guy M  lard, Utilisation approfondie des logiciels statistiques

27

Repr  sentations
graphiques

Nombre de variables

- 1: analyse    une variable ou univari  e ("univariate analysis")
- 2: analyse    deux variables ou bivari  e ("bivariate analysis")
- plusieurs: analyse    plusieurs variables ou multivari  e ("multivariate analysis")

Guy M  lard, Utilisation approfondie des logiciels statistiques

28

Repr  sentations
graphiques

Exemple

-    analyse univari  e (question 1)
-    analyses bivari  es (question 2)

Guy M  lard, Utilisation approfondie des logiciels statistiques

29

Repr  sentations
graphiques

1.2 Les diff  rents types de variables

-    Classification par nature: qualitative, quantitative
-    Classification plus fine bas  e sur l'  chelle de mesure
 -      chelle nominale
 -      chelle ordinale
 -      chelle d'intervalle
 -      chelle de rapport

Guy M  lard, Utilisation approfondie des logiciels statistiques

30

Repr  sentations
graphiques

Echelle NOMINALE

classification en K cat  gories
 exemple: traitement
 \Rightarrow variable qualitative
 cas particulier: DICHOTOMIQUE ($K = 2$)

Guy M  lard, Utilisation approfondie des logiciels statistiques

31

Repr  sentations
graphiques

Echelle ORDINALE

relation d'ordre "a inf  rieur ou   gal    b"
 pour tout $a, b: a \leq b$ ou $a \geq b$
 transitivit  : si $a \leq b$ et $b \leq c$, alors $a \leq c$
 antisym  trie: si $a \leq b$ et $b \leq a$, alors $a = b$
 exemple: s  v  rit  
 \Rightarrow variable qualitative

Guy M  lard, Utilisation approfondie des logiciels statistiques

32

Repr  sentations
graphiques

Echelle d'INTERVALLE

\Rightarrow variable quantitative
 exploiter les diff  rences
 exemple: temp  rature (degr  s Celsius)

Guy M  lard, Utilisation approfondie des logiciels statistiques

33

Repr  sentations
graphiques

Echelle de RAPPORTS

("scale")
 exploiter les rapports
 \Rightarrow z  ro absolu
 exemple: dur  e de vie

Guy M  lard, Utilisation approfondie des logiciels statistiques

34

Repr  sentations
graphiques

Echelles de mesure et op  rations (1/2)

R  gles

- \Rightarrow Ce qui est permis pour une   chelle l'est aussi pour les   chelles plus riches
- \Rightarrow Ce qui est interdit pour une   chelle l'est aussi pour les   chelles moins riches

Guy M  lard, Utilisation approfondie des logiciels statistiques

35

Repr  sentations
graphiques

Echelles de mesure et op  rations (2/2)

Echelle	permis	interdit
Nominale	- Coder par 1, 2, ...	- Effectuer des op��rations sur valeurs cod��es
Ordinale	- Ordonner - D��terminer les rangs - D��terminer la m��diane, les quantiles	- Calculer la moyenne - Utiliser des diff��rences
D'intervalle	- Calculer la moyenne - Utiliser des diff��rences	- Transformer en logarithmes - Utiliser des rapports ou des taux d'accroissement
De rapport	- Transformer en logarithmes - Utiliser des rapports ou des taux d'accroissement	

Exemple (SD-2a)

Guy M  lard, Utilisation approfondie des logiciels statistiques

36

Repr  sentations
graphiques

Remarques

1. Cette classification est controvers  e. Elle sert comme guide, sans plus.
2. Echelle ordinale + codage num  rique   chelle d'intervalle
3. Pr  cision des mesures?
 - ✓ Peu de diff  rence entre variable discr  te (comptage) et continue (autre)
 - ✓ En fait, diff  rence entre cas avec peu et beaucoup de valeurs possibles: tableaux, graphiques.
 - ✓ Encore que ... (ordinateur)

Guy M  lard, Utilisation approfondie des logiciels statistiques

37

Repr  sentations
graphiques

Transformations

But d'une transformation :

⇒ obtenir une distribution

✓ plus sym  trique,

✓ plus normale,

⇒ ou une relation entre variables plus simple   tudier (par exemple: lin  aire)

Guy M  lard, Utilisation approfondie des logiciels statistiques

38

Repr  sentations
graphiques

Soit X :   chelle de rapport
transformation de puissance (pour une variable strictement positive dans le cas o   $c = 0$)

$$\begin{cases} (X + c)^p, (p = 0.5, -1.2, \dots) \\ \log(X + c), \text{ si } p = 0 \end{cases}$$

Guy M  lard, Utilisation approfondie des logiciels statistiques

39

Repr  sentations
graphiques

Proportion $p \Rightarrow$ sym  trie entre p et $1 - p$
transformation logit:

$$\log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p)$$

(pr  serve la sym  trie)

Guy M  lard, Utilisation approfondie des logiciels statistiques

40

Repr  sentations
graphiques

Donn  es manquantes, aberrantes

⇒ Donn  es manquantes: pourquoi? Que faire? Effet?

⇒ Donn  es aberrantes: pourquoi? Que faire? Effet?

N.B. SPSS est le progiciel statistique qui prend le mieux en compte les donn  es manquantes; le probl  me des donn  es aberrantes n  cessite l'emploi de m  thodes statistiques robustes.

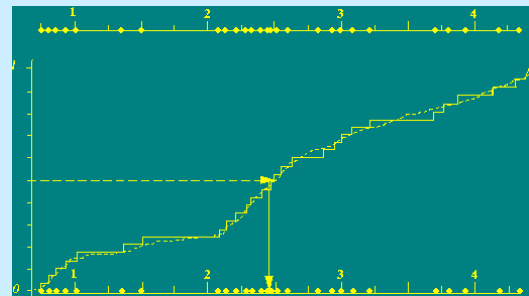
Guy M  lard, Utilisation approfondie des logiciels statistiques

41

Repr  sentations
graphiques

1.3 La fonction de distribution (empirique)

Construction (SD-4ae)



Dosage de 5'NU pour le groupe 2 (contr  le)

Guy M  lard, Utilisation approfondie des logiciels statistiques

42

Repr  sentations
graphiques

Remarques

1. Permet d'estimer directement les probabilit  s
2. Densit   = d  riv  e de la fonction de distribution id  alis  e
(il existe des m  thodes d'estimation de la densit  )
3. Utilisation d'un tableur (milieux des marches)

Guy M  lard, Utilisation approfondie des logiciels statistiques

43

Repr  sentations
graphiques

1.4 Principales caract  ristiques de position, de dispersion et de forme

- ⇒ la moyenne (arithm  tique) D  tails (SD5-a/o)
- ⇒ la m  diane
- ⇒ les quantiles (fractiles) d'ordre p ou    100 p %
- ⇒ la variance et l'  cart-type
- ⇒ autres mesures de dispersion :   tendue,   tendue interquartile
- ⇒ les mesures d'asym  trie (« skewness »)
- ⇒ les mesures d'aplatissement (« kurtosis »)

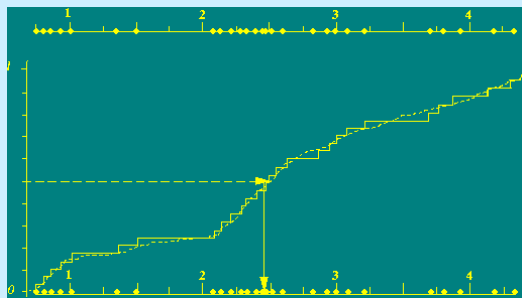
N.B. Les caract  ristiques d'asym  trie et d'aplatissement NE sont PAS utilisables lors d'une analyse exploratoire des donn  es (et ne le sont g  n  ralement pas non plus apr  s)
TOUJOURS INUTILE : sortie de SAS UNIVARIATE

Guy M  lard, Utilisation approfondie des logiciels statistiques

44

Repr  sentations
graphiques

D  termination d'un quantile

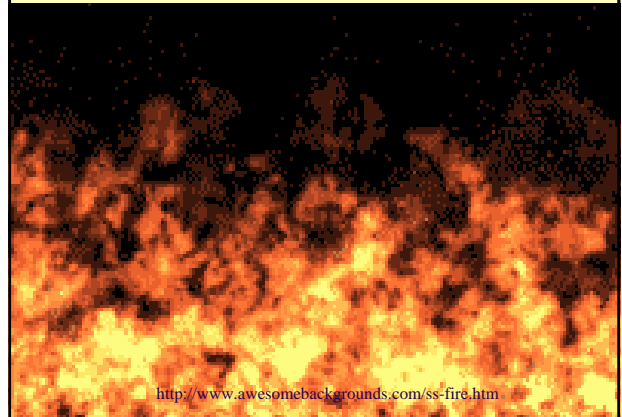


Guy M  lard, Utilisation approfondie des logiciels statistiques

45

Repr  sentations
graphiques

SAS Univariate



<http://www.awesomebackgrounds.com/ss-fire.htm>

A deux dimensions :

D  tails (SD5-A/X)

- ⇒ coefficient de r  gression
- ⇒ coefficient de cor  lation (Pearson)
- ⇒ variance et   cart-type r  siduels

Guy M  lard, Utilisation approfondie des logiciels statistiques

47

Repr  sentations
graphiques

Cas de variables quantitatives

- ⇒    deux dimensions: coefficient de r  gression, coefficient de cor  lation (de Pearson), variance et   cart-type r  siduels; coefficients de cor  lation de rangs
- ⇒    plus de deux dimensions: coefficient de r  gression, coefficient de cor  lation partielle et multiple, variance et   cart-type r  siduels

Guy M  lard, Utilisation approfondie des logiciels statistiques

48

Repr  sentations
graphiques

Cas de variables qualitatives

- ⇒    deux dimensions: mesures d'association
- ⇒    plus de deux dimensions

Guy M  lard, Utilisation approfondie des logiciels statistiques

49

Repr  sentations
graphiques

1.5 Rappels d'inf  rence statistique

Consid  rons l'exemple de la moyenne population μ

On distingue :

- ⇒ le probl  me d'estimation de μ
- ⇒ les propri  t  s de l'estimateur de μ
- ⇒ la construction d'un intervalle de confiance pour μ
- ⇒ le test d'une hypoth  se pour μ , par exemple $\mu = \mu_0$
- ⇒ les probl  mes de comparaison :
 - ✓ cas de 2 populations (comparaison de 2 moyennes)
 - ✓ cas de plusieurs population (analyse de variance)

Remarque : au lieu de la moyenne, on pourrait traiter l'exemple d'une proportion, de la variance ou encore de la r  gression

Guy M  lard, Utilisation approfondie des logiciels statistiques

50

Repr  sentations
graphiques

Estimation statistique

- ⇒ cas le plus simple: celui de la moyenne
- ⇒   quation du mod  le
- ⇒ approche des moindres carr  s
- ⇒ m  thode du maximum de vraisemblance
- ⇒ m  thode du quasi-maximum de vraisemblance

Guy M  lard, Utilisation approfondie des logiciels statistiques

51

Repr  sentations
graphiques

Propri  t  s de l'estimateur

- ⇒ propri  t  s exactes
- ⇒ propri  t  s asymptotiques

Guy M  lard, Utilisation approfondie des logiciels statistiques

52

Repr  sentations
graphiques

Inf  rence statistique

- ⇒ intervalle de confiance
- ⇒ test d'hypoth  se: pas de d  cision mais aide    la d  cision (probabilit   de signification)

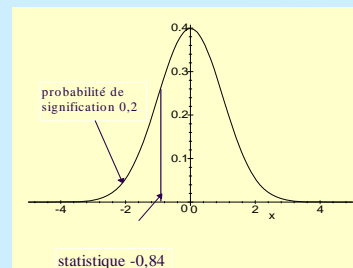
Guy M  lard, Utilisation approfondie des logiciels statistiques

53

Repr  sentations
graphiques

Probabilit   de signification (1)

- ⇒ "P-value"
- ⇒ D  finition: niveau de signification auquel les observations conduiraient de justesse au rejet de H



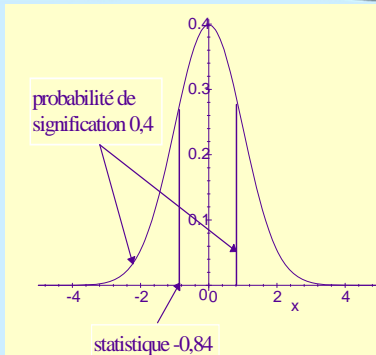
Guy M  lard, Utilisation approfondie des logiciels statistiques

54

Repr  sentations
graphiques

Probabilit   de signification (2)

⇒ Test bilat  ral
(  galit   contre
diff  rence):
doubler



Guy M  lard, Utilisation approfondie des logiciels statistiques

55

Repr  sentations
graphiques

Probabilit   de signification (3)

⇒ Avantages:

- ✓ pas besoin de sp  cifier le niveau de signification de l'utilisateur
- ✓ pas besoin de le communiquer au logiciel statistique
- ✓ r  ponse plus subtile: place    discussion

⇒ En pratique:

- ✓ si $p = 0.001$: pratiquement impossible d'accepter l'hypoth  se
- ✓ si $p = 0.023$: justifi   de rejeter l'hypoth  se (du moins au niveau de 5%)
- ✓ si $p = 0.118$: aucune raison de rejeter l'hypoth  se

Guy M  lard, Utilisation approfondie des logiciels statistiques

56

Repr  sentations
graphiques

1.6 Mod  les statistiques

⇒ d  finition du mod  le

⇒ r  alisme du mod  le

⇒ utilisation du mod  le

Exemple le plus simple de mod  le

$$X = \mu + e$$

avec des hypoth  ses sur le terme d'erreur e

Guy M  lard, Utilisation approfondie des logiciels statistiques

57

Repr  sentations
graphiques

D  finition de mod  le

⇒ description abstraite et grossi  re de la r  alit  

⇒ qui d  pend d'un petit nombre de param  tres

⇒ il est ajust   aux donn  es

⇒ tout se passe comme si les observations   taient obtenues par lui

⇒ mais pas dans les d  tails les plus fins (r  sidus) qui sont consid  r  s comme   tant de nature al  atoire

Guy M  lard, Utilisation approfondie des logiciels statistiques

58

Repr  sentations
graphiques

R  alisme du mod  le

⇒ En supposant en effet une variable al  atoire quelque part permet de se r  f  rer au raisonnement statistique mais il ne faut jamais oublier que ce n'est qu'une hypoth  se

⇒ Les conclusions vont d  pendre de la plausibilit   de cette hypoth  se

Guy M  lard, Utilisation approfondie des logiciels statistiques

59

Repr  sentations
graphiques

Utilisation du mod  le

⇒ estimation statistique

⇒ propri  t  s de l'estimateur

⇒ intervalle de confiance

⇒ test d'hypoth  se

Guy M  lard, Utilisation approfondie des logiciels statistiques

60

Repr  sentations
graphiques

Compl  ment : classification des proc  dures statistiques

- ✓ 1 variable, 1 population: moments, distribution, quantiles
- ✓ 1 variable, 2 populations
- ✓ 1 variable, K populations
- ✓ 2 variables, 1 population
- ✓ p variables, 1 population: techniques multivari  es

Guy M  lard, Utilisation approfondie des logiciels statistiques

61

Repr  sentations
graphiques

1 variable

- ⇒ 1 population: moments, distribution, quantiles
- ⇒ 2 populations
 - ✓ variables qualitatives
 - ◆ dichotomiques: comparaison de proportions
 - ◆ nominales: test χ^2
 - ◆ ordinales: test de Mann-Whitney
 - ✓ variables quantitatives
 - ◆ test t de Student
 - ◆ test d'  galit   de variances
- ⇒ K populations: cfr 2 populations sauf
 - ✓ variables ordinales: test de Kruskal-Wallis
 - ✓ variables quantitatives: ANOVA    1 facteur

Guy M  lard, Utilisation approfondie des logiciels statistiques

62

Repr  sentations
graphiques

2 variables

- ⇒ 1 population
 - ✓ variables qualitatives
 - ◆ test χ^2
 - ◆ mesures d'association
 - ✓ variables ordinales:
 - ◆ cor  lation de rangs (Spearman, Kendall)
 - ✓ variables d'intervalle ou de rapport:
 - ◆ cor  lation
 - ◆ r  gression

Guy M  lard, Utilisation approfondie des logiciels statistiques

63

Repr  sentations
graphiques

p variables

- ⇒ 1 population: techniques multivari  es

Guy M  lard, Utilisation approfondie des logiciels statistiques

64

Repr  sentations
graphiques

Compl  ment : techniques multivari  es

Technique	Nature	
	Var. D��pendante ou groupe 1	Var. explicatives ou groupe 2
R��gression lin��aire	1 quant.	k quant.
Analyse de la variance	1 quant.	k qual.
Analyse de la covariance	1 quant.	k_1 quant., k_2 qual.
R��gression lin��aire multivari��e	m quant.	k quant.
Analyse de corr��l. canonique		
Analyse de la variance mult.	m quant.	k qual.
Analyse de la covariance mult.	m quant.	k_1 quant., k_2 qual.
Analyse en composantes principales		k quant.
Analyse factorielle	k quant.	
Analyse discriminante	1 qual.	k quant.
Analyse de tableaux de contingence	k qual.	
Analyse logit	1 qual.	k qual.

Guy M  lard, Utilisation approfondie des logiciels statistiques

65

Repr  sentations
graphiques

2. Tableaux statistiques

- 2.1 Pr  sentation mat  rielle des tableaux
- 2.2 Objet du tableau
- 2.3 Types de tableaux

Guy M  lard, Utilisation approfondie des logiciels statistiques

66

Repr  sentations
graphiques

2.1 Pr  sentation mat  rielle des tableaux

⇒ Voyons d'abord quelques exemples

Les salaires en Italie (1970-1986)

Ann��e	Trimestre	Donn��e
1970	1	6.88E+02
1970	2	7.62E+02
1970	3	8.35E+02
1970	4	1008.2
1971	1	7.98E+02
1971	2	9.19E+02
1971	3	9.99E+02
1971	4	1163.4
1972	1	9.05E+02
1972	2	1053.8
1972	3	1137.8
1972	4	1353
1973	1	1046
1973	2	1283.4
1973	3	1435.7
1973	4	1643.1
1974	1	1315.4

Ann��e	Trimestre	Donn��e
1970	1	687.6
1970	2	761.6
1970	3	834.5
1970	4	1008.2
1971	1	797.7
1971	2	919.1
1971	3	999.1
1971	4	1163.4
1972	1	904.7
1972	2	1053.8
1972	3	1137.8
1972	4	1353.0
1973	1	1046.0
1973	2	1283.4
1973	3	1435.7
1973	4	1643.1
1974	1	1315.4

✓ "Sur base du tableau et du bon sens, pensez-vous que la s  rie soit annuelle, trimestrielle, mensuelle, autre?"
 ✓ Comparez les deux tableaux. Peut-on am  liorer le tableau de droite?"

Qualit   des tableaux (1/2)

⇒ Les tableaux fournissent souvent des informations utiles    condition que

- ✓ les nombres    comparer soient en colonnes
- ✓ seuls apparaissent les chiffres les plus significatifs
- ✓ une unit   appropri  e soit utilis  e
- ✓ la notation scientifique soit   vit  e

⇒ Mais ils ne remplacent pas les graphiques

Remarque.

95% des nombres des tableaux publi  s par la Banque Nationale de Belgique comprennent au maximum 5 chiffres significatifs.

Qualit   des tableaux (2/2)

⇒ Du point de vue de la lisibilit   et de l'esth  tique, on peut ajouter les condition suivantes :

- ✓   liminer l'inutile
- ✓ faciliter la comparaison dans les colonnes, pas dans les lignes (transposer le tableau au besoin)
- ✓ ordonner colonnes et lignes de mani  re quantitative, pas alphab  tique ni al  atoire
- ✓ regrouper les colonnes par des traits horizontaux et   viter les traits verticaux
- ✓ employer les couleurs avec mod  ration (notamment alternance de deux couleurs une ligne sur deux)

Exemple (TS1-a/b)

2.2 Objet du tableau

⇒ source de donn  es? de moins en moins (bases de donn  es)

⇒ moyen de communication? faciliter celle-ci

2.3 Types de tableaux

⇒ Tableaux de contingence

⇒ Tableaux de caract  ristiques

⇒ Tables de d  cision

Tableaux de contingence

- ⇒ 1, 2, 3, ... dimensions Exemple (TS2-a/b)
- ⇒ parfois de tr  s grande taille (p. ex. SPSS)
- ⇒ effectifs ou fr  quences ou pourcentages
- ⇒ + marges
- ⇒ dans le cas de plus de 2 dimensions, utilit   de pouvoir faire pivoter les dimensions dynamiquement (p. ex. Excel)

Guy M  lard, Utilisation approfondie des logiciels statistiques

73

Repr  sentations
graphiques

T��l��r By T��l��r		T��l��r	
		0	1
Count	7	8	5
Total %	31.62	36.36	68.18
Row %	46.87	53.33	
Col %	100.00	53.33	
Cell Chi^2	1.0394	0.4691	
T��l��r	0	7	7
1	0.00	31.62	31.62
	0.00	100.00	
	0.00	46.87	
	2.2273	1.0394	
	7	15	22
	31.62	68.18	

Guy M  lard, Utilisation approfondie des logiciels statistiques

74

Repr  sentations
graphiques

Tableaux de caract  ristiques

- ⇒ sommes
- ⇒ moyennes
- ⇒ etc.

Fnu By Groupe			
Means for Oneway Anova			
Level	Number	Mean	Std Error
1	16	89.688	23.322
2	27	247.926	17.953
3	0	?	?
4	0	?	?

Std Error uses a pooled estimate of error variance

Guy M  lard, Utilisation approfondie des logiciels statistiques

75

Repr  sentations
graphiques

Tables de d  cision

- ⇒ de plus en plus utilis  es pour r  sumer une analyse
 - ✓ m  dical
 - ✓ marketing
 - ✓ cr  dit
 - ✓ assurance
- ⇒ formule de pr  vision en fonction de variables explicatives (avec possibilit   de Si/alors/sinon embo  t  s)

$$10.75 + \begin{cases} 1.475, & \text{when "gourmet"} \\ -1.475, & \text{when "plain"} \end{cases}$$

Guy M  lard, Utilisation approfondie des logiciels statistiques

76

Repr  sentations
graphiques

3. Graphiques statistiques

- 3.1 Pr  sentation mat  rielle des graphiques
- 3.2 Graphiques pour une variable
- 3.3 Graphiques pour deux variables
- 3.4 Graphiques pour trois variables
- 3.5 Graphiques pour plusieurs variables et plusieurs populations

Guy M  lard, Utilisation approfondie des logiciels statistiques

77

Repr  sentations
graphiques

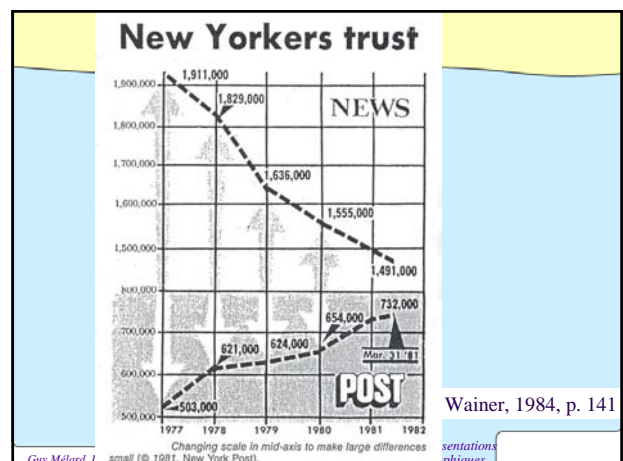
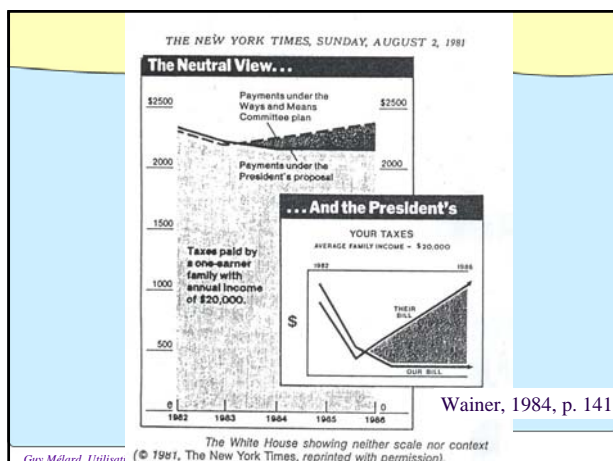
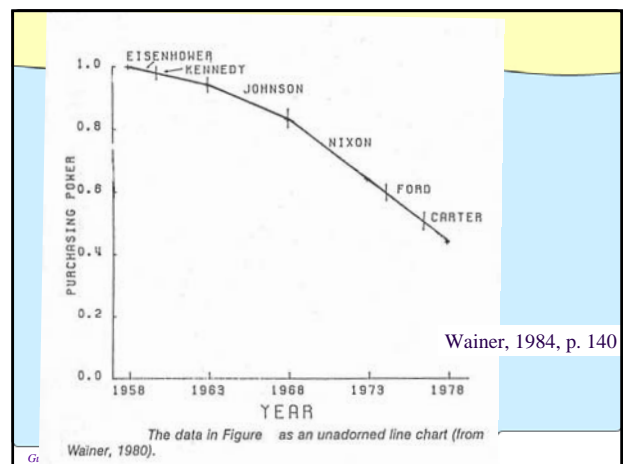
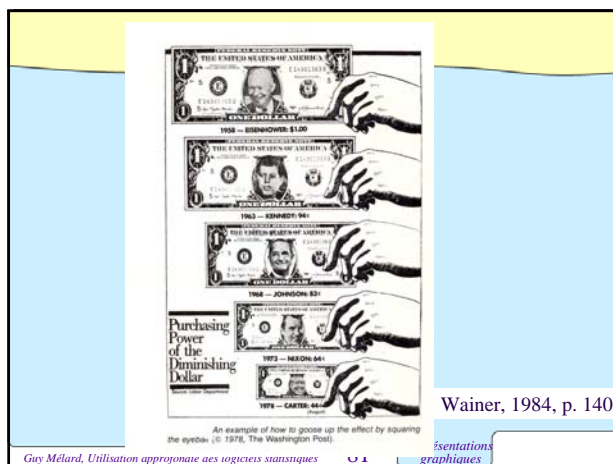
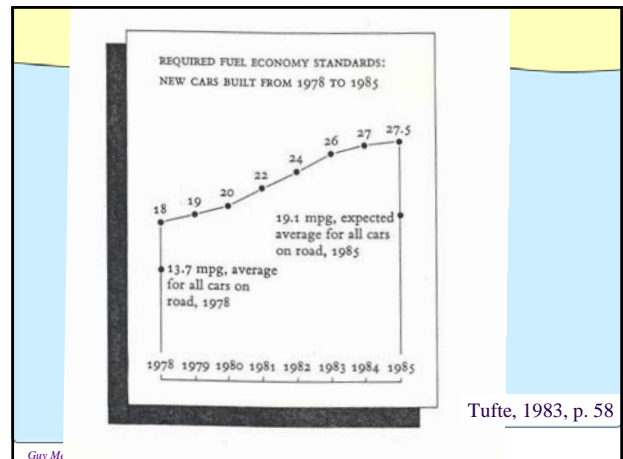
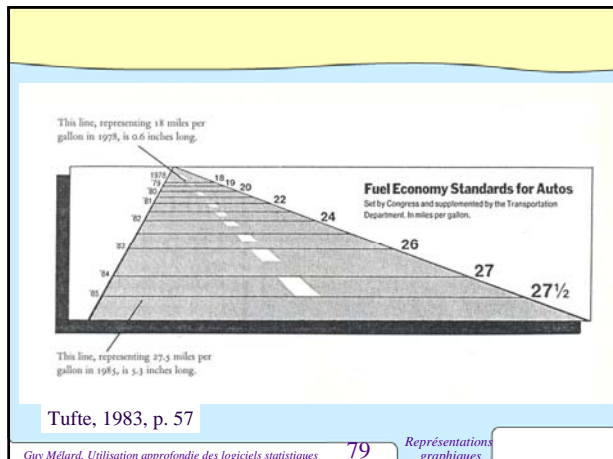
3.1 Pr  sentation mat  rielle des graphiques

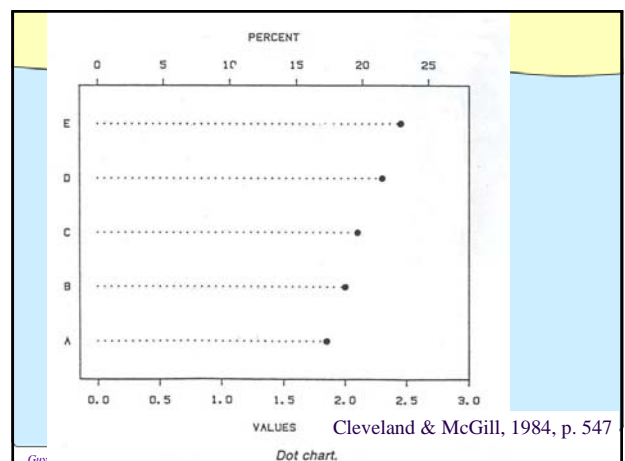
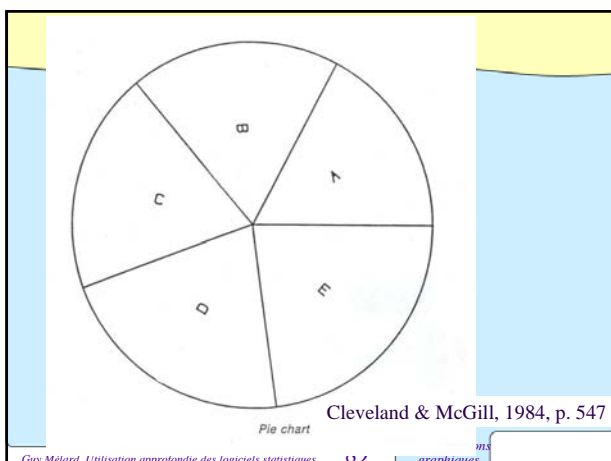
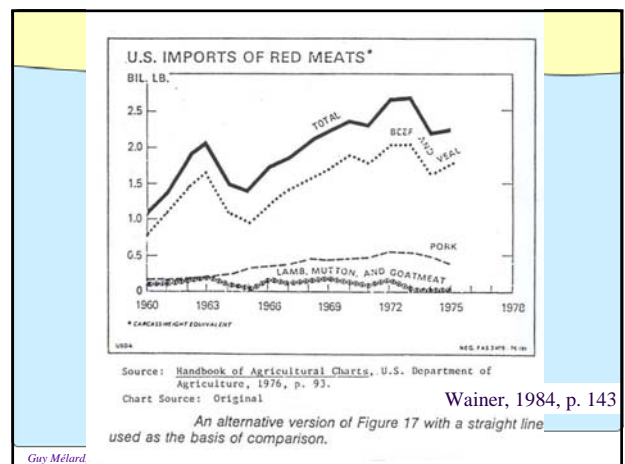
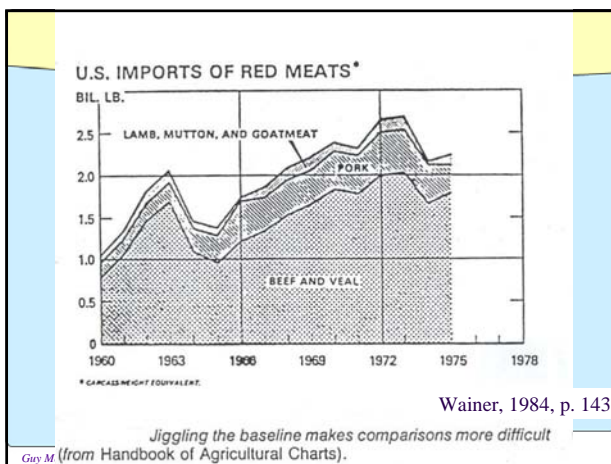
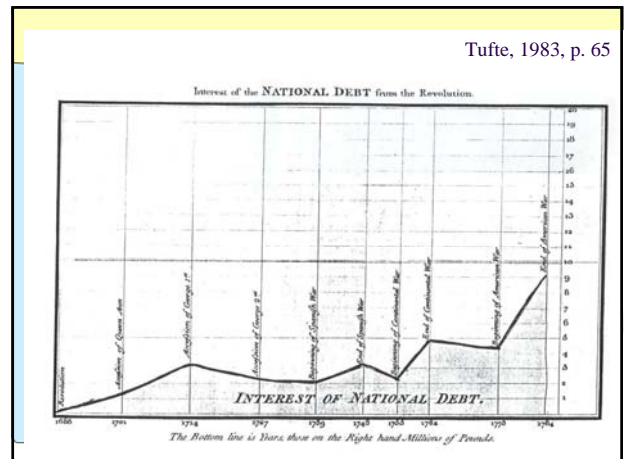
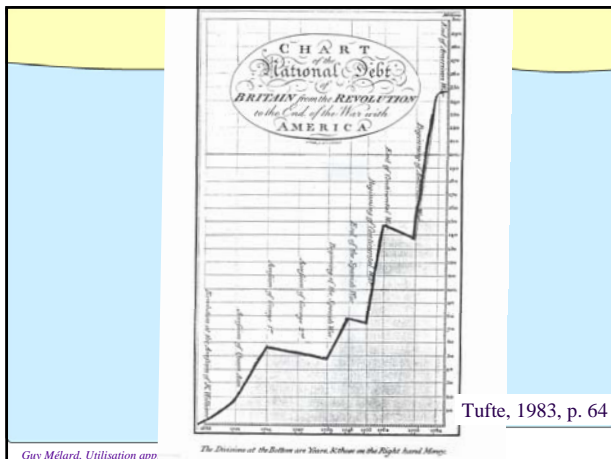
- ⇒ Voyons d'abord quelques exemples

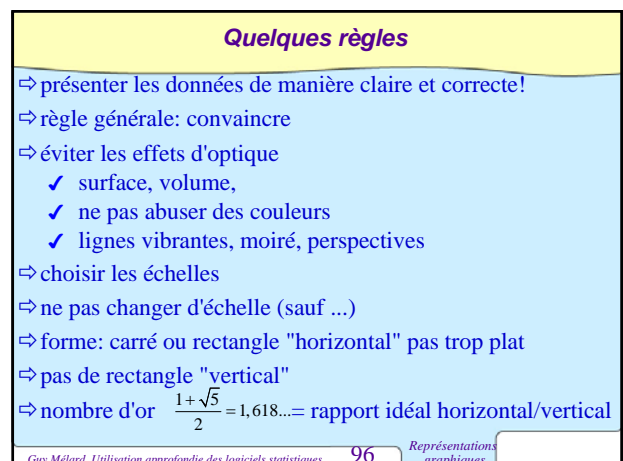
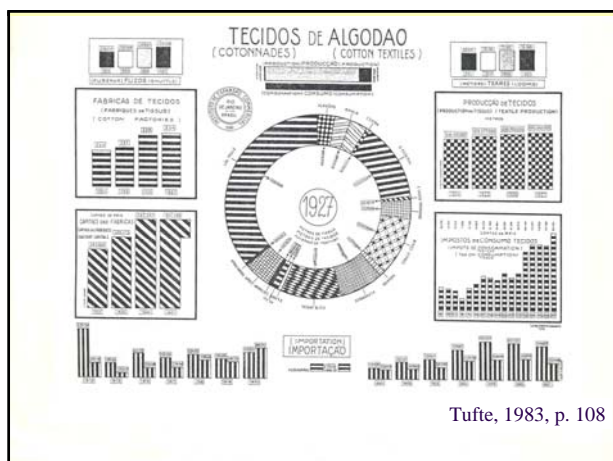
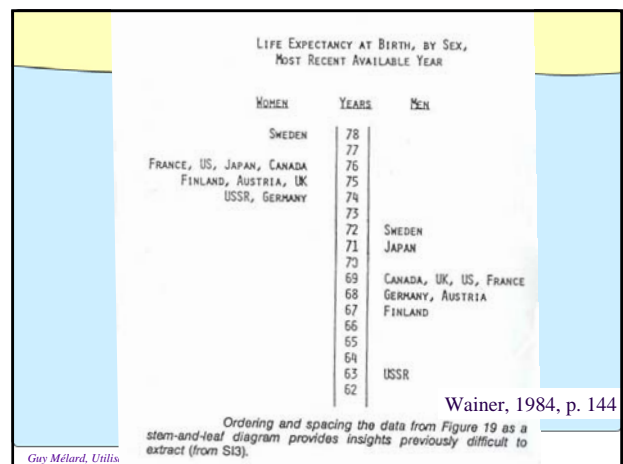
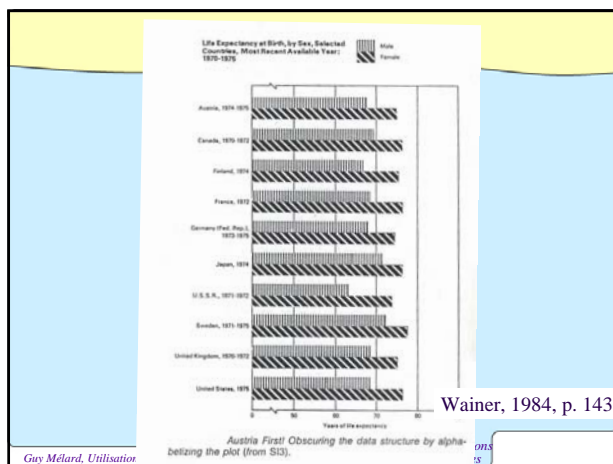
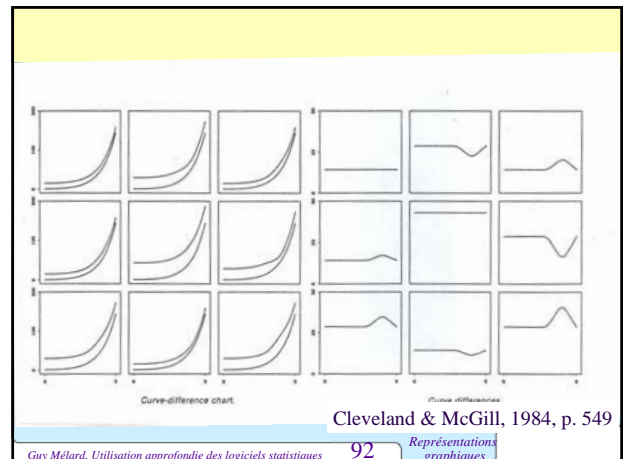
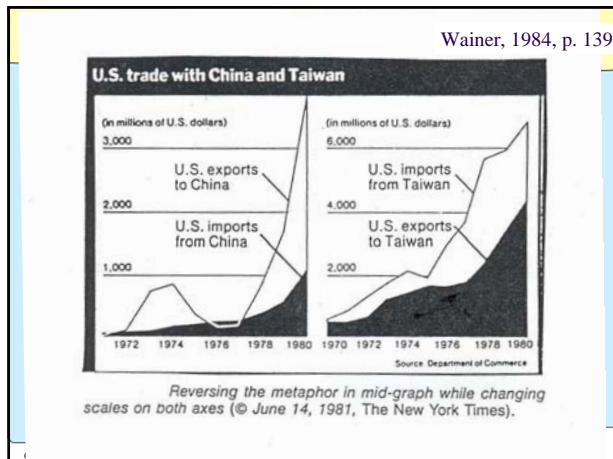
Guy M  lard, Utilisation approfondie des logiciels statistiques

78

Repr  sentations
graphiques







Quelques r  gles

- ⇒   viter les superpositions
 - ✓ seul l'inf  rieur et le total sont visibles
 - ✓ barres divis  es → graphique de points
- ⇒ secteur → graphique de points
- ⇒ attention aux courbes parall  les: diff  rences?
- ⇒ minimiser le rapport encre/information
 - ✓ pas de grille
 - ✓ pas de gadget
- ⇒ si peu de donn  es: tableau
- ⇒ privil  gier les facteurs principaux par rapport aux facteurs accessoires

Guy M  lard, Utilisation approfondie des logiciels statistiques

97

Repr  sentations
graphiques

Graphique ou tableau?

- ⇒ Les graphiques de s  ries peuvent   tre trompeurs
- ⇒ En compl  ment des graphiques, on peut employer — de mani  re interactive — l'identification automatique des points avec la souris
- ⇒ C'est un des outils des graphiques dynamiques
- ⇒ En l'absence, il faut examiner le tableau des donn  es

Ce point
correspond aux valeurs
de deux variables
   une certaine date

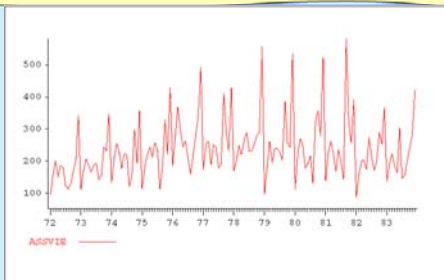


Chapitre 1 Analyse des s  ries temporelles, G. M  lard

98

Repr  sentations
graphiques

La production d'assurance vie d'une compagnie d'assurances (1972-1983)



"Il s'agit de la production d'assurance vie, en nombre de contrats, d'une compagnie d'assurances belge. Sur base du graphique et du bon sens, pensez-vous que la s  rie soit annuelle, trimestrielle, mensuelle, autre?"

Chapitre 1 Analyse des s  ries temporelles, G. M  lard

99

Repr  sentations
graphiques

La production d'assurance vie d'une compagnie d'assurances (1972-1983)

	janv	fevr	mars	avr	mai	juin	juil	ao��t	sept	oct	nov	d��c
1972	94	154	201	151	186	178	122	113	128	166	202	342
1973	111	171	206	186	165	185	193	141	155	244	231	346
1974	133	207	255	226	178	223	218	121	157	299	193	356
1975	114	180	218	243	212	257	234	114	187	329	221	429
1976	185	279	367	298	244	261	208	159	220	278	343	492
1977	172	251	261	190	251	243	179	191	411	291	231	427
1978	170	204	250	221	262	290	232	233	254	279	288	557
1979	96	182	260	195	240	240	227	203	385	258	243	534
1980	113	225	270	252	179	193	216	129	325	356	279	523
1981	139	224	262	225	166	236	198	143	581	322	257	391
1982	87	160	199	203	174	274	213	171	200	289	253	367
1983	137	190	225	186	163	304	146	158	203	245	286	421

- ✓ "Sur base du tableau et du bon sens, pensez-vous que la s  rie soit annuelle, trimestrielle, mensuelle, autre?"
- ✓ "Des contrats promotionnels ont   t   propos  s pendant des mois de septembre de quatre ann  es. Lesquelles?"

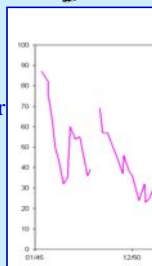
Chapitre 1 Analyse des s  ries temporelles, G. M  lard

100

Repr  sentations
graphiques

Lignes ou points

- ⇒ Dans plusieurs cas, la r  ponse    la question pos  e n'est pas   vidente
- ⇒ En principe les graphiques sont plus informatifs qu'une autre repr  sentation
- ⇒ Une des raisons: l'emploi — traditionnel pour les s  ries temporelles — de lignes pour relier les points
- ⇒ Les points seuls ne permettent pas de bien visualiser les variations
- ⇒ C'est une exception au principe: minimiser le rapport encre/information

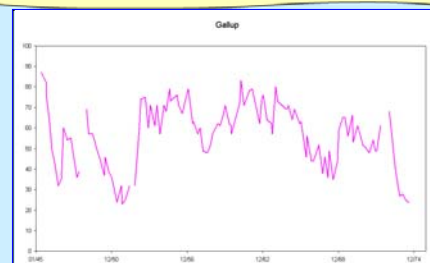


Chapitre 1 Analyse des s  ries temporelles, G. M  lard

101

Repr  sentations
graphiques

La popularit   des pr  sidents des Etats-Unis (1945-1974)

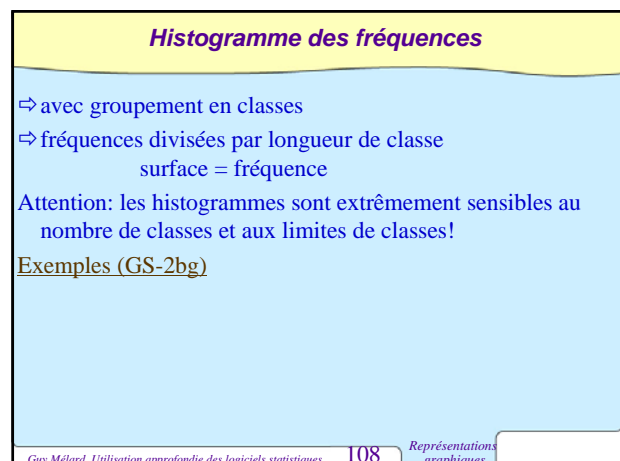
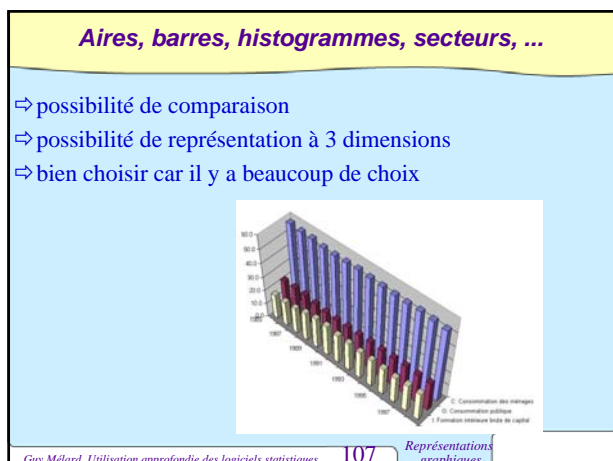
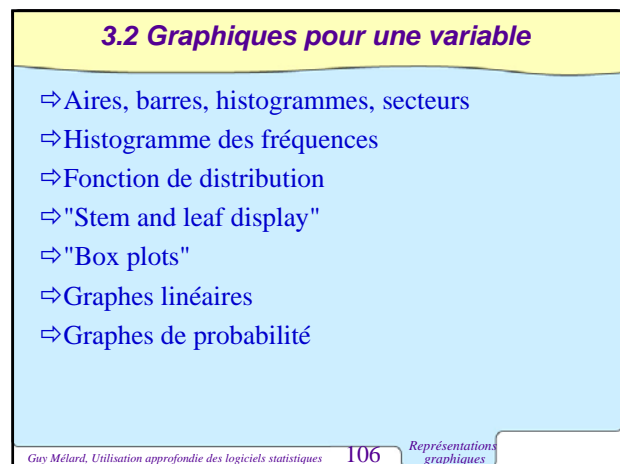
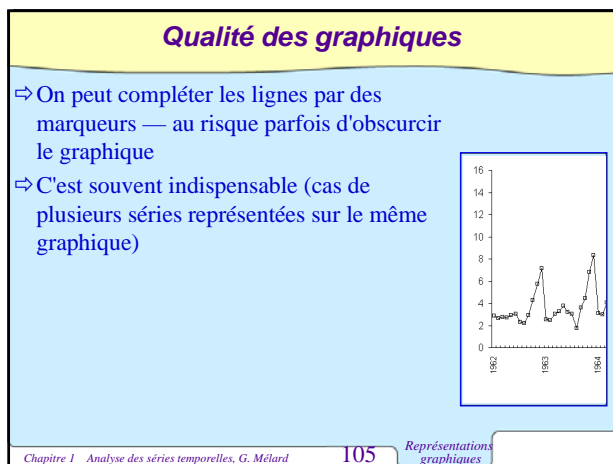
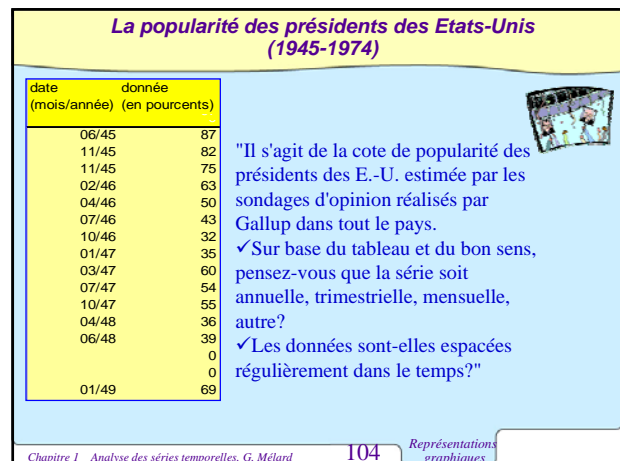
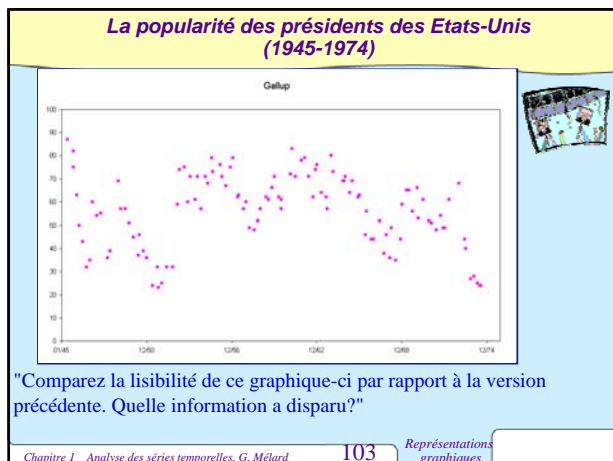


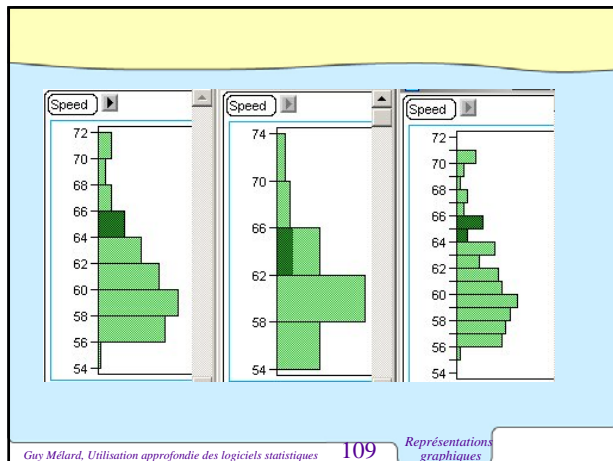
"Il s'agit de la cote de popularit   des pr  sidents des E.-U. estim  e par les sondages d'opinion r  alis  s par Gallup dans tout le pays. Sur base du graphique et du bon sens, pensez-vous que la s  rie soit annuelle, trimestrielle, mensuelle, autre?"

Chapitre 1 Analyse des s  ries temporelles, G. M  lard

102

Repr  sentations
graphiques





"Stem and leaf display"

= diagramme en tiges et feuilles

⇒ = histogramme num  rique

⇒ garder 2 (3) chiffres

⇒ permet de visualiser l'asym  trie

⇒ permet de d  terminer m  diane, quartiles

N.B. L'int  r  t des diagrammes en tiges et feuilles est devenu limit   parce que les autres graphiques peuvent   tre obtenus automatiquement et plus simplement

"Box plots"

= bo  te    moustaches, bo  te    pattes

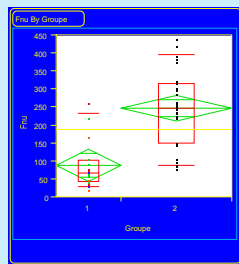
⇒ plusieurs variantes: faire attention    la d  finition

✓ box-plot    quantiles

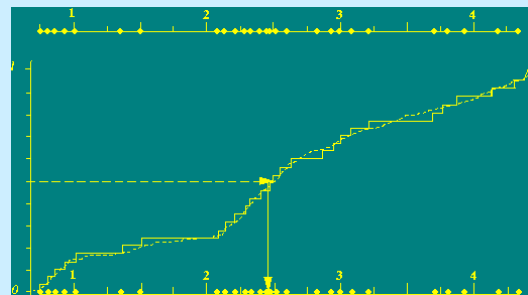
✓ box-plot    valeurs aberrantes

✓ diamants bas  s sur la moyenne et l'  cart-type

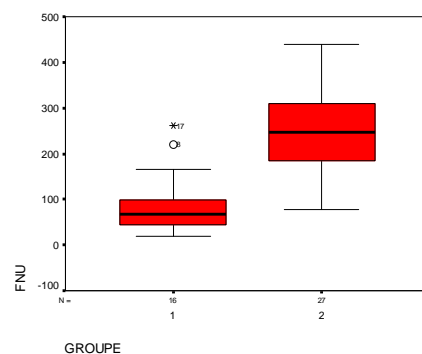
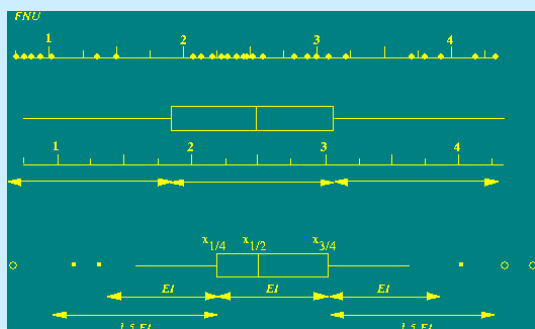
N.B. Ils sont tr  s utiles pour comparer des ensembles de donn  es



Fonction de distribution



Boxplot



Graphe lin  aire

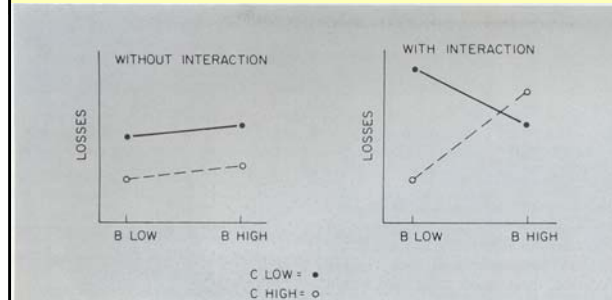
- ⇒ une variable en fonction de l'indice
- ⇒ tr  s bien si l'indice est le temps ou un num  ro d'ordre
- ⇒ comparaison entre deux populations: examen de la diff  rence par rapport    0
- ⇒ avec   ventuellement emploi d'un code ou en joignant les points (m  me si les lignes n'ont pas de sens)
- ⇒ dangereux! Risque de confusion avec les diagrammes x-y

Guy M  lard, Utilisation approfondie des logiciels statistiques

115

Repr  sentations graphiques

Application : interaction en analyse de variance



Graphical interpretation of the interaction between slurry additives B and C. Note that the effect of B going from low to high level depends on whether C is at its high or low level.

Gunter et al., RCA Engineer, 1985, p. 58

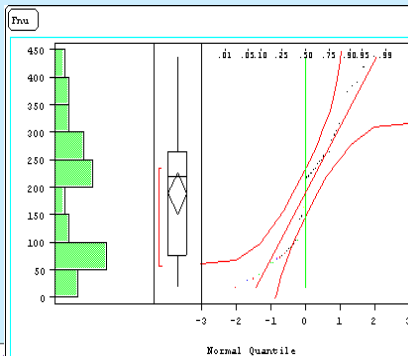
Guy M  lard, Utilisation approfondie des logiciels statistiques

116

Repr  sentations graphiques

Graphes de probabilit  

- ⇒ Graphique quantile-quantile
- ⇒ en particulier "normal probability plot"



Guy M  lard, Utilisation approfondie des logiciels statistiques

3.3 Graphiques pour deux variables

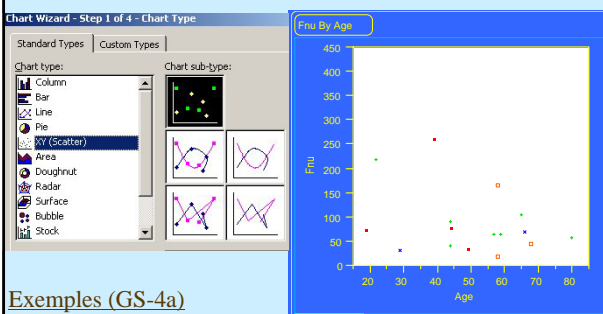
- ⇒ Graphique x-y ("scattergram")
- ⇒ Graphes parall  les
- ⇒ Comparaison entre deux populations
 - ✓ utilisation de la diff  rence    regarder autour de la 1  re bissectrice
 - ✓ m  me le test de Student peut   tre visualis   ainsi

Guy M  lard, Utilisation approfondie des logiciels statistiques

118

Repr  sentations graphiques

Graphique x-y ("scattergram")



Exemples (GS-4a)

Guy M  lard, Utilisation approfondie des logiciels statistiques

119

Repr  sentations graphiques

3.4 Graphiques pour trois variables

- ⇒    3 dimensions (voir les graphiques dynamiques)
- ⇒ graphiques ternaires (pour les proportions)

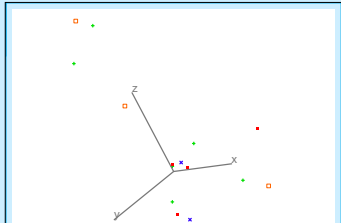
Guy M  lard, Utilisation approfondie des logiciels statistiques

120

Repr  sentations graphiques

Graphiques   3 dimensions

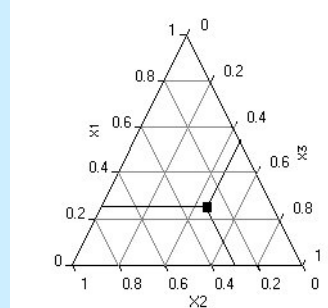
    3 dimensions (voir les graphiques dynamiques)



Guy M  lard, Utilisation approfondie des logiciels statistiques

Graphiques ternaires

  pour les proportions dans le cas de 3 composants d'un produit



Guy M  lard, Utilisation approfondie des logiciels statistiques

122

Repr  sentations graphiques

3.5 Graphiques pour plusieurs variables et plusieurs populations

-   Graphiques pour plusieurs variables
-   Graphiques pour plusieurs populations

Guy M  lard, Utilisation approfondie des logiciels statistiques

123

Repr  sentations graphiques

Plusieurs variables (1/2)

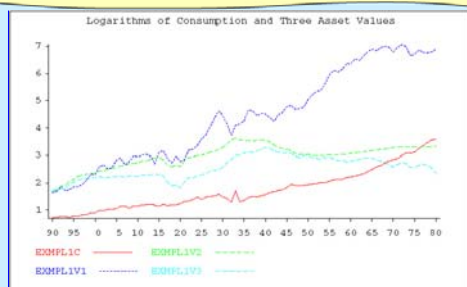
- (  d  faut de pouvoir regarder   plus de 3 dimensions)
-   Aires, histogrammes, barres: empil  es, superpos  es (possibilit   de normalisation   100% pour des fr  quences et effet   3 dimensions)
-   graphes lin  aires (avec superposition et emploi d'un code ou en joignant les points)
-   graphes low-high-close (en plus de la variable qui repr  sente la valeur finale sur une p  riode, on repr  sente le minimum et le maximum sur cette p  riode): cfr bourse
-   tableaux de contingence: mosa  ques

Guy M  lard, Utilisation approfondie des logiciels statistiques

124

Repr  sentations graphiques

Exemple : consommation en logarithmes et les cours de trois actifs



"Qu'est ce qui diff  re par rapport aux exemples pr  c  dents?"

Chapitre 1 Analyse des s  ries temporelles, G. M  lard

125

Repr  sentations graphiques

Exemple : logements commenc  s et ventes de logements aux E.-U.

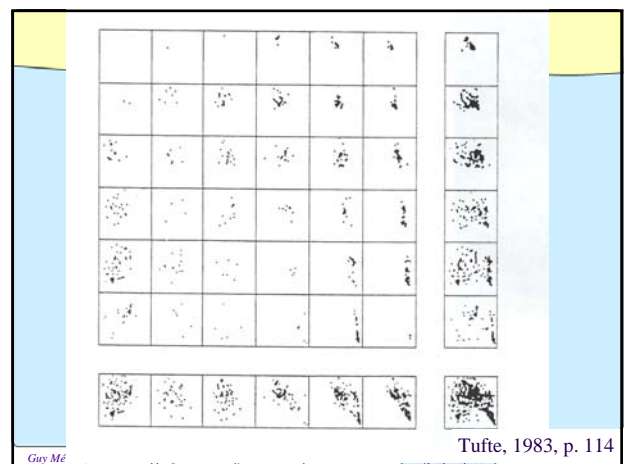
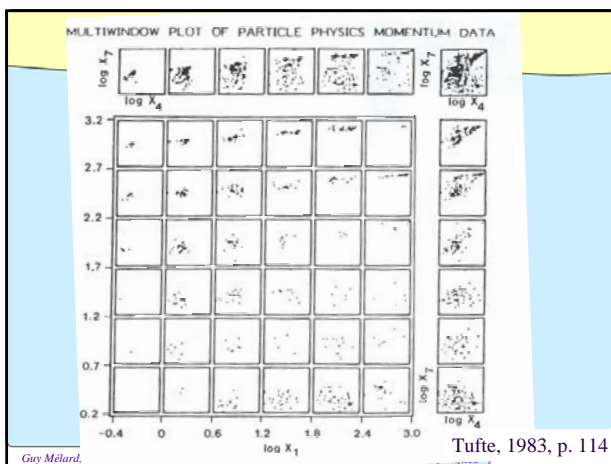
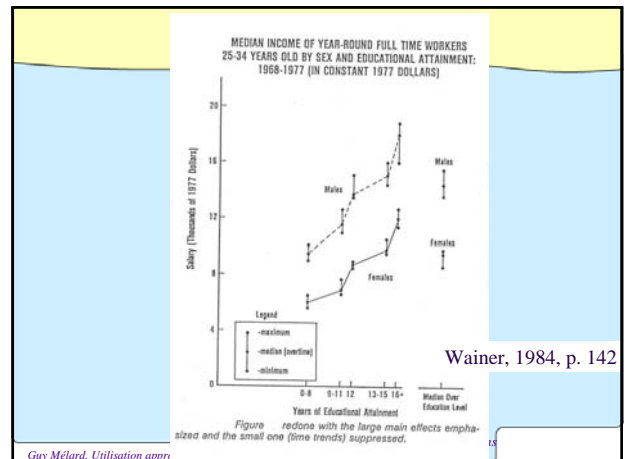
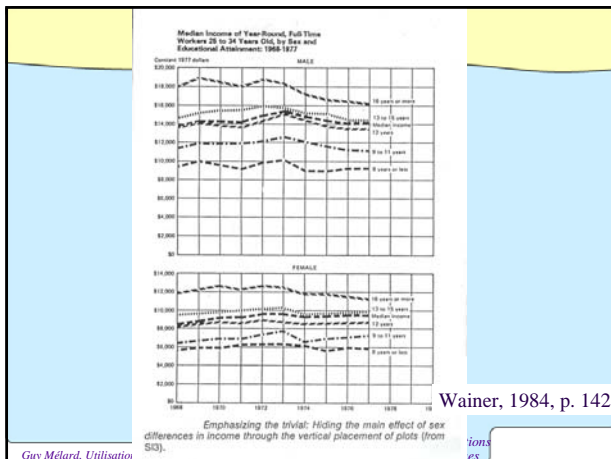


"Qu'est ce qui diff  re par rapport aux exemples pr  c  dents?"

Chapitre 1 Analyse des s  ries temporelles, G. M  lard

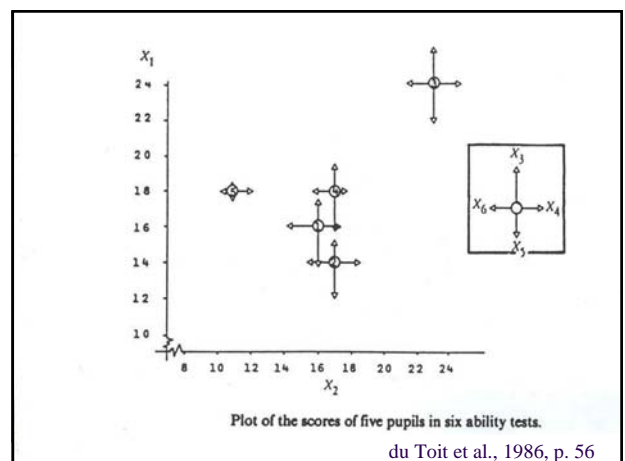
126

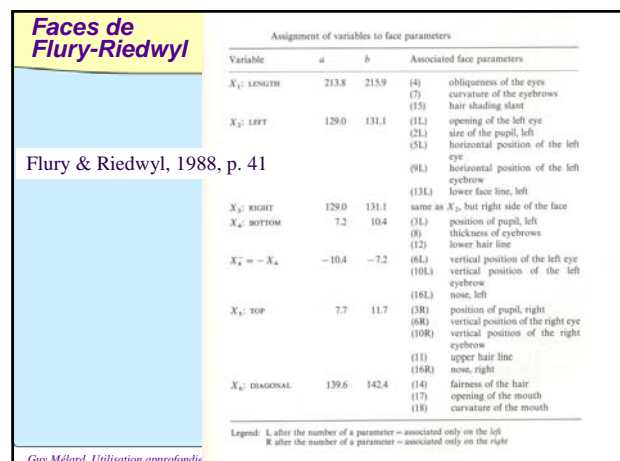
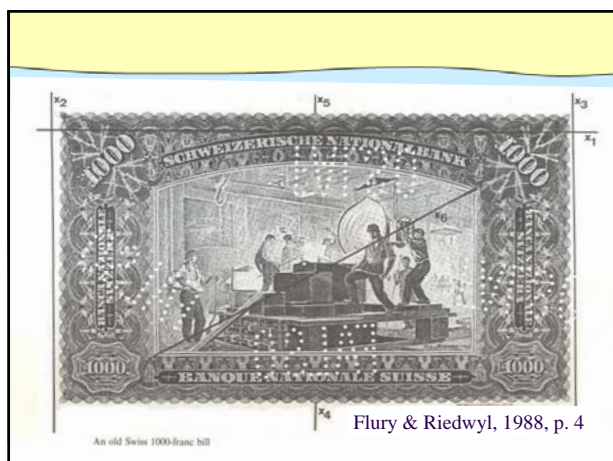
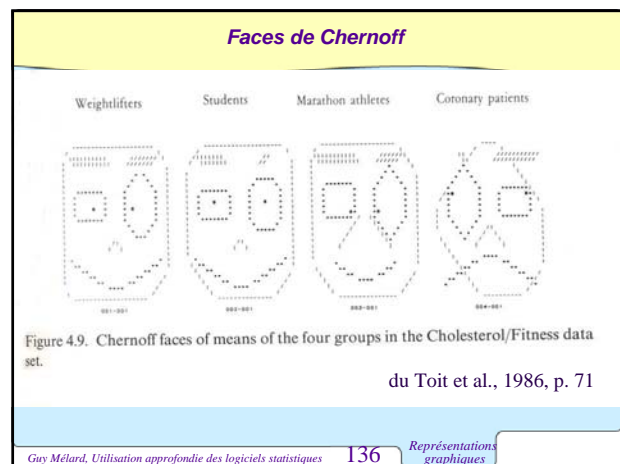
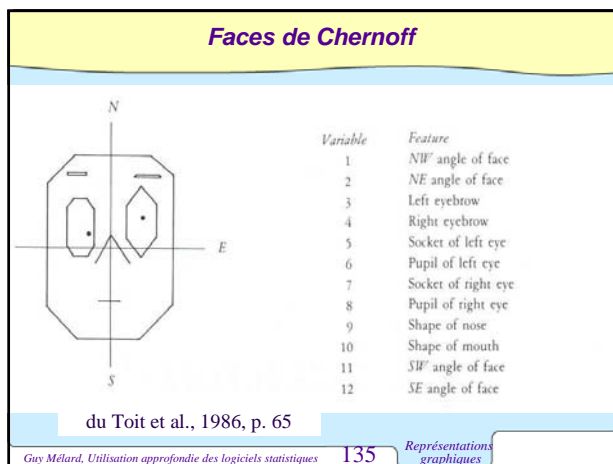
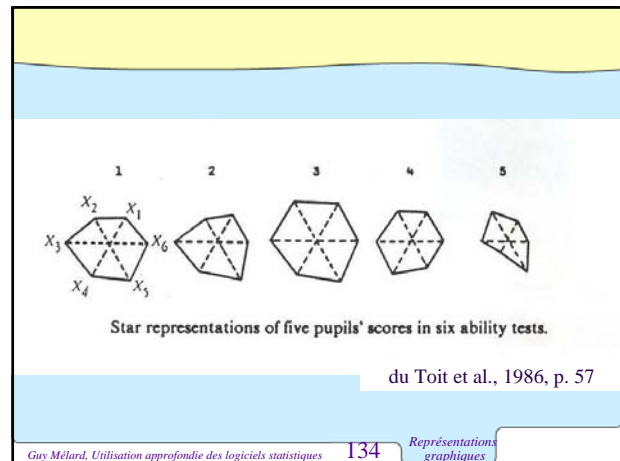
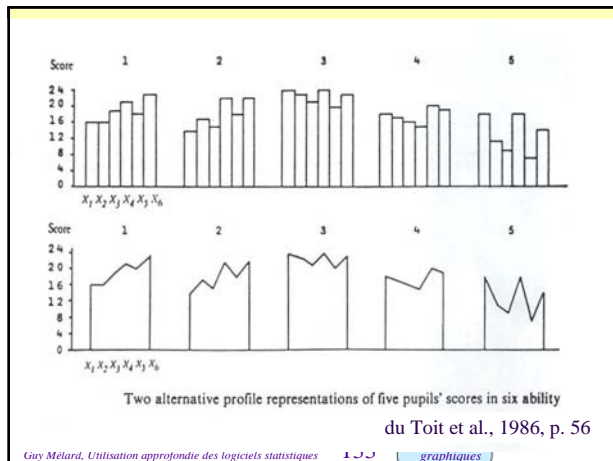
Repr  sentations graphiques

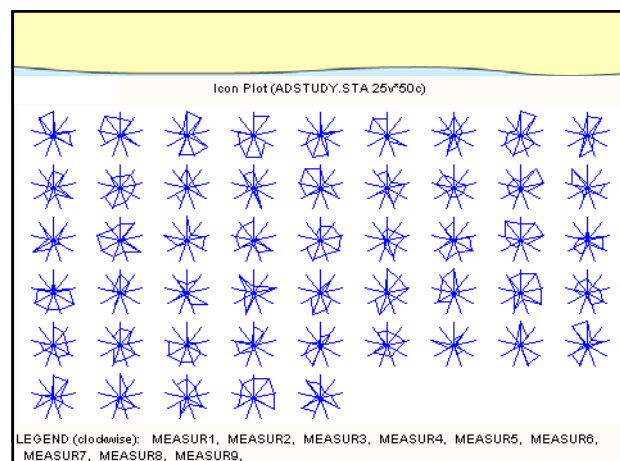
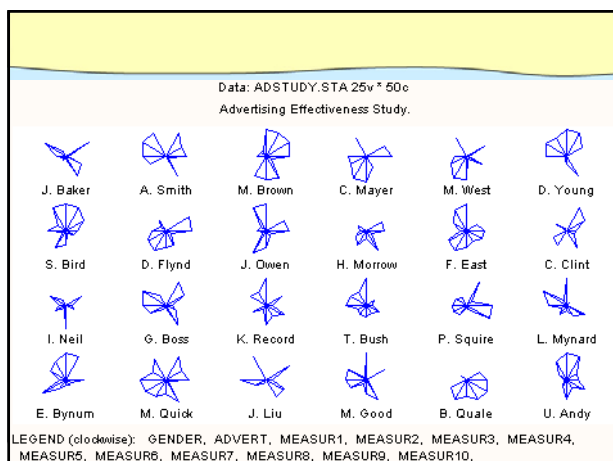
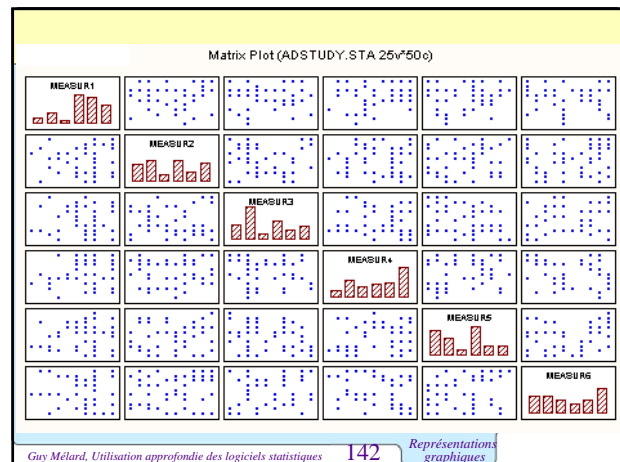
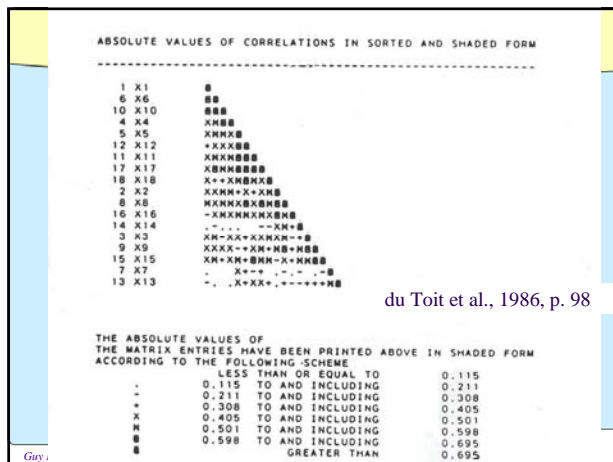
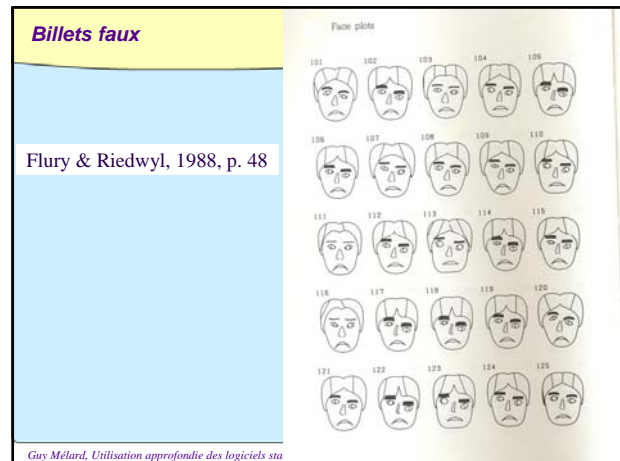
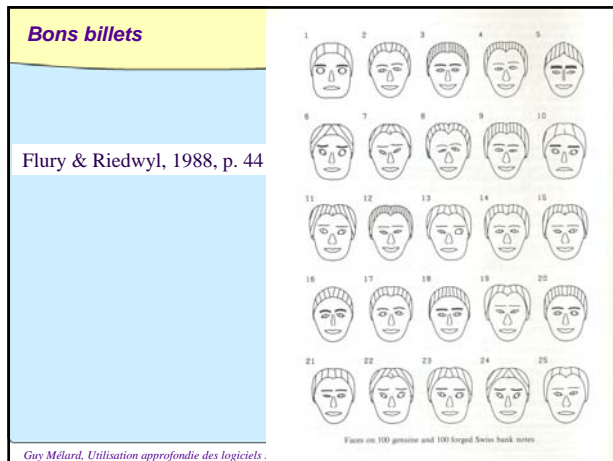


Plusieurs variables (2/2)

- ⇒ profils (pas facile à lire)
- ⇒ radar ou   toiles (limite en nombre de variables)
- ⇒ faces de Chernoff (limite en nombre de variables)
- ⇒ faces de Flury-Riedwyl (jusqu'à 24 variables!)
- ⇒ matrices de corr  lations (repr  sente une corr  lation comme nuance de gris ou couleur)
- ⇒ courbes d'Andrews

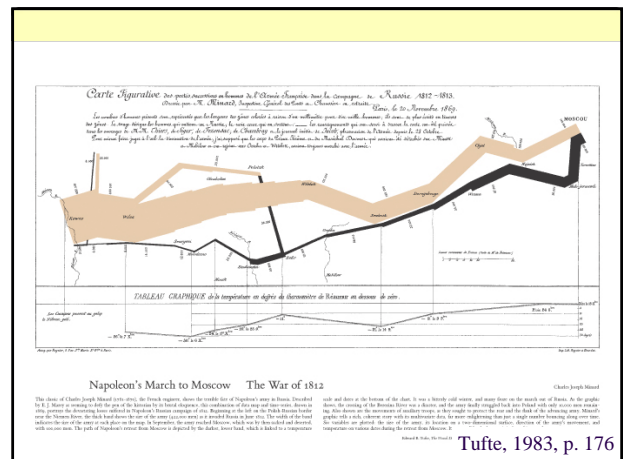
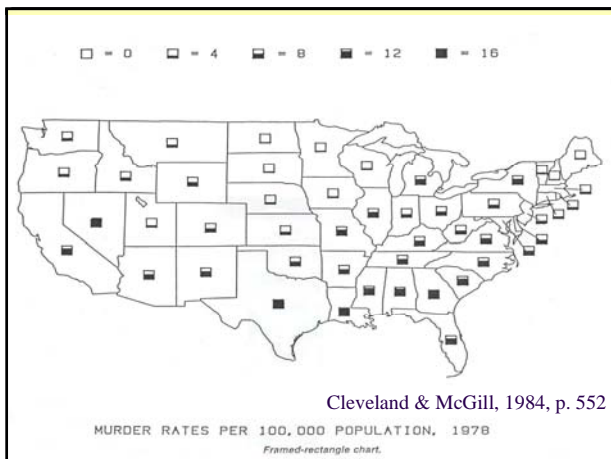
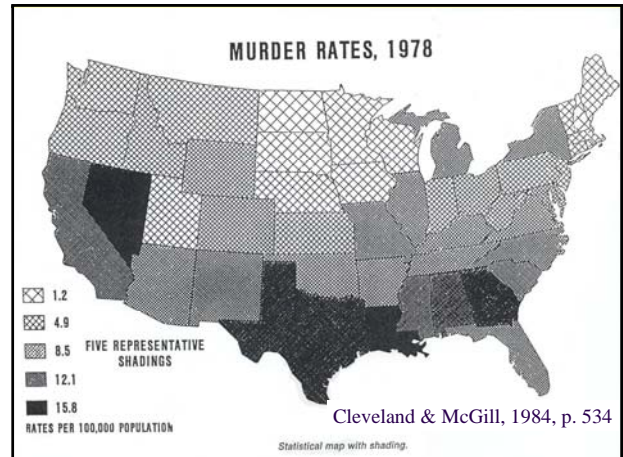






Plusieurs populations

- ⇒ Aires, histogrammes, barres: empil  es, superpos  es (possibilit   de normalisation    100% pour des fr  quences et effet    3 dimensions)
- ⇒ Donn  es r  parties g  ographiquement: cartes avec
 - ✓ codage (patrons, couleurs)
 - ✓ jauges



4. Tableaux et graphiques dynamiques

- 4.1 Objectif
- 4.2 Outils
- 4.3 Tableaux dynamiques
- 4.4 Graphiques dynamiques

Tableaux et graphiques dynamiques

- ⇒ Tableaux dynamiques: p. ex. [Excel](#)
- ⇒ Graphiques dynamiques: p. ex. [JMP](#)

PivotTable

PivotTable Wizard - Step 1 of 4

Where is the data that you want to analyze?

- ☒ Microsoft Excel list or database
- ☐ External data source
- ☐ Multiple consolidation ranges
- ☐ Another PivotTable

Cancel < Back Next > Finish

Guy M  lard, Utilisation approfondie des logiciels statistiques 151 Repr  sentations graphiques

PivotTable Wizard - Step 2 of 4

Where is the data that you want to use?

Range: Browse...

Cancel < Back Next > Finish

Guy M  lard, Utilisation approfondie des logiciels statistiques 152 Repr  sentations graphiques

PivotTable Wizard - Step 3 of 4

Construct your PivotTable by dragging the field buttons on the right to the diagram on the left.

PAGE Mois R  gion COLUMN Mois ROW Sum of Revenu DATA

Cancel < Back Next > Finish

Guy M  lard, Utilisation approfondie des logiciels statistiques 153 Repr  sentations graphiques

PivotTable Field

Source field: Revenu

Name: Sum of Revenu

Summarize by:

- Sum
- Count
- Average
- Max
- Min
- Product
- Count Nums

OK Cancel Delete Number... Options >>

PAGE Mois R  gion COLUMN Mois ROW Sum of Revenu DATA

Cancel < Back Next > Finish

Guy M  lard, Utilisation approfondie des logiciels statistiques 154 Repr  sentations graphiques

PivotTable Wizard - Step 4 of 4

Where do you want to put the PivotTable?

- ☐ New worksheet
- ☒ Existing worksheet

Click Finish to create your PivotTable.

Cancel Options... < Back Next > Finish

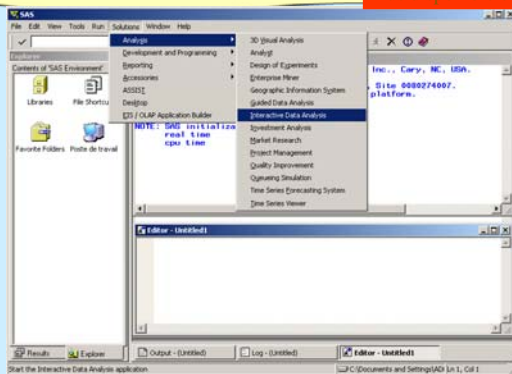
Guy M  lard, Utilisation approfondie des logiciels statistiques 155 Repr  sentations graphiques

	A	B	C	D	E	F
1	Sum of Revenu	R��gion				
2	Mois	Est	Nord	Ouest	Sud	Grand Total
3	Mars	353.1	251.32	246.12	133.55	984.09
4	Avril	165.45	310.54	305.22	215.66	996.87
5	Mai	289.2	264.25	243.99	168.75	966.19
6	Grand Total	807.75	826.11	795.33	517.96	2947.15

Guy M  lard, Utilisation approfondie des logiciels statistiques 156 Repr  sentations graphiques

Annexe. Quelques exemples de SAS/INSIGHT Plus disponible

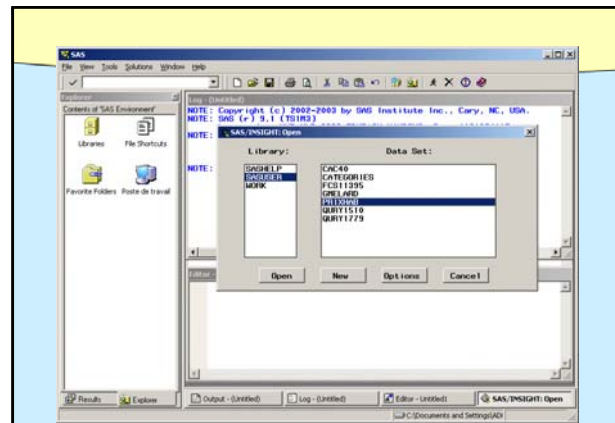
Plus disponible en 2009



Guy Mélard, Utilisation approfondie des logiciels statistiques

157

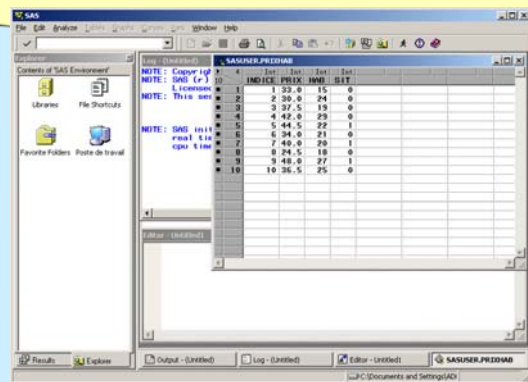
Représentation graphiques



Guy Méléard, Utilisation approfondie des logiciels statistiques

158

Représentation graphiques

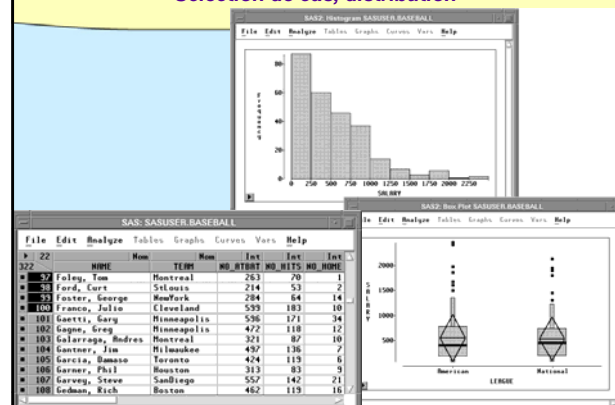


Guy Méléard, Utilisation approfondie des logiciels statistiques

159

Représentation graphique

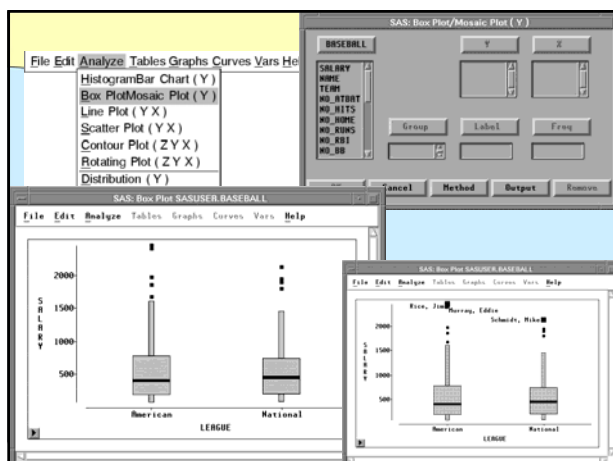
Sélection de cas, distribution



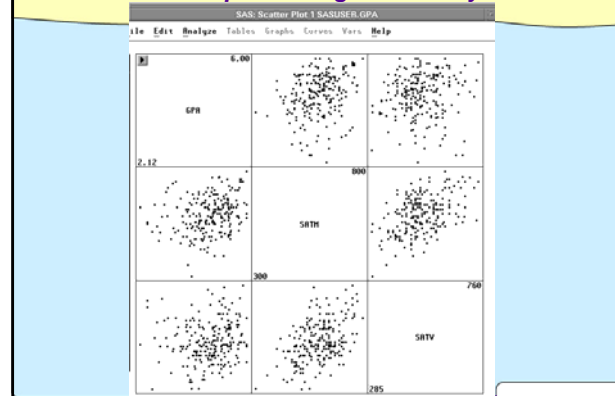
Guy Mèlard, Utilisation approfondie des logiciels statistiques

100

graphiques



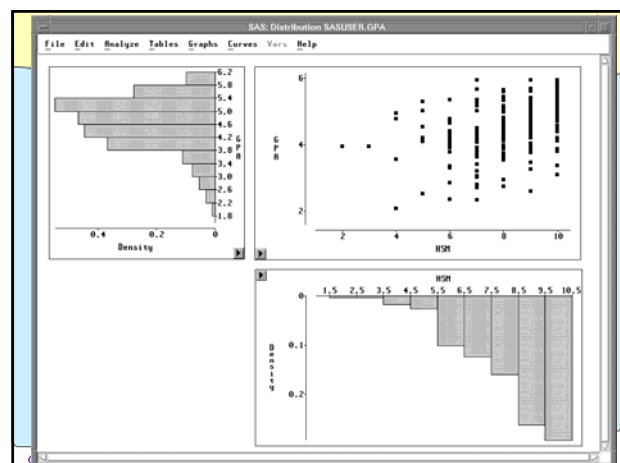
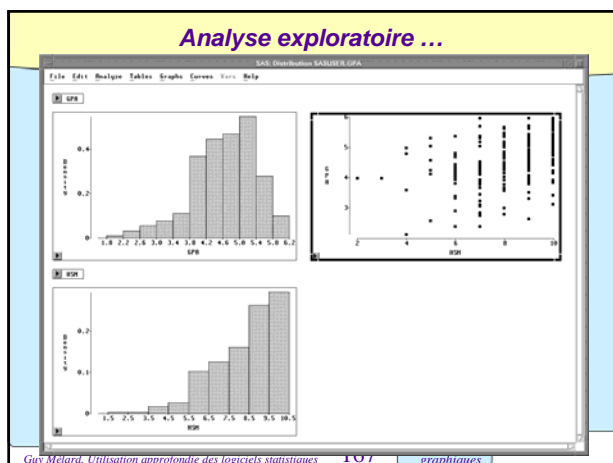
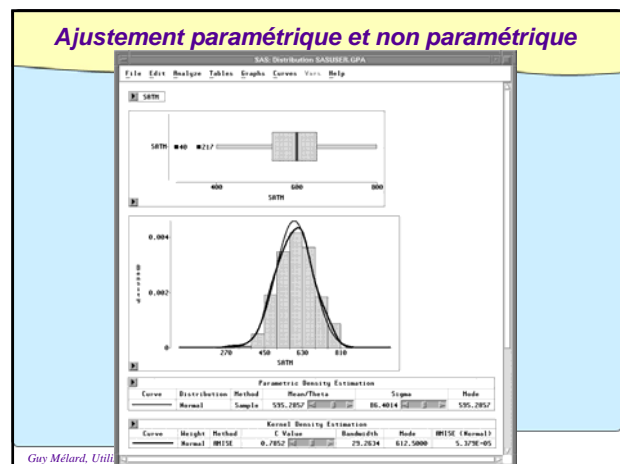
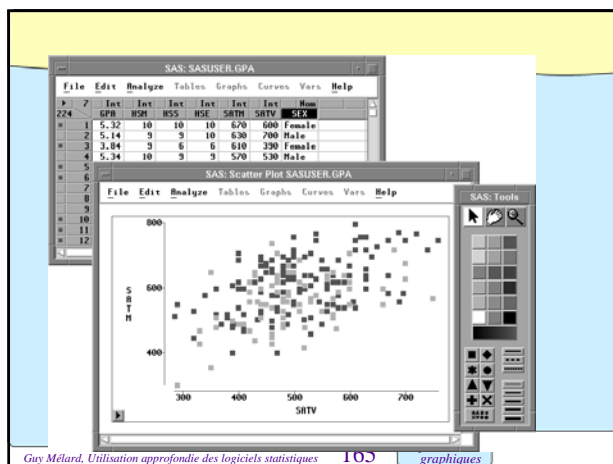
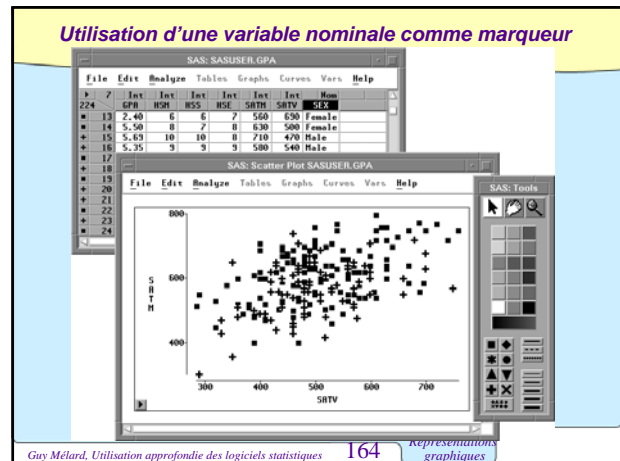
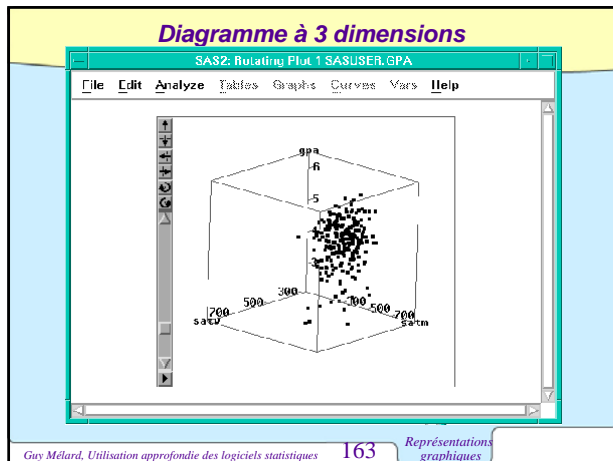
Mosaïque de diagrammes x-y



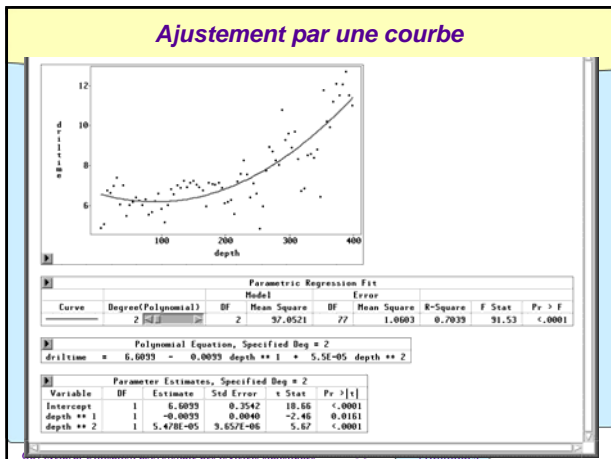
Guy Mélard, Utilisation approfondie des logiciels statistiques

162

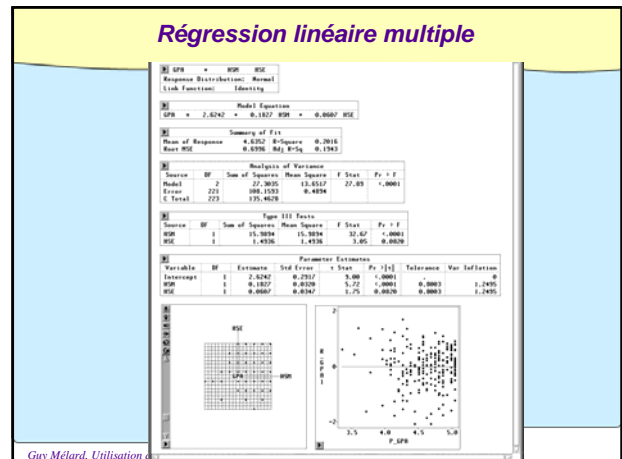
graphiques



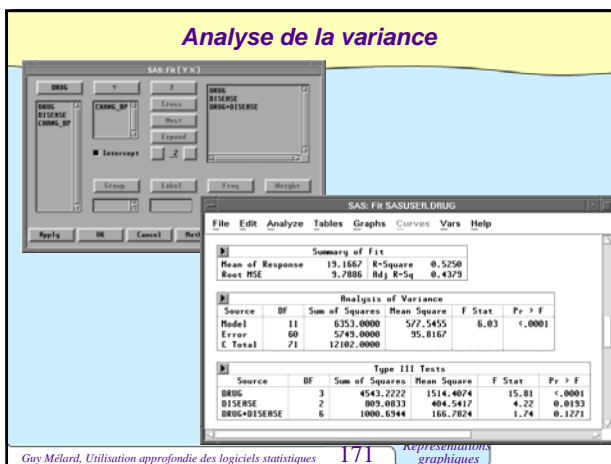
Ajustement par une courbe



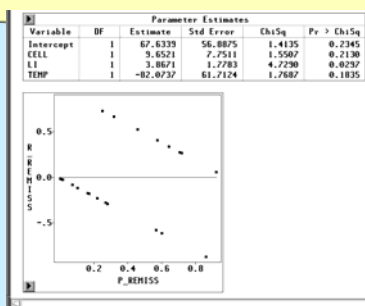
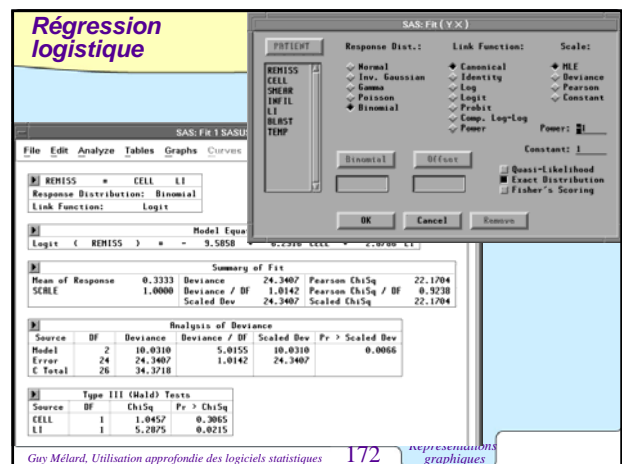
R  gression lin  aire multiple



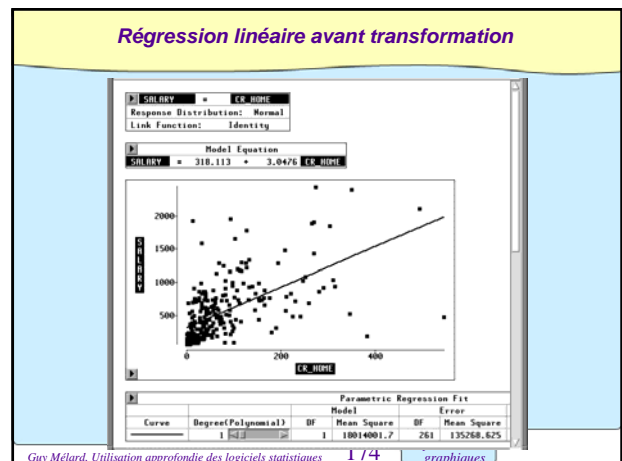
Analyse de la variance

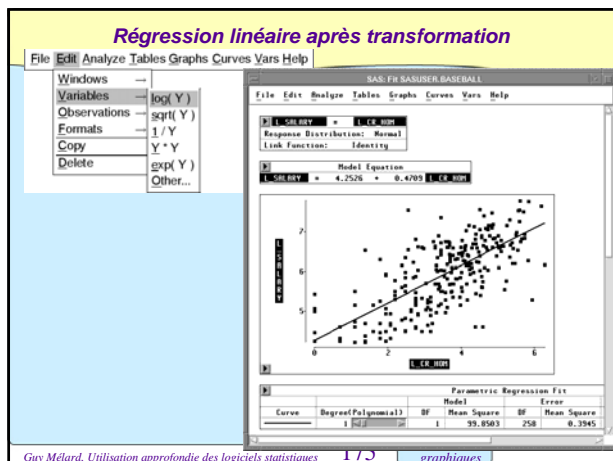


R  gression logistique

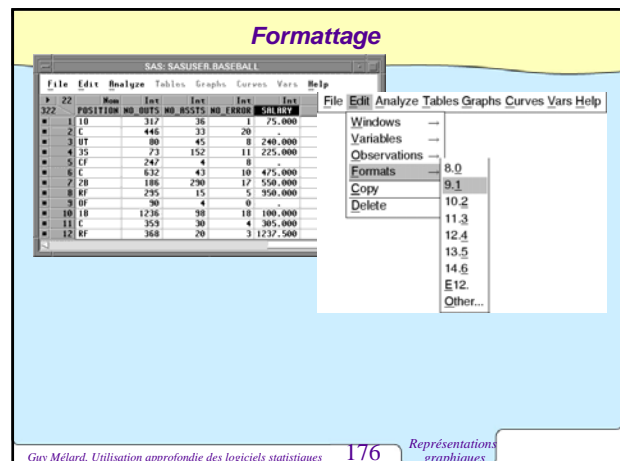


R  gression lin  aire avant transformation

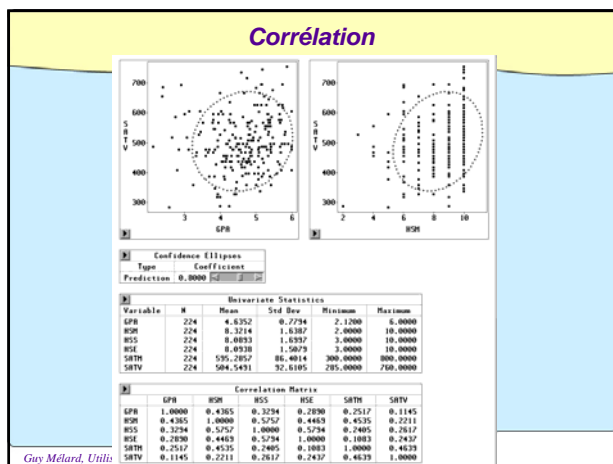




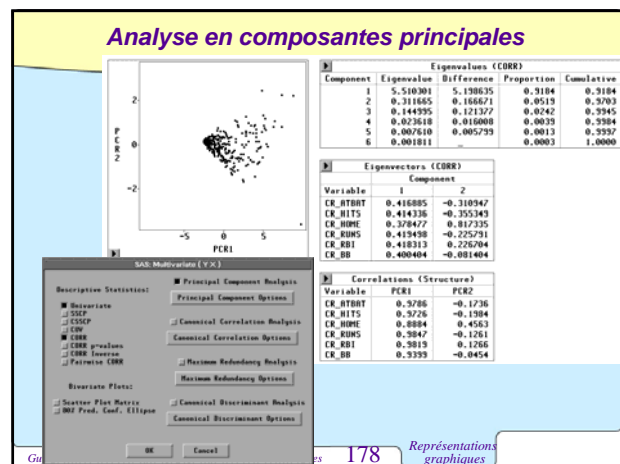
Guy M  lard, Utilisation approfondie des logiciels statistiques 175 graphiques



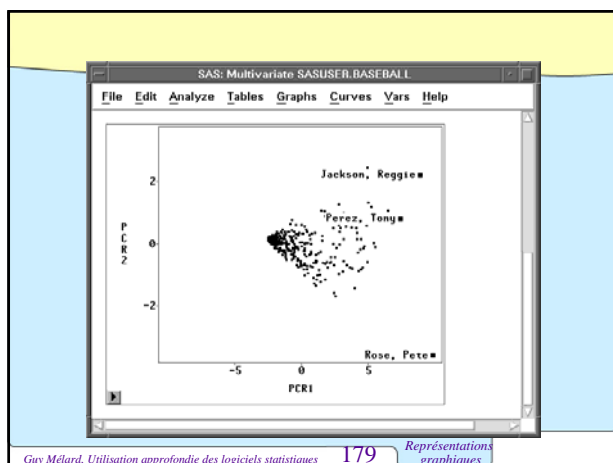
Guy M  lard, Utilisation approfondie des logiciels statistiques 176 Repr  sentations graphiques



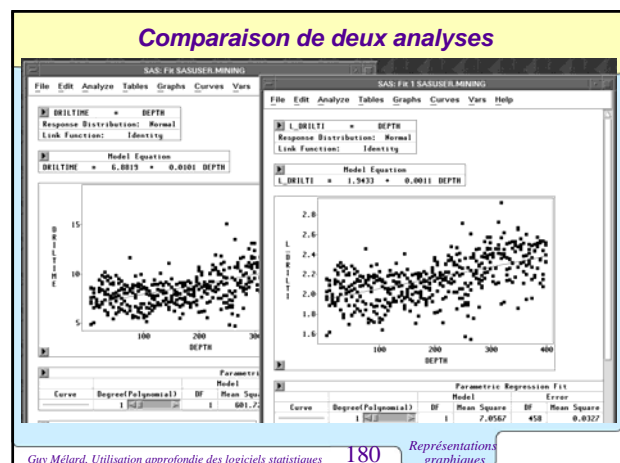
Guy M  lard, Utili 175 graphiques



Guy M  lard, Utili 178 Repr  sentations graphiques



Guy M  lard, Utilisation approfondie des logiciels statistiques 179 Repr  sentations graphiques



Guy M  lard, Utilisation approfondie des logiciels statistiques 180 Repr  sentations graphiques

