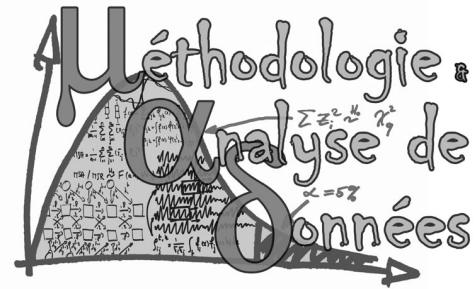




**UNIVERSITÉ  
DE GENÈVE**

**FACULTÉ DE PSYCHOLOGIE  
ET DES SCIENCES DE L'ÉDUCATION**



# **Introduction à l'analyse exploratoire des données avec SPSS**

**Professeurs : O. Renaud et G. Pini**

**Assistante: K. Iglesias**

**Moniteurs : R. Wipfli et B. Ossipow**

**Certaines parties de ce polycopié sont basées sur le  
cours de M.Eid et T. Lischetzke**



# Programme des cours

16 mars	Tous	Intro
23 mars	Jaune et bleu	La démarche de la recherche : où, quand et pourquoi utiliser SPSS ?
30 mars	Rouge et vert	
6 avril	Jaune et bleu	Le questionnaire et la transformation de données
13 avril	Rouge et vert	
27 avril	Jaune et bleu	Analyse des données : Représentations graphiques
4 mai	Rouge et vert	
11 mai	Jaune et bleu	Analyse des données : Les mesures descriptives
18 mai	Rouge et vert	
1 juin	Jaune et bleu	Analyse des données : Corrélation et régression linéaire simple
8 juin	Rouge et vert	
15 juin	Tous	Questions et distribution de l'examen

# Programme des TD

18h-20h	21	mars	Jaune	Découverte de SPSS
12h-14h	22	mars	Bleu	
18h-20h	28	mars	Rouge	
12h-14h	29	mars	Vert	
18h-20h	4	avril	Jaune	Entrées les données à partir du questionnaire
12h-14h	5	avril	Bleu	
18h-20h	11	avril	Rouge	
12h-14h	12	avril	Vert	
18h-20h	25	avril	Jaune	Transformer les données
12h-14h	26	avril	Bleu	
18h-20h	2	mai	Rouge	
12h-14h	3	mai	Vert	
18h-20h	9	mai	Jaune	Représentations graphiques
12h-14h	10	mai	Bleu	
18h-20h	16	mai	Rouge	
12h-14h	17	mai	Vert	
18h-20h	23	mai	Jaune	Mesures descriptives
12h-14h	24	mai	Bleu	
18h-20h	30	mai	Rouge	
12h-14h	31	mai	Vert	
12h-14h	7	juin	Jaune	Corrélation et régression
14h-16h	7	juin	Bleu	
18h-20h	13	juin	Rouge	
12h-14h	14	juin	Vert	

Monitorat

22, 23, 29 et 30 mars, 5, 6, 12, 13, 26 et 27 avril  
3, 4, 10, 11, 17, 18, 24 et 31 mai et 1, 8, 14 et 15 juin

# Informations pratiques

Professeurs : O. Renaud ([Olivier.Renaud@pse.unige.ch](mailto:Olivier.Renaud@pse.unige.ch))  
G. Pini ([Gianreto.Pini@pse.unige.ch](mailto:Gianreto.Pini@pse.unige.ch))  
Assistante : Katia Iglesias ([Katia.Iglesias@pse.unige.ch](mailto:Katia.Iglesias@pse.unige.ch))  
Moniteurs : Baptiste Ossipow ([Ossipow1@etu.unige.ch](mailto:Ossipow1@etu.unige.ch))  
Rolf Wipfli ([Wipflra7@etu.unige.ch](mailto:Wipflra7@etu.unige.ch))

## Indications Générales :

Semestre d'été, 3 crédits

Les étudiant-e-s seront réparti-e-s en 4 groupes (jaune, bleu, rouge et vert). Les cours ont lieu en alternance une semaine sur deux les jeudis de 10h-12h en salle U300 (en alternance jaune-bleu et rouge-vert).

Les TD ont lieu en alternance les mardi de 18h-20h pour les groupes jaunes et rouge et les mercredis de 12h-14h pour les groupes bleu et vert (à l'exception du 7 juin pour les bleus, les TD auront lieu le mardi de 14h-16h au lieu des TP) en salle M4183.

Pour les TP, ils ont lieu les mercredis 14h-16h et jeudis 16h-18h en salle M5183 indépendamment de votre groupe d'appartenance.

## Objectifs du cours :

Se familiariser avec les méthodes de base d'exploration de données de recherche en psychologie et l'utilisation du logiciel statistique SPSS.

## Descriptif du cours :

Le cours aborde tous les aspects de l'analyse de données, aussi bien théoriques que l'application avec un logiciel statistique. Les données utilisées proviennent de recherches dans le domaine de la psychologie. Les étudiant-e-s apprennent comment entrer les données dans l'ordinateur et comment effectuer des calculs élémentaires, qui répondent aux questions de recherche. Le cours donne aussi une introduction aux méthodes statistiques pour décrire la distribution des variables nominales, ordinales et quantitatives et leurs relations.

## **Programme indicatif :**

- La démarche de la recherche : où, quand et pourquoi utiliser SPSS ?
- Le questionnaire et la transformation de données
- Analyse des données : Représentations graphiques
- Analyse des données : Les mesures descriptives
- Analyse des données : Corrélation et régression linéaire simple
- Questions
- Distribution de l'examen

## **Support de cours et des travaux dirigés :**

Il peut être acheté à la centrale des photocopies (ou téléchargé depuis dokeos).

## **Informations et documents :**

Toutes les informations, horaires et support sont sur Dokeos :

<http://dokeos.unige.ch/courses/7111F/>

## **Examen :**

Été 2006 : Distribution, le 15 juin 2006 lors du cours, du ou des jeu/x de donnée pour l'examen. L'examen sera un écrit de 2 heures portant sur les analyses faites à la maison et sur la matière vue en cours et aux travaux dirigés.

Automne 2006 : Disponibilité, le 28 septembre 2006 sur dokeos, du ou des jeu/x de donnée pour l'examen. L'examen sera un écrit de 2 heures portant sur les analyses faites à la maison et sur la matière vue en cours et aux travaux dirigés

Hiver 2007 : examen oral (session extraordinaire).

## **Réception des étudiants :**

Toutes vos questions doivent être posées dans un premier temps aux moniteurs. Si vous avez encore des questions vous prendre rendez-vous durant le TD ou par mail avec Katia Iglesias ([Katia.Iglesias@pse.unige.ch](mailto:Katia.Iglesias@pse.unige.ch)).

Support

de

cours

**Logiciels statistiques et psychologie :**

**chercher l'erreur**

1

**Désolé**

Psychologie rime bien souvent avec Statistique

2



## **Le choc des statistiques**

- Les nombres peuvent effrayer certains.
- Le choc : pour étudier la psychologie, il est nécessaire d'apprendre les statistiques.
- But du cours : faire disparaître la peur des uns et augmenter le plaisir des uns et des autres.
- → nous travaillerons à partir d'un questionnaire et sur vos propres données.

3

## **Connaissances utiles**

---

- Comprendre la littérature, la recherche, lire les résultats, pouvoir être critique, etc.
- 2ème année → votre propre recherche
- Mémoire de fin d'études

4





## Amélioration des débouchés

- Les étudiant-e-s qui disposent d'une bonne connaissance d'analyse des données, de la statistique et de la méthodologie ont de meilleures possibilités de débouchés.
- Ces capacités sont beaucoup valorisées non seulement dans la psychologie scientifique mais également dans les domaines différents de la pratique comme les cliniques et les entreprises.

5

## La psychologie et la recherche

Pour trouver une réponse à nos questions :



Réfléchir sur le monde et le comportement



Récolter des données et interpréter les résultats

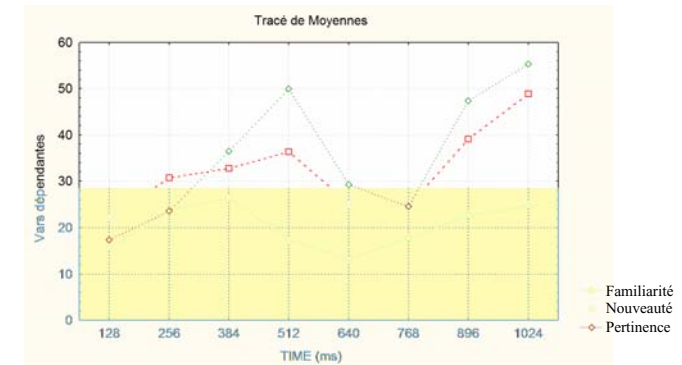
6

## Recherche : électro-encéphalogramme

### Analyses temps-fréquences des potentiels évoqués (EEG)

**But:** Tester l'occurrence temporelle de l'évaluation de la nouveauté versus de la pertinence. Les participants devaient détecter une image-cible (15% des stimuli visuels) dans un flux d'images nouvelles (15%) et familières (70%).

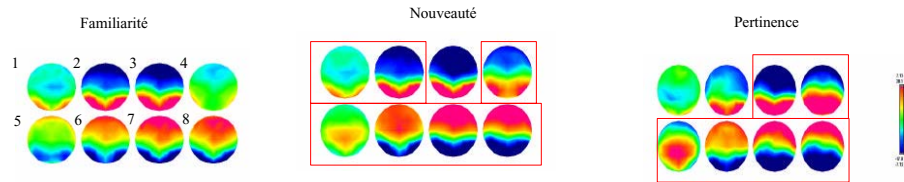
7



Tracé de l'énergie moyenne en fréquence Delta pour toutes les électrodes (N=15) obtenus par décomposition discrète d'ondelette pour les trois conditions expérimentales,

Notez que l'évaluation de la nouveauté (implicite) induit des modifications temporellement antérieures aux modifications associées à la manipulation expérimentale de la pertinence.

8



- 8 cartes topographiques (128 ms par carte) de la bande Delta (rouge = polarité positive, bleu = polarité négative) vue de dessus (partie supérieure de la carte = partie antérieure du scalp).
- Encadré en rouge les périodes durant lesquelles l'énergie dans cette bande est significativement différente comparée à la condition « familiarité » ( $p < 0.05$ ).

Extrait de Grandjean, D., Scherer, K.R. (2003). *Appraisal processes in emotion elicitation: a topographic electrophysiological approach*. Human Brain Mapping, New-York: USA.

9

## Exemple de recherche sur les adolescents

Influences maternelles sur le comportement sexuel des adolescents :

Echantillon : 751 adolescents noirs entre 14 et 17 ans de la région de Philadelphia

10

## Variables d'intérêt

**But** : analyser les différents facteurs pouvant expliquer les variables ci-dessous :

- Sur l'expérience d'une relation sexuelle ou non.
- Pour ceux qui ont déjà eu (au moins) une relation,
- Sur la fréquence de relation sexuelle et
- Sur la fréquence d'utilisation de moyen de contraception.

11

## Variables permettant d'expliquer les variables d'intérêt

3 variables nous intéressent :

- Satisfaction de la relation avec la mère
- Désapprobation de la mère pour des relations sexuelles avant le mariage
- Discussions sur les moyens de contraceptions

Il est connu que d'autres facteurs influencent le comportement sexuel des adolescents. Ces derniers ont été pris en compte :

- Pour la mère : religiosité, niveau d'éducation, âge, niveau social, etc.
- Pour l'enfant : religiosité, moyenne à l'école, etc.

12

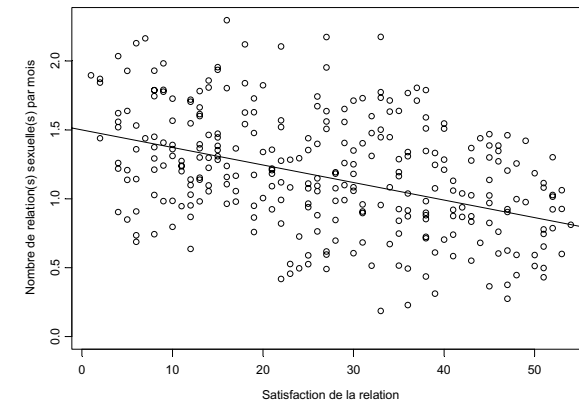
## Un résultat

Pour la fréquence de relation sexuelle, les auteurs ont trouvé :

- Satisfaction de la relation : -2.11 Influence négative forte
- Désapprobation maternelle : -1.74 Influence négative moyenne
- Discussion contraception : 0.83 Pas d'influence

13

S'ils avaient uniquement tenu compte de la relation entre la Satisfaction de la relation avec la mère et la fréquence des relations sexuelles, on aurait obtenu un graphique de ce genre :



Jaccard, J., Dittus, P. J., and Gordon V. V. (1996). Maternal Correlates of Adolescent Sexual and Contraceptive Behavior. *Family Planning Perspectives*, 28, 4, 159-165.

14

## Exemple de recherche en psychologie : problème de méthodologie

Expérience :

- Mesurer l'effet de l'âge sur les capacités cognitives
- Passation des tâches l'après-midi par commodité pour le chercheur

Erreur méthodologique :

- ne pas tenir compte que les conditions optimales de passation pour chacun des groupes sont différentes
- les personnes âgées sont plus performant le matin et les jeunes l'après-midi.

Helmuth, L., (2003). The Wisdom of the Wizen. *Science*, 299, pp. 1300-1302

15

## Informations pratiques

16

## Programme du cours

---

- La démarche de la recherche : où, quand et pourquoi utiliser SPSS ?
- Le questionnaire et la transformation de données
- Analyse des données :
  - Représentations graphiques
  - Les mesures descriptives
  - Corrélation et régression linéaire simple
- Questions
- Distribution de l'examen

17

## Programme du cours

---

- Vu le nombre important d'étudiants suivant ce cours, vous serez séparé en deux groupes pour les cours : jaune-bleu et rouge-vert.
- Le cours a lieu les jeudis de 10h à 12h en salle U300 une semaine sur deux selon le programme et selon le groupe auquel vous appartenez.

18

## Travaux dirigés

---

- Pour les travaux dirigés (TD), vous serez séparés en quatre groupes : jaune, bleu, rouge et vert.
- Ils ont lieu les mardis de 18h à 20h et les mercredis de 12h à 14h une semaine sur deux selon le programme.
- Vous pouvez vous échanger vos groupes en échangeant les programmes de couleurs (ou sur le Forum du cours).
- L'inscription aux TD se fait sur présentation du programme de couleur.

19

## Travaux pratiques

---

- Les travaux pratiques (TP) ont lieu les mercredis de 14h à 16h et les jeudis de 16h à 18h.
- Ils n'y a pas de groupes : chacun peut y aller autant qu'il en a besoin.
- Si vous avez des questions, vous les posez en premiers aux moniteurs, si ce n'est toujours pas clair vous pouvez alors prendre rendez-vous par e-mail avec Katia Iglesias.

20

## Attention

---

- Si vous avez changé de groupe, ce changement est valable aussi bien pour le cours que pour les travaux dirigés

21

## Questionnaire

---

Pour illustrer les différents concepts qui seront vu dans le cours et dans les travaux dirigés, nous utiliserons le questionnaire que nous vous avons distribué.

- Remplir 2 exemplaires par étudiants
- A amener lors du 2<sup>ème</sup> TD

22

## Examen en deux temps :

---

Vous recevez un (ou des) jeu(x) de données environ 10 jours avant l'examen.

- Vous devrez faire des analyses et apporter les résultats le jour de l'examen.
- Ces analyses seront très similaires à celles faites durant les TD et les TP.

Lors de l'examen

- A partir des analyses que vous avez faites à l'avance, vous devrez répondre aux questions.
- Questions théoriques sur le cours.

23

## Sessions d'examen:

---

Sessions usuelles :

- Eté 2006 : Distribution du ou des jeu/x de données le 15 juin 2006 lors du cours
- Automne 2006 : Disponibilité du ou des jeu/x de données le 28 septembre 2006 sur Dokeos
- Modalités : examen écrit de 2 heures

Session extraordinaire :

- Hiver 2007 : examen oral.

24

## Bibliographie

---

Dépelteau, F. (2000), *La démarche d'une recherche en sciences humaines*, Ed. De Boeck Université : Bruxelles.

Howell, D.C. (1998). *Méthodes statistiques en sciences humaines*. Bruxelles : De Boeck Université.

Kinnear, P., Gray, C. (2005). *SPSS facile appliqué à la psychologie et aux sciences sociales : maîtriser le traitement de données*. Bruxelles : De Boeck Université.

Laveault, D., Grégoire, J. (1997). *Introduction aux théories des tests en sciences humaines*. Ed. De Boeck Université : Bruxelles.

25

## Sites web

---

Technique d'analyse en psychologie (université de Montréal):  
<http://www.mapageweb.umontreal.ca/cousined/home/course/PSY1004/index.html>

Annexe 6 : Document général : [.../PSY1004/3-FormationSPSS/PSY1004\\_Annexe6-SPSS.pdf](.../PSY1004/3-FormationSPSS/PSY1004_Annexe6-SPSS.pdf)

Annexe 7 : SPSS avancé : [.../PSY1004\\_Annexe7-SPSS\\_avancée.pdf](.../PSY1004_Annexe7-SPSS_avancée.pdf)

Introduction to SPSS for the PC:

[http://abacus.bates.edu/acad/depts/psychology/SPSSPC/sps\\_spc.html](http://abacus.bates.edu/acad/depts/psychology/SPSSPC/sps_spc.html)

26

## La démarche de la recherche en sciences humaines

27

## Démarche de la recherche

---

Au départ de toute recherche : un questionnaire

Les 4 étapes de la recherche :

- Circonscrire l'objet de recherche
- Construire l'objet de recherche
- Construire les instruments d'observations
- Analyser l'objet d'observations

28

## Circonscrire l'objet de recherche

---

- C'est apprendre à poser et à formuler son questionnement dans le cadre des connaissances actuelles.
- La recherche s'inscrit dans un continuum et se situe par rapport à des courants de pensées qui la précèdent et l'influencent.
- Monter ce qui rapproche notre recherche des autres et ce qui la distingue par rapport à des cadres théoriques connus.

29

## Construire l'objet de recherche

---

- Adopter un cadre théorique : La théorie est une sorte de filtre par lequel on regarde la réalité.
- Opérationnaliser ce cadre théorique.
- « traduire une théorie abstraite et une hypothèse générale en phénomènes concrets et précis qu'on peut expérimenter, c'est-à-dire observer, toucher, sentir, goûter ou entendre » (Dépelteau, 2002, p.170)

30

## Construire l'objet de recherche

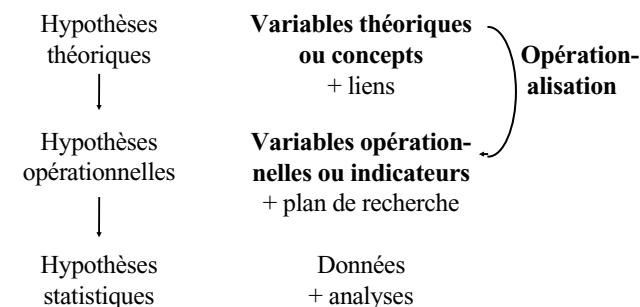
---

- Pour cela il faut déterminer :
  - ◆ les dimensions des concepts
  - ◆ les composantes des dimensions
  - ◆ les indicateurs des composantes
- « Une variable est un élément de l'hypothèse à laquelle on peut attribuer diverses valeurs et qui, comme son nom l'indique, varie » (Dépelteau, 2002, p.163)

31

## Opérationnalisation

---



32

## Construire les instruments d'observation

- Observer

- ◆ « Quoi » : quelles sont les données pertinentes ?
- ◆ « Qui » : qui ou quoi doit être observé ?
- ◆ « Comment » : comment se déroulera le test empirique (mode d'investigation) et comment recueillir les données permettant de tester notre hypothèse de recherche ?

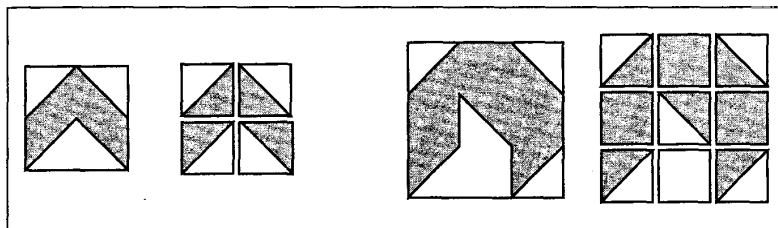
33

## Comment observer ?

- Questionnaire
- Entretien
- Tests
- Observation
- Mesures psychophysiologiques
- Imagerie cérébrale
- Evaluation ambulatoire
- Analyse du texte
- Mesures biochimiques
- ...

34

## Exemple: Test



Exemple de problèmes  
(à gauche la figure à reproduire, à droite la solution)

Figure 4.2  
Test de cubes de Kohs  
(Copyright 1960, Centre de psychologie appliquée)

35

## Exemple: Observation

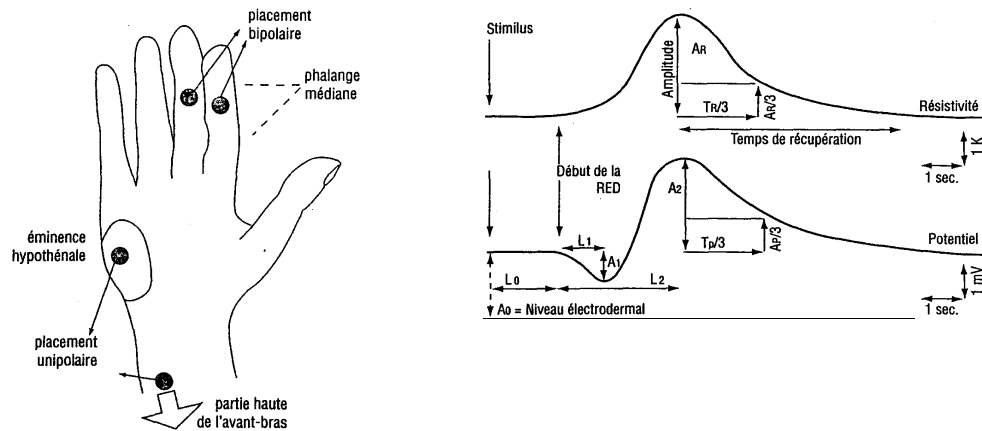
L'échelle HOME (Bradley et Caldwell, 1976) : Réactivité émotionnelle et verbale de la mère

- La mère s'adresse spontanément à l'enfant deux fois au moins pendant la visite.
- La mère répond verbalement à des verbalisations.
- La mère nomme des objets ou des personnes de manière didactique.
- Le langage de la mère est distinct, clair et audible.
- La mère prend l'initiative d'échanges verbaux avec l'observateur, pose des questions, fait des commentaires.
- La mère exprime ses idées librement et facilement et utilise des phrases de longueur appropriées à la conversation (plus que des réponses brèves).
- La mère laisse l'enfant mettre de fouillis en jouant.
- La mère fait l'éloge des qualités ou de la conduite de l'enfant au moins deux fois.
- Lorsque la mère parle à l'enfant, le ton de sa voix est affectueux.
- La mère caresse ou embrasse l'enfant au moins une fois.
- La mère répond positivement aux éloges de l'enfant faits par l'observateur.

36

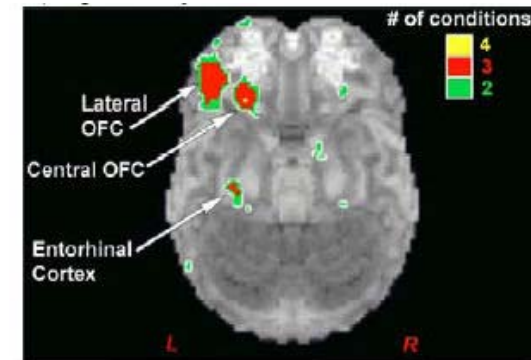


## Exemple: Mesures psychophysiological



37

## Exemple: Imagerie cérébrale



38

## Exemple: Evaluation ambulatoire



39

## Analyser l'objet d'observation

Pourquoi analyser les données récoltées ?

- Pour tester notre hypothèse de départ.

→ besoin d'un logiciel statistique.

40

## Analyser l'objet d'observation : 3 opérations

Préparation des données (décrire et agréger).

Analyse des relations entre les variables.

Comparaison des résultats observés avec les résultats attendus et interprétation de l'écart.

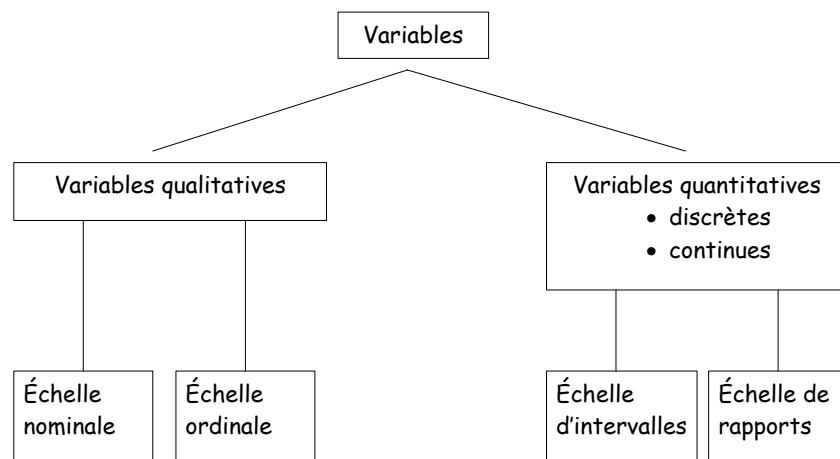
41

## Analyse statistique

- Statistique descriptive - inférentielle.
- Description univariée - bivariée - multivariée.
- Approche exploratoire - confirmatoire.

42

## Les échelles



43

## Variables qualitatives

Modalités : Catégories (mots, lettres), non numériques

- sexe, étude, profession, nationalité, niveau scolaire

Variable dichotomique/binaire : deux catégories

- sexe (féminin, masculin), résolution vs. non-résolution d'une tâche, approbation vs. désapprobation d'un énoncé

Variable polytomique : plusieurs catégories

- degré d'approbation

pas du tout d'accord	peu d'accord	plutôt d'accord	tout à fait d'accord
----------------------	--------------	-----------------	----------------------

44

## Variables qualitatives: Assigner des nombres

### L'échelle nominale

- On utilise les échelles nominales pour représenter la différence entre les sujets.
- Le même nombre est assigné à toutes les personnes ayant les mêmes caractéristiques.
- Les nombres ne possèdent pas de propriétés arithmétiques.

45

## Variables qualitatives: Assigner des nombres

### L'échelle ordinale

- On utilise les échelles ordinales pour représenter la différence et l'ordre entre les sujets (ordre selon la taille, l'intensité, etc.).
- Le même nombre est assigné à toutes les personnes ayant les mêmes caractéristiques.
- Un nombre plus grand est assigné à une personne appartenant à une catégorie plus grande.
- Les nombres ne possèdent pas de propriétés arithmétiques.

46

## Variables quantitatives (variables métriques)

### Les modalités ont des valeurs numériques

- la température, la taille, le revenu, le quotient intellectuel, le nombre de frères

### Variables discrètes : nombre de valeurs possibles dénombrable

- le nombre de frères, le nombre de mots dans un texte, le nombre d'accès de colère dans une semaine

### Variables continues : nombre de valeurs possibles non dénombrable

- La température

47

## Variables quantitatives : Niveaux de mesure

### L'échelle d'intervalles (ou continue)

- Pour représenter la différence et l'ordre entre les sujets. Contrairement à l'échelle ordinale les distances entre les valeurs ont un sens.
  - La température : la différence entre la température aujourd'hui (24° Celsius) et hier (22° Celsius) et la même qu'entre hier (22° Celsius) et avant-hier (20° Celsius) parce que la différence est la même (2° Celsius).
- Le point zéro est déterminé arbitrairement
  - transformation de la température mesurée en l'échelle Celsius en la température mesurée en Fahrenheit :  $\text{Fahrenheit} = 9/5 \text{ Celsius} + 32$

48

## Variables quantitatives : Niveaux de mesure

### L'échelle de rapport

- Pour représenter la différence et l'ordre entre les sujets.
- Les distances et le point de zéro sont significatifs.
  - La longueur

49

## Le questionnaire utilisé dans ce cours

50

## Le questionnaire : Echelle d'expressivité

- Amour
- Joie
- Peur
- Colère
- Tristesse
- Honte

51

## Le questionnaire : Big Five

- Névrosisme (susceptible, sensible, lunatique, pas sûr-e de moi)
- Extraversion (amical-e, sociable, vif/ve, plein-e de tempérament)
- Intellect (malin/gne, intelligent-e, cultivé-e, riche d'esprit)
- Agréabilité (chaleureux/se, plein-e d'égards, de bonne nature, serviable)
- Conscience (travailleur/se, assidu-e, consciencieux/se, ambitieux/se)

52

## Le questionnaire : Echelle d'ouverture

- Ouverture à la rêverie,
- Ouverture à l'esthétique,
- Ouverture aux sentiments,
- Ouverture aux idées,
- Ouverture aux actions et
- Ouverture aux valeurs

53

## Le questionnaire : données sociodémographiques

- Age
- Sexe
- Taille
- Poids
- Nationalité
- Etude
- ...

54

## Différents types de questions

### Questions fermées :

	pas du tout d'accord	un peu d'accord	moyenne- ment d'accord	plutôt d'accord	tout à fait d'accord
J'ai peu d'intérêt pour la réflexion sur la nature de l'univers ou sur la condition humaine.	1	2	3	4	5

### Questions ouvertes :

Qu'est ce que ces loisirs vous apportent ?

---

---

55

## Différents types de questions

Questions à réponse unique : Sexe :  Femme  Homme

### Questions à réponses multiples :

Quelles loisirs avez-vous régulièrement? (*plusieurs réponses sont possibles*)

- Sport
- Rencontrer des amis
- Activités culturelles (cinéma, expositions, concerts, etc.)
- Animal domestique
- Instrument de musique
- Télévision
- Internet
- Bricolage (tricot, jardinage, etc.)
- Autre : \_\_\_\_\_

56

## Du questionnaire à la base de données

- N individus (objets, unités statistiques) qui ont répondu au questionnaire
- P réponses ou variables (caractéristiques, attributs) par questionnaire

→ L' « espace d'observation » comme résultats du produit cartésien entre ces deux ensembles d'éléments : la matrice et le fichier des données.

57

## Du questionnaire à la base de données

Créer un fichier de données → pour cela, il faut élaborer et utiliser un plan de codage

58

### Tableau de codage pour notre questionnaire

Variable	Variable label	Value labels
id	Numéro d'identification	
exp1 – exp12	Expressivité	1 = jamais ou presque jamais 2 = parfois 3 = souvent 4 = presque toujours ou toujours
bf1 – bf20	<i>Big Five (adjectives)</i>	1 = pas du tout 5 = tout à fait
ouv1-ouv30	Ouverture	1 = pas du tout d'accord 5 = tout à fait d'accord
loisirs1 – loisirs8	Loisirs	0 = n'a pas ce loisir 1 = a ce loisir
sexe	Sexe	1 = Femme 2 = Homme
age	Age	

59

### Tableau de codage pour notre questionnaire

natio	Nationalité		
		1 allemande	19 roumaine
		2 coréenne de sud	20 néerlandaise
		3 mexicaine	21 indienne
		4 slovaque	22 grecque
		5 italienne	23 israélienne
		6 taiwanaise	24 belge
		7 finnoise	25 chinoise
		8 russe	26 marocaine
		9 française	27 autrichienne
		10 polonaise	28 espagnole
		11 tchèque	29 bulgare
		12 ukrainienne	30 bolivienne
		13 moldavienne	31 lituanienne
		14 usbekienne	32 palestinienne
		15 turque	33 islandaise
		16 Etats-Unis	34 britannique
		17 luxembourgeoise	35 croate
		18 hongroise	36 suisse

60

## Tableau de codage pour notre questionnaire

etude	Domaine d'études	1 mathématiques	10 économie
		2 psychologie	11 histoire
		3 droit	12 architecture
		4 langues	13 design
		5 géologie	14 informatique
		6 éducation	15 médias
		7 ingénierie	16 théologie
		8 politique	17 philosophie
		9 sociologie	18 ethnologie
taille	Taille en cm		
poids	Poids en kg		
code	Code		

61

## Transformer les données

62

## Transformer les données

Pourquoi transformer les données ?

- Modifier la structure (la forme) d'une variable (recodage) → modifier la manière dont la variable a été initialement codée pour satisfaire aux exigences de certaines analyses.
- Créer une nouvelle variable → création d'une nouvelle variable à l'aide d'opérations arithmétiques et/ou d'opérations logiques sur une ou plusieurs variable(s) d'origine.

63

## Recoder une variable : Exemple 1 : inverser l'échelle

On cherche à mesurer l'ouverture aux sentiments avec les 5 questions suivantes (échelle : 1 pas du tout d'accord à 5 tout à fait d'accord) :

8	La manière dont je sens les choses est importante pour moi.
14	Il est rare que j'accorde beaucoup d'attention à mes sentiments du moment.
19	J'éprouve une grande variété d'émotions ou de sentiments.
24	Je remarque rarement les changements d'humeurs ou les sentiments que provoquent des environnements différents.
27	Je trouve facile d'avoir de l'empathie, c'est-à-dire de ressentir moi-même ce que les autres ressentent.

Les questions 8, 19 et 27 mesurent l'ouverture aux sentiments et les questions 14 et 24 la « fermeture »

64

→ recoder les questions 14 et 24 pour que l'échelle mesure bien l'ouverture aux sentiments

Pour cela, on attribue la valeur 1 à la valeur 5, la valeur 2 à la valeur 4, la valeur 3 à la valeur 3, la valeur 4 à la valeur 2 et la valeur 5 à la valeur 1 :

Anciennes valeurs	Nouvelles valeurs
1	5
2	4
3	3
4	2
5	1
Missing	Missing

65

## Exemple 2 : restreindre l'échelle

a) Passage de 4 modalités à 2 :

	Ancienne variable				→	Nouvelle variable	
	Pas satisfait	Plutôt pas satisfait	Plutôt satisfait	Satisfait		Insatisfait	satisfait
Etes-vous satisfait avec la manière dont vous gérez vos émotions ?	1	2	3	4		1	2

66

## Exemple 2 : restreindre l'échelle

b) Age : variable quantitative allant de 16 à 35 ans

→ Variable recodée en quatre modalités :

- 16 - 20 ans
- 21 - 25 ans
- 26 - 30 ans
- 31 - 35 ans

67

## Exemple 2 : restreindre l'échelle

c) Nationalité :

1 allemand	7 finnois	13 moldavien	19 roumain	25 chinois	31 lituanien
2 coréen du sud	8 russe	14 usbekien	20 néerlandais	26 marocain	32 palestinien
3 mexicain	9 français	15 turque	21 indien	27 autrichien	33 islandais
4 slovaque	10 polonais	16 Etats-Unis	22 grecque	28 espagnol	34 britannique
5 italien	11 tchèque	17 luxembourgeois	23 israélien	29 bulgare	35 croate
6 taiwanais	12 ukrainien	18 hongrois	24 belge	30 bolivien	36 suisse

- Ouest de l'Europe : 1, 5, 7, 9, 17, 20, 22, 24, 27, 28, 33, 36
- Est de l'Europe : 4, 8, 10, 11, 12, 13, 15, 18, 19, 29, 31, 35
- Asie : 2, 6, 14, 15, 21, 23, 25, 32
- Amérique : 3, 16, 21, 30
- Afrique : 26

68



### Créer une nouvelle variable : Ex 1 : une sous-échelle

Pour construire la sous échelle d'ouverture aux sentiments, il a fallu dans un premier temps recoder deux des cinq variables mesurant cette dimension.

A ce stade, nous avons 5 questions qui mesurent l'ouverture aux sentiments (Q : 8, 14r, 19, 24r et 27, où r signifie « renversé »).

- But : passer de 5 mesures à une seule  
→ moyenne de nos 5 questions :  
Mean (ouv08,ouv14r,ouv19,ouv24r,ouv27)

69

### Exemple 2 : l'indice de masse corporelle (IMC)

L' « Indice de masse corporelle » s'obtient en divisant le poids d'un individu (exprimé en kg) par le carré de sa taille (en mètres).

A partir des variables "Poids" et "Taille" figurant dans un fichier, on peut donc créer la nouvelle variable "Masse corporelle" : poids/taille<sup>2</sup>

70

### Exemple : croisement sexe-âge

Connaissant le sexe et l'âge d'un groupe d'individus, on peut créer une variable composite en définissant 4 modalités :

	Homme	Femme
Moins de 25 ans	1	3
25 ans et plus	2	4

- 1 : Hommes d'âge inférieur à 25 ans
- 2 : Hommes de 25 ans ou plus
- 3 : Femmes d'âge inférieur à 25 ans
- 4 : Femmes de 25 ans ou plus

71

## **Analyses des données : représentations graphiques et mesures descriptives**

72

## Description et représentation

Description des caractéristiques d'un ensemble d'observations / d'individus à partir d'une seule variable

→ Deux familles de moyens :

- des tableaux (tableau d'effectifs) et des graphiques (diagramme en bâtons, diagramme circulaire, histogramme).
- des indices numériques d'autre part.

73

## Distribution de fréquences

Fréquence absolue	Fréquence relative	Fréquence cumulée
$\sum_{i=1}^k n_i = n$	$f_i = \frac{n_i}{n}$	$F_1 = f_1$
		$F_2 = f_1 + f_2$
		$F_3 = f_1 + f_2 + f_3$
		⋮
		$F_k = f_1 + \dots + f_k = 1$

$n_i$  : nombre d'éléments ayant pour une modalité / valeur d'une variable X et  $n$  : nombre total d'éléments

Note : Pourcentage = fréquence \* 100

74

Sexe

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Femme	307	63.7	63.7	63.7
Homme	175	36.3	36.3	100.0
Total	482	100.0	100.0	

Joie

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid jamais ou presque jamais	1	.2	.2	.2
parfois	25	5.2	5.3	5.5
souvent	194	40.2	40.8	46.2
presque toujours ou toujours	256	53.1	53.8	100.0
Total	476	98.8	100.0	
Missing System	6	1.2		
Total	482	100.0		

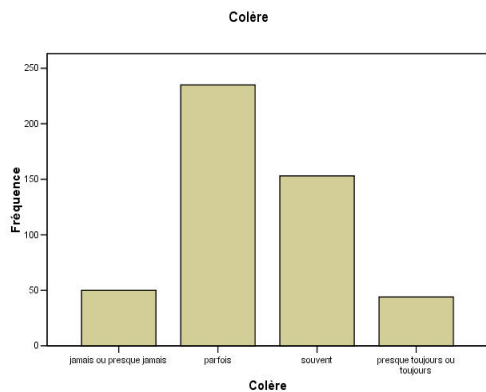
75

Age

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 17	1	.2	.2	.2
18	5	1.0	1.0	1.3
19	36	7.5	7.5	8.8
20	51	10.6	10.6	19.4
21	66	13.7	13.8	33.1
22	81	16.8	16.9	50.0
23	50	10.4	10.4	60.4
24	41	8.5	8.5	69.0
25	49	10.2	10.2	79.2
26	22	4.6	4.6	83.8
27	25	5.2	5.2	89.0
28	15	3.1	3.1	92.1
29	10	2.1	2.1	94.2
30	8	1.7	1.7	95.8
31	5	1.0	1.0	96.9
32	4	.8	.8	97.7
33	3	.6	.6	98.3
35	8	1.7	1.7	100.0
Total	480	99.6	100.0	
Missing System	2	.4		
Total	482	100.0		

76

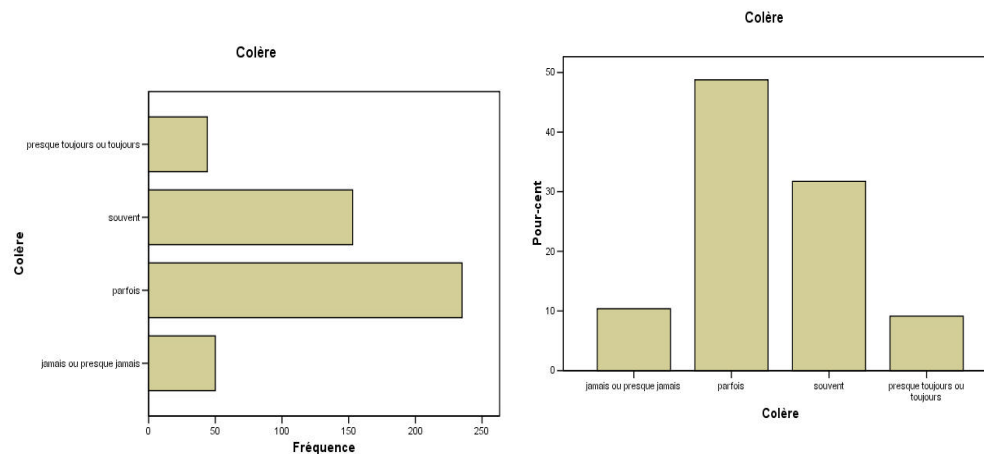
## Diagramme en bâtons (graphique à colonnes, bar plot)



- Une colonne verticale ou horizontale est dessinée pour chaque valeur ou modalité de la variable considérée
- Hauteur/longueur : nombre de membres de chaque classe

77

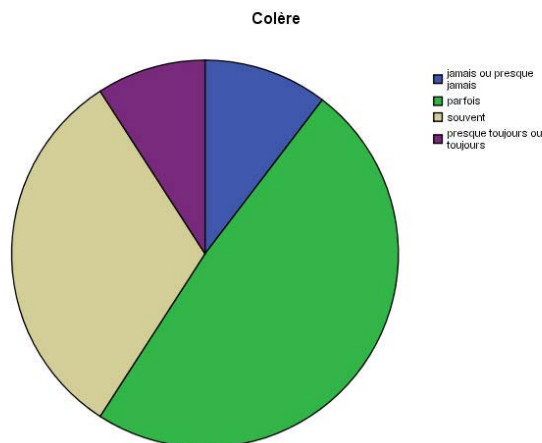
## Diagramme en bâtons



78

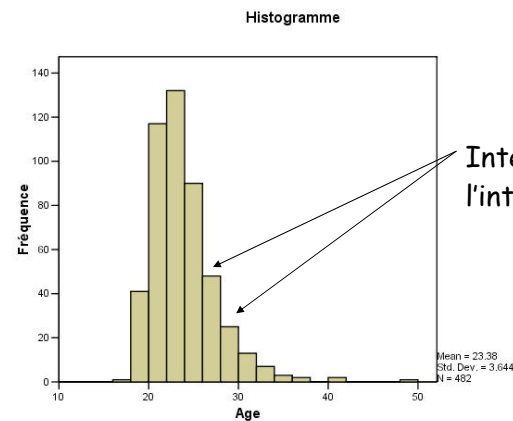
## Diagramme circulaire: camembert, pie-chart

- Cercle : Population totale
- Tranches (aire ou angle) : fréquence relative



79

## Histogramme (vars quantitatives continues)



80

## Histogramme (vars quantitatives continues)

Pour construire un histogramme, il faut :

- Choisir les intervalles (intervalles qui doivent être de mêmes tailles)
- Compter le nombre de cas (observations) dans chaque intervalle

81

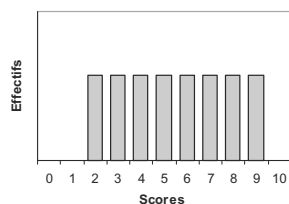
## Mesure de tendance (centrale)

- Mode (*mode*)
- Médiane (*median*)
- Moyenne (*mean, average*)

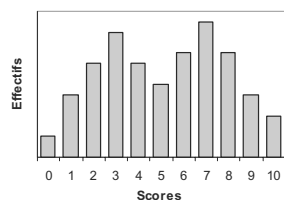
82

## Mode

- Le mode d'une variable qualitative (ou quantitative discrète) est la valeur qui possède la fréquence la plus élevée.
- Il y a des distributions unimodale, bimodale, plurimodale



Distribution uniforme (rectangulaire)



Distribution bimodale

83

## Médiane

- La médiane partage la distribution d'une série d'observations en deux parties égales.
- Les variables doivent être mesurées au moins sur une échelle ordinale.
- Pour calculer la médiane il faut ordonner les valeurs.
- Calcul :
  - Le nombre d'observations est impair : la médiane est la valeur située au milieu de la série
  - Le nombre d'observations est pair : la médiane est la moyenne de la  $\left(\frac{n}{2}\right)^e$  est la  $\left(\frac{n+2}{2}\right)^e$

84

## Moyenne

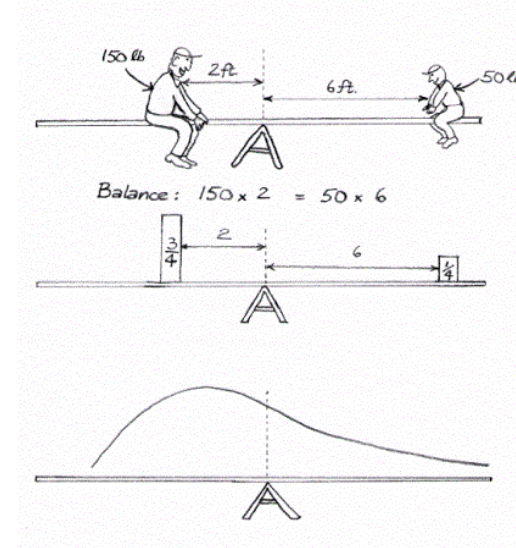
- Moyenne arithmétique

- Valeurs individuelles :

- $x_i$  : valeur individuelle d'un sujet  $i$
- $n$  : nombre de sujets

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

85



Extrait de M. A. Martin, (2003). *Journal of Statistics Education*, 11

86

## Mesure de dispersion et de forme

- Variance et écart-type
- Empan
- Intervalle interquartile
- Box plot
- Coefficient de symétrie et d'aplatissement

87

## Variance (variance)

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Écart-type (*standard error*) :  $s = \sqrt{s^2}$

### Propriétés

- La variance a toujours une valeur non-négative.
- La variance est égale à 0, si toutes les observations sont identiques.

88

## Variance

Estimateur de la variance dans une population grâce à un échantillon :

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

89

## Empan (étendue, range)

Différence entre la valeur la plus élevée et la valeur la plus basse

90

## Intervalle interquartile (*interquartile range*)

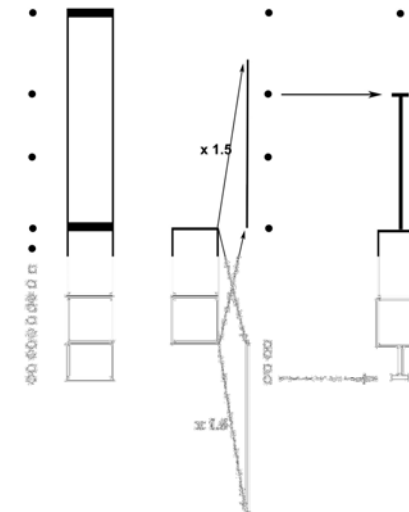
Les quartiles divisent un ensemble d'observations en quatre parties égales (25%).

L'intervalle interquartile est la distance entre le 1er et le 3e quartile.

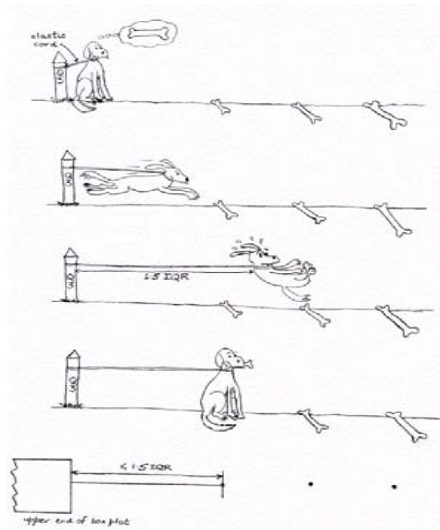
L'intervalle interquartile comprend 50 % des observations les plus au centre de la distribution.

91

## Box plot

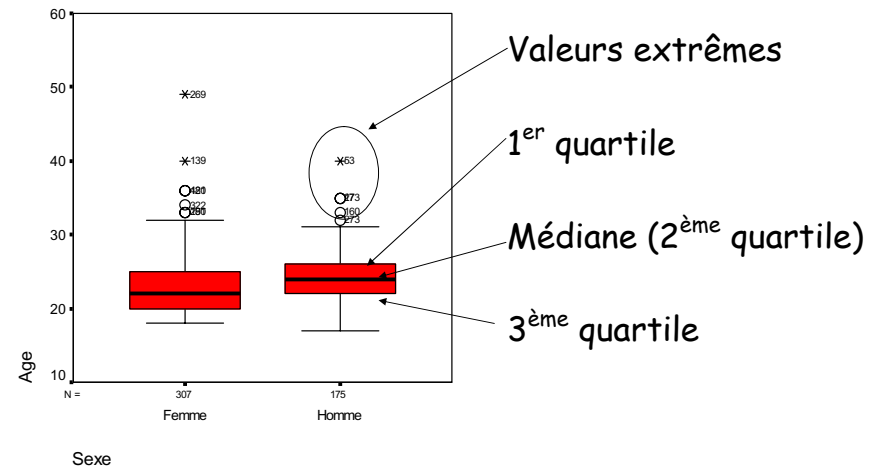


92



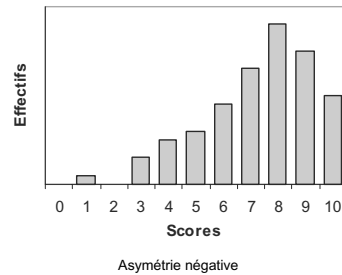
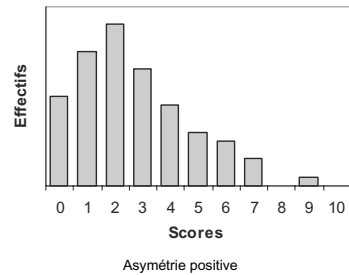
Extrait de M. A. Martin, (2003). *Journal of Statistics Education*, 11

## Box plot



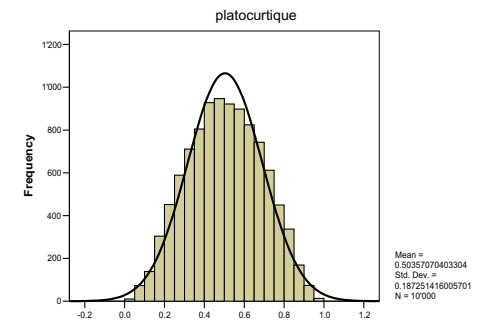
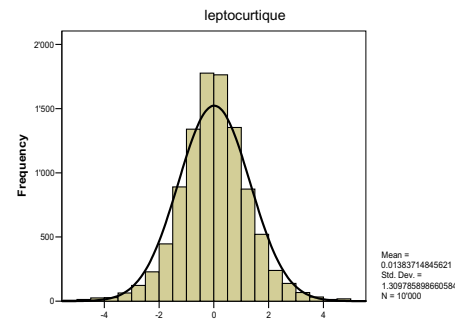
## Mesures de forme : Mesure d'asymétrie

- valeur = 0: symétrie
- valeur > 0: oblique à gauche (ou étalement à droite)
- valeur < 0: oblique à droite (ou étalement à gauche)



## Mesures de forme : Mesure d'aplatissement

- valeur = 0: distribution normale (« cloche de Gauss »)
- valeur > 0: distribution leptocurtique (« tassé »)
- valeur < 0: distribution platocurtique (« plate »)



## Analyses des données bivariées: corrélation et régression linéaire simple

97

## Relation entre deux variables

---

Description des caractéristiques d'un ensemble d'observations / d'individus à partir de deux variables considérées simultanément : étude de la relation entre variables (statistique bivariée).

- Corrélation (Spermann, Kendall, Pearson)
- Régression linéaire simple

98

## Corrélation (*correlation*)

---

La corrélation mesure la relation linéaire entre deux variables

La corrélation peut prendre des valeurs entre -1 et 1 avec

- proche de -1 : forte relation négative,
- proche de 0 : pas de relation linéaire et
- proche de 1 : forte relation positive.

99

## Corrélation

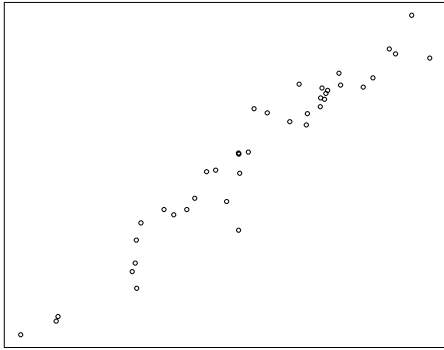
---

Primordial de vérifier la linéarité de la relation, car :

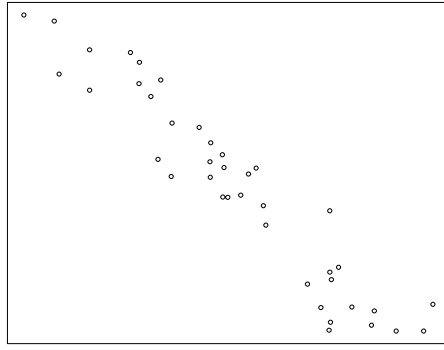
- Corrélation linéaire de valeur nulle ne veut pas dire qu'il n'existe pas de relation entre les variables → p.ex. relation quadratique.
- La mesure de la corrélation est très sensible aux données extrêmes → possible d'obtenir une corrélation linéaire significative même s'il n'y a pas de relation entre les variables.

100

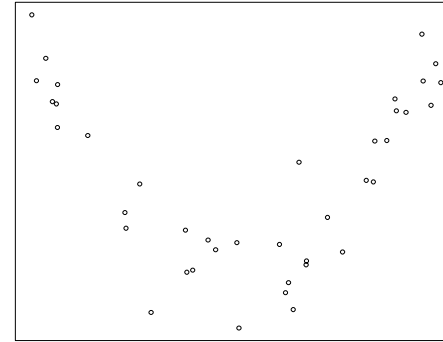




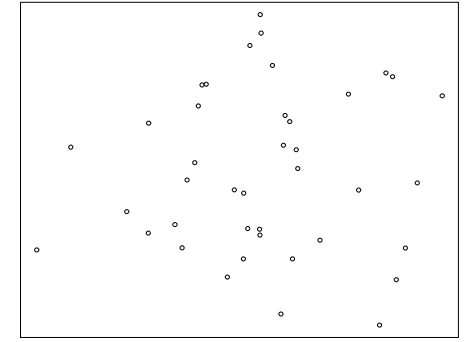
Forte relation linéaire positive



Forte relation linéaire négative



Pas de relation linéaire, mais relation

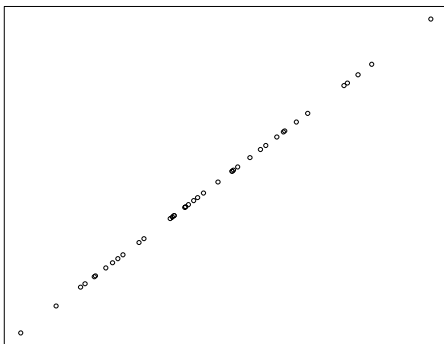


Pas de relation

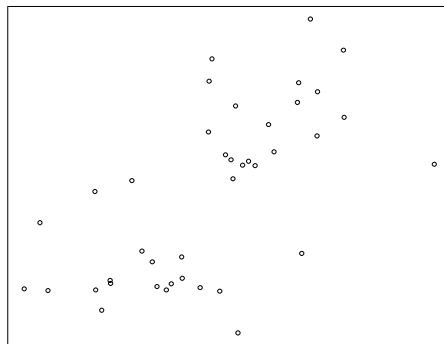
## Relation entre variables quantitatives

Deux variables quantitatives de types d'intervalles ou de rapport :

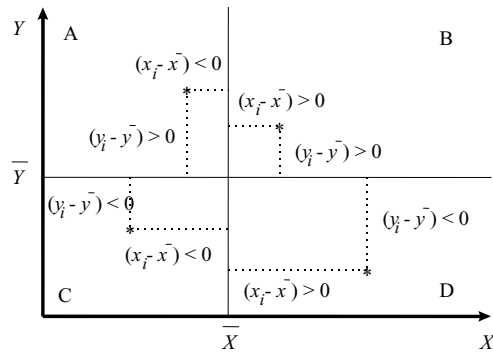
- corrélation avec le coefficient de corrélation linéaire "r" de Bravais-Pearson et le coefficient de détermination



Relation linéaire positive parfaite



Relation linéaire positive faible



- 1) Données sont en majorité dans les quadrants B et C => corrélation positive
- 2) Données sont en majorité dans les quadrants A et D => corrélation négative
- 3) Données sont partout => pas de corrélation

105

## Corrélation de Bravais-Pearson

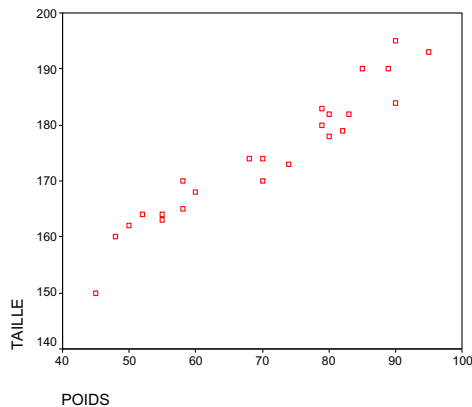
La corrélation  $r$  de Pearson est une mesure d'association paramétrique pour deux variables aléatoires continues

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Avec  $\bar{x}$  la moyenne de  $x$  et  $\bar{y}$  la moyenne de  $y$

106

## Relation taille-poids



Corrélations

	TAILLE	POIDS
TAILLE	Corrélation de Pearson	1
	Sig. (bilatérale)	.961**
	N	24
POIDS	Corrélation de Pearson	.961**
	Sig. (bilatérale)	.000
	N	24

\*\* La corrélation est significative au niveau 0.01

107

## Relation entre variables qualitatives

Deux variables qualitatives de types nominales :

- tableau croisé (ne sera pas traité dans le cours).

Deux variables qualitatives de types ordinales :

- corrélation de rangs avec les coefficients de corrélation rhô de Spearman et tau de Kendall.

108

## Corrélation rhô de Spearman

---

Le rhô de Spearman est une mesure d'association non paramétrique basée sur le rang des valeurs des données.

Il peut être utilisé pour des données qualitatives et pour des valeurs extrêmes. Pour cela, remplacez les valeurs par des rangs, c'est-à-dire remplacez  $x_i$  par  $R_i$  et  $y_i$  par  $S_i$

109

## Corrélation rhô de Spearman

---

$$\rho = \frac{\sum (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum (R_i - \bar{R})^2 \sum (S_i - \bar{S})^2}}$$

Avec  $R_i$  est le rang de la  $i^{\text{ème}}$  valeur de  $x$ ,  $S_i$  est le rang de la  $i^{\text{ème}}$  valeur de  $y$ ,  
 $\bar{R}$  est la moyenne des valeurs  $R_i$  et  $\bar{S}$  est la moyenne des valeurs  $S_i$

110

## Corrélation tau de Kendall

---

Le tau de Kendall est une mesure d'association non paramétrique basée sur le nombre de concordance et de non concordance dans les paires d'observations.

La concordance a lieu lorsque les paires d'observations varient ensemble et la discordance lorsque les paires d'observations varient de manière différente.

111

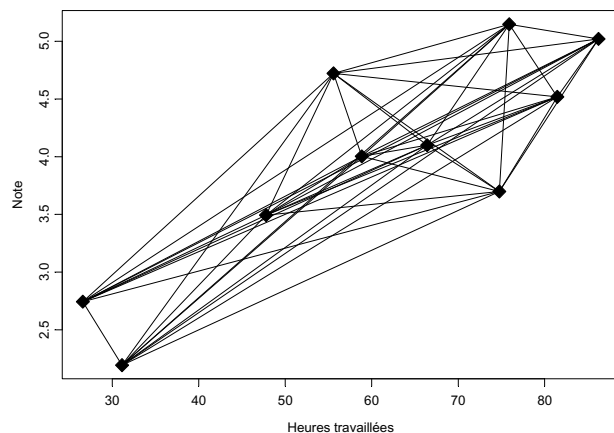
## Corrélation tau de Kendall

---

On compte combien de segments sont orientés vers le haut et combien sont orientés vers le bas :

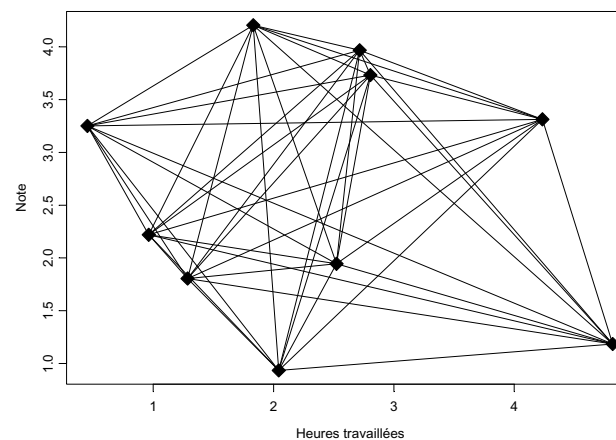
112

## Corrélation tau de Kendall : relation



113

## Corrélation tau de Kendall : pas de relation



114

## Corrélation tau de Kendall

$$\tau = \frac{\sum_{i < j} \text{sgn}(x_i - x_j) \text{sign}(y_i - y_j)}{\sqrt{(T_0 - T_1)(T_0 - T_2)}}$$

$$T_0 = n(n-1)/2$$

$$T_1 = \sum t_i(t_i - 1)/2$$

$$T_2 = \sum u_i(u_i - 1)/2$$

$t_i$  est le nombre des valeurs liées à x dans le  $i^{\text{ème}}$  groupe des valeurs liées à x

$u_i$  est le nombre des valeurs liées à y dans le  $i^{\text{ème}}$  groupe des valeurs liées à y

$n$  le nombre d'observations

$\text{sgn}(z)$  est défini par:

$$\text{sgn}(z) = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{if } z = 0 \\ -1 & \text{if } z < 0 \end{cases}$$

115

## Relation positive

Corrélations

			exp02 Joie	exp08 Bonheur
Tau-B de Kendall	exp02 Joie	Coefficient de corrélation	1,000	,514**
		Sig. (bilatérale)	.	,000
		N	482	482
	exp08 Bonheur	Coefficient de corrélation	,514**	1,000
		Sig. (bilatérale)	,000	.
		N	482	482
Rho de Spearman	exp02 Joie	Coefficient de corrélation	1,000	,541**
		Sig. (bilatérale)	.	,000
		N	482	482
	exp08 Bonheur	Coefficient de corrélation	,541**	1,000
		Sig. (bilatérale)	,000	.
		N	482	482

\*\* . La corrélation est significative au niveau 0,01 (bilatéral).

116

## Relation négative

Corrélations

			ouv14 il est rare que j'accorde beaucoup d'importance aux sentiments que je ressens	ouv27 je trouve facile de sentir de l'empathie
Tau-B de Kendall	ouv14 il est rare que j'accorde beaucoup d'importance aux sentiments que je ressens	Coefficient de corrélation	1,000	-,169**
		Sig. (bilatérale)	.	,000
		N	480	478
Rho de Spearman	ouv27 je trouve facile de sentir de l'empathie	Coefficient de corrélation	-,169**	1,000
		Sig. (bilatérale)	,000	.
		N	478	480

\*\* . La corrélation est significative au niveau 0,01 (bilatéral).

117

## Régression linéaire simple

Permet de mesurer la relation linéaire entre deux variables, tout comme la corrélation, mais elle donne une direction à la relation → c'est la VI qui explique, qui est la cause de la variation de la VD.

Le but de cette mesure est de savoir (1) si cette relation est présente (2) quelle est sa force (quel pourcentage de la variance de la VD, la VI peut expliquer) et (3) comment utiliser ce résultat pour pouvoir prédire la VD sur d'autres sujets pour lesquels on connaît la valeur de la VI.

118

## Régression linéaire simple

- Ex : expliquer le revenu (la VD) par le nombre d'année d'étude (la VI), cela signifie que nous cherchons à connaître le pourcentage de variance du revenu qui peut être expliqué par le fait d'avoir fait peu ou beaucoup d'étude.

119

## Régression linéaire simple

Linéarité de la relation : comme pour la corrélation, la régression mesure la relation linéaire entre deux variables continues.

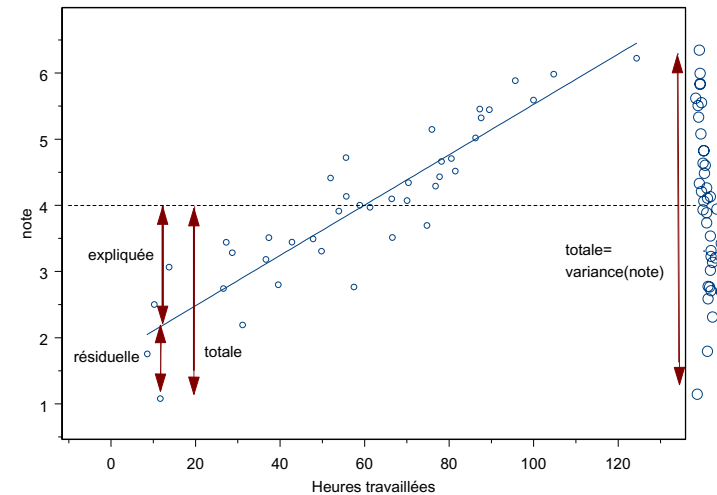
120

## Régression linéaire simple : R-deux ajusté

R-deux ajusté : pourcentage de variance expliquée par la variable indépendante.

R-deux ajusté entre 0 et 1

- proche de 1 : la valeur VI prédit presque exactement la valeur de la VD.
- proche de 0 : la VI n'arrive pas à prédire la VD.



121

122

## Régression linéaire simple

Equation de la droite de régression :

$$Y = a + B \cdot X$$

Y : la variable dépendante (VD)

X : la variable indépendante (VI)

a : ordonnée à l'origine (*intercept*) de la droite de régression

B : pente (*slope*) de la droite de régression

123

## Exemple

Hypothèse : Le nombre d'heure de travail augmente la note obtenue à un examen. Ce qui signifie que le nombre d'heures passées à travailler a un effet sur la note obtenue. Il y a une direction sur l'effet. Nous cherchons l'impact des heures de travail (la VI) sur la note (la VD).

Avant de pouvoir faire l'analyse, il faut vérifier qu'il y a bien une relation linéaire entre nos deux variables, ce qui a déjà été fait précédemment.

124

- 1) La statistique inférencielle nous dirait si on peut affirmer que la relation est significative (cours Stat II).
- 2) Combien le nombre d'heures de travail explique de variance de la note ? → R-deux ajusté est de 82.3% (0.823) dans notre exemple, ce qui très élevé. Les heures de travail permettent d'expliquer presque totalement la variance des notes obtenus.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.909 <sup>a</sup>	.827	.823	.49181	.827	181.846	1	38	.000

a. Predictors: (Constant), heurestrav

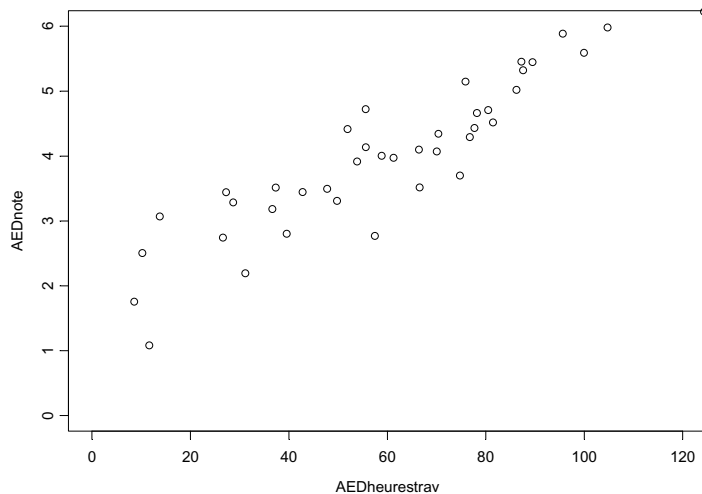
Le tableau ci-dessous permet d'écrire l'équation de régression qui exprime la relation linéaire existant entre la VD et la VI :  
 Note = -27.090 + 21.750\*nb d'heure de travail

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.722	.186		9.249	.000
	heurestrav	.038	.003	.909	13.485	.000

a. Dependent Variable: note

Ce qui signifie que lorsqu'une personne travail une heure de plus, elle augmente en moyenne sa note de 0.038 point. Donc avec 13 de travail supplémentaire, on peut augmenter en moyenne sa note d'un demi point



NATURE DES DONNEES	QUALITATIVES		QUANTITATIVES		
	ECHELLES	NOMINALE	ORDINALE	INTERVALLE	RAPPORT
1 V A R	INDICES DE TENDANCE (DE POSITION)	Mode(s)	Médian	Moyenne arithmétique	
	INDICES DE VARIABILITE (DE DISPERSION)		Ecart interquartile	Variance Ecart-type Coefficient de variation	
2 V A R	INDICES D'ASSOCIATION OU DE DEPENDANCE		Coefficient rho de Spearman Coefficient tau de Kendall	Coefficient r de Bravais-Pearson Analyse de régression	





**Travaux Dirigés**

**et**

**Travaux Pratiques**

## Table des matières

TD1 : Découverte de SPSS .....	4
Qu'est-ce que SPSS ? .....	4
Comment démarrer SPSS .....	4
L'éditeur de données .....	5
Menu File : Ouvrir un fichier de données avec SPSS .....	6
Menu File : Obtenir les informations sur les variables dans un format imprimable .....	9
Menu File : Ouvrir et sauvegarder des fichier de données .....	9
Informations compactes sur les variables dans un fichier de données .....	9
Les différentes fenêtres de SPSS .....	10
Fenêtre de données (Data Editor) .....	10
Fenêtre des résultats (Output Editor) .....	10
Fenêtre de syntaxe (Syntax Editor) .....	10
Menu Graphs : refaire le graphe du cours .....	11
Menu Help .....	13
TD2 : Entrer les données à partir du questionnaire .....	15
Entrer les données directement dans SPSS .....	15
Encoder le questionnaire .....	15
Créer un nouveau fichier de données dans SPSS .....	16
Comment coder les variables alphanumériques/série de caractères ? .....	18
Comment coder les réponses à réponses courtes ? .....	18
Comment coder les réponses multiples ? .....	18
Comment coder les réponses ouvertes ? .....	18
Comment coder les valeurs manquantes ? .....	19
Conseil et remarques .....	19
Réduire les erreurs en entrant les données .....	19
Eliminer et insérer des observations/variables dans la vue de données .....	19
Fusionner des fichiers de données .....	20
<i>Ajouter des observations</i> .....	20
<i>Ajouter des variables</i> .....	21
Utiliser Excel pour entrer les données dans SPSS .....	22
Créer une grille de données avec Excel .....	22
Importer les données dans SPSS .....	22
Utiliser un fichier texte pour entrer les données dans SPSS .....	23
Créer une grille de données avec un fichier texte .....	23
Importer les données dans SPSS .....	23
Eliminer les erreurs de l'entrée de données .....	24
TP2 : Entrées les données à partir du questionnaire .....	25
Exercice 1: Entrer les données .....	25
Exercice 2: Fusionner les données et éliminer les erreurs .....	25
A. Fusionner les données .....	25
B. Eliminer les erreurs .....	25
TD3 : Préparation des données : Transformer les données .....	26
Transformer les données .....	26
Recoder des variables .....	26
Construire des indicateurs .....	27
Transformer les données en utilisant la page de syntaxe .....	28

TP3 : Préparation des données : Transformer les données.....	29
Exercice 3: Créer de nouveaux indicateurs .....	29
Exercice 4 : Recoder et créer de nouvelles variables .....	30
Exercice 5 : Ajouter des variables .....	30
TD4 : Analyse des données : Représentations graphiques.....	31
Fréquences.....	31
Graphiques pour les variables nominales et ordinales (montrant les fréquences) .....	32
Bar (bâtons):.....	32
<i>Distribution des fréquences d'une variable</i> .....	32
<i>Distribution des fréquences d'une variable par sous-groupes</i> .....	33
Pie (Camembert) .....	34
Graphiques pour les variables métriques .....	34
Histogramme .....	34
Modifier les caractéristiques d'un graphique.....	35
Exporter des tableaux et des graphiques dans Word.....	36
TP4 : Analyse des données : Représentations graphiques .....	36
Exercice 6 : Distributions des variables .....	36
TD5 : Analyse des données : Mesures descriptives .....	37
Mesures descriptives .....	37
Graphiques des mesures descriptives .....	38
Bar (bâtons).....	38
Error bar .....	41
Boxplot.....	43
Modifier la présentation des résultats (Output Labels) .....	44
Sélectionner des cas .....	45
Comparer les groupes.....	45
TP5 : Analyse des données : Mesures descriptives.....	45
Exercice 7 : Distributions des variables .....	45
Exercice 8 : Sélectionner et comparer les groupes.....	46
TD6 : Analyse des données : Corrélation et régression .....	47
Corrélation de rang.....	47
Scatterplot.....	47
Corrélation de Bravais-Pearson.....	48
Régression linéaire simple .....	49
Syntaxe : .....	49
SPSS-Viewer (Résultat) .....	49
TP6 : Analyse des données : Corrélation et régression.....	51
Exercice 9 corrélation de rang.....	51
Exercice 10 corrélation et régression linéaire simple.....	51

# TD1 : Découverte de SPSS

## Qu'est-ce que SPSS ?

SPSS signifie « Statistical Package for the Social Sciences ». Son objectif est d'offrir un logiciel permettant de réaliser la totalité des analyses statistiques habituellement utilisées en sciences humaines. C'est un logiciel très complet et dans ce cours, nous ne verrons qu'une très faible partie de ses possibilités. Il existe bien d'autres logiciels comme S-Plus, R ou SAS qui permettent d'atteindre les mêmes buts, c'est-à-dire faire des analyses statistiques.

## Comment démarrer SPSS

Allumer l'ordinateur (si cela est déjà trop difficile, veuillez vous référer à votre cours de SFA !). Une fois l'ordinateur sous tension entrer votre nom d'utilisateur et votre mot de passe. Aller sous « Démarrer » (en bas à gauche de l'écran), puis sous « Programme », puis sous « SPSS for Windows » et finalement sous « SPSS 13.0 for Windows » (cf. figure 1).

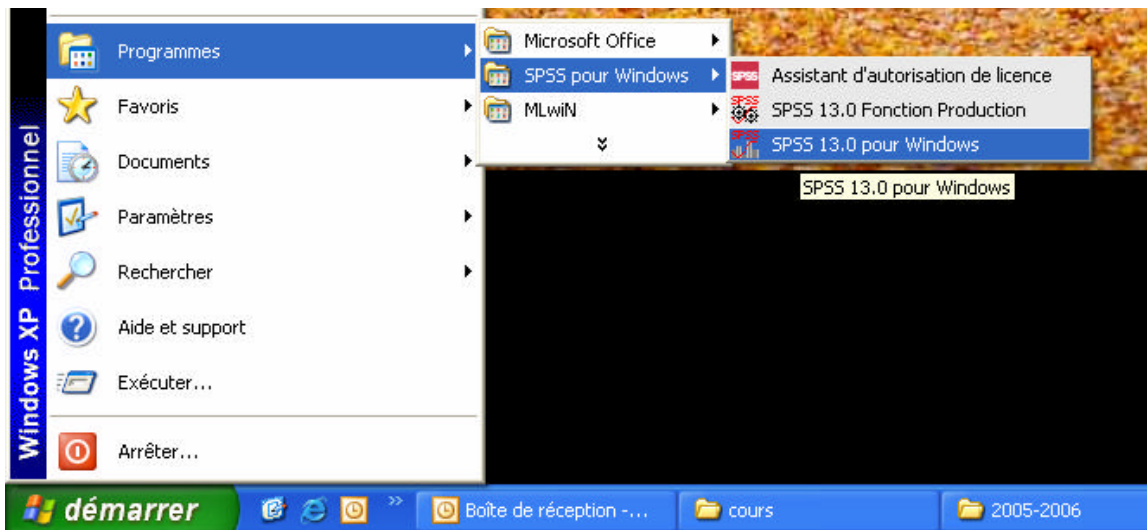


Figure 1

Lorsque SPSS démarre, il ouvre une fenêtre principale qui ressemble un peu à une feuille Excel. Il ouvre devant cette grille, une boîte de dialogue (Figure 2) permettant : d'ouvrir un jeu de données SPSS ou dans un autre format, de créer un nouveau jeu de données, etc. On peut soit utiliser cette boîte de dialogue, soit juste cliquer sur « **Cancel** », car nous pouvons faire exactement les mêmes manipulations directement à l'intérieur de SPSS.

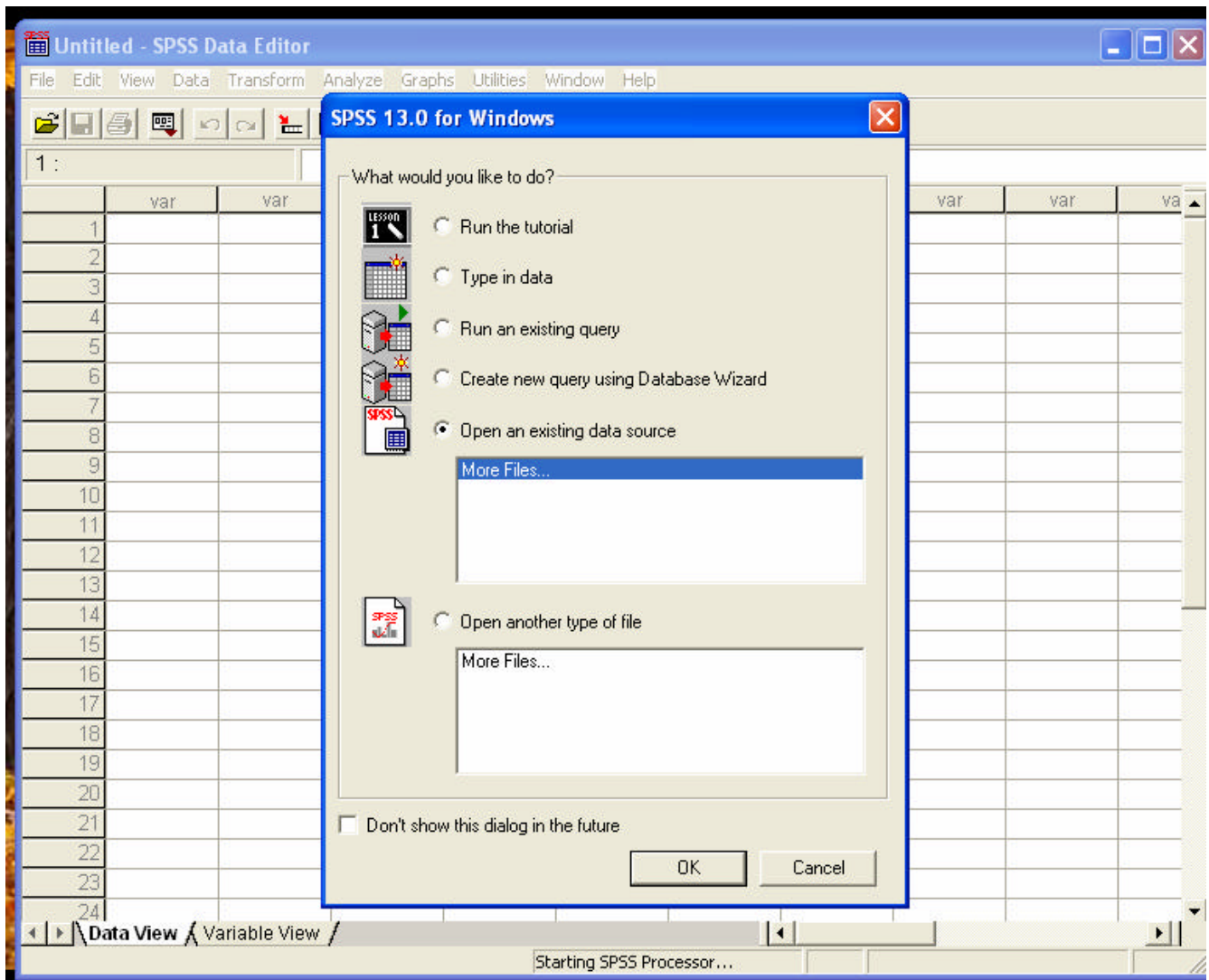


Figure 2

### L'éditeur de données

Une fois cette boîte de dialogue fermée, nous voyons l'éditeur de donnée. L'éditeur de données contient la grille (matrice) de données (vue des données = « **Data View** ») et les descriptions des variables (vue des variables = « **Variable View** »).

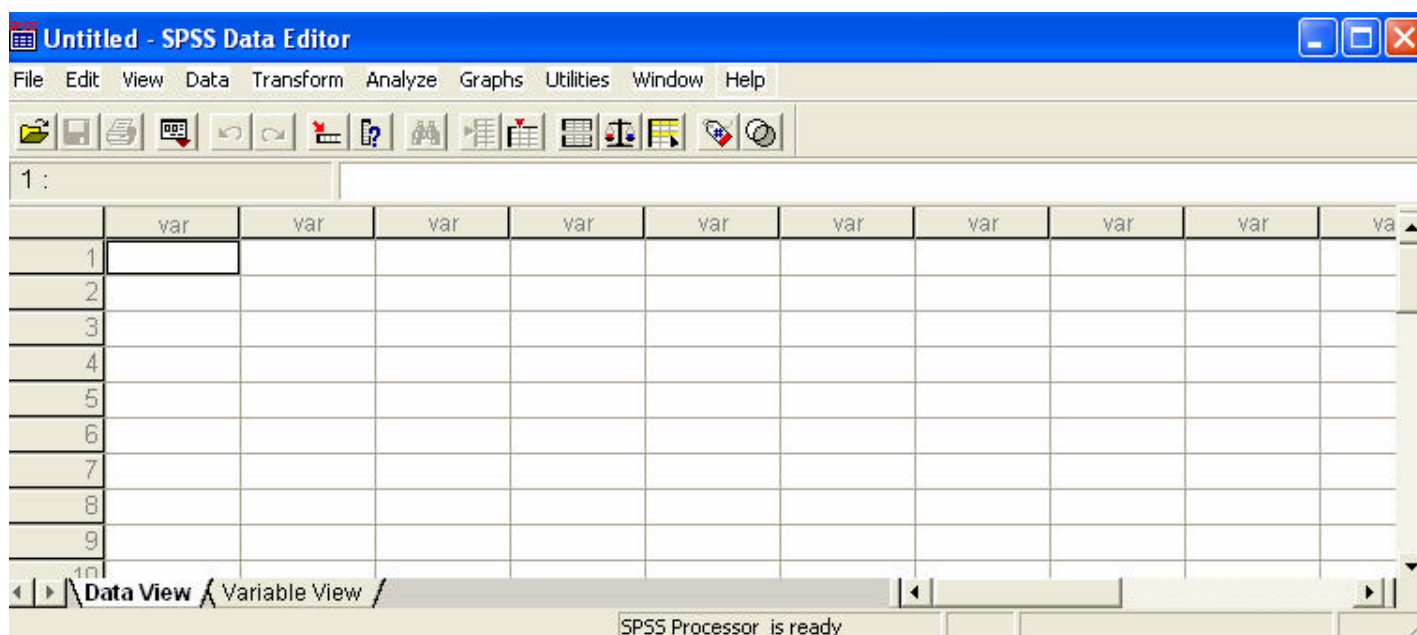


Figure 3

Dans la partie supérieure de l'éditeur de données nous avons, comme dans Word et Excel, des menus déroulants :

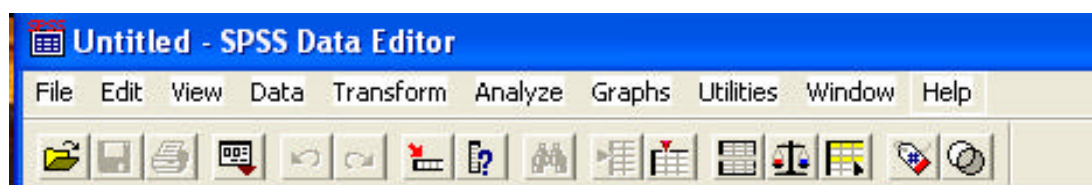


Figure 4

- File:** ... (TP1)
- Edit:** ... (TP1)
- View:** ... (TP1)
- Data:** ... (TP1)
- Transform:** ... (TP3)
- Analyze:** ... (TP5 et TP6)
- Graphs:** ... (TP4)
- Utilities:** ... (TP1)
- Window:** ... (TP1)
- Help:** ... (TP1)

#### Menu File : Ouvrir un fichier de données avec SPSS

Pour ouvrir un fichier de données SPSS (extension .sav), nous allons utiliser les menus déroulant. Cliquer sur « **File** », puis sur « **Open** » et finalement sur « **Data ...** ». Là, vous pouvez chercher votre fichier dans le répertoire où vous l'avez enregistré.

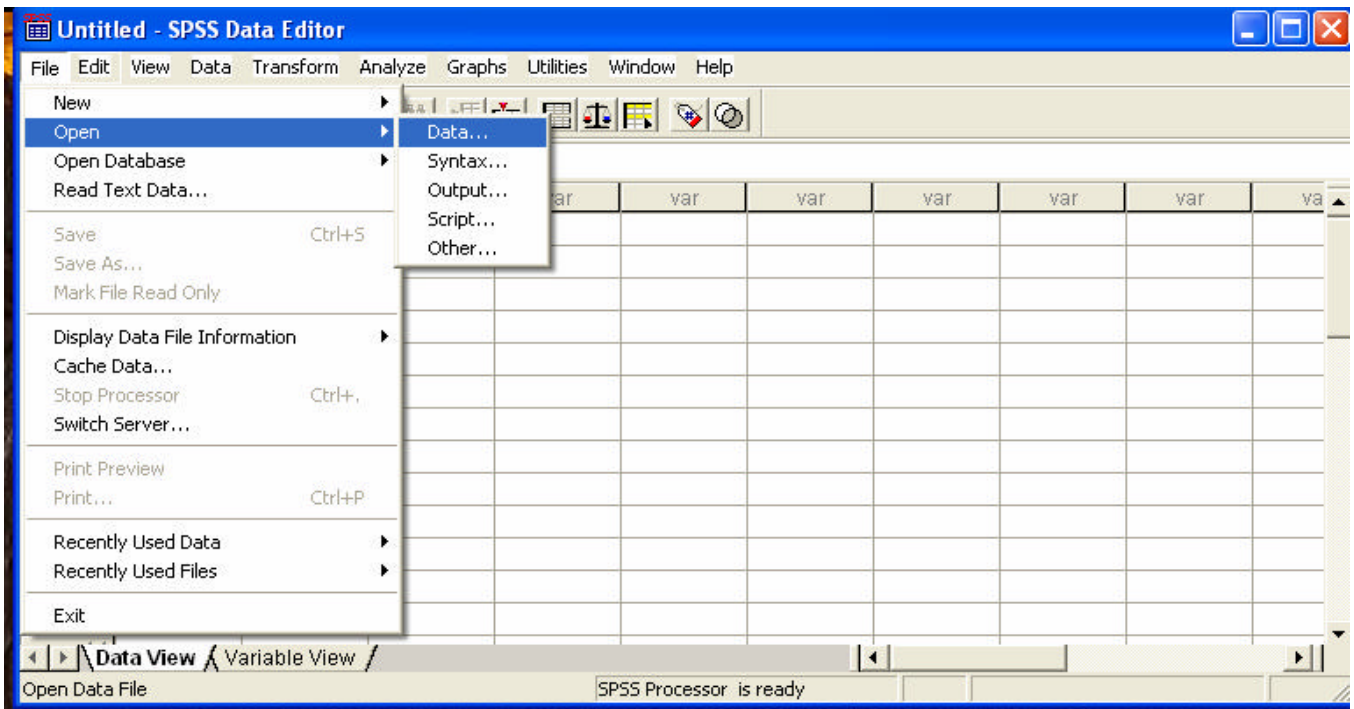


Figure 5

Une fois nos données ouvertes, nous pouvons explorer les deux affichages mentionnés en dessus : l’affichage des données (« **Data View** ») et l’affichage des variables (« **Variable View** ») :

Affichage des données :

- Chaque ligne représente un cas, par exemple un sujet (*case*)
- Chaque colonne représente une variable (*variable*)
- Chaque cellule contient une valeur d’un cas sur une variable

	id	exp01	exp02	exp03	exp04	exp05	exp06	exp07	exp08	exp09
1	1	3	3	1	3	1	1	2	3	
2	2	3	4	2	3	2	2	4	4	
3	3	3	3	2	4	2	2	3	3	
4	4	3	4	2	4	2	4	4	4	
5	5	4	4	3	4	2	4	3	4	
6	6	4	4	1	2	2	2	3	4	
7	7	3	4	2	1	1	2	2	4	
8	8	3	4	2	2	1	2	2	3	
9	9	3	4	2	4	1	2	4	4	
10	10	4	3	2	2	2	3	3	3	
11	11	3	4	2	2	2	3	4	4	
12	12	3	3	3	3	2	3	4	3	
13	13	3	3	2	2	2	2	3	3	

Figure 6

Vue des variables :

- Chaque ligne représente une variable
- Les colonnes décrivent les caractéristiques des variables


	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	3	0	Numéro d'ident	None	None	8	Right	Scale
2	exp01	Numeric	1	0	Affection	{1, jamais ou presque	None	8	Right	Scale
3	exp02	Numeric	1	0	Joie	{1, jamais ou presque	None	8	Right	Scale
4	exp03	Numeric	1	0	Peur	{1, jamais ou presque	None	8	Right	Scale
5	exp04	Numeric	1	0	Colère	{1, jamais ou presque	None	8	Right	Scale
6	exp05	Numeric	1	0	Honte	{1, jamais ou presque	None	8	Right	Scale
7	exp06	Numeric	1	0	Tristesse	{1, jamais ou presque	None	8	Right	Scale
8	exp07	Numeric	1	0	Amour	{1, jamais ou presque	None	8	Right	Scale
9	exp08	Numeric	1	0	Bonheur	{1, jamais ou presque	None	8	Right	Scale
10	exp09	Numeric	1	0	Inquiétude	{1, jamais ou presque	None	8	Right	Scale
11	exp10	Numeric	1	0	Irritation	{1, jamais ou presque	None	8	Right	Scale
12	exp11	Numeric	1	0	Culpabilité	{1, jamais ou presque	None	8	Right	Scale
13	exp12	Numeric	1	0	Depression	{1, jamais ou presque	None	8	Right	Scale
14	bf01	Numeric	1	0	amical(e)	{1, pas du tout}...	None	8	Right	Scale

Figure 7

**Name:** nom de la variable

- le nom de la variable doit être unique
- max 8 caractères (à partir de la version SPSS 12, on peut utiliser plus de caractères : max. 64 !)
- débutant par une lettre
- les autres caractères peuvent être une lettre, un chiffre, un point ou les symbols @ \_ # \$ (tous les autres symbols sont interdits)
- Le nom ne peut pas se terminer par un point
- les espaces vides ne sont pas possibles
- majuscules et minuscules ne sont pas différenciés dans l'appellation d'une variable ou dans celui d'un fichier
- les mots clé de SPSS ne peuvent pas être utilisés (all, and, by, eq, ge, gt, le, lt, ne, not, or, to, with)
- Ne pas mettre de lettre accentuée même si c'est possible

**Type :** nature de la variable (numérique, date, dollar, etc.)

- par défaut, SPSS considérera la variable de type numérique;
- peut être changé dans la boîte de dialogue qui s'ouvre en appuyant sur  (symbole qui apparaît quand on sélectionne la cellule)

**Width :** nombre de chiffres accordés à la donnée (décimaux inclus)

**Decimals :** nombre de décimaux




- Attention : **Width** et **Decimals** définissent le format des données dans la grille de données – la donnée apparaissant dans une cellule peut différer de la donnée exacte (qui est utilisé par SPSS pour les opérations de calcul et qui est visible dans la cellule blanche au-dessous la barre d’outils pour la cellule « active »)

**Label** : étiquette ou description de la variable

- max. 256 caractères

**Values** : valeurs définies et leur description,

- p.ex. 1 = Femme, 2 = Homme
- peuvent être définis dans la boîte de dialogue qui s’ouvre en cliquant sur  - ce qui apparaît quand on choisit la cellule

**Missing**: attribution de certaines valeurs comme codes pour valeurs manquantes

- Attention : « **none** » ne signifie pas qu’il n’y a pas de valeurs manquantes dans les données, ça signifie qu’il n’y a pas de valeurs manquantes définis par l’utilisateur. Par contre, il est possible d’avoir des valeurs **system missing** (correspond aux cellules vides dans la grille de données).

**Columns**: largeur des colonnes dans la vue de données (en augmentant/diminuant le défaut « 8 », plus/moins de caractères de la colonne seront visibles dans la vue de données)

**Align** : Alignement des valeurs des variables dans les cellules de la grille de données (à droite, à gauche, centrées)

**Measure** : Description de l’échelle de mesure (continue (**scale**), **ordinal** ou **nominal**)

- Attention : n’a pas d’influence sur le type d’analyse qu’on peut effectuer avec SPSS !!!

**Menu File** : Obtenir les informations sur les variables dans un format imprimable

Pour obtenir toutes les informations sur les variables dans un format imprimable, il faut cliquer sur **File** | **Display data file information** | **Working file**. Les informations sont transférées dans une nouvelle fenêtre (**SPSS Viewer**). Cette page des résultats peut être sauvegardée et/ou imprimée. Les éléments peuvent être copiés et collés dans un document géré par un autre logiciel (p.ex. Word). Le fichier des résultats (**Output Editor**) porte l’extension « .spo ».

**Menu File** : Ouvrir et sauvegarder des fichiers de données

Dans le menu « **File** », on trouve toutes les commandes pour gérer les fichiers SPSS (ouvrir, sauvegarder, sauvegarder sous...) – pareil aux autres logiciels Windows (p.ex. Word)

**Informations compactes sur les variables dans un fichier de données**

Menu Utilities: Dans le menu **Utilities** | **Variables** on peut consulter les caractéristiques de toutes les variables. Quand vous travaillez avec un fichier avec beaucoup de variables, cette

option a l'avantage de trouver des variables en tapant les premières lettres du nom de la variable. En appuyant sur le bouton « **Go to** » on arrive à la variable cherchée dans la vue de données.

Menu Edit : en général, il est plus facile de trouver des variables dans les listes des boîtes-dialogue quand elles sont affichées par leur nom (**variable name**) et pas par leur étiquette (**variable label**). En plus, il est recommandé de choisir le même ordre que dans le fichier (au lieu d'un ordre alphabétique). Pour cela, il faut aller sous **Edit | Options: General, Variable list : Display names** (au lieu de **labels**), **File** (au lieu de **alphabetical**).

Menu View : pour passer de la vue des données à la vue des variables et vice versa, il faut cliquer sur **Variable View** (respectivement « **Data View** ») dans le coin inférieur gauche de la grille des données ou via le menu **View | Variable** (respectivement **Data**)

## Les différentes fenêtres de SPSS

### Fenêtre de données (Data Editor)

Cette fenêtre permet d'entrer des données, de les modifier ou de les effacer. Il est rare que l'on tape les données manuellement dans SPSS car il y a trop d'erreurs de saisie possibles (on reviendra sur ce point dans le TP2). On va plutôt ouvrir un fichier déjà existant.

### Fenêtre des résultats (Output Editor)

Cette fenêtre apparaît après qu'une commande d'analyse a été effectuée, et contient les résultats de cette analyse. Les résultats apparaissent à droite dans la fenêtre. A gauche, figure une table des matières des résultats générés par SPSS.

Les résultats peuvent être imprimés tels quels (mais avec le risque de voir un tableau s'imprimer sur plusieurs pages. Il est également possible de copier les tableaux qui nous intéressent pour les coller ensuite dans Word, Excel ou dans un autre logiciel. Il est possible de copier un tableau de deux manières. En cliquant sur le tableau en appuyant sur le bouton de droite de la souris, SPSS vous propose de copier (**copy**) ou de copier l'objet (**copy object**). Copier correspond à copier les valeurs, mais lorsqu'il est collé il peut perdre son format (utile pour copier les résultats dans une feuille Excel par exemple). Copier les objets correspond à copier les valeurs et le format du tableau : une fois collé, impossible de modifier les cellules du tableau (utile pour copier les résultats dans Word).

Parfois, le résultat est très long et SPSS n'en montre qu'une partie, suivi d'un triangle rouge. Pour visualiser la suite, il faut double-cliquer sur le texte, puis étendre la taille de la zone de texte vers le bas jusqu'à ce que vous ne voyez plus de texte.

### Fenêtre de syntaxe (Syntax Editor)

Jusqu'à maintenant, nous avons vu comment travailler avec les menus déroulant. Il existe une autre manière de lancer des analyse : passer par la fenêtre de syntaxe. Cette fenêtre permet d'écrire les commandes d'analyses statistiques. Elle fonctionne comme un traitement de texte simple.

Mais pourquoi s'amuser à taper les analyses quand il suffit de faire défiler des menus et cliquer sur des icônes ? Parce que lorsqu'on veut refaire une analyse qu'on a déjà effectuée et juste changer un paramètre, il suffit de reprendre la feuille de syntaxe de l'analyse précédente, changer un mot et renvoyer l'analyse. Sinon, il faudrait recommencer toute l'analyse de A à Z, vu qu'on ne peut pas enregistrer la séquence de clics. De plus, les menus ne donnent pas accès à toutes les commandes SPSS. Finalement, c'est aussi utile pour faire la même analyse sur plusieurs fichiers de données.

Ainsi, une fois la syntaxe faite pour une opération, il est facile d'enregistrer les commandes et de les réutiliser pour différents fichiers de données.

Voici les règles générales pour écrire des commandes dans SPSS :

- Chaque nouvelle commande se trouve en tête de ligne, précédée d'aucun espace.
- Les options qui suivent une commande débutent sur la ligne suivante et sont précédées d'au moins un espace et d'une barre oblique (/).
- Chaque commande doit ABSOLUMENT se terminer par un point.
- Lorsqu'on spécifie un nom de fichier, il doit être "entre guillemets".
- SPSS ne fait pas de différence entre les lettres majuscules et minuscules. Vous pouvez taper les commandes autant d'une manière ou de l'autre.
- De plus, entre les commandes, vous pouvez insérer des lignes vides. SPSS les ignore, mais elles peuvent améliorer la lisibilité des commandes quand il y en a plusieurs dans une fenêtre.

Pour exécuter une commande, il faut sélectionner la commande à exécuter puis choisir dans le menu « **Run : Current** » ou encore Ctrl-R ou encore utiliser le bouton 'run' (? ).

Vous pouvez enregistrer votre script pour le modifier plus tard (menu **File | Save**). Vous pouvez aussi enregistrer la fenêtre de résultats bien que très souvent, il en résulte un fichier énorme (que vous ne pourrez sauvegarder sur une disquette). Cependant, tant et aussi longtemps que vous avez votre syntaxe et vos données, vous pouvez toujours exécuter la syntaxe à nouveau pour revoir les résultats.

### **Menu Graphs : refaire le graphe du cours**

Le menu graphique permet de générer en quelques cliques de souris toute une panoplie de graphiques. Nous reviendrons plus en détail lors du TD 4.

Lors du cours, on vous a présenté les résultats d'une recherche mettant en avant (entre autre) le lien entre la satisfaction des relations avec sa mère et la fréquence des relations sexuelles pour des adolescents noirs de quartiers défavorisés de Philadelphie âgés de 14 à 17 ans.

Le menu **Graphs** permet en quelque clique de souris de reproduire le graphique présenté en cours pour cette expérience. Pour cela, allez chercher sur dokeos la base de données Philadelphie.sav. Enregistrez là sur votre répertoire h:\.... Puis ouvrez SPSS par Démarrer | Programme | SPSS for Windows | SPSS 13.0 for Windows. Allez sous **File | Open | Data...** Cherchez la base de données Philadelphie.sav qui se trouve dans votre répertoire h. Une fois la base de données ouvertes, allez sous **Graphs**.

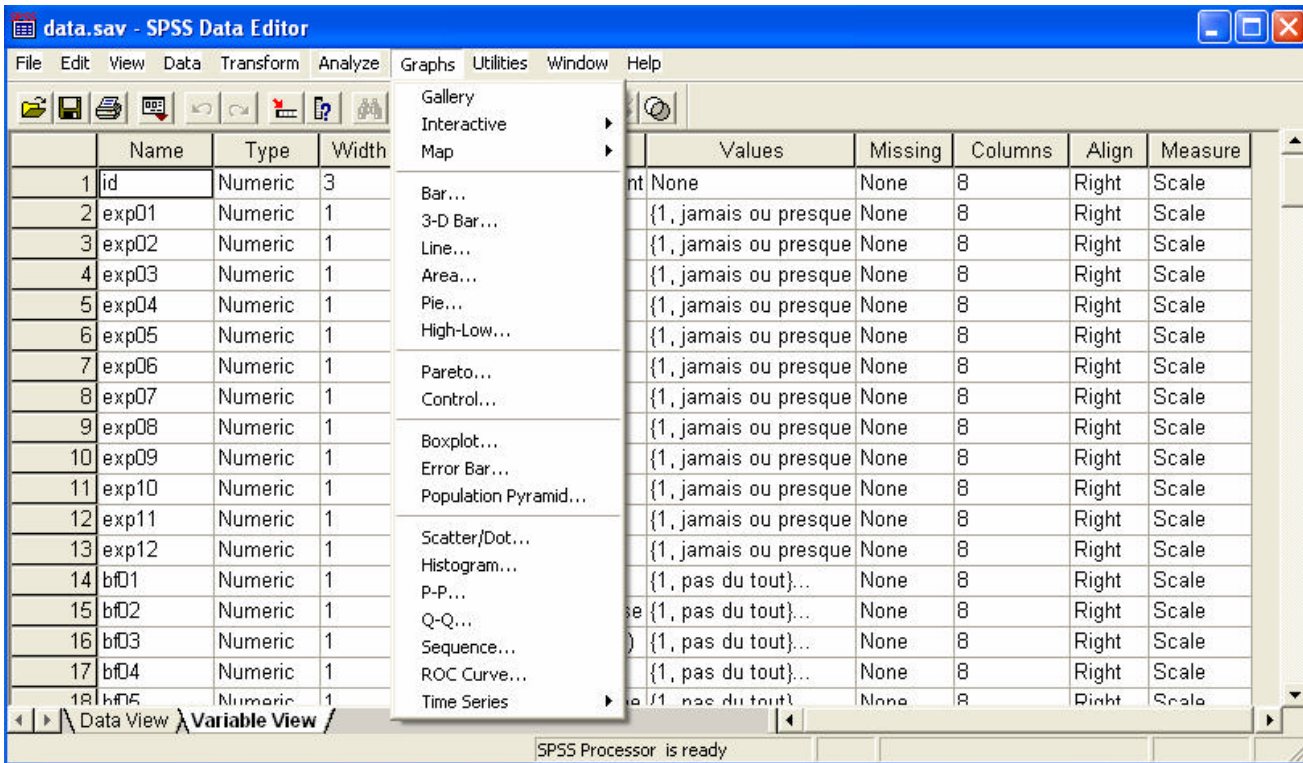


Figure 8

Sélectionnez **Scatter/Dot... | Simple Scatter**. Dans **Y Axis** introduisez la variable de **sexfreq** (fréquence des relations sexuelles) et dans **X Axis**, **satisf** (la variable de la satisfaction de la relation avec la mère). Cliquez sur Ok et vous obtenez le graphique suivant :

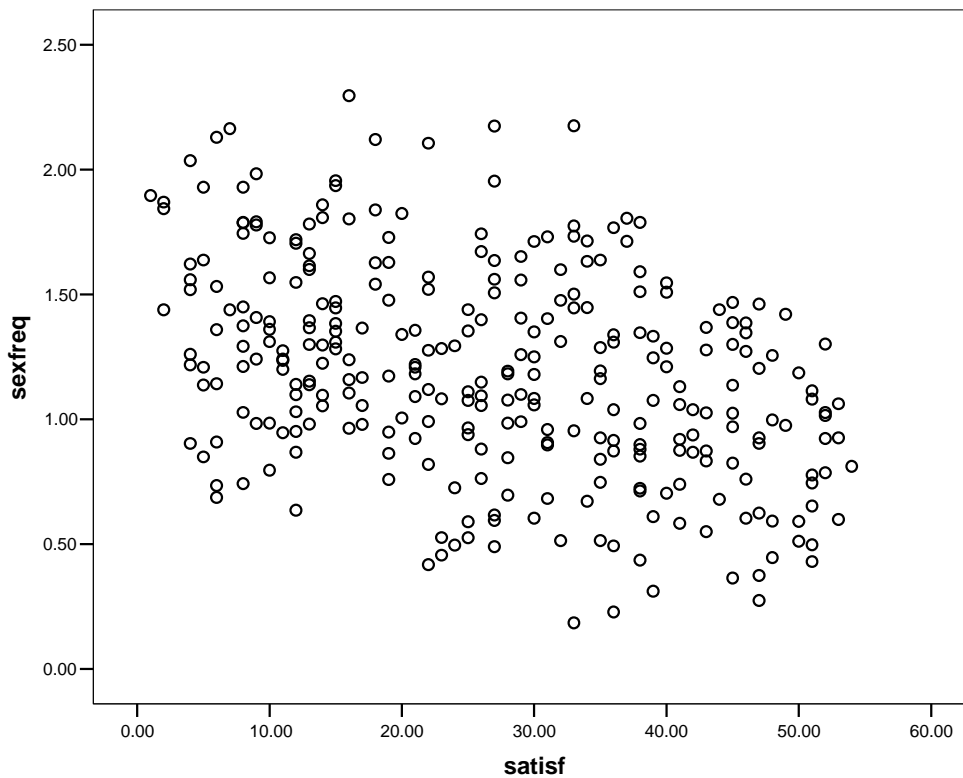


Figure 9

## Menu Help

SPSS offre plusieurs niveaux d'aide (extrait tiré de l'aide de SPSS pour Windows 13.0 en français):

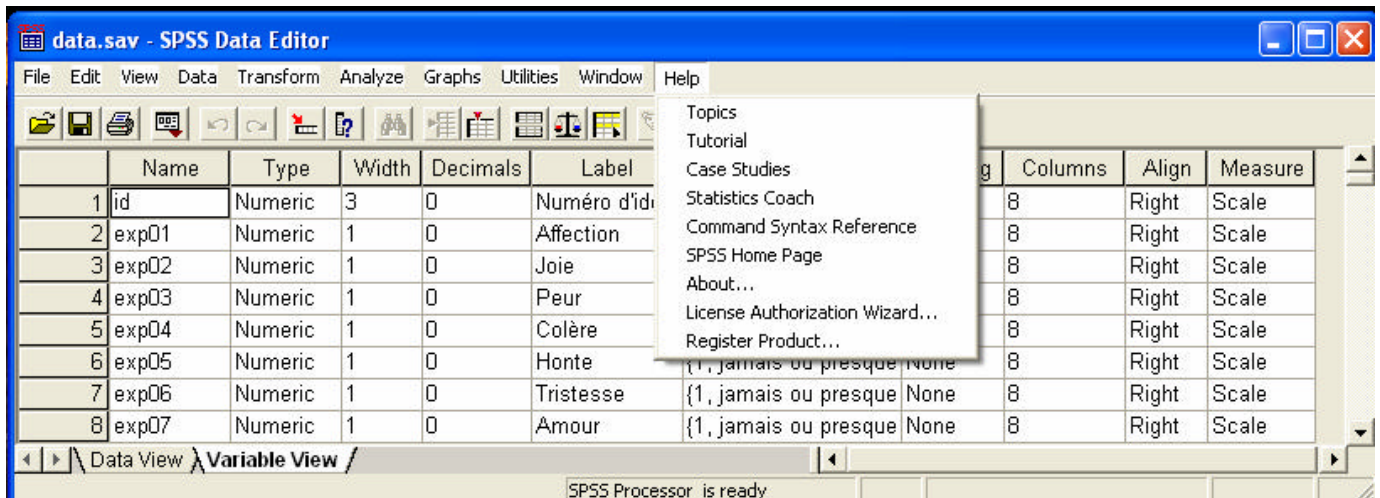


Figure 10

« L'aide apparaît sous plusieurs formes :

Menu Aide : Le menu Aide de la plupart des fenêtres SPSS permet d'accéder au système d'aide principal, ainsi qu'aux didacticiels et aux informations de référence technique.

- Rubriques (Topics) : Les rubriques permettent d'accéder aux **onglets Sommaire, Index et Rechercher**, que vous pouvez utiliser pour chercher des rubriques d'aide particulières.
- Didacticiel (Tutorial) : Instructions illustrées et détaillées étape par étape vous expliquant comment utiliser de nombreuses fonctions de base de SPSS. Il n'est pas nécessaire de visualiser le didacticiel du début à la fin. Vous pouvez choisir les rubriques à visualiser, ignorer et visualiser des rubriques dans l'ordre de votre choix, et utiliser l'index ou le sommaire pour rechercher des rubriques données.
- Etudes de cas (Case Studies) : Exemples pratiques indiquant comment créer différents types d'analyse statistique et comment interpréter les résultats. Les fichiers de données d'exemple utilisés dans ces cas pratiques vous sont également fournis afin que vous puissiez voir exactement comment les résultats ont été générés. Vous pouvez choisir à partir du sommaire les procédures sur lesquelles vous souhaitez obtenir des informations ou rechercher des rubriques appropriées dans l'index.
- Assistant statistique (Statistics Coach) : Vous aide à rechercher la procédure que vous souhaitez utiliser. Une fois vos sélections effectuées, l'Assistant statistique ouvre la boîte de dialogue de la procédure statistique, de rapport ou de diagramme qui correspond aux critères sélectionnés. L'Assistant statistique permet d'accéder à la plupart des procédures statistiques et de rapport du système de base, et à de nombreuses procédures de diagramme.
- Command Syntax Reference : Vous trouverez des informations détaillées de référence de syntaxe des commandes dans le manuel SPSS **Command Syntax Reference**, accessible via le menu Aide.

Aide sensible au contexte : Plusieurs emplacements de l'interface utilisateur vous permettent d'accéder à l'aide contextuelle.

- Boutons Aide de boîte de dialogue : La plupart des boîtes de dialogue disposent d'un bouton **Help** qui vous conduit directement à la rubrique d'aide relative à la boîte de dialogue. La rubrique d'aide fournit des informations générales et propose des liens vers les rubriques apparentées.
- Aide du menu contextuel de la boîte de dialogue : De nombreuses boîtes de dialogue fournissent une aide contextuelle pour les commandes et les fonctions. Cliquez avec le bouton droit sur n'importe quel contrôle d'une boîte de dialogue et sélectionnez **What's this ?** dans le menu contextuel. Une description du contrôle s'affiche alors, ainsi que son mode d'utilisation. (Si **What's this ?** n'apparaît pas dans le menu contextuel, ce type d'aide n'est pas disponible pour cette boîte de dialogue.)
- Aide du menu contextuel du tableau pivotant : Cliquez sur le bouton droit de la souris sur les termes du tableau pivotant dans le Viewer et sélectionnez **What's this ?** dans le menu contextuel afin d'afficher les définitions de ces termes.
- Etudes de cas : Cliquez avec le bouton droit de la souris sur un tableau pivotant, puis sélectionnez **Case Studies** dans le menu contextuel pour accéder directement à un exemple détaillé relatif à la procédure qui a généré ce tableau. (Si **Case Studies** n'apparaît pas dans le menu contextuel, ce type d'aide n'est pas disponible pour cette procédure.)
- Tableaux de syntaxe de commande : Dans une fenêtre de syntaxe de commande, positionnez le curseur n'importe où dans un bloc de syntaxe de commande, puis appuyez sur la touche F1. Vous voyez alors apparaître un tableau de syntaxe de commande complet. » (l'Aide français de SPSS 13.0 pour Windows)

## TD2 : Entrer les données à partir du questionnaire

Il y a plusieurs manières d'entrer les données :

- directement dans SPSS
- dans Excel ; puis nous importons les données dans SPSS
- dans un éditeur de texte, puis nous importons les données dans SPSS (pas recommandé, sujet à erreur!)
- scannage des données : pour cela il est nécessaire d'avoir un hardware avec logiciel et mise en page spécialisés

### Entrer les données directement dans SPSS

#### Encoder le questionnaire

Il est recommandé de résumer les informations les plus importantes sur les variables rassemblées dans un « tableau de codage ». Ce tableau de codage à deux utilités à deux moments bien précis :

- Pendant l'entrée des données : comme règle de codage des valeurs des variables
- Après l'entrée des données : comme description compacte du fichier des données

Le tableau de codage contient les informations suivantes:

- Nom de la variable : les items qui appartiennent au même questionnaire devraient porter le même radical dans leur nom (p.ex., exp1, exp2, exp3... pour un questionnaire mesurant l'expressivité des émotions)
- Etiquette de la variable (variable label)
- Etiquettes des valeurs (value labels) : N'oubliez pas une variable d'identification qui établit une relation entre les documents d'un cas (p.ex. questionnaire, prélèvements physiologiques) et les données dans le fichier.
- Numéro d'identification (ID variable): doit être noté sur les questionnaires, pour que l'on puisse facilement retrouver le document d'un sujet afin de contrôler ou corriger des valeurs dans la base de données
  - ➔ Dans le cas d'une recherche expérimentale, il faut en plus coder la condition dans laquelle on a mesuré les variables dépendantes: variable pour les conditions expérimentales (p. ex. 1 = groupe expérimental, 2 = groupe de contrôle)

*Tableau de codage pour notre questionnaire*

Variable	Variable label	Value labels
id	Numéro d'identification	
exp1 – exp12	Expressivité *	1 = jamais ou presque jamais 2 = parfois 3 = souvent 4 = presque toujours ou toujours
bf1 – bf20	Big Five (adjectives) *	1 = pas du tout 5 = tout à fait
ouv1-ouv30	Ouverture *	1 = pas du tout d'accord 5 = tout à fait d'accord
loisirs1 – loisirs8	Loisirs *	0 = n'a pas ce loisir 1 = a ce loisir
sexe	Sexe	1 = Femme 2 = Homme
age	Age	
natio	Nationalité	1 allemande                      19 roumaine 2 coréenne de sud 20 néerlandaise 3 mexicaine                      21 indienne 4 slovaque                        22 grecque 5 italienne                        23 israélienne 6 taiwanaise                      24 belge 7 finnoise                         25 chinoise 8 russe                              26 marocaine 9 française                        27 autrichienne 10 polonaise                      28 espagnole 11 tchèque                        29 bulgare 12 ukrainienne                    30 bolivienne 13 moldavienne                   31 lituanienne 14 usbekienne                    32 palestinienne 15 turque                         33 islandaise 16 Etats-Unis                     34 britannique 17 luxembourgeoise              35 croate 18 hongroise                      36 suisse
etude	Domaine d'études	1 mathématique                10 économie 2 psychologie                  11 histoire 3 droit                            12 architecture 4 langues                         13 design 5 géologie                        14 informatique 6 éducation                      15 médias 7 ingénierie                      16 théologie 8 politique                       17 philosophie 9 sociologie                      18 ethnologie
code	Code	

\* Pour une vue générale sur les variables qui sont mesurées dans une étude, il suffit de donner le nom du questionnaire dans le tableau de codage (p.e.x, « Big Five » pour les 20 items). Dans le fichier SPSS, on peut entrer le contenu de chaque item dans « **Variable label** » (p.ex., bf1: « amical », bf2: « chaleureux »).

## **Créer un nouveau fichier de données dans SPSS**

- Lorsqu'on démarre SPSS, une fenêtre « **What would you like to do ?** » apparaît par défaut (cf. TD1). On sélectionne « **Type in data** » et on obtient un éditeur de données vide.



- Si on se trouve déjà dans l'éditeur des données (p. ex., un autre fichier de données est ouvert), il faut cliquer sur **File | New | Data**
- Une fois l'éditeur de données ouvert, il faut définir les variables dans la vue des variables (**Variable View**). Pour cela on va s'aider du tableau de codage qu'on a déjà créé à partir de notre questionnaire.
- On commence avec le nom de la première variable. Pour passer à la cellule suivante, appuyez sur TAB ou →.
- Vous pouvez attribuer une ou toutes les caractéristiques d'une variable à une ou plusieurs autres variables !
  - ➔ Une caractéristique : Copiez la cellule (en cliquant droit : **Copy**, ou en cliquant sur le menu **Edit | Copy**) et collez à une ou plusieurs autres cellules (en cliquant droit : **Past**, ou en cliquant sur le menu **Edit | Past**)
  - ➔ Toutes les caractéristiques d'une variable : Copiez et collez toute la variable (en cliquant sur le numéro de la ligne, ce qui la grise)
- La seule chose qui ne peut pas être copiée, c'est le nom d'une variable, parce que chaque variable doit avoir un nom unique, comme mentionné précédemment. C'est pour ça que SPSS nomme une variable nouvelles var00001 (que l'on renomme après)
  - Pour créer plusieurs nouvelles variables avec le même radical dans leur nom (par exemple bf1, bf2, ..., bf20), il faut
  - Entrez la variable bf1 avec son type, son étiquette, etc. (vue : **Data View**)
  - Copiez cette variable.
  - Sélectionnez la ligne du dessous de la variable et cliquez sur le bouton de droite de la souris.
  - Sélectionnez « **Copy variables** ».
  - Dans la boîte de dialogue qui apparaît entrez le nombre de nouvelles variables à créer ici 19 (20 variables big five moins une déjà entrée), leur radical (nom des nouvelles variables : bf) et le numéro de la première variable qui suivra le radical (2, car on a déjà créer bf1).

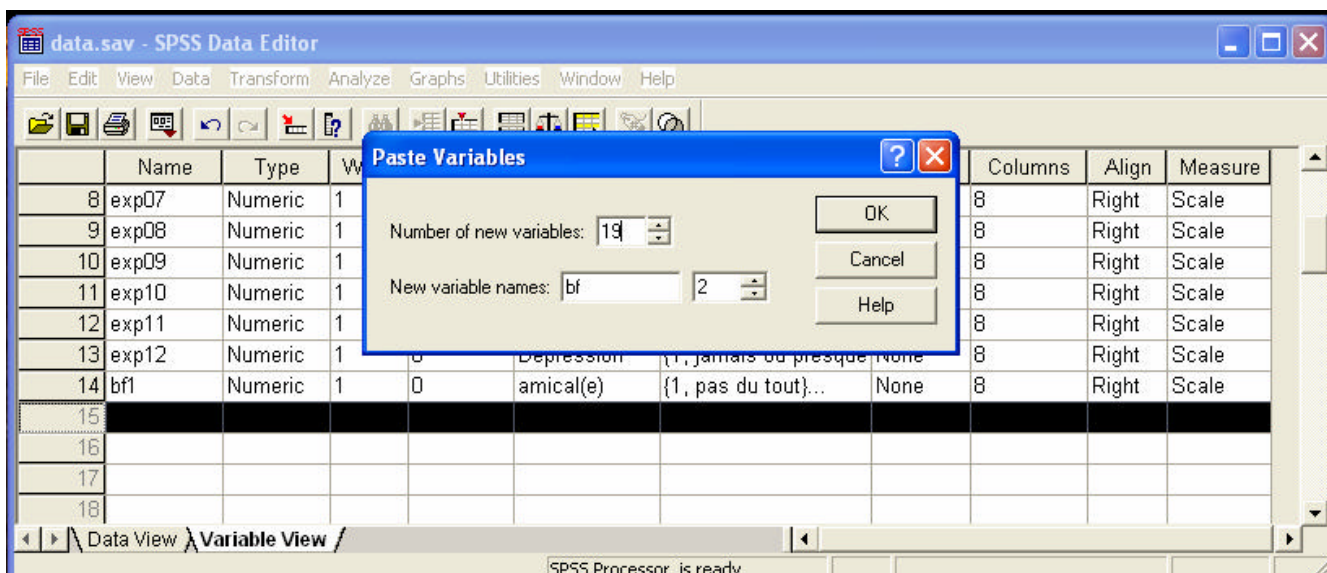


Figure 11

## **Comment coder les variables alphanumériques/série de caractères ?**

Exemple : Code d'un sujet (MAPRGE) dans des études longitudinales

Solution : entrez les caractères et définissez la variable comme chaîne de caractère (**String**)

## **Comment coder les réponses à réponses courtes ?**

Exemple :

Quelle est votre nationalité ? \_\_\_\_\_

Solution : Codez les réponses ouvertes avec des valeurs numérique (1 = suisse, 2 = française, etc.) en faisant une liste

## **Comment coder les réponses multiples ?**

Exemple :

Quelles occupations/loisirs avez-vous régulièrement? (*plusieurs réponses sont possibles*)

Activités culturelles (expositions, concerts, lecture, université 3<sup>e</sup> âge, etc.)

Activités physiques (natation, marche, club de sport, etc.)

Bricolage (tricot, jardinage, etc.)

Animal domestique

Instrument de musique

Internet, e-mail

Autre : \_\_\_\_\_

Solution :

- Créez une variable pour chaque catégorie (p.ex, loisir1 – loisir6), codé par 0 = n'a pas ce loisir, 1 = a ce loisir
- Pour les réponses ouvertes (« autre : »), créez soit une variable alphanumérique (Chaîne de caractère), p.ex. add\_lois, soit une variable numérique en faisant une liste des loisirs qui apparaissent

## **Comment coder les réponses ouvertes ?**

Exemple :

**Qu'est ce que ces loisirs vous apportent ?**

---

---

Solution :

- regrouper l'information en catégorie grâce à l'analyse de contenu. Technique que nous ne verrons pas dans ce cours.
- Donc, pour ce cours, nous laisserons de côté cette question du questionnaire.

## Comment coder les valeurs manquantes ?

- Dès qu'on a entré une donnée, toutes les cellules des autres variables numériques de ce cas sont désignées par un point (= « **Sysmis** », system defined missing) qui est remplacé quand on entre une valeur. Garder le point dans la cellule signifie que la valeur pour cette cellule est manquante et ce type de donnée manquante n'a pas à être définie comme telle car elle est reconnue automatiquement par SPSS comme valeur manquante.
- Garder la cellule vierge pour les variables alphanumériques n'est pas considéré par SPSS comme valeur manquante. Il faut la définir comme telle dans la vue des variables. Pour cela, entrez un espace dans « **discret missing values** ».
- Entrer un chiffre en dehors de l'étendue de valeurs valables (p. ex. 9 ou 99). Il faut définir ces valeurs dans la vue des variables. Pour cela, entrez un espace dans « **discret missing values** ».

## Conseil et remarques

- Il est recommandé d'utiliser le pavé numérique à droite du clavier.
- Quand on utilise la touche „Entrée“ pour confirmer une donnée, la cellule au-dessous devient active (= prochain cas)!

## Réduire les erreurs en entrant les données

- Entrez toujours les données « brutes » telles qu'elles sont. Par exemple, il ne faut pas recoder à la main un item (recodage ou transformation (cf TD3)).
- Si vous avez un format de réponse bipolaire avec des valeurs positives et négatives, il est préférable d'utiliser un codage avec uniquement des valeurs positives. Mais il est important d'avoir défini cela dans le tableau de codage du questionnaire.

<i>Format de réponse:</i>	--	-	0	+	++
<i>sujet à erreur:</i>	-2	-1	0	1	2
<i>mieux:</i>	1	2	3	4	5

## Éliminer et insérer des observations/variables dans la vue de données

- Éliminer un cas : sélectionnez la ligne et appuyez sur « **Delete** » sur le clavier
- Éliminer une variable : sélectionnez la colonne et appuyez sur « **Delete** » sur le clavier
- Insérer un cas entre deux autres cas : sélectionnez la ligne au-dessus de laquelle vous voulez insérer une observation et cliquez sur **Data | Insert cases** ou cliquez droit : **Inserte cases**.
- Insérer une variable entre deux autres variables : sélectionnez la colonne avant celle où vous voulez insérer une variable et cliquez sur **Data | Insert variables** ou cliquez droit : **Inserte Variables**.

## Fusionner des fichiers de données

### *Ajouter des observations*

On a deux fichiers contenant des variables similaires, mais des observations différentes (par exemple, saisie d'une moitié des questionnaires par Paul et saisie de l'autre moitié par Pierre avec pour but d'avoir un et un seul fichier à la fin). Pour cela il faut ouvrir le premier fichier, c'est-à-dire celui qui sera pour nous notre fichier de travail. A partir du menu, faites **Data | Merge files | Add cases** : cherchez votre deuxième fichier.

La boîte de dialogue qui apparaît vérifie si les deux fichiers contiennent les mêmes variables (avec les mêmes noms<sup>1</sup>). Par défaut, toutes les variables de même nom seront incluses. Si ce n'est pas le cas, les variables non appariées se trouvent dans la section « **Unpaired Variables** ».

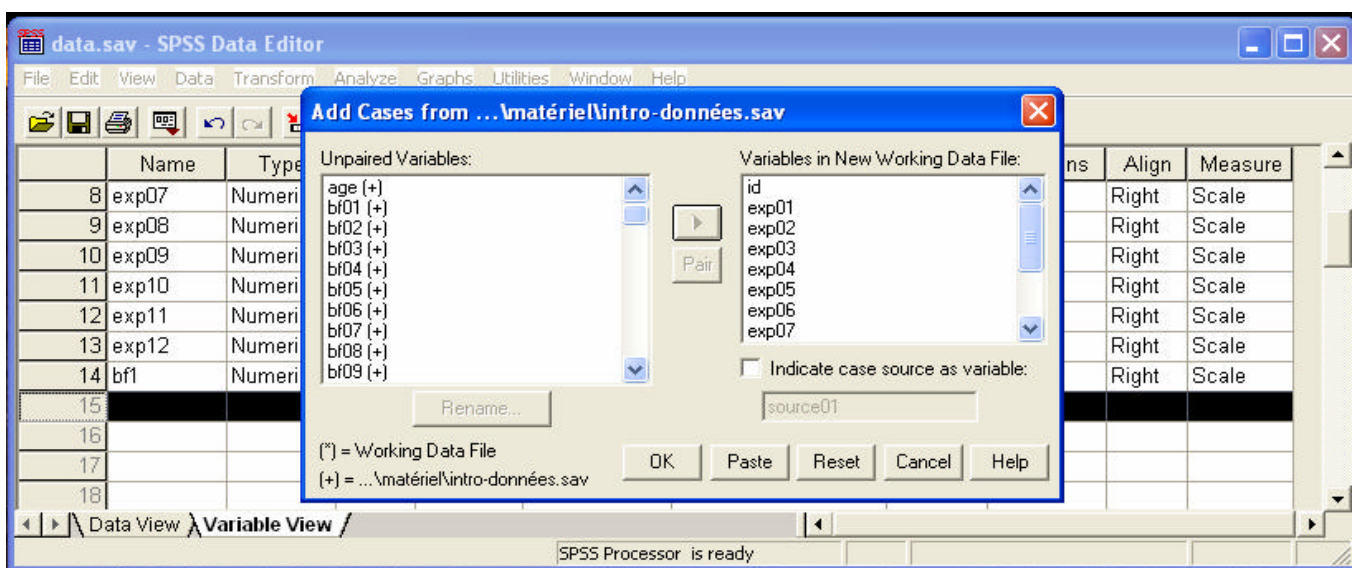


Figure 12

Les variables qui apparaissent seulement dans un fichier de données peuvent quand même être ajoutées en appuyant sur ► (les cas de l'autre fichier qui n'ont pas de valeurs à ces variables reçoivent des valeurs manquantes = **Sysemis**)

S'il y a des variables dans les deux fichiers qui mesurent la même chose mais qui ne portent pas le même nom (p.ex. par erreur), on peut les appairer. Pour faire cela, il faut sélectionner les deux variables (on sélectionne la deuxième variable en pressant sur la touche CTRL), puis appuyer sur « **Paired** », on obtient ainsi dans « **Variables in new working data file** », une nouvelle variable qui se nomme var1 & var2 (dans le fichier fusionné, la variable portera le nom du premier fichier (le fichier de travail)).

<sup>1</sup> Attention, il est important que le nom des variables, ainsi que le type, l'étiquette et les valeurs et les données manquantes soient définis exactement de la même manière dans les deux fichiers, sinon SPSS définira une variable avec même nom comme « Unpaired variables » (fenêtre de gauche) → toujours vérifier que les variables communes aux deux fichiers qui nous intéressent se trouvent bien dans « Variables in new working data file ».

Une fois toutes les variables qui nous intéressent sélectionnées, il faut cliquer sur ok, ce qui ajoute les observations du deuxième fichier au premier. On a maintenant un nouveau fichier de données. Si nous sauvons ce fichier en faisant **File | Save**, cela va écraser le premier fichier. Ce qu'il faut faire c'est **File | Save as** et on obtient ainsi une nouvelle base de données, différentes des deux initialement utilisées pour sa création.

### Ajouter des variables

Si on a deux fichiers contenant les mêmes sujets, mais avec des variables différentes (par exemple, deux expériences différentes avec le même groupe de sujets avec pour but d'avoir des informations supplémentaires sur nos sujets venant de l'autre expérience). Pour cela il faut ouvrir le premier fichier, c'est-à-dire celui qui sera pour nous notre fichier de travail. A partir du menu, faites **Data | Merge files | Add variables** : cherchez votre deuxième fichier.

Pour pouvoir effectuer cette manipulation, il faut impérativement avoir une variable « clé » pour identifier les différentes observations (p. ex. par un code) et pouvoir fusionner les données. Les valeurs de cette variable clé doivent être identiques (p.ex. une variable alphanumérique doit être entrée de la même façon – par rapport aux majuscules/minuscules et nombre de caractères – attention aux espaces après les derniers caractères !)

- Triez les observations selon la variable clé dans tous les fichiers qu'on veut fusionner : **Data | Sort cases**.
- Ouvrez le premier fichier et faites **Data | Merge files | Add variables**: choisissez le deuxième fichier
- Dans la boîte de dialogue qui apparaît, cochez « **Match cases on key variables in sorted files** » - « Les deux fichiers fournissent des observations ». Il faut sélectionner la variable clé (**key variable**) d'appariement dans la section de gauche et appuyer sur ►.

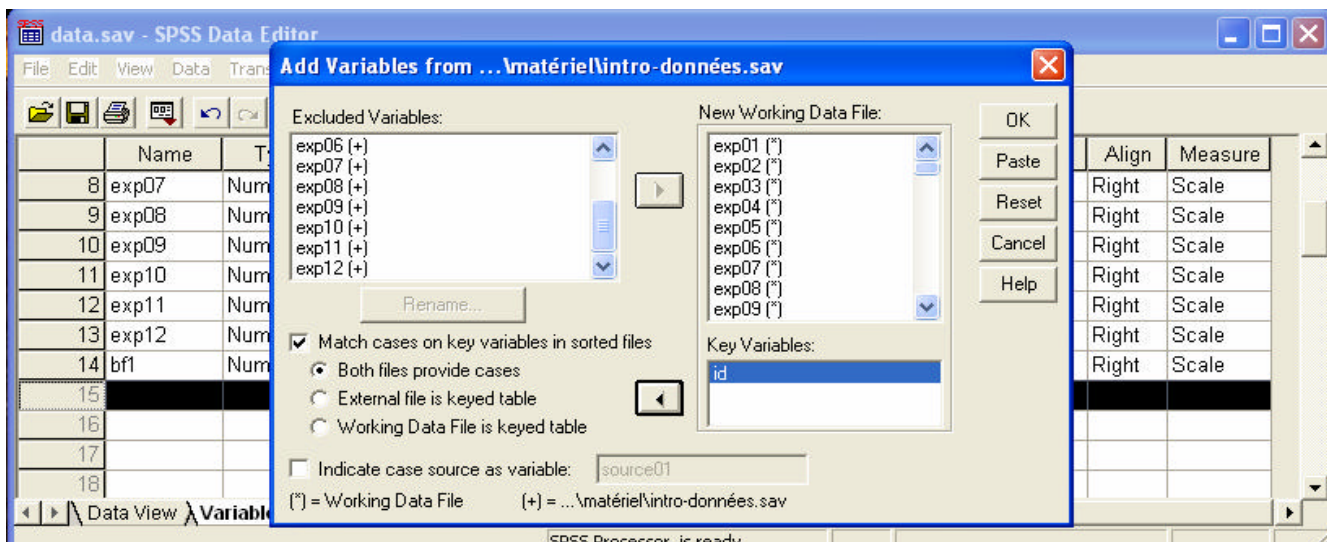


Figure 13

Une fois toutes les variables qui nous intéressent sélectionnées, il faut cliquer sur ok, ce qui ajoute les variables du deuxième fichier sélectionnées au premier. On a maintenant un nouveau fichier de données. Si nous sauvons ce fichier en faisant **File | Save**, cela va écraser le premier fichier. Il est beaucoup plus prudent de sélectionner **File | Save as ...** On obtient ainsi une nouvelle base de données, différentes des deux initialement utilisées pour sa création.

## Utiliser Excel pour entrer les données dans SPSS

### Créer une grille de données avec Excel

- Tapez les noms des variables (pas les étiquettes) dans la première ligne
- Les lignes au-dessous des noms représentent les observations (une observation (cas ou individu) par ligne).
- Sauvegardez les données (**File | Save as ...**) comme classeur Excel (extension .xls)

### Importer les données dans SPSS

- Ouvrir SPSS
- Cliquez sur **File | Open a database | New query**
- Sélectionnez « Fichiers Excel », cliquez sur suivant

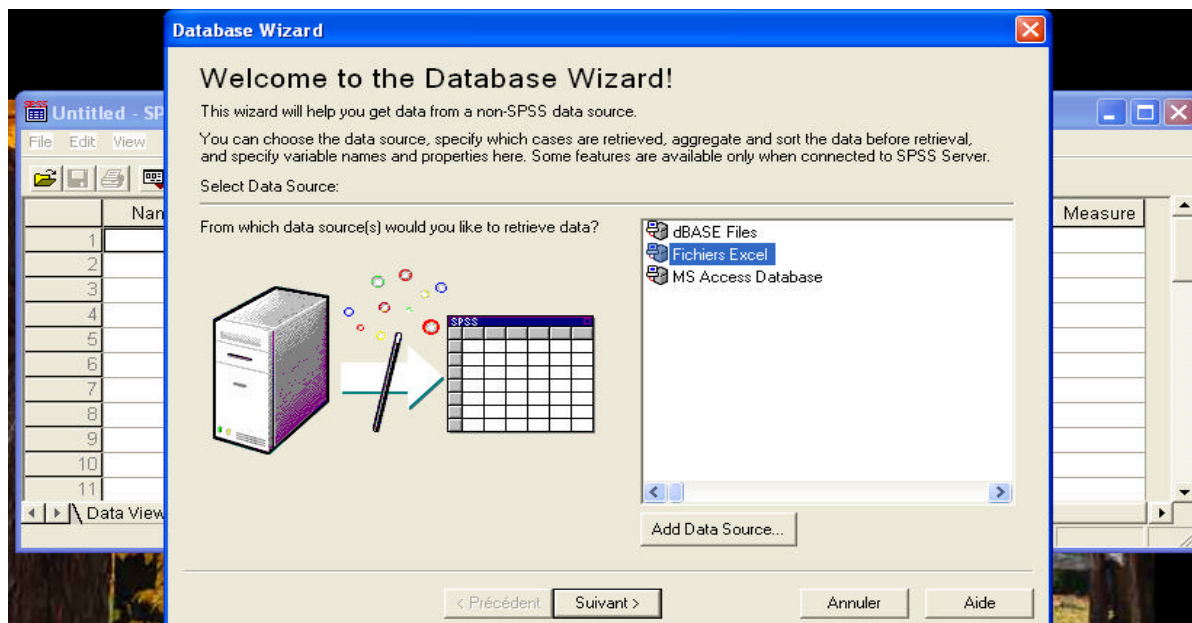


Figure 14

- Sélectionnez la ou les feuilles Excel où se trouvent les données dont vous avez besoin en faisant glisser les feuilles ou les variables dans les feuilles qui nous intéressent dans la fenêtre de gauche vers la fenêtre de droite. Puis cliquez sur suivant.

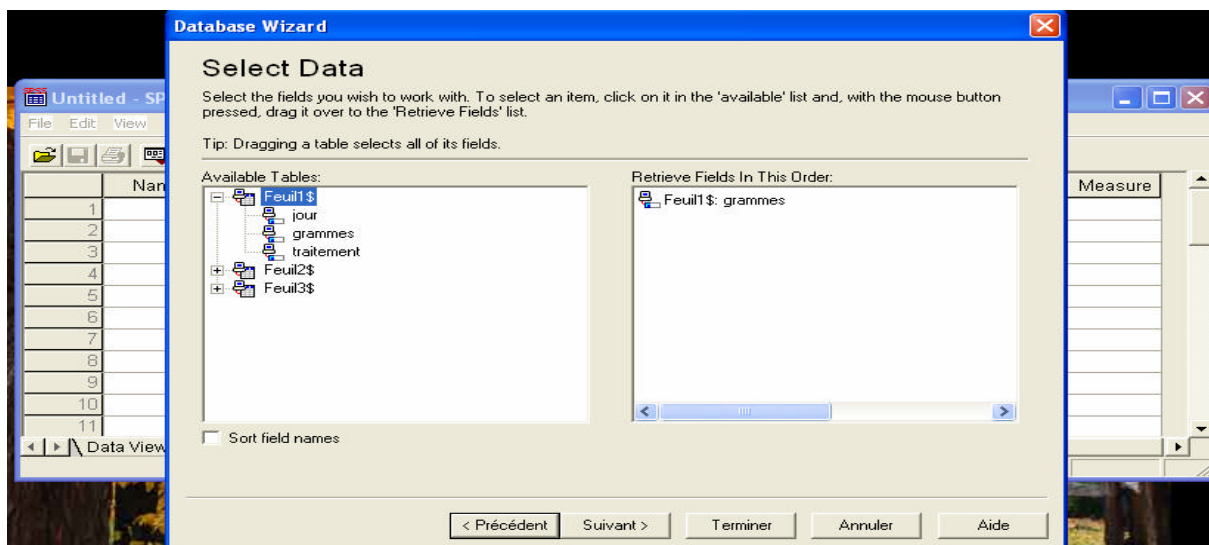


Figure 15

- Définir les variables : type des variables, étiquettes, valeurs, etc.

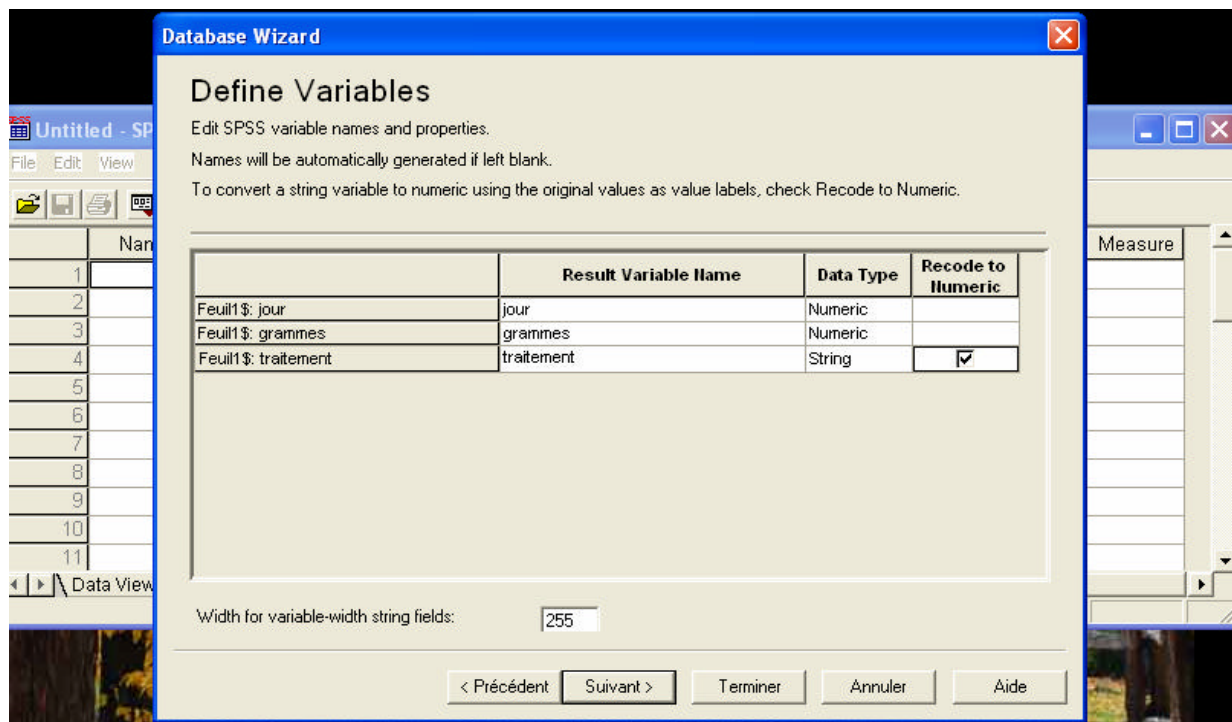


Figure 16

- Terminer et Sauvegardez les données (**File | Save as...**) comme fichier SPSS (extension .sav).

### Utiliser un fichier texte pour entrer les données dans SPSS

#### Créer une grille de données avec un fichier texte

- Ouvrir un fichier texte (p.ex. notepad)
- Tapez les noms des variables dans la première ligne que vous séparer soit par un point-virgule (ou autre séparateur comme tab)
- Les lignes au-dessous des noms représentent les observations (une observation (cas ou individu) par ligne) et chaque valeur de la ligne doit être séparé par un séparateur (toujours utiliser le même séparateur).
- Sauvegardez les données (**File | Save as...**) comme classeur texte (extension .txt)

#### Importer les données dans SPSS

- Ouvrir SPSS
- Cliquez sur **File | Read text data ...**
- Spécifiez l'arrangement de vos variables : types de séparateur, est-ce que chaque ligne correspond à un cas, est-ce que le fichier contient le nom des variables sur la première ligne, etc.
- Terminer et Sauvegardez les données (**File | Save as...**) comme fichier SPSS (extension .sav).

## Éliminer les erreurs de l'entrée de données

Deux types d'erreurs possibles : a) la valeur fautive est une valeur dans l'étendue des valeurs valables ou b) la valeur fautive est une valeur dehors de l'étendue des valeurs valables.

On peut repérer les erreurs du type (a) qu'en comparant les données vraies avec les données entrées. Mais cela est très coûteux, d'où l'importance d'être très minutieux en entrant les données ! Donc seules les erreurs de type (b) peuvent être repérées et éliminées après l'entrée des données.

Pour cela, il faut inspecter les valeurs minimales et maximales de toutes les variables en passant par **Analyze | Descriptive Statistics | Frequencies**. Il faut sélectionner toutes les variables numériques (en utilisant la souris ou les touches SHIFT et ↓) et ajoutez-les dans la section droite ►, puis cliquez dans « **Statistics** », choisissez « **Minimum** » et « **Maximum** »

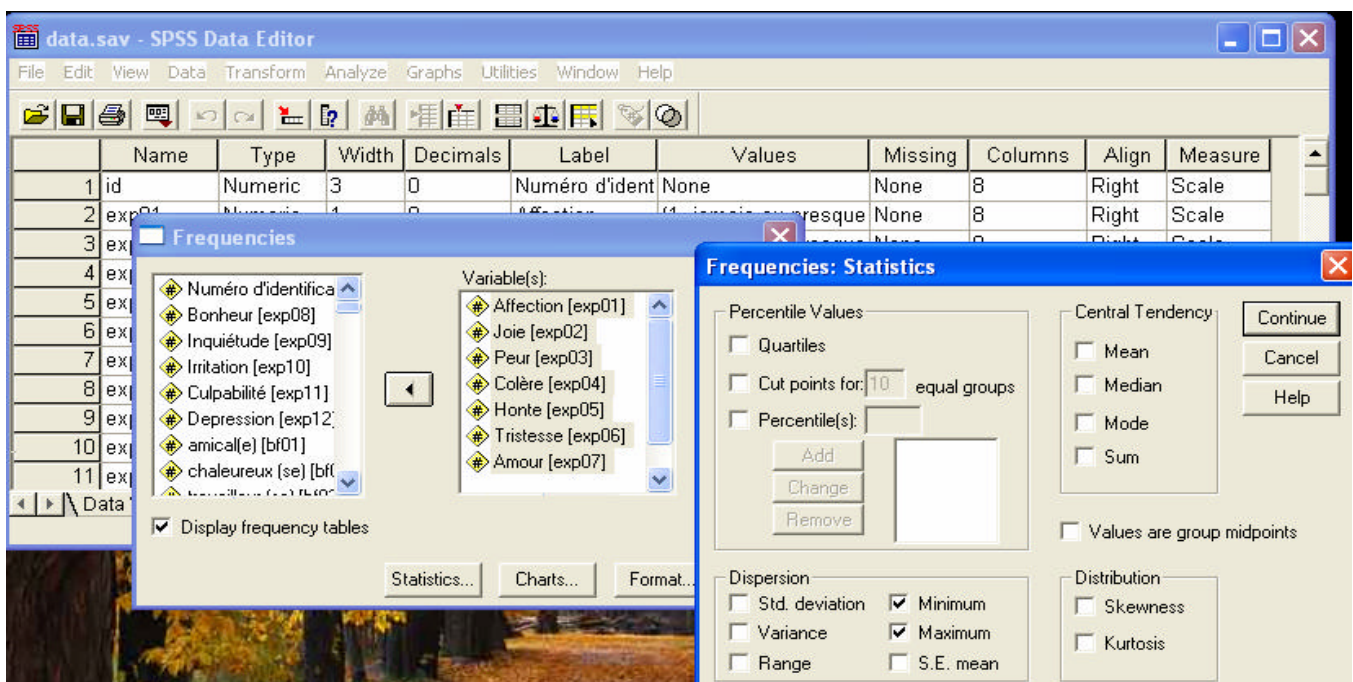


Figure 17

Dans la page des résultats, le tableau « **fréquences** » peut être pivoté en double-cliquant sur le tableau et cliquant sur le menu **Pivot | Transpose rows and columns** ce qui rend le tableau plus facile à lire.

S'il y a des valeurs dehors de l'étendue valable :

- Vérifier combien de valeurs sont fautes : pour chaque variable ayant un minimum trop petit ou un maximum trop grand. Pour cela, il faut inspecter les tableaux de fréquences pour les variables ayant des valeurs erronées et noter le nombre de ces valeurs et leurs valeurs numériques.
- Corriger les valeurs fautes dans la vue des données : pour cela il faut trouver la variable p. ex. à l'aide du menu **Utilities | Variables ... « go to »** et la sélectionner. Puis allez dans **Edit | Find** et entrez la valeur fautive que vous cherchez et cliquez sur « **Find next** ». Activez la cellule. Ensuite, cliquez sur le menu **Utilities | Variables**,



choisissez la variable d'identification et cliquez sur « Go to » : la cellule montre le numéro du cas cherché.

- Vérifiez la valeur correcte dans les documents (p. ex., questionnaire) de ce cas et corrigez-la dans la base de données
  - ➔ Attention: Après avoir trouvé la cellule avec une valeur fautive, il semble peut-être plus facile de noter le numéro de la ligne dans la grille de données (au lieu d'aller via le menu). Mais le numéro de la ligne ne doit pas être identique avec le numéro d'identification du cas !!! (p. ex. si le fichier a été trié par une autre variable ou si la variable d'identification a des valeurs manquantes)

## TP2 : Entrées les données à partir du questionnaire

### Exercice 1: Entrer les données

Préparez l'entrée des données en définissant les variables dans la vue des variables.

Vous avez en votre possession 2 exemplaires du questionnaire remplis. Entrez ces données dans la vue des données.

Créez un classeur « exercices » dans votre domaine (h:\...) et enregistrez-y le fichier sous le nom « exercice1.sav ».

**Envoyez par mail le fichier exercice1.sav à [katia.iglesias@pse.unige.ch](mailto:katia.iglesias@pse.unige.ch) au plus tard une semaine après la date du TD2 afin de pouvoir préparer la base de données qui servira de support lors des TD suivants et des TP.**

### Exercice 2: Fusionner les données et éliminer les erreurs

#### A. Fusionner les données

Deux étudiants ont entré ensemble des données d'une étude – chacun a entré environ la moitié des questionnaires. Les données se trouvent dans deux fichiers qui se trouvent sur dokeos. Fusionnez-les.

- Ouvrez le premier fichier
- « questionnaire-données avec erreurs-1.sav »
- ajoutez les « cases » (sujets) du deuxième fichier
- « questionnaire-données avec erreurs-2.sav »

Enregistrez le fichier fusionné sous le nom « exercice2a.sav » dans le classeur « exercices » sur votre domaine (h:\...)

#### B. Eliminer les erreurs

Vérifiez si les étudiants ont entré les données sans erreurs.

- Utilisez le fichier « exercice2.sav » que vous venez de créer.
- En utilisant **Analyze | Descriptive statistics | Frequencies** trouvez les valeurs fausses.
- Notez le nom de la variable et le numéro d'identification du sujet pour chaque erreur.

Corrigez les erreurs\* et sauvegardez le fichier sous le nom « exercice2b.sav » dans votre domaine (h:\...).

## TD3 : Préparation des données : Transformer les données

### Transformer les données

Avant d'effectuer des analyses statistiques, il est souvent nécessaire de modifier les données brutes :

- recoder des items (regrouper des catégories, inverser le sens de l'échelle, etc.),
- construire des indicateurs (calculer des scores d'échelles, des sous échelles, centrer et réduire un variable, etc.).

Quelques conseils au moment d'effectuer des transformations de données :

- Après avoir éliminé les erreurs, gardez toujours un fichier des données brutes.
- Avant d'effectuer des modifications, sauvegardez les données sous un autre nom pour conserver le fichier original!
- Toutes les variables initiales restent dans le fichier. Les variables recodées obtiennent un nouveau nom.

### Recoder des variables

Pour recoder une variable que ce soit pour regrouper des catégories ou inverser le sens de l'échelle, il faut passer par **Transform | Recode | into different variables**. Il faut éviter d'utiliser **Recode | into same variables**, parce que cette option écraserait votre variable initiale.

Exemple: dans la boîte de dialogue **Transform | Recode | into different variables**, choisissez une (ou plusieurs) variables ►. Ici on choisit ouv02.

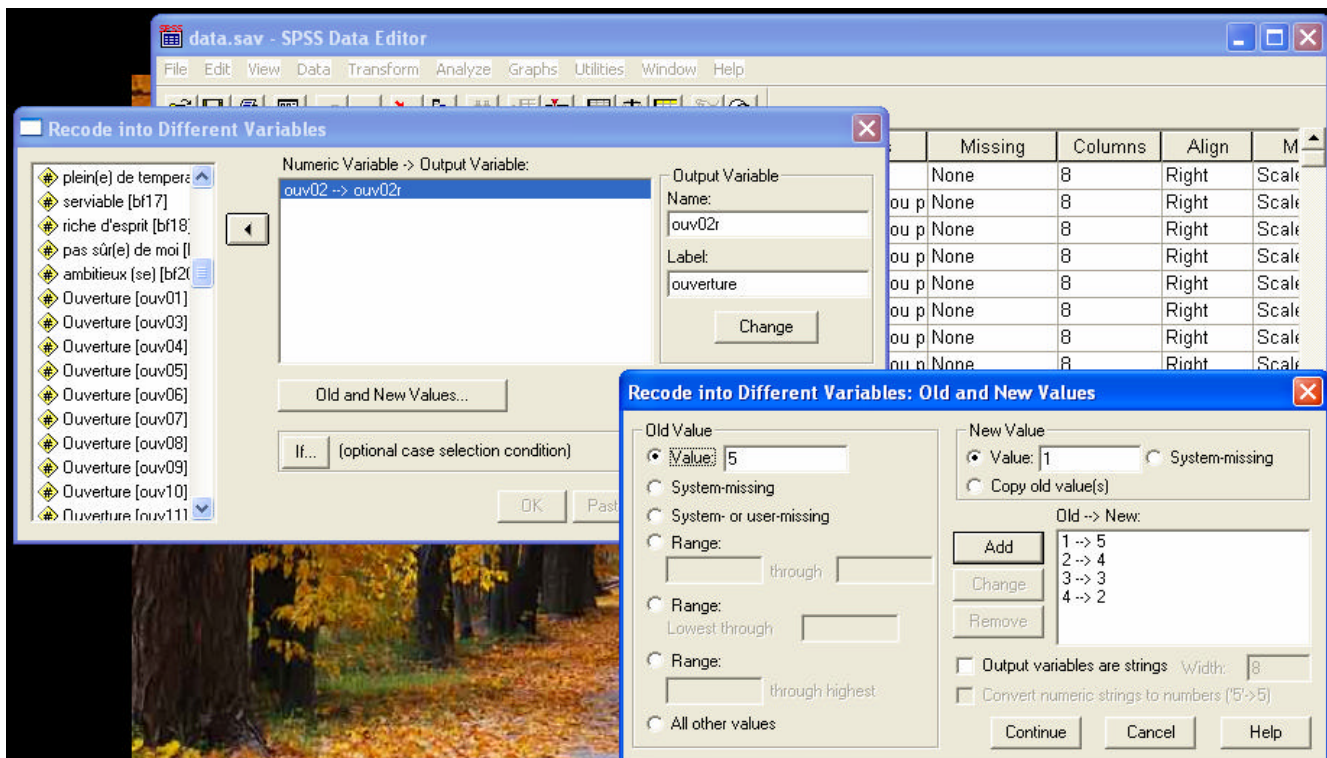


Figure 18

Dans Output variable, on peut choisir le nouveau nom de la variable, ainsi que l'étiquette de la variable. Puis on sélectionne « **change** » pour que SPSS puisse faire le lien entre l'ancienne et la nouvelle variable (ouv02 → ouv02r). Puis on sélectionne « **Old and new values** ». Une nouvelle boîte de dialogue s'ouvre où l'on peut préciser l'ancienne valeur à gauche et la nouvelle valeur à droite. A chaque changement de valeur, il faut appuyer sur « **Add** ».

Lorsqu'il y a des valeurs manquantes dans notre variable initiale, si rien n'est spécifié, alors SPSS attribuera également des valeurs manquantes (sysmis) à la nouvelle variable.

Lorsque dans un questionnaire plusieurs variables d'une échelle sont codées dans le sens inverse des autres (par exemple ouv02 est un élément de l'échelle de mesure de l'ouverture, avec 1 pas du tout ouvert à 5 tout à fait ouvert, mais ouv02 (je suis bien installé dans mes habitudes) va de 1 tout à fait ouvert à 5 pas du tout ouvert), il est possible de recoder toutes les variables dont l'échelle est inversée pour une même dimension en une seule fois.

## Construire des indicateurs

Pour construire une nouvelle variable à partir de plusieurs variables de départ, il faut passer par **Transform | Compute**. Une manière usuelle pour agréger différentes variables, c'est d'utiliser la moyenne ou la somme des différentes variables initiales. Pour construire la sous-échelle d'expression de joie, nous allons utiliser la moyenne des deux éléments qui composent cette sous-échelle : exp02 (joie) et exp08 (bonheur).

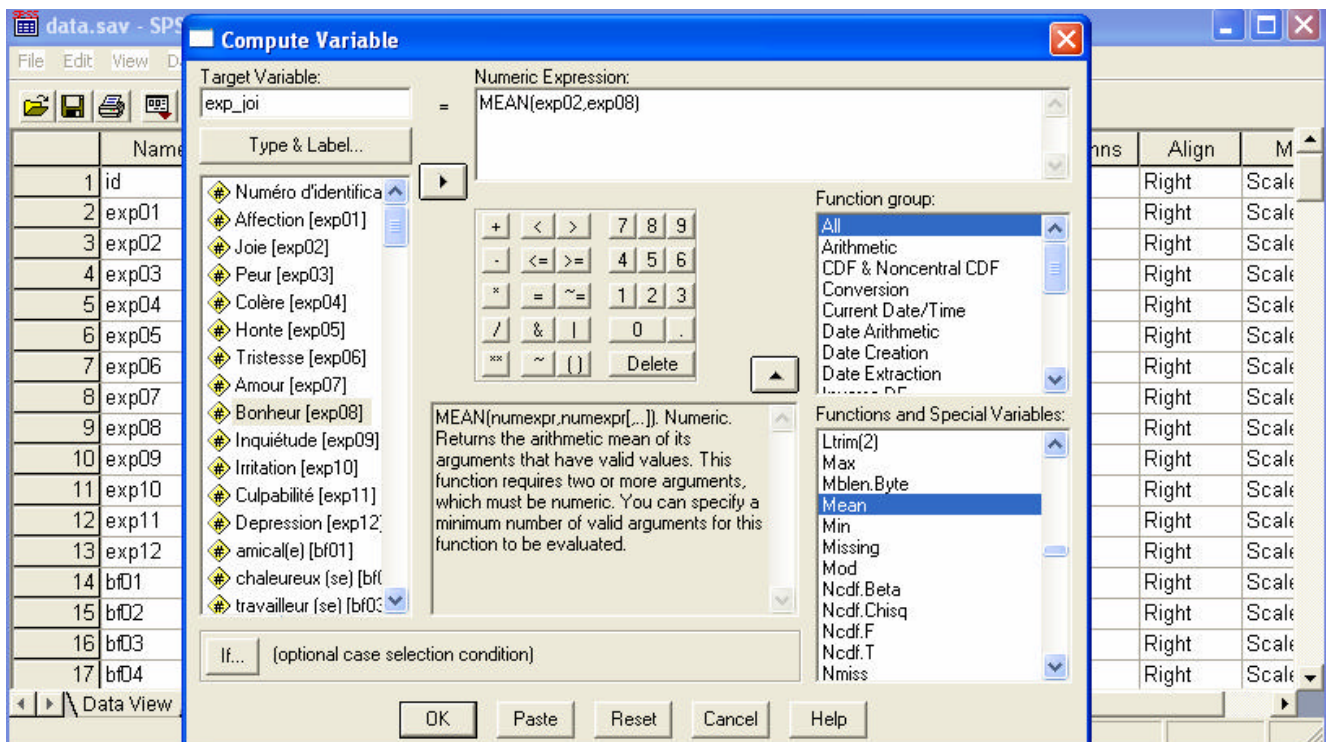


Figure 19

Pour cela, on va sous **Transform | Compute**. On obtient la boîte de dialogue ci-dessus. On définit le nom de la nouvelle variable (« **target variable** »), ainsi que le type, l'étiquette et les valeurs en cliquant en haut à gauche sous le nom de la variable. En haut à droite sous « **numeric expression** » on définit la transformation des variables initiales qu'on souhaite effectuer, ici une moyenne. Les transformations qui peuvent être effectuées sont définies en

bas à droite dans « **functions and special variables** ». En cliquant sur une des fonctions, on obtient au centre de la boîte de dialogue une explication de ce que fait l'opération choisie et comment rentrer les variables initiales (qui se trouvent à gauche dans la boîte).

Au lieu de mettre dans « **numerique expression** » : MEAN(exp02,exp08), il est possible de calculer la moyenne des items seulement pour les personnes qui ont une valeur valable (non manquante) sur au moins « g » de ces items (sinon, la nouvelle valeur sera manquante). Pour cela il faut spécifier le nombre de valeurs valables après l'expression MEAN : MEAN.g(exp02,exp08). Par exemple, MEAN.2(exp02,exp08) ne fait la moyenne que s'il n'y a pas de valeur manquante. N.B. : Il est recommandé d'avoir au moins 80% de valeurs valables pour calculer le score (la moyenne) d'un échelle. Si on choisit MEAN(exp02,exp08), cela correspond à MEAN.1(exp02,exp08), c'est-à-dire qu'on fait la moyenne au moins une variable (mais pas forcément plus).

Attention: On pourrait aussi calculer la somme de tous les items d'une échelle (Sum). Cependant, il n'est pas évident comment gérer les valeurs manquantes en créant la somme des items – si on enlève tous les cas avec les valeurs manquantes on risque de fortement diminuer l'échantillon. Le remplacement des valeurs manquantes par la moyenne de l'échantillon est aussi problématique! C'est pour cela que calculer la moyenne est recommandé.

### Transformer les données en utilisant la page de syntaxe

- Dans la boîte de dialogue d'un menu **Transform** | ... appuyez sur le bouton « **Paste** » au lieu de « **Ok** ». La commande exécutée s'inscrira sur la page de syntaxe.
- La syntaxe est un fichier de texte qui peut être sauvegardé (extension .sps) et imprimé. Les commandes peuvent être copiées, collées et changées.
- On peut ajouter du texte commentaire qui doit être précédé d'un astérisque et suivi par un point.
- Pour exécuter les commandes, on la sélectionne et on envoie la syntaxe en appuyant sur le bouton ►.

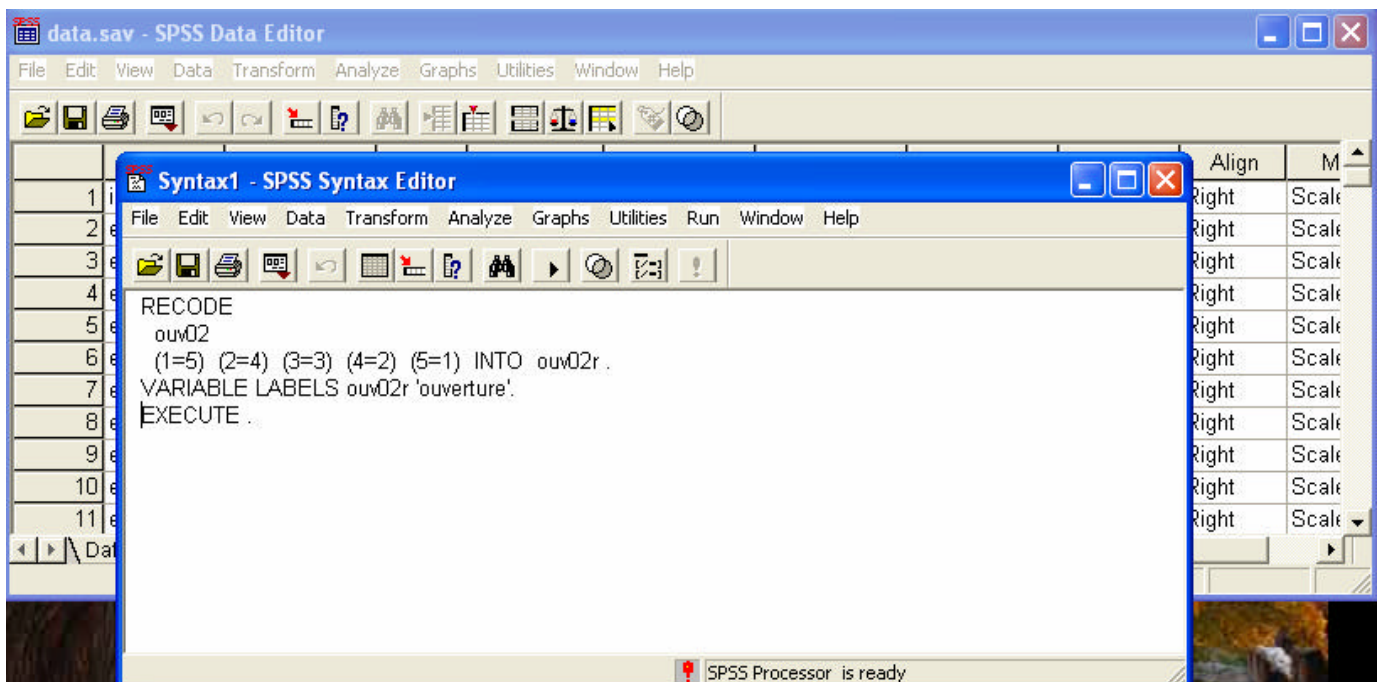


Figure 20

Pourquoi travailler avec la page de syntaxe (cf TD1) :

- **Transparence:** Les commandes peuvent être sauvegardées, imprimées, et contrôlées.
- Les commandes enregistrées peuvent être réutilisées dans les séances de travail ultérieures.
- Les règles les plus importantes à suivre:
- Une commande se compose de son nom (p. ex. recode) et de ses spécifications (p. ex. noms des variables, règle de recodage).
- Chaque commande doit débiter sur une nouvelle ligne et se terminer par un point.
- Il n'y a pas de lignes vides à l'intérieur d'une commande.

Règles de transformation pour notre questionnaire:

Variable	Variable label	Items à recoder	Echelles
exp01 – exp12	Expressivité	-	<b>Amour</b> (exp_amo) : exp01, exp07 <b>Joie</b> (exp_joi): exp02, exp08 <b>Peur</b> (exp_peu) : exp03, exp09 <b>Colère</b> (exp_col) : exp04, exp10 <b>Tristesse</b> (exp_tri) : exp06, exp12 <b>Honte</b> (exp_hon) : exp05, exp11
bf01 – bf20	Big Five	-	<b>Extraversion</b> (bf_ext): bf01, bf06, bf13, bf16 <b>Agréabilité</b> (bf_agr): bf02, bf08, bf15, bf17 <b>Conscience</b> (bf_con): bf03, bf07, bf11, bf20 <b>Névrosisme</b> (bf_nev): bf04, bf10, bf12, bf19 <b>Intelligence</b> (bf_int) : bf05, bf09, bf14, bf18
ouv01 – ouv30	Ouverture	ouv02, ouv03, ouv04, ouv06, ouv07, ouv09, ouv11, ouv12, ouv14, ouv16, ouv18, ouv20, ouv21, ouv22, ouv24, ouv25, ouv28, ouv30,	<b>Ouverture aux rêveries</b> (ouv-rev) : ouv01, ouv04, ouv11, ouv17, ouv22 <b>Ouverture à l'esthétique</b> (ouv-est): ouv05, ouv07, ouv13, ouv18, ouv23 <b>Ouverture aux sentiments</b> (ouv-sen): ouv08, ouv14, ouv19, ouv24, ouv27 <b>Ouverture aux actions</b> (ouv-act): ouv02, ouv09, ouv15, ouv20, ouv28 <b>Ouverture aux idées</b> (ouv-ide): ouv06, ouv10, ouv16, ouv25, ouv29 <b>Ouverture aux valeurs</b> (ouv-val): ouv03, ouv12, ouv21, ouv26, ouv30

## TP3 : Préparation des données : Transformer les données

### Exercice 3: Créer de nouveaux indicateurs

Créez des scores d'échelles représentant les dimensions de l'expressivité des émotions et les dimensions des « Big Five » :

<b>Amour</b> (exp_amo) : moyenne de exp01, exp07 <b>Joie</b> (exp_joi): moyenne de exp02, exp08 <b>Peur</b> (exp_peu) : moyenne de exp03, exp09 <b>Colère</b> (exp_col) : moyenne de exp04, exp10 <b>Tristesse</b> (exp_tri) : moyenne de exp06, exp12 <b>Honte</b> (exp_hon) : moyenne de exp05, exp11	<b>Extraversion</b> (bf_ext): moyenne de bf01, bf06, bf13, bf16 <b>Agréabilité</b> (bf_agr): moyenne de bf02, bf08, bf15, bf17 <b>Conscience</b> (bf_con): moyenne de bf03, bf07, bf11, bf20 <b>Névrosisme</b> (bf_nev): moyenne de bf04, bf10, bf12, bf19 <b>Intelligence</b> (bf_int) : moyenne de bf05, bf09, bf14, bf18
--	---

Démarche :

- Ouvrez le fichier « data.sav »

- Utilisez le menu **Transform | Compute** pour la première variable, entrez toutes les informations nécessaires (nom, label/étiquette, formule) et appuyez sur paste (au lieu de ok) pour transmettre la commande à la syntaxe.
- Dans la syntaxe, copiez, collez et adaptez la commande pour calculer les autres scores d'échelles.
- Acceptez au maximum 20% de données manquantes.

Exécuter les commandes. Sauvegardez la syntaxe sous le nom « exercice3.sps » et les données sous le nom « exercice3.sav » dans votre domaine (h:\...).

#### Exercice 4 : Recoder et créer de nouvelles variables

L'ouverture aux émotions est constituée de 6 dimensions. Pour chacune de ces dimensions certaines variables sont mesurées avec une échelle positive (mesure de l'ouverture) et d'autres sont mesurées avec une échelle négative (mesure de la fermeture). Avant de pouvoir construire nos nouveaux indicateurs d'ouverture, il faut dans d'abord faire en sorte que toutes les variables mesurent l'ouverture.

Variables avec échelle négative	Les 6 dimensions d'ouverture à construire
ouv02, ouv03, ouv04, ouv06, ouv07, ouv09, ouv11, ouv12, ouv14, ouv16, ouv18, ouv20, ouv21, ouv22, ouv24, ouv25, ouv28, ouv30,	<p><b>Ouverture aux rêveries</b> (ouv-rev) : moyenne de ouv01, ouv04, ouv11, ouv17, ouv22</p> <p><b>Ouverture à l'esthétique</b> (ouv-est): moyenne de ouv05, ouv07, ouv13, ouv18, ouv23</p> <p><b>Ouverture aux sentiments</b> (ouv-sen): moyenne de ouv08, ouv14, ouv19, ouv24, ouv27</p> <p><b>Ouverture aux actions</b> (ouv-act): moyenne de ouv02, ouv09, ouv15, ouv20, ouv28</p> <p><b>Ouverture aux idées</b> (ouv-ide): moyenne de ouv06, ouv10, ouv16, ouv25, ouv29</p> <p><b>Ouverture aux valeurs</b> (ouv-val): moyenne de ouv03, ouv12, ouv21, ouv26, ouv30</p>

Démarche :

- Ouvrez le fichier « data.sav »
- Recode : Utilisez le menu Transform | Recode into different variable pour la première variable, entrez toutes les informations nécessaires (nom, label/étiquette, formule) et appuyez sur paste (au lieu de ok) pour transmettre la commande à la syntaxe.
- Dans la syntaxe, copiez, collez et adaptez la commande pour recoder les autres variables
- Construction des indicateurs : Utilisez le menu Transform | Compute pour la première variable, entrez toutes les informations nécessaires (nom, label/étiquette, formule) et appuyez sur paste (au lieu de ok) pour transmettre la commande à la syntaxe.
- Dans la syntaxe, copiez, collez et adaptez la commande pour calculer les autres scores d'échelles.
- Acceptez au maximum 20% de données manquantes.

Exécuter les commandes. Sauvegardez la syntaxe sous le nom « exercice4.sps » et les données sous le nom « exercice4.sav » dans votre domaine (h:\...).

#### Exercice 5 : Ajouter des variables

Ajouter au fichier de données exercice3.sav les nouvelles variables construites. Pour cela, ouvrir exercice3.sav, ajoutez les nouveaux indicateurs qui se trouvent dans exercice4.sav. Enregistrez le fichier fusionné sous le nom « exercice5.sav » dans le classeur « exercices » sur votre domaine (h:\...)

## TD4 : Analyse des données : Représentations graphiques

### Fréquences

Pour obtenir un tableau de fréquences, on utilise le menu: **Analyze | Descriptive Statistics | Frequencies.**

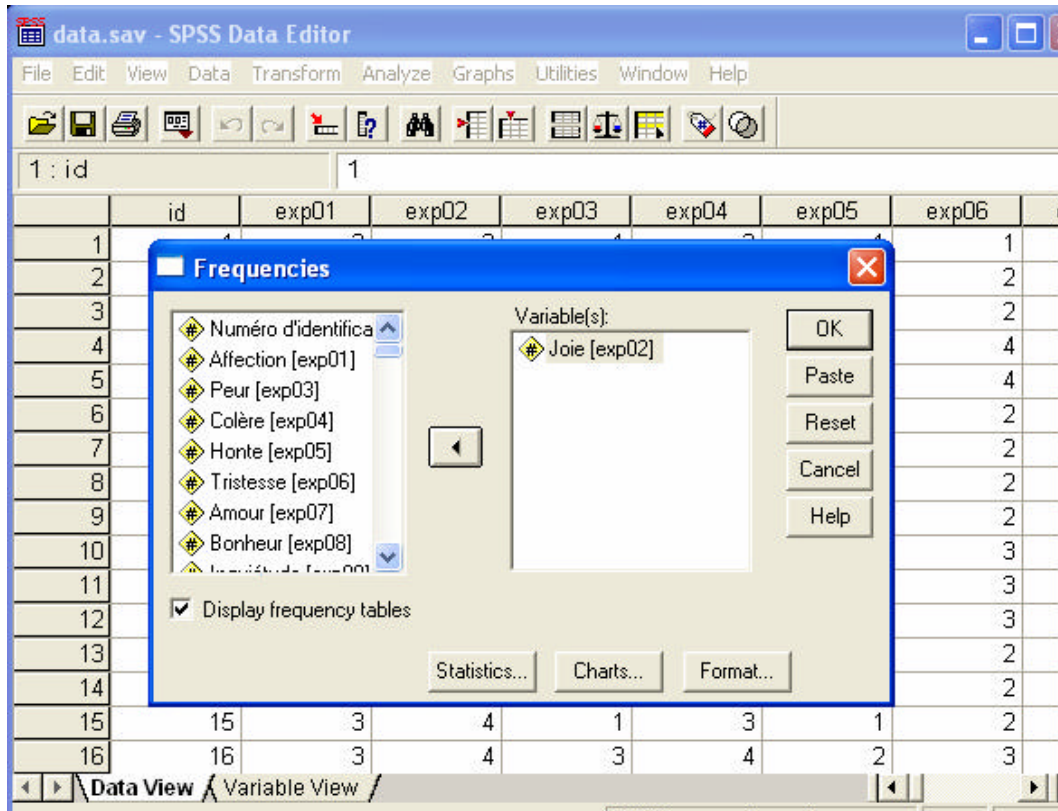


Figure 21

- Choisir une ou plusieurs variables

		Joie			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	jamais ou presque jamais	1	.2	.2	.2
	parfois	25	5.2	5.2	5.4
	souvent	196	40.7	40.7	46.1
	presque toujours ou toujours	260	53.9	53.9	100.0
	Total	482	100.0	100.0	

Figure 22

## Graphiques pour les variables nominales et ordinales (montrant les fréquences)

### Bar (bâtons):

#### *Distribution des fréquences d'une variable*

Présentation graphique de la distribution des fréquences d'une variable dans tout l'échantillon

**Graphs | Bar:** Choisir **Simple** et **Summaries for groups of cases**

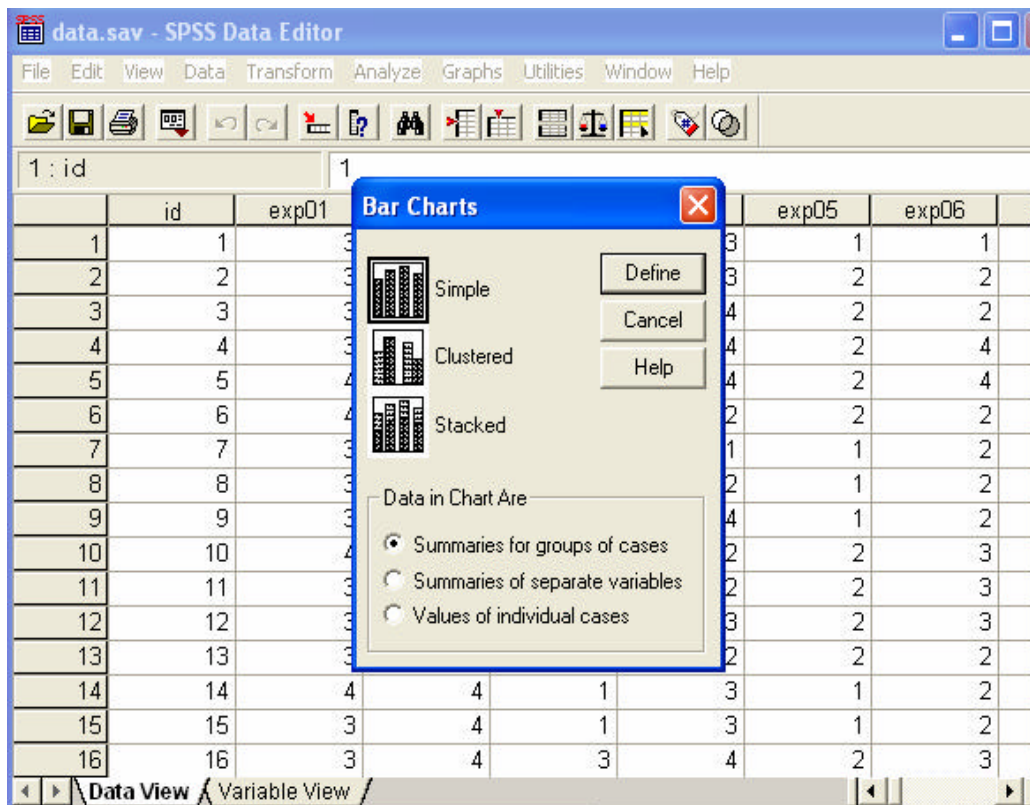


Figure 23

Sous « **Bars represent** », on peut choisir : N of cases (fréquences absolues), % of cases (fréquences relatives), Cum. N (effectif cumulé), Cum. % (pourcentage cumulé) ou un autre type de statistique (que nous verrons dans le TD5)

Sous « **Category axes** », il faut entrer la variable pour laquelle on veut obtenir les fréquences.



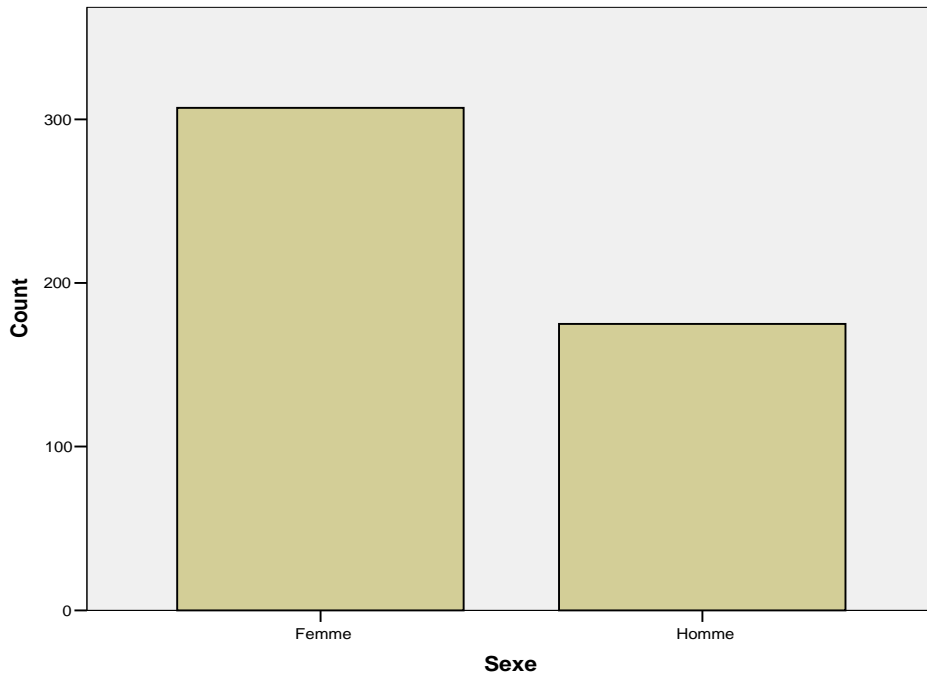


Figure 24

*Distribution des fréquences d'une variable par sous-groupes*

Pour comparer la distribution des fréquences d'une variable dans les sous-groupes, passez par **Graphs | Bar**, choisir « **Clustered** » et « **Summaries for groups of cases** ». Pour « **Bars represent** » choisir **N of cases** (fréquences absolues) ou **% of cases** (fréquences relatives). Dans « **category axes** » entrer la variable pour laquelle on veut obtenir les fréquences et dans « **Define Clusters by** » la variable de groupement pour laquelle on veut obtenir des différents bâtons.

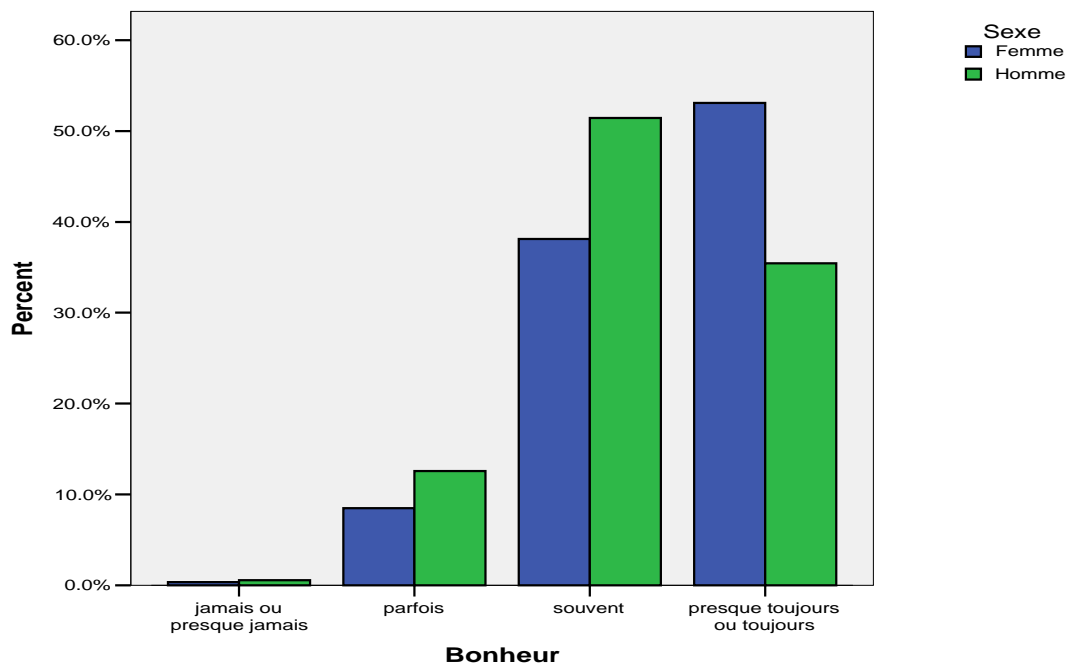


Figure 25

## Pie (Camembert)

Présentation graphique de la distribution des fréquences d'une variable dans tout l'échantillon, on passe par **Graphs | Pie**: « **Summaries for groups of cases** », dans « **Slices represent** » on peut choisir **N of cases** (fréquences absolues) ou **% of cases** (fréquences relatives). Dans « **Define Slices by** » on entre la variable qui nous intéresse ici : je suis travailleur/se

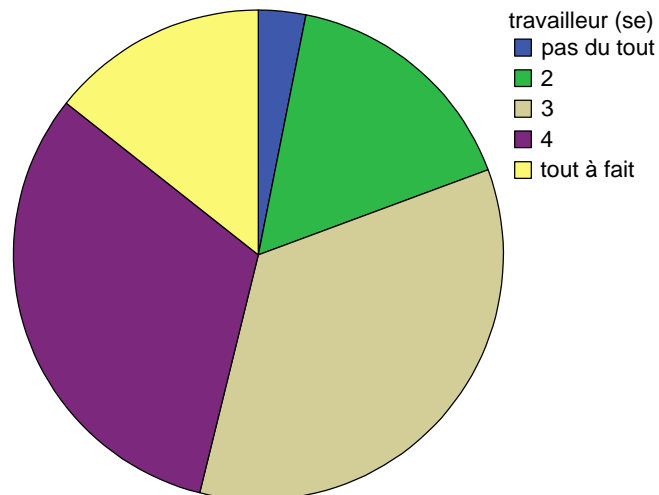


Figure 26

## Graphiques pour les variables métriques

### Histogramme

Pour présenter la distribution des fréquences d'une variable métrique dans tout l'échantillon, on peut faire un histogramme, pour cela, allez sous **Graphs | Histogram**, entrez la variable pour laquelle vous voulez obtenir les fréquences dans « **Variable** », par exemple : ouv \_sen.

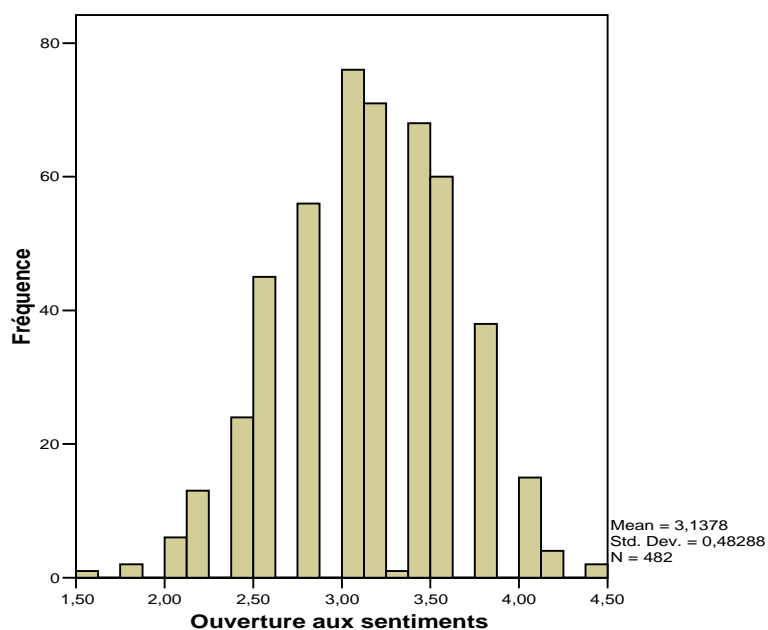


Figure 27

## Modifier les caractéristiques d'un graphique

Pour pouvoir modifier un graphique, il faut commencer par double-cliquer dessus. Dans l'éditeur des graphiques (chart editor) qui s'ouvre on peut modifier différents paramètres :

- les propriétés (p.ex. taille, couleurs, motifs) des différents éléments d'un graphique (p.ex. titre, axes, bâtons) en double-cliquant sur l'élément (après choisir l'onglet pertinent, p.ex. « **Fill and Border** » pour changer les couleurs) ;
- le texte (p.ex. d'un titre ou d'une catégorie) en grisant l'élément (cliquer une fois) et - après une « petite pause » - en cliquant une 2<sup>e</sup> fois dessus (double-cliquer va ouvrir la boîte-dialogue pour changer les propriétés !).

Par exemple pour changer l'échelle de l'axe des ordonnées, il faut, après avoir double-cliquer sur le graphique, double-cliquer sur l'axe. Dans la boîte-dialogue, choisissez l'onglet « **scale** ». Cliquer sur minimum ou maximum (colonne « **auto** ») et remplacez les valeurs dans la colonne « **custom** ». En cliquant sur « **major increment** », on peut changer les intervalles. Et pour transposer les axes, on peut passer par Menu **Chart | Transpose Chart**

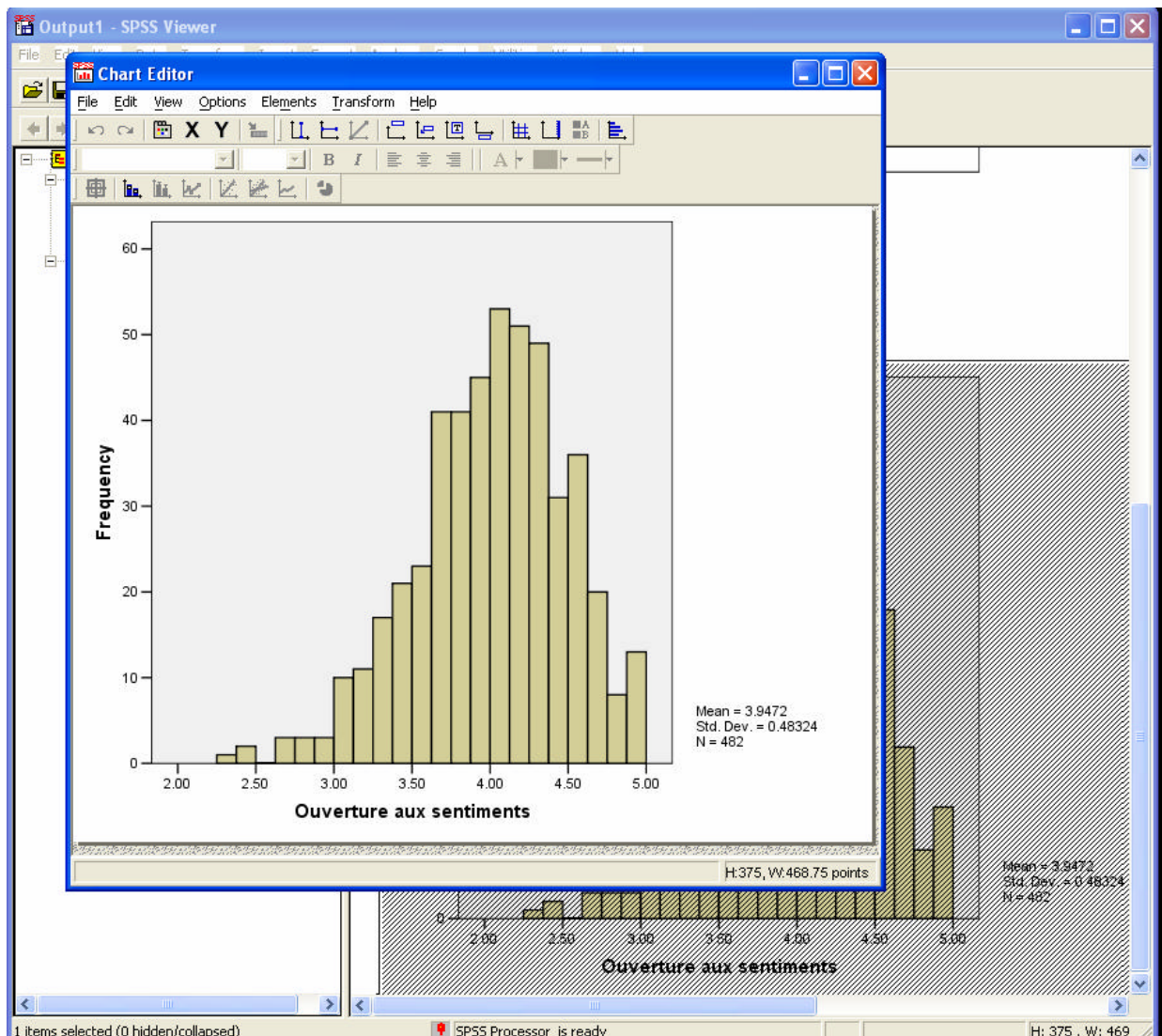


Figure 28

## Exporter des tableaux et des graphiques dans Word

Pour exporter un tableau ou un graphique, sélectionner dans le SPSS-Viewer le tableau et cliquez sur le bouton droit de la souris. Sélectionnez « **Copy** » et les résultats seront collés dans un tableau Word qui peut être modifié ou « **Copy objects** » et les résultats seront collés dans le format du tableau SPSS, ce dernier ne pourra plus être modifié dans Word. Dans Word, cliquez sur le bouton droit de la souris et sélectionnez coller.

La taille du graphique peut être modifiée en traînant un coin du cadre du graphique. Parfois, le graphique n'est pas bien reproduit dans Word – dans ce cas, il faut sauvegarder le graphique comme un fichier et l'insérer après dans Word. Dans le SPSS Viewer, cliquez droit sur le graphique, sélectionnez « **export...** ». Dans la boîte-dialogue, option « **export** », choisir « **charts only** ». Pour « **File Type** » choisissez « **Enhanced Metafile \*.EMF** ». En cliquant sur « **Browse** », choisissez l'endroit où vous voulez sauvegarder le fichier et son nom. Dans Word : menu **Insertion | Image | à partir du fichier**. Si le graphique est de nouveau mal reproduit, faire la même opération, mais Pour « **File Type** » choisir « **JPEG File \*.JPG** ». Ce type de format est très universelle, cependant la définition (en points par pouce) du graphique ne sera pas très bonne et typiquement pas acceptable pour une publication (cf. manuel de l'APA).

## TP4 : Analyse des données : Représentations graphiques

### Exercice 6 : Distributions des variables

Ouvrez le fichier « data1.sav »

- Analysez les fréquences des 4 items mesurant le névrosisme (bf4, bf10, bf12, bf19).
- Pour chaque item, créez des tableaux de fréquences. Modifiez les tableaux de sorte que le nombre de décimales soit égal à 2.
- Pour chaque item, créez un graphique en bâtons représentant les pourcentages de personnes.
- Créez des histogrammes pour les 6 indicateurs construits au TP3 mesurant l'ouverture.

Sauvegardez tous les résultats (« output ») sous le nom « exercice6.spo » dans votre domaine (h:\...).

## TD5 : Analyse des données : Mesures descriptives

### Mesures descriptives

Pour obtenir des informations (en forme de tableaux ou graphiques) sur la distribution d'une variable (fréquences, tendance (mode, médiane, moyenne), dispersion (variance, empan, intervalle interquartile)), on utilise le menu: **Analyze | Descriptive Statistics | Frequencies**

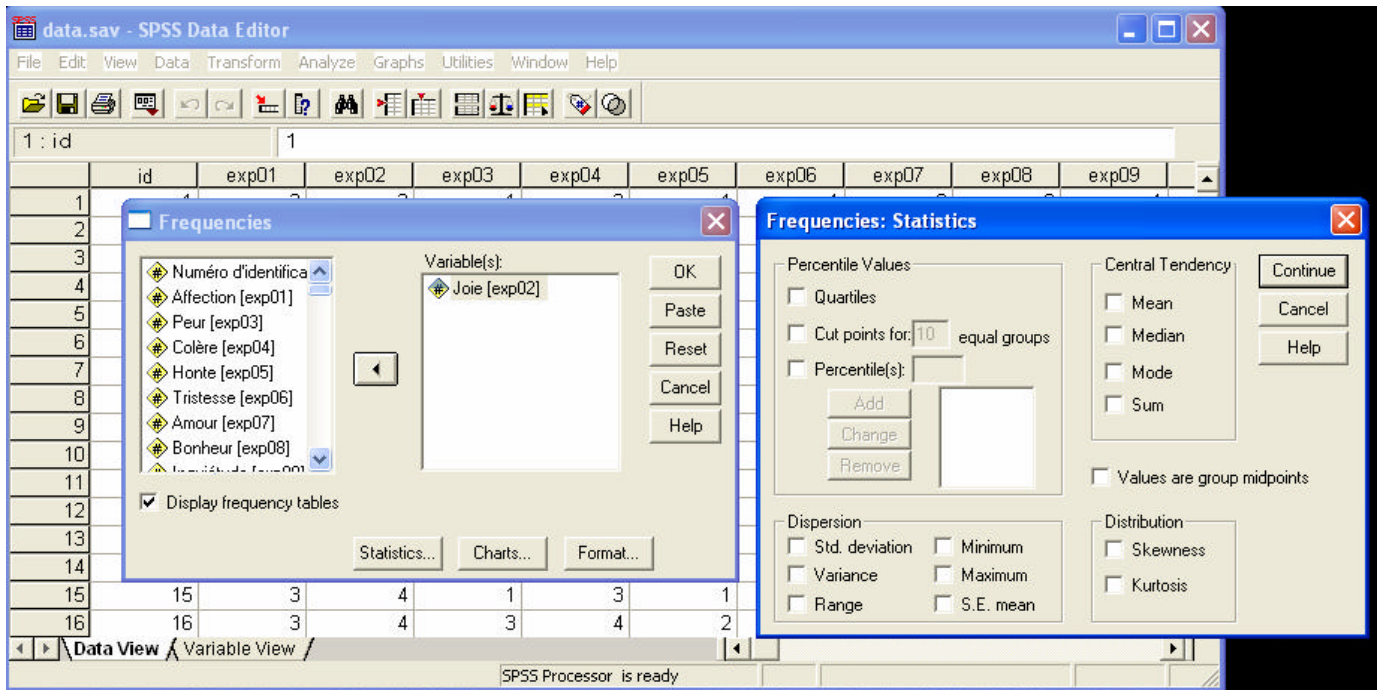


Figure 29

- Choisir une ou plusieurs variables
- *Display frequency tables*: tableaux de distribution de fréquences; il est coché par défaut (peut être enlevé si on souhaite seulement des statistiques et/ou des graphiques)
- *Statistics*: percentiles/quartiles, coefficients de tendance centrale [mean = moyenne, median = médiane, mode = mode], de dispersion [standard deviation = écart-type, variance = variance, range = étendue/empan, Min = minimum, Max = maximum, SE mean = erreur standard de la moyenne] et de distribution [skewness = coefficients d'asymétrie, kurtosis = coefficient d'aplatissement]

**Attention** le choix des statistiques dépend de l'échelle de mesure, mais SPSS calcule tous les coefficients pour toutes les variables choisies - même si ça n'a pas de sens!

- *Charts*: graphiques – on a le choix entre diagramme en bâtons (bar charts) pour les variables nominales ou ordinales, diagramme circulaire / camembert (pie charts) pour les variables nominales ou ordinales et histogramme (histogram) pour les variables métriques.

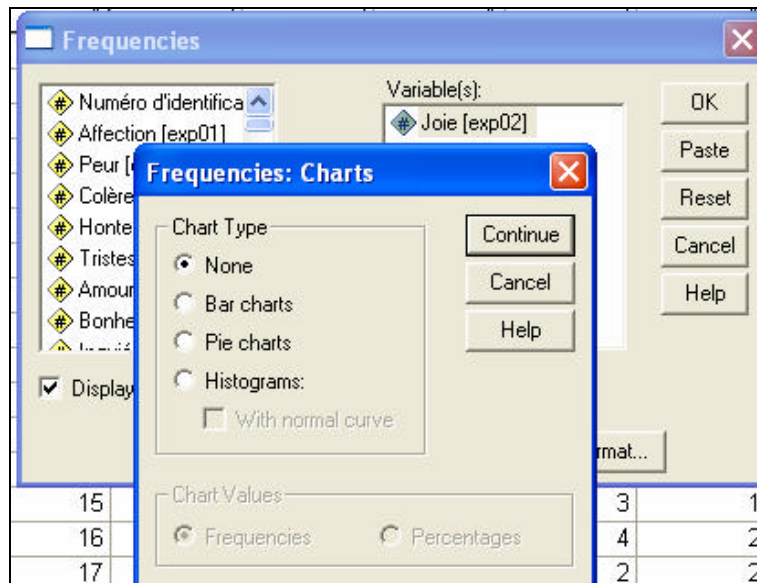


Figure 30

- *Format*: Ici, on peut choisir dans quel ordre les catégories de la variable seront affichées (order by...; default = ascending values) et la forme de présentation des coefficients de plusieurs variables (compare variables = tous les résultats dans un tableau, organize output by variables = tableaux séparés pour les différentes variables).

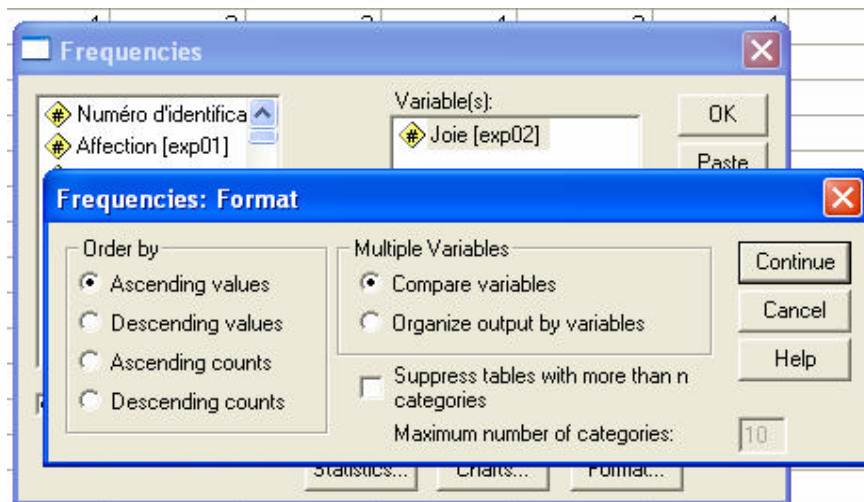


Figure 31

## Graphiques des mesures descriptives

### Bar (bâtons)

Présentation graphique de la moyenne d'une variable métrique dans différents sous-groupes, pour cela cliquez sur **Graphs | Bar**, choisissez « **Simple** » et « **Summaries for groups of cases** ». Dans Bars represent, choisissez « **Other summary function** » variable pour laquelle on veut obtenir les moyennes, MEAN(ouv\_sen) apparaît. Dans « **Category axes** », entrez la variable de groupement, par exemple sexe.

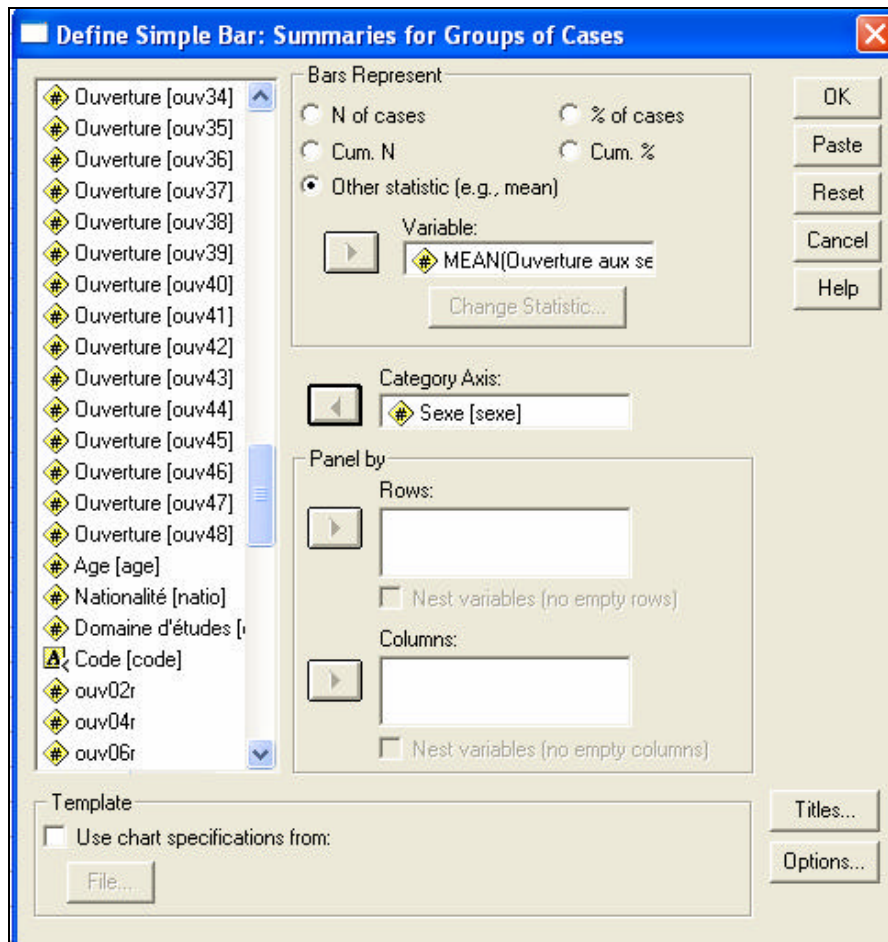


Figure 32

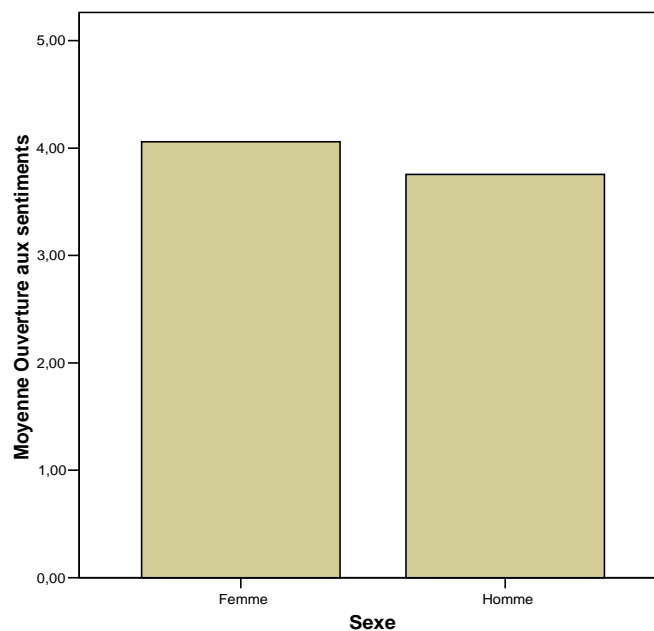


Figure 33

Présentation graphique de la moyenne de plusieurs variables métriques dans tout l'échantillon, pour cela cliquez sur **Graphs** | **Bar**, choisissez « **Simple** » et « **Summaries of separate variables** ». Dans **Bars represent**, entrez les variables pour lesquelles vous voulez obtenir les

moyennes; MEAN(var) apparaît pour chaque variable, comme par exemple la mean pour les 6 dimensions de l'ouverture.

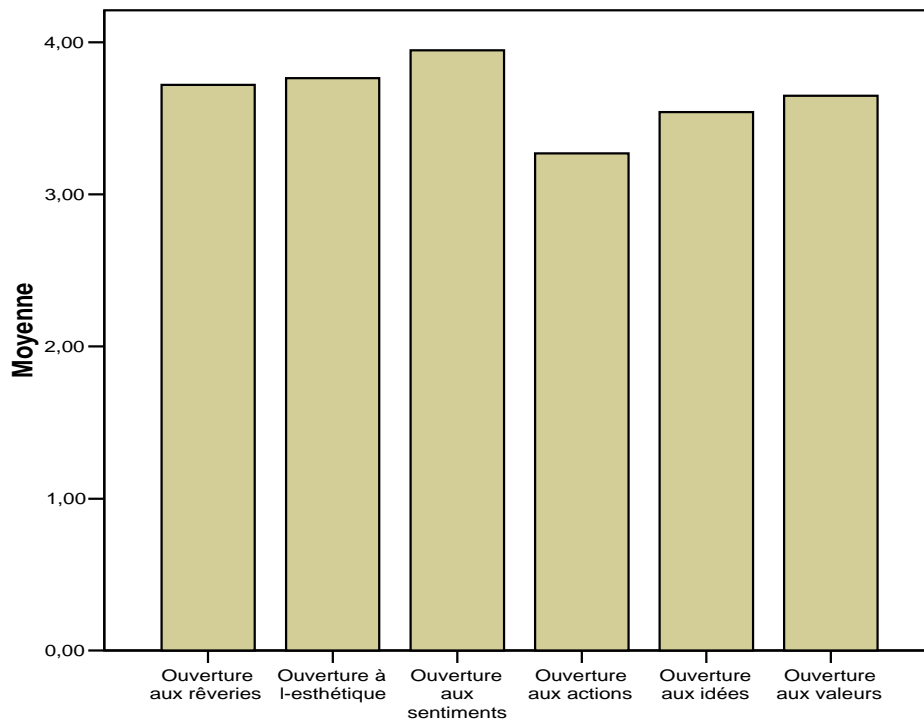


Figure 34

Présentation graphique de la moyenne de plusieurs variables métriques dans différents sous-groupes, pour cela cliquez sur **Graphs | Bar**, choisissez « **Clustered** » et « **Summaries of separate variables** ». Dans **Bars represent**, entrez les variables pour lesquelles vous voulez obtenir les moyennes; MEAN(var) apparaît pour chaque variable. Dans **Category axes**, entrez la variable de groupement, comme par exemple la moyenne de l'ouverture pour deux dimensions en fonction du sexe.

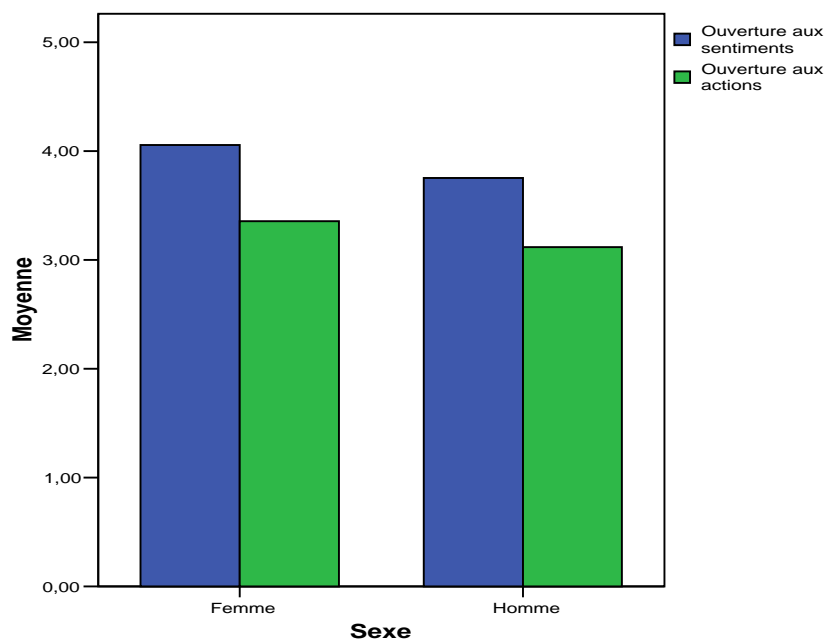


Figure 35



**Remarque** : la moyenne devrait être représentée par un point et non par une hauteur ou une surface comme c'est le cas dans les graphes bar, c'est pourquoi nous vous conseillons d'utiliser les graphes error bar (ci-après) qui sont corrects d'un point de vue strictement du sens de la mesure de la moyenne.

## **Error bar**

L'Error bar permet de présenter la moyenne et la variabilité de variables métriques (moyenne plus ou moins l'écart-type). Pour obtenir un graphique de la moyenne et la variabilité d'une variable métrique dans différents sous-groupes, sélectionnez **Graphs | Error Bar**. Choisissez « **Simple** » et « **Summaries for groups of cases** ». Dans **Variable**, choisissez la variable pour laquelle vous voulez obtenir les moyennes et leur variabilité. Dans **Category axes**, entrez la variable de groupement. Dans **Bars represent**, choisissez « **Standard deviation** » (écart type) [« **Confidence interval for mean** » et « **Standard error of mean** » ne sont pas traité dans ce cours] et multipliez par un. 1

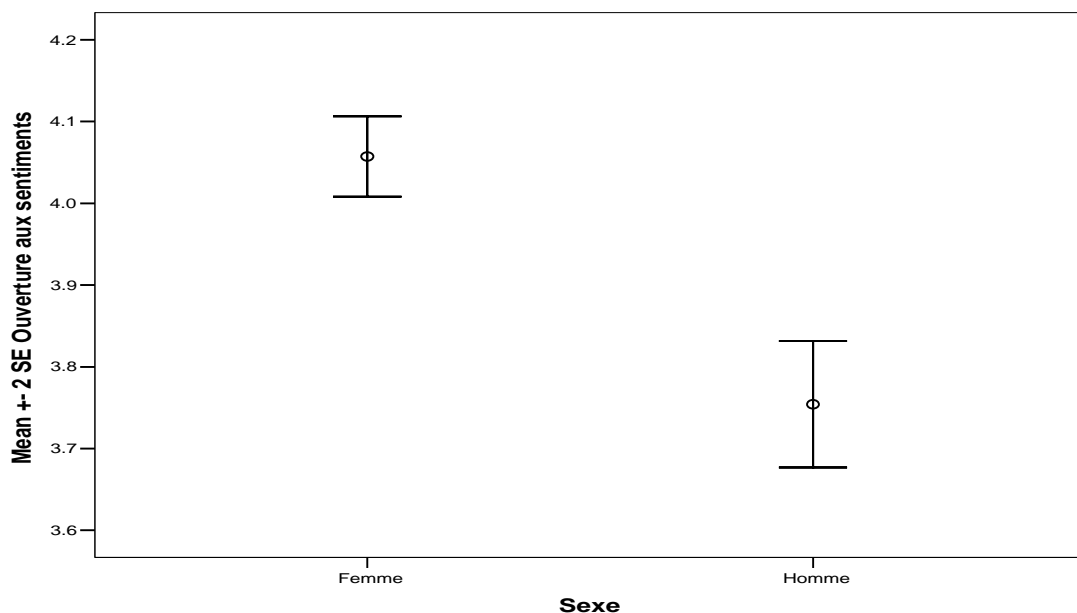


Figure 36

Pour obtenir un graphique de la moyenne et la variabilité de plusieurs variables métriques dans tout l'échantillon, sélectionnez **Graphs | Error Bar**. Choisissez « **Simple** » et « **Summaries of separate variables** ». Dans **Error Bars**, entrez les variables pour lesquelles vous voulez obtenir les moyennes et leur variabilité. Dans **Bars represent**, choisissez « **Standard deviation** » (écart type) comme par exemple les 6 dimensions de l'ouverture.

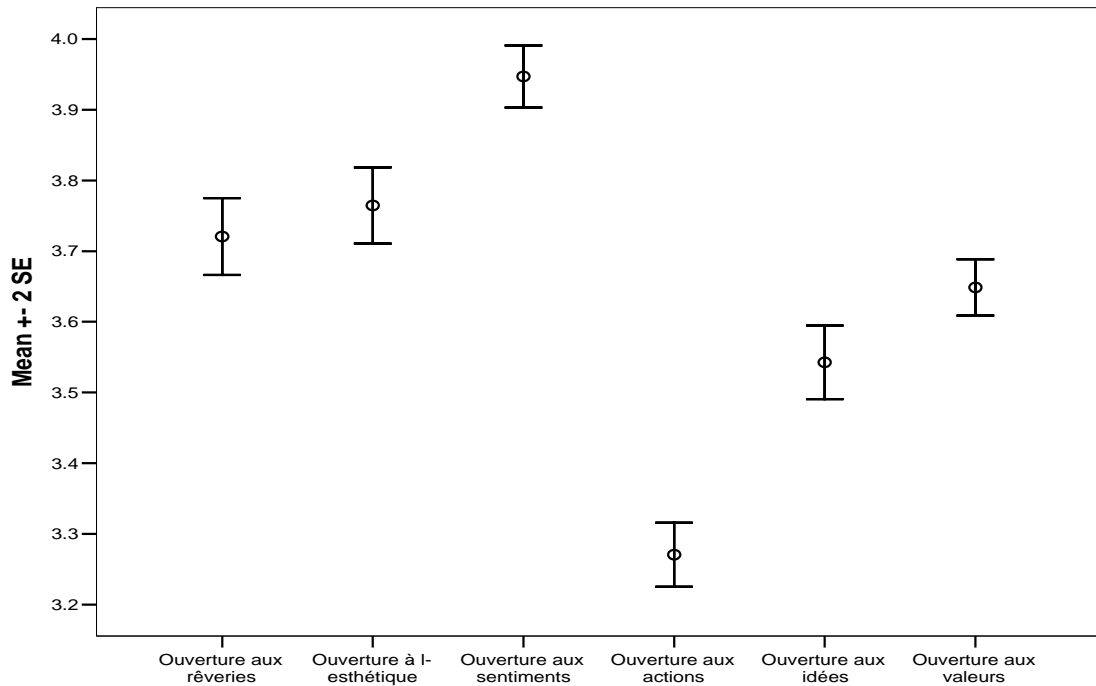


Figure 37

Pour obtenir un graphique de la moyenne et la variabilité de plusieurs variables métriques dans différents sous-groupes, sélectionnez **Graphs | Error Bar**. Choisissez « **Clustered** » et « **Summaries of separate variables** ». Dans **Error Bars**, entrer les variables pour lesquelles vous voulez obtenir les moyennes et leur variabilité, comme par exemple pour les variables ouverture aux sentiment et ouverture aux actions. Dans **Category axes**, entrez la variable de groupement, par exemple la variable sexe.

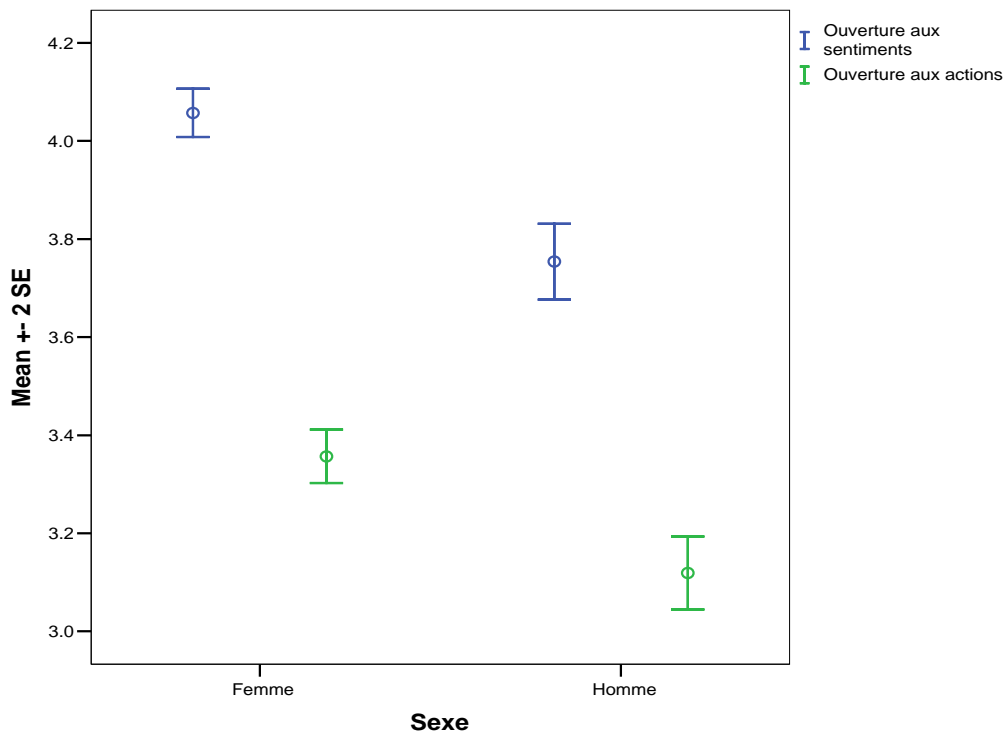


Figure 38

## Boxplot

Le Boxplot est un moyen de représenter graphiquement d'aspects de la distribution d'une variable, comme la médiane et la dispersion. La variable doit être au minimum ordinale :

- Boxplots pour une variable dans différents sous-groupes : sélectionnez **Graphs | Boxplot**. Choisissez « **Simple** » et « **Summaries for groups of cases** ». Dans **Variable**, entrez la variable pour laquelle vous voulez obtenir la médiane et la dispersion, par exemple ouverture aux sentiments. Dans **Category axes**, entrez la variable de groupement, par exemple le sexe. Dans **Label cases by**, si on a une variable d'identification des sujets, on peut l'y entrer pour identifier les valeurs extrêmes (quand on laisse cette option vide, SPSS utilise le numéro de ligne). Une alternative, c'est de passer par **Analyze | Descriptive Statistics | Explore**. Dans « **Dependent list** », entrez la variable pour laquelle vous voulez obtenir le boxplot. Dans « **Factor list** », entrez la variable de groupement.

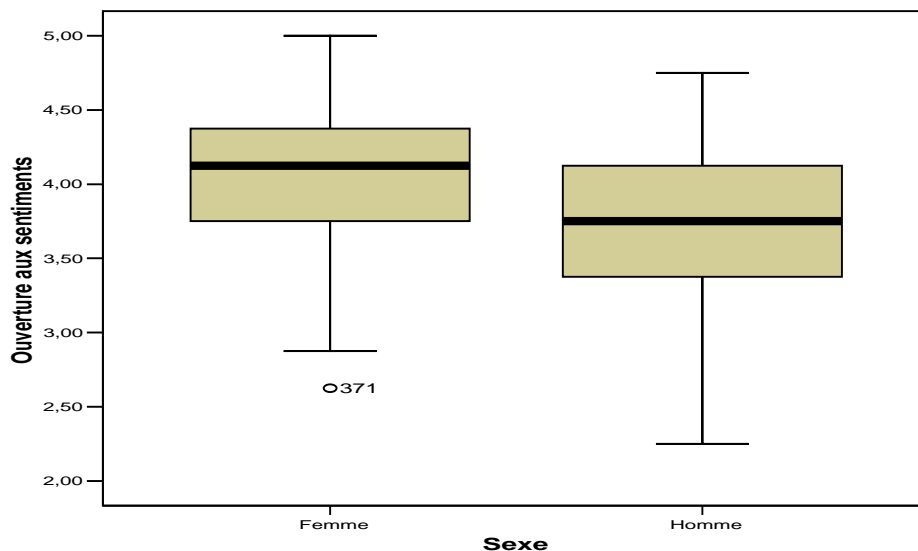


Figure 39

Comment lire le graphique : la hauteur de la boîte correspond à l'intervalle interquartile (IQ), le bord inférieur de la boîte représente le 1<sup>er</sup> et le bord supérieur de la boîte représente le 3<sup>ème</sup> quartile. Le trait traversant la boîte représente la médiane, le trait en dessous de la boîte relie le 1<sup>er</sup> quartile à l'extrême inférieur (c'est-à-dire l'observation égale ou juste supérieure à la valeur du 1<sup>er</sup> quartile moins 1.5 fois l'IQ). Le trait en dessus la boîte relie le 3<sup>ème</sup> quartile à l'extrême supérieur (c'est-à-dire l'observation égale ou juste inférieure à la valeur 3<sup>ème</sup> quartile plus 1.5 fois l'IQ) Les points sont des valeurs extrêmes qui se trouvent entre 1,5 et 3 fois l'IQ au-dessous ou au-dessus de la boîte. Les astérisques sont des valeurs extrêmes qui se trouvent plus de 3 fois l'IQ au-dessous ou au-dessus la boîte. Les chiffres à côté d'un point ou d'une astérisque sont le numéro du sujet.

- Boxplots pour plusieurs variables dans tout l'échantillon : cliquez sur **Graphs | Boxplot**, choisissez « **Simple** » et « **Summaries of separate variables** ». Dans **Boxes represent**, entrez les variables qui vous intéressent comme par exemple les 6 dimension de l'ouverture.

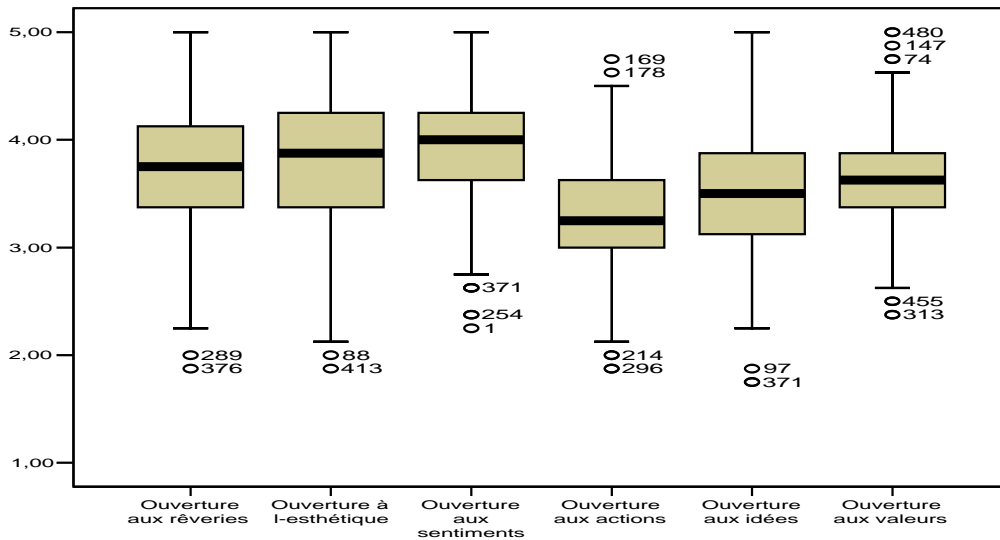


Figure 40

- Boxplots pour plusieurs variables dans différents sous-groupes : cliquez sur **Graphs | Boxplot**, choisissez « **Clustered** » et « **Summaries of separate variables** ». Dans **Boxes represent**, entrez les variables qui vous intéressent, par exemple les variables ouvertures aux sentiments et ouverture aux actions. Dans **Category axes**, entrez la variable de groupement, comme par exemple le sexe.

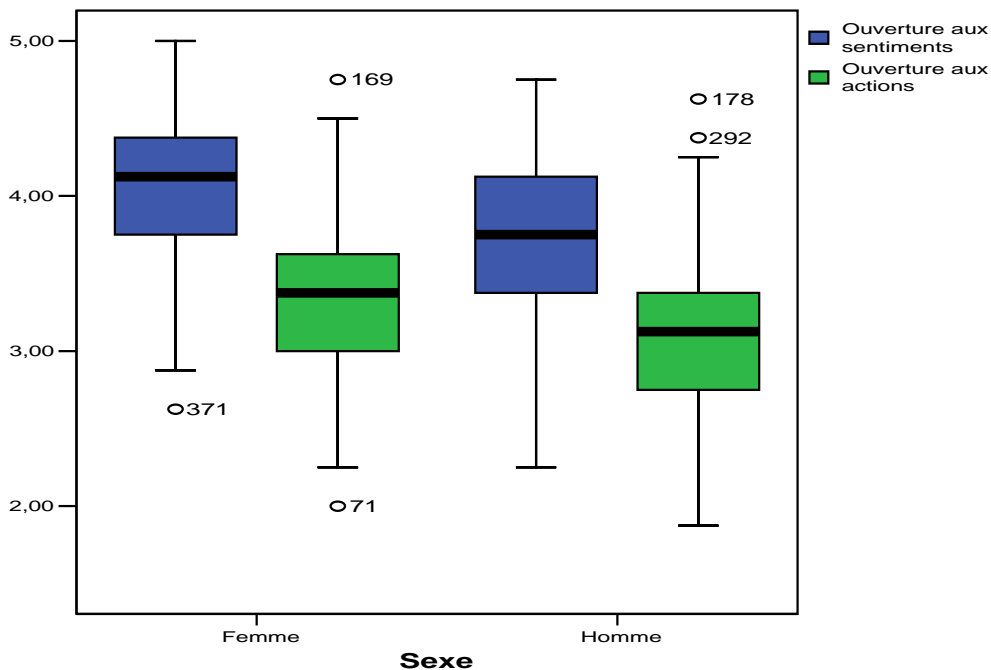


Figure 41

### Modifier la présentation des résultats (Output Labels)

Pour modifier la présentation des résultats, passez par le menu, cliquez dans **Edit | Options**. **Output Labels** permet de modifier la présentation des résultats. Dans la partie **Pivot Table Labeling**, on peut choisir entre (1) les étiquettes des variables (**labels**), (2) les noms des variables (**names**), (3) les noms des variables et leurs étiquettes (**names and labels**).

## Sélectionner des cas

La fonction **Data | Select cases** permet de sélectionner des cas en fonction des valeurs d'une variable (ou de plusieurs variables). C'est à dire qu'on peut limiter les analyses à un sous-groupe de l'échantillon (p.ex. les femmes, un pays, etc.). Pour cela passez par **Data | Select Cases**. Choisissez « **If condition is satisfied** », pressez le bouton « **If...** » et définissez la condition à l'aide de la variable, une valeur, ainsi que des opérateurs relationnels (égal « = », différent de « ≠ », plus petit que « < », plus grand que « > », plus petit ou égal « ≤ », plus grand ou égal « ≥ ») et des opérateurs logiques (et « & » / « **and** », ou « | » / « **or** »), puis pressez **Continue**. Pour **Unselected cases are...**, choisissez l'option « **Filtered** » ( par défaut). Attention : l'option « **deleted** » supprime tous les cas non-sélectionnés!!!

Dans la vue des données, les cas non-sélectionnés sont barrés (dans la première colonne). Dans le coin en bas à droite, SPSS nous informe que la fonction est active (« **Filter on** »). SPSS ajoute une nouvelle variable nommée « **filter\_\$** ». Les valeurs de cette variable filtre sont 1 (Le cas remplit la condition/est sélectionné) ou 0 (Le cas ne remplit pas la condition/n'est pas sélectionné). Une fois cette fonction en marche, seuls les cas sélectionnés (avec une valeur de 1 sur la variable filtre) seront utilisés dans les analyses subséquentes. Si on change la condition sous laquelle les cas sont sélectionnés, SPSS remplace la variable **filter\_\$**. Pour garder une variable filtre, on peut renommer filter\_\$ (p.ex. filt\_usa pour le sous-groupe de personnes des Etats-Unis). Attention : La fonction « **Select cases** » reste active jusqu'à ce qu'on la désactive!!! Pour la désactiver, il faut passer par **Data | Select Cases: All cases**.

## Comparer les groupes

La fonction **Data | Split File** permet de fragmenter un fichier et de comparer des cas en fonction des valeurs d'une variable (ou de plusieurs variables). C'est à dire qu'on peut analyser séparément des sous-groupes de l'échantillon afin de les comparer (p.ex. les femmes et les hommes, les Etats-Unis avec la Chine, etc.). Pour cela il faut passer par les menus déroulant et cliquer sur **Data | Split File**, puis choisir la variable de regroupement. L'option « **compare groups** » donne un tableau commun pour les sous-groupes et l'option « **Organize output by groups** » donne des tableaux séparés pour les sous-groupes.

Dans le coin en bas à droite, SPSS nous informe que la fonction est active (« **Split file on** »). Attention : la fonction « **Split file** » reste active jusqu'à ce qu'on la désactive!!! Pour la désactiver, il faut passer par **Data | Split File, Analyze all cases, do not create groups**.

# TP5 : Analyse des données : Mesures descriptives

## Exercice 7 : Distributions des variables

Ouvrez le fichier « data1.sav »

- Analysez la tendance centrale et la dispersion des variables « Big Five » : névrosisme [bf\_nev], extraversion [bf\_ext], intellect [bf\_int], agréabilité [bf\_agr] et conscience [bf\_con].
- Pour chaque variable, calculez la moyenne, la médiane, le mode, l'écart-type, l'étendue, le minimum et le maximum.

- Créez des histogrammes pour chaque variable.
- Sauvegardez tous les résultats (« output ») sous le nom « exercice7.spo » dans votre domaine (h:\...).

### Exercice 8 : Sélectionner et comparer les groupes

Ouvrez le fichier « data1.sav »

Pour le sous-groupe des allemands, analysez la distribution des « Big Five » de la personnalité (bf\_nev, bf\_ext, bf\_int, bf\_agr et bf\_con) en calculant, pour chaque variable,

- un tableau de fréquences
- la moyenne, la médiane, le mode, la variance, l'écart-type, l'intervalle interquartile, l'asymétrie, l'aplatissement

et en créant, pour chaque variable,

- un histogramme
- un boxplot
- error bar.

Pour toute la population, comparez les hommes et les femmes par rapport aux variables « Big Five » (bf\_nev, bf\_ext, bf\_int, bf\_agr et bf\_con).

- Pour les deux sous-groupes, calculez les moyennes et les écart-types des 5 variables.
- Pour chaque variable « Big Five », créez des boxplots qui représentent leur dispersion dans les deux sous-groupes, c'est-à-dire, créez des graphiques où les catégories de la variable sexe se trouvent sur l'axe des abscisses.

Sauvegardez tous les résultats (« output ») sous le nom « exercice8.spo » dans votre domaine (h:\...).

## TD6 : Analyse des données : Corrélation et régression

### Corrélation de rang

Pour calculer une corrélation de rang, allez sous **Analyze | Correlate | Bivariate**, cochez, dans **Correlation Coefficients**, **Spearman** et/ou **Kendall's tau-b**. Les variables qui sont entrées seront corrélées et on obtient une matrice complète, c'est à dire un tableau avec toutes les corrélations des variables deux à deux.

- Tout ce qui concerne la significativité: pas encore traité en première année
- Options: « **Missing values** » – important seulement si on analyse plusieurs variables en même temps.


### Scatterplot

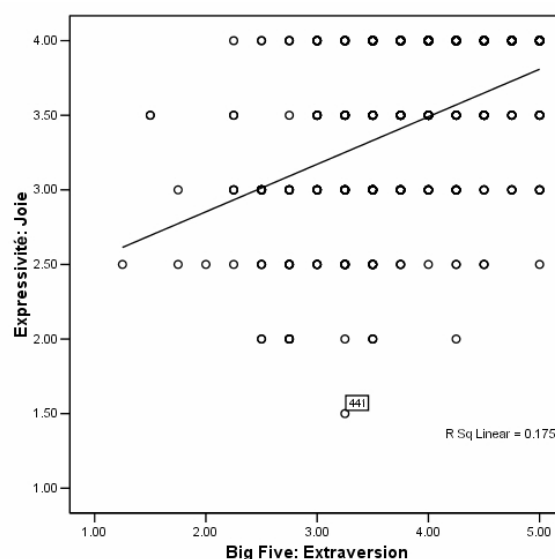
Le Scatterplot permet de présenter la relation entre deux variables métriques. Chaque point dans le graphique représente un cas, c'est à dire un couple de valeurs issu de deux variables. Pour obtenir un scatterplot, passez par **Graphs | Scatter: « Simple »**. Pour cela :

- Choisir les variables : **Y-Axis** pour l'axe des ordonnées et **X-Axis** pour l'axe des abscisses
- **Set markers by** : on peut entrer une variable de groupement (p.ex. le sexe) – les différents sous-groupes seront affichés par des couleurs différentes. Attention : un point dans le Scatterplot peut représenter plusieurs cas – la couleur affichée correspond au sous-groupe du premier cas! L'utilisation de cette option n'est pas recommandée!
- **Label cases by** : on peut entrer la variable ID (numérotation des cas) pour obtenir le numéro d'un cas individuel dans l'éditeur de graphique – quand on n'entre pas de variable ici, SPSS prend la numérotation des lignes (qui ne correspond pas forcément à la variable ID!)
- **Options**: Après avoir entré une variable ID dans « **label cases by** », on peut cocher l'option « **Display chart with case labels** » (en cliquant sur « **Options** »). Ainsi, tous les points seront marqués par leurs numéros, le désavantage, c'est que ce n'est pas très lisible.

Ex : Extraversion et l'expressivité de joie

Pour obtenir le numéro d'un cas individuel dans l'éditeur de graphique :

- double-cliquez sur le graphique pour activer l'éditeur
- Cliquez sur le symbol 



- Cliquez sur le point dont on veut obtenir le numéro
- Cliquez encore une fois au-dessus pour effacer le numéro

D'autres options dans l'éditeur de graphique :

- Pour que le nombre de cas représentés soit indiqué par la taille des points ou par l'intensité de leur couleur, double-cliquez sur les points de données, puis choisissez l'onglet « **Point bins** » et cliquez sur « **Bins** » (défaut : « **marker size** »)
- Pour obtenir la droite de régression, sélectionnez les points de données, après cliquez sur **Chart | Add chart element: « Fit line at total »**

### Corrélation de Bravais-Pearson

Pour calculer une corrélation pour des variables numériques, allez sous **Analyze | Correlate | Bivariate**, le coefficient de corrélation de Pearson est le coefficient par défaut. Les variables qui sont entrées seront corrélées et on obtient une matrice complète. Pour obtenir la moyenne et l'écart-type pour chaque variable (« **means and standard deviations** ») ainsi que les produits croisés et les covariances (« **cross-product deviations and covariances** ») allez sous **Options**. Sous **Options**, vous trouverez également les « **missing values** », cette fonction est importante uniquement si on analyse plusieurs variables en même temps. Si on choisit « **Exclude cases pairwise** » pour chaque coefficient de corrélation, les cas qui n'ont pas de valeurs valables sur le couple de variables seront exclus. Dans le cas où on choisit « **Exclude cases listwise** » alors pour chaque coefficient de corrélation, seulement les cas qui ont des valeurs valables sur toutes les variables seront analysés (même N pour chaque coefficient).

- Tout ce qui concerne la significativité: pas encore traité en première année

Passer par la syntaxe: parfois, on veut obtenir des corrélations entre deux groupes de variables (p.ex., les « Big Five » et les différentes variables de l'expressivité émotionnelle). Dans ce cas, le tableau standard (obtenu par le menu **Correlate**) devient très grand et il y aura des corrélations dont on n'a pas besoin (p.ex. entre les variables « Big Five »). Pour obtenir seulement les corrélations entre les deux groupes, il faut utiliser la syntaxe :

- Dans la boîte-dialogue, entrer toutes les variables et cliquer sur **Paste** (au lieu de **OK**)
- Dans la fenêtre de syntaxe, il faut ajouter le mot **WITH** entre les deux groupes de variables. Exemple:

```
CORRELATIONS
/VARIABLES=bf_ext bf_agr bf_con bf_nev bf_int WITH exp_amo
exp_joi exp_peu exp_col exp_tri
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE .
```



### Correlations

		exp_amo Expressivité: Amour	exp_joi Expressivité: Joie	exp_peu Expressivité: Peur	exp_col Expressivité: Colère	exp_tri Expressivité: Tristesse
bf_ext Big Five: Extraversion	Pearson Correlation	.348**	.418**	.079	.062	.114*
	Sig. (2-tailed)	.000	.000	.085	.175	.013
	N	481	481	481	478	480
bf_agr Big Five: Agréabilité	Pearson Correlation	.245**	.224**	.120**	-.140**	.132**
	Sig. (2-tailed)	.000	.000	.008	.002	.004
	N	482	482	482	479	481
bf_con Big Five: Conscience	Pearson Correlation	.077	.048	.071	-.050	.065
	Sig. (2-tailed)	.094	.291	.119	.280	.157
	N	481	481	481	478	480
bf_nev Big Five: Nevrocisme	Pearson Correlation	-.042	-.055	.141**	.162**	.241**
	Sig. (2-tailed)	.352	.232	.002	.000	.000
	N	481	481	481	478	480
bf_int Big Five: Intelligence	Pearson Correlation	.145**	.094*	.036	.088	.086
	Sig. (2-tailed)	.001	.040	.435	.055	.059
	N	481	481	481	478	480

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

### Régression linéaire simple

Pour effectuer une régression linéaire simple, allez sous **Analyze | Regression | Linear** :

- « **Dependent** » : Entrez la variable Y (variable expliquée), par exemple l'ouverture aux sentiments.
- « **Independent** » : Entrez la variable X (variable explicative), par exemple l'extraversion.
- **Statistics**: on peut cocher **descriptives** pour obtenir des statistiques descriptives pour les variables (moyenne, écart-type, taille de l'échantillon, corrélation des variables)
- Pour la régression simple (avec une variable explicative X), il ne faut pas d'autres options!

### Syntaxe :

```
REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT ouv_sen
/METHOD=ENTER bf_ext .
```

### SPSS-Viewer (Résultat)

« Model summary » :

- R : correspond à la corrélation  $r_{xy}$  dans le cas de la régression simple
- $R^2$  : coefficient de détermination ou pourcentage de la variation de Y expliquée par la régression

### Récapitulatif du modèle

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,325 <sup>a</sup>	,106	,104	,45768

a. Valeurs prédites : (constantes), bf\_ext Big Five: Extraversion

Tableau « ANOVA »: sommes des carrés de la régression

### ANOVA<sup>b</sup>

Modèle		Somme des carrés	ddl	Carré moyen	F	Signification
1	Régression	11,881	1	11,881	56,720	,000 <sup>a</sup>
	Résidu	100,339	479	,209		
	Total	112,220	480			

a. Valeurs prédites : (constantes), bf\_ext Big Five: Extraversion

b. Variable dépendante : ouv\_sen Ouverture aux sentiments

Tableau « Coefficients »

- Ligne « (Constant) » et colonne « B » : l'ordonnée à l'origine de la droite de régression
- Ligne de la variable et colonne « B » (valeur non-standardisée) : pente de la droite de régression
- colonne « Beta » (valeur standardisée) : correspond à la corrélation  $r_{xy}$  dans le cas de la régression simple
- « Standard error » et les valeurs t et sig. concernent la significativité des coefficients: pas encore traité en première année!

### Coefficients<sup>a</sup>

Modèle		Coefficients non standardisés		Coefficients standardisés	t	Signification
		B	Erreur standard	Bêta		
1	(constante)	3,139	,109		28,676	,000
	bf_ext Big Five: Extraversion	,214	,028	,325	7,531	,000

a. Variable dépendante : ouv\_sen Ouverture aux sentiments

## TP6 : Analyse des données : Corrélation et régression

### Exercice 9 corrélation de rang

Ouvrez le fichier « data1.sav » .

- Analysez la corrélation entre les variables d'expressions par dimension.

Sauvegardez tous les résultats (« output ») sous le nom « exercice9.spo » dans votre domaine (h:\...).

### Exercice 10 corrélation et régression linéaire simple

Ouvrez le fichier « data1.sav » .

Recoder la variable âge en quatre groupes contenant chacun 25% de la population. Puis pour chaque groupe d'âge séparément, analysez la relation entre la fréquence expression de la peur et la fréquence d'expression de la colère :

- Créez un scatterplot et rajoutez une droite de régression.
- Analysez la corrélation entre les deux variables.
- Analysez la régression linéaire en utilisant la fréquence d'expression de peur comme variable X (variable explicative) et la fréquence d'expression de colère comme variable Y (variable expliquée).

Sauvegardez tous les résultats (« output ») sous le nom « exercice10.spo » dans votre domaine (h:\...).





**FACULTÉ DE PSYCHOLOGIE ET DES  
SCIENCES DE L'ÉDUCATION**

SECTION DE PSYCHOLOGIE  
UNI MAIL - 40 bd du Pont-d'Arve | CH-1205 Genève

**Prof. G.R. Pini et O. Renaud**

Ass. : K. Iglesias

**Cour « Introduction à l'analyse exploratoire des données »**

**Semestre d'été 2006**

QUESTIONNAIRE

Ci-après, vous trouverez une liste d'émotions. Svp, indiquez pour chaque émotion à quelle fréquence vous **exprimez** cette émotion à d'autres personnes quand vous la ressentez.

	jamais ou presque jamais	par- fois	sou- vent	presque toujours ou toujours
1 Affection	1	2	3	4
2 Joie	1	2	3	4
3 Peur	1	2	3	4
4 Colère	1	2	3	4
5 Honte	1	2	3	4
6 Tristesse	1	2	3	4
7 Amour	1	2	3	4
8 Bonheur	1	2	3	4
9 Inquiétude	1	2	3	4
10 Irritation	1	2	3	4
11 Culpabilité	1	2	3	4
12 Depression	1	2	3	4
	jamais ou presque jamais	par- fois	sou- vent	presque toujours ou toujours

Ci-après, vous trouverez une liste de traits de caractère. Pour chaque mot, veuillez indiquer la réponse qui vous semble le mieux décrire votre personnalité, svp.

**Je suis...**

	pas du tout					tout à fait	
	1	2	3	4	5		
amical(e)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
chaleureux(se)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
travailleur(se)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
susceptible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
malin (maligne)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
sociable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
assidu(e)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
plein(e) d'égards	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
intelligent(e)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
sensible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
	pas du tout					tout à fait	

**Je suis...**

	pas du tout					tout à fait	
	1	2	3	4	5		
conscientieux (se)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
lunatique	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
vif (vive)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
cultivé(e)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
d'un bon naturel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
plein(e) de tempérament	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
serviable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
riche d'esprit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
pas sûr(e) de moi	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
ambitieux(se)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
	pas du tout					tout à fait	

Les questions qui suivent portent sur un aspect de votre personnalité. Pour chaque question, veuillez indiquer la réponse qui vous semble la plus appropriée. Évaluez les affirmations en utilisant l'échelle d'approbation.

		pas du tout d'accord	un peu d'accord	moyenne- ment d'accord	plutôt d'accord	tout à fait d'accord
1	J'ai une imagination très active.	1	2	3	4	5
2	Je suis bien installé(e) dans mes habitudes.	1	2	3	4	5
3	Je crois que laisser les étudiants écouter des orateurs controversés ne peut que les embrouiller et les égarer.	1	2	3	4	5
4	J'essaie de maintenir toutes mes pensées dans une direction réaliste et d'éviter les envolées de l'imagination.	1	2	3	4	5
5	Il m'arrive quelquefois de m'absorber complètement dans la musique que j'écoute.	1	2	3	4	5
6	Je trouve les discussions philosophiques ennuyeuses.	1	2	3	4	5
7	Assister à des spectacles de ballet ou de danse moderne m'ennuie.	1	2	3	4	5
8	La manière dont je sens les choses est importante pour moi.	1	2	3	4	5
9	Une fois que j'ai trouvé la bonne manière de faire quelque chose, je m'y tiens.	1	2	3	4	5
10	Je prends plaisir à résoudre des problèmes ou des énigmes.	1	2	3	4	5
11	Je n'aime pas perdre mon temps à rêvasser.	1	2	3	4	5
12	Je crois que nous devrions nous tourner vers nos autorités religieuses pour les décisions concernant les questions morales.	1	2	3	4	5
13	Je suis intrigué(e) par les formes et les motifs que je trouve dans l'art et dans la nature.	1	2	3	4	5
14	Il est rare que j'accorde beaucoup d'attention à mes sentiments du moment.	1	2	3	4	5
15	J'essaie souvent des plats nouveaux et exotiques.	1	2	3	4	5
16	Je perds quelquefois tout intérêt quand les gens parlent de sujets très abstraits et théoriques.	1	2	3	4	5
17	Je prends plaisir à me concentrer sur une image intérieure ou une rêverie et à explorer toutes ses possibilités en les laissant croître et se développer.	1	2	3	4	5
18	La poésie a peu ou pas d'effet sur moi.	1	2	3	4	5
19	J'éprouve une grande variété d'émotions ou de sentiments.	1	2	3	4	5
20	Je préfère passer mon temps dans un entourage familial.	1	2	3	4	5

21	Je crois que la fidélité à ses propres idéaux et principes est plus importante que « l'ouverture d'esprit ».	1	2	3	4	5
22	Si je sens que mon esprit commence à dériver vers des rêveries, j'ai l'habitude de m'occuper et de me mettre à me concentrer sur un travail ou une activité.	1	2	3	4	5
23	Certains genres de musique exercent sur moi une fascination sans fin.	1	2	3	4	5
24	Je remarque rarement les changements d'humeurs ou les sentiments que provoquent des environnements différents.	1	2	3	4	5

		pas du tout d'accord	un peu d'accord	moyenne-ment d'accord	plutôt d'accord	tout à fait d'accord
25	J'ai peu d'intérêt pour la réflexion sur la nature de l'univers ou sur la condition humaine.	1	2	3	4	5
26	Je me trouve large d'esprit et tolérant(e) pour les façons de vivre des autres.	1	2	3	4	5
27	Je trouve facile d'avoir de l'empathie, c'est-à-dire de ressentir moi-même ce que les autres ressentent.	1	2	3	4	5
28	En vacances, je préfère retourner dans un endroit que je connais bien plutôt qu'aller dans un nouvel endroit.	1	2	3	4	5
29	J'ai beaucoup de curiosité intellectuelle.	1	2	3	4	5
30	Je pense que si les gens ne savent pas à quoi ils croient quand ils ont vingt-cinq ans, il y a quelque chose qui ne tourne pas rond chez eux.	1	2	3	4	5

		Pas satisfait	Plutôt pas satisfait	Plutôt satisfait	Satisfait
1	Etes-vous satisfait avec la manière dont vous gérez vos émotions ?	1	2	3	4

### Quelles loisirs avez-vous régulièrement?

(plusieurs réponses sont possibles)

- Sport
- Rencontrer des amis
- Activités culturelles (cinéma, expositions, concerts, etc.)
- Animal domestique
- Instrument de musique
- Télévision
- Internet
- Bricolage (tricot, jardinage, etc.)
- Autre : \_\_\_\_\_



**Qu'est ce que ces loisirs vous apportent ?**

---

---

---

---

**Variables socio-démographiques :**

**Sexe :**         Femme                     Homme

**Age :** \_\_\_\_\_ ans

**Taille :** \_\_\_\_\_ cm

**Poids :** \_\_\_\_\_ kg

**Nationalité :**  Suisse                     autre: \_\_\_\_\_

Vous êtes étudiant(e) dans quel domaine? \_\_\_\_\_

***Votre code:***

Première lettre du prénom de votre mère:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Dernière lettre du prénom de votre mère:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Première lettre du prénom de votre père:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Dernière lettre du prénom de votre père:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Première lettre de votre lieu de naissance:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Dernière lettre de votre lieu de naissance:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---