

ANALYSE DES DONNÉES

MASTER PRO SOCIALE

J.-B. Légal / A. N'Gbala

2007-2008

Les Objectifs

- Vous permettre de :
 - Déterminer quel type d'analyse réaliser en fonction de la situation
 - Tester statistiquement vos hypothèses
 - (Eventuellement) Vous réconcilier un peu avec les Stats...

Les Moyens

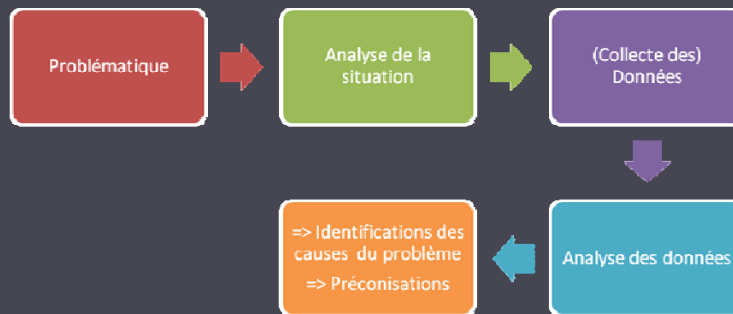
- Approche pragmatique (« *cookbook* »)
- Travail sur logiciel (SPSS)
- Données « réelles » ou adaptées à la pratique professionnelle

Déroulement des sessions

- Pour chaque type d'analyse :
 - ▣ Principe général.
 - ▣ Conditions d'applications : Quand ? Quel type de données ? Pourquoi cette analyse (et ce que l'on va pouvoir en tirer).
 - ▣ Exemple « tutoré » sur SPSS.
 - ▣ Exercice « de consolidation » en autonomie.
 - ▣ Exercice plus poussé posé sous la forme d'une problématique à résoudre + une feuille de données.

Validation

□ Cas pratique :



Programme des séances

- Section 1 : Introduction
 - Présentation du module
 - Rappels & Arbres de décision
 - Prise en main de SPSS
- Section 2 : Travail préalable sur les données
 - Exploration des données
 - Gestion des données extrêmes
 - Fiabilité d'échelle
- Section 3 : à la recherche des différences
 - Chi²
 - T de Student
 - Anova
- Section 4 : Prédire
 - Régression logistique
 - Régressions linéaires simples et multiples
 - Médiation
- Section 5 : Résumer et simplifier des données
 - Analyses factorielles

SECTION 1

Rappels et découverte de SPSS

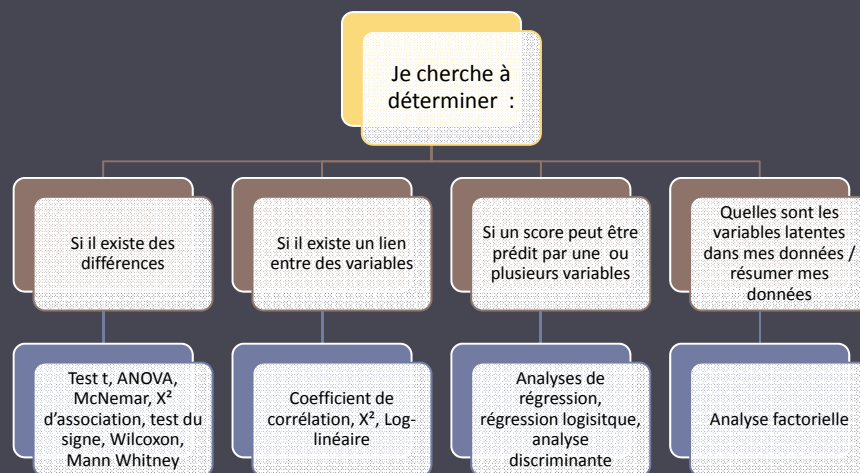
Vocabulaire (rappels)

- Variable:
 - Indépendante (VI) : caractéristique de la personne, de l'environnement, de la situation ou d'un objet qui peut prendre différentes valeurs.
 - Inter : groupes différents assignés aux différents niveaux de la Var
 - Intra : un même groupe passe par tous les niveaux de la Var
 - Dépendante (VD) : la mesure
- Types de variables:
 - Quantitative : peut être mesurée selon une échelle continue (ex : le poids)
 - Qualitative / Catégorielle : modalités non quantifiables (ex : beau / moche)
 - Invoquée : modalités non modifiable par la volonté de l'expérimentateur (ex : sexe des Ps)
 - Manipulée : modalités déterminées par l'expérimentateur (ex : la couleur des murs)
- Types de données
 - Intervalles (ou métriques) : mesures continues (graduations séparées par une même unité ; ex : masse en kg)
 - Ordinales : catégories ordonnées (ex : petit / moyen / grand)
 - Nominale (ou catégorielles) : catégories, attributs ou groupes (ex : homme / femme / enfant)

Vocabulaire (rappels)

- Recherche :
 - Expérimentale : on cherche à mettre en évidence une relation de cause à effet en manipulant une VI
 - Corrélacionnelle : on cherche à mettre en évidence un lien entre des mesures que l'on considère telles qu'elles existent
- Plan :
 - Expérimental : plan contenant uniquement des VI manipulées
 - Quasi-expérimental : contenant au moins une VI invoquée
 - Intra-sujets (*within*) : plan à groupes de mesures liés / appariés (voir VI intra)
 - Inter-sujets (*between*) : plan à groupes de mesures indépendants (voir VI inter)
 - Mixte (*split plot*) : plan contenant au moins une VI inter et une VI intra

Que cherche-t-on à mettre en évidence ?



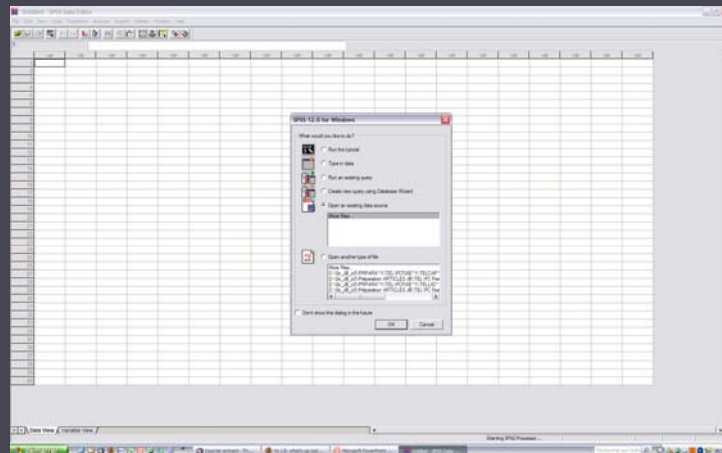
Arbres de décision

- Dans quel cas (quel type de données, de variables, de plan, etc.) peut-on / doit-on appliquer telle ou telle procédure ?
 - => arbres de décision

- ~~/\~~ les arbres de décision constituent une aide certaine, mais le choix d'une analyse plutôt qu'une autre doit être guidée par les objectifs de recherches que l'on s'est fixé.

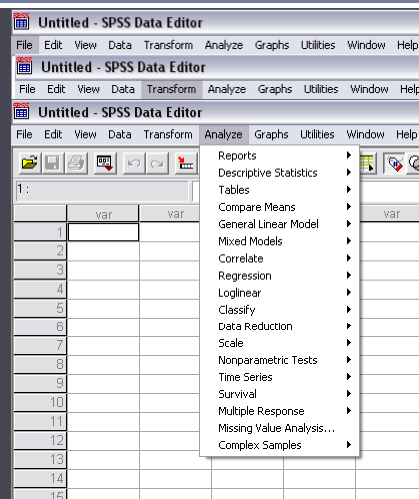
- Doc « arbres »

SPSS : interface générale



SPSS : menus d'intérêt

- ▣ File
- ▣ Transform
- ▣ Analyse



Prise en main de SPSS

- Interface
- Saisie de données
- Importation de données
- Affichage Données / Variables
- Opérations sur les variables
- Ex 1 : entrer des données pour :
 - ▣ une VI inter-sujet
 - ▣ une VI intra-sujets
 - ▣ un plan mixte
- Ex 2 : et pour un plan mixte à plusieurs VI dont 2 intra ?
- Ex. 3 : opérations sur les variables
 - ▣ recodage/recalcul
 - ▣ filtres

Ex. 1 : entrer des données pour...

- Sachant que : le lundi Paul boit 5 cafés, Ginette 1, Pierre 9, Britney 0, Catherine 3, Jacques 7, Toto 2.
- VI : Genre (avec 1 pour Homme ; 2 pour Femme)
- VD : nombre de cafés bus dans la journée de lundi

Ex. 1 : suite...

- Sur la base du fichier qui vient d'être créé, ajouter :
 - ▣ VI : jour de la semaine de lundi à vendredi
 - ▣ VD : nombre de cafés bus

- Q ? : comment entreriez-vous ces données ?

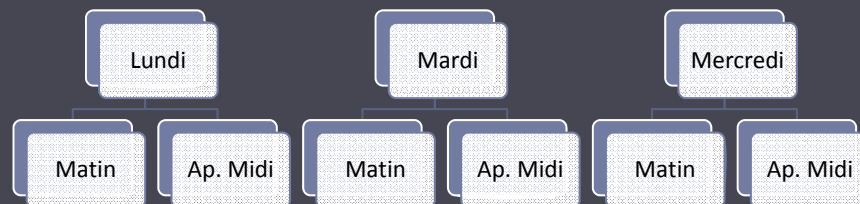
Prénom	Paul	Ginette	Pierre	Britney	Catherine	Jacques	Toto
Genre	1	2	1	2	2	1	1
Lundi	5	1	9	0	3	7	2
Mardi	4	1	1	0	3	5	2
Mercredi	2	1	7	0	1	6	1
Jeudi	4	1	2	0	4	5	0
Vendredi	0	1	11	0	2	8	3

Ex. 2

- Sur la base du fichier de l'exercice 1, comment feriez-vous pour ajouter :
 - ▣ Le moment de la journée où est pris le café en matin vs. après midi ?
 - ▣ Rq : ici VI intra sujets

SPSS

Dans le cas de plusieurs VI intra, un arbre peut aider à représenter les conditions



Ex. 3

- Calcul sur valeurs et création d'une nouvelle variable
- A partir des données de l'exercice 1, créer une nouvelle variable nommée « Somcaf » correspondant au nombre total de cafés bus dans la semaine.

Ex. 3 suite

- Recodage d'une variable continue en variable catégorielle
 - ▣ Recoder la variable Somcaf dans une nouvelle variable « ConsoHeb » à 2 modalités faible / élevée selon une coupe à la médiane
 - Note : la médiane est ici de 13.

Ex. 3 fin

- Utilisation des filtres (sélection de conditions particulières)
 - ▣ Sélectionner les femmes uniquement

SECTION 2

Exploration des données et Fiabilité d'échelle

Au programme :

- Exploration des données
 - ▣ Paramètres utiles

- Identification et gestion des données extrêmes (« *outliers* »)
 - ▣ Diagrammes en tiges et feuilles (« *Stem & leaf* »)
 - ▣ Boîtes à moustaches

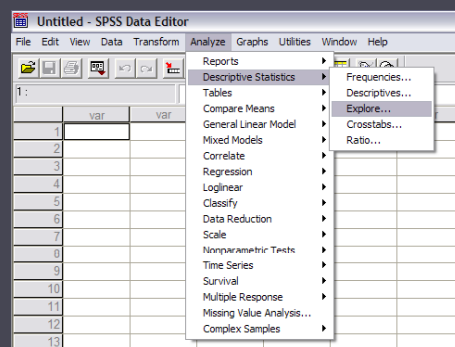
- Fiabilité d'échelle
 - ▣ Alpha de Cronbach

Exploration des données

- Avant de se lancer dans les analyses il est important de vérifier :
 - Que les données ne contiennent pas de réponses aberrantes
 - ex: la réponse de réponses 7 sur une échelle en 5 points
 - Que des valeurs extrêmes (« *outliers* ») ne viennent pas compromettre la validité des analyses.
 - Que les éventuelles valeurs manquantes soient bien traitées comme telles.
 - Que les conditions d'application de l'analyse choisie sont vérifiées.

Exploration des données

- L'option « *Explore* »



Exploration des données

□ Observations

sexe

Case Processing Summary

		Cases					
		Valid		Missing		Total	
Scorestress	sexe	N	Percent	N	Percent	N	Percent
0	0	109	100,0%	0	,0%	109	100,0%
1	1	130	99,2%	1	,8%	131	100,0%

Exploration des données

□ Statistiques descriptives

Moyenne

Médiane

Ecart type

Skewness (asymétrie)

Kurtosis (aplatissement)

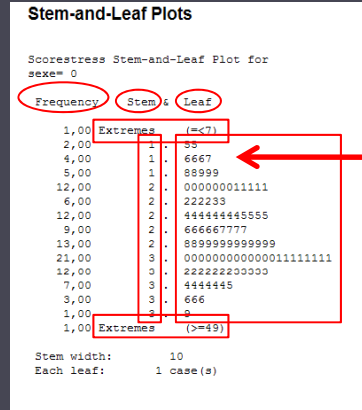
=> Shapiro-Wilks

Descriptives

sexe	Statistic	Std. Error	
Scorestress 0	Mean	26,98	
	95% Confidence Interval for Mean	25,81	
	Lower Bound		
	Upper Bound	28,16	
	5% Trimmed Mean	27,04	
	Median	29,00	
	Variance	38,222	
	Std. Deviation	6,182	
	Minimum	7	
	Maximum	49	
	Range	42	
	Interquartile Range	9	
	Skewness	-,141	,231
	Kurtosis	1,034	,459
1	Mean	28,25	
	95% Confidence Interval for Mean	27,13	
	Lower Bound		
	Upper Bound	29,38	
	5% Trimmed Mean	28,34	
	Median	28,00	
	Variance	42,113	
	Std. Deviation	6,489	
	Minimum	1	
	Maximum	55	
	Range	54	
	Interquartile Range	7	
	Skewness	-,405	,212
	Kurtosis	4,758	,422

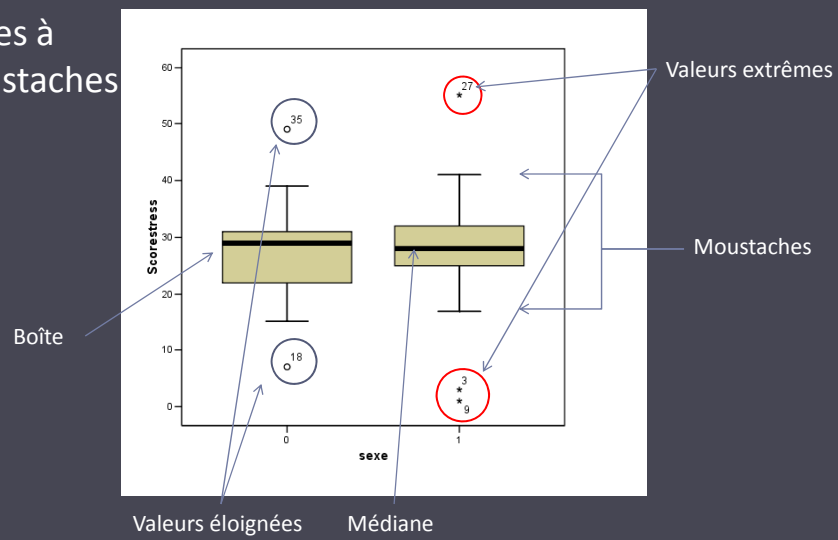
Exploration des données

- Graphs en Tige et Feuilles



Exploration des données

- Boîtes à moustaches



J'ai des *outliers* et/ou une distribution non normale : je fais quoi ?

- Plusieurs solutions :
 - Ne rien faire
 - Solution de facilité
 - Met en péril les analyses et les conclusions
 - Eliminer ou exclure les *outliers* de l'analyse
 - Solution aisée
 - Peu gênant si le nombre de sujets de l'étude est important
 - Améliore la distribution
 - Appliquer une transformation aux données
 - Corrige la distribution
 - « Normalisation » des *outliers*
 - Nouvelles valeurs peuvent être perçues comme sans correspondance « avec le réel »

Transformations

- Quelles transformations et pourquoi :

Problème	Transformation(s)	Effet(s)	Remarque
<i>Outliers</i>	Racine carrée ou Log. ou Inverse	« Réintègre » les <i>outliers</i>	Inverse > Log. > Racine
Asymétrie à droite (positive)	Racine carrée ou Log. ou Inverse	Normalise la distribution	
Asymétrie à gauche (négative)	1) $x - (\text{valeur la plus élevée} + 1)$ 2) Racine carré ou Log. ou Inverse	Normalise la distribution	